ESSAYS ON THE TERM STRUCTURE OF INTEREST RATES

A Dissertation

Presented to

The Faculty of the C.T. Bauer College of Business

University of Houston

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

by

Rui Liu

May, 2017

ACKNOWLEDGMENTS

First and foremost I would like to express my sincere gratitude to my advisor Dr. Kris Jacobs. It has been a great honor to be his Ph.D. student. He has taught me how good finance research is done. I appreciate his guidance, understanding, patience, and most importantly, his mentorship to make my Ph.D. experience productive and stimulating. The insight and enthusiasm he has for his research was contagious and motivational for me, even during tough times in the Ph.D. pursuit. Without his continuous encouragement and support, I would not be able to complete this dissertation.

My sincere gratitude also goes to Dr. Hitesh Doshi, who was always willing to help and give his best suggestions. He has enlightened me the first glance of research and incented me to develop my research from various perspectives. My research would not have been possible without his precious support.

My gratitude is also extended to the rest of my dissertation committee: Dr. Rauli Susmel, Dr. Sang Byung Seo, and Dr. Tong Lu. Without their assistance and dedicated involvement in every step throughout the process, this dissertation would not have been accomplished.

I would like to acknowledge the Department of Finance, C. T. Bauer College of Business. My graduate experience benefited greatly from the courses I took, the opportunities to serve as an instructor, and the high-quality seminars that the department organized. I thank all of the department faculty members for providing me with the foundation for becoming an academic researcher. Special thanks to Dr. Praveen Kumar, Dr. Thomas George, Dr. Stuart Turnbull, Dr. Vijay Yerramilli, Dr. Paul Povel, and Dr. James Yae.

Getting through my dissertation required more than academic support, I have been blessed with a friendly and cheerful group of fellow Ph.D. students. I am grateful for their friendship, discussion, encouragement, and companion.

Last but not the least, I owe an enormous debt to my parents for their love, support and sacrifices during this long journey. This dissertation is dedicated to them.

ESSAYS ON THE TERM STRUCTURE OF INTEREST RATES

Abstract

Rui Liu

May, 2017

This dissertation consists of three essays on the term structure of interest rates. In the first essay, I provide evidence on the existence of unspanned macro risk. I investigate the usefulness of unspanned macro information for forecasting bond risk premia in a macro-finance term structure model from the perspective of a bond investor. I account for model uncertainty by combining forecasts with and without unspanned output and inflation risks optimally from the forecaster's objective, and I take advantage of the no-arbitrage condition by imposing risk premium restrictions for the purpose of forecasting. Incorporating macro information generates significant gains in forecasting bond risk premia relative to yield curve information at long forecast horizons, especially when allowing for time-varying combination weight. These gains in predictive accuracy significantly improve investor utility.

Cochrane and Piazzesi (2005) and Duffee (2011a) find that information not captured by traditional term structure factors helps predict excess bond returns. The second essay shows that when estimating no-arbitrage affine term structure models, aligning in-sample and out-of-sample objective functions results in term structure factors that capture similar information that remains hidden from existing approaches. Specifically, the estimates of the third term structure factor radically differ. Consistent with Cochrane and Piazzesi (2005), this factor confirms the importance of the fourth principal component of yields for forecasting the term structure. The new objective function leads to substantial improvements in forecasting performance. Model term premiums are higher and expected future short rates are lower.

The third essay proposes a no-arbitrage term structure model with a Taylor rule and two macroeconomic variables, real activity growth and inflation, that each contain long-run and short-run components. Variance decompositions and impulse responses indicate that the impact of macroeconomic variables on the term structure differs from existing models. For short maturities, inflation is relatively more important than real activity growth at short forecast horizons. For longer maturity yields, the long-run component of inflation explains most of the long-horizon forecast variance, but real activity growth matters for short

forecast horizons. Unlike existing macro models, the model implies plausible term premia and expectations of short rates. The long-run components also improve the prediction of bond excess returns relative to information in the yield curve and macro variables. Measures of in-sample and out-of-sample fit confirm the benefits of allowing for long- and short-run components.

Contents

List of Tables

List of Figures

Chapter 1

Forecasting Bond Risk Premia with Unspanned Macroeconomic Information

1.1 Introduction

Do macroeconomic fundamentals contain information about future bond yields and bond risk premia that is not fully embedded in current bond yields? An extensive literature based on ad-hoc predictive regressions and dynamic macro-finance term structure models (MTSMs) shows that variables reflecting the state of the economy help predict interest rates and bond risk premia.[1] However, some concerns have been raised about the robustness of the documented predictive evidence (Duffee, 2007, 2013a, 2013b; Bauer and Hamilton, 2016).[2]

[1]For examples of predictive regressions, see among others Ludvigson and Ng (2009, 2010), Cooper and Priestley (2009), and Greenwood and Vayanos (2014). Examples of MTSMs include Ang and Piazzesi (2003), Bikbov and Chernov (2010), DeWachter and Lyrio (2006), Ang, Dong, and Piazzesi (2007), and Mőnch (2008). Additionally, studies based on the dynamic Nelson-Siegel (1987) model also show that incorporating macro information is beneficial for explaining the yield curve: Diebold, Rudebusch, and Aruoba (2006) and Mőnch (2012).

[2]Duffee (2007, 2013b) and Bauer and Hamilton (2016) express concerns regarding the statistical reliability of the inferences based on the regression models. Small-sample distortion can be particularly serious, since the yield factors are not strictly exogenous, and both the yield factors and the macro variables are highly persistent. Duffee (2013a) expresses a particular concern that the predictive evidence in Ludvigson and Ng

Furthermore, Joslin, Priebsch, and Singleton (2014) argue that existing MTSMs impose inconsistent restrictions on the joint distribution of bond yields and the macroeconomy. The MTSMs implicitly indicate that macro variables can be inverted as a linear combination of contemporaneous yield curve. This invertibility (Duffee, 2011a) or spanning condition seems inconsistent with evidence on the predictive power of macroeconomic variables.

In this paper, I analyze the predictive power of macro information in a consistent framework by using the new family of unspanned MTSMs proposed by Joslin, Priebsch, and Singleton (2014). Specifically, I investigate the performance of the MTSM with unspanned output and inflation risks from the perspective of a bond investor in an out-of-sample exercise. One key innovation of this study is to take into account model uncertainty by combining forecasts with and without unspanned macro information. Bond prices and returns are sensitive to monetary policy and inflation prospects, both of which are known to shift over time.[3] Therefore, it is critical to adopt a framework which accounts for the possibility that the forecasting model may change over time.

I optimally combine the forecasts from the perspective of the forecaster's evaluation loss function. I consider a commonly used sum squared error loss function and a loss function based on the mean–variance utility of a bond portfolio. Following Carriero and Giacomini (2011), I treat the optimal combination weight as a measure of the usefulness of the unspanned macro information in forecasting bond excess returns, and conduct formal statistical tests on this measure.[4] I allow the combination weight to be time-varying to accommodate changes in the relative performance of term structure models with and without unspanned macro information.

I analyze the role of macro information in a no-arbitrage framework, because restrictions

(2009) is sample-specific.

[3]See among others Cogley and Sargent, (2001, 2005), Erceg and Levin (2003), Boivin and Giannoni (2006), Stock and Watson (2007), Bikbov and Chernov (2008), Cogley, Primiceri, and Sargent (2010), Faust and Wright (2013).

[4]Carriero and Giacomini (2011) develop a general framework for analyzing the usefulness of imposing parameter restrictions on a forecasting model. They apply the methodology to study the usefulness of no-arbitrage restrictions in affine term structure models from out-of-sample forecasting perspective.

on market prices of risk typically improve inference on risk premia by making better use of the information in the cross section of yields (e.g., Cochrane and Piazessi, 2008; Duffee, 2011a; Joslin, Priebsch, and Singleton, 2014; Bauer, 2016). I apply a novel framework to examine the risk premium restrictions for the purpose of forecasting. I extend the model selection method in Joslin, Singleton, and Zhu (2011) to unspanned MTSMs. I impose risk premium restrictions that depend on the number of term structure factors (level, slope, and curvature) which have time-varying risk premia. This model selection method only requires the estimation of three alternative specifications, which provides computational advantages when conducting a recursive out-of-sample exercise.[5] By incorporating the no-arbitrage condition into the model selection procedure, I can capture the predictability that cannot be identified through regression method.

My empirical findings suggest that the time variation in bond risk premia depends on different term structure factors over the estimation sample period (1981:11 to 2007:12). The preferred unspanned model, which takes into account these changes in the structure of risk premia, provides better forecasting performance than the model with only yield curve information at long forecast horizons (1-, 2-, and 5-year). For example, at the five-year horizon, the improvement in the root mean squared error (RMSE) from using unspanned macro information in forecasting bond excess returns is about 11% on average across maturities, which corresponds to an out-of-sample R-square of 21%. The superior predictive power of the preferred unspanned macro model can be translated into economic benefits for bond investors with mean-variance utility at long forecast horizons. For example, at the five-year horizon, the annualized certainty equivalent return (CER) for a mean-variance investor with low risk aversion level is 2.21% by using macro information relative to 1.46% by using yield curve information only.

The forecasting performance of macro variables can be further improved by optimally

[5] In Joslin, Priebsch, and Singleton (2014), $524,288$ possible parameter restrictions need to be examined. Brute-force estimation of this amount of term structure models is computationally costly in a recursive out-of-sample exercise.

combining forecasts with and without unspanned macro information, especially when allowing for time-varying combination weights. For example, at the five-year horizon, the greatest improvement in RMSE from using unspanned macro information in the forecast combinations is about 14% on average across maturities, which corresponds to an out-of-sample R-square of 26%. For mean-variance bond investors with low risk aversion level, the largest annualized CER from the forecast combinations at the five-year horizon is 3.98% by using macro information relative to 1.46% by using yield curve information only. These results are robust to using a power utility function or alternative performance measures.

Furthermore, formal statistical tests on the optimal combination weights suggest that the usefulness of the unspanned macro information is time-varying, and the hypothesis that unspanned macro information is consistently useless for forecasting bond risk premia is strongly rejected. These results hold for all the horizons and maturities under consideration, and they are also robust to both statistical loss function and utility-based loss function. Overall, I conclude that there is substantial out-of-sample evidence against the spanning hypothesis, which states that all the relevant information for predicting future yields and returns is spanned by level, slope and curvature of the yield curve.

The rest of this paper proceeds as follows. Section 2 summarizes the related literature. Section 3 compares the specifications of the term structure models with and without unspanned macro information. Section 4 discusses the parameterization for the market price of risk from the perspective of forecasting. Section 5 presents the framework to combine the forecasts with and without unspanned macro information optimally for the forecaster's objective. Section 6 describes the data and the estimation method. Section 7 presents the model selection results. Section 8 reports the out-of-sample forecasting results evaluated with sum squared error loss and mean-variance utility loss. Section 9 contains robustness analyses. Section 10 concludes.

1.2 Overview of Related Literature

While several studies have used unspanned MTSMs to investigate the relation between bond yields and the state of the economy (see among others Wright, 2011; Priebsch, 2014; Chernov and Mueller, 2012), only Bauer and Rudebusch (2016) directly focus on testing the validity of unspanned MTSMs. In contrast to Joslin, Priebsch, and Singleton (2014), Bauer and Rudebusch (2016) argue that unspanned models are nested by canonical MTSMs.[6] They perform direct likelihood-ratio test of the restrictions required for unspanned model and conclude that these restrictions are rejected by the data. The focus of this paper is not on determining which term structure model is a superior specification for incorporating macro information. Instead I investigate if macro information is useful for forecasting bond returns given information in the current yield curve. I approach this question using unspanned MTSMs, because both Joslin, Priebsch, and Singleton (2014) and Bauer and Rudebusch (2016) claim that the predictive power of macro variables documented in the literature is consistent with this model specification. Moreover, unspanned MTSMs incorporate macro information in a more parsimonious way than canonical MTSMs, which may be useful for out-of-sample forecasting.

This paper is also related to the existing work that uses combination schemes to forecast bond yields. Issues like structural breaks and misspecification biases make it difficult to find a model that consistently outperforms all others. De Pooter, Ravazzolo, and Van Dijk (2010) find that incorporating macro information into autoregression and vector autoregression models, dynamic Nelson-Siegel (1987) class of models, affine models, and their combinations improves interest rates forecasts. Byrne, Cao, and Korobilis (2015) document improved yields forecasts using dynamic Nelson-Siegel models with macroeconomic and financial variables, in which time-varying parameters, stochastic volatility and dynamic model averaging are considered. These studies have considered the predictive content of unspanned information

[6] Joslin, Priebsch, and Singleton (2014) treat the unspanned model as the nesting model. The spanned model is the nested model with only yield factors. The spanned model is rejected in favor of the unspanned model based on a likelihood-ratio test.

in dynamic factor models, but they do not directly test the spanning hypothesis. The forecast combination schemes employed in these papers are not designed to examine the usefulness of unspanned macro information in forecasting bond risk premia.

This study also contributes to the literature that investigates the economic predictability of bond risk premia. Thornton and Valente (2012) show that predictive models using long-term forward rates are unable to generate economic value over no-predictability expectation hypothesis benchmark. Sarno, Schneider, and Wagner (2014) reach a similar conclusion through an estimation strategy for affine term structure models, which jointly fits yields and bond excess returns. On the other hand, Gargano, Pettenuzzo, and Timmermann (2015) show that combining return-forecasting factors, such as Fama and Bliss (1987) forward spread, Cochrane and Piazzesi (2005) factor, and Ludvigson and Ng (2009) macro factor, in a regression model that allows for time-varying parameters and stochastic volatility leads to substantial gains in out-of-sample forecasting accuracy compared with the expectation hypothesis. These studies focus on the null hypothesis that bond excess returns are unpredictable. Nevertheless, they do not test the more relevant null hypothesis that bond risk premia is uncorrelated with the macroeconomy conditioning on the information in current yield curve, which is the purpose of my paper.[7] To the best of my knowledge, this is the first study to evaluate the unspanned macro model using an economic loss function.

More recently, the literature has uncovered a wide array of macro-based factors that forecast bond excess returns. Cieslak and Povala (2015) construct a cyclical factor based on yields and inflation. Huang and Shi (2011) extract an aggregate factor from a large sample of macro variables using a model selection method that takes into account the forecasting objective. Eriksen (2015) takes a forward-looking perspective by using the survey forecasts of macro variables. Coroneo, Giannone, and Modugno (2016) extract unspanned macro

[7] Gargano, Pettenuzzo, and Timmermann (2015) show that their forecasting models with Ludvigson and Ng (2009) macro factor generate more accurate forecasts than the expectation hypothesis benchmark, because some, but not all, of the information in the macro factor is spanned by the yield curve. Based on Duffee (2011), the hidden or unspanned risk factors affect the expected future short rates and the term premia in opposite directions. Their forecasts of bond excess returns are negatively correlated with the survey-based forecasts of future short rates.

factors from the cross sections of macroeconomic indicators and yields using dynamic factor model. Additionally, examples of other unspanned return-forecasting factors include jump risk (Wright and Zhou, 2009), option prices (Almeida, Graveline, and Joslin, 2011), and technical indicators (Goh, Jiang, Tu, and Zhou, 2013). This paper does not directly focus on the identification of new return-forecasting factors and the conventional measures of real economic activity and inflation. Instead the paper's contribution is to provide a framework to test the spanning hypothesis out of sample, both statistically and economically. This framework can be applied to any return-forecasting factors.

1.3 The Model

1.3.1 MTSMs with Unspanned Macro Factors

In this section, I describe the main aspects of unspanned MTSMs. For further details I refer to Joslin, Priebsch, and Singleton (2014, henceforth referred to as JPS). The canonical MTSMs generally have three components: an equation relating the short rate to the risk factors, a time series model for the risk factors, and a dynamic specification for the risk factors under risk-neutral pricing measure. These components are given by

$$r_t = \rho_0 + \rho_1 Z_t, \tag{1.1}$$

$$Z_t = K_0^P + K_1^P Z_{t-1} + \Sigma \varepsilon_t^P, \tag{1.2}$$

$$Z_t = K_0^Q + K_1^Q Z_{t-1} + \Sigma \varepsilon_t^Q. \tag{1.3}$$

Following the normalization in Joslin, Le, and Singleton (2013, henceforth referred to as JLS), the state variables that determine yields Y_t are denoted by Z_t including both macro factors M_t and yield factors PC_t^l. Y_t is a $N \times 1$ vector, N denotes the number of available yields in the term structure. M_t is a $m \times 1$ vector, PC_t^l is the first l principal components of the observed yield curve $PC_t^l = W^l Y_t$, W^l is the corresponding portfolio weights, a $l \times N$ matrix. $k = m + l$ risk factors $Z_t = [PC_t^{l\prime}, M_t^{\prime}]^{\prime}$ are all observable. The cross section of bond yields is well described by low-dimensional factor models, I therefore focus on the model with three yield factors ($l = 3$) and two macro variables ($m = 2$). r_t is the one-period spot interest rate, ε_t^P and ε_t^Q are assumed to be distributed $N(0,\, I_k)$, and $\Sigma\Sigma^{\prime}$ is the conditional covariance matrix of Z_t.

Given the dynamics in equations (1.1) to (2.5), the model-implied continuously compounded yields $\widehat{Y}_t$ are (see Duffie and Kan, 1996)

$$\widehat{Y}_t = A(\Theta^Q) + B(\Theta^Q) Z_t, \tag{1.4}$$

where the $N \times 1$ vector $A(\Theta^Q)$, and the $N \times k$ matrix $B(\Theta^Q)$ are functions of the parameters under the Q-dynamics, $\Theta^Q = \{K_0^Q, K_1^Q, \rho_0, \rho_1, \Sigma\}$, through a set of recursive equations. Equations (2.7) can be used to invert for Z_t, and in particular for macro factors. This is the invertibility (Duffee, 2011a, 2013b) or spanning condition in canonical MTSMs. Appendix A provides details on the bond pricing and the implied spanning condition.

The new class of MTSMs proposed by JPS imposes knife-edge restrictions on the canonical model to avoid theoretical macro-spanning, which contradicts with the empirical regression evidence (Ludvigson and Ng, 2009; Cooper and Priestley, 2009; Greenwood and Vayanos, 2014). [8] For the short rate process, instead of using equation (1.1), the unspanned macro model assumes that the short rate depends only on yield factors

$$r_t = \rho_0 + \underbrace{\begin{bmatrix} \rho_{PC}, & 0_m \end{bmatrix}}_{\rho_1} \underbrace{\begin{bmatrix} PC_t^l \\ M_t \end{bmatrix}}_{Z_t}. \tag{1.5}$$

[8] For an in-depth analysis on the inconsistency between the theoretical spanning condition and empirical regression evidence, I refer to Bauer and Rudebusch (2016).

The dynamics of the risk factors under Q-measure is given by

$$PC_t^l = \underbrace{K_{0PC}^Q}_{K_0^Q} + \underbrace{\begin{bmatrix} K_{1PC}^Q, & 0_{l\times m} \end{bmatrix}}_{K_1^Q} \underbrace{\begin{bmatrix} PC_{t-1}^l \\ M_{t-1} \end{bmatrix}}_{Z_{t-1}} + \Sigma_{PC}\varepsilon_{PCt}^Q, \tag{1.6}$$

where ε_{PCt}^Q is assumed to be distributed $N(0,\, I_l)$. $\Sigma_{PC}\Sigma_{PC}'$ is the upper $l \times l$ block of $\Sigma\Sigma'$. The risk-neutral dynamics implies that yields only depend on yield factors but not on macro factors. Therefore, the model-implied continuously compounded yields $\widehat{Y}_t$ are given by

$$\widehat{Y}_t = A(\Theta^Q) + \underbrace{\begin{bmatrix} B_{PC}(\Theta^Q), & 0_{m\times m} \end{bmatrix}}_{B(\Theta^Q)} \underbrace{\begin{bmatrix} PC_t^l \\ M_t \end{bmatrix}}_{Z_t}. \tag{1.7}$$

The above equation implies that in the unspanned model, there is no direct link from macro variables to contemporaneous yields. The macro variation is not fully captured by model-implied yields, hence the unspanned macro variation can have predictive power for future yields and returns.

For the P-measure, the dynamics of the risk factors are the same as in equation (2.4). I can rewrite it as follows

$$\underbrace{\begin{bmatrix} PC_t^l \\ M_t \end{bmatrix}}_{Z_t} = \underbrace{\begin{bmatrix} K_{0PC}^P \\ K_{0M}^P \end{bmatrix}}_{K_0^P} + \underbrace{\begin{bmatrix} K_{PC}^P & K_{PCM}^P \\ K_{MPC}^P & K_M^P \end{bmatrix}}_{K_1^P} \underbrace{\begin{bmatrix} PC_{t-1}^l \\ M_{t-1} \end{bmatrix}}_{Z_{t-1}} + \Sigma \underbrace{\begin{bmatrix} \varepsilon_{PCt}^P \\ \varepsilon_{Mt}^P \end{bmatrix}}_{\varepsilon_t^P}. \tag{1.8}$$

The stochastic discount factor (SDF) is therefore given by

$$-\log m_{PC,t+1} = r_t + \frac{1}{2}\lambda_{PCt}'\lambda_{PCt} + \lambda_{PCt}'\varepsilon_{PCt+1}^P, \tag{1.9}$$

where the market prices of risks are affine function of the factors

$$\lambda_{PCt} = \Sigma_{PC}^{-1}(\lambda_{0PC} + \lambda_{1PC} Z_t). \tag{1.10}$$

λ_{0PC} and λ_{1PC} are given by

$$\lambda_{0PC} = K_{0PC}^{P} - K_{0PC}^{Q}, \tag{1.11}$$

$$\lambda_{1PC} = \left[K_{PC}^{P}, K_{PCM}^{P}\right] - \left[K_{1PC}^{Q}, 0_{l\times m}\right]. \tag{1.12}$$

In this specification, the unspanned macro risk is built into the term structure model by construction. The only way macro variables enter the unspanned model is as additional predictors in the VAR process under P-measure, equation (1.8). Therefore, they affect real-world expectations of future interest rates and term premia, but they do not directly affect current yields.

1.3.2 Spanned Term Structure Models with Yield Factors

The spanned model enforces spanning of the forecasts of output and inflation by the yield factors (Bernanke, Reinhar, and Sack, 2004; Kim and Wright, 2005; JPS). It can be obtained by imposing the following restrictions on the unspanned macro model

$$K_{PCM}^{P} = 0_{l\times m}, \tag{1.13}$$

$$K_{M}^{P} = 0_{m\times m}. \tag{1.14}$$

The spanned model has the same Q-dynamics as the unspanned macro model, but different P-specifications. For model with three yield factors and two macro factors, ten zero-restrictions are required. These restrictions are referred to as spanning restrictions in the paper.[9] The spanned model corresponds to the yield factors only model. Following Joslin, Singleton, and Zhu (2011, henceforth referred to as JSZ), the dynamics of the yield factors only model is given by setting $Z_t = PC_t^l$ in equations (1.1) to (2.5). The only difference between the spanned model and the yield factors only model is that the spanned model includes two VAR equations for forecasting macro variables using yield factors. But for bond yields and risk premia, the two models have the exact same observational implications.

1.3.3 Bond Risk Premia

Bond excess return rx_{t+h}^n is the continuously compounded return of holding an n-maturity bond for h periods, in excess of the h-period rate. The fitted excess return $\widehat{rx}_{t+h}^n$ can be computed using model-implied yields

$$\widehat{rx}_{t+h}^n = -(n-h)\widehat{Y}_{t+h}^{n-h} + n\widehat{Y}_t^n - \widehat{Y}_t^h, \tag{1.15}$$

where $\widehat{Y}_{t+h}^{n-h}$, $\widehat{Y}_t^n$ and $\widehat{Y}_t^h$ are given by equation (1.7).

Model-implied bond risk premia is the conditional expectation of bond excess return[10]

$$E_t\left[\widehat{rx}_{t+h}^n\right] = -(n-h)\widehat{Y}_{t+h|t}^{n-h} + n\widehat{Y}_t^n - \widehat{Y}_t^h, \tag{1.16}$$

where $\widehat{Y}_{t+h|t}^{n-h}$, model's prediction of the h-period ahead $(n-h)$-maturity yield computed using

[9]Bauer and Rudebusch (2016) treat the JLS canonical macro model as the spanned model, and it nests the unspanned model with eight unspanned restrictions for model with three yield factors and two macro factors. The unspanned macro model and the JLS model only differ in Q-specifications. The JLS model uses both yield factors and macro variables to explain the cross-sectional fit of the yield curve, while the unspanned model uses only yield factors. The yield factors: level, slope, and curvature, have already explained almost all the variation of the yield curve (Litterman and Scheinkman, 1991). The JLS model tends to imply implausibly large Sharpe ratios that arise from overfitting the pricing distribution of the risk factors (Duffee 2010).

[10]Appendix B provides more discussions on the model's implications for bond risk premia.

parameters estimated at time t is given by

$$\begin{aligned}\widehat{Y}_{t+h|t}^{n-h}(\Theta) &= A_{n-h}(\Theta^Q) + B_{n-h}(\Theta^Q)\widehat{Z}_{t+h|t} \\ &= A_{n-h}(\Theta^Q) + B_{n-h}(\Theta^Q) f(Z_t, h; K_0^P, K_1^P),\end{aligned} \tag{1.17}$$

where $\Theta = \{\Theta^Q, K_0^P, K_1^P\}$, and f is given by

$$f(Z_t, h; K_0^P, K_1^P) = (I_k + K_1^P + ... + (K_1^P)^{h-1})K_0^P + (K_1^P)^h Z_t.$$

1.4 No-Arbitrage and Risk Premium Restrictions

Absence of arbitrage creates a link between the cross-sectional variation of interest rates and their time-series variation through a risk premium adjustment. Cochrane and Piazzesi (2008) and Bauer (2016) demonstrate that imposing restrictions on market prices of risks can help to identify the time-series estimation by making better use of the information in the cross section of bond yields. Moreover, the number of parameters in MTSMs is large and there is a risk of overfiting the joint dynamics of yield curve and macro variables (Kim, 2007).[11] Imposing restrictions on market prices of risks can provide a more parsimonious MTSMs. Motivated by these insights, I discuss the parameterizations of market prices of risks in this section. I first discuss the risk prices restrictions of dynamic term structure models (DTSMs) in existing literature. Subsequently I describe the method used in this paper to select risk premium restrictions from out-of-sample forecasting perspective.

1.4.1 Risk Premium Restrictions in DTSMs

In existing studies, zero restrictions on risk price parameters are generally imposed in an ad-hoc approach, which is based on the t-statistics from in-sample estimation (see Dai and Sin-

[11]The unspanned macro model has 30 parameters governing the P-dynamics (5 for K_0^P, and 25 for K_1^P), and 19 additional parameters governing the Q-dynamics (1 for r_∞^Q, 3 for λ^Q, and 15 for Σ). Section 1.6.2 discusses model normalization.

gleton, 2002; Duffee, 2002; Ang and Piazzesi, 2003; Kim and Wright, 2005). This approach is unsatisfactory since joint restrictions are imposed without considering joint significance. JPS estimate their unspanned macro model under all possible zero-restrictions on risk price parameters, and select one specification based on standard information criteria (e.g., Akaike (1973) information criteria and Schwarz (1978) Baysian information criteria, referred to as AIC and SBIC henceforth). The scaled market price of risk in equation (1.10) depends on 18 parameters (3 for λ_{0PC} and 15 for λ_{1PC}). There are 2^{18} possible risk price restrictions to be considered. Each of the 2^{18} models are further examined with and without largest eigenvalue constraint of K_1^P, for a total of $2^{19} = 524,288$ specifications to be estimated using maximum likelihood.[12] Even though JPS normalization scheme allows rapid convergence to the global optimum of the likelihood estimation, it is still computationally costly to estimate such a large number of models recursively in the out-of-sample exercise.

One might argue that the model selected based on the in-sample procedure of JPS should be used to investigate the out-of-sample forecasting performance, because in-sample tests generally have strong statistic power (Hansen and Timmermann, 2015; Diebold, 2015). However, the model selected based on the in-sample tests might not be always optimal for out-of-sample forecasting of yields. For example, Christensen, Diebold, and Rudebusch (2011) find that the independent-factor arbitrage free Nelson-Siegel (1987) model (AFNS) provides better forecasts than the AFNS model with correlated factors, even though the independent-factor model is rejected based on the in-sample likelihood ratio test. The independent-factor model is statistically rejected but economically important for forecasting yields.[13]

Moreover, it is crucial to allow the risk premium structure to change over time. Monetary policy shifts can affect agent's risk assessments in the future, hence the source of time-varying

[12]JPS also constrain the largest eigenvalue of K_1^P to be the same as the largest eigenvalue of K_{1PC}^Q. This constraint is motivated by the near unit-root behavior of yields under both P- and Q-measures. The estimates of K_1^P from a vector autoregression (VAR) process are generally biased toward a dynamic that displays less persistence than the true process.

[13]I conduct out-of-sample forecasting exercise for the preferred model in JPS. Same sample period and same risk price restrictions are applied as in JPS. I find that the RMSEs of the predicted bond excess returns from the preferred model is even larger than those from the maximally flexible unspanned model.

risk premia may differ over time. For example, Ang, Boivin, Dong, and Loo-Kung (2011) find that agents tend to assign a risk discount to monetary policy shifts and are willing to pay to be exposed to activist monetary policy. A large literature on learning also demonstrate that investors in government bond markets adjust their forecasts of bond yields and assessments of required compensations for bearing relevant factor risks when the Federal Reserve change its operating procedures and disclosure policies (see for instance Orphanides and Wei, 2012; Laubach, Tetlow, and Williams, 2007; Cogley 2005; Piazzesi and Schneider, 2007; Giacoletti, Laursen, and Singleton, 2014). These findings suggest that in-sample model selection method may not be plausible, because it can easily overlook important variations around the turning points of monetary policy and economic conditions.

1.4.2 Risk Premium Restrictions for Forecasting

In this paper, I focus on a different type of risk premium restrictions: the number of term structure factors that have time-varying risk premia, which is directly against the constant risk premium as claimed by the weak expectation hypothesis. JSZ propose a novel test based on the rank of the risk price parameter in their canonical specification. I extend this method to the unspanned macro model. More importantly, this method can be easily implemented at each recursion, which is well suited for out-of-sample forecasting exercise.

To interpret the rank restriction, I write the one-period expected excess returns on portfolios of bonds with payoffs tracking the first three principal components PC_t as follows (Cox and Huang, 1989 and JPS)

$$\begin{aligned} xPC_t &= \lambda_{0PC} + \lambda_{1PC} Z_t, \\ &= \left(K_{0PC}^P - K_{0PC}^Q\right) + \left(\left[K_{PC}^P, K_{PCM}^P\right] - \left[K_{1PC}^Q, 0_{l\times m}\right]\right) Z_t. \end{aligned} \tag{1.18}$$

The component of $\left(\left[K_{PC}^P, K_{PCM}^P\right] - \left[K_{1PC}^Q, 0_{l\times m}\right]\right) Z_t$ is the source of the time-varying risk premium for exposure to Z_t. Therefore, the constraint that the one-period expected excess returns on bond portfolios are driven by ζ linear combinations of the pricing factors Z_t amounts to the constraint that the rank of λ_{1PC} is ζ. The constraints on the rank of the risk price matrix essentially can be translated into parameter space. I provide details of the translation in Appendix C. Under the weak expectation hypothesis, λ_{1PC} is zero, expected excess returns and term premia are constant.

Existing studies find that ζ is usually smaller than three. Level, slope, and curvature risks are not all priced in the economy. Duffee (2002) and Bauer (2016) show that conditioning on yield curve information, only variation in the price of level risk is necessary to capture the failure of the expectation hypothesis. This variation is typically captured by linking the price of level risk to the slope of the term structure. Duffee (2011a) proposes a dynamic term structure model with information hidden to the cross-section of yields, in which a single linear combination of state variables determine the required risk premia. JPS find that market prices of risks depend on both level and slope factors, and are earned in compensation for exposure to macroeconomic shocks. Motivated by these findings, I focus on three specifications with $\zeta = 1$ (model M_1), $\zeta = 2$ (model M_2), and $\zeta = 3$ (the unrestricted model M_0). By focusing on the reduced-rank restriction, only three alternative models need to be estimated recursively, which provides significant efficiency in the forecasting process. Duffee (2010) notes that unconditional Shape ratios are higher for short-maturity bonds than long-maturity bonds. This observation is consistent with nonzero unconditional mean of the risk prices for both level and slope. I therefore treat λ_{0PC} as free parameters.

1.5 Forecast Combination with Respect to Loss Functions

I test the spanning hypothesis by examining the performance of the unspanned macro model relative to that of the spanned model out of sample, both statistically and economically. Following Carriero and Giacomini (2011), I employ a forecast combination and estimate the combination weight by optimizing the out-of-sample evaluation loss function. The optimal weight associated with the unspanned model measures the usefulness of the macro information in forecasting bond excess returns. This combination weight can further be used to produce combined forecasts that optimally exploit the macro information for the forecast loss function. Moreover, the combination weight can be time-varying to accommodate changes in the usefulness of the unspanned macro information over time. I also formally test the hypothesis that the macro information is useless by an out-of-sample encompassing test. In this section, I first layout the framework to estimate and test the combination weight with respect to the forecast loss function of interest. Subsequently, I discuss other related forecast combination methods.

1.5.1 General Set-Up

The forecasts are obtained by a m-year rolling window estimation scheme.[14] At each time t, I estimate three specifications of the unspanned macro models M_1, M_2, and M_0 using data from $t-m+1$ to t. I rank these models according to standard information criteria and choose the most plausible model to produce h-period ahead forecasts of bond excess return using equation (1.16). Our first estimation starts at $t = J$, and G is the length of the out-of-sample period. This gives a sequence of forecasts from the preferred unspanned macro model: $\widehat{rx}^{M_{us}}_{t+h|t}$. The spanning restrictions are enforced on the preferred unspanned model to generate the benchmark forecasts $\widehat{rx}^{M_{span}}_{t+h|t}$. I suppress the maturity dependence of bond excess return for notational simplicity throughout.

Consider the following combination forecast for forecasting horizon h

$$f_{t+h|t} = \beta_h \widehat{rx}^{M_{us}}_{t+h|t} + (1-\beta_h)\widehat{rx}^{M_{span}}_{t+h|t}, \tag{1.19}$$

[14] We use a rolling window to update the parameter estimates of models, because parameters estimated using data from very long-ago periods may not be necessarily useful to make current out-of-sample predictions due to the shifts in Fed policy and macro environment.

where β_h is the weight on the preferred unspanned macro model. The optimal combination weight can be estimated by minimizing the expected out-of-sample evaluation loss function

$$\widehat{\beta}_h = \underset{\beta_h \in R}{\arg\min} E[L(rx_{t+h}, f_{t+h|t})], \tag{1.20}$$

where L denotes the evaluation loss function. L can be either the sum squared error loss or the classic Markowitz (1952) mean-variance utility loss. rx_{t+h} represents the observed h-period bond excess return. The estimated weight $\widehat{\beta}_h$ is the measure of the usefulness of unspanned macro information for forecasting h-period bond excess returns under a given loss function L. A large $\widehat{\beta}_h$ suggests that the unspanned macro information is useful for forecasting h-period bond excess returns. Following Carriero and Giacomini (2011), I construct formal statistical tests on whether the unspanned macro information is useful for forecasting ($H_0 : \widehat{\beta}_h = 0$). Appendix D presents the derivation of the tests. The estimated combination weight can further be used to produce combined forecasts $f^*_{t+h|t}$, which exploit the unspanned macro information in a way that is optimal for the forecaster's loss function.

The above framework can be extended to allow for time-varying combination weight. $\beta_{t,h}$ measures the usefulness of unspanned macro information for forecasting h-period bond excess returns under a given loss function L conditional on time t. The consistent estimator of $\beta_{t,h}$ can be obtained by solving the optimization problem over a smoothing window of size d using data up to time t.[15]

$$\widehat{\beta}_{t,h} = \underset{\beta_{t,h} \in R}{\arg\min} \sum_{j=t-d+1-h}^{t} E[L(rx_{j+h}, f_{j+h|t})], \tag{1.21}$$

$$\begin{aligned} \text{for } j &= t-d+1-h, ..., t-h, \\ t &= J+d-1+h, ..., J+G. \end{aligned}$$

[15]The window size d is a constant fraction of the out-of-sample size G. In the empirical exercise, we choose $d = G \times 0.2$, which corresponds to a critical value of 3.179 for 95% confidence bands, according to Carriero and Giacomini (2011).

A plot of the sample path of $\widehat{\beta}_{t,h}$ can reveal possible time-variation in the usefulness of the unspanned macro information. Appendix D shows how to test the hypothesis that the preferred unspanned forecast is consistently useless ($H_0 : \widehat{\beta}_{t,h} = 0$) or the spanned forecast is consistently useless ($H_0 : \widehat{\beta}_{t,h} = 1$) over time.

In the following two subsections, I present the closed-form solution to the optimization problem in equation (1.20) regarding the loss functions under consideration.

1.5.2 Sum Squared Error Loss

The sum squared error (SSE) loss is defined as

$$L(rx_{t+h}, f_{t+h|t}) = (rx_{t+h} - f_{t+h|t})^2. \tag{1.22}$$

The consistent estimator of β_h can be seen as the estimator of β_h in the following regression

$$rx_{t+h} - \widehat{rx}^{M_{us}}_{t+h|t} = \beta_h(\widehat{rx}^{M_{us}}_{t+h|t} - \widehat{rx}^{M_{span}}_{t+h|t}) + \eta_{t+h}, \tag{1.23}$$

$$\text{for } t = J, ..., J + G.$$

Note that β_h is maturity specific for the SSE loss, since the forecast errors are evaluated for each maturity respectively. The combined weight for n-maturity bond at h-horizon is denoted as β^n_h. In the presence of time variation, the consistent estimator of $\beta_{t,h}$ can be obtained by estimating equation (1.23) over each smoothing window. Note that $\beta_{t,h}$ is also maturity specific, I denote it as $\beta^n_{t,h}$.

1.5.3 Mean-Variance Utility Loss

The portfolio utility loss function considers the asset allocation decisions of an investor with investment horizon h. The investor chooses to allocate his or her wealth between h-maturity bond and $\widetilde{N}$ bonds with maturities greater than h. Since the maturity of the shorter-term bond matches the investment horizon, the h-period bond represents the risk-free asset. The longer-term bonds represent the risky assets. I evaluate the bond portfolios using the classic mean-variance utility loss[16]

$$L(rx_{t+h}, f_{t+h|t}) = -E_t[W_{t+h}] + \frac{\gamma}{2} var_t[W_{t+h}], \tag{1.24}$$

where $f_{t+h|t}$ represents the $\widetilde{N} \times 1$ vector of conditional expectations of returns for risky bonds generated by the forecast combination. rx_{t+h} stands for the $\widetilde{N} \times 1$ vector of observed h-period excess returns for $\widetilde{N}$ risky bonds. W_{t+h} is the wealth at time $t+h$ for an investor with relative risk aversion γ. $W_{t+h} = Y_{t+h}^h + \omega_{t,h} rx_{t+h}$, where $\omega_{t,h}$ is a $1 \times \widetilde{N}$ vector representing weights of the risky bonds at time t for forecast horizon h.

Given the classic solution of $\omega_{t,h}$, I minimize equation (1.24) with respect to β_h. More details on the optimization problem are provided in Appendix D. The closed-form solution to this problem is given by[17]

$$\widehat{\beta}_h = \frac{2E\left[Y_{t+h}^h + \left(rx_{t+h} - \widehat{rx}_{t+h|t}^{M_{span}}\right)(\widehat{rx}_{t+h|t}^{M_{us}} - \widehat{rx}_{t+h|t}^{M_{span}})\right]}{E\left[\widehat{rx}_{t+h|t}^{M_{us}} - \widehat{rx}_{t+h|t}^{M_{span}}\right]}, \tag{1.25}$$

$$\text{for } t = J, ..., J+G.$$

In the empirical exercise, I also consider the situation when $\omega_{t,h}$ is constrained. I rule out short selling ($\omega_{t,h}^s \geq 0, s = 1, ..., \widetilde{N}$) and also impose constraints on total portfolio leverage ($\sum_{s=1}^{\widetilde{N}} \omega_{t,h}^s - 1 \leq leverage$) in order to avoid implausible positions. However, no closed-form solution is available when $\omega_{t,h}$ is constrained. I therefore numerically optimize investor's utility (1.24) to solve for constrained $\omega_{t,h}$ for each iteration of β_h. When time-variation is

[16] Thornton and Valente (2012), Sarno, Schneider, and Wagner (2014), Eriksen (2016), and Della Corte, Sarno, and Thornton (2008) also use quadratic utility framework to study the economic value in Treasury bond market.

[17] The estimated combination weight is irrelevant to investor's risk aversion and the variance-covariance matrix of bond excess returns in the close form. While with constrained $\omega_{t,h}$, investor's risk aversion and the variance-covariance matrix do affect the estimation of the combination weight.

allowed, the consistent estimator of $\beta_{t,h}$ can be obtained using equation (1.25) over each smoothing window. For constrained $\omega_{t,h}$, I numerically optimize equation (1.24) over each smoothing window for each iteration of $\beta_{t,h}$.

1.5.4 Related Combination Scheme

This combination method as discussed in Carriero and Giacomini (2011) is related to Bayesian model averaging (BMA), in which combination weights are posterior probabilities. However, the limitation of BMA is that it assumes that the true model is contained in the set of models under consideration. The optimal weights are constrained to be between 0 and 1. The combination weights depend on variances and correlation between the forecasts, and theoretically they can be negative or greater than 1. By allowing $\beta_h \in R$ and $\beta_{t,h} \in R$, I account for the possibility that both models are misspecified, but their forecasts can be useful if combined in an optimal way.

This combination method also relates to the optimal predictive pool (OPP) proposed by Geweke and Amisano (2011). The combination weights under OPP is determined by recursively maximizing the log score at each time period. Recursively updating the combination weights is empirically important for bond return forecasting, since it accommodates changes in the relative performance of different models (see Gargano, Pettenuzzo, and Timmermann, 2015). I follow the combination methodology in Carriero and Giacomini (2011), because the combination weight can be used to construct formal statistical tests on the usefulness of the unspanned macro forecasts. Moreover, this method can be easily tailored to portfolio utility loss function.

1.6 Data and Estimation

1.6.1 Data

I use continuously compounded monthly zero-coupon bond yields from the Gürkaynak, Sack, and Wright (2007, GSW) dataset. The sample period is from November 1971 to December 2012.[18] For the estimation, I use data up to December 2007 to avoid the recent zero lower bound episode, which is troublesome for affine models (Bauer and Rudebusch, 2013). I use ten maturities in the estimation, and use 24 maturities to construct 36 horizon (h) and maturity ($n-h$) combinations of bond excess returns. Following JPS, I use three-month moving average of the Chicago Fed National Activity Index (CFNAI) as a measure of real activity growth (GRO). Inflation is defined as the 12-month moving average of core CPI inflation.[19]

Table 1 presents summary statistics for the term structure of yields, macro variables and excess bond returns. Panel A reports the results for the first three principal components of yields with 24 maturities covering 1, 2, 3, 4, 6, 7, 9, 12, 13, 15, 18, 24, 25, 27, 30, 36, 48, 60, 61, 63, 66, 72, 84, and 120 months. Panel A of Figure 1 displays the time series of the first three principal components, which are typical level, slope and curvature factors. Level factor is highly persistent with mild excess kurtosis and positive skewness. The yield curve on average is upward sloping but with slightly negative skewed slope. Curvature factor has relatively lower persistence than the other two factors.

The macroeconomic data is obtained from the Federal Reserve Economic Data. The CFNAI is a weighted average of 85 existing monthly indicators of national economic activity.[20]

[18]The GSW dataset is compiled by the Federal Reserve. November 1971 is the earliest date with consistent availability of 10-year yield data for the dataset. The GSW data is available at `http://www.federalreserve.gov/pubs/feds/2006/200628/200628abs.html`

[19]JPS use expected rate of inflation from surveys of professional forecasters by Blue Chip Financial Forecasters, but Blue Chip inflation forecasts are not available prior to the early 1980s. Instead, I define inflation based on core CPI inflation. Core price indexes, which exclude the volatile food and energy components, are more stable and display a much stronger link to monetary policy actions and interest rates (Ajello, Benzoni, and Chyruk, 2012; Bauer and Rudebusch, 2015).

[20]The 85 economic indicators that are included in the CFNAI are drawn from four broad categories of data: production and income; employment, unemployment, and hours; personal consumption and housing;

It is constructed to have an average value of zero and a standard deviation of one. Panel B of Table 1 indicates that the GRO on average is -0.01% in my sample, which means that overall the economic growth is below the trend growth rate. Panel B of Figure 1 plots the time series of the macro data with National Bureau of Economic Research (NBER) recessions as shaded areas. GRO is quite negative around recession. Inflation is highly persistent with first-order autocorrelation of 0.99. At one-year displacement, inflation has much larger autocorrelation than GRO. As shown in Panel B of Figure 1, inflation attained twin peaks around the time of the 1970s oil shocks. It fell sharply during the Volcker disinflation, and then settled down around two percent after the mid-1990s. It dropped slightly during the financial crisis, but increased back to around two percent after 2010.

Bond excess returns are computed using all the 24 maturities. I focus on six holding periods h (1-, 3-, 6-, 12-, 24-, and 60-month) and six maturities $n - h$ (1-, 3-, 6-, 12-, 24-, and 60-month). The means, standard deviations, and Sharpe ratios are annualized and reported in Panel C of Table 1. For one-year holding period, mean excess bond returns range from 0.10% for the three-month bond to 1.78% for the five-year bond and corresponding standard deviations range from 0.19% to 7.19%, suggesting that investors require a higher premium for investing in longer-maturity (riskier) Treasury bonds. The return and volatility values correspond to Sharpe ratios ranging from 0.53 to 0.25. Similar patterns are observed across holding periods: Sharpe ratios are relatively larger for short-maturity bonds, which is consistent with Duffee (2010). Moreover, higher Shapre ratios are earned for short-holding period relative to long-holding period. Returns for short-holding period (1- and 3-month) are right-skewed, and more so for short-maturity bonds. Medium- and Long-horizon returns (6-month, 1-, 2-, and 5-year) are negatively skewed for most maturities. Short-horizon returns are more fat-tailed than long-horizon returns, especially for short-maturity bonds. Long-horizon returns display stronger serial correlation than short-horizon returns due to the overlapping nature of computing the holding period returns.

sales, orders, and inventories.

Figure 2 plots nine forecast horizon h (1-, 12-, and 60-months) and maturity $n - h$ (1-, 12-, and 60-months) combinations of annualized bond excess returns. The plot indicates that excess returns have common business-cycle movements. The patterns appear more clearly for one-year horizon (the middle panels).[21] This is consistent with existing predictability literature which documents that expected excess returns on bonds are countercyclical at longer holding periods, such as one year (e.g., Cochrane and Piazzesi, 2005). In particular, excess returns are high right after recession troughs. At one-year horizon, returns are high for all the three maturities around and after the 1974, 1982, 1991, and 2001 recessions. Returns are also high around 1984, which is the year with slower growth as indicated by GRO although it is not classified as recession. For one-month and five-year holding periods, the countercyclical patterns are much weaker. However, I note that the return series are high around the changing term of Fed Chairmen, especially for the beginning term of Volcker and Greenspan.

1.6.2 Estimation

The estimation of affine term structure models is challenging due to the high level of non-linearity in the parameters (Duffee, 2011b; Duffee and Stanton, 2012; Kim and Orphanides, 2012). Normalization assumptions are needed to identify the parameters of the model. I follow the parameterization of JPS and JSZ in which the yield factors are the first three principal components of the observed yield curve. The first three principal components are denoted as PC_t. This observable assumption substantially simplifies the estimation, since no filtering is necessary. More importantly, the parameters governing the P-dynamics $\Theta^P = \{K_0^P, K_1^P\}$ can be concentrated out of the likelihood function. They can be estimated simply from ordinary least squares (OLS). Given K_0^P and K_1^P from the OLS estimation, the sample conditional variance $\widehat{\Sigma\Sigma'}$ is computed and used as initial value for the optimization

[21] Annual returns have been the focus of most studies in the literature on bond return predictability (e.g., Ludvigson and Ng, 2009; Cooper and Priestley, 2009; Cieslak and Povala, 2015; Eriksen, 2016; Greenwood and Vayanos, 2014), while Gargano, Pettenuzzo, and Timmermann (2015) focus on monthly bond returns predictability.

of the likelihood function. The Q-parameters determining $A(\Theta^Q)$ and $B(\Theta^Q)$ are ultimately functions of $\Theta^Q = \{r^Q_\infty, \lambda^Q, \Sigma\}$, where r^Q_∞ is a scalar related to the long-run mean of the short rate under risk neutral measure, and λ^Q is a vector represents the ordered eigenvalues of K_1^Q. Appendix E provides further details about this transformation. Yield factors PC_t and macro variables M_t are assumed to be measured without errors. Yields are assumed to be measured with errors $Y_t = \widehat{Y}_t + e_t$, where e_t is a vector of measurement errors that is assumed to be $i.i.d.$ normal. I further assume that the errors on each maturity have equal variance σ_e^2 so that the likelihood tries equally hard to match the yield curve.

The conditional likelihood function of the yield curve Y_t is

$$f(Y_t|Y_{t-1},\Theta) = f(Y_t|PC_t, r^Q_\infty, \lambda^Q, \Sigma, \sigma_e^2) \times f(Z_t|Z_{t-1}, K_0^P, K_1^P, \Sigma), \tag{1.26}$$

where $\Theta = \{r^Q_\infty, \lambda^Q, \Sigma, \sigma_e^2, K_0^P, K_1^P\}$. The first term is the Q-likelihood capturing the cross-sectional dependence of yields on risk factors and its logarithm is given by

$$\log f(Y_t|PC_t, r^Q_\infty, \lambda^Q, \Sigma, \sigma_e^2) = const - (N-k)\log(\sigma_e^2) - 0.5\frac{\|e_t\|^2}{\sigma_e^2}. \tag{1.27}$$

Recall that N is the number of yields in the term structure, k is the number of risk factors under Q-measure. $\|e_t\|$ denotes the Euclidean norm of the vector of measurement errors, which $e_t = Y_t - \widehat{Y}_t = Y_t - (A(\Theta^Q) + B(\Theta^Q)Z_t)$. The second term of equation (1.26) is the P-likelihood capturing the time-series dynamics of the risk factors. It corresponds to the likelihood of a Gaussian VAR process. The logarithm is given by

$$\log f(Z_t|Z_{t-1}, K_0^P, K_1^P, \Sigma) = const - 0.5\log(\Sigma\Sigma') - 0.5\left\|\Sigma^{-1}(Z_t - K_0^P - K_1^P Z_{t-1})\right\|^2. \tag{1.28}$$

The unspanned model with reduced-rank risk premium restrictions can be easily estimated through the concentration of the likelihood. The parameters of the Q-distribution are determined largely by the cross section of bond yields, and not by their time-series properties under the P-dynamics. Therefore, any differences in the risk premium structure must be attributable to differences in the P-distribution. Given $\{r_\infty^Q, \lambda^Q, \Sigma, \sigma_e\}$, the estimates of K_1^P are computed in a novel way as proposed by JSZ. Based on equation (1.18), the estimates of the risk price parameters can be seen as the estimate of θ_1 in the following regression

$$PC_{t+1} - \left(K_{0PC}^Q + \left[K_{1PC}^Q, 0_{l\times m}\right] Z_t\right) = \theta_0 + \theta_1 Z_t + \varepsilon_{PCt}^P, \tag{1.29}$$

where the volatility matrix Σ_{PC} of errors ε_{PCt}^P are fixed. Recall that $\Sigma_{PC}\Sigma_{PC}'$ is the upper $l \times l$ block of $\Sigma\Sigma'$. θ_1 is restricted to have rank ζ. The estimates of the constrained regression can be computed in closed form. I refer Appendix C for details on the estimation of the constrained regression. Given the estimated risk price parameters $\widehat{\theta}_0$ and $\widehat{\theta}_1$, the P-dynamics can be computed as

$$K_{0PC}^P = K_{0PC}^Q + \widehat{\theta}_0, \tag{1.30}$$

$$\left[K_{PC}^P, K_{PCM}^P\right] = \left[K_{1PC}^Q, 0_{l\times m}\right] + \widehat{\theta}_1. \tag{1.31}$$

The elements in K_0^P and K_1^P that relate to the macro variables can be identified from the estimation of the VAR process. For the spanned model, additional spanning restrictions are imposed in the estimation of θ_1.

1.7 Model Selection Results

I estimate unspanned macro models M_1, M_2, and M_0 within each rolling window and rank these models according to standard information criteria. Figure 3 plots the differences in AIC and SBIC between these models. I compute the information criteria based on 10-year rolling window estimation scheme over the estimation period 1981:11-2007:12 for M_1, M_2, and M_0. The solid line in Panel A plots the differences in AIC between M_0 and M_2. The dotted line in Panel A plots the differences in AIC between M_0 and M_1. Panel B plots the results for SBIC. At each time period, positive values for both lines in the figure indicate that the maximally flexible model is preferred. Other than that, smaller values suggest the corresponding reduced-rank restricted model to be the preferred model. AIC and SBIC are negative for most of the sample, with the solid line for M_2 displaying more negative values than the dotted line for M_1. Based on these results, I choose the most plausible model at each time to forecast bond excess returns.[22]

Panels A and B of Table 3 presents the annualized RMSEs for the 36 horizon (h) and maturity ($n - h$) combinations of bond excess returns using unspanned macro models. I compare the forecasts from the preferred unspanned macro model M_{us} to those of model M_0. Panel C presents the RMSE ratios of the two models. An RMSE ratio less than one indicates that model M_{us} provides improvements in forecasting relative to M_0. The statistical significance of the relative forecasting performance is evaluated using the Diebold and Mariano (1995) t-statistics computed with a serial correlation robust variance and the small sample adjustment of Harvey, Leybourne, and Newbold (1997).[23] In general, model M_{us} outperforms M_0 except for short-maturity bonds at very short forecast horizons (e.g., $n - h =$ 1-, 3-, 6-month and $h =$ 1-month). The improvements in forecast performance are greatest for long forecast horizons (e.g., $h =$ 2- and 5-year). For the five-year forecast horizon, the improvement in the forecasting RMSEs from imposing risk premium restrictions is approximately 16% on average across maturities.

To provide an intuition on how the P-dynamics could differ under different risk premium restrictions, I report the in-sample parameter estimates for models M_1, M_2, and M_0 in Table 2. I also report the estimates for the nested model of M_0 that enforces the spanning of macro variables by the yield curve M_{0span} as a comparison. Standard errors are computed using the outer product of the gradient of the likelihood function, and they are given in parentheses.

[22]AIC and SBIC provide consistent comparison results for most of the sample. When there is a difference in the comparisons of the two criteria, SBIC are used, since AIC may asymptotically over fit.

[23]To ensure that the Diebold-Mariano test is valid for the setup, I examine the loss differential series based on the forecasts using the forecasting and standard loss functions. The augmented Dickey–Fuller tests reject the null of a unit root for all maturities at each forecast horizon.

Parameter estimates that are statistically significant are shown in boldface.

The maximum likelihood estimates of parameters governing the Q-distribution of the risk factors are displayed in Panel A. Consistent with the literature (JSZ, JLS, and JPS), the Q-estimates are virtually indistinguishable from the four models. The notable differences are observed in the estimated feedback matrix under P-measure: K_1^P. Panel B of Table 2 reports the moduli of the eigenvalues of K_1^P: $\left|\lambda^P\right|$. Comparing across models, I find that enforcing reduced-rank restrictions in the unspanned macro model increases the largest eigenvalue of K_1^P from 0.9876 (M_0) to 0.9989 (M_2). The high persistence in model M_2 is able to generate more variable expectations of future short rates, which is consistent with survey-based expectations of inflation and policy rates (Kim and Orphanides, 2012). In addition, the models with unspanned macro risks provide more persistent estimates than M_{span}, but only when risk premium restrictions are imposed.

The estimates of the risk premium parameters (λ_{0PC} and λ_{1PC}) are shown in Panel C of Table 2. Macro risks GRO and inflation are both priced, but the signs of the coefficients differ across the unspanned models. For example, shocks to GRO induce pro- (counter-) cyclical movements in the risk premia associated with exposures to $PC2$ under model M_0 (M_1 and M_2). The unconditional means of the risk prices for both level and slope are significant under all models. With risk premium restrictions, the risk price associated with level factor is negative.

I also note that the zero-restrictions on risk price parameters, which are generally imposed based on the t-statistics from in-sample estimation, are not consistent with the reduced-rank restrictions. For example, in model M_2, two sources of term structure risks are priced. However, the in-sample t-statistics indicates that all three risk factors are significantly priced. This result confirms the insight in Bauer (2016), which suggests that joint significance is needed when joint risk premium restrictions are imposed.

1.8 Out-of-Sample Forecasting Performance

This section describes the out-of-sample forecasting performance of models. I first report the forecasting results evaluated using SSE loss. Subsequently I present the results with economic evaluation. In the empirical exercise, the forecasts are generated based on a 10-year rolling window. The first estimation starts using data from 1971:11 to 1981:11, and the estimates are used to forecast returns h-period ahead of 1981:11. I proceed recursively with estimation and forecasting, each time moving one period ahead. I continue to update the estimation and forecasts in this way until 2007:12. For each model, 315 forecasts are produced through this way for each forecasting horizon h. As mentioned in Section 1.6.1, I focus on the performance of 36 horizon and maturity combinations of bond excess returns. The holding periods (h) and the maturities ($n-h$) examined in the paper are 1 month, 3, 6, 12, 24, and 60 months.

Seven different forecasts of bond excess returns are considered in the empirical studies: the forecasts from the preferred unspanned macro model ($M_{us}, \beta_h = 1$), the benchmark forecasts from the nested model that enforces the spanning restrictions ($M_{span}, \beta_h = 0$), and five forecast combinations using the two forecasts. I include the combination with equal weights as comparison ($M^{0.5}_{com}, \beta_h = 0.5$), because empirical studies often find that simple equal-weighted forecast combinations perform very well compared with more sophisticated combination schemes that rely on estimated combination weights (see Elliott and Timmermann, 2016; Timmermann, 2006). It is possible that errors introduced by estimation of the combination weights could overwhelm gains from using optimal weights over equal weights. In my case, the optimal combination weights by construction should not underperform the equal weights, since the estimation of the weights are directly aligned with forecaster's evaluation loss function. Moreover, the optimal weights are updated recursively, which accommodates the changes in relative model performance. This is empirically important as we shall see. Still, it may be interesting to see whether the optimal weights are statistically different from 0.5 and how large the gains in using optimal rather than equal weights. Another four combinations are constructed based on the estimated weights as discussed in Section 1.5. I consider both constant ($M_{com1}, \beta_h = \widehat{\beta}_h \in R$) and time-varying optimal weights ($M^{tv}_{com2}, \beta_{t,h} = \widehat{\beta}_{t,h} \in R$), and they can be constrained to $[0,1]$ ($M_{com3}, \beta_h = \widehat{\beta}_h \in [0,1]$; $M^{tv}_{com4}, \beta_{t.h} = \widehat{\beta}_{t,h} \in [0,1]$) or not.

1.8.1 Results for Sum Squared Error Loss

I first present the forecasting results evaluated using the SSE loss function as in equation (1.22). The RMSEs for the 36 horizon (h) and maturity ($n-h$) combinations of bond excess returns are reported in Table 4, each panel presents the results for one forecast horizon. Within each panel, each column corresponds to one model specification, and each row corresponds to one maturity. The annualized RMSEs in basis points for the benchmark spanned model are reported in the first column. For all the other specifications, percentage gain or loss relative to the spanned model are reported. Negative (Positive) values suggest that the alternative specification provide smaller (larger) out-of-sample RMSEs than the spanned model. The RMSEs are computed using the forecasts of annualized bond excess returns at different horizons. The statistical significance of the relative model performance is evaluated using the Diebold and Mariano (1995) t-statistics computed with a serial correlation robust variance and the small sample adjustment of Harvey, Leybourne, and Newbold (1997).

The forecasting performance of the preferred unspanned macro model M_{us} is not significantly different from M_{span} at short forecast horizons (1-, 3-, and 6-month), but model M_{us} outperforms model M_{span} significantly at long forecast horizons (1-, 2-, and 5-year). For the two- and five-year forecast horizons, M_{us} outperforms M_{span} for all maturities. For example, at the five-year horizon, the improvement of using unspanned macro information in forecasting returns on average across maturities is about 11%. In the forecasting literature, the out-of-sample R-square is often considered, which is defined as $1 - (MSE^{\text{Benchmark}}/MSE^{\text{Alternative}})$, where $MSE^{\text{Benchmark}}$ refers to the mean squared errors (MSE) of the benchmark spanned model and $MSE^{\text{Alternative}}$ to MSE of the alternative model. This gives an out-of-sample R-square of $1-(1-0.11)^2 = 21\%$. These findings are consistent with the theoretical evidence in the macroeconomics literature (Backus and Zin, 1993) which suggests that macro variables such as inflation has long memory effect on the term structure resulting from the aggregation across agents with heterogeneous beliefs in the monetary policy target.

Equal-weighted forecast combination $M_{com}^{0.5}$ provides significant gains for some maturities (e.g., $n-h=$ 1-, 6-, 12-, and 60-month) across different forecast horizons relative to model M_{span}. However, this combination forecasts underperform M_{us} at long forecast horizons.

The combination forecasts with constant optimal weights (M_{com1} and M_{com3}) provide better performance than M_{span}. At the five-year horizon, the improvements are significant for all maturities. Models M_{com1} and M_{com3} also provide better forecasts than equally weighted forecast combination $M_{com}^{0.5}$, especially at long forecast horizons.

I present the estimated constant optimal weights $\widehat{\beta}_h^n$ for the preferred unspanned model using SSE loss function in Panel A of Table 5. Larger optimal weights indicate that the unspanned macro information is more useful in forecasting yields. The $[0,1]$ restriction seems to be irrelevant for the estimation of $\widehat{\beta}_h^n$. The estimates are larger than one only at the five-year forecast horizon. I further assess whether these estimated weights are statistically different from 0.5 using Carriero and Giacomini (2011) tests. The optimal weights at long forecast horizons are statistically different from 0.5, except for very long-maturity yields at one- and two-year horizons.

The best performers in Table 4 are models M_{com2}^{tv} and M_{com4}^{tv}, combination forecasts based on time-varying weights. For long forecast horizons, models M_{com2}^{tv} and M_{com4}^{tv} provide significantly improvements across all maturities. For example, at the five-year horizon, the improvement of using unspanned macro information in forecasting returns on average across maturities is about 14% by using unconstrained time-varying combination weights (M_{com2}^{tv}), which corresponds to an out-of-sample R-square of 26%.

I also test whether the unspanned macro information is consistently useless or not in forecasting bond excess returns. I plot the estimated weights $\widehat{\beta}^n_{t,h} \in R$ and the 95% confidence bands for 9 horizon (h = 1-, 12-, and 60-month) and maturity ($n - h$ = 1-, 12-, and 60-month) combinations of bond excess returns in Figure 4.[24] The usefulness of unspanned macro information is not constant over time. The null hypothesis that $H_0 : \widehat{\beta}^n_{t,h} = 0$ can be rejected if there exists at least one t at which zero falls outside the confidence bands. The null hypothesis is rejected for all horizon and maturity combinations. In particular, for the five-year forecast horizon, $\widehat{\beta}^n_{t,h}$ is statistically different from zero through the entire sample, and statistically different from 0.5 for some period of the sample. Figure 4 also suggests that the unspanned macro information are more useful at the beginning and the end of the sample, especially for long forecast horizons. I obtain very similar results for $\widehat{\beta}^n_{t,h} \in [0, 1]$, which are reported in Figure A1.

In summary, I find that incorporating macro information generates significant gains in forecasting bond risk premia relative to yield curve information at long forecast horizons, especially when I allow the combination weight to be time varying. Moreover, the preferred unspanned model that takes into account risk-price restrictions provides better forecasting performance than the spanned model at long forecast horizons. This result is consistent with the literature, which documents that risk premium restrictions affect the expectation of future short rate and term premia (Cochrane and Piazessi, 2008; Bauer, 2016; JPS). I confirm the finding from an out-of-sample forecasting perspective. More importantly, this result suggests that the ability of unspanned macro variables to forecast bond excess returns cannot be fully captured in a regression sense.

1.8.2 Results for Mean-Variance Utility Loss

In this section, I investigate whether the statistical predictive power of the unspanned macro information can be used to improve the economic utility of a bond investor. The evaluation

[24]Due to space constraints, I only report the results for 9 forecast horizon and maturity combinations in this figure. For the other combinations considered in this paper, similar results are obtained.

loss function of interest in this case is specified in equation (1.24). The economic value of return forecasts is measured with certainty equivalent returns (CERs). Under mean-variance utility, the CER is computed simply as the average realized utility over the out-of-sample period (e.g., Cenesizoglu and Timmermann 2012). Using notation in Subsection 1.5.3, the CER for forecast horizon h is given by

$$CER_h = \frac{1}{G} \sum_{t=J+h}^{J+h+G} U(W_{t+h}^*). \tag{1.32}$$

Table 6 presents the annualized CERs for portfolio decisions based on out-of-sample forecasts of bond excess returns from seven different forecasts. I present the results for mean-variance investors with different levels of risk aversion $\gamma = 3$, 5, 10, which respectively represent low, moderately and highly risk averse investors. To test if the annualized CERs of the alternative models (M_{us}, $M_{com}^{0.5}$, M_{com1}, M_{com2}^{tv}, M_{com3}, M_{com4}^{tv}) are statistically different from the CERs of the benchmark model M_{span}, I use Diebold and Mariano (1995) t-statistics computed with a serial correlation robust variance and the small sample adjustment of Harvey, Leybourne, and Newbold (1997). The benchmark utility is generated by the predicted returns using only information in the yield curve, this is the null of the spanning hypothesis as in Bauer and Hamilton (2016). This benchmark differs from the no-predictability expectation hypothesis, which has been examined in Thornton and Valente (2012), Gargano, Pettenuzzo, and Timmermann (2015), and Sarno, Schneider, and Wagner (2014). Because the focus here is the economic gains from the macro information conditioning on current yield curve information.

I find significantly improved economic performance of the unspanned macro model M_{us} relative to the spanned model M_{span} at long forecast horizons (1-, 2-, and 5-year) for all risk aversion levels. The only exception is at the two-year forecast horizon for $\gamma = 10$. For equal-weighted combination forecasts, the improvements are modest, especially for $\gamma = 10$. For the forecast combination with optimal weights, the improvements are much greater at long forecast horizons. The best combination forecasts are obtained for model M^{tv}_{com2}, with unconstrained time-varying combination weights. For example, a low risk-averse investor can earn an annualized CER of 3.98% for a five-year investment horizon by using model M^{tv}_{com2} relative to an annualized CER of 1.46% by using model M_{span}. This finding is consistent with Gargano, Pettenuzzo, and Timmermann (2015), who show that accounting for model instability by recursively updating combination weights can improve the economic value of forecasting substantially.

Panel B of Table 5 presents the estimated weights $\widehat{\beta}_h$ for model M_{us} using mean-variance utility loss function. Consistent with the results in Panel A of Table 5, which is based on the SSE loss, the $[0, 1]$ constraint only matters at long forecast horizons. The larger weights at long forecast horizons generate better economic performance as shown in Table 6. For example, a low risk-averse investor can earn an annualized CER of 3.69% for a five-year investment horizon by using model M_{com1} relative to an annualized CER of 2.21% by using model M_{com3}. The weights at long forecast horizons are statistically different from 0.5 based on Carriero and Giacomini (2011) test. Recall from equation (1.25), the estimated weights under mean-variance utility are irrelevant to investor's risk aversion level when risky portfolio weights are not constrained.

Moreover, I investigate the usefulness of the unspanned macro information over time. Figure 5 plots the estimated weights $\widehat{\beta}_{t,h}$ and the 95% confidence bands for each forecast horizon. The null hypothesis that $H_0 : \widehat{\beta}^n_{t,h} = 0$ can be rejected if there exists at least one t at which zero falls outside the bands. The null hypothesis that model M_{us} is consistently useless for all forecast horizons is rejected. For long forecast horizons, this null hypothesis is rejected over the entire sample, and the null hypothesis that $\widehat{\beta}_{t,h}$ is consistently 0.5 is also rejected. At short forecast horizons (1-, 3-, and 6-month), the unspanned macro information is generally more useful after the recession periods. I also present the results for $\widehat{\beta}_{t,h} \in [0, 1]$ in Figure A2, similar conclusions are obtained.

The above analyses are based on unconstrained weights $\omega^*_{t,h}$ for the risky bonds in the asset allocation problem. To avoid implausible positions, I rule out short selling and also impose constraints on total portfolio leverage. The annualized CERs of the interested models under this situation are reported in Table 7. Each panel presents the results for one leverage ratio limit. The leverage ratios of 0, 50%, and 100% (e.g., $leverage = 50\%$ indicates that the investor cannot borrow more than 50% of his or her total wealth) are considered. In general, the numbers in Table 7 are smaller than those in Table 6 due to position constraints. The results in Table 7 confirm the results from Table 6. Model M_{us} provides higher CERs relative to model M_{span} at long forecast horizons for almost all risk aversion levels. For example, a high risk-averse investor with 50% leverage ratio constraint can earn an annualized CER of 1.08% for a five-year investment horizon by using model M_{us} relative to an annualized CER of 0.28% by using model M_{span}. The best performer is the forecast with unconstrained time-varying combination weights, M^{tv}_{com2}, for all risk aversion levels. For example, a moderate risk-averse investor with 50% leverage ratio constraint can earn an annualized CER of 2.15% for a one-year investment horizon by using model M^{tv}_{com2} relative to an annualized CER of 0.99% by using model M_{span}. These findings are consistent under different leverage restrictions.

Overall the results based on the portfolio utility loss function are consistent with those based on the SSE loss. The statistical forecast accuracy of the unspanned macro information can be used to improve the economic utilities of bond investors.

1.9 Robustness

To corroborate my findings, in this section, I perform various robustness analyses. I show that the conclusions on the predictive power of unspanned macro information are robust to power utility function and other performance measure that is widely used in the literature.

1.9.1 Power Utility

I firstly consider the economic values of the forecasts using unspanned macro information for an investor with constant relative risk aversion (CRRA) utility function

$$U(W_{t+h}) = \frac{W_{t+h}^{1-\gamma}}{1-\gamma}, \tag{1.33}$$

where $W_{t+h} = Y_{t+h}^h + \omega_{t,h} f_{t+h|t}$, and γ is the investor's coefficient of relative risk aversion. In a plausible portfolio choice model, absolute risk aversion should decline or, at the very least, should not increase with wealth, while relative risk aversion should be independent of wealth: this favors power utility (Campbell and Viceira, 2002). I therefore repeat the exercise in section 1.8.2 using power utility function. I focus on the constrained asset allocation problem, since I want to further confirm that the improvements in the economic value generated by the unspanned macro information is not driven by implausible trading positions. Under CRRA utility, for each forecast horizon h, the CER is computed as

$$CER_h = (1-\gamma)\frac{1}{G}\sum_{t=J+h}^{J+h+G} U(W_{t+h}^*)^{1/(1-\gamma)} - 1. \tag{1.34}$$

The annualized CERs for the interested models are reported in Table 8 with each panel corresponding to one leverage ratio limit. The results are similar to those for the mean-variance investors. The preferred unspanned model M_{us} provides higher CERs at long forecast horizons (1-, 2-, and 5-year) for almost all risk aversion levels. For example, a high risk-averse investor with 50% leverage ratio constraint can earn an annualized CER of 1.04% for a five-year investment horizon by using model M_{us} relative to an annualized CER of 0.24% by using model M_{span}. The best performer M_{com2}^{tv} provides significantly larger CERs than M_{span} at long forecast horizons for all risk aversion levels. For example, a moderate risk-averse investor with 50% leverage ratio constraint can earn an annualized CER of 1.90% for a one-year investment horizon by using model M_{com2}^{tv} relative to an annualized CER of

0.70% by using model M_{span}. These results are robust to different leverage ratios. Overall, both mean-variance investors and CRRA investors can achieve significantly economic gains at long forecast horizons by using the unspanned macro information in their prediction of bond excess returns.

1.9.2 Performance Measure

Instead of using the CER to measure the economic value generated by the forecasts with unspanned macro variables, I follow Thornton and Valente (2012) and Sarno, Schneider, and Wagner (2014) to compute the manipulation-proof performance measure (MPPM) for the forecasts under consideration (Goetzmann, Ingersoll, Spiegel, and Welch, 2007). Following the notation in Subsection 1.5.3, the MPPM at forecast horizon h for the alternative model relative to the benchmark spanned model is defined as

$$\begin{aligned} \text{MPPM}_h &= \frac{12}{(1-\gamma)h}\log\left(\frac{1}{G}\sum_{t=J+h}^{J+G}\left[\frac{1+Y_{t+h}^h+\omega_{t,h}^* f_{t+h|t}}{1+Y_{t+h}^h}\right]^{1-\gamma}\right) \\ &\quad -\frac{12}{(1-\gamma)h}\log\left(\frac{1}{G}\sum_{t=J+h}^{J+G}\left[\frac{1+Y_{t+h}^h+\omega_{t,h}^* \widehat{rx}_{t+h|t}^{M_{span}}}{1+Y_{t+h}^h}\right]^{1-\gamma}\right). \end{aligned} \tag{1.35}$$

The MPPM_h denotes the annualized risk-adjusted return earned by the portfolio based on forecast $f_{t+h|t}$ in excess of forecast from the spanned model at forecast horizon h. In contrast to another commonly used measure, Sharpe ratio, MPPM can alleviate concerns related to non-normal distribution of bond returns. Most importantly, Sharpe ratio is more likely to subject to manipulation in various ways as discussed in Goetzmann, Ingersoll, Spiegel, and Welch (2007).

Table 9 reports the annualized MPPM in percentage for mean-variance investors under different leverage constraints. Model M_{us} in general provides higher MPPM than M_{span} at long forecast horizons. Guiding investment decisions using combination forecasts with time-varying weights M_{com2}^{tv} results in the largest excess return to mean-variance investors, especially at long forecast horizons. For example, M^{tv}_{com2} generates annualized excess return of 0.95% for low risk averse investor at the five-year horizon with 50% leverage constraint. These results are robust to risk aversion levels and leverage ratio constraints. The conclusion that unspanned macro information helps forecasting bond excess returns is robust to MPPM.

1.10 Conclusion

In this paper, I study the predictability of bond risk premia using macro information in the new family of unspanned MTSMs. I examine the out-of-sample performance of the MTSM with unspanned output and inflation risks from the perspective of a bond investor. I account for model uncertainty by combining the forecasts with and without unspanned macro information optimally for the forecasting objective, both statistical and economic measures of forecasting accuracy are considered.

I also conduct a novel model selection method to impose risk premium restrictions on the unspanned macro model for the purpose of forecasting. Using the preferred unspanned model in the forecast combination provides substantially better forecasting performance than the model with only yield curve information at long forecast horizons, both statistically and economically. The improvement is more pronounced when allowing for time-varying combination weight.

I perform formal statistical tests on the combination weight, which can be seen as a measure of the usefulness of the unspanned macro information in forecasting bond excess returns. The null hypothesis: unspanned macro information is consistently useless for forecasting bond risk premia, can be rejected for all the horizons and maturities under consideration, both statistically and economically. Overall, this study presents empirical evidence against the spanning hypothesis in an out-of-sample context.

This work can be extended in several directions. First, the proposed framework can be applied to examine the usefulness of other unspanned return-forecasting factors. This extension can offer insights on how bond risk premia relates to the information on other financial assets. Second, I can extend this framework to international bond markets and other interest rate instruments, which could lead to new visions regarding the comovement of the yield curve across different currencies and segments of the fixed income market. Third, my study seems to suggest the need for a model with both spanned and unspanned macro risks. Such a model might be sufficient to answer most questions about macro-yield interactions. I leave this open question for future research.

Appendices

A Affine Bond Pricing

Under canonical MTSMs, the stochastic discount factor (SDF), which defines the change of probability measure between P and Q, is specified as exponentially affine

$$-\log m_{t+1} = r_t + \frac{1}{2}\lambda_t'\lambda_t + \lambda_t'\varepsilon_{t+1}^P. \tag{A.1}$$

The time-varying market prices of risk are evolving with the risk factors as

$$\lambda_t = \Sigma^{-1}(\lambda_0 + \lambda_1 Z_t), \tag{A.2}$$

This is the essentially affine risk price specification of Duffee (2002). The risk price parameters are given by

$$\begin{aligned}\lambda_0 &= K_0^P - K_0^Q, \\ \lambda_1 &= K_1^P - K_1^Q.\end{aligned} \tag{A.3}$$

The model-implied continuously compounded yields are given by equation (2.7). Premultiply equation (2.7) with a $k \times N$ matrix, W^k, to select k linear combinations of model-implied yields, $PO_t^k = W^k A(\Theta^Q) + W^k B(\Theta^Q) Z_t$. This equation can be inverted for Z_t, and in particular for macro factors

$$M_t = \delta_0 + \delta_1 PO_t^k. \tag{A.4}$$

The above equation implies that macro variables are essentially a linear function of yields.

The model-implied price of n-maturity bond at time t is exponentially affine function of risk factors (Duffie and Kan, 1996)

$$\widehat{P}_t^n = \exp(\overline{A}_n(\Theta^Q) + \overline{B}_n(\Theta^Q) Z_t),$$

where $\overline{A}_n$ is a function of $\{K_0^Q, K_1^Q, \rho_0, \rho_1, \Sigma\}$, and $\overline{B}_n$ is a function of $\{K_1^Q, \rho_1\}$ following the recursive relations

$$\overline{A}_n = -\rho_0 + \overline{A}_{n-1} + \overline{B}_{n-1} K_0^Q + \frac{1}{2} \overline{B}_{n-1} \Sigma \Sigma' \overline{B}'_{n-1}, \tag{A.5}$$

$$\overline{B}_n = -\rho_1 + \overline{B}_{n-1} K_1^Q, \tag{A.6}$$

with starting values $\overline{A}_1 = -\rho_0$, and $\overline{B}_1 = -\rho_1$. The model-implied continuously compounded n-maturity yield $\widehat{Y}_t^n$ is given by

$$\widehat{Y}_t^n = -\frac{log \widehat{P}_t^n}{n} = -\frac{\overline{A}_n(\Theta^Q) + \overline{B}_n(\Theta^Q) Z_t}{n}. \tag{A.7}$$

Match the above equation to equation (2.7), $A_n(\Theta^Q) = -\frac{\overline{A}_n(\Theta^Q)}{n}$, the n^{th} element of the $N \times 1$ vector $A(\Theta^Q)$, and $B_n(\Theta^Q) = -\frac{\overline{B}_n(\Theta^Q)}{n}$, the n^{th} row of the $N \times k$ matrix $B(\Theta^Q)$. Recall that k is the number of risk factors.

B Model's Implications on Bond Risk Premia

The unspanned macro model implies different estimates of P-parameters, $\Theta^P = \{K_0^P, K_1^P\}$ from the spanned model, or alternatively different estimates of market prices of risks (λ_{PCt}), since the two models have the same Q-distribution. I consider the one-period risk premium for n-maturity bond as an example to provide some intuition on the model-implied risk premia. Using equations (A.5), (A.6), and (1.16), I can rewrite the one-period expected excess returns as follows

$$
\begin{aligned}
E_t\left[\widehat{rx}_{t+1}^n\right] &= \overline{A}_{n-1} - \overline{A}_n + \overline{A}_1 + \overline{B}_{n-1}\widehat{Z}_{t+1|t} - (\overline{B}_n - \overline{B}_1)Z_t, \qquad \text{(B.1)}\\
&= \overline{A}_{n-1} - (-\rho_0 + \overline{A}_{n-1} + \overline{B}_{n-1}K_0^Q + \frac{1}{2}\overline{B}_{n-1}\Sigma\Sigma'\overline{B}_{n-1}') - \rho_0 + \overline{B}_{n-1}\widehat{Z}_{t+1|t} - \\
&\quad (-\rho_1 + \overline{B}_{n-1}K_1^Q + \rho_1)Z_t,\\
&= -\frac{1}{2}\overline{B}_{n-1}\Sigma\Sigma'\overline{B}_{n-1}' + \overline{B}_{n-1}\widehat{Z}_{t+1|t} - \overline{B}_{n-1}(K_0^Q + K_1^Q Z_t),\\
&= -\frac{1}{2}\overline{B}_{n-1}\Sigma\Sigma'\overline{B}_{n-1}' + \overline{B}_{n-1}(\widehat{Z}_{t+1|t} - \widehat{Z}_{t+1|t}^Q)],\\
&= -\frac{1}{2}\overline{B}_{n-1}\Sigma\Sigma'\overline{B}_{n-1}' + \overline{B}_{n-1}\Sigma\lambda_{PCt}.
\end{aligned}
$$

The first term corresponds to $\frac{1}{2}var_t(\widehat{rx}_{t+1}^n)$, conditional variance of excess returns, which is due to convexity. The second term captures the actual risk compensation for level, slope, and curvature risks. This can be rewritten as $\lambda_{PCt}Cov_t(\varepsilon_{t+1}^P, \widehat{rx}_{t+1}^n)$, the product of the prices of risks and the quantities of risks. The risk prices λ_{PCt} measure the additional expected return required per unit of risk in each of the shocks in ε_{t+1}^P. In Gaussian term structure model, quantities of risks are constant, and time-variation in expected returns is due exclusively to movements in λ_{PCt}, which depends on the dynamics of the risk factors as in equation (1.10). Parameters λ_{0PC} and λ_{1PC} determine the risk adjustments. Under the weak expectation hypothesis, λ_{1PC} is zero, expected excess returns and term premia are constant.

The model-implied bond risk premia as in equation (1.16) can be rewritten as

$$E_t\left[\widehat{rx}^n_{t+h}\right] = \overline{A}_{n-h}(\Theta^Q) - \overline{A}_n(\Theta^Q) + \overline{A}_h(\Theta^Q) + \overline{B}_{n-h}(\Theta^Q)\widehat{Z}_{t+h|t} - (\overline{B}_n(\Theta^Q) - \overline{B}_h(\Theta^Q))Z_t, \quad \text{(B.2)}$$

The fitted excess return as in equation (1.15) can be computed as follows

$$\widehat{rx}^n_{t+h} = \overline{A}_{n-h}(\Theta^Q) - \overline{A}_n(\Theta^Q) + \overline{A}_h(\Theta^Q) + \overline{B}_{n-h}(\Theta^Q)Z_{t+h} - (\overline{B}_n(\Theta^Q) - \overline{B}_h(\Theta^Q))Z_t, \quad \text{(B.3)}$$

where $\overline{A}_n(\Theta^Q) = -nA_n(\Theta^Q)$ and $\overline{B}_n(\Theta^Q) = -nB_n(\Theta^Q)$. $A_n(\Theta^Q)$ is the n^{th} element of the $N \times 1$ vector $A(\Theta^Q)$, and $B_n(\Theta^Q)$ is the n^{th} row of the $N \times k$ matrix $B(\Theta^Q)$. Using equations (B.2) and (B.3), I can obtain the surprise component of the excess return as in Duffee (2011a)

$$\widehat{rx}^n_{t+h} - E_t\left[\widehat{rx}^n_{t+h}\right] = \overline{B}_{n-h}(\Theta^Q)(Z_{t+h} - \widehat{Z}_{t+h|t}) = \overline{B}_{n-h}(\Theta^Q)\sum_{i=1}^{h}(K_1^P)^{h-1}\Sigma\varepsilon^P_{t+i}. \quad \text{(B.4)}$$

C Estimation of Reduced-Rank Regressions

The constraints on the rank of the risk price matrix can be mapped into the parameter space. I impose restrictions on the rank (ζ) of $\lambda_{1PC} = \left[K^P_{PC}, K^P_{PCM}\right] - \left[K^Q_{1PC}, 0_{l\times m}\right]$ to be $\zeta = 1$, 2, or 3. Following Joslin, Singleton and Zhu (2011), I start with singular value decomposition of λ_{1PC}, UDV', where U is $l \times l$ unitary matrix, V is $k \times k$ unitary matrix, and D is $l \times k$ rectangular diagonal matrix with the diagonal sorted in decreasing order. Equation (1.18) can be rewritten as follows

$$xPC_t = \left(K^P_{0PC} - K^Q_{0PC}\right) + U_{\bullet 1}\left(D_{11}\sum_{l=1}^{k} V'_{l1} Z_{lt}\right) + U_{\bullet 2}\left(D_{22}\sum_{l=1}^{k} V'_{l2} Z_{lt}\right) + U_{\bullet 3}\left(D_{33}\sum_{l=1}^{k} V'_{l3} Z_{lt}\right). \tag{C.1}$$

where $U_{\bullet 1}$, $U_{\bullet 2}$, $U_{\bullet 3}$ are columns of U. The time-varying components of xPC_t are linear combination of pricing factors Z_t. All parameters in equation (C.1) are econometrically identified by using the properties of the unitary matrix, $U'U = I_l$, and $V'V = I_k$. In our empirical studies, $l = 3$ and $k = 5$. For reduce rank ζ, I set the singular values of D to 0 for $l > \zeta$ and $k > \zeta$, this corresponds to $(l - \zeta)(k - \zeta)$ cross-equation restrictions on the parameters of λ_{1PC}. When $\zeta = 1$, eight restrictions on risk premium parameters are imposed. When $\zeta = 2$, three restrictions on risk premium parameters are imposed.

To estimate the regression in equation (1.29) with reduced-rank constraint, I consider the regression in general form $y_t = \theta_0 + \theta_1 Z_t + \epsilon_t$ for notation simplicity. θ_1 is subject to have rank ζ. $\epsilon_t \sim N(0, \Sigma)$ i.i.d. with Σ known. I need to solve the following optimization

$$(\theta_0, \theta_1) = \underset{rank(\theta_1)=\zeta}{\arg\min} \sum_t (y_t - \theta_0 - \theta_1 Z_t)' \Sigma\Sigma' (y_t - \theta_0 - \theta_1 Z_t) \tag{C.2}$$

Without loss of generality, I first de-mean the variables and assume $\theta_0 = 0$. Furthermore, by transforming the variables, I assume $\Sigma = I$ and $\sum_t Z_t Z_t' = I$. With these assumptions, the optimization problem can be expressed as follows

$$\begin{aligned}
\theta_1 &= \underset{rank(\theta_1)=\zeta}{\arg\min}\, trace\left((y-\theta_1 Z)'(y_t-\theta_1 Z)\right) \qquad \text{(C.3)}\\
&= \underset{rank(\theta_1)=\zeta}{\arg\min}\, trace\left((y-\theta_1^{OLS} Z)'(y_t-\theta_1^{OLS} Z)\right) - 2trace\left(Z(y_t-\theta_1^{OLS}Z)'(\theta_1-\theta_1^{OLS})\right)\\
&+ trace\left(ZZ'(\theta_1'-\theta_1'^{OLS})(\theta_1-\theta_1^{OLS})\right)\\
&= \underset{rank(\theta_1)=\zeta}{\arg\min}\, \left\|\theta_1-\theta_1^{OLS}\right\|_F,
\end{aligned}$$

where y and Z are $(l \times T)$ and $(k \times T)$ matrices with the time series stacked horizontally, $\theta_1^{OLS} = (Z'Z)^{-1}Z'y$, and F denotes the Frobenius norm: $\|A\|_F^2 = \sum_{i,j} |A_{i,j}|^2$. As in Keller (1962), this minimization problem has solution $\theta_1^* = UD_\zeta^* V'$, where D_ζ^* is the same as D except for setting all of the singular values for $l > \zeta$ and $k > \zeta$ to 0.

D Statistical Inference of the Combination Weights

I provide the main framework to derive the statistical inference of the estimated combination weights. For more details, I refer to Carriero and Giacomini (2011). The asymptotic standard deviation that is needed to construct the statistical inference are given by

$$\widehat{\sigma}_h = \sqrt{\widehat{H}_h^{-1}\widehat{\Xi}\widehat{H}_h^{-1}}. \tag{D.1}$$

Following the notation in Section 1.5, I have

$$\widehat{H}_h^{-1} = E\left[\bigtriangledown_{\beta_h\beta_h}\Gamma_h(\widehat{\beta}_h)\right],$$

$$\widehat{\Xi}_h = \sum_{j=-p_{n+1}}^{p_{n-1}}(1-\left|\frac{j}{p_n}\right|)G^{-1}\sum_{t=J+j}^{J+G}s_t(\widehat{\beta}_h)s_{t-j}(\widehat{\beta}_h),$$

$$s_t(\widehat{\beta}_h) = \bigtriangledown_{\beta_h}\Gamma_h(\widehat{\beta}_h).$$

$\Gamma_h(\widehat{\beta}_h) = E[L(rx_{t+h}, f_{t+h|t})]$, L can be the SSE loss or the mean-variance utility loss. $\bigtriangledown_{\beta_h}$ and $\bigtriangledown_{\beta_h\beta_h}$indicate the first and second derivatives with respect to β_h, the bandwidth is $p_n = 2(h-1)$.

To be more specific, for the SSE loss, the asymptotic standard deviation is derived as follows

$$\widehat{\sigma}_h = \left(2G\sum_{t=J}^{J+G}\left(\widehat{rx}_{t+h|t}^{M_{us}} - \widehat{rx}_{t+h|t}^{M_{span}}\right)^2\right)^{-1} \times \tag{D.2}$$
$$\sqrt{\sum_{j=-p_{n+1}}^{p_{n-1}}(1-\left|\frac{j}{p_n}\right|)G^{-1}\sum_{t=J+j}^{J+G}\left(\widehat{rx}_{t+h|t}^{M_{us}} - \widehat{rx}_{t+h|t}^{M_{span}}\right)\widehat{\eta}_{t+h}\left(\widehat{rx}_{t+h-j|t-j}^{M_{us}} - \widehat{rx}_{t+h-j|t-j}^{M_{span}}\right)\widehat{\eta}_{t+h-j}}.$$

Under the mean-variance framework, the classic solution of $\omega_{t,h}$ is given by

$$\omega_{t,h}^* = \frac{f_{t+h|t}^{'}\Omega^{-1}}{\gamma}, \tag{D.3}$$

where $\Omega_{t+h|t}$ is the $\widetilde{N} \times \widetilde{N}$ variance-covariance matrix of $f_{t+h|t}$. The term structure models considered in this paper do not specify the dynamics of conditional variance, I therefore use the unconditional variance-covariance matrix Ω. Given $W_{t+h} = Y^h_{t+h} + \omega_{t,h} rx_{t+h}$, the utility loss in equation (1.24) can then be computed as

$$\begin{aligned} L(rx_{t+h}, f_{t+h|t}) &= -Y^h_{t+h} - \omega^*_{t,h} f_{t+h} + \frac{\gamma}{2}\omega^*_{t,h}\Omega\omega^{*'}_{t,h}, \\ &= -Y^h_{t+h} - \frac{f'_{t+h|t}\Omega^{-1}}{\gamma} rx_{t+h} + \frac{f'_{t+h|t}\Omega^{-1}f_{t+h|t}}{2\gamma}. \end{aligned} \tag{D.4}$$

With estimated optimal weight $\widehat{\beta}_h$, $\Gamma_h(\widehat{\beta}_h)$ is given by

$$\begin{aligned} \Gamma_h(\widehat{\beta}_h) = G^{-1}\sum_{t=J+j}^{J+G}\left[-Y^h_{t+h} - \frac{\left(\widehat{\beta}_h\widehat{rx}^{M_{us}}_{t+h|t} + (1-\widehat{\beta}_h)\widehat{rx}^{M_{span}}_{t+h|t}\right)'\Omega^{-1}}{\gamma} rx_{t+h}\right] + \\ G^{-1}\sum_{t=J+j}^{J+G}\left[\frac{\left(\widehat{\beta}_h\widehat{rx}^{M_{us}}_{t+h|t} + (1-\widehat{\beta}_h)\widehat{rx}^{M_{span}}_{t+h|t}\right)'\Omega^{-1}\left(\widehat{\beta}_h\widehat{rx}^{M_{us}}_{t+h|t} + (1-\widehat{\beta}_h)\widehat{rx}^{M_{span}}_{t+h|t}\right)}{2\gamma}\right]. \end{aligned} \tag{D.5}$$

The asymptotic variance for the utility loss function is obtained using

$$\begin{aligned} s_t(\widehat{\beta}_h) = G^{-1}\sum_{t=J+j}^{J+G}\left[-Y^h_{t+h} - \frac{\Omega^{-1}}{\gamma} rx_{t+h}(\widehat{rx}^{M_{us}}_{t+h|t} - \widehat{rx}^{M_{span}}_{t+h|t})\right] + \\ G^{-1}\sum_{t=J+j}^{J+G}\left[\frac{\left(\widehat{\beta}_h\widehat{rx}^{M_{us}}_{t+h|t} + (1-\widehat{\beta}_h)\widehat{rx}^{M_{span}}_{t+h|t}\right)'}{2\gamma}\Omega^{-1}(\widehat{rx}^{M_{us}}_{t+h|t} - \widehat{rx}^{M_{span}}_{t+h|t})\right], \end{aligned} \tag{D.6}$$

and

$$\frac{\partial s_t(\widehat{\beta}_h)}{\partial \beta_h} = G^{-1}\sum_{t=J+j}^{J+G}\left[\frac{\left(\widehat{rx}^{M_{us}}_{t+h|t} - \widehat{rx}^{M_{span}}_{t+h|t}\right)'\Omega^{-1}(\widehat{rx}^{M_{us}}_{t+h|t} - \widehat{rx}^{M_{span}}_{t+h|t})}{2\gamma}\right]. \tag{D.7}$$

The test statistic is given by

$$t_h = \frac{\sqrt{G}(\widehat{\beta}_h - \beta_h)}{\widehat{\sigma}_h}. \tag{D.8}$$

Then the null hypothesis $\widehat{\beta}_h = \beta_h$ can be rejected at a given significance level α when $|t_h| > c_{\alpha/2}$, with $c_{\alpha/2}$ indicating critical value of a $N(0,1)$ distribution. Recall that $\widehat{\beta}_h$ is maturity specific for the SSE loss, I therefore compute $\widehat{\sigma}_h$ and t_h for each bond maturity.

The confidence bands of the time-varying optimal weights $\widehat{\beta}_{t,h}$ for a given significance level α are the following

$$(\widehat{\beta}_{t,h} - k_{\alpha,d}\frac{\widehat{\sigma}_h}{\sqrt{d}}, \widehat{\beta}_{t,h} + k_{\alpha,d}\frac{\widehat{\sigma}_h}{\sqrt{d}}), \text{for} \quad t = J + d + h - 1, ..., J + G. \tag{D.9}$$

In the empirical exercise, I choose a smoothing window with size $d = G \times 0.2$, which corresponds to a critical value of 3.179 for 95% confidence bands, according to Carriero and Giacomini (2011). Recall that G is the length of the out-of-sample period. The null hypothesis that the unspanned forecast is consistently useless ($H_0 : \widehat{\beta}_{t,h} = 0$) or the spanned forecast is consistently useless ($H_0 : \widehat{\beta}_{t,h} = 1$) can be rejected if there exists at least one t at which 1 or 0 falls outside the bands respectively.

E The JSZ Canonical Form

The unspanned macro model by JPS and the spanned JSZ model have the same Q-dynamics, this appendix provides the mapping of $\Theta^Q = \{K_0^Q, K_1^Q, \rho_0, \rho_1, \Sigma\}$ to $\Theta^Q = \{r_\infty^Q, \lambda^Q, \Sigma\}$. I focus on a simple case where the eigenvalues of K_1^Q are real, distinct, and nonzero. This follows Joslin, Singleton and Zhu (2011), who demonstrate the result for all cases including zero, repeated and complex eigenvalues.

As specified in Section 1.3.1, the model-implied continuously compounded yields $\widehat{Y}_t$ are given by

$$\widehat{Y}_t = A(\Theta^Q) + B(\Theta^Q) Z_t.$$

$Z_t = PC_t^l$ for both models. The first l principal components of the observed yield curve, $PC_t^l = W^l Y_t$, are priced without error, usually $l = 3$. I denote the first three principal components as PC_t.

There exists a matrix C such that $K_1^Q = Cdiag(\lambda^Q)C^{-1} + I_l$. Define $D = Cdiag(\rho_1)C^{-1}$, $D^{-1} = Cdiag(\rho_1)^{-1}C^{-1}$ and

$$\begin{aligned} L_t &= D(PC_t + (K_1^Q - I_l)^{-1} K_0^Q), \\ &\Rightarrow PC_t = D^{-1} L_t - (K_1^Q - I_l)^{-1} K_0^Q. \end{aligned}$$

Then the dynamic of L_t under the Q-measure is given by

$$\begin{aligned} \Delta L_{t+1} &= D \Delta PC_{t+1} \qquad \text{(E.1)} \\ &= D[K_0^Q + (K_1^Q - I_l)(D^{-1} L_t - (K_1^Q - I_l)^{-1} K_0^Q) + \Sigma \varepsilon_{t+1}^P] \\ &= diag(\lambda^Q) L_t + D \Sigma \varepsilon_{t+1}^P, \end{aligned}$$

and the dynamic of L_t under the P-measure is as follows

$$\begin{aligned}\Delta L_{t+1} &= D\Delta PC_{t+1} \qquad \text{(E.2)}\\ &= D[K_0^P + (K_1^P - I_l)(D^{-1}L_t - (K_1^Q - I_l)^{-1}K_0^Q) + \Sigma\varepsilon_{t+1}^P]\\ &= DK_0^P + D(K_1^P - I_l)D^{-1}L_t - D(K_1^P - I_l)(K_1^Q - I_M)^{-1}K_0^Q + D\Sigma\varepsilon_{t+1}^P.\end{aligned}$$

The dynamic of the short rate is

$$\begin{aligned}r_t &= \rho_0 + \rho_1 PC_t \qquad \text{(E.3)}\\ &= \rho_0 + \rho_1(D^{-1}L_t - (K_1^Q - I_l)^{-1}K_0^Q)\\ &= \rho_0 - \rho_1(K_1^Q - I_l)^{-1}K_0^Q + \rho_1 D^{-1}L_t\\ &= r_\infty^Q + \tau L_t,\end{aligned}$$

where $r_\infty^Q = \rho_0 - \rho_1(K_1^Q - I_l)^{-1}K_0^Q$, and τ is a row of l ones. Given the dynamics in equations (E.1)-(E.3), the model-implied continuously compounded yields $\widehat{Y}_t$ are given by

$$\widehat{Y}_t = A(\Theta_L^Q) + B(\Theta_L^Q)L_t,$$

where $\Theta_L^Q = \{r_\infty^Q, \lambda^Q, \Sigma\}$. The first l principal components of the observed yield curve are perfectly priced and can be written as

$$\begin{aligned}PC_t &= WY_t\\ &= W(A(\Theta_L^Q) + B(\Theta_L^Q)L_t)\end{aligned}$$

If the model is non-redundant, $WB(\Theta_L^Q)$ is invertible,

$$L_t = (WB(\Theta_L^Q))^{-1}PC_t - (WB(\Theta_L^Q))^{-1}WA(\Theta_L^Q).$$

Then the dynamic of PC_t under the Q-measure can be rewritten as follows

$$\begin{aligned}\Delta PC_{t+1} &= WB(\Theta_L^Q)\Delta L_{t+1} \\ &= WB(\Theta_L^Q)(diag(\lambda^Q)L_t + D\Sigma\varepsilon_{t+1}^P) \\ &= WB(\Theta_L^Q)\{diag(\lambda^Q)[(WB(\Theta_L^Q))^{-1}PC_t - (WB(\Theta_L^Q))^{-1}WA(\Theta_L^Q)] + D\Sigma\varepsilon_{t+1}^P\}.\end{aligned} \tag{E.4}$$

Comparing the coefficients in equations (E.4) and (1.6),

$$\begin{aligned}K_1^Q &= WB(\Theta_L^Q)diag(\lambda^Q)(WB(\Theta_L^Q))^{-1} + I_l, \\ K_0^Q &= -WB(\Theta_L^Q)diag(\lambda^Q)(WB(\Theta_L^Q))^{-1}WA(\Theta_L^Q).\end{aligned} \tag{E.5}$$

The dynamic of the short rate can be rewritten as

$$\begin{aligned}r_t &= r_\infty^Q + \tau L_t \\ &= r_\infty^Q + \tau[(WB(\Theta_L^Q))^{-1}PC_t - (WB(\Theta_L^Q))^{-1}WA(\Theta_L^Q)] \\ &= r_\infty^Q - \tau(WB(\Theta_L^Q))^{-1}WA(\Theta_L^Q) + \tau(WB(\Theta_L^Q))^{-1}PC_t.\end{aligned} \tag{E.6}$$

Comparing the coefficients in equations (E.6) and (1.5),

$$\begin{aligned}\rho_0 &= r_\infty^Q - \tau(WB(\Theta_L^Q))^{-1}WA(\Theta_L^Q), \\ \rho_1 &= \tau(WB(\Theta_L^Q))^{-1}.\end{aligned} \tag{E.7}$$

Bibliography

[1] Ajello, A., Benzoni, L. , and Chyruk, O. (2012). Core and 'crust': consumer prices and the term structure of interest rates. Working paper, Federal Reserve Bank of Chicago.

[2] Akaike, H. (1973). Information theory and an extension of the likelihood principle, in Boris Petrov and Frigyes Csa´ki, eds.: Proceedings of the Second International Symposium of Information Theory (Akademiai Kiado, Budapest).

[3] Almeida, C., Graveline, J. J., and Joslin, S. (2011). Do interest rate options contain information about excess returns? Journal of Econometrics, 164(1), 35-44.

[4] Ang, A., Boivin, J., Dong, S., and Loo-Kung, R. (2011). Monetary policy shifts and the term structure. The Review of Economic Studies, 78(2), 429-457.

[5] Ang, A., Dong, S., and Piazzesi, M. (2007). No-arbitrage Taylor rules. National Bureau of Economic Research.

[6] Ang, A. ,and Piazzesi, M. (2003). A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables. Journal of Monetary Economics, 50(4), 745-787.

[7] Backus, D., and Zin, S. E. (1993). Long-memory inflation uncertainty: Evidence from the term structure of interest rates. Journal of Money Credit and Banking, 25(3), 681-700.

[8] Bauer, M. D. (2016). Restrictions on risk prices in dynamic term structure models. Journal of Business & Economic Statistics, Forthcoming.

[9] Bauer, M. D., and Hamilton, J. D. (2016). Robust bond risk premia. Working paper, Federal Reserve Bank of San Francisco.

[10] Bauer, M. D., and Rudebusch, G. D. (2013). Monetary policy expectations at the zero lower bound. Working paper, Federal Reserve Bank of San Francisco.

[11] Bauer, M. D., and Rudebusch, G. D. (2016). Resolving the spanning puzzle in macro-finance term structure models. Review of Finance, Forthcoming.

[12] Bernanke, B., Reinhart, V., and Sack, B. (2004). Monetary policy alternatives at the zero bound: An empirical assessment. Brookings papers on economic activity, 2004(2), 1-100.

[13] Bikbov, R., and Chernov, M. (2008). Monetary policy regimes and the term structure of interest rates. Working paper.

[14] Bikbov, R. and Chernov, M. (2010). No-Arbitrage macroeconomic determinants of the yield curve, Journal of Econometrics, 159, 166-182.

[15] Boivin, J., and Giannoni, M. P. (2006). Has monetary policy become more effective? The Review of Economics and Statistics, 88(3), 445-462.

[16] Byrne, J. P., Cao, S., and Korobilis, D. (2015). Term structure dynamics, macro-finance factors and model uncertainty. Working paper, University of Glasgow.

[17] Campbell, J. Y., and Viceira, L. M. (2002). Strategic asset allocation: portfolio choice for long-term investors. Oxford University Press, USA.

[18] Carriero, A., and Giacomini, R. (2011). How useful are no-arbitrage restrictions for forecasting the term structure of interest rates? Journal of Econometrics, 164(1), 21-34.

[19] Cenesizoglu, T., and Timmermann, A. (2012). Do return prediction models add economic value? Journal of Banking and Finance, 36(11), 2974-2987.

[20] Chernov, M., and Mueller, P. (2012). The term structure of inflation expectations. Journal of Financial Economics, 106(2), 367-394.

[21] Christensen, J. H., Diebold, F. X., and Rudebusch, G. D. (2011). The affine arbitrage-free class of Nelson–Siegel term structure models. Journal of Econometrics, 164(1), 4-20.

[22] Cieslak, A., and Povala, P. (2015). Expected returns in Treasury bonds. Review of Financial Studies, 28(10), 2859-2901.

[23] Cochrane, J. H., and Piazzesi, M. (2005). Bond risk premia. American Economic Review, 95(1), 138-160.

[24] Cochrane, J., and Piazzesi, M. (2008). Decomposing the yield curve. Working paper, University of Chicago.

[25] Cogley, T. (2005). Changing beliefs and the term structure of interest rates: Cross-equation restrictions with drifting parameters. Review of Economic dynamics, 8(2), 420-451.

[26] Cogley, T., and Sargent, T. J. (2001). Evolving post-world war II US inflation dynamics. In NBER Macroeconomics Annual 2001, 16, 331-388. MIT Press.

[27] Cogley, T., and Sargent, T. J. (2005). Drifts and volatilities: monetary policies and outcomes in the post WWII US. Review of Economic dynamics, 8(2), 262-302.

[28] Cogley, T., Primiceri, G. E., and Sargent, T. J. (2010). Inflation-gap persistence in the US. American Economic Journal: Macroeconomics, 2(1), 43-69.

[29] Cooper, I., and Priestley, R. (2009). Time-varying risk premiums and the output gap. Review of Financial Studies, 22(7), 2801-2833.

[30] Coroneo, L., Giannone, D., and Modugno, M. (2016). Unspanned macroeconomic factors in the yield curve. Journal of Business & Economic Statistics, 34(3),472-485.

[31] Dai, Q., and Singleton, K. J. (2002). Expectation puzzles, time-varying risk premia, and affine models of the term structure. Journal of financial Economics, 63(3), 415-441.

[32] Della Corte, P., Sarno, L., and Thornton, D. L. (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.

[33] De Pooter, M., Ravazzolo, F., and Van Dijk, D. J. (2010). Term structure forecasting using macro factors and forecast combination. FRB International Finance Discussion Paper, (993).

[34] Dewachter, H., and Lyrio, M. (2006). Macro factors and the term structure of interest rates. Journal of Money, Credit and Banking, 119-140.

[35] Diebold, F. X. (2015). Comparing predictive accuracy, twenty years later: A personal perspective on the use and abuse of Diebold–Mariano tests. Journal of Business and Economic Statistics, 33(1), 1-1.

[36] Diebold, F. X., and Mariano, R. (1995). Comparing predictive accuracy. Journal of Business and Economic Statistics 13(3), 253-63.

[37] Diebold, F. X., Rudebusch, G. D., and Aruoba, S. B. (2006). The macroeconomy and the yield curve: a dynamic latent factor approach. Journal of Econometrics, 131(1), 309-338.

[38] Duffee, G. R. (2002). Term premia and interest rate forecasts in affine models. The Journal of Finance, 57(1), 405-443.

[39] Duffee, G. R. (2007). Are variations in term premia related to the macroeconomy. Working papers, Johns Hopkins University, Department of Economics.

[40] Duffee, G. R. (2010). Sharpe ratios in term structure models (No. 575). Working papers, Johns Hopkins University, Department of Economics.

[41] Duffee, G. R. (2011a). Information in (and not in) the term structure. Review of Financial Studies, 24(9), 2895-2934.

[42] Duffee, G. R. (2011b). Forecasting with the term structure: The role of no-arbitrage restrictions. Working paper, Johns Hopkins University.

[43] Duffee, G. R. (2013a). Bond pricing and the macroeconomy. Constantinides, G. M., Harris, M., and Stulz, R. M. Edition, Handbook of the Economics of Finance, Volume 2B: Asset Pricing, 907-967. North Holland.

[44] Duffee, G. R. (2013b). Forecasting Interest Rates. Elliott, G., and Timmermann, A. Edition, Handbook of Economic Forecasting, Vol. 2, Part A: Elsevier, pp. 385–426.

[45] Duffee, G. R., and Stanton, R. H. (2012). Estimation of dynamic term structure models. The Quarterly Journal of Finance, 2(2).

[46] Elliott, G., and Timmermann, A. (2016). Economic forecasting. Princeton University Press.

[47] Erceg, C. J., and Levin, A. T. (2003). Imperfect credibility and inflation persistence. Journal of Monetary Economics, 50(4), 915-944.

[48] Eriksen, J. N. (2015). Expected business conditions and bond risk premia. Journal of Financial and Quantitative Analysis, Forthcoming.

[49] Faust, J., and Wright, J. H. (2013). Forecasting inflation. Handbook of Economic Forecasting, 2(Part A), 3-56.

[50] Gargano, A., Pettenuzzo, D., and Timmermann, A. G. (2015). Bond return predictability: Economic value and links to the macroeconomy. Working paper.

[51] Geweke, J., and Amisano, G. (2011). Optimal prediction pools. Journal of Econometrics, 164(1), 130-141.

[52] Giacoletti, M., Laursen, K. T., and Singleton, K.J. (2014). Learning the macro-dynamics of US Treasury yields with aribitrage-free term structure models. Working paper, Stanford University.

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

[54] Goh, J., Jiang, F., Tu, J., and Zhou, G. (2013). Forecasting government bond risk premia using technical indicators. Working paper.

[55] Greenwood, R., and Vayanos, D. (2014). Bond supply and excess bond returns. Review of Financial Studies, 27(3), 663-713.

[56] Gürkaynak, R. S., Sack, B., and Wright, J. H. (2007). The US Treasury yield curve: 1961 to the present. Journal of Monetary Economics, 54(8), 2291-2304.

[57] Hansen, P. R., and Timmermann, A. (2015). Equivalence Between Out-of-Sample Forecast Comparisons and Wald Statistics. Econometrica, 83(6), 2485-2505.

[58] Harvey, D., Leybourne, S., and Newbold, P. (1997). Testing the equality of prediction mean squared errors. International Journal of Forecasting, 13(2), 281-291.

[59] Huang, J. Z., and Shi, Z. (2011). Determinants of bond risk premia. Working paper, Penn State University.

[60] Joslin, S., Le, A., and Singleton, K. J. (2013). Why Gaussian macro-finance term structure models are (nearly) unconstrained factor-VARs. Journal of Financial Economics, 109(3), 604-622.

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

[62] Joslin, S., Singleton, K. J., and Zhu, H. (2011). A new perspective on Gaussian dynamic term structure models. Review of Financial Studies, 24(3), 926-970.

[63] Keller, J. B. (1962). Factorization of matrices by least-squares. Biometrika, 49(1/2), 239-242.

[64] Kim, D. H. (2007). Challenges in macro-finance modeling. BIS Working papers 240, Bank for International Settlements.

[65] Kim, D. H., and Orphanides, A. (2012). Term structure estimation with survey data on interest rate forecasts. Journal of Financial and Quantitative Analysis, 47(01), 241-272.

[66] Kim, D. H., and Wright, J. H. (2005). An arbitrage-free three-factor term structure model and the recent behavior of long-term yields and distant-horizon forward rates. Finance and Economics Discussion Series 2005-33, Federal Reserve Board of Governors.

[67] Laubach, T., Tetlow, R. J., and Williams, J. C. (2007). Learning and the role of macroeconomic factors in the term structure of interest rates. Manuscript, Frankfurt.

[68] Litterman, R. B., and Scheinkman, J. (1991). Common factors affecting bond returns. The Journal of Fixed Income, 1(1), 54-61.

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

[70] Ludvigson, S. C., and Ng, S. (2010). A factor analysis of bond risk premia. Handbook of Empirical Economics and Finance, 313.

[71] Markowitz, H. (1952). Portfolio selection. The Journal of Finance, 7(1), 77-91.

[72] Mönch, E. (2008). Forecasting the yield curve in a data-rich environment: A no-arbitrage factor-augmented VAR approach. Journal of Econometrics, 146(1), 26-43.

[73] Mönch, E. (2012). Term structure surprises: the predictive content of curvature, level, and slope. Journal of Applied Econometrics, 27(4), 574-602.

[74] Nelson, C. R., and Siegel, A. F. (1987). Parsimonious modeling of yield curves. Journal of Business, 60, 473-489.

[75] Orphanides, A., and Wei, M. (2012). Evolving macroeconomic perceptions and the term structure of interest rates. Journal of Economic Dynamics and Control, 36(2), 239-254.

[76] Piazzesi, M., and Schneider, M. (2007). Equilibrium yield curves. In NBER Macroeconomics Annual 2006, 21, 389-472. MIT Press.

[77] Priebsch, M. (2014). (Un)Conventional monetary policy and the yield curve. Working Paper, Federal Reserve Board, Washington, D.C.

[78] Sarno, L., Schneider, P., and Wagner, C. (2014). The economic value of predicting bond risk premia: Can anything beat the expectations hypothesis? Working paper, University of Lugano.

[79] Stock, J. H., and Watson, M. W. (2007). Why has US inflation become harder to forecast? Journal of Money, Credit and banking, 39(s1), 3-33.

[80] Schwarz, G. (1978). Estimating the dimension of a model, Annals of Statistics 6, 461–464.

[81] Thornton, D. L., and Valente, G. (2012). Out-of-sample predictions of bond excess returns and forward rates: An asset allocation perspective. Review of Financial Studies, 25(10), 3141-3168.

[82] Timmermann, A. (2006). Forecast combinations. Handbook of Economic Forecasting, 1, 135-196.

[83] Wright, J. H. (2011). Term premia and inflation uncertainty: Empirical evidence from an international panel dataset. The American Economic Review, 101(4), 1514-1534.

[84] Wright, J. H., and Zhou, H. (2009). Bond risk premia and realized jump risk. Journal of Banking and Finance, 33(12), 2333-2345.

Table 1.1 Summary Statistics

	Mean (%)	St.Dev (%)	Skewness	Kurtosis	AC(1)	AC(12)	
Panel A: Term Structure of Yields							
	Mean (%)	St.Dev (%)	Skewness	Kurtosis	AC(1)	AC(12)	
PC1	2.93	1.55	0.37	3.07	0.99	0.85	
PC2	0.40	0.22	-0.18	2.45	0.94	0.41	
PC3	0.13	0.06	0.02	4.59	0.81	0.22	
Panel B: Macro Variables							
	Mean (%)	St.Dev (%)	Skewness	Kurtosis	AC(1)	AC(12)	
GRO	-0.01	0.92	-1.44	6.77	0.94	0.12	
Inflation	4.24	2.68	1.33	4.00	0.99	0.84	
Panel C: Bond Excess Returns							
One-Month Horizon	Mean (%)	St.Dev (%)	Skewness	Kurtosis	AC(1)	AC(12)	Sharpe Ratio
n-h Yields=1 month	0.36	0.20	2.84	17.59	0.35	-0.03	1.81
3 month	0.60	0.52	2.49	19.84	0.18	-0.13	1.14
6 month	0.79	0.97	1.52	17.05	0.19	-0.08	0.81
12 month	1.23	1.91	0.95	16.08	0.18	-0.04	0.64
24 month	1.47	3.44	0.37	14.96	0.15	0.00	0.43
60 month	2.32	7.12	0.14	7.37	0.11	0.02	0.33
Three-Month Horizon	Mean (%)	St.Dev (%)	Skewness	Kurtosis	AC(1)	AC(12)	Sharpe Ratio
n-h Yields=1 month	0.20	0.20	2.28	15.46	0.70	-0.04	1.00
3 month	0.41	0.56	1.61	14.05	0.71	-0.02	0.74
6 month	0.58	1.09	1.12	12.70	0.70	0.04	0.54
12 month	0.93	2.08	0.62	9.75	0.71	0.10	0.45
24 month	1.23	3.70	0.38	8.04	0.70	0.09	0.33
60 month	2.08	7.41	0.11	4.99	0.68	0.07	0.28
Six-Month Horizon	Mean (%)	St.Dev (%)	Skewness	Kurtosis	AC(1)	AC(12)	Sharpe Ratio
n-h Yields=1 month	0.13	0.18	0.14	6.42	0.79	0.07	0.73
3 month	0.29	0.52	-0.14	6.92	0.83	0.10	0.56
6 month	0.46	1.05	-0.43	6.09	0.85	0.18	0.43
12 month	0.76	1.99	-0.19	4.79	0.86	0.19	0.38
24 month	1.10	3.55	-0.24	4.65	0.85	0.15	0.31
60 month	1.94	7.11	-0.14	3.87	0.85	0.07	0.27
One-Year Horizon	Mean (%)	St.Dev (%)	Skewness	Kurtosis	AC(1)	AC(12)	Sharpe Ratio
n-h Yields=1 month	0.10	0.19	-0.14	5.40	0.77	0.28	0.53
3 month	0.23	0.53	-0.03	4.27	0.91	0.27	0.44
6 month	0.39	1.04	-0.04	3.56	0.93	0.27	0.37
12 month	0.58	1.93	-0.11	3.30	0.93	0.21	0.30
24 month	1.01	3.52	-0.09	3.29	0.93	0.15	0.29
60 month	1.78	7.19	0.06	3.32	0.92	0.04	0.25
Two-Year Horizon	Mean (%)	St.Dev (%)	Skewness	Kurtosis	AC(1)	AC(12)	Sharpe Ratio
n-h Yields=1 month	0.06	0.17	-0.44	3.44	0.83	0.43	0.39
3 month	0.16	0.49	-0.62	3.31	0.94	0.51	0.34
6 month	0.30	0.98	-0.62	3.47	0.96	0.51	0.31
12 month	0.54	1.87	-0.49	3.21	0.96	0.49	0.29
24 month	0.94	3.37	-0.40	3.43	0.96	0.47	0.28
60 month	1.55	6.79	-0.30	4.22	0.96	0.44	0.23
Five-Year Horizon	Mean (%)	St.Dev (%)	Skewness	Kurtosis	AC(1)	AC(12)	Sharpe Ratio
n-h Yields=1 month	0.04	0.14	-0.66	6.76	0.78	0.53	0.32
3 month	0.12	0.38	-0.92	5.23	0.90	0.64	0.32
6 month	0.24	0.73	-0.98	4.24	0.95	0.72	0.32
12 month	0.43	1.41	-0.84	3.64	0.97	0.74	0.31
24 month	0.74	2.68	-0.96	3.84	0.97	0.71	0.27
60 month	1.26	5.98	-0.89	3.75	0.98	0.71	0.21

This table presents the summary statistics for the data used in estimation and evaluation. The sample period is from 1971:11 to 2012:12. I present sample mean, standard deviation, skewness, kurtosis, and autocorrelations (1- and 12-order) for the term structure of yields (Panel A), macro variables (Panels B), and bond excess returns (Panel C). The yields are continuously compounded monthly zero-coupon bond yields from Gürkaynak, Sack, and Wright (2007) dataset. Panel A reports the results for the first three principal components (PC1, PC2, PC3) of yields with 24 different maturities ranging from one-month to ten-year. Following Joslin, Priebsch, and Singleton (2014), I use three-month moving average of the Chicago Fed National Activity Index as a measure of real activity growth (GRO). Inflation is defined as the 12-month moving average of core CPI inflation. I construct 36 horizon (h) and maturity (n-h) combinations of bond excess returns using all the yields. The means and standard deviations reported in Panel C are annualized. I also report the annualized Sharpe ratio in the last column of Panel C.

Table 1.2 Parameter Estimates

					Panel A: Q-Parameters Persistence					
r_∞^Q	**0.0119**									
	(0.0008)									
λ^Q	**0.9978**	**0.9454**	**0.7514**							
	(0.0001)	(0.001)	(0.0058)							
					Panel B: P-Parameters Persistence					
		$\mathbf{M}_{0span}$						$\mathbf{M}_0$		
$\lvert\lambda^P\rvert$	**0.9936**	**0.9490**	**0.7685**			**0.9876**	**0.9876**	**0.9415**	**0.9415**	**0.7648**
	(0.0001)	(0.0021)	(0.0108)			(0.0003)	(0.0012)	(0.0031)	(0.0017)	(0.0091)
		$\mathbf{M}_1$						$\mathbf{M}_2$		
$\lvert\lambda^P\rvert$	**0.9986**	**0.9903**	**0.9779**	**0.9779**	**0.8123**	**0.9989**	**0.9754**	**0.9754**	**0.9722**	**0.7893**
	(0.0001)	(0.0022)	(0.0013)	(0.0034)	(0.0078)	(0.0003)	(0.0023)	(0.0042)	(0.0013)	(0.0082)
					Panel C: Risk Premium Parameters					
		$\mathbf{M}_{0span}$						$\mathbf{M}_0$		
	PC1	PC2	PC3			PC1	PC2	PC3	GRO	Inflation
λ_{0PC}*1200	**0.4259**	**2.3817**	2.1127			**0.3736**	**2.2877**	2.1334		
	(0.0707)	(0.3112)	(2.1294)			(0.0779)	(0.1151)	(2.0298)		
λ_{1PC}										
PC1	**-0.0060**	**0.0159**	**-0.0148**			**-0.0334**	**0.0161**	0.0272	**0.2778**	**0.1145**
	(0.0019)	(0.0019)	(0.0061)			(0.0126)	(0.0061)	(0.0852)	(0.0124)	(0.0187)
PC2	**-0.0032**	-0.0596	**-0.2831**			-0.0066	**-0.0592**	**-0.2609**	**0.0764**	**0.0121**
	(0.0011)	(0.0663)	(0.1232)			(0.0174)	(0.0163)	(0.1234)	(0.0303)	(0.0015)
PC3	0.0004	-0.0064	-0.2128			-0.0006	-0.0066	-0.2169	-0.0042	0.0045
	(0.0029)	(0.0053)	(0.1531)			(0.0046)	(0.0043)	(0.1622)	(0.0241)	(0.0091)
		$\mathbf{M}_1$						$\mathbf{M}_2$		
	PC1	PC2	PC3	GRO	Inflation	PC1	PC2	PC3	GRO	Inflation
λ_{0PC}*1200	**-2.1313**	**2.2377**	2.2233			**-0.1056**	**1.8427**	2.0252		
	(0.1352)	(0.3154)	(2.0457)			(0.0514)	(0.3517)	(2.1721)		
λ_{1PC}										
PC1	0.0000	**-0.0026**	0.0318	**0.0031**	**-0.0014**	**-0.0222**	**0.0761**	0.1279	**0.1531**	**0.0910**
	(0.0021)	(0.0011)	(0.0731)	(0.0013)	(0.0005)	(0.0132)	(0.0112)	(0.0852)	(0.0175)	(0.0215)
PC2	-0.0002	0.0121	**-0.1488**	**-0.0147**	**0.0066**	0.0039	-0.0022	**-0.1653**	**-0.0420**	**-0.0102**
	(0.0121)	(0.0213)	(0.0834)	(0.0102)	(0.002)	(0.0181)	(0.0151)	(0.0925)	(0.0168)	(0.0011)
PC3	-0.0002	0.0151	-0.1844	-0.0182	0.0082	0.0020	**0.0073**	-0.1935	-0.0331	-0.0009
	(0.0021)	(0.0112)	(0.1322)	(0.0241)	(0.0093)	(0.0031)	(0.0031)	(0.1298)	(0.0315)	(0.0041)

This table presents the maximum likelihood estimates of the persistence parameters (Panels A and B) and the risk premium parameters (Panel C) for the maximally flexible unspanned macro model (M_0), the nested model of M_0 enforcing the spanning restrictions (M_{0span}), and the unspanned model with reduced-rank restrictions $\zeta = 1$ (M_1) and $\zeta = 2$ (M_2). I report the long-run mean of the short rate under Q (r_∞^Q), the eigenvalues of the feedback matrix under Q (λ^Q), the moduli of the eigenvalues of the P-feedback matrix ($|\lambda^P|$), and the market prices of risks (λ_{0PC}, λ_{1PC}). λ_{0PC} is scaled by $1,200$ to convert to annualized percentage points. Standard errors are computed using the outer product of the gradient of the likelihood function, and they are given in parentheses. Parameter estimates that are statistically significant are shown in boldface.

Table 1.3 Out-of-Sample RMSEs of the Unspanned Macro Model

Panel A: Preferred Unspanned Macro Model

Forecast Horizon	1 month	3 month	6 month	12 month	24 month	60 month
n-h Yields						
1 month yield	34.30	63.20	90.11	115.68	168.41	234.46
3 month yield	56.00	57.87	74.18	109.94	168.93	240.71
6 month yield	31.93	47.50	73.35	111.43	171.15	239.70
12 month yield	18.41	44.72	75.84	116.97	175.92	236.83
24 month yield	22.60	46.45	78.64	118.33	173.98	219.69
60 month yield	22.76	45.79	74.29	107.97	153.41	173.94

Panel B: Maximally Flexible Unspanned Macro Model

Forecast Horizon	1 month	3 month	6 month	12 month	24 month	60 month
n-h Yields						
1 month yield	32.41	54.48	77.09	116.10	186.37	284.30
3 month yield	48.71	51.07	77.78	120.27	190.53	288.88
6 month yield	27.10	50.88	79.48	125.28	193.90	270.51
12 month yield	27.07	50.86	79.50	124.93	188.98	281.39
24 month yield	28.57	51.99	78.21	120.20	175.79	259.09
60 month yield	27.66	49.28	72.38	105.54	163.51	218.83

Panel C: RMSE Ratio

Forecast Horizon	1 month	3 month	6 month	12 month	24 month	60 month
n-h Yields						
1 month yield	1.06*	1.16*	1.17*	1.00	0.90**	0.82**
3 month yield	1.15*	1.13**	0.95*	0.91**	0.89*	0.83***
6 month yield	1.18**	0.93*	0.92*	0.89**	0.88**	0.88*
12 month yield	0.68***	0.88***	0.95*	0.94**	0.93*	0.84**
24 month yield	0.79**	0.89*	1.01	0.98	0.99	0.83**
60 month yield	0.82*	0.93**	1.03	1.02	0.94*	0.79**

This table presents the out-of-sample root mean squared errors (RMSEs) for 36 horizon (h) and maturity (n-h) combinations of bond excess returns using unspanned macro models. Panel A is for the preferred unspanned macro model (M_{us}). Panel B is for the maximally flexible unspanned macro model (M_0). RMSEs are reported in basis points. Panel C shows the ratios of the RMSEs in Panel A and Panel B. The out-of-sample results are obtained by estimating the models recursively at each month using a rolling window of the most recent 10 years of observations. The statistical significance of the relative forecasting performance is evaluated using the Diebold and Mariano (1995) t-statistics computed with a serial correlation robust variance and with the small sample adjustment of Harvey, Leybourne, and Newbold (1997). The significance is denoted by *, ** and ***, corresponding to significance levels of 10%, 5% and 1% respectively.

Table 1.4 Out-of-Sample Forecasting Performance: RMSEs

n-h Yields	RMSE	Forecast Error Relative to the Spanned Model (%)					
	M_{span}	M_{us}	$M^{0.5}_{com}$	M_{com1}	M^{tv}_{com2}	M_{com3}	M^{tv}_{com4}
	$\beta^n_h = 0$	$\beta^n_h = 1$	$\beta^n_h = 0.5$	$\widehat{\beta}^n_h \in R$	$\widehat{\beta}^n_{t,h} \in R$	$\widehat{\beta}^n_h \in [0,1]$	$\widehat{\beta}^n_{t,h} \in [0,1]$
Panel A: One-Month Forecast Horizon							
1 month	42.71	-19.68	-17.37	-20.85	-31.33	-20.85	-28.18
3 month	44.12	26.95	-7.44	-10.30**	-15.78*	-10.30**	-15.78*
6 month	30.88	3.39	-9.79**	-9.86**	-11.07**	-9.86**	-11.07**
12 month	17.64	4.37	-3.73*	-3.95*	-5.13*	-3.95*	-5.12*
24 month	25.32	-10.76	-11.02	-12.40	-15.49	-12.40	-15.49
60 month	21.81	4.37	-4.59**	-4.78**	-5.40	-4.78**	-5.40
Panel B: Three-Month Forecast Horizon							
1 month	73.76	-14.31	-12.75	-15.17	-21.94	-15.17	-19.44
3 month	50.42	14.77	-5.52	-6.78**	-9.48**	-6.78**	-9.48**
6 month	45.84	3.61	-5.86**	-5.98**	-6.94**	-5.98**	-6.94**
12 month	44.48	0.56	-4.37*	-4.37*	-4.65**	-4.37*	-4.60**
24 month	49.84	-6.81	-7.33	-8.10	-11.61	-8.10	-11.10
60 month	44.22	3.56	-2.01*	-2.23*	-3.36*	-2.23*	-3.21*
Panel C: Six-Month Forecast Horizon							
1 month	102.46	-12.05	-11.13	-13.01	-15.63	-13.01	-15.08
3 month	73.81	0.50	-7.76**	-7.77**	-8.66**	-7.77**	-8.66**
6 month	74.36	-1.36	-6.54**	-6.56**	-6.80**	-6.56**	-6.80**
12 month	77.85	-2.59	-5.51*	-5.61	-6.69**	-5.61	-6.63**
24 month	83.04	-5.30	-6.29	-6.80	-10.57*	-6.80	-9.95*
60 month	72.47	2.51	-2.01*	-2.13*	-4.74*	-2.13*	-4.09**
Panel D: One-Year Forecast Horizon							
1 month	132.34	-12.59**	-10.46*	-12.91**	-15.34***	-12.91**	-14.18***
3 month	115.94	-8.15*	-5.17	-8.87*	-10.64**	-8.87*	-9.53**
6 month	116.80	-7.00**	-4.60	-8.30*	-9.22**	-8.30*	-9.14**
12 month	121.42	-5.71	-3.66	-6.14	-10.02**	-6.14	-8.99**
24 month	122.47	-5.19*	-3.38	-5.91	-9.36**	-5.91	-9.10**
60 month	104.74	-1.78*	-1.55*	-1.75*	-6.47**	-1.75*	-3.68*
Panel E: Two-Year Forecast Horizon							
1 month	186.20	-9.55***	-8.50**	-10.09**	-12.88***	-10.09**	-12.09***
3 month	180.11	-7.55*	-6.21	-8.12*	-10.96**	-8.12*	-9.96**
6 month	184.19	-7.59***	-6.08	-8.41*	-11.74**	-8.41*	-10.47**
12 month	187.77	-7.15*	-6.31	-7.50	-10.88**	-7.50	-9.70**
24 month	182.60	-5.86*	-4.72	-6.28	-9.25**	-6.28	-8.21**
60 month	153.07	-4.22**	-3.24*	-3.24*	-7.31**	-3.24*	-5.20**
Panel F: Five-Year Forecast Horizon							
1 month	268.41	-12.65**	-8.71*	-12.79***	-14.27***	-12.65**	-12.78***
3 month	265.16	-9.22**	-7.27	-9.29*	-13.74***	-9.29*	-11.31***
6 month	267.37	-10.35**	-7.53	-10.36**	-13.70**	-10.35**	-11.91**
12 month	265.53	-10.81**	-7.60*	-10.88**	-14.07**	-10.81**	-12.24**
24 month	247.31	-11.17**	-7.86*	-11.23**	-14.80**	-11.17**	-12.62**
60 month	196.43	-11.45**	-8.08*	-11.51**	-13.70**	-11.45**	-12.69**

This table presents the out-of-sample forecasting performance for 36 horizon (h) and maturity (n-h) combinations of bond excess returns using the preferred unspanned macro model (M_{us}), the spanned model (M_{span}) and their combinations. Each panel displays the results for one forecast horizon. I report the annualized root mean squared errors (RMSEs) in basis points for M_{span} in Column 2. Column 3-8 presents the improvement in out-of-sample forecast accuracy relative to M_{span}. The improvement is measured by the percentage change in RMSE. For example, a number of -19.68 (26.95) means that the alternative model has -19.68% smaller (26.95% larger) out-of-sample RMSE than the benchmark model. Column 3 is for M_{us}. Column 4 is for the combination forecasts with equal weights ($M^{0.5}_{com}$). Column 5-8 are for the combination forecasts based on the estimated optimal weights with respect to the sum squared error (SSE) loss function. I consider both constant (M_{com1}, $\widehat{\beta}^n_h \in R$) and time-varying optimal weights (M^{tv}_{com2}, $\widehat{\beta}^n_{t,h} \in R$), and they can be constrained to $[0,1]$ (M_{com3}, $\widehat{\beta}^n_h \in [0,1]$; M^{tv}_{com4}, $\widehat{\beta}^n_{t,h} \in [0,1]$) or not. The out-of-sample results are obtained by estimating the models recursively at each month using a rolling window of the most recent 10 years of observations. The statistical significance of the RMSE gains or losses is evaluated using the Diebold and Mariano (1995) t-statistics computed with a serial correlation robust variance and with the small sample adjustment of Harvey, Leybourne, and Newbold (1997). The significance is denoted by *, ** and ***, corresponding to significance levels of 10%, 5% and 1% respectively.

Table 1.5 Constant Optimal Weights of the Preferred Unspanned Macro Model

Panel A: Sum Squared Error Loss Function

n-h Yields	$\widehat{\beta}_h^n \in R$	$\widehat{\beta}_h^n \in [0,1]$	n-h Yields	$\widehat{\beta}_h^n \in R$	$\widehat{\beta}_h^n \in [0,1]$
One-Month Forecast Horizon			**Three-Month Forecast Horizon**		
1 month	0.82	0.82	1 month	0.81	0.81
3 month	0.33*	0.33*	3 month	0.35*	0.35*
6 month	0.46	0.46	6 month	0.44	0.44
12 month	0.41	0.41	12 month	0.48	0.48
24 month	0.74	0.74	24 month	0.72	0.72
60 month	0.42	0.42	60 month	0.38	0.38
Six-Month Forecast Horizon			**One-Year Forecast Horizon**		
1 month	0.79*	0.79*	1 month	0.87**	0.87**
3 month	0.49	0.49	3 month	0.62*	0.62*
6 month	0.53	0.53	6 month	0.62*	0.62*
12 month	0.58	0.58	12 month	0.72*	0.72*
24 month	0.68	0.68	24 month	0.62	0.62
60 month	0.40	0.40	60 month	0.37	0.37
Two-Year Forecast Horizon			**Five-Year Forecast Horizon**		
1 month	0.82**	0.82**	1 month	1.12***	1.00**
3 month	0.68*	0.68*	3 month	0.92**	0.92**
6 month	0.72**	0.72**	6 month	1.02***	1.00**
12 month	0.72*	0.72*	12 month	1.08***	1.00**
24 month	0.67*	0.67*	24 month	1.08**	1.00*
60 month	0.49	0.49	60 month	1.07**	1.00**

Panel B: Mean-Variance Utility Loss Function

Forecast Horizon	$\widehat{\beta}_h \in R$	$\widehat{\beta}_h \in [0,1]$
h=1 month	0.60	0.60
h=3 month	0.68	0.68
h=6 month	0.77	0.77
h=12 month	1.58***	0.94*
h=24 month	1.72***	1.00*
h=60 month	1.81***	1.00**

This table presents the constant optimal weights of the preferred unspanned macro model (M_{us}) with respect to the sum squared error (SSE) loss function $\widehat{\beta}_h^n$ (Panel A) and the mean-variance utility loss function $\widehat{\beta}_h$ (Panel B). The constant optimal weights are computed with respect to the interested loss function using the entire out-of-sample forecasts of bond excess returns. The optimal combination weights can be constrained to $[0,1]$ or not. For the SSE loss function, I report the results for 36 horizon (h) and maturity (n-h) combinations of bond excess return forecasts. For the mean-variance utility loss function, I report the results for six forecast horizons. The results are irrelevant to investor's risk aversion level. For an investment horizon h, the investor selects bond portfolios with $(h+1)$-, $(h+3)$-, $(h+6)$-, $(h+12)$-, $(h+24)$-, $(h+60)$-, and h-month maturities, I therefore report results based on forecast horizons only. The statistical significance of the estimated optimal weight against 0.5 is evaluated using Carriero and Giacomini (2011) test. The significance is denoted by *, ** and ***, corresponding to significance levels of 10%, 5% and 1% respectively.

Table 1.6 Economic Value of Bond Excess Return Forecasts: Mean-Variance: Certainty Equivalent Returns (%)

	M_{span} $\beta_h = 0$	M_{us} $\beta_h = 1$	$M^{0.5}_{com}$ $\beta_h = 0.5$	M_{com1} $\widehat{\beta}_h \in R$	M^{tv}_{com2} $\widehat{\beta}_{t,h} \in R$	M_{com3} $\widehat{\beta}_h \in [0,1]$	M^{tv}_{com4} $\widehat{\beta}_{t,h} \in [0,1]$
γ**=3**							
h=1 month	6.47	6.47	6.48	6.48	6.60	6.48	6.50
h=3 month	6.07	6.04	6.08	6.08	6.47	6.08	6.14
h=6 month	5.44	5.46	5.48	5.48	6.42*	5.48	5.65
h=12 month	2.16	3.19*	2.66	3.70**	3.89**	3.19*	3.74**
h=24 month	1.15	2.18*	1.96*	3.69***	3.84***	2.18*	2.63**
h=60 month	1.46	2.21*	1.61	3.69***	3.98***	2.21*	2.71**
γ**=5**							
h=1 month	5.46	5.46	5.47	5.47	5.56	5.47	5.48
h=3 month	5.12	5.10	5.13	5.13	5.45	5.13	5.19
h=6 month	4.60	4.61	4.62	4.63	5.39*	4.63	4.76
h=12 month	1.94	2.87*	2.45	3.12*	3.42**	2.87*	3.99***
h=24 month	1.10	2.15*	1.85*	3.12***	3.53***	2.15*	2.58**
h=60 month	1.03	2.01*	1.87*	3.10***	3.65***	2.01*	2.94**
γ**=10**							
h=1 month	4.16	4.16	4.17	4.17	4.25	4.17	4.17
h=3 month	3.90	3.88	3.91	3.91	4.13	3.91	3.95
h=6 month	3.50	3.51	3.52	3.52	4.08	3.52	3.63
h=12 month	1.21	2.08*	1.71	2.38*	2.66**	2.08*	2.31*
h=24 month	1.25	1.93	1.46	2.38*	2.59**	1.93	2.32*
h=60 month	0.74	1.63*	1.20	2.30**	3.01***	1.63*	1.99**

This table presents the annualized certainty equivalent returns (CERs) for portfolio decisions based on out-of-sample forecasts of bond excess returns. For an investment horizon h, the mean-variance investors with different levels of risk aversion (γ=3, 5, 10) select $(h+1)$-, $(h+3)$-, $(h+6)$-, $(h+12)$-, $(h+24)$-, $(h+60)$-month bonds and h-month bond based on the predictive returns implied by a given model. I report results for six investment horizons based on seven different forecasts: the preferred unspanned macro model (M_{us}), the spanned model (M_{span}), and the forecast combinations of the two models. The forecast combinations with equal weights ($M^{0.5}_{com}$). The forecast combinations based on the estimated optimal weights with respect to mean-variance loss function. I consider both constant (M_{com1}, $\widehat{\beta}_h \in R$) and time-varying optimal weights (M^{tv}_{com2}, $\widehat{\beta}_{t,h} \in R$), and they can be constrained to $[0,1]$ (M_{com3}, $\widehat{\beta}_h \in [0,1]$; M^{tv}_{com4}, $\widehat{\beta}_{t,h} \in [0,1]$) or not. The constant optimal weights are computed using the entire out-of-sample forecasts of returns. The time-varying optimal weights are computed using the out-of-sample forecasts of returns with a smoothing window of size $d = 315 \times 0.2 = 63$. All CERs are reported in percentage. The statistical significance of the utility differences relative to M_{span} is evaluated using the Diebold and Mariano (1995) t-statistics computed with a serial correlation robust variance and with the small sample adjustment of Harvey, Leybourne, and Newbold (1997). The significance is denoted by *, ** and ***, corresponding to significance levels of 10%, 5% and 1% respectively.

Table 1.7 Economic Value of Bond Excess Return Forecasts: Mean-Variance: Certainty Equivalent Returns (%): Leverage and Short-Selling Constraints

	M_{span} $\beta_h = 0$	M_{us} $\beta_h = 1$	$M_{com}^{0.5}$ $\beta_h = 0.5$	M_{com1} $\widehat{\beta}_h \in R$	M_{com2}^{tv} $\widehat{\beta}_{t,h} \in R$	M_{com3} $\widehat{\beta}_h \in [0,1]$	M_{com4}^{tv} $\widehat{\beta}_{t,h} \in [0,1]$
				Panel A: Leverage = 0%			
γ**=3**							
h=1 month	3.26	3.51	3.43	3.74	3.95	3.51	3.78
h=3 month	1.62	1.68	1.96	2.06	2.19	1.68	2.07
h=6 month	1.33	1.35	1.35	1.71	1.83	1.35	1.44
h=12 month	1.04	1.61	1.23	1.85*	2.52**	1.61	2.28**
h=24 month	0.72	1.25	0.95	1.38	1.96**	1.25	1.32
h=60 month	0.36	1.37*	1.28*	1.46*	1.88**	1.37*	1.42*
γ**=5**							
h=1 month	2.61	3.12	2.93	3.02	3.19	3.12	3.09
h=3 month	1.31	1.69	1.62	1.87	1.99	1.69	1.90
h=6 month	1.08	1.50	1.16	1.59	1.67	1.50	1.61
h=12 month	0.83	1.29	1.09	1.41	2.20**	1.29	1.39
h=24 month	0.58	1.09	0.92	1.61*	1.99**	1.09	1.17
h=60 month	0.31	1.03*	0.94	1.18*	2.02**	1.03*	1.18*
γ**=10**							
h=1 month	1.77	1.97	1.95	2.17	2.33	1.97	2.27
h=3 month	0.91	1.01	0.99	1.25	1.42	1.01	1.13
h=6 month	0.73	0.90	0.74	1.07	1.17	0.90	1.13
h=12 month	0.56	1.18	1.02	1.26	2.07**	1.18	1.90**
h=24 month	0.34	1.10*	0.81	1.15*	1.65**	1.10*	1.52*
h=60 month	0.23	0.79	0.68	0.96*	1.01*	0.79	0.89

	M_{span} $\beta_h = 0$	M_{us} $\beta_h = 1$	$M_{com}^{0.5}$ $\beta_h = 0.5$	M_{com1} $\widehat{\beta}_h \in R$	M_{com2}^{tv} $\widehat{\beta}_{t,h} \in R$	M_{com3} $\widehat{\beta}_h \in [0,1]$	M_{com4}^{tv} $\widehat{\beta}_{t,h} \in [0,1]$
Panel B: Leverage $<$ 50%							
γ=3							
h=1 month	4.13	4.47	4.64	4.64	4.77	4.47	4.62
h=3 month	2.07	2.20	2.13	2.36	2.49	2.20	2.35
h=6 month	1.71	2.12	1.85	2.25	2.31	2.12	2.26
h=12 month	1.32	2.08*	1.48	2.14*	2.95**	2.08*	2.16*
h=24 month	0.92	1.54	1.38	1.55	2.12*	1.54	1.70*
h=60 month	0.48	1.39*	1.23*	1.41*	2.05**	1.39*	1.40*
γ=5							
h=1 month	3.14	3.34	3.19	3.39	3.71	3.34	3.41
h=3 month	1.59	2.04	1.78	2.15	2.22	2.04	2.19
h=6 month	1.29	1.67	1.43	1.73	1.90	1.67	1.86
h=12 month	0.99	1.78*	1.41	1.99*	2.15*	1.78*	1.72*
h=24 month	0.65	1.30	0.87	1.67*	2.04**	1.30	1.45*
h=60 month	0.40	1.21*	0.75	1.27*	1.67**	1.21*	1.42*
γ=10							
h=1 month	2.05	2.37	2.27	2.41	2.51	2.37	2.39
h=3 month	1.09	1.35	1.16	1.36	1.54	1.35	1.41
h=6 month	0.87	1.01	0.96	1.12	1.34	1.01	1.26
h=12 month	0.64	1.08	0.87	1.27	2.01**	1.08	1.17
h=24 month	0.37	1.16*	1.15*	1.18*	1.58**	1.16*	1.32*
h=60 month	0.28	1.08*	0.98	1.12*	1.25*	1.08*	1.20*

	M_{span} $\beta_h = 0$	M_{us} $\beta_h = 1$	$M_{com}^{0.5}$ $\beta_h = 0.5$	M_{com1} $\widehat{\beta}_h \in R$	M_{com2}^{tv} $\widehat{\beta}_{t,h} \in R$	M_{com3} $\widehat{\beta}_h \in [0,1]$	M_{com4}^{tv} $\widehat{\beta}_{t,h} \in [0,1]$
				Panel C: Leverage $<$ 100%			
γ=3							
1 month	4.75	5.11	5.30	5.35	5.43	5.11	5.32
3 month	2.39	2.95	2.72	2.76	2.85	2.95	2.81
6 month	1.96	2.17	2.16	2.22	2.52	2.17	2.23
12 month	1.50	2.29*	1.92	2.44*	3.06**	2.29*	2.89**
24 month	1.04	1.60	1.59	1.68	2.39**	1.60	1.80*
60 month	0.58	1.42*	1.35*	1.56*	3.13***	1.42*	1.79**
γ=5							
1 month	3.53	4.01	3.90	4.11	4.16	4.01	4.12
3 month	1.82	2.08	1.99	2.41	2.50	2.08	2.36
6 month	1.46	1.76	1.68	2.08	2.12	1.76	2.01
12 month	1.12	1.78	1.69	2.09*	2.14*	1.78	1.88*
24 month	0.69	1.39	1.22	1.69*	1.99**	1.39	1.79*
60 month	0.46	1.26*	1.14	1.48*	1.72**	1.26*	1.51*
γ=10							
1 month	2.25	2.53	2.49	2.51	2.86	2.53	2.71
3 month	1.23	1.46	1.30	1.75	1.88	1.46	1.80
6 month	0.97	1.10	1.06	1.24	2.12*	1.10	1.40
12 month	0.69	1.27	1.13	1.39	2.04**	1.27	1.83*
24 month	0.39	1.17*	1.14*	1.21*	1.64**	1.17*	1.42*
60 month	0.32	1.13*	1.19*	1.20*	1.27*	1.13*	1.26*

This table presents the annualized certainty equivalent returns (CERs) for portfolio decisions based on out-of-sample forecasts of bond excess returns. For an investment horizon h, the mean-variance investors with different levels of risk aversion (γ=3, 5, 10) select $(h+1)$-, $(h+3)$-, $(h+6)$-, $(h+12)$-, $(h+24)$-, $(h+60)$-month bonds and h-month bond based on the predictive returns implied by a given model. I rule out short-selling (negative portfolio weights), and also impose constraints on total portfolio leverage in order to avoid implausible positions. We consider leverage ratio of 0 (Panel A), 50% (Panel B), and 100% (Panel C). For example, leverage=50% indicates that the investor cannot borrow more than 50% of his or her total wealth. We report results for six investment horizons based on seven different forecasts: the preferred unspanned macro model (M_{us}), the spanned model (M_{span}), and the forecast combinations of the two models. The forecast combinations with equal weights ($M_{com}^{0.5}$). The forecast combinations based on the estimated optimal weights with respect to mean-variance loss function. I consider both constant (M_{com1}, $\widehat{\beta}_h \in R$) and time-varying optimal weights (M_{com2}^{tv}, $\widehat{\beta}_{t,h} \in R$), and they can be constrained to $[0,1]$ (M_{com3}, $\widehat{\beta}_h \in [0,1]$; M_{com4}^{tv}, $\widehat{\beta}_{t,h} \in [0,1]$) or not. All CERs are reported in percentage. The statistical significance of the utility differences relative to M_{span} is evaluated using the Diebold and Mariano (1995) t-statistics computed with a serial correlation robust variance and with the small sample adjustment of Harvey, Leybourne, and Newbold (1997). The significance is denoted by *, ** and ***, corresponding to significance levels of 10%, 5% and 1% respectively.

Table 1.8 Economic Value of Bond Excess Return Forecasts: Power Utility: Certainty Equivalent Returns (%): Leverage and Short-Selling Constraints

	M_{span} $\beta_h = 0$	M_{us} $\beta_h = 1$	$M_{com}^{0.5}$ $\beta_h = 0.5$	M_{com1} $\widehat{\beta}_h \in R$	M_{com2}^{tv} $\widehat{\beta}_{t,h} \in R$	M_{com3} $\widehat{\beta}_h \in [0,1]$	M_{com4}^{tv} $\widehat{\beta}_{t,h} \in [0,1]$
Panel A: Leverage = 0%							
γ**=3**							
h=1 month	2.58	2.96	3.16	3.18	3.21	3.16	3.20
h=3 month	1.94	2.55	2.30	2.36	2.49	2.55	2.42
h=6 month	1.21	1.24	1.24	1.63	1.78	1.24	1.34
h=12 month	0.71	1.29	0.91	1.56*	2.25**	1.29	1.97**
h=24 month	0.45	0.98	0.68	1.14	1.73**	0.98	1.06
h=60 month	0.25	1.26*	1.18*	1.38*	1.84**	1.26*	1.33*
γ**=5**							
h=1 month	1.71	2.22	2.10	2.32	2.38	2.22	2.35
h=3 month	1.64	1.98	1.86	2.11	2.32	1.98	2.27
h=6 month	1.16	1.58	1.24	1.70	1.81	1.58	1.70
h=12 month	0.68	1.14	0.94	1.29	2.11**	1.14	1.25
h=24 month	0.42	0.93	0.75	1.48*	1.87**	0.93	1.02
h=60 month	0.24	0.96*	0.87	1.14*	1.99**	0.96*	1.13*
γ**=10**							
h=1 month	1.31	1.66	1.61	1.91	2.00	1.66	1.97
h=3 month	1.60	1.73	1.70	2.00	2.19	1.73	1.86
h=6 month	1.05	1.23	1.07	1.43	1.54	1.23	1.47
h=12 month	0.62	1.24	1.08	1.35*	2.17**	1.24	1.97**
h=24 month	0.37	1.13*	0.84	1.21*	1.72**	1.13*	1.56*
h=60 month	0.22	0.78	0.67	0.98*	1.04*	0.78	0.89

	M_{span} $\beta_h = 0$	M_{us} $\beta_h = 1$	$M_{com}^{0.5}$ $\beta_h = 0.5$	M_{com1} $\widehat{\beta}_h \in R$	M_{com2}^{tv} $\widehat{\beta}_{t,h} \in R$	M_{com3} $\widehat{\beta}_h \in [0,1]$	M_{com4}^{tv} $\widehat{\beta}_{t,h} \in [0,1]$
				Panel B: Leverage < 50%			
γ=3							
h=1 month	4.81	5.19	5.37	5.38	5.46	5.37	5.36
h=3 month	2.03	2.19	2.11	2.40	2.55	2.19	2.35
h=6 month	1.29	1.50	1.49	1.59	1.93	1.50	1.57
h=12 month	0.75	1.52*	0.91	1.62*	2.44**	1.52*	1.61*
h=24 month	0.48	1.10	0.94	1.15	1.75**	1.10	1.28*
h=60 month	0.30	1.21*	1.05*	1.26*	1.93**	1.21*	1.23*
γ=5							
h=1 month	2.60	2.81	2.67	2.87	3.01	2.81	2.90
h=3 month	1.82	2.31	2.03	2.45	2.55*	2.31	2.47
h=6 month	1.18	1.56	1.32	1.67	1.85	1.56	1.77
h=12 month	0.70	1.49*	1.12	1.74*	1.90*	1.49*	1.44*
h=24 month	0.43	1.08	0.65	1.48*	1.87**	1.08	1.24*
h=60 month	0.28	1.09*	0.63	1.17*	1.59**	1.09*	1.31*
γ=10							
h=1 month	1.39	1.88	1.65	1.99	2.07	1.88	1.97
h=3 month	1.60	1.86	1.70	2.07	2.13	1.86	1.91
h=6 month	1.07	1.21	1.16	1.35	1.61	1.21	1.48
h=12 month	0.63	1.07	0.85	1.29	2.04**	1.07	1.17
h=24 month	0.37	1.16*	1.15*	1.21*	1.65**	1.16*	1.33*
h=60 month	0.24	1.04*	0.94	1.11*	1.25*	1.04*	1.17*

	M_{span} $\beta_h = 0$	M_{us} $\beta_h = 1$	$M^{0.5}_{com}$ $\beta_h = 0.5$	M_{com1} $\widehat{\beta}_h \in R$	M^{tv}_{com2} $\widehat{\beta}_{t,h} \in R$	M_{com3} $\widehat{\beta}_h \in [0,1]$	M^{tv}_{com4} $\widehat{\beta}_{t,h} \in [0,1]$
				Panel C: Leverage $<$ 100%			
γ=**3**							
1 month	5.34	5.60	5.52	5.87	5.80	5.60	5.69
3 month	2.08	2.16	2.43	2.57	2.73	2.43	2.56
6 month	1.26	1.67	1.40	1.85	1.94	1.67	1.82
12 month	0.78	1.58*	1.20	1.74*	2.38**	1.58*	2.19**
24 month	0.50	1.06	1.05	1.16	1.90**	1.06	1.28*
60 month	0.34	1.18*	1.11*	1.33*	2.92***	1.18*	1.56**
γ=**5**							
1 month	4.38	4.91	4.73	4.87	4.94	4.91	4.95
3 month	1.92	2.32	2.24	2.37	2.49	2.32	2.35
6 month	1.20	1.50	1.42	1.85	1.92*	1.50	1.77
12 month	0.72	1.38	1.28	1.69*	1.76*	1.38	1.49*
24 month	0.44	1.14	0.97	1.45*	1.77**	1.14	1.55*
60 month	0.30	1.10*	0.98	1.33*	1.60**	1.10*	1.37*
γ=**10**							
1 month	2.03	2.37	2.26	2.49	2.70	2.37	2.40
3 month	1.58	1.88	1.68	1.92	2.13	1.88	1.95
6 month	1.08	1.21	1.17	1.37	2.28*	1.21	1.53
12 month	0.63	1.21	1.07	1.34*	2.00**	1.21	1.78*
24 month	0.38	1.16*	1.13*	1.20*	1.65**	1.16*	1.42*
60 month	0.24	1.06*	1.12*	1.14*	1.22*	1.06*	1.20*

This table presents the annualized certainty equivalent returns (CERs) for portfolio decisions based on out-of-sample forecasts of bond excess returns. For an investment horizon h, investors with power utility at different levels of risk aversion (γ=3, 5, 10) select $(h+1)$-, $(h+3)$-, $(h+6)$-, $(h+12)$-, $(h+24)$-, $(h+60)$-month bonds and h-month bond based on the predictive returns implied by a given model. I rule out short-selling (negative portfolio weights), and also impose constraints on total portfolio leverage in order to avoid implausible positions. We consider leverage ratio of 0 (Panel A), 50% (Panel B), and 100% (Panel C). For example, leverage=50% indicates that the investor cannot borrow more than 50% of his or her total wealth. I report results for six investment horizons based on seven different forecasts: the preferred unspanned macro model (M_{us}), the spanned model (M_{span}), and the forecast combinations of the two models. The forecast combinations with equal weights ($M^{0.5}_{com}$). The forecast combinations based on the estimated optimal weights with respect to mean-variance loss function. I consider both constant (M_{com1}, $\widehat{\beta}_h \in R$) and time-varying optimal weights (M^{tv}_{com2}, $\widehat{\beta}_{t,h} \in R$), and they can be constrained to $[0,1]$ (M_{com3}, $\widehat{\beta}_h \in [0,1]$; M^{tv}_{com4}, $\widehat{\beta}_{t,h} \in [0,1]$) or not. All CERs are reported in percentage. The statistical significance of the utility differences relative to M_{span} is evaluated using the Diebold and Mariano (1995) t-statistics computed with a serial correlation robust variance and with the small sample adjustment of Harvey, Leybourne, and Newbold (1997). The significance is denoted by *, ** and ***, corresponding to significance levels of 10%, 5% and 1% respectively.

Table 1.9 Economic Value of Bond Excess Return Forecasts: Mean-Variance: Manipulation-Proof Performance Measure (%): Leverage and Short-Selling Constraints

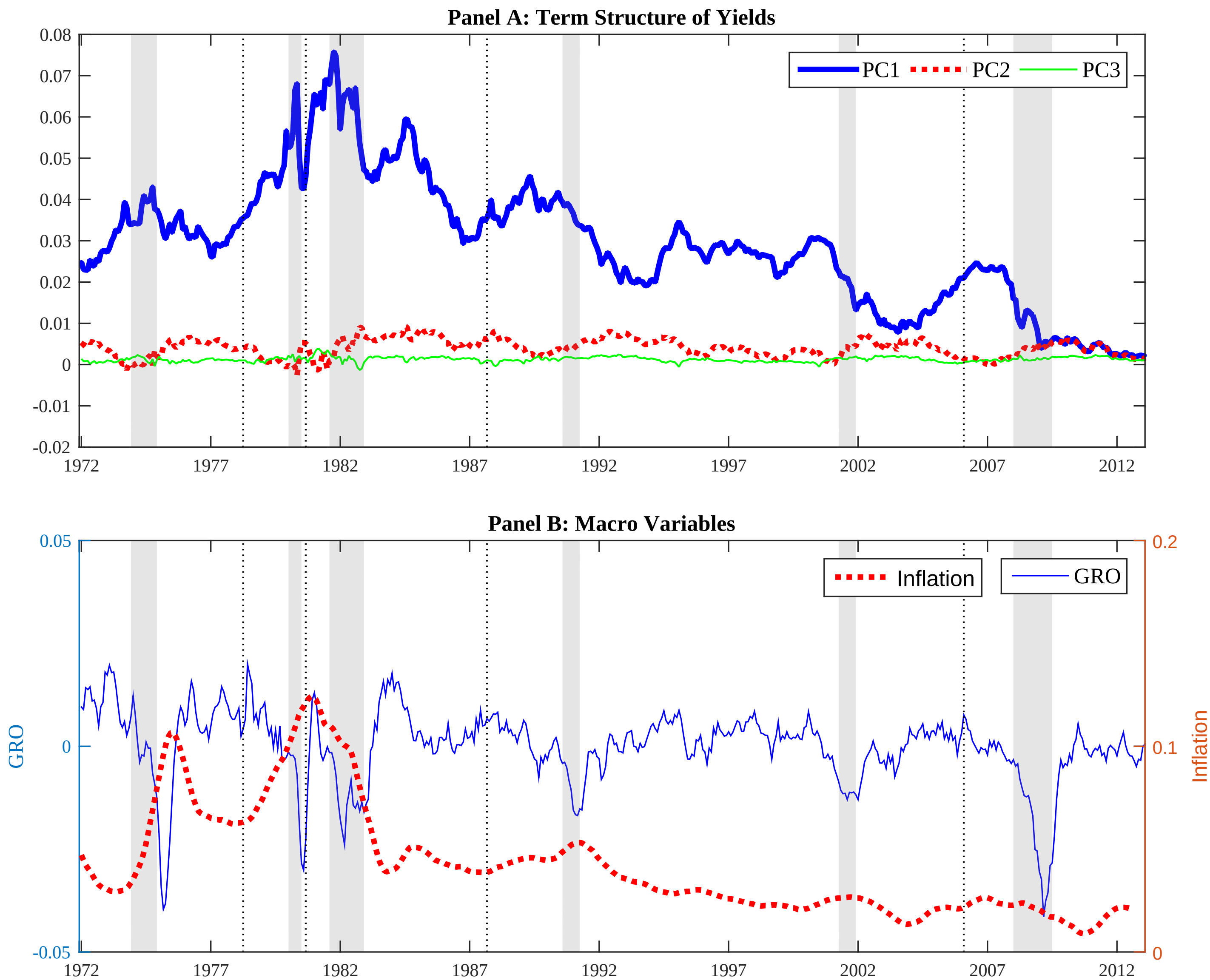
	M_{us} $\beta_h = 1$	$M_{com}^{0.5}$ $\beta_h = 0.5$	M_{com1} $\widehat{\beta}_h \in R$	M_{com2}^{tv} $\widehat{\beta}_{t,h} \in R$	M_{com3} $\widehat{\beta}_h \in [0,1]$	M_{com4}^{tv} $\widehat{\beta}_{t,h} \in [0,1]$
			Panel A: Leverage = 0%			
γ=3						
h=1 month	-0.04	0.14	0.19	0.22	0.14	0.20
h=3 month	0.15	0.13	0.17	0.25	0.15	0.20
h=6 month	-0.09	-0.05	0.18	0.27	-0.05	0.01
h=12 month	0.25	0.07	0.45	0.86	0.25	0.70
h=24 month	0.25	0.10	0.38	0.73	0.25	0.33
h=60 month	0.55	0.54	0.66	0.93	0.55	0.63
γ=5						
h=1 month	0.11	0.10	0.22	0.26	0.11	0.24
h=3 month	0.08	0.08	0.23	0.34	0.08	0.31
h=6 month	0.15	0.00	0.27	0.33	0.15	0.27
h=12 month	0.20	0.12	0.32	0.81	0.20	0.30
h=24 month	0.25	0.17	0.61	0.84	0.25	0.33
h=60 month	0.39	0.36	0.52	1.03	0.39	0.52
γ=10						
h=1 month	0.06	0.08	0.25	0.31	0.08	0.29
h=3 month	-0.02	0.02	0.19	0.30	0.02	0.11
h=6 month	0.03	-0.03	0.19	0.25	0.03	0.21
h=12 month	0.31	0.25	0.41	0.90	0.31	0.78
h=24 month	0.41	0.27	0.49	0.79	0.41	0.70
h=60 month	0.30	0.26	0.44	0.48	0.30	0.39

	M_{us} $\beta_h = 1$	$M_{com}^{0.5}$ $\beta_h = 0.5$	M_{com1} $\widehat{\beta}_h \in R$	M_{com2}^{tv} $\widehat{\beta}_{t,h} \in R$	M_{com3} $\widehat{\beta}_h \in [0,1]$	M_{com4}^{tv} $\widehat{\beta}_{t,h} \in [0,1]$
			Panel B: Leverage < 50%			
γ**=3**						
h=1 month	0.02	0.12	0.13	0.18	0.12	0.12
h=3 month	-0.01	-0.06	0.11	0.20	-0.01	0.08
h=6 month	0.06	0.03	0.12	0.30	0.06	0.11
h=12 month	0.39	0.03	0.45	0.94	0.39	0.45
h=24 month	0.32	0.23	0.35	0.71	0.32	0.43
h=60 month	0.52	0.43	0.55	0.95	0.52	0.53
γ**=5**						
h=1 month	-0.03	-0.12	0.00	0.11	-0.03	0.02
h=3 month	0.21	0.05	0.29	0.35	0.21	0.31
h=6 month	0.16	0.02	0.22	0.33	0.16	0.29
h=12 month	0.42	0.20	0.57	0.67	0.42	0.39
h=24 month	0.36	0.10	0.59	0.83	0.36	0.45
h=60 month	0.47	0.19	0.51	0.76	0.47	0.60
γ**=10**						
h=1 month	0.18	0.05	0.24	0.29	0.18	0.22
h=3 month	0.10	0.00	0.21	0.26	0.10	0.13
h=6 month	0.04	0.01	0.12	0.27	0.04	0.20
h=12 month	0.23	0.10	0.36	0.81	0.23	0.29
h=24 month	0.45	0.45	0.48	0.74	0.45	0.56
h=60 month	0.47	0.41	0.51	0.59	0.47	0.54

	M_{us} $\beta_h = 1$	$M_{com}^{0.5}$ $\beta_h = 0.5$	M_{com1} $\widehat{\beta}_h \in R$	M_{com2}^{tv} $\widehat{\beta}_{t,h} \in R$	M_{com3} $\widehat{\beta}_h \in [0,1]$	M_{com4}^{tv} $\widehat{\beta}_{t,h} \in [0,1]$
Panel C: Leverage $<$ 100%						
γ**=3**						
1 month	-0.07	-0.09	0.09	0.06	-0.07	-0.01
3 month	-0.02	0.09	0.16	0.25	0.09	0.16
6 month	0.13	-0.01	0.23	0.30	0.13	0.21
12 month	0.40	0.18	0.50	0.88	0.40	0.77
24 month	0.28	0.28	0.34	0.78	0.28	0.41
60 month	0.48	0.44	0.57	1.52	0.48	0.70
γ**=5**						
1 month	0.14	0.04	0.13	0.17	0.14	0.17
3 month	0.13	0.09	0.19	0.26	0.13	0.18
6 month	0.11	0.06	0.32	0.35	0.11	0.27
12 month	0.34	0.28	0.53	0.57	0.34	0.41
24 month	0.38	0.28	0.57	0.76	0.38	0.63
60 month	0.46	0.38	0.59	0.75	0.46	0.62
γ**=10**						
1 month	0.09	0.03	0.14	0.28	0.09	0.12
3 month	0.11	0.00	0.16	0.28	0.11	0.18
6 month	0.03	0.00	0.12	0.67	0.03	0.22
12 month	0.31	0.23	0.39	0.79	0.31	0.65
24 month	0.45	0.43	0.47	0.74	0.45	0.61
60 month	0.47	0.51	0.52	0.57	0.47	0.56

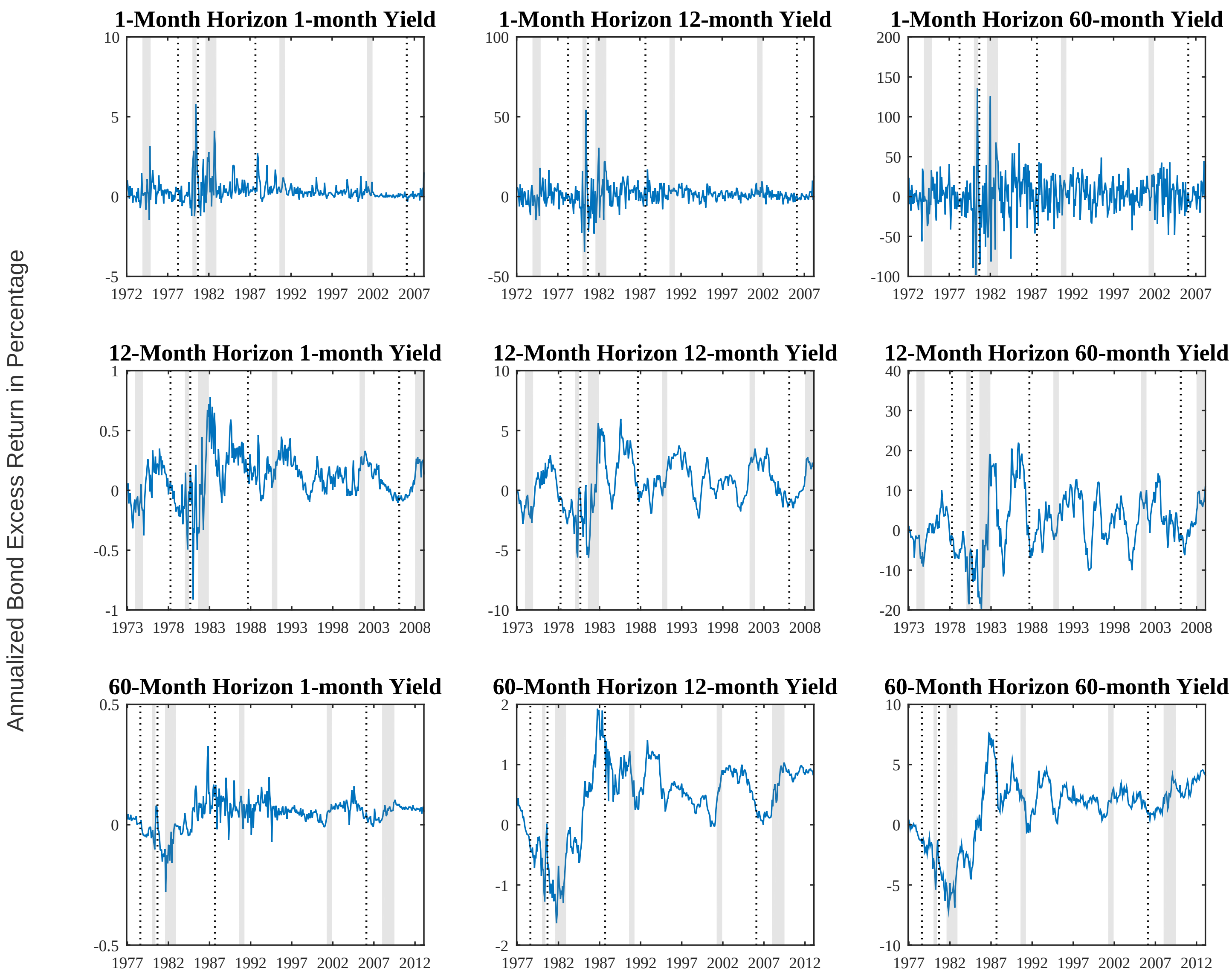
This table presents the annualized manipulation-proof performance measure (MPPM, Goetzmann, Ingersoll, Spiegel, and Welch, 2007) for portfolio decisions based on out-of-sample forecasts of bond excess returns. For an investment horizon h, the mean-variance investors with different levels of risk aversion (γ=3, 5, 10) select $(h+1)$-, $(h+3)$-, $(h+6)$-, $(h+12)$-, $(h+24)$-, $(h+60)$-month bonds and h-month bond based on the predictive returns implied by a given model. I rule out short-selling (negative portfolio weights), and also impose constraints on total portfolio leverage in order to avoid implausible positions. We consider leverage ratio of 0 (Panel A), 50% (Panel B), and 100% (Panel C). For example, leverage=50% indicates that the investor cannot borrow more than 50% of his or her total wealth. I report results for six investment horizons based on six alternative models relative to the benchmark spanned model (M_{span}): the preferred unspanned macro model (M_{us}), and the forecast combinations of the two models. The forecast combinations with equal weights ($M_{com}^{0.5}$). The forecast combinations based on the estimated optimal weights with respect to mean-variance loss function. I consider both constant (M_{com1}, $\widehat{\beta}_h \in R$) and time-varying optimal weights (M_{com2}^{tv}, $\widehat{\beta}_{t,h} \in R$), and they can be constrained to $[0,1]$ (M_{com3}, $\widehat{\beta}_h \in [0,1]$; M_{com4}^{tv}, $\widehat{\beta}_{t,h} \in [0,1]$) or not. Positive numbers indicate that portfolio based on the forecasts from the alternative model can earn risk-adjusted return in excess of portfolio based on the forecasts from model M_{span}. All MPPMs are reported in percentage.

Figure 1.1 Term Structure of Yields and Macro Variables



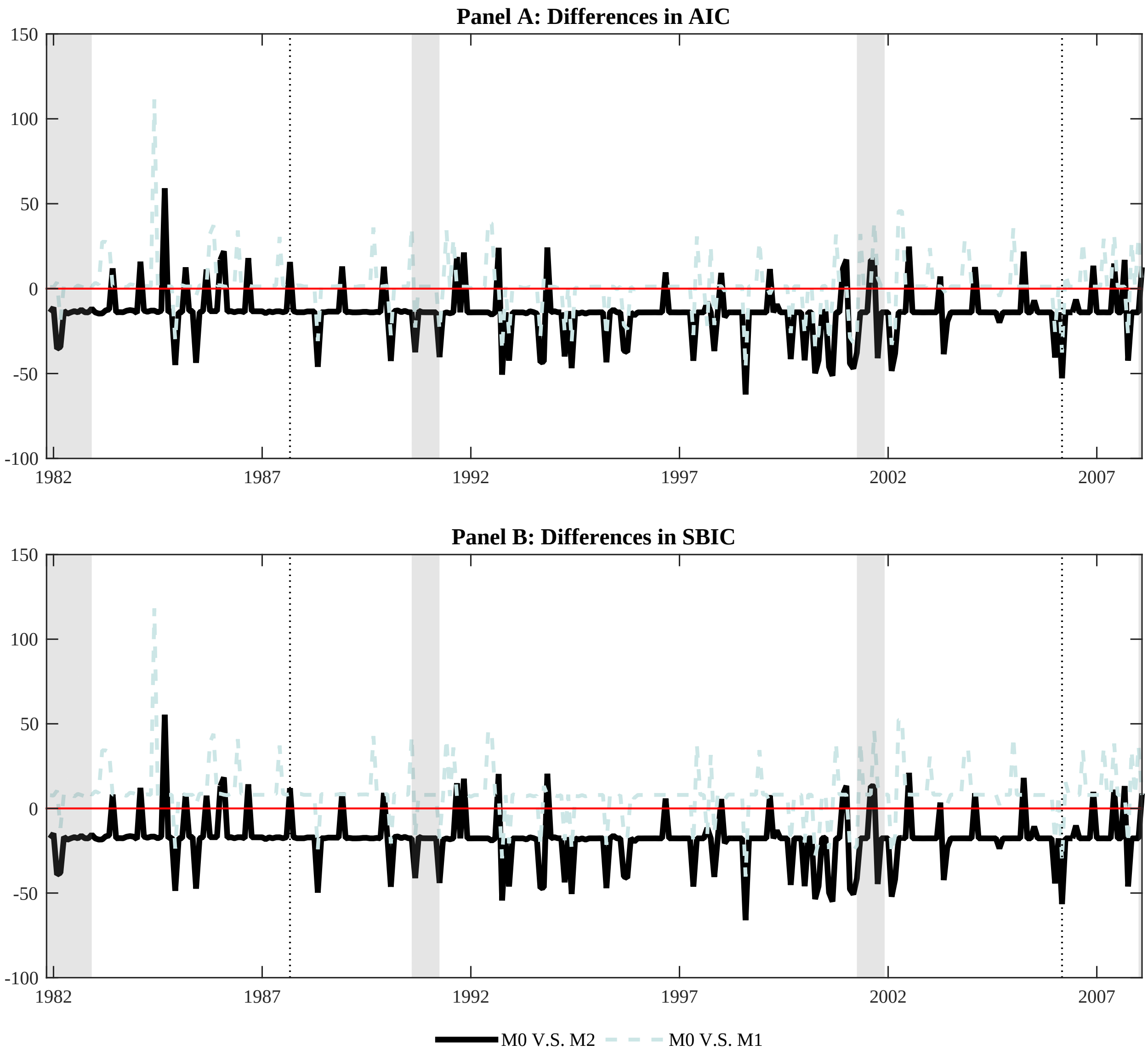
This figure plots the first three principal components (PC1, PC2, PC3) of the term structure of zero coupon yields (Panel A) and the macro variables: real activity growth and inflation (Panel B). Following Joslin, Priebsch, and Singleton (2014), I use three-month moving average of the Chicago Fed National Activity Index as a measure of real activity growth (GRO). Inflation is defined as the 12-month moving average of core CPI inflation. The shaded areas are the National Bureau of Economic Research (NBER) dated recessions. The vertical lines indicate the beginning of the term by subsequent Fed Chairmen: Miller, Volcker, Greenspan, and Bernanke.

Figure 1.2 Bond Excess Returns



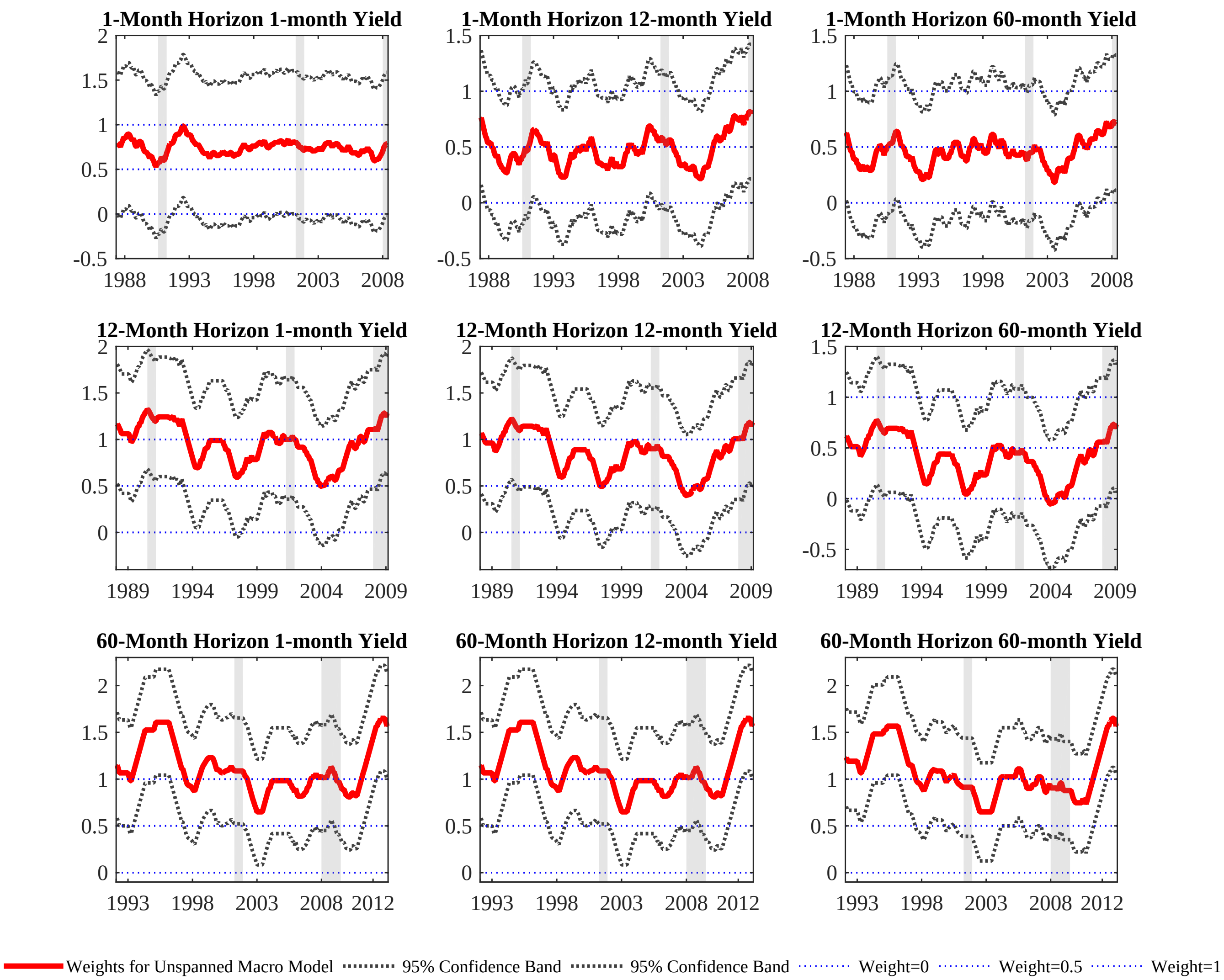
This figure plots nine forecast horizon h (1-, 12-, 60-month) and maturity $n-h$ (1-, 12-, 60-month) combinations of realized bond excess returns. The panels in each row represent the results for a given forecast horizon, and the panels in each column represent the results for a given maturity. The returns are annualized and expressed in percentage. The shaded areas are the National Bureau of Economic Research (NBER) dated recessions. The vertical lines indicate the beginning of the term by subsequent Fed Chairmen: Miller, Volcker, Greenspan, and Bernanke.

Figure 1.3 Model Selection: Risk Premium Restrictions



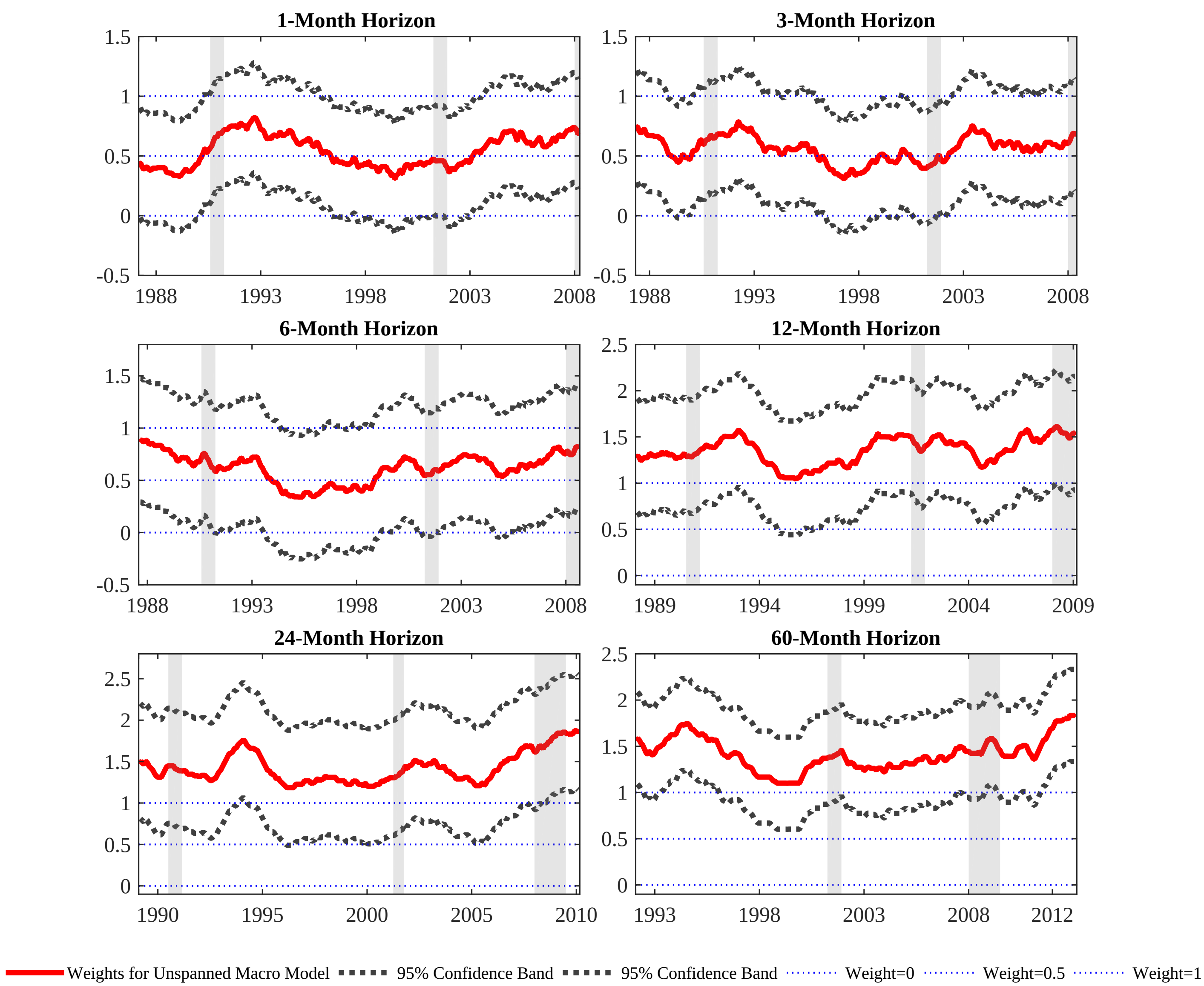
This figure plots the differences in Akaike (1973) information criteria (AIC) and Schwarz (1978) Baysian information criteria (SBIC) between different specifications of the unspanned macro model. We compute AIC and SBIC based on 10-year rolling window estimation scheme for the maximally flexible unspanned model (M_0), the unspanned model with reduced-rank restrictions $\zeta = 1$ (M_1) and $\zeta = 2$ (M_2). The solid line in Panel A plots the differences in AIC between M_0 and M_2. The dotted line in Panel A plots the differences in AIC between M_0 and M_1. Panel B plots the results for SBIC. Positive values indicate that M_0 is preferred. Negative values indicate that reduced-rank restricted model is superior. The shaded areas are the National Bureau of Economic Research (NBER) dated recessions. The vertical lines indicate the beginning of the term by subsequent Fed Chairmen: Greenspan and Bernanke.

Figure 1.4 Unconstrained Time-Varying Optimal Weights: Sum Squared Error Loss



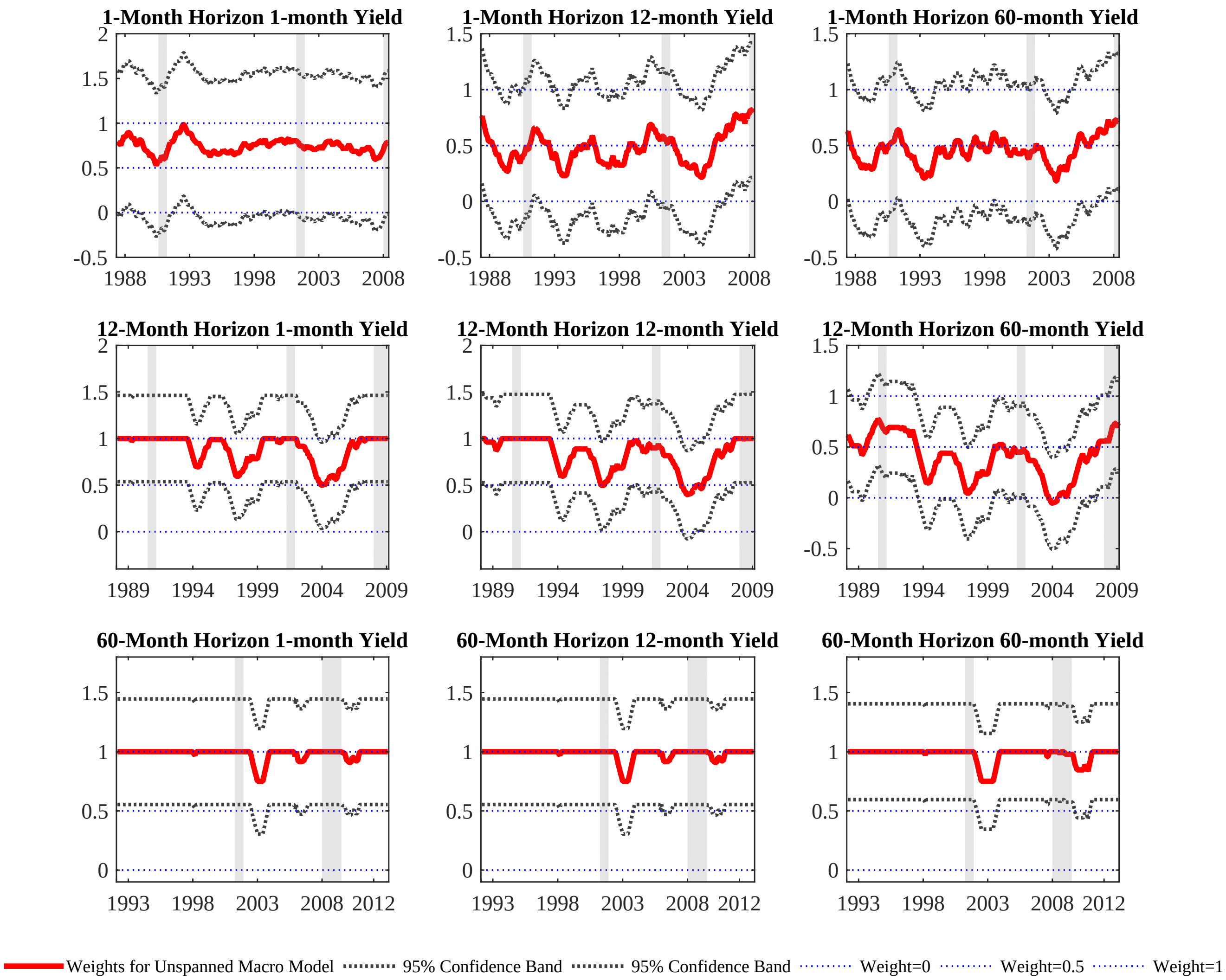
This figure plots the time-varying optimal weights $\widehat{\beta}^n_{t,h}$ for the preferred unspanned macro model (M_{us}) in the forecast combination with respect to sum squared error (SSE) loss function. I plot the estimated weights and the 95% confidence bands for 9 forecast horizon h (1-, 12-, 60-month) and maturity $n-h$ (1-, 12-, 60-month) combinations of bond excess returns. $\widehat{\beta}^n_{t,h}$ is computed using the out-of-sample h-period forecasts of n-maturity bond excess returns with a smoothing window of size $d = 315 \times 0.2 = 63$, which corresponds to a critical value of 3.179 for 95% confidence bands as shown in Carriero and Giacomini (2011). $\widehat{\beta}^n_{t,h} \in R$ in the estimation. The panels in each row represent the results for a given forecast horizon, and the panels in each column represent the results for a given maturity. The shaded areas are the National Bureau of Economic Research (NBER) dated recessions.

Figure 1.5 Unconstrained Time-Varying Optimal Weights: Mean-Variance Utility Loss



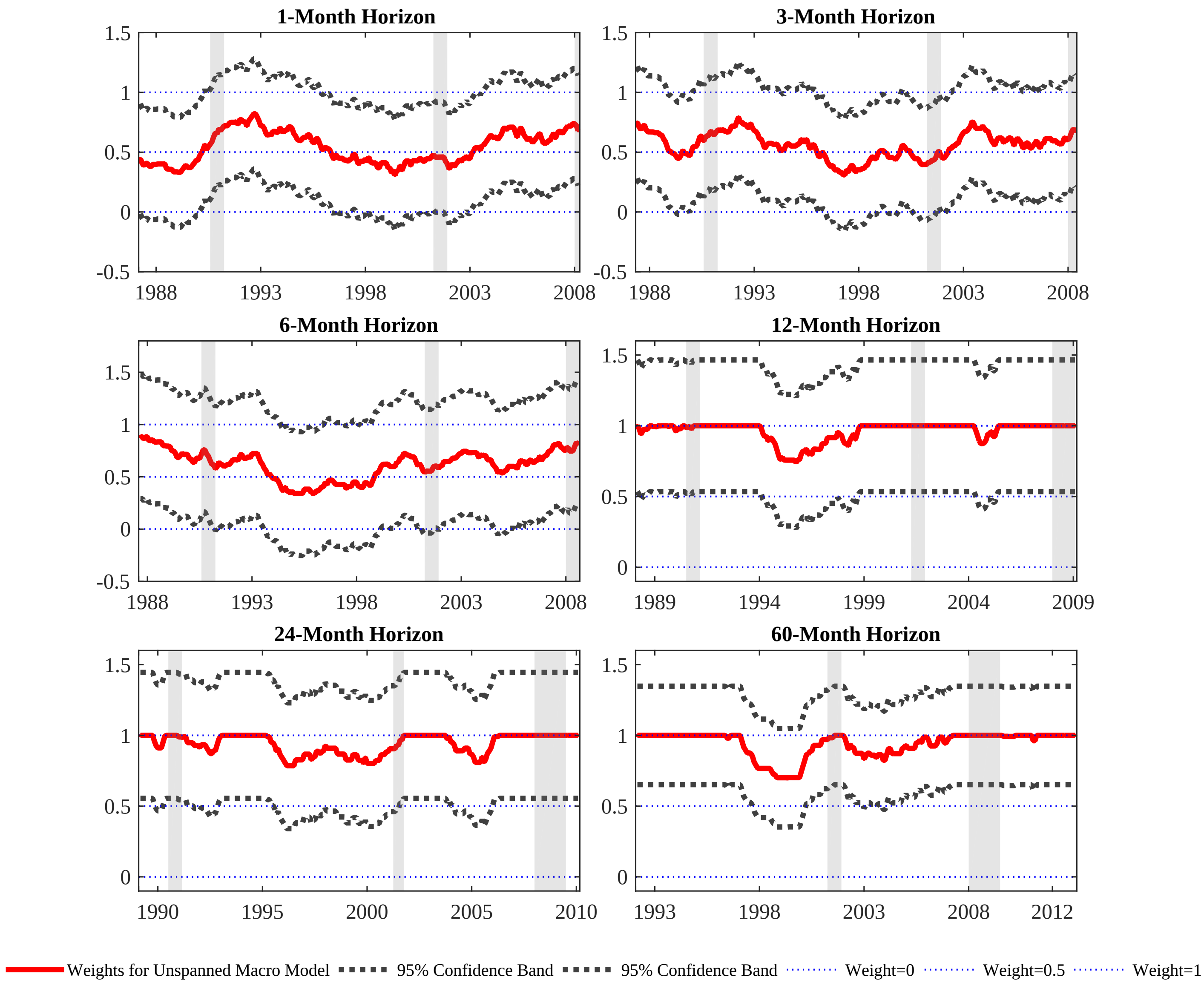
This figure plots the time-varying optimal weights $\widehat{\beta}_{t,h}$ for the preferred unspanned macro model (M_{us}) in the forecast combination with respect to the mean-variance utility loss function. I plot the estimated weights and the 95% confidence bands for different forecast horizons. Each panel represents the results for one forecast horizon. $\widehat{\beta}_{t,h}$ is computed using the out-of-sample h-period forecasts of bond portfolio excess returns with a smoothing window of size $d = 315 \times 0.2 = 63$, which corresponds to a critical value of 3.179 for 95% confidence bands as shown in Carriero and Giacomini (2011). $\widehat{\beta}_{t,h} \in R$ in the estimation. The shaded areas are the National Bureau of Economic Research (NBER) dated recessions.

Figure 1.A.1 Constrained Time-Varying Optimal Weights: Sum Squared Error Loss.



This figure plots the time-varying optimal weights $\widehat{\beta}^n_{t,h}$ for the preferred unspanned macro model (M_{us}) in the forecast combination with respect to sum squared error (SSE) loss function. I plot the estimated weights and the 95% confidence bands for 9 forecast horizon h (1-, 12-, 60-month) and maturity $n-h$ (1-, 12-, 60-month) combinations of bond excess returns. $\widehat{\beta}^n_{t,h}$ is computed using the out-of-sample h-period forecasts of n-maturity bond excess returns with a smoothing window of size $d = 315 \times 0.2 = 63$, which corresponds to a critical value of 3.179 for 95% confidence bands as shown in Carriero and Giacomini (2011). $\widehat{\beta}^n_{t,h}$ is constrained to $[0, 1]$ in the estimation. The panels in each row represent the results for a given forecast horizon, and the panels in each column represent the results for a given maturity. The shaded areas are the National Bureau of Economic Research (NBER) dated recessions.

Figure 1.A.2 Constrained Time-Varying Optimal Weights: Mean-Variance Utility Loss.



This figure plots the time-varying optimal weights $\widehat{\beta}_{t,h}$ for the preferred unspanned macro model (M_{us}) in the forecast combination with respect to the mean-variance utility loss function. I plot the estimated weights and the 95% confidence bands for different forecast horizons. Each panel represents the results for one forecast horizon. $\widehat{\beta}_{t,h}$ is computed using the out-of-sample h-period forecasts of bond portfolio excess returns with a smoothing window of size $d = 315 \times 0.2 = 63$, which corresponds to a critical value of 3.179 for 95% confidence bands as shown in Carriero and Giacomini (2011). $\widehat{\beta}_{t,h}$ is constrained to $[0, 1]$ in the estimation. The shaded areas are the National Bureau of Economic Research (NBER) dated recessions.

Chapter 2

Information in the Term Structure: A Forecasting Perspective

2.1 Introduction

Modeling the term structure is a topic of great practical importance for investors as well as monetary policy makers. The term structure contains valuable information about expectations of future interest rates and term premia, and the identification of these components critically rests on the specification of the term structure model. Much of the term structure literature therefore focuses on model specification and estimation.

We depart from the existing literature by documenting the importance of the information used in the estimation of the term structure model. Specifically, while the cross-sectional and time-series dimension of our sample is very similar to the existing literature, we differ by considering a different objective (loss) function, which is explicitly motivated from a forecasting perspective. This approach is motivated by the literature on bond return prediction. Cochrane and Piazzesi (2005) and Duffee (2011a), among others, argue that a hidden factor not captured by the traditional level, slope, and curvature factors helps in predicting excess bond returns. We build on this insight and estimate no-arbitrage term structure models by specifying a loss function at the estimation stage which is motivated from a forecasting perspective and aligned with the out-of-sample evaluation measure. Our null hypothesis is that when we estimate the no-arbitrage model using this alternative loss function, the model properties will capture and reflect this hidden information.

It is well known in the statistics literature that the specification of the loss function critically influences model estimates and properties. Indeed, the specification of a loss function implicitly amounts to the specification of a statistical model, because the loss function determines how different forecast errors are weighted (see Engle, 1993; Granger, 1993; Weiss, 1996; Elliott and Timmermann, 2008). Since the loss function is an important element in the process of delivering a forecast, it is an integral part of model specification. If a particular criterion is used to evaluate forecasts, it should also be used at the estimation stage.[1] Estimating a model under one loss function and evaluating it under another amounts to changing the model specification without allowing the parameter estimates to adjust.

Motivated by these insights, we estimate Affine Term Structure Models (ATSMs) while aligning the loss functions for in-sample estimation and out-of-sample evaluation.[2] We estimate the model by minimizing the squared forecasting errors for a given forecast horizon, and we refer to these estimates as based on the forecasting loss function. We compare the implications of these estimates with the implications of estimates obtained by minimizing the mean-squared error loss function based on current yields, which we refer to as the standard loss function. We analyze how the choice of loss function affects model implications for factor dynamics, expected short rates and term premia, as well as its out-of-sample forecasting performance. Our empirical analysis focuses on three-factor ATSMs. We first present results using the Joslin, Singleton, and Zhu (2011, henceforth referred to as JSZ) canonical form of

[1] An extensive literature studies the theoretical properties of optimal forecasts under asymmetric loss functions and documents that forecast errors have different properties under different loss functions. See for example Patton and Timmermann (2007a, 2007b), Elliott, Komunjer, and Timmermann (2005, 2008), and Christoffersen and Diebold (1996, 1997).

[2] The empirical literature on ATSMs is very extensive. See Vasicek (1977), Cox, Ingersoll, and Ross (1985), Chen and Scott (1992), Longstaff and Schwartz (1992), Duffie and Kan (1996), and Dai and Singleton (2000) for important contributions.

the Gaussian three-factor model, and show the robustness of our results using the arbitrage-free Nelson Siegel (AFNS) specification proposed by Christensen, Diebold, and Rudebusch (2011, henceforth referred to as CDR), as well as stochastic volatility models.

We compare the factor dynamics implied by the forecasting and standard loss functions. We confirm that with the forecasting loss function, the factors differ from the traditional level, slope, and curvature factors, especially for the curvature (third) factor. Consistent with Cochrane and Piazzesi (2005), this factor confirms the importance of the fourth principal component of yields for forecasting the term structure. This information remains hidden when estimating using a standard loss function. Term premia obtained using the forecasting loss function are countercyclical, which is consistent with the literature (see Campbell and Cochrane, 1999; Wachter, 2006; Harvey, 1989; Cochrane and Piazzesi, 2005), but the forecasting loss function generates larger term premia than the standard loss function, especially at the time of the 1970s oil shocks.

When using the forecasting loss function in estimation, we find substantial improvements in out-of-sample forecasting performance, especially for longer forecast horizons and for short and medium maturities. For example, for the Gaussian model, the improvement in the root mean square error (RMSE) for short maturity yields (three-month, six-month, and one-year yields) is about 14% on average across different forecast horizons, which corresponds to an out-of-sample R-square of 26%. For longer forecast horizons (nine- and twelve-month forecast horizons), the improvement is about 11% on average across maturities, which corresponds to an out-of-sample R-square of 21%. These results confirm the insights of Granger (1993) and Engle (1993) that the choice of loss function is an integral part of model specification and that aligning the estimation loss function with the loss function used for out-of-sample model evaluation improves out-of-sample forecasting performance. See also Cochrane (2005, p. 298) for a related discussion. Consistent with these insights, we also confirm a trade-off between in-sample and out-of-sample fit: the parameters estimated using the forecasting loss function do not improve the in-sample fit based on parameters estimated using the standard loss function. This trade-off is especially relevant at longer forecast horizons.

Our results contribute to several strands of literature. Much of the recent literature on the estimation of ATSMs focuses on innovative estimation approaches to address well-known identification problems inherent in the estimation of ATSMs.[3] We take a different perspective and analyze how the choice of loss function affects a given model's implications for factor dynamics, expected short rates and term premia, as well as its out-of-sample forecasting performance.[4] Our contribution is therefore complementary to most of the recent literature on ATSMs, because the insight that estimation using the forecasting loss function will lead to better out-of-sample performance is valid regardless of the estimation method. The existing literature on forecasting Treasury yields focuses almost exclusively on comparisons of the forecasting performance for alternative specifications of the term structure model.[5] The forecasting exercise is not explicitly taken into account at the estimation stage.

We also contribute to the empirical literature that documents the importance of the loss function. Part of this literature focuses on autoregressive specifications for economic variables such as inflation and unemployment and compares iterated forecasts, made using a one-period ahead model, with direct forecasts, made using a horizon-specific model.[6] These studies also consider a horizon-specific loss function but differ from our approach because they compare two loss functions that are both designed for forecasting. This literature concludes that the performance of the horizon-specific loss function is mixed at best. We find instead that the horizon-specific loss function performs much better. This may be due to the fact

[3]On identification problems in these models, see for example Duffee (2011b), Duffee and Stanton (2012), and Hamilton and Wu (2012). For examples of methods that address these identification problems, see JSZ (2011), Hamilton and Wu (2012), Adrian, Crump, and Moench (2013), Diez de los Rios (2014), Bauer, Rudebusch, and Wu (2012), and Creal and Wu (2015).

[4]Adrian, Crump, and Moench (2013) and Sarno, Schneider, and Wagner (2014) estimate model parameters in ATSMs using an objective function that takes into account excess returns for different horizons. This approach is similar to ours in the sense that the implied loss function is different from the standard loss function based on yields. However, their implied loss function is different from ours, and not necessarily optimal from a forecasting perspective.

[5]See Duffee (2002), Ang and Piazzesi (2003), Diebold and Li (2006), Bowsher and Meeks (2008), and Christensen, Diebold, and Rudebusch (2011) for examples of studies that focus on point forecasts. See Hong, Li, and Zhao (2004), Egorov, Hong, and Li (2006), and Shin and Zhong (2013) for studies that focus on density forecasts.

[6]See Ang, Piazzesi, and Wei (2005), Kang (2003), Liu (1996), and Marcellino, Stock, and Watson (2006).

that we model yields and autoregressive representations of latent state variables rather than macroeconomic variables. Also, in our implementation the competing specification is not designed for forecasting.

The paper proceeds as follows. Section 2 compares the forecasting loss function with the standard loss function based on yields. Section 3 presents the data. Section 4 compares the forecasting performance of different loss functions based on the estimation of a Gaussian model using the JSZ canonical form. Section 5 discusses the impact of the loss function on implied state variables, parameter estimates, and estimates of future short rates and term premia. Section 6 presents robustness results using the Gaussian AFNS model and stochastic volatility models with latent factors, and also presents robustness results using a mean absolute error loss function. Section 7 concludes.

2.2 Loss Functions for Term Structure Estimation

Given term structure data for months $t = 1, ..., T$ on maturities $n = 1, ..., N$, the parameters Θ of a term structure model are typically estimated using a loss function that minimizes a well-defined distance between the observed yields y_t^n and the model yield, which we denote here by $\widehat{y}_{t|t}^n(\Theta)$ to emphasize that the model yield is computed using the state variables at time t. In general, the notation $\widehat{y}_{t+k|t}^n$ indicates a model-implied yield at time $t+k$ computed using information up to time t. We use this type of loss function as a benchmark. Several such loss functions can in principle be used, but we limit ourselves to loss functions that are based on the difference between observed and model yields.[7] We estimate the term structure parameters Θ by minimizing the root-mean-squared-error based on observed and

[7] Alternatively, loss functions based on relative errors or other transformations of yields can be studied, but in the term structure literature this is less critical than for other applications, such as derivative securities.

model yields:[8]

$$RMSE(\Theta) = \sqrt{\frac{1}{NT}\sum_{n=1}^{N}\sum_{t=1}^{T}(\widehat{y}^n_{t|t}(\Theta) - y^n_t)^2}. \tag{2.1}$$

Estimating the model parameters by optimizing the log likelihood or the root-mean-squared-error provides the best possible in-sample fit. Our focus is not on in-sample fit but rather on forecasting. To improve forecasting performance, we deviate from the benchmark implementation by aligning the loss functions for in-sample and out-of-sample evaluation, as suggested by Granger (1993) and Weiss (1996). The choice of loss function at the estimation stage should therefore reflect that out-of-sample forecasting is the objective of the empirical exercise. The out-of-sample forecasting performance for the n-maturity yield with forecast horizon k is evaluated using

$$RMSE\_OS_{n,k} = \sqrt{\frac{1}{T-k}\sum_{t=1}^{T-k}(\widehat{y}^n_{t+k|t}(\Theta) - y^n_{t+k})^2}, \tag{2.2}$$

where y^n_{t+k} is the observed n-maturity yield at time $t+k$ and $\widehat{y}^n_{t+k|t}(\Theta)$ is the model-predicted k-period ahead n-maturity yield based on the parameter set Θ, which is estimated at time t.

To align the loss function at the estimation stage with the out-of-sample loss function, we therefore estimate the models for a given forecast horizon k by minimizing the following loss function:

$$OS\_RMSE_k(\Theta) = \sqrt{\frac{1}{N(T-k)}\sum_{n=1}^{N}\sum_{t=k+1}^{T}(\widehat{y}^n_{t|t-k}(\Theta) - y^n_t)^2}. \tag{2.3}$$

We use the terminology standard loss function for (2.1) and forecasting loss function for (2.3) to emphasize that only one of the loss functions is designed for forecasting. A related litera-

[8]In-sample estimation of term structure models usually maximizes the log likelihood. We use the root-mean-squared error instead to facilitate the comparison with the forecasting loss function. If the measurement errors are normally distributed and constant across maturities, the likelihood simply scales the mean-squared error. For other cases, optimizing the likelihood and the mean squared error gives very similar results.

ture (Kang, 2003; Liu, 1996; Marcellino, Stock, and Watson, 2006) forecasts macroeconomic variables using autoregressive specifications. These studies refer to the forecast from the horizon-specific loss function as the direct forecast and to the forecast from the one-period ahead model as the iterated forecast.

2.3 Data

We use monthly data on continuously compounded zero-coupon bond yields with maturities of three and six months, and one, two, three, four, five, ten and twenty years, for the period April 1953 to December 2012. Our results are qualitatively similar if we omit the longer maturities. The three- and six-months yields are obtained from the Fama CRSP Treasury Bill files, and the one- to five-year bond yields are obtained from the Fama CRSP zero coupon files. The ten- and twenty-year maturity zero-coupon yields are obtained from the H.15 data release of the Federal Reserve Board of Governors.[9]

Table 1 shows that, on average, the yield curve is upward sloping, and the volatility of yields is relatively lower for longer maturities. The yields for all maturities are highly persistent, with slightly higher autocorrelation for long-term yields than for short-term yields. Yields exhibit mild excess kurtosis and positive skewness for all maturities.

2.4 Results for the Gaussian Model

Our empirical analysis compares the forecasting performance of the standard loss function (2.1) and the forecasting loss function (2.3). The choice of loss function is potentially important for all term structure models, but we limit ourselves to three-factor ATSMs because

[9]The Federal Reserve database provides constant maturity treasury (CMT) rates for different maturities. The ten- and twenty-year CMT rates are converted into zero-coupon yields using the piecewise cubic polynomial. Data on 20-year yields are not available from January 1987 through September 1993. We fill this gap by computing the 20-year CMT forward yield using 10-year and 30-year CMT yields. Our long maturity dataset is comparable to the Treasury yield curve published by the Federal Reserve (Gürkaynak, Sack, and Wright, 2007), but our approach allows for a longer sample period.

of their central place in the literature and their tractability.

The estimation of ATSMs is challenging due to the high level of nonlinearity in the parameters (Duffee, 2011b; Duffee and Stanton, 2012; Kim and Orphanides, 2012). In recent work, JSZ (2011) develop a canonical representation that allows for stable and tractable estimation of the $A_0(3)$ model and addresses these identification problems. We therefore begin our analysis using their representation of the Gaussian model. In this section, we first present the general framework of the $A_0(3)$ model and provide a brief discussion of the $A_0(3)$ canonical representation in JSZ. Subsequently, we present the empirical results. Finally, we discuss the trade-off between in-sample and out-of-sample fit.

2.4.1 The Three-Factor Affine Gaussian Model

In the term structure literature, affine term structure models (ATSMs) have received significant attention because of their rich structure and tractability. The existing literature has concluded that at least three factors are needed to explain term structure dynamics (see for example Litterman and Scheinkman, 1991; Knez, Litterman, and Scheinkman, 1994). A three-factor affine Gaussian model $A_0(3)$ can be written as

$$X_{t+1} = K_0^P + K_1^P X_t + \Sigma \varepsilon_{t+1}^P, \tag{2.4}$$

$$X_{t+1} = K_0^Q + K_1^Q X_t + \Sigma \varepsilon_{t+1}^Q, \tag{2.5}$$

$$r_t = \rho_0 + \rho_1 X_t. \tag{2.6}$$

The state variables X_t follow a first-order Gaussian vector autoregression under P- and Q-measures. ε_t^P and ε_t^Q are assumed to be distributed $N(0, I_3)$. r_t is the one-period spot interest rate, and $\Sigma\Sigma'$ is the conditional covariance matrix of X_t. Assuming absence of arbitrage, the stochastic discount factor is an exponentially affine function of the state variables. We adopt the essentially affine specification for the price of risk, as in Duffee (2002).

With these dynamics, the model-implied continuously compounded yields $\widehat{y}_t$ are given by (see Duffie and Kan, 1996)

$$\widehat{y}_t = A(\Theta^Q) + B(\Theta^Q)X_t, \tag{2.7}$$

where the $N \times 1$ vector $A(\Theta^Q)$, and the $N \times 3$ matrix $B(\Theta^Q)$ are functions of the parameters under the Q-dynamics, $\Theta^Q = \{K_0^Q, K_1^Q, \rho_0, \rho_1, \Sigma\}$, through a set of recursive equations. Recall that N denotes the number of available yields in the term structure.

For estimation using the forecasting loss function (2.3), we need the model's prediction of the k-period ahead n-maturity yield, based on parameter estimates at time t, which is given by

$$\begin{aligned} \widehat{y}^n_{t+k|t}(\Theta) &= A_n(\Theta^Q) + B_n(\Theta^Q)\widehat{X}_{t+k|t} \\ &= A_n(\Theta^Q) + B_n(\Theta^Q) f(X_t, k; K_0^P, K_1^P), \end{aligned} \tag{2.8}$$

where $A_n(\Theta^Q)$ is the n^{th} element of $A(\Theta^Q)$, $B_n(\Theta^Q)$ is the n^{th} row of $B(\Theta^Q)$, $\Theta = \{\Theta^Q, K_0^P, K_1^P\}$ and f is given by

$$f(X_t, k; K_0^P, K_1^P) = (I_3 + K_1^P + ... + (K_1^P)^{k-1})K_0^P + (K_1^P)^k X_t.$$

Under the forecasting loss function, we predict the state variables k-period ahead given information up to time t using the P-parameters, $\Theta^P = \{K_0^P, K_1^P\}$, through function f. The forecasting loss function is therefore tailored to the forecast horizon.[10]

[10]Horizon-specific loss functions are also used in the multiperiod time series forecasting literature (Weiss, 1991, 1996; Lin and Granger, 1994; Marcellino, Stock, and Watson, 2006; Ang, Piazzesi, and Wei, 2005). In this literature, the theoretical studies emphasize the advantages of the horizon-specific forecast (Cox, 1961). The empirical evidence is mixed (Kang, 2003; Marcellino, Stock, and Watson, 2006). See also Cochrane (2005, p. 298) for a related discussion.

2.4.2 The JSZ Canonical Form

The estimation of ATSMs is challenging due to the high level of nonlinearity in the parameters (Duffee, 2011b; Duffee and Stanton, 2012; Kim and Orphanides, 2012).[11] Identification problems are especially relevant for our analysis because they can easily distort comparisons of loss functions. If the estimation using the standard loss function (2.1) does not lead to the global optimum, we may overestimate the advantages provided by the forecasting loss function (2.3). The reverse is of course also possible. We adopt the canonical representation of JSZ (2011) which allows for stable and tractable estimation of the $A_0(3)$ model and addresses these identification problems. We briefly discuss their approach and refer to Appendix A and JSZ (2011) for details.

The state variables X_t under the JSZ normalization are the perfectly priced portfolios of yields, $PO_t = Wy_t$. W denotes the portfolio weights, a $3 \times N$ matrix. PO_t are observable, and thus the parameters governing the P-dynamics can be estimated separately from the parameters governing the Q-dynamics. JSZ demonstrate that the ordinary least squares (OLS) estimates of K_0^P and K_1^P from the observed factors PO_t nearly recover the maximum likelihood (ML) estimates of K_0^P and K_1^P from the P- and Q-dynamics jointly, to the extent that $Wy_t \approx W\widehat{y}_t$. As noted by JSZ, the best approximation is obtained by choosing W_0 such that $W_0 y_t = PC_t$, the first three principal components of the observed term structure of yields.[12] Therefore, under JSZ canonical form $X_t = PC_t$.

JSZ show that $A(\Theta^Q)$ and $B(\Theta^Q)$ are ultimately functions of $\Theta^Q = \{r_\infty^Q, \lambda^Q, \Sigma\}$, where r_∞^Q is a scalar related to the long-run mean of the short rate under risk neutral measure and λ^Q, a 3×1 vector, represents the ordered eigenvalues of K_1^Q. For a three-factor model, there are in total $10 + N$ parameters to be estimated (1 for r_∞^Q, 3 for λ^Q, 6 for Σ, and N for the

[11] Dai and Singleton (2000) argue that not all parameters are well identified, and that rotation and normalization restrictions need to be imposed. Even with the Dai-Singleton normalization, it is possible to end up within a parameter space that is locally unidentified. See for instance the discussions in Hamilton and Wu (2012), Collin-Dufresne, Goldstein, and Jones (2008) and Aït-Sahalia and Kimmel (2010).

[12] Strictly speaking, the OLS estimates are exactly equal to the ML estimates only if one assumes that the yields are measured without errors. Empirically, JSZ show that the use of the principal components ensures that the OLS estimates and ML estimates are nearly identical.

diagonal variance-covariance matrix of the measurement errors).

The JSZ canonical form provides important computational advantages, because it allows the estimation to be performed directly on the principal components of the observed yields, which in turn allows factorization of the likelihood and isolates the subset of parameters governing the Q-dynamics. This canonical form therefore dramatically reduces the difficulties that typically arise in the search for the global optimum.

When implementing the JSZ canonical form using the forecasting loss function, we estimate Θ^P and Θ^Q simultaneously by minimizing the forecasting loss function (2.3). Note that the P-parameters determine the properties of the state variables, which are important for forecasting yields, as seen in equation (2.8). In contrast, these parameters do not play a role under the standard loss function (2.1) for the JSZ normalization.[13] This is a critical difference between the loss functions. When using the forecasting loss function, we cannot determine K_0^P and K_1^P from the OLS estimates, because the forecasting loss function depends on all parameters simultaneously. This makes it harder to implement the model, but our main interest is if it improves the model's forecasting performance.

2.4.3 Recursive Estimation

We provide an empirical comparison of the forecasting performance of the forecasting loss function (2.3) and the standard loss function (2.1). Our procedure for examining the out-of-sample forecasts of the JSZ with forecasting loss function is as follows. We proceed recursively with estimation and forecasting, each time adding one month to the estimation sample. At each time t and for each forecast horizon k, we estimate the model using data up to and including t. Our first estimation uses the first half of the data, up to December 1982. We estimate the parameter sets Θ^P and Θ^Q for forecast horizon k by minimizing the k-period ahead squared forecasting errors, as expressed in equation (2.3), and forecast the k-period

[13] Note that JSZ do not minimize the mean-squared error but instead use maximum likelihood. However, the same argument applies: these parameters play no role in the standard likelihood function under the JSZ normalization.

ahead yields based on equation (2.8).

The recursion then proceeds: we add one month of data, re-estimate the parameters using information up to and including time $t+1$, and forecast the k-period ahead yields. We continue to update the sample in this way until time $T-k$, where T is the end of the sample, December 2012. Note that the estimation based on the forecasting loss function is forecast-horizon specific. At each time t, we have a different parameter set for each k.

The procedure for the JSZ with standard loss function (2.1) follows the same procedure, but this procedure is by definition not horizon-specific. Instead, at each time t one set of parameters is estimated that is used to generate forecasts for different horizons.

2.4.4 Variable Portfolio Weights

We find that the improvement in out-of-sample forecasting performance resulting from the forecasting loss function is small for the JSZ canonical specification. However, the implementation using the JSZ canonical form imposes a very important restriction. The portfolio weights W in $PO_t = Wy_t$ are given by W_0 such that $W_0 y_t = PC_t$. JSZ show that this restriction does not affect in-sample estimation results.[14] However, from a forecasting perspective, imposing these restrictions may mean that the parameters governing the dynamics of the state variables, K_0^P and K_1^P, do not have a strong incentive to move away from the OLS estimates, even if the OLS estimates are not optimal from the perspective of out-of-sample forecasting performance. Moreover, the time-series dynamics of the first three principal components may also be sub-optimal from a forecasting perspective. Therefore, we modify the estimation in JSZ by allowing the portfolio weights W to be estimated as parameters.

This approach is motivated by the literature on forecasting bond returns. Cochrane and Piazzesi (2005) suggest that the fourth principal component of the yield curve accounts for a large part of the predictability in bond returns. Moreover, the literature on the predictabil-

[14] We confirm this by performing a full sample one-time estimation of the JSZ with standard loss function and variable weights. The portfolio weights W converge to W_0. The first three principal components provide the best in-sample fit.

ity of bond excess returns shows that other variables, such as forward rates (Cochrane and Piazzesi, 2005), macroeconomic variables (Ludvigson and Ng, 2009; Cooper and Priestley, 2009; Cieslak and Povala, 2015; Joslin, Priebsch, and Singleton, 2014), and a hidden factor (Duffee, 2011a) also help predict excess bond returns. By allowing the weights to be free parameters, the estimation based on the forecasting loss function has more flexibility to search for the best possible state variables for the purpose of forecasting. This parameterization thus provides more flexibility to the forecasting loss function to determine the state variables that are best suited for out-of-sample forecasting.

The resulting econometric problem is somewhat more complex, and it is worth outlining it in more detail. First, consider the model-predicted k-period ahead n-maturity yield given parameter estimates $\Theta = \{\Theta^P, \Theta^Q, W\}$ at time t, which can be written as follows

$$\begin{aligned} \widehat{y}^n_{t+k|t}(\Theta) &= A_n(\Theta^Q) + B_n(\Theta^Q)\widehat{PO}_{t+k|t} \\ &= A_n(\Theta^Q) + B_n(\Theta^Q) f(y_t, k; K_0^P, K_1^P, W), \end{aligned} \tag{2.9}$$

where $A_n(\Theta^Q)$ is the n^{th} element of $A(\Theta^Q)$, $B_n(\Theta^Q)$ is the n^{th} row of $B(\Theta^Q)$, and f is given by

$$f(y_t, k; K_0^P, K_1^P, W) = (I_3 + K_1^P + ... + (K_1^P)^{k-1})K_0^P + (K_1^P)^k W y_t.$$

We estimate the JSZ representation with variable portfolio weights for each forecast horizon k by minimizing the forecasting loss function (2.3) with respect to Θ. By varying W, we construct the state variables as linear combinations of the observed term structure of yields, but they are not restricted to be the first three principal components of the observed yields.

We implement this estimation using a two-step procedure, taking full advantage of the estimation method proposed by JSZ, which typically converges in a few seconds. We start our estimation based on the forecasting loss function in equation (2.3) by using the converged JSZ estimates from the standard loss function (2.1) as initial values. Given these initial Θ^P and Θ^Q, the estimation is performed using the following steps.

1. For a given Θ^P and Θ^Q, we search for the best possible weights W among the linear combinations of yields that provide the lowest squared forecasting error in equation (2.3). The starting values for W are obtained as follows. For a given set of parameters (Θ^P and Θ^Q), if we assume that the forecast errors are all zeros, the portfolio weights can be inverted from the observed term structure of yields as shown in Appendix B. This linear mapping is not exact since not all yields can be forecasted perfectly, but these are useful starting values for faster convergence.

2. Once we obtain a W in step 1, we fix it and solve for the parameter set Θ^P and Θ^Q by minimizing the squared forecasting error. Note that in this step, the identification of Σ is achieved by using the time-series information of the state variables following JSZ (2011) and Joslin, Le, and Singleton (2013). Specifically, for a given weighting matrix W, we compute the linear combination of yields and treat the state variables as observables. Given Θ^P, we can then infer Σ from the time series of the observed factors.

3. Once we obtain the converged Θ^P and Θ^Q from the previous step, we go back to the first step, and the optimization goes back and forth between the two steps until it converges.

2.4.5 Empirical Results

Table 2 provides the empirical results. Panel A of Table 2 provides the RMSEs resulting from the JSZ canonical specification with forecasting loss function (2.3) and variable portfolio weights. Panel B presents the RMSEs from the JSZ empirical implementation with the standard loss function (2.1) and fixed portfolio weights.[15] We report the results for one-month to six-month, nine-month, and twelve-month forecast horizons, for all nine maturities

[15] We also investigate the implications of the loss function using a traditional implementation of the $A_0(3)$ model with latent factors. The empirical results are reported in Table A1 in the Appendix.

used in estimation.[16]

One might argue that the benchmark specification should also allow the portfolio weights to be free parameters. However, we know from JSZ that this is irrelevant under the standard loss function, since W_0 gives the optimal results for in-sample fit. This suggests that allowing the portfolio weights to be free parameters under the standard loss function yields the same parameter estimates as the JSZ model with fixed weights, and therefore also the same out-of-sample performance. We verified that this is indeed the case.[17] Henceforth, when we refer to the JSZ model under the standard loss function, it is always implemented with fixed weights, unless otherwise mentioned.

Panel C of Table 2 presents the ratio of the out-of-sample RMSEs. The RMSE ratios are defined as the ratio of the RMSE obtained using the forecasting loss function and the RMSE obtained using the standard loss function. An RMSE ratio less than one indicates that the forecasting loss function provides improvements in forecasting relative to the benchmark standard loss function. The statistical significance of the relative forecasting performance is evaluated using the Diebold and Mariano (1995) t-statistics computed with a serial correlation robust variance and the small sample adjustment of Harvey, Leybourne, and Newbold (1997).[18]

The improvements in forecasting RMSEs are substantial for three- to six-, nine-, and twelve-month forecast horizons. The improvement in the RMSEs is about 11% on average across maturities for the longer forecast horizons (nine- and twelve-month forecast horizons). In the forecasting literature, the out-of-sample R-square is often considered, which is defined as $1 - (MSE^{FL}/MSE^{SL})$, where SL refers to the benchmark model with standard loss function and FL to the alternative model with forecasting loss function. For longer forecast

[16]We report results for the entire sample. Year-by-year RMSEs indicate that the results are robust over time.

[17]Hamilton and Wu (2014) also find that the first three pricncipal components lead to a better fit than any other linear combination of yields.

[18]To ensure that the Diebold-Mariano test is valid for our setup, we examine the loss differential series based on the forecasts using the forecasting and standard loss functions. The augmented Dickey–Fuller tests reject the null of a unit root for all maturities at each forecast horizon.

horizons in Table 2, this gives on average $1 - (1 - 0.11)^2 = 0.21$. The improvement in forecasting RMSE therefore corresponds to an out-of-sample R-square of 21%. For short- and medium-term yields (three- and six-month yields, and one- and five-year yields), the forecasting loss function outperforms the standard loss function at all forecast horizons. For example, the improvement in the RMSEs for short maturity yields (three-month, six-month, and one-year yields) is about 14% on average across different forecast horizons, which corresponds to an out-of-sample R-square of 26%.

These results suggest that when using the JSZ canonical form, the time-series properties of the state variables are critically important to achieve better out-of-sample forecasting performance, which can be achieved using the forecasting loss function. It is imperative to free the portfolio weights to give the forecasting loss function more power to search for the best possible state variables for the purpose of out-of-sample forecasting. The forecasting loss function implicitly places a relatively higher weight on the time-series information of the state variables. This contrasts with in-sample estimation, where fixing the portfolio weights is optimal, as demonstrated by JSZ. Henceforth, when we discuss the JSZ specification with the forecasting loss function, it is always implemented with variable portfolio weights, unless otherwise mentioned.

2.4.6 In-Sample Fit

It is to be expected that the parameter estimates based on the forecasting loss function give rise to an in-sample fit that is worse than that for the standard loss function, because the latter loss function selects the parameters to provide the best possible in-sample fit. We document this trade-off between in-sample and out-of-sample fit. We illustrate these issues using the estimates for the JSZ canonical specification.

Table 3 reports the in-sample RMSEs for the JSZ model with forecasting loss function and standard loss function. We estimate these specifications using the entire sample and compute the in-sample RMSEs for different forecast horizons. Recall that the resulting estimates for the specification with forecasting loss function are forecast-horizon specific. We therefore report RMSEs for each forecast horizon.

The results in Panels A and B of Table 3 indicate a clear trade-off between in-sample and out-of-sample fit. The JSZ model with standard loss function (in Panel B) provides a better in-sample fit than the model with forecasting loss function (in Panel A). The JSZ model with forecasting loss function provides a higher RMSE, especially for longer forecast horizon. This result is of course not surprising, since the parameters for the JSZ model with forecasting loss function are chosen to optimally fit yields k periods ahead. These results therefore simply reflect a trade-off between in-sample and out-of-sample fitting. When using the forecasting loss function, results strongly differ from the JSZ model with standard loss function both in-and out-of-sample. Presumably these differences are due to differences in estimated parameters and implied state variables. We now investigate these differences in more detail.

2.5 Model Properties

We find that out-of-sample forecasting can be substantially improved by aligning the loss functions for in-sample and out-of-sample evaluation, as suggested by Granger (1993) and Weiss (1996). Presumably this finding results from differences in parameter estimates and implied state variables. In this section we document and discuss these differences. We then examine differences in the model-implied estimates of expected future short-term interest rates and term premia.

2.5.1 Loss Functions and State Variables

We examine the time-series properties of the state variables under the standard and forecasting loss functions. Figure 1 is based on the JSZ model with standard loss function. Panel A shows the time series of the first three principal components (PC): level, slope and curvature. Panel B presents the factor loadings $B(\Theta^Q)$ on the yield curve. Panel C shows the portfolio weights W_0 that ensure $W_0 y_t = PC_t$. For the JSZ with standard loss function, we obtain the customary level, slope and curvature factors.

Figures 2-4 are based on the JSZ model with forecasting loss function. To emphasize the differences resulting from the use of different loss functions, we present the resulting differences between the state variables, factor loadings, and portfolio weights, rather than the levels. Because the estimation is forecast-horizon specific, each figure has six panels, one for each forecast horizon k.

Figure 2 shows the differences in the time series of the state variables, $Wy_t - PC_t$, where W is estimated using the forecasting loss function. Note that the magnitude of the third factor is on average smaller than that of the curvature factor in the JSZ with standard loss function, regardless of the forecast horizon. The magnitudes of the first two factors on average are larger than the level and slope factors in the JSZ with standard loss function for longer forecast horizons.

Figure 3 plots the differences between the estimated factor loadings $B(\Theta^Q)$ from the JSZ model with forecasting loss function and standard loss function. For the first factor, the loadings are exactly the same for all forecast horizon estimations. For the second factor, the estimated factor loadings are very similar, except for medium- and long-maturity yields. The most pronounced differences are observed for the third factor. For all forecast horizons, the estimated loadings for the JSZ with forecasting loss function differ greatly from those with standard loss function at the short end of the yield curve. The estimated loadings are smaller for six-month, one-year, and two-year yields, but larger for very short-maturity yield (three-month yield). For intermediate- and long-maturity yields, the estimated factor loadings differ slightly from those obtained from standard loss function. The estimated loadings are larger for medium-term yields, but smaller for the twenty-year yield.

Figure 4 shows the differences in portfolio weights, $W - W_0$. These differences are similar across forecast horizons. The JSZ with forecasting loss function implies a different linear combination of yields, especially for the third factor. The resulting time series of the third state variable differs from the traditional curvature factor.

Table 4 presents the correlations between the state variables from the JSZ model with forecasting loss function and the first five principal components of the yield curve. Because the estimation is forecast-horizon specific, we report the results for each forecast horizon k. The third factor in the JSZ with forecasting loss function is highly correlated with the fourth principal component of the yield curve, whereas standard loss functions result in a third factor that is highly correlated with the third principal component. This result is related to the findings of Cochrane and Piazzesi (2005), who find that the fourth principal component accounts for a large part of the predictability in bond returns, even though it explains only a small part of in-sample variability. The third factor in the JSZ with forecasting loss function captures information that is hidden from the current yield curve, and this results in gains in out-of-sample forecasting performance.

2.5.2 Loss Functions and Parameter Estimates

We now compare the parameter estimates from the JSZ model with different loss functions. Table 5 presents the estimates of the parameters governing the state variables under the P- and Q-measures (K_0^P, K_1^P, K_0^Q, K_1^Q, $\Sigma\Sigma'$) for both specifications. Panel A of Table 5 reports the estimates for the JSZ model with forecasting loss function, which are different for each forecast horizon k. In the JSZ model with standard loss function, K_0^P and K_1^P are the OLS estimates, as shown in Panel B of Table 5.

The most interesting observations are related to the dynamic properties of the model. Regardless of the model and the forecast horizon, under both measures the first factor is the most persistent and the third factor is the least persistent. To assess the persistence properties of the model, we need to inspect the eigenvalues of K_1. The dominant eigenvalue in Panel B under the Q-measure is equal to one, whereas under the P-measure it is slightly smaller than one. In Panel A it is slightly smaller than one under both the P- and

Q-measures. A plausible explanation for this finding is that when computing yields, the dominant eigenvalue can be larger than one under Q if it is not too high. For in-sample fitting, a dominant eigenvalue of one may help fit yields better, but this may not necessarily be the case when the yield forecast is explicitly considered in the loss function.

Another substantial difference between Panels A and B is with respect to the third factor. The persistence of the third factor in Panel A is much smaller than that in Panel B under both the P- and Q-measures. Additionally, the third factor is less persistent under the Q-measure relative to the P-measure in Panel A. The $(1, 3)$ entry of the feedback matrix, which governs how the third factor this period forecasts the first factor next period differs greatly between the two panels as well. The relative impact of the third factor on the first factor is higher in the model with forecasting loss function. A similar finding obtains for the $(2, 3)$ entry of the feedback matrix.[19] The variance of the third factor is smaller in Panel A than that in Panel B. The covariances of the third factor with the other two factors also differ between the two panels. These results are consistent with the results in Figure 2: the third factor behaves differently for different loss functions.

We conclude that the improvement in forecasting performance is driven by the time-series properties and the parameter estimates of the state variables. It is important to use variable portfolio weights, which allow the forecasting loss function to search for the best possible state variables from the out-of-sample forecasting perspective.

2.5.3 Decomposing Forward Rates

The differences in parameter estimates and implied state variables also generate different implications for term premia and policy expectations. We compare the estimates of expected future short-term interest rates and term premia obtained using the JSZ model with different loss functions. We decompose the forward rate $F_t^{n_1,n_2}$ that one can lock in at time t for a

[19]Joslin and Le (2013) discuss estimation of the feedback matrix in ATSMs with stochastic volatility. They show that the implicit restriction on the relation betwee K_1^P and K_1^Q causes the estimates of K_1^P to differ from the OLS estimates.

$(n_2 - n_1)$-period loan starting in n_1 periods into the expectations of future short rates and the forward term premium $FTP_t^{n_1,n_2}$:

$$F_t^{n_1,n_2} = FTP_t^{n_1,n_2} + \underbrace{E_t \sum_{i=n_1}^{n_2-1} y_{t+i}^1}_{\substack{\text{Expectation} \\ \text{(Risk-Neutral Yield)}}} \quad . \tag{2.10}$$

We illustrate our results using five- to ten-year forward rates ($n_2 = 10$ years and $n_1 = 5$ years). Appendix C provides additional details on the computation of the model-implied term premia.

The top panels of Figure 5 compare the time series of model-implied term premia from the JSZ model with different loss functions. Because the estimation with the forecasting loss function is forecast-horizon specific, we plot the results for different forecast horizons. The left panels are based on the one-month forecast horizon, and the right panels are based on the twelve-month forecast horizon. The estimation with forecasting loss function implies larger term premia than the estimation with standard loss function at both short- and long-forecast horizons. The estimated term premia from both loss functions are countercyclical, consistent with existing studies (Campbell and Cochrane, 1999; Wachter, 2006; Harvey, 1989; Cochrane and Piazzesi, 2005). The bottom panels plot the differences between the term premia from the two loss functions. They are relatively large around the time of the 1970s oil shocks. In general, the differences seem to increase in recession periods. The middle panels display the sum of the expectations of future short rates in equation (3.16), i.e., the risk-neutral yields. The estimated risk-neutral yield from the forecasting loss function is smaller than that of the standard loss function at short- and long-forecast horizons.

2.6 Robustness

In this section, we examine the robustness of our results using the three-factor Gaussian AFNS model of CDR (2011), as well as stochastic volatility models with latent factors. We also show robustness results using a mean absolute error loss function. We follow the same procedure as in Section 2.4.3 to examine the out-of-sample forecasts.[20]

2.6.1 Results for the Gaussian AFNS Model

The JSZ canonical form is the maximally flexible specification of the $A_0(3)$ model. Another strand of the ATSM literature has considered imposing constraints on the factor structure to improve identification. One example of this approach is the AFNS model of CDR (2011). By imposing the Nelson-Siegel structure on the $A_0(3)$ model, the AFNS model greatly facilitates the estimation, and improves the predictive performance relative to the canonical $A_0(3)$ model.

For a three-factor Gaussian AFNS model, the state variables are governed by the following Q-dynamics

$$\begin{bmatrix} X_{1t+1} \\ X_{2t+1} \\ X_{3t+1} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \lambda & 1-\lambda \\ 0 & 0 & \lambda \end{bmatrix} \begin{bmatrix} X_{1t} \\ X_{2t} \\ X_{3t} \end{bmatrix} + \Sigma\varepsilon_{t+1}^{Q}, \tag{2.11}$$

$$r_t = X_{1t} + X_{2t}, \tag{2.12}$$

CDR (2011) show that for this particular Q-dynamics, the model-implied continuously compounded n-maturity yield matches the Nelson-Siegel yield function as follows

$$\widehat{y}_t^n = X_{1t} + (\frac{1-e^{-\lambda n}}{\lambda n})X_{2t} + (\frac{1-e^{-\lambda n}}{\lambda n} - e^{-\lambda n})X_{3t} + A_n(\Theta^Q), \tag{2.13}$$

[20] Because of space constraints, this section focuses on the out-of-sample forecasting performance of these models. We report the in-sample performance of these models in Tables A2-A4 in the Appendix. We find a trade-off between the in-sample and out-of-sample fit, similar to the results for the Gaussian model.

where $\Theta^Q = \{\lambda, \Sigma\}$. Note that the factor loadings exactly match the Nelson-Siegel factor structure. The factors are therefore identified as level, slope and curvature. The P-dynamics of the state variables are the same as in equation (2.4). For estimation using the forecasting loss function in equation (2.3), we need the model's prediction of the k-period ahead n-maturity yield, based on parameter estimates at time t. This is given by equation (2.8) with

$$B_n(\Theta^Q) = \left[\begin{array}{ccc} 1, & \frac{1-e^{-\lambda n}}{\lambda n}, & \frac{1-e^{-\lambda n}}{\lambda n} - e^{-\lambda n} \end{array}\right].$$

Since the state variables in the model are latent, we estimate the parameters $\Theta = \{K_0^P, K_1^P, \lambda, \Sigma\}$ and filter the state variables X_t by minimizing the forecasting loss function (2.3). We compare the out-of-sample forecasting performance of the Gaussian AFNS model estimated using forecasting loss function (2.3) relative to the model estimated using standard loss function (2.1).

Table 6 reports the out-of-sample RMSEs. Panel A presents the RMSEs for forecasting loss function (2.3), and Panel B presents the RMSEs for standard loss function (2.1). Consistent with the results for other specifications of Gaussian model, the RMSE ratios in Panel C indicate that the improvements in forecast performance are greatest for medium and long forecast horizons (five-, six-, nine-, and twelve-month forecast horizons). The improvement in forecasting RMSEs from using forecasting loss function (2.3) is on average across maturities approximately 14% at the twelve-month forecast horizon. This corresponds to an out-of-sample R-square of 27%. The improvements are also more pronounced for short- and medium-term yields. For example, for the five-year yield, the improvement in RMSEs is approximately 19% on average across forecast horizons, which corresponds to an out-of-sample R-square of 34%.

The Gaussian AFNS model is an invariant transformation of a special case of the JSZ canonical form with additional restrictions on the Q-parameters $\lambda^Q = [1, \lambda, \lambda]$ and $r_\infty^Q = 0$.[21]

[21]Since the AFNS model is an invariant transformation of a special case of the JSZ canonical form, we

JSZ (2011) show that these additional constraints on the Q-parameters of the JSZ canonical form do not significantly affect forecasts of bond yields.[22] Comparing Panel B of Table 6 with Panel B of Table 2 confirms that the forecasting performance of the maximally flexible AFNS model is similar to that of the JSZ canonical form under the standard loss function.

This comparison under the standard loss function may not necessarily extend to the forecasting loss function. Indeed, a comparison of Panel A of Table 6 with Panel A of Table 2 suggests that the Q-restrictions in the AFNS model improve the forecasting performance for medium-term yields (one- to five-year yields) under the forecasting loss function, although the improvements are modest. This different result obtains because under the forecasting loss function, estimation depends on Q- and P-parameters simultaneously. Therefore, restrictions on Q-parameters when using the forecasting loss function may affect the estimation of P-parameters, which in turn affect the forecasting performance.

We also consider a parsimonious version of the AFNS model, the AFNS model with independent factors. Under this specification, the state variables follow an independent first-order autoregression, i.e. K_1^P and Σ are diagonal matrices in equation (2.4). CDR (2011) show that this parsimonious version of the AFNS model exhibits significantly better out-of-sample forecast performance than the maximally flexible specification of the AFNS model, i.e. the AFNS model with correlated factors. We report the out-of-sample RMSEs of the independent factor AFNS model using forecasting and standard loss functions in Table A5 in the Appendix. The results for the independent factor AFNS model are consistent with the results for the AFNS model with correlated factors: using the forecasting loss function at estimation stage improves the forecasting performance for the Gaussian AFNS model.

In summary, we conclude that the results based on the Gaussian AFNS model are consistent with the results obtained using the JSZ canonical form. Aligning the loss functions

can also estimate a restricted JSZ canonical form with observed state variables directly. Using the standard loss function, we verified that the resulting estimates are similar to those obtained using the Kalman filter. When using the forecasting loss function with variable portfolio weights, this approach is not available.

[22]More generally, JSZ (2011) show that constraints on Q-parameters within an identified dynamic term structure model with latent factors cannot affect the forecasts of bond yields relative to an unconstrained VAR process.

for in-sample estimation and out-of-sample evaluation improves the forecasting performance of various implementations of Gaussian models.

2.6.2 Results for Stochastic Volatility Models

We compare the forecasting performance based on the standard loss function (2.1) and the forecasting loss function (2.3) for three-factor affine term structure models with stochastic volatility. We follow the Dai and Singleton (2000) normalization for the stochastic volatility models, $A_j(3)$, with $j = 1, 2$ or 3 factors driving the conditional variance of the state variables. The latent stochastic volatility model can be expressed using a state-space representation. The P- and Q-parameters are estimated simultaneously by applying the Kalman filter to the state-space representation. Appendix D provides details on the specification and estimation of the $A_j(3)$ model. For additional details, we refer to Dai and Singleton (2000).

Table 7 reports the ratio of the out-of-sample forecast RMSEs from the forecasting loss function (2.3) and the standard loss function (2.1) for the $A_1(3)$, $A_2(3)$, and $A_3(3)$ stochastic volatility models. The results based on the stochastic volatility models are consistent with the results for the Gaussian model. Aligning loss functions for in-sample estimation and out-of-sample forecasting evaluation provides improvements in out-of-sample forecasting performance. The improvements in RMSEs are greatest for longer forecast horizons (nine- and twelve-month forecast horizons). For example, for the $A_1(3)$, at the twelve-month forecast horizon, the improvement in forecasting RMSEs from using forecasting loss function (2.3) is on average across maturities approximately 19%, which corresponds to an out-of-sample R-square of 35%. For five-year yields, the forecasting loss function outperforms the standard loss function at all forecast horizons. The improvement in the RMSEs is approximately 16% on average. This corresponds to an out-of-sample R-square of 29%.

2.6.3 Results for Alternative Loss Functions

The standard and forecasting loss functions discussed so far are quadratic. We provide robustness analysis using absolute error loss. This is motivated by Diebold and Shin (2015), who demonstrate that the absolute error loss is essentially equivalent to the loss evaluation based on the entire distribution of the forecasting errors. Following the notation in Section 2.2, we express the standard absolute error loss function as follows

$$MAE(\Theta) = \frac{1}{NT}\sum_{n=1}^{N}\sum_{t=1}^{T}\left|\widehat{y}^{n}_{t|t}(\Theta) - y^{n}_{t}\right|. \tag{2.14}$$

The forecasting absolute error loss function is given by

$$OS\_MAE_{k}(\Theta) = \frac{1}{N(T-k)}\sum_{n=1}^{N}\sum_{t=k+1}^{T}\left|\widehat{y}^{n}_{t|t-k}(\Theta) - y^{n}_{t}\right|. \tag{2.15}$$

The out-of-sample forecasting performance for the n-maturity yield with forecast horizon k is evaluated by the mean absolute error (MAE)

$$MAE\_OS_{n,k} = \frac{1}{T-k}\sum_{t=1}^{T-k}\left|\widehat{y}^{n}_{t+k|t}(\Theta) - y^{n}_{t+k}\right|. \tag{2.16}$$

We conduct the robustness exercise using the JSZ canonical form. Table 8 reports the forecasting MAEs. Panels A and B of Table 8 present the MAEs using equations (2.15) and (2.14). Panel C presents the MAE ratios. The improvements in forecasting MAEs are substantial for three- to six-, nine-, and twelve-month forecast horizons. For example, at the six-month forecast horizon, the improvement in MAEs is about 13% on average across maturities, which corresponds to an out-of-sample R-square of 24%. The improvements are greatest for short- and medium-term yields (three- and six-month yields, and one- and five-year yields). For example, for the six-month yield, the forecasting loss function outperforms the standard loss function at all forecast horizons. The improvement in the MAEs is about 20% on average, which corresponds to an out-of-sample R-square of 37%.

The results are similar in magnitude to the results in Table 2, which are based on the RMSEs. Overall, these results confirm that our finding that aligning the loss functions for in-sample estimation and out-of-sample evaluation improves out-of-sample forecasting is not limited to one particular loss function.

2.7 Conclusion

Information not captured by traditional term structure factors helps predict excess bond returns. We propose estimating the parameters of no-arbitrage affine term structure models by aligning the objective functions for in-sample estimation and out-of-sample evaluation, instead of the traditional optimization of the likelihood criterion or the mean-squared error. Our approach amounts to letting the data determine the state variables that are best suited for out-of-sample forecasting. This results in term structure factors that capture similar information that remains hidden from existing approaches. Specifically, the estimates of the third term structure factor radically differ. Consistent with Cochrane and Piazzesi (2005), this factor confirms the importance of the fourth principal component of yields for forecasting the term structure. The new objective function leads to substantial improvements in forecasting performance, especially for long forecast horizons. Model term premiums are higher and expected future short rates are lower. We document a trade-off between in-sample and out-of-sample fit.

Our results may be extended in several ways. Most importantly, the question arises if our results generalize to other affine and non-affine term structure models. It may be challenging to address this issue because of the presence of identification problems. Using currently available estimation techniques, addressing these identification problems is harder than in the case of the affine model. Developing improved estimation methods for these alternative models is therefore also critically important.

Appendices

A The JSZ Canonical Form

Given the dynamics in equations (2.4)-(3.1), the model-implied continuously compounded yields $\widehat{y}_t$ are given by

$$\widehat{y}_t = A(\Theta^Q) + B(\Theta^Q)PO_t.$$

Now consider M linear combinations of N yields, $PO_t = Wy_t$, that are priced without error. We focus on a simple case where the eigenvalues of K_1^Q are real, distinct, and nonzero. This follows Joslin, Singleton and Zhu (2011), who demonstrate the result for all cases including zero, repeated and complex eigenvalues.

There exists a matrix C such that $K_1^Q = Cdiag(\lambda^Q)C^{-1} + I_M$. Define $D = Cdiag(\rho_1)C^{-1}$, $D^{-1} = Cdiag(\rho_1)^{-1}C^{-1}$ and

$$\begin{aligned} L_t &= D(PO_t + (K_1^Q - I_M)^{-1}K_0^Q), \\ &\Rightarrow PO_t = D^{-1}L_t - (K_1^Q - I_M)^{-1}K_0^Q. \end{aligned}$$

Then we have the dynamic of L_t under the Q-measure

$$\begin{aligned} \Delta L_{t+1} &= D\Delta PO_{t+1} \\ &= D[K_0^Q + (K_1^Q - I_M)(D^{-1}L_t - (K_1^Q - I_M)^{-1}K_0^Q) + \Sigma\varepsilon_{t+1}^P] \\ &= diag(\lambda^Q)L_t + D\Sigma\varepsilon_{t+1}^P, \end{aligned} \tag{A.1}$$

and the dynamic of L_t under the P-measure

$$\begin{aligned} \Delta L_{t+1} &= D\Delta PO_{t+1} \\ &= D[K_0^P + (K_1^P - I_M)(D^{-1}L_t - (K_1^Q - I_M)^{-1}K_0^Q) + \Sigma\varepsilon_{t+1}^P] \\ &= DK_0^P + D(K_1^P - I_M)D^{-1}L_t - D(K_1^P - I_M)(K_1^Q - I_M)^{-1}K_0^Q + D\Sigma\varepsilon_{t+1}^P, \end{aligned} \tag{A.2}$$

The dynamic of the short rate is

$$
\begin{aligned}
r_t &= \rho_0 + \rho_1 PO_t \qquad \text{(A.3)} \\
&= \rho_0 + \rho_1(D^{-1}L_t - (K_1^Q - I_M)^{-1}K_0^Q) \\
&= \rho_0 - \rho_1(K_1^Q - I_M)^{-1}K_0^Q + \rho_1 D^{-1}L_t \\
&= r_\infty^Q + \tau L_t,
\end{aligned}
$$

where $r_\infty^Q = \rho_0 - \rho_1(K_1^Q - I_M)^{-1}K_0^Q$, and τ is a row of M ones. Given the dynamics in equations (A.1)-(A.3), the model-implied continuously compounded yield $\widehat{y}_t$ is given by

$$\widehat{y}_t = A(\Theta_L^Q) + B(\Theta_L^Q)L_t,$$

where $\Theta_L^Q = \{r_\infty^Q, \lambda^Q, \Sigma\}$. The M linear combinations of N yields are perfectly priced and can be written as

$$
\begin{aligned}
PO_t &= Wy_t \\
&= W(A(\Theta_L^Q) + B(\Theta_L^Q)L_t)
\end{aligned}
$$

If the model is non-redundant, $WB(\Theta_L^Q)$ is invertible, and we have

$$L_t = (WB(\Theta_L^Q))^{-1}PO_t - (WB(\Theta_L^Q))^{-1}WA(\Theta_L^Q).$$

Then we can rewrite the dynamic of PO_t under the Q-measure as follows

$$
\begin{aligned}
\Delta PO_{t+1} &= WB(\Theta_L^Q)\Delta L_{t+1} \qquad \text{(A.4)} \\
&= WB(\Theta_L^Q)(diag(\lambda^Q)L_t + D\Sigma\varepsilon_{t+1}^P) \\
&= WB(\Theta_L^Q)\{diag(\lambda^Q)[(WB(\Theta_L^Q))^{-1}PO_t - (WB(\Theta_L^Q))^{-1}WA(\Theta_L^Q)] + D\Sigma\varepsilon_{t+1}^P\}.
\end{aligned}
$$

Comparing the coefficients in equations (A.4) and (2.5), we have

$$
\begin{aligned}
K_1^Q &= WB(\Theta_L^Q)diag(\lambda^Q)(WB(\Theta_L^Q))^{-1} + I_M, \\
K_0^Q &= -WB(\Theta_L^Q)diag(\lambda^Q)(WB(\Theta_L^Q))^{-1}WA(\Theta_L^Q).
\end{aligned}
\tag{A.5}
$$

We can also rewrite the dynamic of the short rate as follows

$$
\begin{aligned}
r_t &= r_\infty^Q + \tau L_t \\
&= r_\infty^Q + \tau[(WB(\Theta_L^Q))^{-1}PO_t - (WB(\Theta_L^Q))^{-1}WA(\Theta_L^Q)] \\
&= r_\infty^Q - \tau(WB(\Theta_L^Q))^{-1}WA(\Theta_L^Q) + \tau(WB(\Theta_L^Q))^{-1}PO_t.
\end{aligned}
\tag{A.6}
$$

Comparing the coefficients in equations (A.6) and (3.1), we have

$$
\begin{aligned}
\rho_0 &= r_\infty^Q - \tau(WB(\Theta_L^Q))^{-1}WA(\Theta_L^Q), \\
\rho_1 &= \tau(WB(\Theta_L^Q))^{-1}.
\end{aligned}
\tag{A.7}
$$

B Identification of Portfolio Weights

We show that for a given set of parameters Θ and forecast horizon k, the portfolio weights can be inverted from the observed term structure of yields assuming that the forecast errors are all zero. The model's prediction of the k-period ahead yield curve, based on parameter estimates at time t, is given by

$$\begin{aligned}\widehat{y}_{t+k|t}(\Theta) &= A(\Theta^Q) + B(\Theta^Q)\widehat{PO}_{t+k|t} \\ &= A(\Theta^Q) + B(\Theta^Q) f(y_t, k; K_0^P, K_1^P, W),\end{aligned} \tag{B.1}$$

where f is given by

$$f(y_t, k; K_0^P, K_1^P, W) = (I_3 + K_1^P + ... + (K_1^P)^{k-1})K_0^P + (K_1^P)^k W y_t.$$

Given term structure data for months $t = 1, ..., T$ on maturities $n = 1, ..., N$, the prediction of the yield curve $\widehat{y}$ is given by

$$\widehat{y} = A^*(\Theta^Q) + B(\Theta^Q)((I_3 + K_1^P + ... + (K_1^P)^{k-1})K_0^P + (K_1^P)^k W L^k(y)), \tag{B.2}$$

where $\widehat{y}$ collects the predicted yield curve between time k and T, which is a $N \times (T-k)$ matrix. $A(\Theta^Q)$ is a 9×1 vector, which can be replicated $T-k$ times to create a $N \times (T-k)$ matrix $A^*(\Theta^Q)$. $B(\Theta^Q)$ is a $N \times 3$ matrix since we have three state variables. K_0^P is a 3×1 vector, K_1^P is a 3×3 matrix, W is a $3 \times N$ matrix. L^k is the k-period lag operator, and $L^k(y)$ is also a $N \times (T-k)$ matrix.

Assuming zero forecast errors, $\widehat{y} = y$, we can rewrite equation (B.2) as follows

$$y - A^*(\Theta^Q) - M = B(\Theta^Q)(K_1^P)^k W L^k(y).$$

Note that $B(\Theta^Q)((I_3 + K_1^P + ... + (K_1^P)^{k-1})K_0^P)$ in equation (B.2) is replicated $T-k$ times to create a $N \times (T-k)$ matrix M. Defining $\widetilde{y} = y - A^{*}(\Theta^{Q}) - M$, the portfolio weights are computed as follows

$$W = (B^{'}(\Theta^{Q})B(\Theta^{Q})(K_1^{P})^{k})^{-1}B^{'}(\Theta^{Q})\widetilde{y}(L^{k}(y))^{'}(L^{k}(y)(L^{k}(y))^{'-1}. \tag{B.3}$$

This linear mapping is not exact, since not all yields can be forecasted perfectly. However, the inverted weights can be used as the initial value for the first step estimation of the JSZ canonical form with forecasting loss function.

C Model-Implied Term Premia

As discussed in Section 3.7.1, we can write the forward term premium $FTP_t^{n_1,n_2}$ as follows

$$FTP_t^{n_1,n_2} = F_t^{n_1,n_2} - \underbrace{E_t \sum_{i=n_1}^{n_2-1} y_{t+i}^1}_{\substack{\text{Expectation} \\ \text{(Risk-Neutral Yield)}}} . \tag{C.1}$$

The forward rate $F_t^{n_1,n_2}$ that one can lock in at time t for a $(n_2 - n_1)$-period loan starting in n_1 periods is computed as

$$F_t^{n_1,n_2} = \log[p_t^{n_1}/p_t^{n_2}], \tag{C.2}$$

where $p_t^{n_1}$ denotes time t price of a n_1-period zero coupon bond, and $\log(p_t^{n_1}) = -n_1 y_t^{n_1}$.

Given the model-implied yield for a one-period bond

$$y_t^1 = A_1 + B_1 X_t, \tag{C.3}$$

the expectation term in equation (C.1) can be rewritten as follows

$$\begin{aligned} E_t \sum_{i=n_1}^{n_2-1} y_{t+i}^1 &= \sum_{i=n_1}^{n_2-1} E_t[y_{t+i}^1] \\ &= \sum_{i=n_1}^{n_2-1} E_t[A_1 + B_1 X_{t+i}] \\ &= \sum_{i=n_1}^{n_2-1} A_1 + B_1 \sum_{i=n_1}^{n_2-1} E_t[X_{t+i}], \end{aligned} \tag{C.4}$$

where $E_t[X_{t+i}]$ is a function of the parameters K_0^P and K_1^P

$$E_t[X_{t+i}] = (I_3 + K_1^P + ... + (K_1^P)^{i-1})K_0^P + (K_1^P)^i X_t. \tag{C.5}$$

D Latent Stochastic Volatility Models

Using the classification of Dai and Singleton (2000), we focus on $A_j(3)$ models with $j = 1, 2$ or 3 factors driving the conditional variance of the state variables, which are given by

$$dX_t = (K^P_{0\Delta} + K^P_{1\Delta} X_t)dt + \Sigma\sqrt{S_t} dW^P_{t+1}, \tag{D.1}$$

$$dX_t = (K^Q_{0\Delta} + K^Q_{1\Delta} X_t)dt + \Sigma\sqrt{S_t} dW^Q_{t+1}, \tag{D.2}$$

$$r_t = \rho_0 + \rho_1 X_t, \tag{D.3}$$

where W^P_{t+1} and W^Q_{t+1} are three-dimensional independent standard Brownian motions under physical measure P and risk-neutral measure Q respectively, r_t is the instantaneous spot interest rate, and $\Sigma S_t \Sigma'$ is the conditional covariance matrix of X_t. S_t is a 3×3 diagonal matrix with the ith diagonal element given by

$$[S_t]_{ii} = \alpha_i + \beta_i^{'} X_t, \tag{D.4}$$

where α_i is a scalar, and β_i is a 3×1 vector. $\alpha = [\alpha_1, \alpha_2, \alpha_3]^{'}$is a 3×1 vector. $\beta = [\beta_1, \beta_2, \beta_3]$ is a 3×3 matrix. We follow the Dai and Singleton identification scheme to ensure the $[S_t]_{ii}$ are strictly positive for all i.

The model-implied continuously compounded yields $\widehat{y}_t$ are given by (see Duffie and Kan, 1996)

$$\widehat{y}_t = A(\Theta^Q) + B(\Theta^Q) X_t, \tag{D.5}$$

where the $N \times 1$ vector $A(\Theta^Q)$, and the $N \times 3$ matrix $B(\Theta^Q)$ are functions of the parameters under the Q-dynamics, $\Theta^Q = \{K^Q_{0\Delta}, K^Q_{1\Delta}, \rho_0, \rho_1, \Sigma, \alpha, \beta\}$, through a set of Ricatti ordinary differential equations. Recall that N denotes the number of available yields in the term structure. We adopt the essentially affine specification for the price of risk, as in Duffee (2002).

The affine dynamic for X_t in equation (D.1) implies that the one-period ahead conditional expectation of X_t under the P measure, $\widehat{X}_{t+\Delta|t} = \text{constant} + \Delta e^{K^P_{1\Delta}} X_t$, where $\Delta = 1/12$. Thus X_t follows a first order VAR when sampled monthly. Similarly, the affine dynamic in equation (D.2) under the Q measure implies a first order VAR for X_t sampled at the monthly frequency. For estimation based on the forecasting loss function in equation (2.3), we need the model's prediction of the k-period ahead n-maturity yield, based on parameter estimates at time t. This is given by

$$\begin{aligned} \widehat{y}^n_{t+k|t}(\Theta) &= A_n(\Theta^Q) + B_n(\Theta^Q)\widehat{X}_{t+k|t} \\ &= A_n(\Theta^Q) + B_n(\Theta^Q) f(X_t, k; K^P_0, K^P_1), \end{aligned} \tag{D.6}$$

where $A_n(\Theta^Q)$ is the n^{th} element of $A(\Theta^Q)$, $B_n(\Theta^Q)$ is the n^{th} row of $B(\Theta^Q)$, $\Theta = \{\Theta^Q, K^P_0, K^P_1\}$, and f is given by

$$f(X_t, k; K^P_0, K^P_1) = (I_3 + K^P_1 + ... + (K^P_1)^{k-1})K^P_0 + (K^P_1)^k X_t.$$

where K^P_0 and K^P_1 are the parameters for the VAR(1) process of X_t under the P measure, which can be mapped to $K^P_{0\Delta}$ and $K^P_{1\Delta}$ respectively in equation (D.1) through the nonlinear relations $K^P_1 = e^{\Delta K^P_{1\Delta}}$ and $K^P_0 = K^P_{0\Delta} \int_0^\Delta e^{sK^P_{1\Delta}} ds$. In particular, for small Δ, $K^P_0 \approx \Delta K^P_{0\Delta}$ and $K^P_1 \approx I_3 + \Delta K^P_{1\Delta}$. We can view K^P_0 and $K^P_{0\Delta}$, and K^P_1 and $K^P_{1\Delta}$, as interchangeable. Similarly, K^Q_0 and $K^Q_{0\Delta}$, and K^Q_1 and $K^Q_{1\Delta}$ are interchangeable.

A three factor latent model can be expressed using a state-space representation. Using equation (D.1) and an Euler discretization, the state equation can be written as $X_{t+1} = K^P_0 + K^P_1 X_t + \varepsilon^P_{t+1}$, where $\varepsilon^P_{t+1|t}$ is assumed to be distributed $N(0,\, \Sigma S_t \Sigma')$. The observed yield curve $y_t = \widehat{y}_t + e_t$ is the measurement equation, where $\widehat{y}_t$ is the model-implied yield as specified in equation (D.5), and e_t is a vector of measurement errors that is assumed to be *i.i.d.* normal with diagonal covariance matrix R. The estimates of the P-parameters, $\Theta^P = \{K_0^P, K_1^P\}$ are related to the Q-parameters, since the pricing model is required to filter the latent factors X_t. We therefore need to estimate the P- and Q-parameters simultaneously. We do this by applying the Kalman filter to the state-space representation. We estimate the parameters $\Theta = \{K_0^P, K_1^P, K_0^Q, K_1^Q, \rho_0, \rho_1, \Sigma, \alpha, \beta\}$ and filter the state variables X_t by minimizing the forecasting loss function, equation (2.3). We compare the results obtained from the forecasting loss function with the estimation of the latent stochastic volatility models based on the standard loss function equation (2.1).

When estimating models with latent factors, the numerical implementation is important because of the existence of identification problems. We follow the implementation of Hamilton and Wu (2012). We extract the first three principal components from the observed term structure of yields, and normalize each principal component to have zero mean and unit variance. We estimate the dynamics of the normalized first three principal components through OLS, and use the OLS estimates of K_0^P and K_1^P as initial values.

We obtain the initial values for ρ_0 and ρ_1 by regressing one-month yields on the normalized principal components. To get the initial value for the Q parameters, we regress the observed term structure of yields on the normalized principal components to get estimated loadings $\widehat{A}$ and $\widehat{B}$. The Q parameters enter the loadings $A(\Theta^Q)$ and $B(\Theta^Q)$, which are defined recursively. Subsequently we obtain an initial guess for the Q parameters by minimizing the distances from $A(\Theta^Q)$ and $B(\Theta^Q)$ to the estimated loadings $\widehat{A}$ and $\widehat{B}$.

With this set of initial values, we find Θ by optimizing the standard log likelihood function using the fminsearch algorithm in MATLAB. We compute the 99% confidence interval, $[\Theta_-, \Theta^-]$, for the converged values of Θ. Then we generate another 100 different sets of Θ from the uniform distributions $U[\Theta_-, \Theta^-]$. We rank these different sets of Θ by the implied likelihood, and use the top 10 ranked sets of Θ as initial values for another round of numerical search. We choose the converged sets of Θ based on the likelihood, and form the new range of the parameter set using the chosen sets of Θ. We continue generating different sets of initial values until they converge to very similar values.

Bibliography

[1] Adrian, T., Crump, R. K., and Moench, E. (2013). Pricing the term structure with linear regressions. Journal of Financial Economics, 110(1), 110-138.

[2] Ang, A., and Piazzesi, M. (2003). A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables. Journal of Monetary Economics, 50(4), 745-787.

[3] Ang, A., Piazzesi, M., and Wei, M. (2006). What does the yield curve tell us about GDP growth?. Journal of Econometrics, 131(1), 359-403.

[4] Aït-Sahalia, Y., and Kimmel, R. L. (2010). Estimating affine multifactor term structure models using closed-form likelihood expansions. Journal of Financial Economics, 98(1), 113-144.

[5] Bauer, M. D., Rudebusch, G. D., and Wu, J. C. (2012). Correcting estimation bias in dynamic term structure models. Journal of Business and Economic Statistics, 30(3), 454-467.

[6] Bowsher, C. G., and Meeks, R. (2008). The dynamics of economic functions: modeling and forecasting the yield curve. Journal of the American Statistical Association, 103(484), 1419-1437.

[7] Campbell, J. Y. and Cochrane, J. H. (1999). By force of habit: a consumption-based explanation of aggregate stock market behavior. Journal of Political Economy, 107(2), 205–251.

[8] Chen, R. R., and Scott, L. (1992). Pricing interest rate options in a two-factor Cox-Ingersoll-Ross model of the term structure. Review of Financial Studies, 5(4), 613-636.

[9] Christensen, J. H., Diebold, F. X., and Rudebusch, G. D. (2011). The affine arbitrage-free class of Nelson–Siegel term structure models. Journal of Econometrics, 164(1), 4-20.

[10] Christoffersen, P. F., and Diebold, F. X. (1996). Further results on forecasting and model selection under asymmetric loss. Journal of Applied Econometrics, 11(5), 561-571.

[11] Christoffersen, P. F., and Diebold, F. X. (1997). Optimal prediction under asymmetric loss. Econometric Theory, 13(6), 808-817.

[12] Cieslak, A., and Povala, P. (2015). Expected returns in Treasury bonds. Review of Financial Studies.

[13] Cochrane, J. H. (2005). Asset pricing. Princeton University Press.

[14] Cochrane, J. H., and Piazzesi, M. (2005). Bond risk premia. American Economic Review, 95(1), 138-160.

[15] Collin-Dufresne, P., Goldstein, R. S., and Jones, C. S. (2008). Identification of maximal affine term structure models. The Journal of Finance, 63(2), 743-795.

[16] Cooper, I., and Priestley, R. (2009). Time-varying risk premiums and the output gap. Review of Financial Studies, 22(7), 2801-2833.

[17] Cox, D. R. (1961). Prediction by exponentially weighted moving averages and related methods. Journal of the Royal Statistical Society. Series B (Methodological), 414-422.

[18] Cox, J. C., Ingersoll Jr, J. E., and Ross, S. A. (1985). A theory of the term structure of interest rates. Econometrica, 53(2), 385-408.

[19] Creal, D. D., and Wu, J. C. (2015). Estimation of affine term structure models with spanned or unspanned stochastic volatility. Journal of Econometrics, 185(1), 60-81.

[20] Dai, Q., and Singleton, K. J. (2000). Specification analysis of affine term structure models. The Journal of Finance, 55(5), 1943-1978.

[21] Diebold, F. X. (2015). Comparing predictive accuracy, twenty years later: A personal perspective on the use and abuse of Diebold–Mariano tests. Journal of Business & Economic Statistics, 33(1), 1-1.

[22] Diebold, F. X., and Li, C. (2006). Forecasting the term structure of government bond yields. Journal of Econometrics, 130(2), 337-364.

[23] Diebold, F. X., and Mariano, R. (1995). Comparing predictive accuracy. Journal of Business and Economic Statistics 13(3), 253-63.

[24] Diebold, F. X., and Shin, M. (2015). Assessing point forecast accuracy by stochastic loss distance. Economics Letters, 130, 37-38.

[25] Diez de Los Rios, A. (2014). A new linear estimator for Gaussian dynamic term structure models. Journal of Business and Economic Statistics, 33(2), 282-295.

[26] Duffee, G. R. (2002). Term premia and interest rate forecasts in affine models. The Journal of Finance, 57(1), 405-443.

[27] Duffee, G. R. (2011a). Information in (and not in) the term structure. Review of Financial Studies, 24(9), 2895-2934.

[28] Duffee, G. R. (2011b). Forecasting with the term structure: The role of no-arbitrage restrictions. Working Paper. Johns Hopkins University.

[29] Duffee, G. R., and Stanton, R. H. (2012). Estimation of dynamic term structure models. The Quarterly Journal of Finance, 2(2).

[30] Duffie, D., and Kan, R. (1996). A yield-factor model of interest rates. Mathematical Finance, 6(4), 379-406.

[31] Egorov, A. V., Hong, Y., and Li, H. (2006). Validating forecasts of the joint probability density of bond yields: Can affine models beat random walk? Journal of Econometrics, 135(1), 255-284.

[32] Elliott, G., Komunjer, I., and Timmermann, A. (2005). Estimation and testing of forecast rationality under flexible loss. The Review of Economic Studies, 72(4), 1107-1125.

[33] Elliott, G., Komunjer, I., and Timmermann, A. (2008). Biases in macroeconomic forecasts: irrationality or asymmetric loss? Journal of the European Economic Association, 6(1), 122-157.

[34] Elliott, G., and Timmermann, A. (2008). Economic forecasting. Journal of Economic Literature, 46(1), 3-56.

[35] Engle, R. (1993). A comment on Hendry and Clements on the limitations of comparing mean square forecast errors. Journal of Forecasting 12(8), 642-644.

[36] Gourieroux, C., Monfort, A., and Polimenis, V. (2006). Affine models for credit risk analysis. Journal of Financial Econometrics, 4(3), 494-530.

[37] Granger, C. W. (1993). On the limitations of comparing mean square forecast errors: Comment. Journal of Forecasting, 12(8), 651-652.

[38] Gürkaynak, R. S., Sack, B., and Wright, J. H. (2007). The US Treasury yield curve: 1961 to the present. Journal of Monetary Economics, 54(8), 2291-2304.

[39] Hamilton, J. D., and Wu, J. C. (2012). Identification and estimation of Gaussian affine term structure models. Journal of Econometrics, 168(2), 315-331.

[40] Hamilton, J. D., and Wu, J. C. (2014). Testable implications of affine term structure models. Journal of Econometrics, 178, 231-242.

[41] Harvey, C. R. (1989). Time-varying conditional covariances in tests of asset pricing models. Journal of Financial Economics, 24(2), 289-317.

[42] Harvey, D., Leybourne, S., and Newbold, P. (1997). Testing the equality of prediction mean squared errors. International Journal of Forecasting, 13(2), 281-291.

[43] Hong, Y., Li, H., and Zhao, F. (2004). Out-of-sample performance of discrete-time spot interest rate models. Journal of Business & Economic Statistics, 22(4), 457-473.

[44] Joslin, S., and Le, A. (2013). Interest rate volatility and no-arbitrage affine term structure models. Working Paper. University of Southern California.

[45] Joslin, S., Le, A., and Singleton, K. J. (2013). Why Gaussian macro-finance term structure models are (nearly) unconstrained factor-VARs. Journal of Financial Economics, 109(3), 604-622.

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

[47] Joslin, S., Singleton, K. J., and Zhu, H. (2011). A new perspective on Gaussian dynamic term structure models. Review of Financial Studies, 24(3), 926-970.

[48] Kang, I. B. (2003). Multi-period forecasting using different models for different horizons: an application to US economic time series data. International Journal of Forecasting, 19(3), 387-400.

[49] Knez, P. J., Litterman, R., and Scheinkman, J. (1994). Explorations into factors explaining money market returns. The Journal of Finance, 49(5), 1861-1882.

[50] Kim, D. H., and Orphanides, A. (2012). Term structure estimation with survey data on interest rate forecasts. Journal of Financial and Quantitative Analysis, 47(01), 241-272.

[51] Le, A., Singleton, K. J., and Dai, Q. (2010). Discrete-time affine term structure models with generalized market prices of risk. Review of Financial Studies, 23(5), 2184-2227.

[52] Lin, J. L., and Granger, C. W. (1994). Forecasting from non-linear models in practice. Journal of Forecasting, 13(1), 1-9.

[53] Litterman, R. B., and Scheinkman, J. (1991). Common factors affecting bond returns. The Journal of Fixed Income, 1(1), 54-61.

[54] Liu, S.-I. (1996). Model selection for multiperiod forecasts. Biometrika, 83, 861-873.

[55] Longstaff, F. A., and Schwartz, E. S. (1992). Interest rate volatility and the term structure: A two-factor general equilibrium model. The Journal of Finance, 47(4), 1259-1282.

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

[57] Marcellino, M., Stock, J. H., and Watson, M. W. (2006). A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series. Journal of Econometrics, 135(1), 499-526.

[58] Nelson, C. R., and Siegel, A. F. (1987). Parsimonious modeling of yield curves. Journal of Business, 60, 473-489.

[59] Patton, A. J., and Timmermann, A. (2007a). Properties of optimal forecasts under asymmetric loss and nonlinearity. Journal of Econometrics, 140(2), 884-918.

[60] Patton, A. J., and Timmermann, A. (2007b). Testing forecast optimality under unknown loss. Journal of the American Statistical Association, 102(480), 1172-1184.

[61] Sarno, L., Schneider, P., and Wagner, C. (2014). The economic value of predicting bond risk premia: Can anything beat the expectations hypothesis? Working Paper. University of Lugano.

[62] Shin, M., and Zhong, M. (2013). Does realized volatility help bond yield density prediction? Working Paper. University of Pennsylvania.

[63] Vasicek, O. (1977). An equilibrium characterization of the term structure. Journal of Financial Economics, 5(2), 177-188.

[64] Wachter, J. A. (2006). A consumption-based model of the term structure of interest rates. Journal of Financial economics, 79(2), 365-399.

[65] Weiss, A. A. (1991). Multi-step estimation and forecasting in dynamic models. Journal of Econometrics, 48(1), 135-149.

[66] Weiss, A. A. (1996). Estimating time series models using the relevant cost function. Journal of Applied Econometrics, 11(5), 539-560.

Table 2.1 Summary Statistics

		Central Moments			Autocorrelation		
	Mean	St.Dev	Skewness	Kurtosis	Lag 1	Lag 12	Lag 30
3 month yield	0.0450	0.0290	0.8938	4.3247	0.9773	0.7944	0.5197
6 month yield	0.0479	0.0305	0.8717	4.2283	0.9850	0.8126	0.5359
1 year yield	0.0516	0.0306	0.6980	3.6594	0.9857	0.8317	0.5891
2 year yield	0.0536	0.0301	0.6734	3.4957	0.9878	0.8509	0.6485
3 year yield	0.0554	0.0294	0.6703	3.4460	0.9884	0.8611	0.6782
4 year yield	0.0569	0.0288	0.6903	3.4142	0.9882	0.8655	0.7000
5 year yield	0.0579	0.0282	0.7270	3.3717	0.9890	0.8739	0.7183
10 year yield	0.0617	0.0275	0.9148	3.5853	0.9890	0.8739	0.7183
20 year yield	0.0638	0.0265	0.9158	3.5373	0.9930	0.8936	0.7724

Notes to Table: We present summary statistics for the data used in estimation. We present the sample mean, standard deviation, skewness, kurtosis, and autocorrelations for each of the yields. The yields are continuously compounded monthly zero-coupon bond yields. The sample period is from 1953:04 to 2012:12.

Table 2.2 Out-of-Sample RMSEs: JSZ Canonical Form

Panel A: Forecasting Loss Function

Forecast Horizon k	1 month	2 month	3 month	4 month	5 month	6 month	9 month	12 month
3 month yield	37.02	46.51	56.77	67.12	76.45	87.09	114.96	146.76
6 month yield	27.31	43.64	57.05	70.74	81.84	94.58	125.89	160.31
1 year yield	36.37	53.92	67.82	80.75	91.26	103.03	131.51	163.32
2 year yield	48.85	63.84	76.13	87.09	96.26	106.25	130.05	158.13
3 year yield	45.40	59.87	71.53	81.66	90.23	99.15	120.06	145.65
4 year yield	38.01	53.53	65.15	74.73	82.72	90.95	109.95	133.65
5 year yield	31.29	48.20	60.44	70.36	78.60	86.54	104.62	126.41
10 year yield	35.39	48.74	58.54	66.76	74.17	80.70	96.60	114.17
20 year yield	50.51	57.97	64.86	70.28	76.01	80.68	92.52	105.21

Panel B: Standard Loss Function

Forecast Horizon k	1 month	2 month	3 month	4 month	5 month	6 month	9 month	12 month
3 month yield	38.72	55.31	69.90	80.40	91.57	103.00	136.41	162.37
6 month yield	33.68	52.77	70.13	84.76	98.56	111.71	146.05	174.08
1 year yield	37.51	59.02	78.09	94.16	108.31	121.03	153.24	179.91
2 year yield	39.83	63.66	82.94	98.76	112.23	124.11	152.92	177.25
3 year yield	37.95	60.60	78.53	93.15	105.56	116.32	142.01	164.41
4 year yield	34.56	55.80	72.41	85.65	96.97	106.86	130.12	150.98
5 year yield	32.00	52.34	68.19	80.92	91.77	101.07	122.68	142.36
10 year yield	32.62	49.44	61.88	72.43	81.54	89.17	107.30	124.79
20 year yield	26.92	41.27	52.09	61.43	69.49	76.19	92.18	107.75

Panel C: RMSE Ratio

Forecast Horizon k	1 month	2 month	3 month	4 month	5 month	6 month	9 month	12 month
3 month yield	0.96	0.84***	0.81***	0.83***	0.83***	0.85***	0.84***	0.90***
6 month yield	0.81***	0.83***	0.81***	0.83***	0.83***	0.85***	0.86***	0.92***
1 year yield	0.97	0.91***	0.87***	0.86***	0.84***	0.85***	0.86***	0.91***
2 year yield	1.23***	1.00	0.92***	0.88***	0.86***	0.86***	0.85***	0.89***
3 year yield	1.20***	0.99	0.91***	0.88***	0.85***	0.85***	0.85***	0.89***
4 year yield	1.10***	0.96**	0.90***	0.87***	0.85***	0.85***	0.84***	0.89***
5 year yield	0.98	0.92***	0.89***	0.87***	0.86***	0.86***	0.85***	0.89***
10 year yield	1.08**	0.99	0.95*	0.92***	0.91***	0.90***	0.90***	0.91***
20 year yield	1.88***	1.40***	1.25***	1.14***	1.09***	1.06*	1.00	0.98

Notes to Table: We present the out-of-sample RMSEs for the JSZ canonical form with forecasting loss function and standard loss function. Panel A reports on the JSZ canonical form with forecasting loss function and variable portfolio weights. At each time t and for each forecast horizon k, we estimate the model using data up to and including t, and forecast k periods ahead. Panel B reports on the JSZ canonical form with standard loss function and fixed portfolio weights. We estimate the model using data up to and including t at each month t, and forecast one to twelve months ahead. RMSEs are reported in basis points. Panel C shows the ratios of the RMSEs in Panel A and Panel B. The statistical significance of the relative forecasting performance is evaluated using the Diebold and Mariano (1995) test statistic computed with a serial correlation robust variance and the small sample adjustment of Harvey, Leybourne, and Newbold (1997). The significance is denoted by *, ** and ***, corresponding to significance levels of 10%, 5% and 1% respectively.

Table 2.3 In-Sample RMSEs: JSZ Canonical Form

				Panel A: Forecasting Loss Function				
Forecast Horizon k	1 month	2 month	3 month	4 month	5 month	6 month	9 month	12 month
3 month yield	31.12	32.81	33.48	37.30	34.51	38.77	40.33	54.96
6 month yield	18.64	18.57	17.75	18.29	18.34	18.83	18.00	38.60
1 year yield	17.39	17.59	18.57	20.46	21.41	24.39	23.13	41.08
2 year yield	23.14	23.26	23.96	24.39	24.20	26.03	25.48	31.28
3 year yield	24.24	24.14	24.64	24.88	24.59	25.77	24.77	26.97
4 year yield	21.34	21.20	21.69	21.87	21.55	22.65	21.39	22.68
5 year yield	16.95	16.62	17.10	16.87	17.18	17.66	16.98	17.89
10 year yield	19.28	19.79	19.61	19.72	19.27	20.35	20.66	21.75
20 year yield	34.89	35.31	35.20	35.62	35.41	36.46	38.09	41.61
				Panel B: Standard Loss Function				
3 month yield					15.18			
6 month yield					13.99			
1 year yield					15.04			
2 year yield					8.55			
3 year yield					6.87			
4 year yield					9.86			
5 year yield					10.68			
10 year yield					11.62			
20 year yield					11.54			

Notes to Table: We present the in-sample RMSEs for the JSZ canonical form with forecasting loss function and variable portfolio weights (Panel A), and the JSZ canonical form with standard loss function and fixed portfolio weights (Panel B). The estimates of the specification with forecasting loss function are forecast-horizon specific, so we report the in-sample RMSEs for each forecast horizon. RMSEs are reported in basis points.

Table 2.4 Correlations Between State Variables and Principal Components

Principal Components	PC1	PC2	PC3	PC4	PC5
			1 Month Horizon		
Factor 1	1.0000	-0.0024	0.0066	-0.0029	-0.0014
Factor 2	-0.0136	0.9999	0.0041	0.0015	0.0008
Factor 3	-0.0386	0.0653	0.0114	0.9971	-0.0001
			2 Month Horizon		
Factor 1	1.0000	-0.0042	0.0082	-0.0060	-0.0032
Factor 2	-0.0105	0.9999	0.0083	0.0008	-0.0001
Factor 3	-0.0668	0.0564	-0.0297	0.9959	-0.0059
			3 Month Horizon		
Factor 1	1.0000	-0.0087	0.0138	-0.0090	-0.0046
Factor 2	0.0043	0.9999	0.0138	0.0002	-0.0008
Factor 3	-0.1452	0.1145	-0.0495	0.9823	-0.0180
			4 Month Horizon		
Factor 1	1.0000	-0.0088	0.0154	-0.0135	-0.0078
Factor 2	0.0459	0.9987	0.0162	-0.0012	-0.0012
Factor 3	-0.2001	0.1854	-0.0728	0.9611	-0.0081
			5 Month Horizon		
Factor 1	0.9999	-0.0146	0.0182	-0.0164	-0.0093
Factor 2	0.0622	0.9977	0.0200	-0.0015	-0.0025
Factor 3	-0.1749	0.3172	-0.0499	0.9324	-0.0165
			6 Month Horizon		
Factor 1	0.9999	-0.0204	0.0237	-0.0184	-0.0109
Factor 2	0.0751	0.9967	0.0140	-0.0017	-0.0027
Factor 3	-0.2396	0.2713	-0.0696	0.9324	-0.0357
			9 Month Horizon		
Factor 1	0.9999	-0.0189	0.0272	-0.0289	-0.0151
Factor 2	0.0773	0.9964	0.0186	-0.0004	-0.0029
Factor 3	-0.1679	0.4725	-0.1468	0.8531	-0.0220
			12 Month Horizon		
Factor 1	0.9998	-0.0234	0.0379	-0.0401	-0.0213
Factor 2	0.2431	0.9677	0.0542	0.0022	-0.0011
Factor 3	-0.2270	0.4452	-0.1344	0.8596	-0.0147

Notes to Table: We present the correlations between the state variables from the JSZ canonical form with forecasting loss function and the first five principal components (PC) of the yield curve. The state variables are estimated using the entire sample. The estimates are forecast-horizon specific.

Table 2.5 Parameter Estimates: JSZ Canonical Form

Panel A: Forecasting Loss Function

	P-Dynamics					Q-Dynamics							
Forecast Horizon	K_0^P	K_1^P			Eigenvalues	K_0^Q	K_1^Q			Eigenvalues	$\Sigma\Sigma' \times 1e4$		
1 month													
	-0.0015	0.9941	0.0551	-1.0073	0.9949	-0.0040	1.0292	0.0723	-1.9989	0.9991	0.9618	0.2112	0.0065
	0.0018	0.0018	0.9328	0.5121	0.9417	0.0043	-0.0330	0.9735	2.1475	**0.9341**	0.2112	0.2351	0.0306
	-0.0008	0.0004	0.0094	0.4795	**0.4698**	-0.0017	0.0157	-0.0148	0.0019	**0.0714**	0.0065	0.0306	0.0421
2 month													
	-0.0015	0.9938	0.0551	-0.9093	0.9947	-0.0038	1.0281	0.0735	-1.9313	0.9991	0.9578	0.2122	0.0072
	0.0018	0.0018	0.9336	0.5120	0.9438	0.0041	-0.0319	0.9722	2.0793	**0.9339**	0.2122	0.2316	0.0298
	-0.0008	0.0004	0.0094	0.5334	**0.5223**	-0.0016	0.0152	-0.0142	0.0338	**0.1010**	0.0072	0.0298	0.0417
3 month													
	-0.0014	0.9941	0.0551	-0.9144	0.9950	-0.0040	1.0298	0.0717	-2.0402	0.9991	0.9505	0.2090	0.0034
	0.0018	0.0018	0.9327	0.5128	0.9429	0.0043	-0.0336	0.9741	2.1891	**0.9340**	0.2090	0.2316	0.0285
	-0.0008	0.0004	0.0094	0.5317	**0.5207**	-0.0017	0.0159	-0.0150	-0.0139	**0.0568**	0.0034	0.0285	0.0426
4 month													
	-0.0014	0.9938	0.0547	-1.0049	0.9946	-0.0039	1.0298	0.0720	-2.0462	0.9991	1.0043	0.2197	0.0199
	0.0018	0.0018	0.9314	0.5031	0.9399	0.0042	-0.0337	0.9740	2.1917	**0.9337**	0.2197	0.2183	0.0294
	-0.0009	0.0004	0.0096	0.4578	**0.4485**	-0.0017	0.0160	-0.0150	-0.0142	**0.0567**	0.0199	0.0294	0.0448
5 month													
	-0.0014	0.9934	0.0566	-1.0069	0.9942	-0.0039	1.0293	0.0726	-2.0114	0.9993	0.9647	0.2175	0.0020
	0.0018	0.0019	0.9266	0.5078	0.9356	0.0042	-0.0331	0.9733	2.1611	**0.9341**	0.2175	0.2299	0.0220
	-0.0008	0.0003	0.0093	0.4835	**0.4737**	-0.0016	0.0157	-0.0145	0.0013	**0.0705**	0.0020	0.0220	0.0445
6 month													
	-0.0013	0.9938	0.0559	-1.0066	0.9945	-0.0038	1.0293	0.0727	-2.0201	0.9992	0.9778	0.2196	0.0022
	0.0018	0.0019	0.9290	0.5099	0.9382	0.0041	-0.0331	0.9732	2.1647	**0.9337**	0.2196	0.2241	0.0213
	-0.0008	0.0004	0.0094	0.4907	**0.4807**	-0.0016	0.0157	-0.0145	0.0013	**0.0708**	0.0022	0.0213	0.0469
9 month													
	-0.0013	0.9934	0.0539	-1.1175	0.9937	-0.0038	1.0293	0.0726	-2.0318	0.9991	1.0363	0.2149	0.0021
	0.0018	0.0017	0.9314	0.5117	0.9426	0.0041	-0.0331	0.9734	2.1703	**0.9347**	0.2149	0.2050	0.0124
	-0.0008	0.0004	0.0095	0.5490	**0.5375**	-0.0016	0.0157	-0.0142	0.0007	**0.0696**	0.0021	0.0124	0.0436
12 month													
	-0.0012	0.9933	0.0529	-1.1491	0.9935	-0.0038	1.0299	0.0712	-2.0807	0.9992	1.0113	0.2024	0.0011
	0.0017	0.0019	0.9286	0.4964	0.9401	0.0041	-0.0339	0.9736	2.2265	**0.9331**	0.2024	0.2049	0.0131
	-0.0008	0.0004	0.0097	0.5545	**0.5428**	-0.0016	0.0160	-0.0146	-0.0153	**0.0558**	0.0011	0.0131	0.0417

Panel B: Standard Loss Function

P-Dynamics					Q-Dynamics							
K_0^P	K_1^P			Eigenvalues	K_0^Q	K_1^Q			Eigenvalues	$\Sigma\Sigma' \times 1e4$		
-0.0021	0.9940	0.0549	0.3129	0.9948	0.0004	1.0052	0.1039	-0.2569	1.0000	0.9893	0.2297	0.0513
0.0004	0.0017	0.9337	0.1538	0.9274	-0.0003	-0.0073	0.9370	0.2717	0.9648	0.2297	0.2544	0.0238
0.0012	-0.0002	-0.0042	0.8084	0.8139	0.0003	0.0042	0.0136	0.8685	0.8458	0.0513	0.0238	0.1203

Notes to Table: We present the estimated P- and Q-parameters governing the state variables in the JSZ canonical form with forecasting loss function and standard loss function. The parameters are estimated using the entire sample. Panel A reports on the JSZ canonical form with forecasting loss function and variable portfolio weights. These estimates are forecast-horizon specific. Panel B reports on the JSZ canonical form with standard loss function. We conduct likelihood ratio tests on the eigenvalues estimated from different loss functions. The eigenvalues shown in boldface in Panel A are significantly different from the corresponding eigenvalues in Panel B.

Table 2.6 Out-of-Sample RMSEs: AFNS

Panel A: Forecasting Loss Function

Forecast Horizon k	1 month	2 month	3 month	4 month	5 month	6 month	9 month	12 month
3 month yield	36.12	46.23	56.23	67.45	79.12	88.45	118.35	142.12
6 month yield	27.68	43.48	55.12	67.29	78.09	89.13	121.22	150.24
1 year yield	36.39	51.91	65.33	76.87	86.87	98.15	124.90	152.97
2 year yield	45.85	60.75	72.38	82.62	90.60	100.90	122.71	146.88
3 year yield	42.29	57.20	68.23	77.77	85.22	94.25	113.45	134.92
4 year yield	36.45	51.87	62.90	71.96	79.13	86.94	104.56	123.89
5 year yield	30.57	46.72	58.58	67.94	75.61	82.73	99.72	117.36
10 year yield	38.30	54.53	59.76	65.39	73.64	79.09	95.46	106.86
20 year yield	51.13	60.13	68.06	75.18	79.50	84.24	93.32	105.71

Panel B: Standard Loss Function

Forecast Horizon k	1 month	2 month	3 month	4 month	5 month	6 month	9 month	12 month
3 month yield	40.49	57.84	73.01	84.12	95.69	107.43	141.22	167.42
6 month yield	33.48	52.14	69.17	83.55	97.18	110.22	144.37	172.31
1 year yield	37.09	56.92	75.04	90.40	104.03	116.35	148.01	174.43
2 year yield	38.15	61.79	80.97	96.67	110.02	121.83	150.57	174.87
3 year yield	38.43	61.33	79.48	94.27	106.82	117.69	143.54	166.01
4 year yield	36.37	57.97	74.88	88.42	99.99	110.08	133.60	154.58
5 year yield	34.98	55.50	71.56	84.56	95.65	105.14	127.03	146.83
10 year yield	34.24	51.02	63.49	74.06	83.19	90.87	109.12	126.70
20 year yield	28.18	41.00	50.92	59.64	67.20	73.49	88.64	103.57

Panel C: RMSE Ratio

Forecast Horizon k	1 month	2 month	3 month	4 month	5 month	6 month	9 month	12 month
3 month yield	0.89***	0.80**	0.77***	0.80	0.83***	0.82***	0.84***	0.85***
6 month yield	0.83***	0.83***	0.80***	0.81***	0.80***	0.81***	0.84***	0.87***
1 year yield	0.98***	0.91***	0.87***	0.85***	0.84***	0.84**	0.84**	0.88***
2 year yield	1.20**	0.98**	0.89**	0.85***	0.82***	0.83***	0.81***	0.84***
3 year yield	1.10***	0.93***	0.86**	0.83**	0.80*	0.80***	0.79**	0.81***
4 year yield	1.00	0.89***	0.84***	0.81***	0.79**	0.79*	0.78***	0.80**
5 year yield	0.87**	0.84***	0.82***	0.80**	0.79**	0.79***	0.79***	0.80**
10 year yield	1.12***	1.07***	0.94***	0.88***	0.89***	0.87***	0.87***	0.84***
20 year yield	1.81**	1.47***	1.34*	1.26**	1.18***	1.15***	1.05***	1.02***

Notes to Table: We present the out-of-sample RMSEs for the maximally flexible specification of the Gaussian AFNS model with forecasting loss function and standard loss function. Panel A reports on the AFNS model with forecasting loss function. At each time t and for each forecast horizon k, we estimate the model using data up to and including t, and forecast k periods ahead. Panel B reports on the AFNS model with standard loss function. We estimate the model using data up to and including t at each month t, and forecast one to twelve months ahead. RMSEs are reported in basis points. Panel C shows the ratios of the RMSEs in Panel A and Panel B. The statistical significance of the relative forecasting performance is evaluated using the Diebold and Mariano (1995) test statistic computed with a serial correlation robust variance and the small sample adjustment of Harvey, Leybourne, and Newbold (1997). The significance is denoted by *, ** and ***, corresponding to significance levels of 10%, 5% and 1% respectively.

Table 2.7 Out-of-Sample RMSE Ratios: Stochastic Volatility Models

Panel A: $A_1(3)$ with Latent Factors

Forecast Horizon	1 month	2 month	3 month	4 month	5 month	6 month	9 month	12 month
3 month yield	0.99*	0.98*	1.00*	0.98**	0.95**	0.93**	0.89**	0.84***
6 month yield	1.05	0.97*	0.99*	0.98**	0.96**	0.94**	0.84*	0.92***
1 year yield	1.01	0.98	0.98*	0.93**	0.95**	0.93**	0.91***	0.88***
2 year yield	1.23*	0.94*	0.95*	0.92**	0.90***	0.88***	0.82***	0.83***
3 year yield	1.18	0.91**	0.91**	0.88***	0.86***	0.85***	0.79***	0.80***
4 year yield	1.07*	0.90***	0.89**	0.84**	0.83***	0.81***	0.79***	0.76***
5 year yield	0.88**	0.88*	0.90**	0.88**	0.84***	0.83***	0.76***	0.76***
10 year yield	0.97**	0.89*	0.89*	0.90**	0.83**	0.83**	0.76***	0.72***
20 year yield	1.83**	1.33	1.12	1.00	0.92	0.86*	0.78***	0.76***

Panel B: $A_2(3)$ with Latent Factors

Forecast Horizon	1 month	2 month	3 month	4 month	5 month	6 month	9 month	12 month
3 month yield	0.99	1.00	0.99**	0.95***	0.93***	0.90***	0.85***	0.79***
6 month yield	1.05*	0.98**	1.00*	0.96***	0.93***	0.90***	0.79***	0.86***
1 year yield	0.99	0.96**	0.97**	0.95***	0.92***	0.94***	0.92***	0.88***
2 year yield	1.21	0.86*	0.96***	0.95***	0.94***	0.92***	0.83***	0.84***
3 year yield	1.14	0.82***	0.91***	0.89***	0.89***	0.87***	0.80***	0.81***
4 year yield	1.01**	0.83**	0.88***	0.84***	0.84***	0.82***	0.80***	0.74***
5 year yield	0.94*	0.85**	0.92***	0.92***	0.87***	0.86***	0.76***	0.77***
10 year yield	0.93**	0.84*	0.89**	0.95***	0.84***	0.85***	0.77***	0.71***
20 year yield	1.77**	1.25	1.05	0.93	0.85*	0.79*	0.74***	0.79***

Panel C: $A_3(3)$ with Latent Factors

Forecast Horizon	1 month	2 month	3 month	4 month	5 month	6 month	9 month	12 month
3 month yield	0.99**	0.99***	0.96***	0.96***	0.93***	0.91***	0.86***	0.80***
6 month yield	1.04***	0.99***	1.00	0.97***	0.93***	0.91**	0.79***	0.87***
1 year yield	1.00*	0.97***	0.99***	0.98***	0.94***	0.95***	0.96***	0.89***
2 year yield	1.20	0.86***	0.96***	0.96***	0.95***	0.93***	0.83***	0.84***
3 year yield	1.14**	0.82***	0.91***	0.90***	0.89***	0.87***	0.81***	0.82***
4 year yield	1.02**	0.83***	0.88***	0.85***	0.85***	0.82***	0.81***	0.75***
5 year yield	0.93***	0.85***	0.93***	0.92***	0.88***	0.87***	0.77***	0.78***
10 year yield	0.94**	0.84***	0.89***	0.96***	0.85***	0.86***	0.78***	0.72***
20 year yield	1.62***	1.21***	1.03	0.91**	0.83***	0.77***	0.75***	0.81***

Notes to Table: We present the out-of-sample RMSE ratios for the stochastic volatility models with forecasting loss function and standard loss function. Panel A reports on the $A_1(3)$ model, Panel B reports on the $A_2(3)$ model, and Panel C reports on the $A_3(3)$ model. For the forecasting loss function, at each time t and for each forecast horizon k, we estimate the models using data up to and including t, and forecast k periods ahead. For the standard loss function, we estimate the models using data up to and including t at each month t, and forecast one to twelve months ahead. The statistical significance of the relative forecasting performance is evaluated using the Diebold and Mariano (1995) test statistic computed with a serial correlation robust variance and the small sample adjustment of Harvey, Leybourne, and Newbold (1997). The significance is denoted by *, ** and ***, corresponding to significance levels of 10%, 5% and 1% respectively.

Table 2.8 Out-of-Sample MAEs: JSZ Canonical Form

Panel A: Forecasting Loss Function

Forecast Horizon k	1 month	2 month	3 month	4 month	5 month	6 month	9 month	12 month
3 month yield	26.77	30.31	36.95	45.64	55.85	66.85	95.06	119.68
6 month yield	17.80	28.90	38.82	50.46	62.78	75.30	104.70	129.76
1 year yield	26.96	41.60	52.12	62.80	74.13	85.44	111.76	135.70
2 year yield	36.82	49.82	60.10	70.03	79.83	89.08	111.90	132.78
3 year yield	34.33	47.25	56.60	66.07	74.55	82.95	103.16	122.35
4 year yield	29.01	42.93	51.96	60.00	66.92	74.74	92.92	111.32
5 year yield	24.81	39.04	48.44	56.61	63.01	70.11	85.99	103.78
10 year yield	27.98	38.22	46.33	52.64	58.40	63.40	74.86	88.16
20 year yield	41.36	48.18	52.95	56.84	60.64	64.10	71.78	81.17

Panel B: Standard Loss Function

Forecast Horizon k	1 month	2 month	3 month	4 month	5 month	6 month	9 month	12 month
3 month yield	26.62	38.94	50.45	60.44	71.37	81.41	109.79	131.29
6 month yield	24.41	39.89	53.84	66.38	78.76	90.22	118.78	140.71
1 year yield	29.20	46.52	62.23	75.78	88.55	99.91	125.80	146.25
2 year yield	31.94	50.90	66.51	80.31	92.32	102.68	126.89	146.24
3 year yield	30.42	48.51	62.85	75.72	86.91	96.88	119.29	137.17
4 year yield	27.37	44.30	57.55	68.97	79.33	88.91	110.18	126.74
5 year yield	25.49	41.68	54.09	64.93	74.74	84.12	104.19	119.86
10 year yield	25.65	38.72	48.91	58.02	65.95	72.67	89.15	102.80
20 year yield	20.91	32.15	40.82	48.58	55.16	60.30	72.45	84.56

Panel C: MAE Ratio

Forecast Horizon k	1 month	2 month	3 month	4 month	5 month	6 month	9 month	12 month
3 month yield	1.01	0.78***	0.73***	0.76***	0.78***	0.82***	0.87***	0.91***
6 month yield	0.73***	0.72***	0.72***	0.76***	0.80***	0.83***	0.88***	0.92***
1 year yield	0.92*	0.89***	0.84***	0.83***	0.84***	0.86***	0.89***	0.93***
2 year yield	1.15	0.98	0.90***	0.87***	0.86***	0.87***	0.88***	0.91***
3 year yield	1.13**	0.97***	0.90***	0.87***	0.86***	0.86***	0.86***	0.89***
4 year yield	1.06**	0.97*	0.90***	0.87***	0.84***	0.84***	0.84***	0.88***
5 year yield	0.97	0.94**	0.90***	0.87***	0.84***	0.83***	0.83***	0.87***
10 year yield	1.09*	0.99	0.95	0.91**	0.89***	0.87***	0.84***	0.86***
20 year yield	1.98***	1.50*	1.30**	1.17***	1.10**	1.06	0.99	0.96*

Notes to Table: We present the out-of-sample MAEs for the JSZ canonical form with forecasting loss function and standard loss function. Panel A reports on the JSZ canonical form with forecasting loss function and variable portfolio weights. At each time t and for each forecast horizon k, we estimate the model using data up to and including t, and forecast k periods ahead. Panel B reports on the JSZ canonical form with standard loss function and fixed portfolio weights. We estimate the model using data up to and including t at each month t, and forecast one to twelve months ahead. MAEs are reported in basis points. Panel C shows the ratios of the MAEs in Panel A and Panel B. The statistical significance of the relative forecasting performance is evaluated using the Diebold and Mariano (1995) test statistic computed with a serial correlation robust variance and the small sample adjustment of Harvey, Leybourne, and Newbold (1997). The significance is denoted by *, ** and ***, corresponding to significance levels of 10%, 5% and 1% respectively.

Table 2.A.1 Out-of-Sample RMSEs: $A_0(3)$ with Latent Factors

Panel A: Forecasting Loss Function

Forecast Horizon k	1 month	2 month	3 month	4 month	5 month	6 month	9 month	12 month
3 month yield	39.24	56.64	70.37	80.35	90.75	101.65	133.55	158.00
6 month yield	34.18	54.44	71.25	85.20	98.20	110.87	136.98	180.50
1 year yield	38.60	61.01	79.85	94.82	107.94	119.97	154.49	182.17
2 year yield	51.21	64.62	83.59	97.54	110.05	121.12	145.67	177.11
3 year yield	47.40	60.41	77.75	90.06	101.70	111.63	134.90	164.27
4 year yield	39.44	55.08	70.79	81.39	92.10	101.11	125.24	147.93
5 year yield	31.38	51.20	67.98	80.18	89.40	98.91	117.93	143.79
10 year yield	38.75	50.39	64.41	77.58	83.66	92.91	111.43	130.02
20 year yield	43.11	57.40	65.76	73.09	78.31	81.67	94.18	107.69

Panel B: Standard Loss Function

Forecast Horizon k	1 month	2 month	3 month	4 month	5 month	6 month	9 month	12 month
3 month yield	42.43	57.80	73.42	88.17	104.39	121.31	150.07	165.45
6 month yield	36.97	56.15	70.62	86.27	102.28	120.48	152.18	187.41
1 year yield	38.95	60.58	80.87	96.05	109.16	122.05	157.64	188.16
2 year yield	41.28	65.69	83.41	98.41	112.67	128.05	153.59	189.39
3 year yield	38.79	62.79	77.24	93.07	110.19	126.17	154.13	178.94
4 year yield	35.39	58.78	73.27	89.29	104.76	119.57	132.66	165.83
5 year yield	34.09	53.12	70.81	86.26	101.18	115.26	140.62	162.65
10 year yield	33.40	50.65	63.46	77.44	89.47	101.59	136.72	155.50
20 year yield	31.59	44.73	59.00	72.21	85.31	97.55	121.19	134.25

Panel C: RMSE Ratio

Forecast Horizon k	1 month	2 month	3 month	4 month	5 month	6 month	9 month	12 month
3 month yield	0.92	0.98*	0.96***	0.91***	0.87***	0.84***	0.89***	0.95***
6 month yield	0.92*	0.97**	1.01***	0.99***	0.96***	0.92***	0.90**	0.96***
1 year yield	0.99	1.01*	0.99***	0.99***	0.99***	0.98***	0.98***	0.97***
2 year yield	1.24	0.98	1.00*	0.99***	0.98***	0.95***	0.95***	0.94***
3 year yield	1.22	0.96*	1.01***	0.97***	0.92***	0.88***	0.88***	0.92***
4 year yield	1.11	0.94**	0.97***	0.91***	0.88***	0.85***	0.94***	0.89***
5 year yield	0.92*	0.96***	0.96***	0.93***	0.88***	0.86***	0.84***	0.88***
10 year yield	1.16	0.99*	1.01***	1.00	0.94***	0.91***	0.82***	0.84***
20 year yield	1.36***	1.28	1.11	1.01	0.92	0.84**	0.78***	0.80***

Notes to Table: We present the out-of-sample RMSEs for the $A_0(3)$ latent model with forecasting loss function and standard loss function. Panel A reports on the $A_0(3)$ model with forecasting loss function. At each time t and for each forecast horizon k, we estimate the model using data up to and including t, and forecast k periods ahead. Panel B reports on the $A_0(3)$ model with standard loss function. We estimate the model using data up to and including t at each month t, and forecast one to twelve months ahead. RMSEs are reported in basis points. Panel C shows the ratios of the RMSEs in Panel A and Panel B. The statistical significance of the relative forecasting performance is evaluated using the Diebold and Mariano (1995) test statistic computed with a serial correlation robust variance and the small sample adjustment of Harvey, Leybourne, and Newbold (1997). The significance is denoted by *, ** and ***, corresponding to significance levels of 10%, 5% and 1% respectively.

Table 2.A.2 In-Sample RMSEs: AFNS

				Panel A: Forecasting Loss Function				
Forecast Horizon k	1 month	2 month	3 month	4 month	5 month	6 month	9 month	12 month
3 month yield	36.28	39.04	41.13	43.20	50.28	50.23	50.01	47.61
6 month yield	21.60	23.89	25.21	26.38	20.13	20.11	19.97	28.36
1 year yield	18.35	19.43	20.13	20.80	18.95	18.95	19.01	22.72
2 year yield	28.06	31.98	34.55	36.93	28.91	28.86	28.72	41.56
3 year yield	27.26	29.89	31.80	33.64	30.38	30.32	30.07	37.01
4 year yield	22.68	24.06	25.07	26.06	28.74	28.70	28.46	27.73
5 year yield	16.19	16.63	16.97	17.31	26.66	26.64	26.48	17.71
10 year yield	22.26	24.75	26.29	27.70	24.24	24.40	24.67	30.24
20 year yield	40.99	43.34	45.15	46.95	75.70	75.70	75.37	50.15
				Panel B: Standard Loss Function				
3 month yield					18.56			
6 month yield					15.87			
1 year yield					19.90			
2 year yield					10.81			
3 year yield					9.18			
4 year yield					10.27			
5 year yield					11.64			
10 year yield					12.67			
20 year yield					10.00			

Notes to Table: We present the in-sample RMSEs for the maximally flexible specification of the AFNS model with forecasting loss function (Panel A) and standard loss function (Panel B). The estimates of the specification with forecasting loss function are forecast-horizon specific, so we report the in-sample RMSEs for each forecast horizon. RMSEs are reported in basis points.

Table 2.A.3 In-Sample RMSEs: AFNS with Independent Factors

				Panel A: Forecasting Loss Function				
Forecast Horizon k	1 month	2 month	3 month	4 month	5 month	6 month	9 month	12 month
3 month yield	43.91	46.78	53.67	53.56	67.65	63.44	63.29	62.97
6 month yield	24.60	24.56	24.50	24.43	22.79	24.36	24.22	23.96
1 year yield	24.60	24.62	24.65	24.68	18.64	24.73	24.85	25.01
2 year yield	35.77	35.69	35.63	35.56	37.72	35.48	35.41	35.27
3 year yield	38.88	38.75	38.64	38.52	35.62	38.37	38.18	37.84
4 year yield	37.69	37.57	37.46	37.33	27.68	37.18	36.94	36.52
5 year yield	33.55	33.51	33.43	33.36	19.48	33.29	33.09	32.69
10 year yield	29.08	29.76	30.16	30.63	26.83	31.38	31.75	31.96
20 year yield	51.45	51.40	61.24	61.11	71.69	81.00	85.52	89.72
				Panel B: Standard Loss Function				
3 month yield					23.80			
6 month yield					15.72			
1 year yield					27.61			
2 year yield					11.38			
3 year yield					7.59			
4 year yield					12.95			
5 year yield					18.03			
10 year yield					20.23			
20 year yield					24.45			

Notes to Table: We present the in-sample RMSEs for the independent-factor AFNS model with forecasting loss function (Panel A) and standard loss function (Panel B). The estimates of the specification with forecasting loss function are forecast-horizon specific, so we report the in-sample RMSEs for each forecast horizon. RMSEs are reported in basis points.

Table 2.A.4 In-Sample RMSEs: Stochastic Volatility Models

Panel A: $A_1(3)$ with Latent Factors

	Forecasting Loss Function								**Standard Loss Function**
Forecast Horizon k	1 month	2 month	3 month	4 month	5 month	6 month	9 month	12 month	
3 month yield	35.30	43.35	44.36	38.47	42.75	48.67	54.68	58.54	18.21
6 month yield	21.58	20.05	20.48	20.74	21.27	23.74	30.34	41.92	15.66
1 year yield	20.17	22.44	22.14	23.33	26.54	24.24	35.47	46.59	17.26
2 year yield	26.10	27.69	27.44	27.63	28.70	29.08	32.17	35.89	11.01
3 year yield	27.68	28.60	28.45	28.41	28.94	30.00	30.10	30.88	7.83
4 year yield	24.63	25.14	25.23	24.94	25.43	27.42	26.10	25.83	10.43
5 year yield	20.20	19.84	19.81	19.90	20.36	25.05	20.15	20.11	12.12
10 year yield	22.49	21.82	21.95	21.74	21.83	28.08	24.15	25.59	12.82
20 year yield	40.08	39.98	39.50	39.63	40.78	47.67	43.23	46.95	15.34

Panel B: $A_2(3)$ with Latent Factors

	Forecasting Loss Function								**Standard Loss Function**
Forecast Horizon k	1 month	2 month	3 month	4 month	5 month	6 month	9 month	12 month	
3 month yield	34.96	35.61	37.63	40.04	64.62	44.21	51.89	44.88	18.18
6 month yield	21.32	20.42	21.26	20.46	25.42	23.03	26.15	29.86	16.77
1 year yield	19.69	21.39	21.48	22.10	26.29	23.45	31.00	37.55	18.51
2 year yield	26.59	27.12	27.09	27.49	28.61	28.46	29.64	32.18	14.07
3 year yield	27.98	28.24	28.35	28.32	29.50	29.30	28.57	29.69	12.57
4 year yield	24.78	24.91	25.22	24.84	26.68	26.35	24.97	25.99	11.79
5 year yield	19.85	20.21	21.40	19.91	21.15	23.22	19.39	22.15	10.77
10 year yield	21.77	22.03	23.27	22.02	23.59	25.44	23.26	26.69	11.74
20 year yield	39.55	40.23	40.48	40.13	39.55	43.39	42.94	48.83	16.42

Panel C: $A_3(3)$ with Latent Factors

	Forecasting Loss Function								**Standard Loss Function**
Forecast Horizon k	1 month	2 month	3 month	4 month	5 month	6 month	9 month	12 month	
3 month yield	35.28	35.44	40.43	36.05	57.04	38.87	78.56	54.56	19.29
6 month yield	19.27	18.96	21.85	20.48	23.21	19.84	40.10	33.59	17.79
1 year yield	19.61	21.35	20.92	20.82	24.40	23.11	37.90	38.47	19.51
2 year yield	25.52	26.60	27.38	25.99	26.88	26.88	31.64	32.43	14.92
3 year yield	26.78	27.45	28.55	26.92	27.37	27.61	29.77	29.68	13.34
4 year yield	23.68	24.14	25.49	23.69	24.20	24.95	26.26	25.70	12.50
5 year yield	18.79	19.60	22.00	18.93	18.86	22.28	21.06	20.51	11.42
10 year yield	21.01	21.07	22.22	20.90	21.74	24.94	22.61	22.22	11.39
20 year yield	37.98	38.52	38.88	38.09	38.92	41.96	41.11	41.27	18.99

Notes to Table: We present the in-sample RMSEs for the $A_1(3)$, $A_2(3)$ and $A_3(3)$ stochastic volatility models with forecasting loss function and standard loss function. For the models with forecasting loss function, the estimates are forecast-horizon specific, so we report the in-sample RMSEs for each forecast horizon. RMSEs are reported in basis points.

Table 2.A.5 Out-of-Sample RMSEs: AFNS Model with Independent Factors

Panel A: Forecasting Loss Function

Forecast Horizon k	1 month	2 month	3 month	4 month	5 month	6 month	9 month	12 month
3 month yield	36.47	45.70	57.08	66.56	82.69	88.66	117.24	138.26
6 month yield	27.19	41.32	53.54	65.12	73.99	85.41	114.56	139.89
1 year yield	35.63	48.71	61.25	71.07	80.26	90.77	115.08	138.85
2 year yield	42.08	56.41	66.99	76.17	82.68	93.04	112.24	132.02
3 year yield	38.46	53.33	63.36	72.01	78.07	86.98	103.93	120.83
4 year yield	34.21	49.08	59.17	67.43	73.55	80.73	96.47	111.03
5 year yield	29.20	44.16	55.30	63.83	70.71	76.81	92.27	105.37
10 year yield	39.09	51.01	58.66	67.12	73.10	77.48	90.06	103.74
20 year yield	52.04	65.42	71.97	79.83	82.87	86.55	94.12	106.21

Panel B: Standard Loss Function

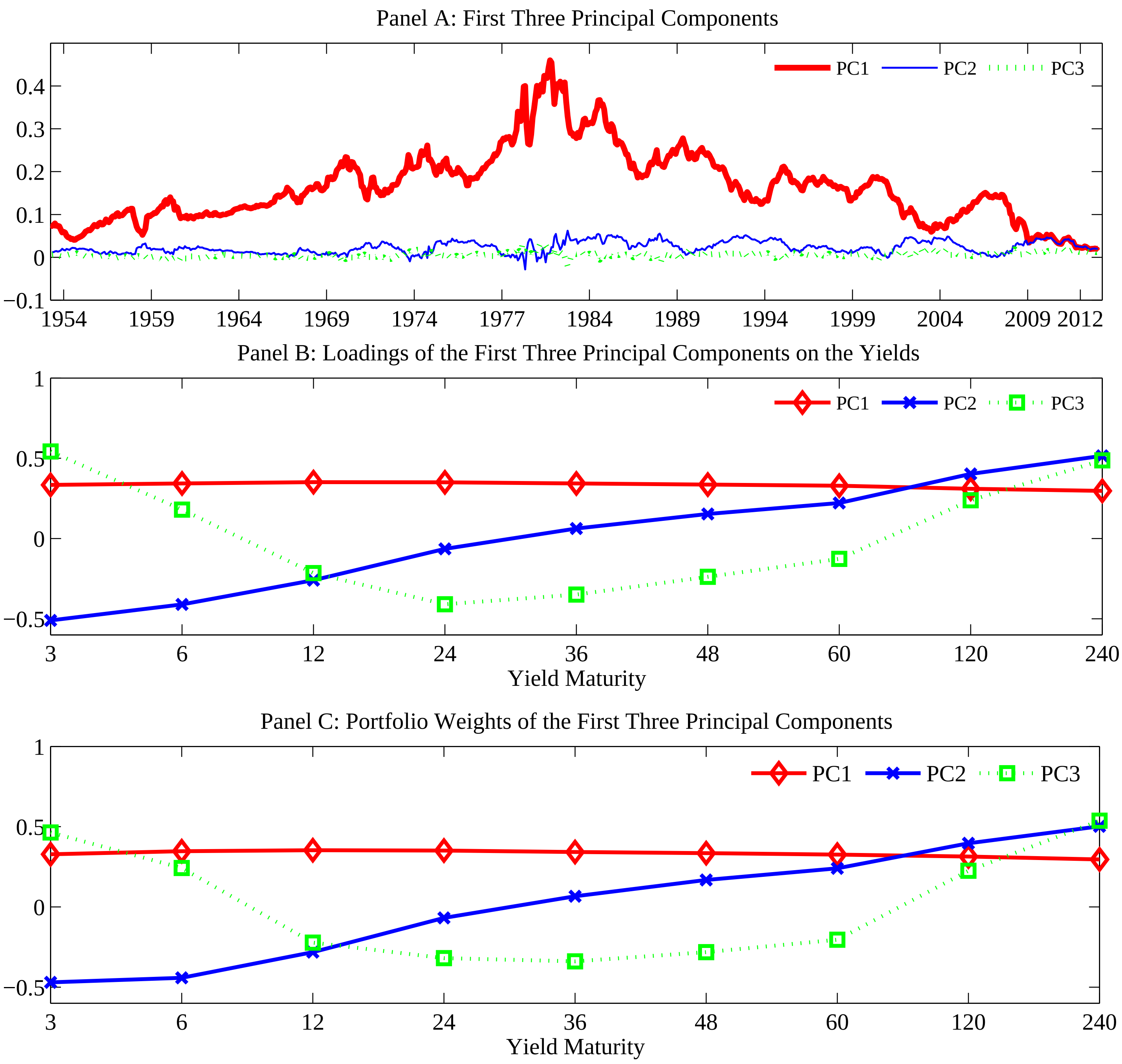
Forecast Horizon k	1 month	2 month	3 month	4 month	5 month	6 month	9 month	12 month
3 month yield	38.65	52.85	64.87	73.96	84.19	94.37	123.39	149.18
6 month yield	29.80	43.90	57.40	69.31	80.60	91.45	121.66	149.24
1 year yield	36.67	50.97	64.18	76.14	86.58	96.19	123.08	148.72
2 year yield	34.27	52.17	66.18	78.20	88.36	97.24	121.17	144.86
3 year yield	33.14	51.07	64.60	76.11	85.69	93.78	115.11	137.18
4 year yield	33.57	50.98	63.61	74.13	82.87	90.37	109.34	129.73
5 year yield	34.38	51.21	63.69	74.10	82.72	89.84	107.55	126.27
10 year yield	37.84	51.71	62.13	71.28	78.97	85.10	100.39	115.98
20 year yield	45.64	59.50	69.61	77.72	84.05	88.62	97.99	106.57

Panel C: RMSE Ratio

Forecast Horizon k	1 month	2 month	3 month	4 month	5 month	6 month	9 month	12 month
3 month yield	0.94**	0.86**	0.88**	0.90***	0.98**	0.94**	0.95***	0.93***
6 month yield	0.91**	0.94**	0.93***	0.94**	0.92***	0.93**	0.94***	0.94***
1 year yield	0.97**	0.96**	0.95**	0.93**	0.93**	0.94***	0.94**	0.93***
2 year yield	1.23**	1.08**	1.01**	0.97***	0.94**	0.96**	0.93***	0.91**
3 year yield	1.16**	1.04***	0.98***	0.95***	0.91**	0.93**	0.90***	0.88***
4 year yield	1.02**	0.96**	0.93**	0.91**	0.89***	0.89**	0.88***	0.86***
5 year yield	0.85***	0.86***	0.87**	0.86***	0.85***	0.85***	0.86***	0.83***
10 year yield	1.03***	0.99***	0.94***	0.94***	0.93***	0.91***	0.90***	0.89***
20 year yield	1.14***	1.10***	1.03***	1.03***	0.99***	0.98***	0.96***	1.00

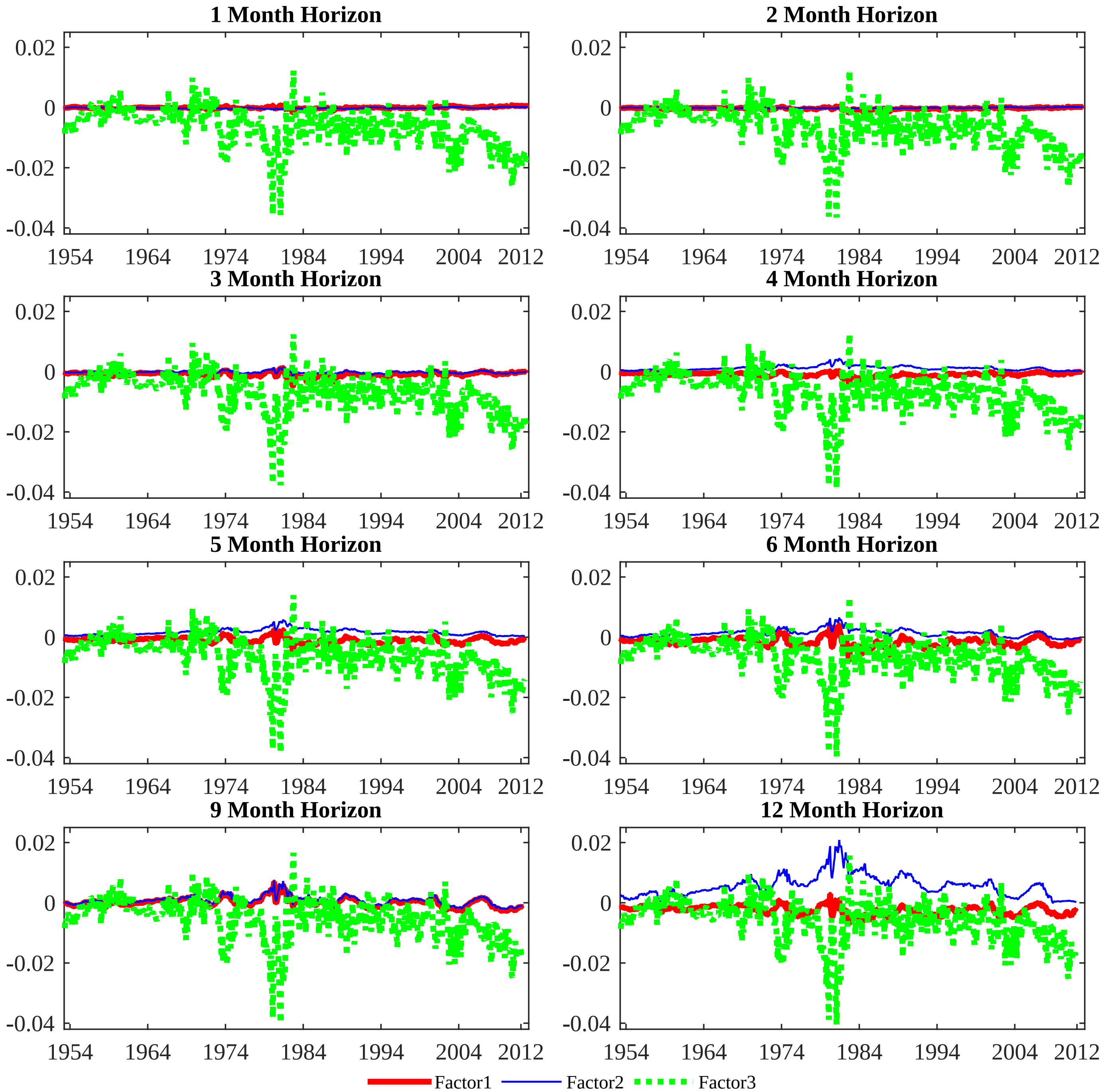
Notes to Table: We present the out-of-sample RMSEs for the independent-factor AFNS model with forecasting loss function and standard loss function. Panel A reports on the independent-factor AFNS model with forecasting loss function. At each time t and for each forecast horizon k, we estimate the model using data up to and including t, and forecast k periods ahead. Panel B reports on the independent-factor AFNS model with standard loss function. We estimate the model using data up to and including t at each month t, and forecast one to twelve months ahead. RMSEs are reported in basis points. Panel C shows the ratios of the RMSEs in Panel A and Panel B. The statistical significance of the relative forecasting performance is evaluated using the Diebold and Mariano (1995) test statistic computed with a serial correlation robust variance and the small sample adjustment of Harvey, Leybourne, and Newbold (1997). The significance is denoted by *, ** and ***, corresponding to significance levels of 10%, 5% and 1% respectively.

Figure 2.1 State Variables for the JSZ Canonical Form with Standard Loss Function



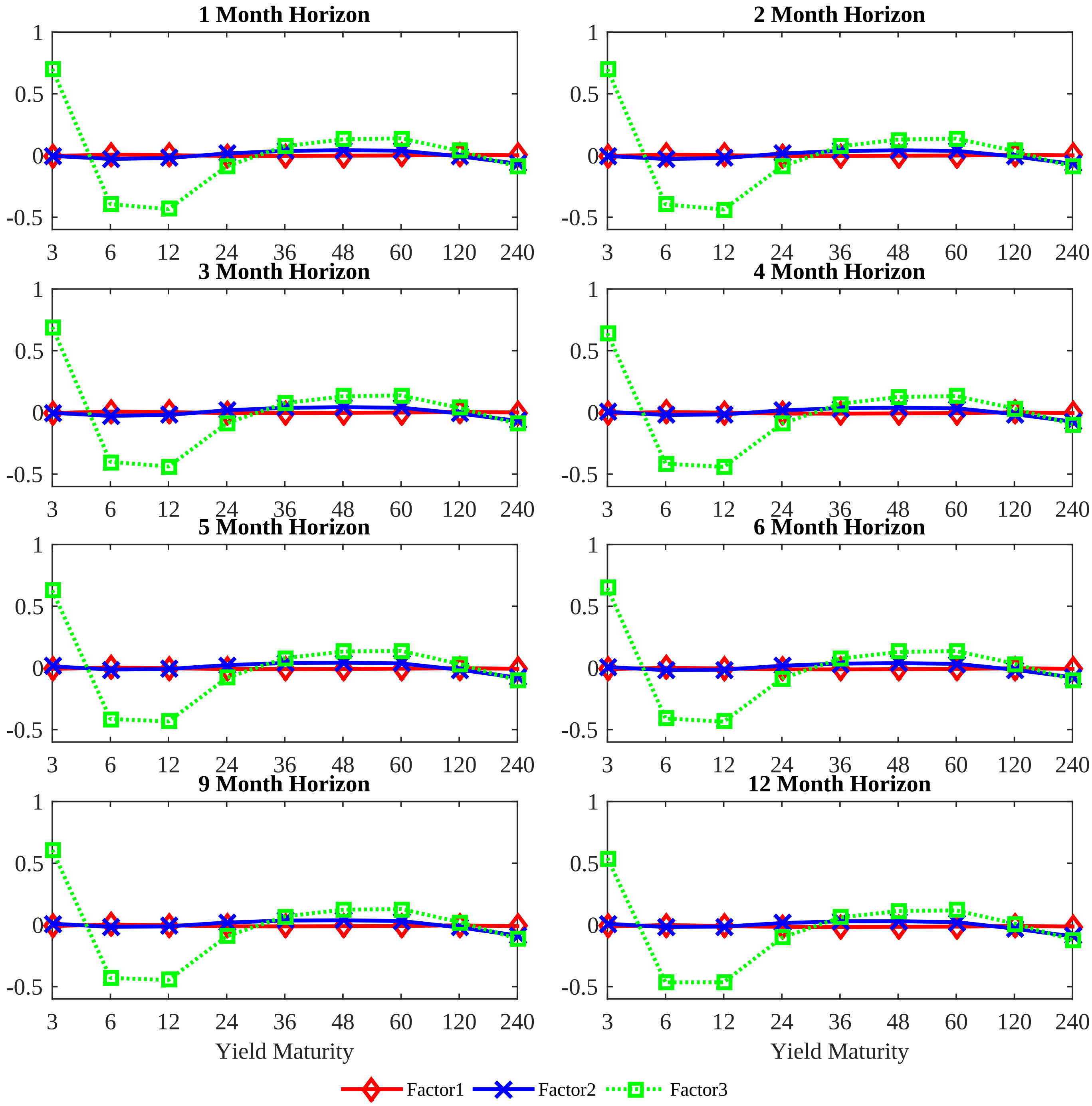
Notes to Figure: For the JSZ canonical form with standard loss function, Panel A shows the time series of the first three principal components PC, Panel B shows the factor loadings $B(\Theta^Q)$ on the yield curve, and Panel C shows the portfolio weights W_0.

Figure 2.2 Differences Between State Variables Using Forecasting and Standard Loss Functions



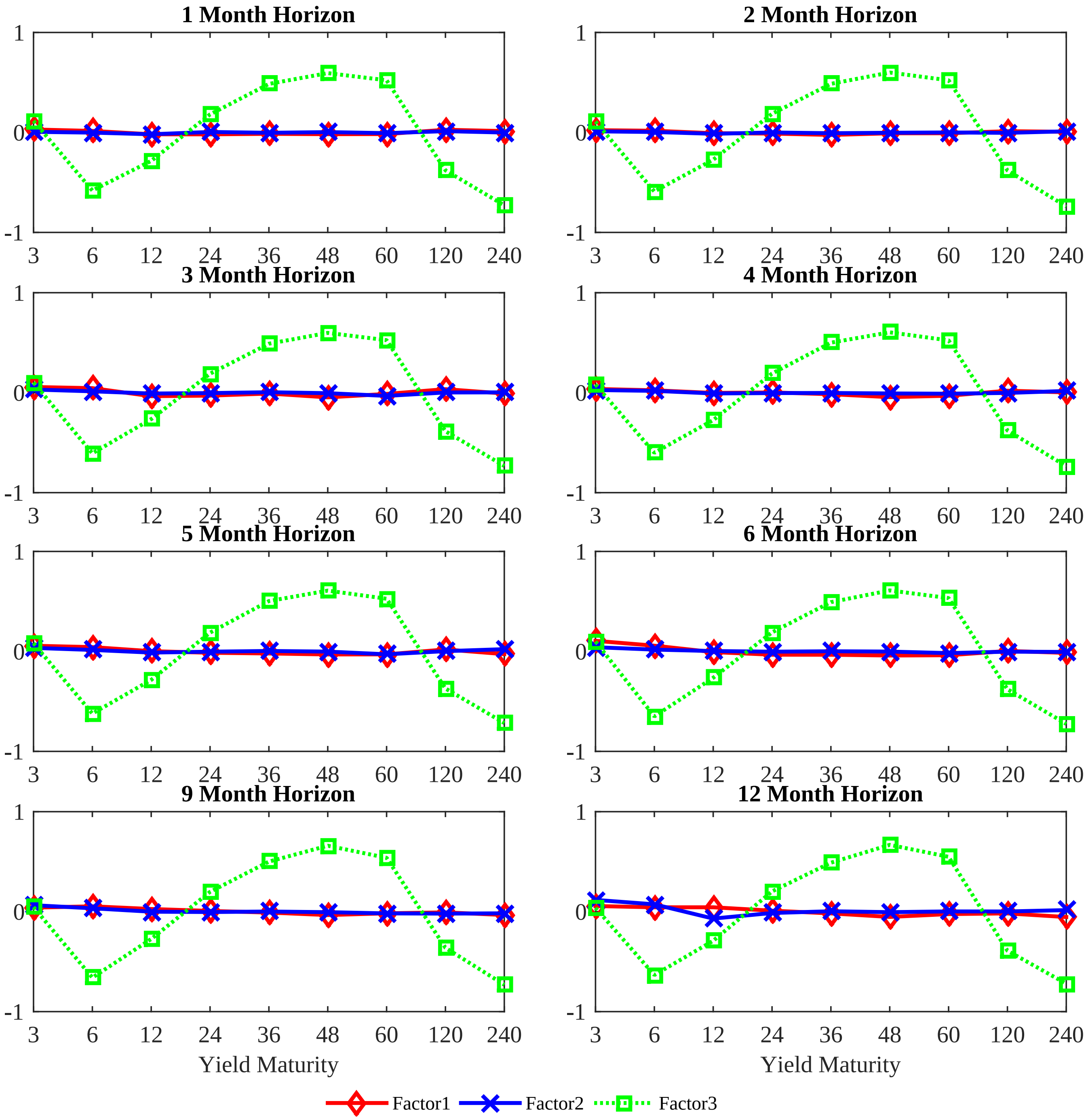
Notes to Figure: We plot the differences between the state variables from the JSZ canonical form with forecasting loss function and variable portfolio weights and the state variables from the JSZ canonical form with standard loss function. The state variables are estimated using the entire sample. The panels represent the differences in the estimates for different forecast horizons.

Figure 2.3 Differences Between Factor Loadings Using Forecasting and Standard Loss Functions



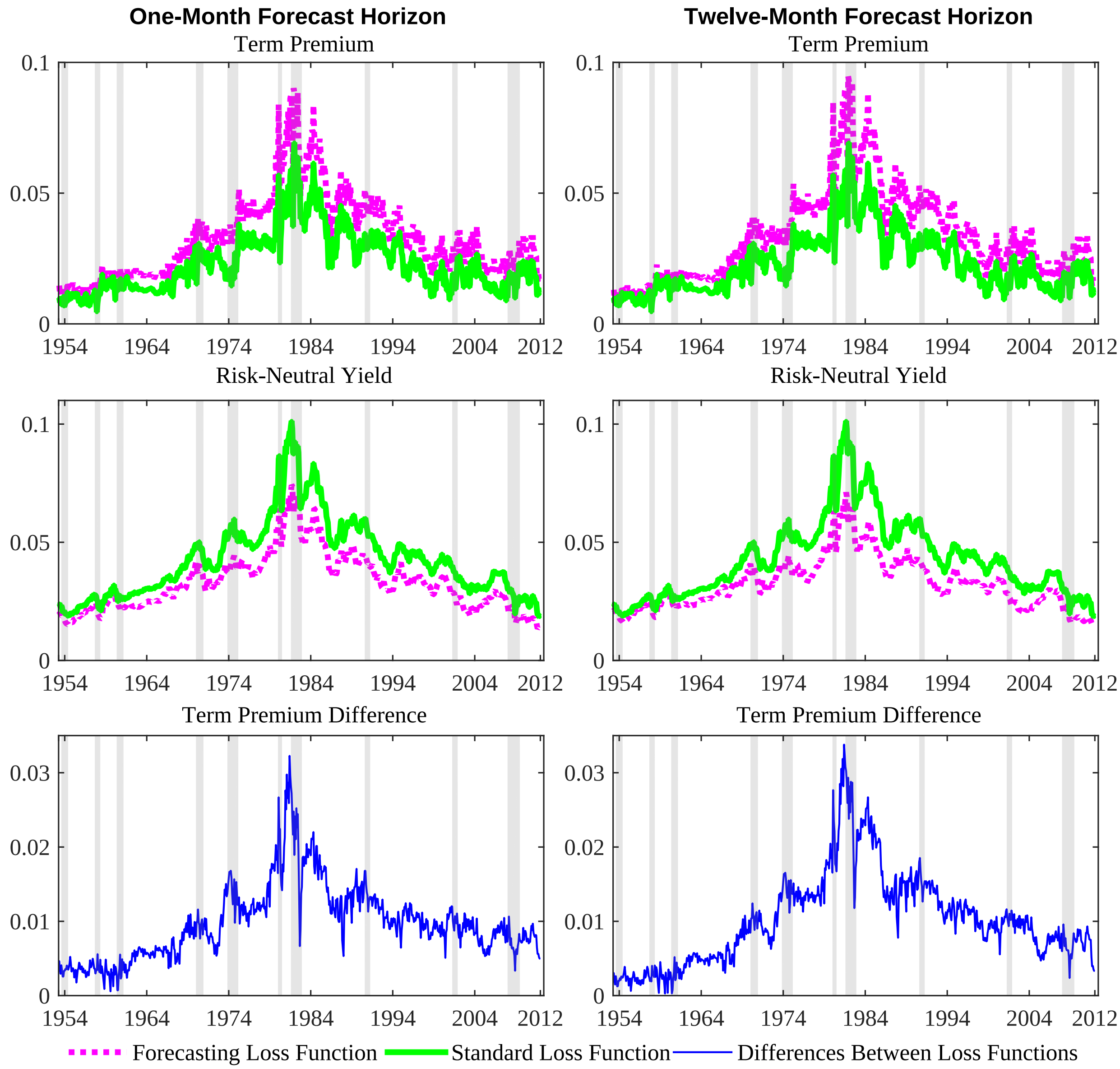
Notes to Figure: We plot the differences between the factor loadings from the JSZ canonical form with forecasting loss function and variable portfolio weights and the factor loadings from the JSZ canonical form with standard loss function. The factor loadings are estimated using the entire sample. The panels represent the differences in the estimates for different forecast horizons.

Figure 2.4 Differences Between Portfolio Weights Using Forecasting and Standard Loss Functions



Notes to Figure: We plot the differences between the variable portfolio weights from the JSZ canonical form with forecasting loss function and the fixed portfolio weights from the JSZ canonical form with standard loss function. The portfolio weights are estimated using the entire sample. The panels represent the differences in the estimates for different forecast horizons.

Figure 2.5 Decomposing Forward Rates



Notes to Figure: We plot model estimates of term premia and risk-neutral yields (the expectations component of yields) for five-to-ten-year forward rates. The highlighted line shows the estimates from the JSZ canonical form with standard loss function. The dotted line shows the estimates from the JSZ canonical form with forecasting loss function and variable portfolio weights. The solid line in the bottom two panels represents the differences between the term premia from the JSZ canonical form with forecasting loss function and variable portfolio weights and the term premia from the JSZ canonical form with standard loss function. The panels on the left plot results for the one-month forecast horizon, and the panels on the right plot results for the twelve-month forecast horizon. The shaded areas indicate recessions as defined by the National Bureau of Economic Research (NBER).

Chapter 3

Macroeconomic Determinants of the Term Structure: Long-Run and Short-Run Dynamics

3.1 Introduction

The literature on modeling the term structure of default-free interest rates is extensive. Many existing term structure models use dynamic no-arbitrage models with latent factors to explain the movements of the yield curve. Early studies include Vasicek (1977) and Cox, Ingersoll, and Ross (1985), which are special cases of the affine class described by Duffie and Kan (1996). The drawback of these models is that they do not readily provide information about the underlying economic determinants of the yield curve. More recently, a burgeoning literature incorporates macroeconomic factors, specifically real activity growth and inflation, into no-arbitrage term structure models.[1] These studies confirm our intuition

[1]The literature on term structure models with macroeconomic variables is too large to cite in full here. Some important papers include Ang and Piazzesi (2003), Diebold, Rudebusch, and Aruoba (2006), Ang, Piazzesi, and Wei (2006), Lu and Wu (2009), Ang, Boivin, Dong, and Loo-Kung (2011), Joslin, Priebsch, and Singleton (2014), and Ajello, Benzoni, and Chyruk (2012). See Duffee (2013) and Gürkaynak and Wright (2012) for an overview of this literature.

that macroeconomic fluctuations are an important source of uncertainty that affect bond yield dynamics.

This paper contributes to the literature on the term structure of interest rates with macroeconomic variables. We propose a no-arbitrage VAR model of term structure dynamics with two macroeconomic variables, real activity growth and inflation. The macroeconomic variables contain long-run and short-run components, which amounts to autoregressive specifications for the macro variables with a time-varying stochastic mean. This specification is motivated by a strand of the macroeconomic literature which argues that macroeconomic variables, such as inflation, follow first-order autoregressive processes with time-varying parameters. The macroeconomic variables contain a slowly moving trend component that is related to the monetary policy target, and the trending component is approximated by the long-run mean of the macro variables. This decomposition has been widely used in the macroeconomic literature, see for instance Cogley and Sargent (2001, 2005), Erceg and Levin (2003), Stock and Watson (2007), Cogley, Primiceri, and Sargent (2010), and Faust and Wright (2011).

We incorporate these stylized features of macroeconomic variables into the modeling of the yield curve. Our main objective is to study the impact of long-run and short-run shocks to the macroeconomic variables on the term structure of yields. We find that the impact of a shock to short-run inflation decays very quickly to zero, while a shock to short-run real activity growth survives much longer for all maturities. On the other hand, a shock to long-run inflation dominates a shock to long-run real activity growth and survives at very long horizons.

We decompose the proportion of total forecast variance that can be attributed to each of the four state variables at different maturities, and a complex picture emerges. By ignoring the component structure of inflation and real activity growth, one would conclude that inflation is the main determinant of the variance at all maturities and for all forecast horizons. In the component model, the relative importance of inflation and real activity growth critically depends on the maturity and the forecast horizon. The short-run component of inflation explains a large proportion of the short-horizon forecast variance at the short end of the term structure. The long-run component of inflation on the other hand explains most of the long-horizon forecast variance for long-maturity yields.

Interest rate expectations and term premia implied by the model with long-run and short-run components are more plausible than those implied by existing models. Model-implied short-term interest rate expectations exhibit a substantial decline over the last two decades, consistent with survey-based expectations of inflation and policy rates (see for example Kozicki and Tinsley, 2001a; Kim and Orphanides, 2012; Wright, 2011). Term premia, estimated as the difference between long-term rates and expected future short-term rates, are more stable in the component model. This finding is particularly noteworthy because the excessive variability of term premia constitutes a puzzle in the literature on dynamic term structure models (Kim and Orphanides, 2012).

We examine the predictability of bond excess returns using macro variables and their components filtered from the component model. Incorporating the long-run components in standard predictive regressions improves the forecast performance relative to models with macro variables or relative to information in the current yield curve, especially for bonds with two- to five-year maturities. These results suggest that the filtered long-run components help uncover information that is also captured by models with hidden or unspanned risk factors (Duffee, 2011; Joslin, Priebsch, and Singleton, 2014).

We compare the in-sample and out-of-sample performance of the newly proposed component model with several alternative specifications that are widely used in the literature. The component model provides substantial improvements in in-sample fit relative to the benchmark macro model without long-run components, especially for long-maturity yields, and it has uniformly better out-of-sample performance than the benchmark model. For long maturities, the component model also outperforms models with three latent factors and models with three latent factors and macroeconomic variables. These results suggest that the long-run component of inflation is important for yield curve modeling, both for improving in-sample and out-of-sample performance.

Our analysis is related to several other existing strands of the fixed income literature. Some studies integrate state variables with stochastic means, which are usually referred to as shifting endpoints, into no-arbitrage models. Part of this literature focuses on latent models, see for instance Kozicki and Tinsley (1998, 2001a), who find that models with shifting endpoints perform much better for explaining long-term yields. Other studies use no-arbitrage term structure models with macroeconomic variables that contain shifting endpoints. These models embed various economic restrictions, and are mainly focused on the role of the term structure in identifying the monetary policy rule and forecasting macroeconomic variables.[2] In contrast, the economic structure of our model is simpler, consisting of a simple Taylor rule as in Ang and Piazzesi (2003), and our focus is on the in- and out-of-sample performance of the term structure model and the impact of macroeconomic shocks on the term structure. On the other hand, our model is more complex with respect to the richness of the risk neutralization and the structure of the shocks to the macro variables.

Our analysis is also related to fixed-income studies that model change in the structure of the model and the parameters. One strand of the literature formulates and estimates regime-switching models.[3] Another literature models the effect of learning, often in the context of a more elaborate economic model.[4] While the long-run mean of inflation and real activity growth in our model can be thought of as the result of a learning process, our results may just as well be interpreted as resulting from an improved statistical specification of the macro variables.

The remainder of the paper is organized as follows. Section 2 presents the term structure model with long-run components. Section 3 describes the data and the estimation strategy.

[2] See for instance Dewachter and Lyrio (2006, 2008), Dewachter, Lyrio, and Maes (2006), and Berardi (2009).

[3] See for instance Hamilton (1988), Ang and Bekaert (2002), Ang, Bekaert, and Wei (2008), Bikbov and Chernov (2010), Bansal and Zhou (2002), and Dai, Singleton, and Yang (2007).

[4] See for instance Dewachter and Lyrio (2008), Laubach, Tetlow, and Williams (2007), and Orphanides and Wei (2012).

Section 4 presents in-sample estimation results. Sections 5, 6, and 7 discuss the economic implications of the models. Section 8 presents out-of-sample results, and Section 9 concludes.

3.2 The Model

In this section, we describe the pricing model for default-free bonds. Let r_t denote the instantaneous default-free interest rate. We assume that r_t has a linear specification given by

$$r_t = \alpha_0 + \alpha_1 g_t + \alpha_2 \pi_t, \tag{3.1}$$

where g_t and π_t are real activity growth and inflation respectively. This specification of the short rate is motivated by the monetary policy rule proposed by Taylor (1993). A large literature in macroeconomics has modeled real activity growth and inflation using long-run and short-run components (see for example Cogley and Sargent, 2001, 2005; Erceg and Levin, 2003). Following this literature, we specify the dynamics of real activity growth and inflation as follows

$$\begin{aligned} x_{t+1} &= \mu_{t+1} + \rho x_t + \Sigma e_{t+1}, \\ \mu_{t+1} &= \theta + \phi \mu_t + \Omega v_{t+1}, \end{aligned} \tag{3.2}$$

where $x_t = [g_t,\ \pi_t]'$, $\mu_t = [\mu_{gt},\ \mu_{\pi t}]'$ and $\theta = [\theta_g,\ \theta_\pi]'$ are 2×1 vectors, ρ, ϕ, Σ, and Ω are 2×2 diagonal matrices, μ_{gt} and $\mu_{\pi t}$ are the time-varying stochastic intercepts of the short-run components of real activity growth and inflation respectively, and $e_t = [e_{gt},\ e_{\pi t}]'$ and $v_t = [v_{gt},\ v_{\pi t}]'$ are the 2×1 vectors of short-run and long-run residual shocks, where the subscript g is for real activity growth and subscript π is for inflation. Both the short-run and long-run shocks are assumed to be distributed $N(0,\ I)$, where I is a 2×2 identity matrix. Given the dynamics of the covariates specified in equation (3.2), the long-run components of real activity growth and inflation are given by

$$
\begin{aligned}
g_t^* &= \frac{\mu_{gt}}{1-\rho_g}, \\
\pi_t^* &= \frac{\mu_{\pi t}}{1-\rho_\pi},
\end{aligned}
\tag{3.3}
$$

where g_t^* and π_t^* are the long-run components of real activity growth and inflation respectively, and ρ_g and ρ_π are their respective persistences, the diagonal elements of the matrix ρ.

The above representation of the dynamics of the state variables can be restated in matrix notation. We have four covariates, g_t, π_t, μ_{gt} and $\mu_{\pi t}$ in the model, and we can define the factors, the shocks and the variance matrix SS' as

$$
F_t = \begin{pmatrix} g_t \\ \pi_t \\ \mu_{gt} \\ \mu_{\pi t} \end{pmatrix}, \quad Z_t = \begin{pmatrix} e_{gt} \\ e_{\pi t} \\ v_{gt} \\ v_{\pi t} \end{pmatrix}, \quad \text{and } S = \begin{pmatrix} \Sigma & \Omega \\ 0_{2,2} & \Omega \end{pmatrix}.
$$

Therefore, the dynamic of the state variables can be represented as follows

$$
F_{t+1} = \psi + KF_t + SZ_{t+1}, \tag{3.4}
$$

where

$$
K = \begin{pmatrix} \rho & \phi \\ 0_{2,2} & \phi \end{pmatrix},
$$

and

$$
\psi = \begin{pmatrix} \theta_g & \theta_\pi & \theta_g & \theta_\pi \end{pmatrix}'.
$$

So far, we have discussed the dynamics of the covariates under the physical measure. Now we specify the market prices of risk and the pricing kernel to price the default-free bonds.

To change from the physical to risk-neutral measure, we specify the pricing kernel to take the form

$$m_{t+1} = \exp\left(-r_t - \frac{1}{2}\lambda_t'\lambda_t - \lambda_t' Z_{t+1}\right), \tag{3.5}$$

where λ_t is a 4×1 vector. We follow the term structure literature and assume time-varying prices of risk that evolve with the covariates:

$$\lambda_t = \lambda_0 + \lambda_1 F_t, \tag{3.6}$$

where λ_0 is a 4×1 vector and λ_1 is a 4×4 matrix, which can be expressed in the following matrix form

$$\lambda_0 = \begin{pmatrix} \lambda_{01} \\ \lambda_{02} \end{pmatrix}, \text{ and } \lambda_1 = \begin{pmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \lambda_{22} \end{pmatrix}$$

where λ_{01} and λ_{02} are 2×1 vectors, and λ_{11}, λ_{12}, λ_{21} and λ_{22} are 2×2 diagonal matrices. The dynamics of the covariates under the risk-neutral measure can therefore be written as

$$\begin{aligned} x_{t+1} &= \mu_{t+1} + \rho^Q x_t + \Sigma e_{t+1}, \\ \mu_{t+1} &= \theta^Q + \phi^Q \mu_t + \Omega v_{t+1}, \end{aligned} \tag{3.7}$$

where ρ^Q, θ^Q, and ϕ^Q are given by

$$\begin{aligned} \rho^Q &= \rho - \Sigma\lambda_{11}, \\ \theta^Q &= \theta - \Omega\lambda_{02}, \\ \phi^Q &= \phi - \Omega\lambda_{22}. \end{aligned} \tag{3.8}$$

The dynamics under the risk-neutral measure can also be expressed in matrix form

$$F_{t+1} = \psi^Q + K^Q F_t + S Z_{t+1}, \tag{3.9}$$

where ψ^Q and K^Q are given by

$$\begin{aligned} \psi^Q &= \psi - S\lambda_0, \\ K^Q &= K - S\lambda_1, \end{aligned} \tag{3.10}$$

and the elements of λ_0 and λ_1 need to satisfy the following restrictions

$$\begin{aligned} \lambda_{01} &= 0_{2,1}, \\ \lambda_{12} &= \lambda_{21} = 0_{2,2}. \end{aligned}$$

The price of a default-free zero coupon bond at time t that matures in h periods is given by

$$P(t, t+h) = E_t^Q \left[\exp \left(-\sum_{j=0}^{h-1} r_{t+j} \right) \right],$$

where E_t^Q is the expectation under the risk-neutral measure. The short rate process (3.1) can also be written as a function of the four covariates

$$r_t = \delta_0 + \delta_1 F_t, \tag{3.11}$$

where $\delta_0 = \alpha_0$ and $\delta_1 = [\alpha_1, \alpha_2, 0, 0]$. Given the dynamics in equations (3.11) and (3.9), the price of a zero coupon bond can be written as

$$P(t, t+h) = \exp(A_h + B_h' F_t),$$

where A_h and B_h satisfy the following recursive relations

$$A_h = -\delta_0 + A_{h-1} + B_{h-1}^{'} \psi^Q + \frac{1}{2} B_{h-1}^{'} S S^{'} B_{h-1}, \tag{3.12}$$

$$B_h = -\delta_1 + B'_{h-1}K^Q, \quad (3.13)$$

where $A_1 = -\delta_0$, and $B_1 = -\delta'_1$. The derivation of the recursive relations is provided in Appendix A. The continuously compounded yield $y(t, t+h)$ for a zero coupon bond that matures in h periods is given by

$$y(t, t+h) = -\frac{logP(t, t+h)}{h} = -\frac{A_h + B'_h F_t}{h}. \quad (3.14)$$

Yields are affine functions of the covariates and their long-run components, which allows us to study the impact of the shocks to the long-run components on the term-structure of interest rates.

3.3 Data and Estimation Method

3.3.1 Data

We use monthly data on continuously compounded zero-coupon bond yields with maturities of three and six months, and one, two, three, four, five, ten and twenty years, for the period April 1953 to December 2012. The three- and six-months yields are obtained from the Fama CRSP Treasury Bill files, and the one- to five-year bond yields are obtained from the Fama CRSP zero coupon files. The ten- and twenty-year maturity zero-coupon yields are obtained from the H.15 data release of the Federal Reserve Board of Governors.[5] Table 1 shows that the yields have mild excess kurtosis and positive skewness at all maturities. On average the yield curve is upward sloping, and the volatility of yields is relatively smaller for longer maturities. The yields for all maturities are highly persistent with long-term yields slightly more autocorrelated than short-term yields.

[5] We include these long maturities to facilitate the identification of the long-run components. The Federal Reserve database provides constant maturity treasury (CMT) rates for different horizons, which are converted into ten- and twenty-year maturity zero-coupon yields using the piecewise cubic polynomial.

The macroeconomic data is obtained from the Federal Reserve Economic Data. We use the Consumer Price Index for all urban consumers to calculate inflation. Following Ang and Piazzesi (2003), we extract the first principal component from a group of monthly variables that capture real activity growth: unemployment, the growth rate of payroll employment, the growth rate of employment and the growth rate of industrial production.[6] All growth rates, including inflation, are measured as the difference in logs of the index in month t and $t-12$. Table 1 indicates that both macro variables are highly persistent, but inflation has much larger autocorrelations than real activity growth at long lags.

3.3.2 Estimation method

The long-run components of real activity growth and inflation in our model are latent. We use the Kalman filter to filter the latent long-run components (g_t^* and π_t^*) in the estimation. In our setup, the state propagation equation (3.2) is Gaussian. The measurement equation in the state-space system is denoted by

$$Y_t = G(F_t) + \varepsilon_t, \tag{3.15}$$

where Y_t is the vector of observables, and F_t is a vector of latent state variables and the observable covariates. We assume an additive normally distributed error vector, ε_t, and we assume that the covariance matrix $E[\varepsilon_t \varepsilon_t^T]$ is diagonal.

To estimate the model, we assume that all yields and macro variables are measured with errors. The measurement equation (3.15) thus includes the bond yields with different maturities and the macro variables, real activity growth g_t and inflation π_t, which are also linear in the state variables. This means that Y_t is a vector of dimension 11. This approach should allow for improved identification of the long-run components of the macro variables,

[6] Alternatively one could use a measure of the capacity utilization rate, as in Rudebusch and Wu (2008) and Kozicki and Tinsley (2001b). However, the capacity utilization rate is subject to substantial revisions and covers a small and decreasing share of the real economy.

which are typically estimated using macroeconomic data only.

Following Duffee and Stanton (2012), we apply the Kalman filter to the state-space representation of the model and we estimate the model parameters and filter the long-run components of the macro variables via maximum likelihood.[7] The likelihood function is derived in Appendix B.

3.4 Parameter Estimates and Model Fit

3.4.1 Parameter Estimates

Table 2 presents parameter estimates for two models estimated using the full sample from April 1953 to December 2012.[8] Standard errors are computed using the outer product of the gradient, and all parameter estimates are statistically significant for both models. Panel A of Table 2 presents the parameter estimates for the model with long-run and short-run components for inflation and real activity growth dynamics. Henceforth we refer to this model as the component model. Panel B of Table 2 is a special case of the model presented in equation (3.2), where we shut off the stochastic long-run components by setting ϕ and Ω equal to zero. Henceforth we refer to this model as the benchmark macro model. For both specifications, the short rate is given by a version of Taylor's (1993) rule (3.1), where the monetary authority sets the short rate as a function of real activity growth and inflation. Both panels report the loadings of the short rate on real activity growth and inflation. The panels also report the dynamics of the covariates under the physical measure and the corresponding market prices of risk.

Table 2 shows that the loadings on real activity growth and inflation are similar for both

[7]Other available estimation strategies for these term structure models include the exact inversion likelihood appraoch of Chen and Scott (1993), and Pearson and Sun (1994), closed-form approximate likelihood (Aït-Sahalia and Kimmel, 2010), simulated maximum likelihood (Brandt and He, 2002), and Bayesian Markov Chain Monte Carlo (MCMC, Collin-Dufresne, Goldstein, and Jones, 2008).

[8]Rather than simply reporting the P and Q parameters, for convenience we also report the prices of risk implied by the mapping between the P and Q parameters.

specifications. The loading on real activity growth is smaller than the loading on inflation, which means inflation has more weight in the policy rule because the two variables are roughly of the same order of magnitude. Table 2 also reports the dynamics of the covariates for both specifications. Recall that we set the off-diagonal elements of ρ, ϕ, Σ, and Ω in equation (3.2) equal to zero. We therefore estimate and report only the diagonal elements of these matrices. Panel B shows that both the real activity growth and inflation are highly persistent under the physical measure in the benchmark macro model. When we allow for long-run components, the long-run inflation component captures the high persistence of inflation and the persistence of the short-run inflation component is substantially lower. Both the long-run and the short-run components of real activity growth are highly persistent.

The estimates of the risk parameters λ_0 and λ_1 imply that all risk factors are priced in both the component model and the benchmark macro model. In the component model, the estimated market prices of risk for the unconditional mean of the long-run components are not significant. However, the long- and short-run macro risks both have statistically significant effects on the premia. In the benchmark macro model, the unconditional mean of the risk prices for the macro variables are statistically significant. In this model, the macro risks have negative effects on the risk prices of the macro variables.

Figure 2 plots the time-varying market price of risk λ_t for the two specifications, Panel A for the component model and Panel B for the benchmark macro model. In the benchmark macro model, the price of risk associated with real activity growth is mostly positive and countercyclical, and the price of risk associated with inflation is mostly negative and procyclical. In the component model, the price of risk for the long-run and short-run components of these variables often have opposite signs. The price for the long-run (short-run) component of real activity growth is pro- (counter-) cyclical, while the price of the long-run (short-run) component of inflation is counter- (pro-) cyclical.

3.4.2 In-Sample Model Fit

Our main objective is to study the impact of long-run shocks to inflation and real activity growth on the term structure of interest rates, but it is important that the model adequately captures the stylized facts in the data. We therefore report the fit of our model and compare its performance with that of several other specifications widely used in the literature.

Table 3 compares the in-sample fit of six different models. Model 1 is our newly proposed component model. Model 2 is the benchmark macro model. Model 1* is similar to the component model, but we restrict the persistence of the long-run components to be equal to one. This kind of specification for the macro variables is widely used in the macroeconomic literature (Cogley and Sargent, 2001, 2005; Cogley, Primiceri, and Sargent, 2010; Stock and Watson, 2007). Model 3 is a specification with three latent factors, which is the standard specification used in the literature to model default-free interest rates. To ensure comparability with the component model, we do not allow for correlation between the factors. Model 4 is similar to the model in Ang and Piazzesi (2003), with two macro variables and three latent factors. Model 4* is identical to model 4 except that the model is estimated in two steps. The first step involves estimating the model with macro variables only. In the second step, we fix the parameters estimated in the first step and estimate only the parameters associated with the latent factors. This is similar to the implementation in Ang and Piazzesi (2003). For all models with latent factors, we use the Kalman filter together with maximum likelihood to estimate the models. We recursively estimate the model each month using the previous 60 months of data. The in-sample root mean squared errors (RMSEs) in Table 3 are computed as the average RMSEs across all recursive samples. We report the in-sample RMSEs based on recursive estimation to allow comparison with the out-of-sample RMSEs, which we discuss in Section 3.8. Note that other tables and figures are based on the parameters in Table 2 rather than the average of the recursively estimated parameters.

Table 3 presents the in-sample RMSEs for all six specifications. The component model performs substantially better than the benchmark macro model, especially for long-maturity yields. Figure 1 graphs the model-implied yields for the component model together with the benchmark macro model, as well as the data. The model-implied 3-month and 6-month yields do not differ much across the two models. However, for longer-maturity yields, the component model clearly provides better in-sample fit.

Remarkably, our newly proposed component model outperforms all other specifications at the long end of the yield curve (10-year and 20-year yields), even the models with latent factors. This suggests that adding time-varying means of macro variables may not just be useful to provide more economic intuition, it may also resolve some problems of standard models to fit the long end of the yield curve. Intuitively, movements at the long end of the yield curve are related to learning about monetary policy target changes, and presumably this is what the long-run components in our model manage to capture.

In Table 4, we report autocorrelations for the actual data yields (repeated from Table 1 for convenience) and model-implied yields. Panel A shows the autocorrelations of the data yields, Panel B presents the implied autocorrelations for the component model, Panel C for the benchmark macro model, and Panel D for the model with three latent factors. The component model captures the autocorrelation patterns in the data much better than the benchmark macro model for the long-maturity yields. The implied estimates from the component model are similar to the estimates from the three-factor latent model. In summary, the proposed component model adequately captures the persistence of the yields in the data, especially for long maturity yields.

The model with unit root long-run components is of interest, because it is parsimonious. Table 3 indicates that the performance of the model with unit root long-run components is comparable to that of the model with mean-reverting long-run components.

3.5 Analyzing the Long-Run Components

In this section, we first examine the time-series properties of the filtered long-run components for inflation and real activity growth. Subsequently we document the model-implied short rate and its long-run component, as well as the model-implied long-run and short-run components of the term structure of yields.

3.5.1 The Long-Run Components of the Macro Variables

Figure 3 plots the time series of the model-implied and observed inflation and real activity growth. In both cases, the component model performs well in capturing the time-variation in the macro variables. In the case of real activity growth, the fitting errors are very small and the time series for the data and the model are hard to distinguish. Joslin, Le, and Singleton (2013) find that the use of macro variables in a Kalman filter setup results in improved fitting of yields at the cost of the fit of the macro variables. We instead obtain good fit for both yields and macro variables in the context of the component model. We show below that the implied long-run components of the macro variables are intuitively plausible and consistent with other proxies of long-run components. This suggests that in our specification, using information from both yields and macro variables to filter the latent long-run components results in improved estimates of the long-run components and the responses of the yields to macro shocks.

Figure 3 also plots the long-run components of the macro variables estimated from the component model. The long-run components are obtained as the time-varying expectation of the macro variables, as specified in equation (3.3). Panel A shows that the long-run component of inflation moves closely with the inflation series. There is a small time lag between the two series, with inflation leading the long-run component.

Inflation was low and steady in the early 1960s, it began rising in the mid-1960s, and it attained twin peaks around the time of the 1970s oil shocks. It fell sharply during the Volcker disinflation, and then settled down around two percent after the mid-1990s. It dropped slightly during the financial crisis, but increased back to around two percent after 2010. The long-run component of inflation captures the time-varying expectations of economic agents in the midst of these policy changes, and we can interpret the difference between inflation and the long-run component of inflation as the forecast errors of the agents. These forecast errors are quite significant from the late 1970s to the mid-1980s. These positive errors reflect substantial doubts about whether the Federal Reserve would continue to pursue a disinflationary policy. During the Volcker disinflation period, policy tightening began in October 1979, the expected inflation started to decrease slightly. But the actual inflation continued to rise over the next few months. As a result, the market lost confidence in the Fed's commitment to its policy stance, because the Fed had shown a high degree of tolerance for rising inflation in the 1970s. Following continued Federal Reserve tightening in the 1980s however, inflation and expected inflation began to fall, suggesting the market's improved confidence in the Fed's policy stance. Nevertheless, it is apparent that realized inflation initially fell much more rapidly than inflation expectations.

The importance of the Fed's credibility during this period has been discussed by Goodfriend (1993) and Erceg and Levin (2003). Goodfriend (1993) emphasizes the slow convergence of long-term nominal interest rates during the Volcker disinflation period. He interprets the temporary spike in nominal interest rates during 1980s as reflecting a continued lack of faith on the part of market participants in the Fed's willingness to maintain a tight hold on inflation.

The component term structure model successfully captures this credibility concern. Figure 4 plots the long-run component of inflation together with the model yields and the observed yields for various maturities. The long-run component of inflation moves closely with yields. As the correlation of the two series increases, especially during the Volcker disinflation period, the model-implied yields get closer to the actual yields. This is because the long-run component of inflation manages to capture the spike in long-term yields during the Volcker disinflation period.

The long-run component of inflation is also related to widely used proxies for agents' inflation expectations. Figure 5 plots the filtered long-run component of inflation together with different proxies for long-run inflation expectations. Following the adaptive learning literature (Ljungqvist and Sargent, 2004), Panel A constructs a proxy for agents' long-run inflation expectations using a ten-year discounted moving average of past CPI inflation. The long-run component is highly correlated with inflation expectations, with a correlation coefficient of 88%. Panel B uses the median ten-year CPI inflation expectations from the Livingston survey and the Blue Chip Economic Indicators survey as a proxy for long-run inflation expectations.[9] This graph also shows that the filtered long-run component captures agents' inflation expectations. Note that the sample periods in Panels A and B are different due to data availability.

In Figure 3, the small time lag between the long-run component of inflation and the inflation data indicates that a high expected inflation today is due to the fact that inflation was high in the recent past. These findings are consistent with studies that have applied the shifting endpoints approach to the term structure of interest rates. Kozicki and Tinsley (2001a, 2001b) find that the source of endpoint shifts is due to agents' long-run inflation expectations. They also show that the endpoint shifts lag changes in actual inflation, which is consistent with our finding.

Panel B of Figure 3 presents the long-run component of real activity growth together with overall real activity growth. The long-run component of real activity growth is less volatile than overall real activity growth. In our implementation, we noticed that the long-run component of inflation is more robustly identified than the long-run component of real

[9]The time series of the ten-year CPI inflation expectations data are obtained from the Philadelphia Fed. These data are for the period 1979:12 to 2012:12 and combine the Philadelphia Fed's Livingston Survey with the surveys from Blue Chip Economic Indicators. The forecast is for the average rate of CPI inflation for the next ten years. The Livingston survey data are available for June and December. The long-term inflation forecasts from Blue Chip Economic Indicators are available for March and October. To obtain monthly data, we interpolate the semi-annual observations. Survey forecasts are widely adopted as proxy for inflation expectations. See for instance Duffee (2014), Ang, Bekaert, and Wei (2007), and Chernov and Mueller (2012).

activity growth. There is some evidence in the literature that it is difficult to identify the long run component of real activity growth, but this evidence is mostly about the level of real activity rather than the growth rate.[10]

Overall, the long-run components of inflation and real activity growth filtered using our model are intuitively plausible and consistent with alternative proxies of long-run components and existing results. This suggests that our model provides a useful framework to study the impact of the macro variables on the term structure of interest rates.

3.5.2 The Long-Run Components of the Short Rate and Yields

Figure 6 shows the time series of the short rate, together with its long-run component implied by the component model. The short rate is a linear function of the macro variables, as specified in equation (3.1). Given the dynamics of the macro variables as shown in equation (3.2), the long-run component of short rate can be obtained as a linear function of the long-run components of inflation and real activity growth.

Figure 6 shows that the long-run component of the short rate is less volatile than the overall short rate. The implied short rate moves closely with the short end of the yield curve, as shown in Panel A of Figure 1. The correlation with the three-month yield is 90%. The model-implied short rate displays several spikes during the 1970s, drops sharply during the Volcker disinflation period, and then settles down around four percent after the mid-1990s. It briefly becomes negative during the financial crisis, but increases back to around two percent after 2010. The long-run component of the short rate never becomes negative.

Figure 7 plots the data and the model-implied long-run and short-run yield components for various maturities. Appendix C provides the derivation of the long-run and short-run yield components. The implied long-run components of medium- and long-maturity yields

[10] To extract trend components, the macroeconomics literature uses unobserved components models (Harvey, 1985, Watson, 1986, and Clark, 1987), Beveridge-Nelson (1981) decompositions, as well as the Hodrick and Prescott (1997) filter. Morley, Nelson, and Zivot (2003) find that some of the differences in extracting trends from the level of real activity are due to different correlation assumptions made by the unobserved components and Beveridge-Nelson methods.

are more volatile than the long-run component of the short-maturity yields. Most of the variation in the medium- and long-maturity yields comes from the long-run component. The short-run component accounts for most of the variation at the short end of the yield curve. The implied short-run component is much less volatile at the intermediate and long end than at the short end of the yield curve.

3.6 Impulse Response Functions and Variance Decompositions

In this section, we discuss the impact of the shocks to inflation and real activity growth on the yield curve. Appendix D discusses the computation of the impulse response function for different horizons and maturities. We first describe the initial responses of the term structure of interest rates to shocks to inflation and real activity growth. Subsequently we describe the propagation of the shocks for different horizons and maturities. Finally, we also investigate the relative contributions of each of the factors to the forecast variance. Appendix E presents the technical details of the variance decomposition.

3.6.1 Initial Responses

Figure 8 plots the initial response of the yield curve to shocks to inflation and real activity growth. Panels A and B present the initial responses using the component model. Panel C presents the initial responses using the benchmark macro model. In all panels, the solid line represents real activity growth and the dotted line represents inflation. Maturities are on the x-axis. All response coefficients are scaled to correspond to movements of one standard deviation in the factors, and are annualized and expressed in percentages by multiplying by 1200.

Panel A of Figure 8 presents the initial responses of the term structure of interest rates to short-run shocks to inflation and real activity growth. It shows that the impact of short-run shocks to real activity growth and inflation decreases with maturity. The impact of the short-run shock to real activity growth is less than half the impact of the short-run shock to inflation for shorter maturity yields. For longer maturity yields, the impact of a short-run shock to real activity growth is larger than the impact of a short-run shock to inflation.

Panel B presents the initial responses of the term structure of interest rates to long-run shocks to inflation and real activity growth. Long-run shocks to inflation have a much larger effect on the yield curve than long-run shocks to real activity growth. The impact of the long-run shock to real activity growth increases with maturities up to 10 years and then decreases slightly. The impact of the long-run shock to inflation also increases with maturities up to four years and then decreases slightly for longer maturities. Nevertheless, the impact of a long-run inflation shock is larger for longer maturity yields than for shorter maturity yields. The results are consistent with the extant literature, which shows that long-maturity yields contain information about future inflation.[11]

Panel C shows the effects of shocks to real activity growth and inflation using the benchmark macro model. The impact of shocks to inflation is larger than the impact of shocks to real activity growth for all maturities. The impact of shocks to real activity growth and inflation is larger for shorter maturity yields. A comparison of Panel C with the evidence in Panels A and B clearly shows how the benchmark model is restricted, because it only has one factor to capture the very different structure of shocks to long-term and short-term components as a function of maturity evident from Panels A and B.

3.6.2 Shock Propagation Over Time

We now investigate how shocks to macro variables propagate over time for various maturities. Figure 9 presents the evidence. Figure 9a shows the impulse response functions for different

[11] Mishkin (1990) shows that the term structure of nominal interest rates provides information about the future path of inflation for longer maturities of nine and twelve months. Fama (1990) finds that the term structure contains information about the future path of inflation at even longer maturities of 5 years. Gürkaynak, Sack, and Swanson (2005) also find that the movements in long rates reflect fluctuations in inflation perceptions. Hőrdahl, Tristani, and Vestin (2006) show that changes in the perceived inflation target tend to have a stronger impact on longer term yields based on German data.

maturities for the component model. The panels on the left provide the impulse responses for shocks to the short-run components of real activity growth and inflation. The panels on the right provide the impulse responses for shocks to the long-run components of real activity growth and inflation. Figure 9b shows the impulse response function for different maturity yields using the benchmark macro model. The x-axis represents months since the initial shock in both figures. The solid lines are for real activity growth and the dotted lines are for inflation. Note that the initial responses for each maturity are identical to the ones reported in Figure 8.

The left-side panels in Figure 9a show that the short-run shocks to inflation and real activity growth have a large initial impact on the yield curve, but the impact of the short-run component quickly dies off. The impact of the short-run shock to inflation dies off even quicker than the impact of the short-run shock to real activity growth, for all maturities. For example, for three-month yields, the impact of the short-run shock to inflation starts and peaks at 3.89% and drops to 0.01% after one year. The impact of a short-run shock to real activity growth starts and peaks at 1.79% and drops to 0.01% after more than ten years. This is partly due to the fact that the short-run component of inflation has lower persistence than the short-run component of real activity growth. The impact of a short-run shock to real activity growth is larger than that of a short-run shock to inflation for all horizons except the very short ones.

The right-side panels of Figure 9a show that the impulse responses for the long-run shocks to inflation are larger than those for the long-run shocks to real activity growth, for all maturities. The responses of short and intermediate maturity yields to the long-run shocks to real activity growth are hump-shaped with respect to the forecast horizon. The initial responses gradually increase, and achieve maximums around the 5-year forecast horizon. The long-run components of both macro variables are highly persistent. This high persistence of the long-run components results in much slower decay of the responses to long-run shocks. The right-side panels in Figure 9a show that the responses to long-run shocks of real activity growth and inflation last for much longer horizons for all maturities.

Figure 9b present the impulse responses for shocks to inflation and real activity growth using the benchmark macro model, i.e. this model has only one shock for each macro variable. The impulse responses for inflation shocks are much larger than those for real activity growth shocks, for all forecast horizons. This result is consistent with Ang and Piazzesi (2003), who show that inflation shocks have much larger impact across the yield curve. The response of the 3-month yield to a one-standard deviation inflation shock peaks at 3.72% and the response of the same yield to a one-standard deviation real activity growth shock peaks at 1.48%.

The response of the yield curve to a one-standard deviation inflation shock remains at least 1% even after 20 years. These findings are consistent with existing theoretical frameworks in the macroeconomics literature (Granger and Joyeux, 1980; Backus and Zin, 1993) which explain long memory in inflation as the result of aggregation across agents with heterogeneous beliefs. The impact of a shock to real activity growth also slowly dies off. The response of the yield curve to a one-standard deviation real activity growth shock remains at least 0.04% after 10 years. Once again, the impulse responses for the one-shock model can be seen to reflect a compromise between the very different persistence of the long-term and short-term components evident from Figure 9a.

The responses to macro shocks decrease as maturity increases in both models. This finding confirms the existing literature which documents that monetary policy shocks affect short-maturity rates more than long-maturity rates (Cochrane, 1989; Evans and Marshall, 1998; Kuttner, 2001; Piazzesi, 2005). Macroeconomic shocks induce changes in real activity growth and inflation, and systematic monetary policy adjusts the funds rate according to the Taylor rule.

3.6.3 Variance Decompositions

3.6.3.1 The Benchmark Model

Panel A of Table 5 presents the variance decomposition for the benchmark macro model. The table presents the variance decomposition for different maturity yields and various forecast horizons. The proportion of the variance accounted for by real activity growth is decreasing with yield maturity and forecast horizon. The proportion of the variance accounted for by inflation is increasing with yield maturity and forecast horizon.

In this model, inflation explains a large proportion of total variance for all maturities and forecast horizons. To illustrate this, the largest contribution of real activity growth is 13.62%, for the 3-month yield and the 1-month ahead forecast. This means 13.62% of the one-step ahead forecast variance of the 3-month yield is explained by real activity growth and 86.38% by inflation. This is consistent with the impulse response functions, which show that the responses of the yield curve from inflation shocks are much larger than that from the real activity growth shocks. Ang and Piazzesi (2003) also find that inflation contributes more to the forecast variance than real activity growth at all points of the yield curve and for all forecast horizons.

3.6.3.2 The Component Model

Panel B of Table 5 presents the variance decomposition for the component model for different maturity yields at various forecast horizons. We decompose the total variance for each maturity yield and forecast horizon into the proportion explained by the short-run and long-run components of real activity growth and inflation, but we also report the sum of the short- and long-run proportions of each macro variable, which can be compared to the numbers obtained from the benchmark macro model.

Consistent with its intuitive interpretation, the short-run component of inflation explains a large proportion of the forecast variance for short forecast horizons. This explanatory power is much larger for shorter maturity yields and declines for longer maturity yields regardless of the forecast horizons. The long-run component of inflation explains a much larger proportion of forecast variance for long-maturity yields. Moreover, the explanatory power of the long-run component of inflation increases as the forecast horizon increases for all maturity yields, because the short-run component of inflation mean-reverts at a faster rate than the long-run component. As discussed earlier, these findings are consistent with the extant macroeconomic literature which demonstrates the relationship between long-term yields and the future path of inflation. The long-run component of inflation seems to capture agents' evolving beliefs about the future path of inflation, and therefore it performs better for long forecast horizons.

Panel B of Table 5 also shows that for short maturities, the explanatory power of the short-run component of real activity growth is a hump-shaped function of the forecast horizon. For medium- and long-term yields, the short-run component of real activity growth explains a larger proportion of the short horizon forecast variance. Finally, the explanatory power of the long-run component of real activity growth increases with maturity. Moreover, the explanatory power of the long-run component of real activity growth increases as the forecast horizon increases for short- and medium- term yields. For long maturity yields, the explanatory power of the long-run component of real activity growth is a slightly hump-shaped function with respect to the forecast horizon.

3.6.3.3 Comparing the Models

The last two columns in Panel B of Table 5 contain the total contribution of real activity growth and inflation to the forecast variance for different forecast horizons and yield maturities. A comparison of these numbers with those in Panel A highlights substantial differences between the two models. Inflation is still the most important determinant of variation in the term structure, but in the component model the long-run and short-run components of real activity growth account for a substantial part of the forecast variance for short-maturity yields at longer forecast horizons, and for long-maturity yields at shorter forecast horizons.

3.7 Term Premia and the Predictability of Bond Returns

In this section, we examine if the component model can generate plausible term premia and short rate expectations. We also examine the predictability of bond excess returns using the components filtered from the component model.

3.7.1 Term Premia and Short Rate Expectations

The dynamics of term premia greatly influence the monetary policy transmission mechanism. The central bank controls short-term interest rates, but the entire term structure of interest rates is relevant for households' spending decisions. Therefore, the expectations about future policy rates and the time variation of term premia, and their impact on the yield curve, are of paramount importance for both researchers and policy makers (see among others Cochrane and Piazzesi, 2008; Wright, 2011; Joslin, Priebsch, and Singleton 2014; Bauer, Rudebusch and Wu 2012; Bauer, 2016). In this section, we discuss short-rate expectations and term premia implied by the component model and the benchmark macro model.

The estimation of term premia amounts to the estimation of expectations of future short-term interest rates. We focus on the forward term premium $FTP_t^{h_1,h_2}$, which is defined as the difference between the forward rate $F_t^{h_1,h_2}$ that one can lock in at time t for a (h_2-h_1)-period loan starting in h_1 periods, and the expected yield on a (h_2-h_1)-period bond purchased h_1 periods from now.

$$FTP_t^{h_1,h_2} = F_t^{h_1,h_2} - \underbrace{E_t \sum_{i=h_1}^{h_2-1} y(t+i, t+i+1)}_{\text{Expectation (Risk-Neutral Yield)}} . \tag{3.16}$$

Appendix F provides additional details on the construction of these model-implied term premia. We illustrate our results using five- to ten-year forward term premia ($h_2 = 10$ years and $h_1 = 5$ years).

Panel A of Figure 10 plots the forward rate from the data together with the estimates from the benchmark macro model and the component model. The component model provides a substantially better fit of forward rates than the benchmark macro model, especially during the 1980s.

Panel B compares the time series of model-implied term premia from the component model and the benchmark macro model. The most striking difference is that the component model implies larger and less volatile term premia than the benchmark macro model. The differences between the term premia from the two models are relatively large during the 1970s, the conundrum period, and the financial crisis.[12] In particular, the term premia in the benchmark macro model decline during the 1970s. Several existing studies have argued that term premia are slow-moving and countercylical (Campbell and Cochrane, 1999; Wachter, 2006; Cochrane and Piazzesi, 2005). Consistent with these studies, term premia from the component model are slow-moving (i.e., less variable) relative to the benchmark model. However, term premia are not unambiguously countercyclical in either model. While term premia are high during the second recession of the early 1980s in both models, the movements around other recessions are more muted for the component model. For the benchmark macro model, term premia exhibit a pronounced procyclical pattern during the recessions of the 1970s.

Panel C displays the sum of the expectations of future short term rates in equation (3.16), i.e., the risk-neutral yields. The risk-neutral yield from the component model is smaller than that of the benchmark macro model. Panels B and C suggest that before the 1980s, the component model attributes increases in forward rates to increasing term

[12] From June 2004 to December 2005, the Federal Open Market Committee thirteen times increased the target for the federal funds rate by 25 basis points. However, long-term interest rates actually declined during this period. Fed Chairman Alan Greenspan referred to this puzzling behavior of interest rates as a "conundrum" (Greenspan, 2005).

premia, while the benchmark model attributes them to increases in expectations of future short-term rates. After 1980, the benchmark macro model attributes more of the gradual decline in the forward rate to falling term premia, while the component model attributes it to a decline in expectations. The results from the component model are consistent with the sizable decreases in survey-based forecasts of inflation and policy rates over the last 20 years (Kim and Orphanides, 2012; Kozicki and Tinsley, 2001a; Wright, 2011).

In summary, our results suggest that term premia and expected future short rates in the component model are more plausible and consistent with the existing literature than the corresponding estimates in the benchmark model.

3.7.2 The Predictability of Bond Returns

The existing literature documents that bond excess returns can be predicted using variables such as forward spreads (Fama and Bliss, 1987), yield spreads (Campbell and Shiller, 1991; Keim and Stambaugh, 1986), forward rates (Cochrane and Piazzesi, 2005) and macroeconomic variables (Ludvigson and Ng, 2009; Cooper and Priestley, 2009; Greenwood and Vayanos, 2014).[13] In this section, we analyze return predictability using the risk factors extracted from the component macro model and several alternative models.

We run predictive regressions using the risk factors F_t for the period April 1953 to December 2012. The regression specification is similar to that of Cochrane and Piazzesi (2005):

$$rx_{t,t+12}^{h} = a^{h} + b^{h} F_t + \epsilon_t^{h}, \tag{3.17}$$

where $rx_{t,t+12}^{h}$ is the annual holding-period return, in excess of the one-year yield, on a bond

[13]Ludvigson and Ng (2009) show that linear and nonlinear combinations of principal components extracted from a large dataset of macro variables forecast future excess returns. Cooper and Priestley (2008) find that the output gap helps predict excess bond returns. Greenwood and Vayanos (2014) document that measures of Treasury bond supply help forecast yields and returns.

with maturity h,

$$rx^h_{t,t+12} = -(h-12)y(t+12,t+h) + hy(t,t+h) - y(t,t+12), \tag{3.18}$$

and ϵ^h_t is the prediction error. The parameters a^h and b^h are maturity-dependent. Note that under the expectation hypothesis $b^h = 0$: Bond risk premia are unpredictable and equal to a constant a^h. The component model implies that the risk factors are real activity growth, inflation and their long-run components, $F_t = [g_t,\, \pi_t, \mu_{gt}, \mu_{\pi t}]'$. For the benchmark macro model, $F_t = [g_t,\, \pi_t]'$. We also consider forecasting with the first three principal components of the yield curve PC_t, which correspond to the risk factors in a fully latent model.

The predictive power of the risk factors is measured using the regression's adjusted R^2. Table 6 reports the R^2 for bonds with maturities of two to five, ten and twenty years. The risk factors based on the benchmark model explain 4-6% of the one-year ahead variation in bond risk premia across the maturity spectrum. The explanatory power increases substantially after including the long-run components of the macro variables. The macro variables together with their long-run components explain between 16% and 22% of the variation of one-year ahead excess returns, with the largest explained proportion for two to five-year bonds.

The risk factors from the component model also outperform the first three principal components of the yield curve for two to five year maturity bonds. The R^2 associated with the macro variables and their long-run components exceeds the R^2 for the principal components by about 3 to 8 percent for two to five year maturity bonds. This finding is consistent with Cieslak and Povala (2015), who show that trend inflation measured by the discounted moving average of past CPI inflation has additional predictive content for bond returns relative to yield curve information. These findings suggest that the long-run components filtered from the information in the yield curve and macro variables may be capturing information that is also captured by models with hidden or unspanned risk factors (Duffee, 2011; Joslin, Priebsch, and Singleton, 2014).

3.8 Out-of-Sample Performance

In this section, we compare the out-of-sample performance of our component model with that of alternative specifications. Table 6 reports the results for seven specifications, including the benchmark macro model, the fully latent model with three latent factors, and the models with macro variables and three latent factors. These are the same models we use for the in-sample exercise. We also include the random walk model, which is referred to as Model 5.[14]

We compare the out-of-sample performance of these models by computing the forecast RMSEs for one-month, three-month, and six-month ahead yields for all nine yields used in estimation. Our procedure for examining out-of-sample forecasts is as follows. Each month t, we estimate each of the models using the previous sixty months of data. We then forecast the yields one month, three months, and six months ahead. Note that we also forecast the state variables, unlike some other studies with macro variables which use the subsequent realizations of these variables. We report the RMSE of the one-month ahead forecast in Panel A of Table 6, the RMSE of the three-month ahead forecast in Panel B, and the RMSE of the six-month ahead forecast in Panel C. The random walk model outperforms all other models for all maturities at all forecast horizons. This is not surprising. The literature has documented that it is difficult to beat a random walk when forecasting yields (see for example Orphanides and Wei, 2012; Van Dijk, Koopman, Van der Wel, and Wright, 2014). The random walk is therefore useful as a benchmark to rank and compare the underperformance of economic models.

The component model forecasts uniformly better than the benchmark macro model, especially for long-maturity yields. It outperforms all models except for the random walk at the long end of the yield curve (10-year and 20-year yields) for all forecast horizons.

These results are particularly interesting since the component model is a rather restrictive and parsimonious specification relative to some of the other models. Even if one interprets

[14] Appendix G considers the performance of other related models.

it as a model with latents, it includes only two latent factors compared to the three latent factors for some of the other specifications. Moreover, the latents in this model are of course restricted by their relation to inflation and real activity growth.

Finally, as the forecast horizon increases, the RMSEs increase, but the ranking of the models is similar and the conclusions remain the same. Overall, the main conclusion from the out-of-sample exercise is that the component model results in substantial improvements in out-of-sample forecasting performance at the intermediate and long end of the yield curve.

3.9 Conclusion

We study the impact of long-run and short-run shocks to inflation and real activity growth on the term structure of interest rates. We model the term structure of interest rates using a reduced form no-arbitrage model with macro variables. The macro variables in our specification follow autoregressive dynamics with time-varying stochastic long-run means.

We estimate our model using the Kalman filter to identify the latent long-run components of both macro variables. The model provides good in-sample and out-of-sample fit relative to several alternative specifications considered in the literature, especially for long-maturity yields. The improved performance of our model for long-maturity yields is mainly driven by the long-run component of inflation.

We show that the filtered long-run component of inflation is intuitively plausible and consistent with alternative proxies of long-run inflation from survey data. We investigate the impact of long-run and short-run shocks on yields over time and across maturities. The impact of long-run shocks is relatively small for short maturities, but shocks to long-run inflation greatly impact long-maturity yields, especially over long horizons.

The decomposition of the forecast variance reveals a complex pattern across forecast horizons and maturities, and illustrates the need to separate the long- and short-run components of the macro variables. Most importantly, in the benchmark macro model, real activity growth hardly matters for long maturities, and this is not the case in the component model, at least for short forecasting horizons.

Our model suggests that the decline in forward rates over the last two decades mainly results from lower expectations of future short rates, which is consistent with survey-based forecasts of inflation and policy rates. The corresponding forward term premia are slow-moving, which is intuitively plausible.

The filtered long-run components of macro variables help predict bond excess returns over and above the information in the current yield curve. These results suggest that the filtered long-run components help uncover information that is also captured by models with hidden or unspanned risk factors (Duffee, 2011; Joslin, Priebsch, and Singleton, 2014).

Overall, our results emphasize the importance of distinguishing between long-run and short-run components, and highlight the importance of the long-run component of inflation for modeling and predicting the yield curve.

Appendices

A The Default-Free Model

To derive the recursions in equations (3.12) and (3.13), we first note that the price of a one-period bond, $h = 1$, is as follows

$$\begin{aligned} P(t, t+1) &= E_t^Q[\exp(-r_t)] \\ &= \exp(-\delta_0 - \delta_1 F_t). \end{aligned} \tag{A.1}$$

Suppose that the price of a h-period bond is given by $P(t, t+h) = \exp(A_h + B_h^{'} F_t)$. Matching coefficients gives $A_1 = -\delta_0$ and $B_1 = -\delta_1^{'}$. In order to solve for A_h and B_h, we derive the bond price under the risk neutral probability measure

$$\begin{aligned} P(t, t+h) &= E_t^Q[\exp(-r_t) P(t+1, t+h)] \\ &= E_t^Q[\exp(-\delta_0 - \delta_1 F_t) \exp(A_{h-1} + B_{h-1}^{'} F_{t+1})] \\ &= \exp(-\delta_0 - \delta_1 F_t + A_{h-1}) E_t^Q[\exp(B_{h-1}^{'} F_{t+1})] \\ &= \exp(-\delta_0 - \delta_1 F_t + A_{h-1}) E_t^Q\{\exp[B_{h-1}^{'}(\psi^Q + K^Q F_t + S Z_{t+1})]\} \\ &= \exp(-\delta_0 - \delta_1 F_t + A_{h-1}) \exp[B_{h-1}^{'}(\psi^Q + K^Q F_t)] E_t^Q[\exp(B_{h-1}^{'} S Z_{t+1})] \\ &= \exp[(-\delta_0 - \delta_1 F_t + A_{h-1}) + B_{h-1}^{'}(\psi^Q + K^Q F_t)] \; \exp(\frac{1}{2} B_{h-1}^{'} S S^{'} B_{h-1}) \\ &= \exp[-\delta_0 - \delta_1 F_t + A_{h-1} + B_{h-1}^{'}(\psi^Q + K^Q F_t) + \frac{1}{2} B_{h-1}^{'} S S^{'} B_{h-1}]. \end{aligned} \tag{A.2}$$

Matching coefficients results in the recursive relations in equations (3.12) and (3.13).

B The Likelihood Function

We have data on zero coupon bond yields for nine different maturities, as well as the macro variables. y_t represents the term structure of yields at time t, a 9×1 vector. x_t represents the macro variables (real activity growth g_t and inflation π_t) at time t. $x_t = [g_t,\, \pi_t]'$ is a 2×1 vector. Stacking macro variables x_t and bond yields y_t, we have the vector of observables Y_t, which is 11×1. As shown in equation (3.15), the vector of observables Y_t is linear in the state variables F_t, where $F_t = [x_t^{'},\, \mu_t^{'}]^{'}$. The time-varying intercepts μ_t are the unobserved factors. As described in equation (3.14), the h-period bond yield at time t takes the following form

$$y(t, t+h) = -\frac{A_h + B_h' F_t}{h}.$$

We can rewrite it as follows

$$y(t, t+h) = \overline{A}_h + \overline{B}_h^{'} F_t = \overline{A}_h + \overline{B}_h^{1'} x_t + \overline{B}_h^{2'} \mu_t,$$

where $\overline{A}_h = -\frac{A_h}{h}$ and $\overline{B}_h^{'} = -\frac{B_h'}{h}$. $\overline{A}_h$ is a scalar, and $\overline{B}_h^{'}$ is a 1×4 vector. $\overline{B}_h^{'} = [\overline{B}_h^{1'},\, \overline{B}_h^{2'}]$. $\overline{B}_h^{1'}$ and $\overline{B}_h^{2'}$ are 1×2 vectors. The term structure of zero coupon bond yields y_t is given by

$$y_t = \overline{A} + \overline{B}^{'} F_t = \overline{A} + \overline{B}^{1'} x_t + \overline{B}^{2'} \mu_t, \tag{B.1}$$

where $\overline{A}$ is a 9×1 vector, and $\overline{B}^{'}$ is a 9×4 matrix. $B' = [\overline{B}^{1'},\, \overline{B}^{2'}]$. $\overline{B}^{1'}$ and $\overline{B}^{2'}$ are 9×2 matrices.

To estimate the model, we assume that all yields and macro variables are measured with errors ε_t, where ε_t is a 11×1 vector. We assume an additively normally distributed error vector, ε_t, and we assume that the unconditional covariance matrix $E[\varepsilon \varepsilon_t^T]$ is diagonal. The parameter vector is $\eta = (\psi,\, K,\, S,\, \delta_0,\, \delta_1,\, \lambda_0,\, \lambda_1)$. We apply the Kalman filter to the state-space representation of the model, estimating the model parameters η and filtering the latent stochastic means of the macro variables μ_t via maximum likelihood. Denoting the normal density functions of the state variables F_t and the errors ε_t as f_F and f_ε respectively, the joint likelihood $L(\eta)$ of the observed data is given by

$$L(\eta) = \prod_{t=2}^{T} f(Y_t|Y_{t-1})$$

$$\begin{aligned}
\log(L(\eta)) &= \sum_{t=2}^{T} - \log|\det(J)| + \log f_F(x_t, \mu_t | x_{t-1}, \mu_{t-1}) + \log f_\varepsilon(\varepsilon_t) \qquad \text{(B.2)}\\
&= -(T-1)\log|\det(J)| - (T-1)\frac{1}{2}\log(\det(SS^{'}))\\
&- \frac{1}{2}\sum_{t=2}^{T}(F_t - \psi - KF_{t-1})^{'}(SS^{'})^{-1}(F_t - \psi - KF_{t-1})\\
&- \frac{T-1}{2}\log\sum_{n=1}^{11}\omega_n^2 - \frac{1}{2}\sum_{t=2}^{T}\sum_{n=1}^{11}\frac{(\varepsilon_{t,n})^2}{\omega_n^2},
\end{aligned}$$

where ω_n^2 is the variance of the n^{th} measurement error $\varepsilon_{t,n}$, and the Jacobian term is given by

$$J = \begin{pmatrix} I_4 & 0_{4\times 11} \\ B_J & \omega \end{pmatrix},$$

where $B_J = \begin{pmatrix} \rho & \phi \\ \overline{B}^{1'} & \overline{B}^{2'} \end{pmatrix}$, and ω is a 11×11 diagonal matrix.

C Long-Run and Short-Run Components of Yields

As shown in Appendix B, the h-period bond yield at time t takes the following form

$$y(t, t+h) = \overline{A}_h + \overline{B}_h^{'} F_t = \overline{A}_h + \overline{B}_h^{1'} x_t + \overline{B}_h^{2'} \mu_t. \tag{C.1}$$

where $\overline{A}_h = -\frac{A_h}{h}$ and $\overline{B}_h^{'} = -\frac{B_h'}{h}$. $\overline{A}_h$ is a scalar, and $\overline{B}_h^{'}$ is a 1×4 vector. $\overline{B}_h^{'} = [\overline{B}_h^{1'}, \overline{B}_h^{2'}]$. $\overline{B}_h^{1'}$ and $\overline{B}_h^{2'}$ are 1×2 vectors. A_h is a scalar, and it satisfies the following recursive relation as specified in equation (3.12)

$$A_h = -\delta_0 + A_{h-1} + B_{h-1}^{'} \psi^Q + \frac{1}{2} B_{h-1}^{'} S S^{'} B_{h-1},$$

where $A_1 = -\delta_0$. Therefore $\overline{A}_h$ satisfies the following recursive relation

$$\overline{A}_h = -\frac{1}{h}(-\delta_0 + A_{h-1} + B_{h-1}^{'} \psi^Q + \frac{1}{2} B_{h-1}^{'} S S^{'} B_{h-1}),$$

where $\overline{A}_1 = \frac{\delta_0}{h}$. To separate the long-run and short-run components in $\overline{A}_h$, we rewrite it as follows

$$\begin{aligned} \overline{A}_h &= -\frac{1}{h}(-\delta_0 + A_{h-1} + B_{h-1}^{'} \psi^Q + \frac{1}{2} B_{h-1}^{1'} \Sigma \Sigma^{'} B_{h-1}^{1} + \frac{1}{2}(B_{h-1}^{1'} + B_{h-1}^{2'}) \Omega \Omega^{'} (B_{h-1}^{1} + B_{h-1}^{2})) \\ &= \delta_0 - \frac{1}{h}(\sum_{s=2}^{h} (B_{s-1}^{'} \psi^Q + \frac{1}{2}(B_{s-1}^{1'} + B_{s-1}^{2'}) \Omega \Omega^{'} (B_{s-1}^{1} + B_{s-1}^{2}) + \frac{1}{2} B_{s-1}^{1'} \Sigma \Sigma^{'} B_{s-1}^{1})) \text{for } h \geq 2, \end{aligned}$$

where $B_h^{'}$ is a 1×4 vector. $B_h^{'} = [B_h^{1'}, B_h^{2'}]$ where $B_h^{1'}$ and $B_h^{2'}$ are 1×2 vectors.

Based on the above expression of $\overline{A}_h$ and equation (C.1), the long-run (subscript lr) and short-run (subscript sr) components of the h-period bond yield at time t are as follows

$$y(t, t+h)_{lr} = \overline{B}_h^{2'} \mu_t + \delta_0 - \frac{1}{h}(\sum_{s=2}^{h} (B_{s-1}^{'} \psi^Q + \frac{1}{2}(B_{s-1}^{1'} + B_{s-1}^{2'}) \Omega \Omega^{'} (B_{s-1}^{1} + B_{s-1}^{2}))), \tag{C.2}$$

$$y(t,t+h)_{sr} = \overline{B}_h^{1'} x_t - \frac{1}{h}(\sum_{s=2}^{h}(\frac{1}{2}B_{s-1}^{1'}\Sigma\Sigma^{'}B_{s-1}^{1})). \tag{C.3}$$

D Impulse Response Functions

To derive the impulse responses of yields to short-run and long-run shocks to the macro variables, we rewrite the dynamics of the state variables under the risk-neutral measure in Wold $VMA(\infty)$ form

$$F_t = (I_4 - K^Q)^{-1}\psi^Q + \sum_{k=0}^{\infty} \Gamma_k S Z_{t-k}, \tag{D.1}$$

where I_4 is a 4×4 identity matrix, $\Gamma_0 = I_4$, and $\Gamma_k = (K^Q)^k$. We assume $\sum_{k=0}^{\infty} \Gamma_k \Gamma_k^{'}$ converges. Bond yields are given by equation (B.1). Substituting equation (D.1) into equation (B.1), and rewriting the bond yields in Wold $VMA(\infty)$ form we get

$$y_t = \overline{A} + \overline{B}^{'}[(I_4 - K^Q)^{-1}\psi^Q + \sum_{k=0}^{\infty} \Gamma_k S Z_{t-k}]. \tag{D.2}$$

Therefore, the impulse responses of bond yields to shocks to the state variables at horizon k can be presented as follows

$$\zeta_k = \overline{B}^{'} \Gamma_k S. \tag{D.3}$$

To separate the effects of the different components, we focus on the following expression of bond yields in equation (B.1)

$$y_t = \overline{A} + \overline{B}^{1'} x_t + \overline{B}^{2'} \mu_t.$$

The long-run components are given by equation (3.3). Substituting equation (3.3) into the above equation, and rewriting the bond yields as follows

$$y_t = \overline{A} + \overline{B}^{1'} x_t + \overline{B}^{2*'} x_t^*, \tag{D.4}$$

where $x_t^* = [g_t^*, \pi_t^*]^{'}$ is a 2×1 vector, $\overline{B}^{2*'} = \overline{B}^{2'}(I_2 - \rho)$, and I_2 is a 2×2 identity matrix.

The time-varying intercept can be represented in the following Wold $MA(\infty)$ form

$$\mu_{it} = \frac{\theta_i}{1-\phi_{ii}} + \Omega_{ii} v_{it}(1 + \phi_{ii} L + \phi_{ii}^2 L^2 + ...), \tag{D.5}$$

where L refers to the lag operator, and $i = g$ for real activity growth, $i = \pi$ for inflation. ϕ_{ii} corresponds to the diagonal terms of companion matrix ϕ. Note that the off-diagonal terms are zero because we did not consider correlation between the state variables. Ω_{ii} corresponds to the diagonal terms of matrix Ω. Based on equation (D.5) and equation (3.3), we can express the long-run component x_t^* in the following Wold form

$$\begin{aligned} g_t^* &= \frac{\theta_g}{(1-\phi_{gg})(1-\rho_{gg})} + \frac{\Omega_{gg}}{(1-\rho_{gg})} v_{gt}(1 + \phi_{gg} L + \phi_{gg}^2 L^2 + ...), \\ \pi_t^* &= \frac{\theta_\pi}{(1-\phi_{\pi\pi})(1-\rho_{\pi\pi})} + \frac{\Omega_{\pi\pi}}{(1-\rho_{\pi\pi})} v_{\pi t}(1 + \phi_{\pi\pi} L + \phi_{\pi\pi}^2 L^2 + ...). \end{aligned} \tag{D.6}$$

Substituting equation (D.5) into the dynamics of the macro variables, and rewriting the macro variables as follows

$$\begin{aligned} x_{it} &= \frac{\mu_{it}}{1-\rho_{ii} L} + \Sigma_{ii} e_{it}(1 + \rho_{ii} L + \rho_{ii}^2 L^2 + ...) \\ &= (\frac{\theta_i}{1-\phi_{ii} L} + \Omega_{ii} v_{it}(1 + \phi_{ii} L + \phi_{ii}^2 L^2 + ...))(1 + \rho_{ii} L + \rho_{ii}^2 L^2 + ...) \\ &+ \Sigma_{ii} e_{it}(1 + \rho_{ii} L + \rho_{ii}^2 L^2 + ...), \end{aligned} \tag{D.7}$$

where ρ_{ii} corresponds to the diagonal elements of companion matrix ρ, and Σ_{ii} corresponds to the diagonal elements of matrix Σ.

The short-run and long-run shocks at time t are given by e_t and v_t. Let ς_k^{sr} represent the impulse responses of the yield curve at horizon k resulting from the shocks to the short-run components, and let ς_k^{lr} represent the impulse responses of the yield curve at horizon k resulting from the shocks to the long-run components. According to equations (D.4), (D.6), and (D.7) , we can write the impulse responses to the short-run shocks as follows

$$
\begin{aligned}
\varsigma_0^{sr} &= \overline{B}^{1'}\Sigma \\
\varsigma_1^{sr} &= \overline{B}^{1'}\Sigma\rho \\
&... \\
\varsigma_k^{sr} &= \overline{B}^{1'}\Sigma\rho^k \text{for } k \geq 1,
\end{aligned}
\tag{D.8}
$$

where ς_k^{sr} is a 9×2 matrix, $\varsigma_k^{sr} = [\varsigma_k^{sr,g}, \varsigma_k^{sr,\pi}]$, $\varsigma_k^{sr,g}$ and $\varsigma_k^{sr,\pi}$ are 9×1 vectors, with $\varsigma_k^{sr,g}$ the impulse responses of the yield curve to the shocks of the short-run component of real activity growth at horizon k, and $\varsigma_k^{sr,\pi}$ the impulse responses of the yield curve to the shocks of the short-run component of inflation at horizon k. Each row of ς_k^{sr} corresponds to the impulse responses of a certain maturity yield to short-run shocks to real activity growth and inflation at horizon k.

Also based on equations (D.4), (D.6), and (D.7), we can express the impulse responses to the long-run shocks as follows

$$
\begin{aligned}
\varsigma_0^{lr} &= \overline{B}^{2'}\Omega + \overline{B}^{1'}\Omega \\
\varsigma_1^{lr} &= \overline{B}^{2'}\Omega\phi + \overline{B}^{1'}\Omega(\rho + \phi) \\
\varsigma_2^{lr} &= \overline{B}^{2'}\Omega\phi^2 + \overline{B}^{1'}\Omega(\rho^2 + \rho\phi + \phi^2) \\
&... \\
\varsigma_k^{lr} &= \overline{B}^{2'}\Omega\phi^k + \overline{B}^{1'}\Omega(\rho^k + \rho^{k-1}\phi + \rho^{k-2}\phi^2 + ... + \rho^2\phi^{k-2} + \rho\phi^{k-1} + \phi^k) \text{ for } k \geq 1,
\end{aligned}
\tag{D.9}
$$

where ς_k^{lr} is a 9×2 matrix, $\varsigma_k^{lr} = [\varsigma_k^{lr,g}, \varsigma_k^{lr,\pi}]$, $\varsigma_k^{lr,g}$ and $\varsigma_k^{lr,\pi}$ are 9×1 vectors, with $\varsigma_k^{lr,g}$ the impulse responses of the yield curve to the shocks of the long-run component of real activity growth at horizon k, and $\varsigma_k^{lr,\pi}$ the impulse responses of the yield curve to the shocks of the long-run component of inflation at horizon k. Each row of ς_k^{lr} corresponds to the impulse responses of a certain maturity yield to the long-run shocks of real activity growth and inflation at horizon k.

E Variance Decomposition

Working with the Wold $VMA(\infty)$ representation in equation (D.2), the error of the optimal k-step ahead forecast of the term structure of yields at time t, $\widehat{y}_{t+k|t}$, is given by

$$\widehat{y}_{t+k|t} \;-\; y_{t+k} \;= \sum_{s=0}^{k-1} (\varsigma_s^{sr} e_{t+k-s} + \varsigma_s^{lr} \upsilon_{t+k-s}), \tag{E.1}$$

where $\widehat{y}_{t+k|t}$ is a 9×1 vector. Denoting the mean squared error of $\widehat{y}_{t+k|t}$ by $MSE(\widehat{y}_{t+k|t})$, we have

$$MSE(\widehat{y}_{t+k|t}) = \sum_{s=0}^{k-1} [(\varsigma_s^{sr,g})^2 + (\varsigma_s^{sr,\pi})^2 + (\varsigma_s^{lr,g})^2 + (\varsigma_s^{sr,\pi})^2]. \tag{E.2}$$

The contribution of the short-run components to the MSE of the k-step ahead forecast of the yield curve is given by

$$\Xi_k^{sr,g} = \frac{\sum_{s=0}^{k-1} (\varsigma_s^{sr,g})^2}{MSE(\widehat{y}_{t+k|t})}, \tag{E.3}$$

$$\Xi_k^{sr,\pi} = \frac{\sum_{s=0}^{k-1} (\varsigma_s^{sr,\pi})^2}{MSE(\widehat{y}_{t+k|t})}, \tag{E.4}$$

where $\Xi_k^{sr,g}$ and $\Xi_k^{sr,\pi}$ are 9×1 vectors. $\Xi_k^{sr,g}$ represents the contributions of the short-run component of real activity growth to the k-step ahead forecast variance of the yield curve. $\Xi_k^{sr,\pi}$ represents the contributions of the short-run component of inflation to the k-step ahead forecast variance of the yield curve.

The contribution of the long-run components to the MSE of the k-step ahead forecast of the yield curve is given by

$$\Xi_k^{lr,g} = \frac{\sum_{s=0}^{k-1} (\varsigma_s^{lr,g})^2}{MSE(\widehat{y}_{t+k|t})}, \tag{E.5}$$

$$\Xi_k^{lr,\pi} = \frac{\sum_{s=0}^{k-1} (\varsigma_s^{lr,\pi})^2}{MSE(\widehat{y}_{t+k|t})}, \tag{E.6}$$

where $\Xi_k^{lr,g}$ and $\Xi_k^{lr,\pi}$ are 9×1 vectors. $\Xi_k^{lr,g}$ represents the contributions of the long-run component of real activity growth to the k-step ahead forecast variance of the yield curve. $\Xi_k^{lr,\pi}$ represents the contributions of the long-run component of inflation to the k-step ahead forecast variance of the yield curve.

F Model-Implied Term Premia

As discussed in Section 3.7.1, we can write the forward term premium $FTP_t^{h_1,h_2}$ as follows

$$FTP_t^{h_1,h_2} = F_t^{h_1,h_2} - \underbrace{E_t \sum_{i=h_1}^{h_2-1} y(t+i, t+i+1)}_{\text{Expectation (Risk-Neutral Yield)}} . \tag{F.1}$$

The forward rate $F_t^{h_1,h_2}$ that one can lock in at time t for a $(h_2 - h_1)$-period loan starting in h_1 periods is computed as

$$F_t^{h_1,h_2} = \log[p(t, t+h_1)/p(t, t+h_2)], \tag{F.2}$$

where $p(t, t+h_1)$ denotes time t price of a h_1-period zero coupon bond.

Given model-implied yield for one-period bond

$$y(t, t+1) = -A_1 - B_1' F_t, \tag{F.3}$$

the expectation term in equation (F.1) can be rewritten as follows

$$\begin{aligned} E_t \sum_{i=h_1}^{h_2-1} y(t+i, t+i+1) &= \sum_{i=h_1}^{h_2-1} E_t[y(t+i, t+i+1)] \\ &= -\sum_{i=h_1}^{h_2-1} E_t[A_1 + B_1' F_{t+i}] \\ &= -\sum_{i=h_1}^{h_2-1} A_1 - B_1' \sum_{i=h_1}^{h_2-1} E_t[F_{t+i}], \end{aligned} \tag{F.4}$$

where $E_t[F_{t+i}]$ is a function of parameters ψ and K given F_t and i

$$E_t[F_{t+i}] = (I_4 + K + ... + (K)^{i-1})\psi + (K)^i F_t. \tag{F.5}$$

G The Performance of Related Models

Our model is related to an extensive literature on term structure modeling, the relation between macroeconomics and the term structure, and the usefulness of information extracted from yields for forecasting inflation and real activity growth. A substantial part of the term structure literature uses models with latent factors.

Over the last decade, a growing literature has emerged that models the interaction between the term structure and macroeconomic variables within the context of a no-arbitrage setup. This literature has not only focused on building better models of the term structure, but also on using term-structure information to improve forecasts of macroeconomic variables.

Our model can be most usefully thought of as an extension of the classical Ang and Piazzesi (2003) setup, where the two macroeconomic variables have a stochastic long-run mean. We therefore have included this model as a benchmark in the empirical exercises. In Ang and Piazzesi (2003), the macroeconomic variables influence the term structure through their impact on the short rate via a simple Taylor rule. There is a rich literature that uses other policy rules together with a variety of no-arbitrage models. The studies that are most closely related to ours use macroeconomic variables with shifting endpoints. Van Dijk, Koopman, Van der Wel, and Wright (2014) use a Nelson-Siegel (1987) model with shifting endpoints. Dewachter and Lyrio (2008) use New Keynesian models with shifting endpoints. Finally Dewachter and Lyrio (2006), Dewachter, Lyrio, and Maes (2006), Berardi (2009), and Orphanides and Wei (2012) use affine models with a variety of economic restrictions and policy rules that impose restrictions among the state variables.

We do not attempt a full-fledged horserace between these models; we merely want to illustrate how the main differences with our model setup influence the ability of these models to fit and forecast yields. Consider the models by Dewachter and Lyrio (2006) and Dewachter, Lyrio, and Maes (2006), who were the first to propose a shifting endpoints approach in a model with macro variables, using a continuous time framework. In order to compare the class of models they propose with our model more directly, we implement a discrete-time counterpart of their model using our sample.[15] Our implementation also differs from the implementation in Dewachter and Lyrio (2006) and Dewachter, Lyrio, and Maes (2006) in some other relatively minor ways. However, we found that our main point does not depend on the implementation: while the shifting endpoints approach in the existing literature allows for adequate modeling of the long end of the term structure, the absence of additional economic restrictions in our model leads to a better out-of-sample fit. This simply reflects a different objective. We are more focused on the impact of the long-run components of the macro variables on the term structure, while the existing papers are more interested in what the term structure can tell us about the macroeconomy.

Dewachter and Lyrio (2006) consider a no-arbitrage term structure model in which the state variables are real short rate, the output gap and inflation, and also the shifting endpoints of these macro variables. The model in Dewachter, Lyrio, and Maes (2006) is similar to the one in Dewachter and Lyrio (2006). They include shifting endpoints of the real short rate as another state variable in the bond pricing. In both papers, the real short rate and the macro variables are potentially correlated, which is equivalent to allowing for off-diagonal terms in the companion matrix for the factors in the discrete time VAR process. This channel introduces feedback from the latent factors to the macro variables. Ang and Piazzesi (2003) only allow for unidirectional dynamics in the no-arbitrage model: macro variables determine yields but not vice versa. They find that macro variables do not account for the movements at the long end of the yield curve, while the level factor does. Later work by Diebold, Rudebusch, and Aruoba (2006) and Ang, Dong, and Piazzesi (2007) find that these interactions are important. The amount of yield variation that can be attributed to macro factors depends on whether or not the system allows for bidirectional linkages. Therefore the improvement in in-sample model fit at the long end of the yield curve in Dewachter and Lyrio (2006) may not entirely be due to the trending component of inflation but rather to

[15] We have also implemented these models in continuous time, and the results are similar.

the correlations.

We therefore implement the models in Dewachter, Lyrio, and Maes (2006) and Dewachter and Lyrio (2006) using our sample with unidirectional and bidirectional dynamics. In the following, we layout the main framework of the models. Dewachter and Lyrio (2006) consider a no-arbitrage term structure model in which the state variables are the real short rate ρ_t, macro variables (the output gap y_t and inflation π_t), and the shifting endpoints of the macro variables (y_t^* and π_t^*). They assume the following dynamics for the state variables

$$\begin{aligned} y_{t+1} &= \kappa_{yy}(y_t - y_t^*) + \kappa_{y\pi}(\pi_t - \pi_t^*) + \kappa_{y\rho}(\rho_t - \rho_t^*) + \sigma_y e_{y_{t+1}}, \\ \pi_{t+1} &= \kappa_{\pi y}(y_t - y_t^*) + \kappa_{\pi\pi}(\pi_t - \pi_t^*) + \kappa_{\pi\rho}(\rho_t - \rho_t^*) + \sigma_\pi e_{\pi_{t+1}}, \\ \rho_{t+1} &= \kappa_{\rho y}(y_t - y_t^*) + \kappa_{\rho\pi}(\pi_t - \pi_t^*) + \kappa_{\rho\rho}(\rho_t - \rho_t^*) + \sigma_\rho e_{\rho_{t+1}}. \end{aligned} \tag{G.1}$$

The shifting endpoints y_t^* and π_t^* are assumed to follow a random walk process

$$\begin{aligned} y_{t+1}^* &= y_t^* + \sigma_{y^*} e_{y_{t+1}^*}, \\ \pi_{t+1}^* &= \pi_t^* + \sigma_{\pi^*} e_{\pi_{t+1}^*}. \end{aligned} \tag{G.2}$$

Dewachter and Lyrio (2006) simplify the empirical model by concentrating only on the shifting endpoints of inflation π_t^*. They use the Hodrick-Prescott filter to pre-filter the output data and make use of the transitory component of the series. The long-run expectation of the output gap y_t^* is set equal to zero. They assume the central bank follows a rule based on the long-run expectations of the macro variables. More specifically, the policy rule is formalized as

$$\rho_{t+1}^* = \gamma_0 + \gamma_{y^*} y_{t+1}^* + \gamma_{\pi^*} \pi_{t+1}^*. \tag{G.3}$$

Since they set y_t^* equal to zero, ρ_t^* is just a linear function of π_t^*. The nominal short rate is defined as follows

$$r_t = \pi_t + \rho_t. \tag{G.4}$$

The above representation of the dynamics of the economy can be restated in matrix notation. There are four factors in the model, and we can define the factors, the shocks and the variance matrix SS' as

$$F_t = \begin{pmatrix} y_t \\ \pi_t \\ \rho_t \\ \pi_t^* \end{pmatrix}, \quad Z_t = \begin{pmatrix} e_{y_t} \\ e_{\pi_t} \\ e_{\rho_t} \\ e_{\pi_t^*} \end{pmatrix}, \quad \text{and } S = diag \begin{pmatrix} \sigma_y & \sigma_\pi & \sigma_\rho & \sigma_{\pi^*} \end{pmatrix}.$$

Therefore, the dynamic of the state variables can be represented as follows

$$F_{t+1} = \psi + KF_t + SZ_{t+1}, \tag{G.5}$$

where

$$K = \begin{pmatrix} \kappa_{yy} & \kappa_{y\pi} & \kappa_{y\rho} & -\kappa_{y\pi} - \kappa_{y\rho}\gamma_{\pi^*} \\ \kappa_{\pi y} & \kappa_{\pi\pi} & \kappa_{\pi\rho} & -\kappa_{\pi\pi} - \kappa_{\pi\rho}\gamma_{\pi^*} \\ \kappa_{\rho y} & \kappa_{\rho\pi} & \kappa_{\rho\rho} & -\kappa_{\rho\pi} - \kappa_{\rho\rho}\gamma_{\pi^*} \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

and

$$\psi = \begin{pmatrix} -\kappa_{y\rho}\gamma_0 & -\kappa_{\pi\rho}\gamma_0 & -\kappa_{\rho\rho}\gamma_0 & 0 \end{pmatrix}'.$$

To develop the term structure model, Dewachter and Lyrio (2006) use the no-arbitrage assumption to guarantee the existence of an equivalent martingale measure Q. The pricing kernel takes the standard form as in equation (3.5) in our model. Given these assumptions on the dynamics of the state variables and the market price of risk, the yields can be written as an affine function of the factors.

The model in Dewachter, Lyrio, and Maes (2006) is similar to the one in Dewachter and Lyrio (2006). The main differences are as follows. First, they do not consider the shifting endpoints of the output gap y_t^* in the dynamics of the state variables.[16] Instead, they include the shifting endpoints of the real short rate ρ_t^* as another state variable in the bond pricing. Second, they do not consider the restrictions on the relation between ρ_t^* and π_t^*. Finally, they adopt a mean reverting process for the shifting endpoints ρ_t^* and π_t^*. The resulting specification has five state variables and the dynamics of the state variables are as follows

$$\begin{aligned} y_{t+1} &= \kappa_{yy} y_t + \kappa_{y\pi}(\pi_t - \pi_t^*) + \kappa_{y\rho}(\rho_t - \rho_t^*) + \sigma_y e_{y_{t+1}}, \\ \pi_{t+1} &= \kappa_{\pi y} y_t + \kappa_{\pi\pi}(\pi_t - \pi_t^*) + \kappa_{\pi\rho}(\rho_t - \rho_t^*) + \sigma_\pi e_{\pi_{t+1}}, \\ \rho_{t+1} &= \kappa_{\rho y} y_t + \kappa_{\rho\pi}(\pi_t - \pi_t^*) + \kappa_{\rho\rho}(\rho_t - \rho_t^*) + \sigma_\rho e_{\rho_{t+1}}. \end{aligned} \tag{G.6}$$

The shifting endpoints follow stationary processes

$$\begin{aligned} \pi_{t+1}^* &= \phi_{\pi^*}(\pi_t^* - \mu_{\pi^*}) + \sigma_{\pi^*} e_{\pi_{t+1}^*}, \\ \rho_{t+1}^* &= \phi_{\rho^*}(\rho_t^* - \mu_{\rho^*}) + \sigma_{\rho^*} e_{\rho_{t+1}^*}. \end{aligned} \tag{G.7}$$

The nominal short rate is defined in the same way as in Dewachter and Lyrio (2006)

$$r_t = \pi_t + \rho_t. \tag{G.8}$$

Again, the above representation of the dynamics of the state variables can be restated in matrix notation. We define the factors, the shocks and the variance matrix SS' as

$$F_t = \begin{pmatrix} y_t \\ \pi_t \\ \rho_t \\ \pi_t^* \\ \rho_t^* \end{pmatrix}, \; Z_t = \begin{pmatrix} e_{y_t} \\ e_{\pi_t} \\ e_{\rho_t} \\ e_{\pi_t^*} \\ e_{\rho_t^*} \end{pmatrix}, \text{ and } S = diag\begin{pmatrix} \sigma_y & \sigma_\pi & \sigma_\rho & \sigma_{\pi^*} & \sigma_{\rho^*} \end{pmatrix}.$$

[16] Dewachter and Lyrio (2006) set y_t^* equal to zero in the estimation. Therefore, removing y_t^* from the dynamic term structure model should not make any difference.

Therefore, the dynamic of the state variables can be represented as follows

$$F_{t+1} = \psi + KF_t + SZ_{t+1}, \tag{G.9}$$

where

$$K = \begin{pmatrix} \kappa_{yy} & \kappa_{y\pi} & \kappa_{y\rho} & -\kappa_{y\pi} & -\kappa_{y\rho} \\ \kappa_{\pi y} & \kappa_{\pi\pi} & \kappa_{\pi\rho} & -\kappa_{\pi\pi} & -\kappa_{\pi\rho} \\ \kappa_{\rho y} & \kappa_{\rho\pi} & \kappa_{\rho\rho} & -\kappa_{\rho\pi} & -\kappa_{\rho\rho} \\ 0 & 0 & 0 & \phi_{\pi^*} & 0 \\ 0 & 0 & 0 & 0 & \phi_{\rho^*} \end{pmatrix},$$

and

$$\psi = \begin{pmatrix} 0 & 0 & 0 & -\phi_{\pi^*}\mu_{\pi^*} & -\phi_{\rho^*}\mu_{\rho^*} \end{pmatrix}'.$$

Table A1 reports the in- and out-of-sample results, where a star refers to the model without correlations. The DLM model is the best performer both in- and out-of-sample, and we therefore limit ourselves to a comparison between the DLM model and our model. There is an interesting contrast between in- and out-of-sample performance when comparing the DLM model with our model. In-sample, the DLM model performs slightly worse than our model for intermediate and longer maturities. Out-of-sample, the performance of our models is better at all maturities, and the DLM model underperforms rather spectacularly at the short end. One possible explanation for these findings is that the richer economic restrictions in the DLM model make it more difficult to fit the short end of the term structure. In-sample this shortcoming is overcome by the fact that the model is rather richly parameterized, but the restrictions affect the model's out-of-sample performance.[17] Our prior is that this argument will hold a fortiori for other models with shifting endpoints in the macro-finance literature,

[17]Our model has 30 parameters including the measurement error standard deviations for the yields, and two latent factors. The model in Dewachter and Lyrio (2006) has 35 parameters and two latent factors. When restricting the correlations between real short rate and macro variables to be zero, it has 27 parameters and two latent factors. The model in Dewachter, Lyrio, and Maes (2006) has 40 parameters and three latent factors. With zero correlations, it has 32 parameters and three latent factors.

such as the New Keynesian models, because these models build in even more economic restrictions. Note however that this should not be thought of as a shortcoming of these models, but rather as a trade-off. When modeling the economic variables, these restrictions are presumably helpful, while for forecasting the term structure they may hurt.[18] A more thorough investigation of this trade-off would be interesting, but we leave this for future work.

[18]The different data used in the component model and our implementation of DL and DLM may explain some of the differences in fit. For the component model, we follow Ang and Piazzesi (2003), who use real activity growth. For the DL and DLM implementation, we construct the data for the output gap by interpolating the quarterly real GDP data to get monthly data, and applying the Hodrick-Prescott filter to the logarithm of monthly real GDP. The residuals are taken as the measure of output gap. The correlation between this measure of the output gap and real activity growth is approximately 0.7.

Bibliography

[1] Ajello, A., Benzoni, L. and Chyruk, O. 2012. Core and Crust: Consumer Prices and the Term Structure of Interest Rates, Working Paper, Federal Reserve Bank of Chicago.

[2] Ang, A. and Bekaert, G. 2002. Regime Switches in Interest Rates, Journal of Business and Economic Statistics, 20, 163-182.

[3] Ang, A., Bekaert, G. and Wei, M. 2007. Do Macro Variables, Asset Markets, or Surveys Forecast Inflation Better?, Journal of Monetary Economics, 54, 1163-1212.

[4] Ang, A., Bekaert, G. and Wei, M. 2008. The Term Structure of Real Rates and Expected Inflation, The Journal of Finance, 63, 797-849.

[5] Ang, A., Boivin, J., Dong, S. and Loo-Kung, R. 2011. Monetary Policy Shifts and the Term Structure, The Review of Economic Studies, 78, 429-457.

[6] Ang, A., Dong, S. and Piazzesi, M. 2007. No-Arbitrage Taylor Rules, Working Paper, National Bureau of Economic Research.

[7] Ang, A. and Piazzesi, M. 2003. A No-Arbitrage Vector Autoregression of Term Structure Dynamics with Macroeconomic and Latent Variables, Journal of Monetary Economics, 50, 745-787.

[8] Ang, A., Piazzesi, M. and Wei, M. 2006. What Does the Yield Curve Tell Us About GDP Growth?, Journal of Econometrics, 131, 359-403.

[9] Aït-Sahalia, Y. and Kimmel, R. L. 2010. Estimating Affine Multifactor Term Structure Models Using Closed-Form Likelihood Expansions, Journal of Financial Economics, 98, 113-144.

[10] Backus, D. K. and Zin, S. E. 1993. Long Memory Inflation Uncertainty: Evidence from the Term Structure of Interest Rates, Journal of Money Credit and Banking, 25, 681-700.

[11] Bansal, R. and Zhou, H. 2002. Term Structure of Interest Rates with Regime Shifts, The Journal of Finance, 57, 1997-2043.

[12] Bauer, M. D. 2016. Restrictions on Risk Prices in Dynamic Term Structure Models. Journal of Business & Economic Statistics, Forthcoming.

[13] Bauer, M. D., Rudebusch, G. D. and Wu, J. C., 2012. Correcting Estimation Bias in Dynamic Term Structure Models. Journal of Business & Economic Statistics, 30, 454-467.

[14] Berardi, A. 2009. Term Structure, Inflation, and Real Activity, Journal of Financial and Quantitative Analysis, 44, 987-1011.

[15] Beveridge, S. and Nelson, C. R. 1981. A New Approach to Decomposition of Economic Time Series into Permanent and Transitory Components with Particular Attention to Measurement of the 'Business Cycle', Journal of Monetary Economics, 7, 151-174.

[16] Bikbov, R. and Chernov, M. 2010. No-Arbitrage Macroeconomic Determinants of the Yield Curve, Journal of Econometrics, 159, 166-182.

[17] Brandt, M. W. and He, P. 2002. Simulated likelihood Estimation of Affine Term Structure Models from Panel Data, Working Paper, Wharton.

[18] Campbell, J. Y. and Cochrane, J. H. 1999. By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior, Journal of Political Economy, 107, 205–251.

[19] Campbell, J. Y. and Shiller, R. J. 1991. Yield Spreads and Interest Rate Movements: A Bird's Eye View. The Review of Economic Studies, 58, 495-514.

[20] Chen, R. R. and Scott, L. 1993. Maximum Likelihood Estimation for a Multifactor Equilibrium Model of the Term Structure of Interest Rates, The Journal of Fixed Income, 3, 14-31.

[21] Chernov, M. and Mueller, P. 2012. The Term Structure of Inflation Expectations, Journal of Financial Economics, 106, 367-394.

[22] Cieslak, A., and Povala, P. 2015. Expected Returns in Treasury Bonds. Review of Financial Studies, 28, 2859-2901.

[23] Clark, P. K. 1987. The Cyclical Component of U.S. Economic Activity, The Quarterly Journal of Economics, 102, 797-814.

[24] Cochrane, J. H. 1989. The Return of the Liquidity Effect: A Study of the Short-Run Relation between Money Growth and Interest Rates, Journal of Business and Economic Statistics, 7, 75-83.

[25] Cochrane, J. H. and Piazzesi, M. 2005. Bond Risk Premia. The American Economic Review, 95, 138-160.

[26] Cochrane, J. H. and Piazzesi, M. 2008. Decomposing the Yield Curve. Working paper, University of Chicago.

[27] Cogley, T. and Sargent, T. J. 2001. Evolving Post World War II U.S. Inflation Dynamics, NBER Macroeconomics Annual, 16, 331-373.

[28] Cogley, T. and Sargent, T. J. 2005. Drifts and Volatilities: Monetary Policies and Outcomes in the Post World War II U.S., Review of Economic Dynamics, 8, 262-302.

[29] Cogley, T., Primiceri, G. E. and Sargent, T. J. 2010. Inflation-Gap Persistence in the U.S., American Economic Journal: Macroeconomics, 2, 43-69.

[30] Collin-Dufresne, Goldstein, R. S. and Jones, C. S. 2008. Identification of Maximal Affine Term Structure Models, The Journal of Finance, 63, 743-795.

[31] Cooper, I. and Priestley, R. 2009. Time-Varying Risk Premiums and the Output Gap. Review of Financial Studies, 22, 2801-2833.

[32] Cox, J. C., Ingersoll Jr, J. E. and Ross, S. A. 1985. A Theory of the Term Structure of Interest Rates, Econometrica, 385-407.

[33] Dai, Q., Singleton, K. J., and Yang, W. 2007. Regime Shifts in a Dynamic Term Structure Model of US Treasury Bond Yields, Review of Financial Studies, 20, 1669-1706.

[34] Dewachter, H. and Lyrio, M. 2006. Macro Factors and the Term Structure of Interest Rates, Journal of Money Credit and Banking, 38, 119-140.

[35] Dewachter, H. and Lyrio, M. 2008. Learning, Macroeconomic Dynamics and the Term Structure of Interest Rates. In John Campbell (ed.): Asset Prices and Monetary Policy, University of Chicago Press, 191-245.

[36] Dewachter, H., Lyrio, M. and Maes, K. 2006. A Joint Model for the Term Structure of Interest Rates and the Macroeconomy, Journal of Applied Econometrics, 21, 439-462.

[37] Diebold, F., Rudebusch, G. D. and Aruoba, S. B. 2006. The Macroeconomy and the Yield Curve: A Dynamic Latent Factor Approach, Journal of Econometrics, 131, 309-338.

[38] Duffee, G. R. 2011. Information in (and not in) the Term Structure. Review of Financial Studies, 24, 2895-2934.

[39] Duffee, G. R. 2013. Bond Pricing and the Macroeconomy. Constantinides, G. M., Harris, M., and Stulz, R. M. Edition, Handbook of the Economics of Finance, Volume 2B: Asset Pricing, 907-967. North Holland.

[40] Duffee, G. R. 2014. Expected Inflation and Other Determinants of Treasury Yields, Working Paper, Johns Hopkins University.

[41] Duffee, G. R. and Stanton, R. H. 2012. Estimation of Dynamic Term Structure Models, The Quarterly Journal of Finance, 2.

[42] Duffie, D. and Kan, R. 1996. A Yield Factor Model of Interest Rates, Mathematical Finance, 6, 379-406.

[43] Erceg, C. J. and Levin, A. T. 2003. Imperfect Credibility and Inflation Persistence, Journal of Monetary Economics, 50, 915-944.

[44] Evans, C. L. and Marshall, D. A. 1998. Monetary Policy and the Term Structure of Nominal Interest Rates: Evidence and Theory, Carnegie-Rochester Conference Series on Public Policy, 49, 53-111.

[45] Fama, E. F. 1990. Term Structure Forecasts of Interest Rates, Inflation, and Real Returns, Journal of Monetary Economics, 25, 59-76.

[46] Fama, E. F. and Bliss, R.R. 1987. The Information in Long-Maturity Forward Rates. The American Economic Review, 680-692.

[47] Faust, J. and Wright, J. 2011. Forecasting Inflation, Working Paper, Johns Hopkins University.

[48] Goodfriend, M. 1993. Interest Rate Policy and the Inflation Scare Problem: 1979-1992, FRB Richmond Economic Quarterly, 79, 1-23.

[49] Granger, C. W. J. and Joyeux, R. 1980. An Introduction to Long Memory Time Series Models and Fractional Differencing, Journal of Time Series Analysis, 1, 15-29.

[50] Greenspan, A. 2005. Federal Reserve Board's semiannual monetary policy report to the Congress. January, 16.

[51] Greenwood, R. and Vayanos, D. 2014. Bond Supply and Excess Bond Returns. Review of Financial Studies, 27, 663-713.

[52] Gürkaynak, R. S., Sack, B. and Swanson, E. 2005. The Sensitivity of Long Term Interest Rates to Economic News: Evidence and Implications for Macroeconomic Models, American Economic Review, 95, 425-436.

[53] Gürkaynak, R. S. and Wright, J. H. 2012. Macroeconomics and the Term Structure, Journal of Economic Literature, 50, 331-367.

[54] Hamilton, J. D. 1988. Rational-Expectations Econometric Analysis of Changes in Regime: An Investigation of the Term Structure of Interest Rates, Journal of Economic Dynamics and Control, 12, 385-423.

[55] Harvey, A. C. 1985. Trends and Cycles in Macroeconomic Time Series, Journal of Business and Economic Statistics, 3, 216-227.

[56] Hodrick, R. J. and Prescott, E. C. 1997. Postwar US Business Cycles: An Empirical Investigation, Journal of Money, Credit, and Banking, 29, 1-16.

[57] Hördahl, P., Tristani, O. and Vestin, D. 2006. A Joint Econometric Model of Macroeconomic and Term-Structure Dynamics, Journal of Econometrics, 131, 405-444.

[58] Joslin, S., Le, A., and Singleton, K. J. 2013. Why Gaussian Macro-Finance Term Structure Models Are (Nearly) Unconstrained Factor-VARs, Journal of Financial Economics, 109, 604-622.

[59] Joslin, S., Priebsch, M., and Singleton, K. J. 2014. Risk Premiums in Dynamic Term Structure Models with Unspanned Macro Risks, The Journal of Finance, 69, 1197-1233.

[60] Keim, D. B. and Stambaugh, R. F. 1986. Predicting Returns in the Stock and Bond Markets. Journal of Financial Economics, 17, 357-390.

[61] Kim, D. H. and Orphanides, A. 2012. Term Structure Estimation with Survey Data on Interest Rate Forecasts. Journal of Financial and Quantitative Analysis, 47, 241-272.

[62] Kozicki, S. and Tinsley, P. A. 1998. Moving Endpoints and the Internal Consistency of Agents' Ex Ante Forecasts, Computational Economics, 11, 21-40.

[63] Kozicki, S. and Tinsley, P. A. 2001a. Shifting Endpoints in the Term Structure of Interest Rates, Journal of Monetary Economics, 47, 613-652.

[64] Kozicki, S. and Tinsley, P. A. 2001b. Term Structure Views of Monetary Policy Under Alternative Models of Agent Expectations, Journal of Economic Dynamics and Control, 25, 149-184.

[65] Kuttner, K. N. 2001. Monetary Policy Surprises and Interest Rates: Evidence from the Fed Funds Futures Market, Journal of Monetary Economics, 47, 523-544.

[66] Laubach, T., Tetlow, R. J., and Williams, J. C. 2007. Learning and the Role of Macroeconomic Factors in the Term Structure of Interest Rates, Working Paper, Federal Reserve Board of Governors.

[67] Litterman, R. B. and Scheinkman, J. 1991. Common Factors Affecting Bond Returns. The Journal of Fixed Income, 1, 54-61.

[68] Ljungqvist, L. and Sargent, T. J. 2004. Recursive Macroeconomic Theory, The MIT Press.

[69] Lu, B. and Wu, L. 2009. Macroeconomic Releases and the Interest Rate Term Structure, Journal of Monetary Economics, 56, 872-884.

[70] Ludvigson, S. C. and Ng, S. 2009. Macro Factors in Bond Risk Premia. Review of Financial Studies, 22, 5027-5067.

[71] Mishkin, F. 1990. The Information in the Longer Maturity Term Structure about Future Inflation, The Quarterly Journal of Economics, 105, 815-828.

[72] Morley, J. C., Nelson, C. R., and Zivot, E. 2003. Why are the Beveridge-Nelson and Unobserved-Components Decompositions of GDP so Different?, Review of Economics and Statistics, 85, 235-243.

[73] Nelson, C. R. and Siegel, A. F. 1987. Parsimonious Modeling of Yield Curves, Journal of Business, 473-489.

[74] Orphanides, A. and Wei, M. 2012. Evolving Macroeconomic Perceptions and the Term Structure of Interest Rates, Journal of Economic Dynamics and Control, 36, 239-254.

[75] Pearson, N. D. and Sun, T. S. 1994. Exploiting the Conditional Density in Estimating the Term Structure: An Application to the Cox, Ingersoll, and Ross Model, The Journal of Finance, 49, 1279-1304.

[76] Piazzesi, M. 2005. Bond Yields and the Federal Reserve, Journal of Political Economy, 113, 311-344.

[77] Rudebusch, G. D. and Wu, T. 2008. A Macro Finance Model of the Term Structure, Monetary Policy and the Economy, The Economic Journal, 118, 906-926.

[78] Stock, J. H. and Watson, M. W. 2007. Why Has Inflation Become Harder to Forecast?, Journal of Money Credit and Banking, 39, 3-34.

[79] Taylor, J. B. 1993. Discretion Versus Policy Rules in Practice, Carnegie-Rochester Conference Series on Public Policy, 39, 195-214.

[80] Van Dijk, D., Koopman, S. J., Van der Wel, M. and Wright, J. H. 2014. Forecasting Interest Rates with Shifting Endpoints, Journal of Applied Econometrics.

[81] Vasicek, O. 1977. An Equilibrium Characterization of the Term Structure, Journal of Financial Economics, 5, 177-188.

[82] Wachter, J. A. 2006. A Consumption-Based Model of the Term Structure of Interest Rates. Journal of Financial Economics, 79, 365-399.

[83] Watson, M. W. 1986. Univariate Detrending Methods with Stochastic Trends, Journal of Monetary Economics, 18, 49-75.

[84] Wright, J. H. 2011. Term Premia and Inflation Uncertainty: Empirical Evidence From An International Panel Dataset. The American Economic Review, 101, 1514-1534.

Table 3.1 Summary Statistics

	Central Moments				Autocorrelations		
	Mean	St.Dev	Skewness	Kurtosis	Lag 1	Lag 12	Lag 30
3 month yield	0.0450	0.0290	0.8938	4.3247	0.9773	0.7944	0.5197
6 month yield	0.0479	0.0305	0.8717	4.2283	0.9850	0.8126	0.5359
1 year yield	0.0516	0.0306	0.6980	3.6594	0.9857	0.8317	0.5891
2 year yield	0.0536	0.0301	0.6734	3.4957	0.9878	0.8509	0.6485
3 year yield	0.0554	0.0294	0.6703	3.4460	0.9884	0.8611	0.6782
4 year yield	0.0569	0.0288	0.6903	3.4142	0.9882	0.8655	0.7000
5 year yield	0.0579	0.0282	0.7270	3.3717	0.9890	0.8739	0.7183
10 year yield	0.0617	0.0275	0.9148	3.5853	0.9890	0.8739	0.7183
20 year yield	0.0638	0.0265	0.9158	3.5373	0.9930	0.8936	0.7724
Inflation	0.0361	0.0272	1.4029	5.1277	0.9901	0.7529	0.4403
Real Activity Growth	0.0296	0.0433	1.3115	5.0337	0.9685	-0.0221	-0.0649
UE	0.0595	0.0162	0.6661	3.0617	0.9889	0.7095	0.3149
IP	0.0143	0.0165	-0.7045	3.9698	0.9548	0.2188	-0.1159
EMPLOY	0.0280	0.0537	-0.7254	4.3417	0.9638	-0.0759	-0.0569
PayrollEMPLOY	0.0167	0.0211	-0.7578	3.4009	0.9842	0.2515	-0.0816

Notes to Table: We present summary statistics for the data used in estimation. We present the sample mean, standard deviation, skewness, kurtosis, and autocorrelations for each of the yields and macro variables. The yields are continuously compounded monthly zero-coupon bond yields. We use the Consumer Price Index for all urban consumers to calculate inflation. Following Ang and Piazzesi (2003), we extract the first principal component from a group of monthly variables that capture real activity growth: unemployment, the growth rate of industrial production, the growth rate of employment, and the growth rate of payroll employment. All growth rates are measured as the log difference of the index in months t and t-12. The sample period is from 1953:04 to 2012:12.

Table 3.2 Parameter Estimates

Panel A: Component Model

	Real Activity Growth	Inflation
Short Rate Parameters		
$\alpha_0 \times 10^2$	0.0313	
	(0.0174)	
α_1	0.0087	
	(0.0020)	
α_2	0.0914	
	(0.0038)	
VAR Parameters P dynamics		
	Real Activity Growth	Inflation
ρ	0.9291	0.6020
	(0.0219)	(0.0321)
Σ	0.0147	0.0045
	(0.0028)	(0.0004)
$\theta \times 10^2$	-0.0072	0.0105
	(0.0005)	(0.0008)
ϕ	0.9915	0.9988
	(0.0042)	(0.0015)
$\Omega \times 10^2$	0.0212	0.0253
	(0.0019)	(0.0019)
VAR Parameters Q dynamics		
	Real Activity Growth	Inflation
ρ	0.9690	0.6062
	(0.0075)	(0.1428)
$\theta \times 10^2$	-0.0060	0.0109
	(0.0006)	(0.0010)
ϕ	0.9912	0.9986
	(0.0090)	(0.0036)
Risk Premium Parameters		
	Real Activity Growth	Inflation
λ_{02}	-0.0584	-0.0167
	(0.0512)	(0.0115)
λ_{11}	-2.7055	-0.9257
	(0.4988)	(0.2625)
λ_{22}	1.2793	0.8646
	(0.7432)	(0.4743)

Panel B: Benchmark Macro Model

	Real Activity Growth	Inflation
Short Rate Parameters		
$\alpha_0 \times 10^2$	0.1347	
	(0.0195)	
α_1	0.0075	
	(0.0012)	
α_2	0.0730	
	(0.0025)	
VAR Parameters P dynamics		
	Real Activity Growth	Inflation
ρ	0.9731	0.9903
	(0.0089)	(0.0049)
Σ	0.0139	0.0036
	(0.0015)	(0.0004)
$\mu \times 10^2$	0.0135	0.0386
	(0.0015)	(0.0043)
VAR Parameters Q dynamics		
	Real Activity Growth	Inflation
ρ	0.9823	0.9965
	(0.0115)	(0.0821)
$\mu \times 10^2$	0.0669	0.0463
	(0.0203)	(0.0122)
Risk Premium Parameters		
	Real Activity Growth	Inflation
λ_0	-0.0384	-0.0216
	(0.0145)	(0.0054)
λ_1	-0.6591	-1.7470
	(0.1660)	(0.6420)

Notes to Table: We present the parameter estimates for two different specifications. Panel A is for the component model. λ_{02} is the market price of risk for the unconditional mean of the long-run components of the macro variables. We present the diagonal elements of λ_{11}, which represents the market price of risk for the short-run components of the macro variables. We present the diagonal elements of λ_{22}, which represents the market price of risk for the long-run components of the macro variables. Panel B is for the benchmark macro model. λ_0 denotes the market price of risk for the unconditional mean of the macro variables. λ_1 represents the market price of risk for the macro variables. Here we report the diagonal elements of λ_1. Standard errors are reported in parentheses.

Table 3.3 In-Sample Model Fit (RMSE)

	Model 1	Model 1*	Model 2	Model 3	Model 4	Model 4*
3 month yield	122.71	113.75	159.25	43.76	67.21	44.18
6 month yield	81.23	89.00	156.12	37.64	61.75	41.27
1 year yield	54.81	59.16	148.68	31.35	53.50	30.49
2 year yield	41.33	40.26	137.24	34.14	54.30	34.04
3 year yield	35.16	39.92	128.39	34.68	51.23	32.35
4 year yield	34.27	38.18	122.26	34.96	45.97	33.03
5 year yield	34.17	34.12	117.34	33.40	42.35	31.10
10 year yield	39.77	41.54	123.42	46.45	54.30	44.52
20 year yield	41.25	43.68	111.08	50.02	62.74	49.67

Notes to Table: We present the in-sample root mean square error (RMSE) in basis points. Model 1: Component model. Model 1*: Unit root component model. Model 2: Benchmark macro model. Model 3: Model with three latent factors. Model 4: Model with macro variables and three latent factors. Model 4*: Model with macro variables and three latent factors estimated using two-step estimation method.

Table 3.4 Model-Implied Autocorrelations

	Panel A: Data			Panel B: Component Model		
	Lag 1	Lag 12	Lag 30	Lag 1	Lag 12	Lag 30
3 month yield	0.9773	0.7944	0.5197	0.9902	0.8015	0.6044
6 month yield	0.9850	0.8126	0.5359	0.9920	0.8515	0.6853
1 year yield	0.9857	0.8317	0.5891	0.9928	0.8763	0.7184
2 year yield	0.9878	0.8509	0.6485	0.9931	0.8858	0.7285
3 year yield	0.9884	0.8611	0.6782	0.9932	0.8890	0.7321
4 year yield	0.9882	0.8655	0.7000	0.9932	0.8906	0.7340
5 year yield	0.9890	0.8739	0.7183	0.9931	0.8914	0.7349
10 year yield	0.9890	0.8739	0.7183	0.9926	0.8918	0.7320
20 year yield	0.9930	0.8936	0.7724	0.9907	0.8886	0.7176

	Panel C: Benchmark Macro Model			Panel D: Latent-Factor Model		
	Lag 1	Lag 12	Lag 30	Lag 1	Lag 12	Lag 30
3 month yield	0.9851	0.7228	0.4693	0.9797	0.7930	0.5006
6 month yield	0.9851	0.7241	0.4691	0.9857	0.8199	0.5509
1 year yield	0.9849	0.7264	0.4688	0.9858	0.8310	0.5874
2 year yield	0.9846	0.7304	0.4680	0.9860	0.8443	0.6329
3 year yield	0.9842	0.7335	0.4671	0.9865	0.8543	0.6660
4 year yield	0.9838	0.7359	0.4660	0.9871	0.8620	0.6910
5 year yield	0.9833	0.7378	0.4649	0.9875	0.8678	0.7101
10 year yield	0.9802	0.7422	0.4594	0.9887	0.8816	0.7565
20 year yield	0.9707	0.7394	0.4506	0.9893	0.8873	0.7775

Notes to Table: We present data and model-implied autocorrelations for each of the yields. Panel A is based on the data. Panel B is based on the component model. Panel C is based on the benchmark macro model. Panel D is based on the model with three latent factors.

Table 3.5 Variance Decompositions (in Percentages)

Panel A: Benchmark Macro Model

	Forecast Horizon	Macro Factors	
		Real Activity Growth	Inflation
3 month yield			
	1 month	13.62	86.38
	12 months	11.95	88.05
	60 months	7.48	92.52
	120 months	5.18	94.82
6 month yield			
	1 month	13.13	86.87
	12 months	11.51	88.49
	60 months	7.19	92.81
	120 months	4.97	95.03
3 year yield			
	1 month	9.22	90.78
	12 months	8.04	91.96
	60 months	4.95	95.05
	120 months	3.40	96.60
5 year yield			
	1 month	7.09	92.91
	12 months	6.16	93.84
	60 months	3.77	96.23
	120 months	2.58	97.42
10 year yield			
	1 month	4.04	95.96
	12 months	3.49	96.51
	60 months	2.11	97.89
	120 months	1.44	98.56
20 year yield			
	1 month	1.88	98.12
	12 months	1.62	98.38
	60 months	0.97	99.03
	120 months	0.66	99.34

Notes to Table: We list the contributions of different factors to the forecast variance of yields for different maturities and forecast horizons.

Panel B: Component Model							
		Short-Run Component		Long-Run Component			
	Forecast Horizon	Real Activity Growth	Inflation	Real Activity Growth	Inflation	Real Activity Growth Total	Inflation Total
3 month yield							
	1 month	17.21	81.27	0.03	1.49	17.24	82.76
	12 months	46.09	39.61	1.44	12.85	47.54	52.46
	60 months	36.59	17.07	16.83	29.51	53.42	46.58
	120 months	25.54	11.65	25.09	37.71	50.64	49.36
6 month yield							
	1 month	32.35	62.62	0.18	4.85	32.53	67.47
	12 months	58.14	20.48	2.58	18.80	60.72	39.28
	60 months	38.40	7.34	20.05	34.21	58.45	41.55
	120 months	25.67	4.80	27.88	41.65	53.55	46.45
3 year yield							
	1 month	52.76	33.68	1.05	12.51	53.81	46.19
	12 months	46.57	1.17	15.83	36.43	62.40	37.60
	60 months	22.23	0.30	32.64	44.83	54.87	45.13
	120 months	13.86	0.18	35.46	50.50	49.31	50.69
5 year yield							
	1 month	59.51	13.44	4.54	22.51	64.05	35.95
	12 months	31.63	0.51	25.12	42.74	56.75	43.25
	60 months	14.07	0.12	37.08	48.73	51.15	48.85
	120 months	8.66	0.07	37.08	54.18	45.74	54.26
10 year yield							
	1 month	54.18	7.49	9.30	29.02	63.49	36.51
	12 months	13.74	0.17	34.19	51.90	47.93	52.07
	60 months	5.83	0.04	37.91	56.22	43.74	56.26
	120 months	3.55	0.02	34.69	61.74	38.24	61.76
20 year yield							
	1 month	46.82	4.90	14.33	33.94	61.16	38.84
	12 months	5.28	0.06	29.57	65.09	34.85	65.15
	60 months	2.20	0.01	28.80	68.99	31.00	69.00
	120 months	1.31	0.01	24.73	73.96	26.03	73.97

Table 3.6 Return Predictability

	Benchmark Macro Model	**Component Model**	**PCs From Latent-Factor Model**
2 year yield	4.90	21.63	12.39
3 year yield	6.14	19.97	11.65
4 year yield	5.91	18.78	13.87
5 year yield	6.29	17.41	14.65
10 year yield	3.10	15.99	17.12
20 year yield	3.74	16.02	16.93

Notes to Table: We present the adjusted R^2s for the predictive regressions of annual excess returns on risk factors implied from different term structure models: inflation and real activity growth from the benchmark macro model, inflation, real activity growth, and their long-run components from the component model, and the first three principal components from the latent-factor model. We report R^2s in percentages for bonds with maturities of two to five, ten and twenty years.

Table 3.7 Forecast Comparisons (RMSE)

Panel A: 1-Month Ahead Forecast

	Model 1	Model 1*	Model 2	Model 3	Model 4	Model 4*	Model 5
3 month yield	129.12	121.40	165.00	64.49	79.71	64.16	60.78
6 month yield	88.60	92.69	162.12	52.10	66.88	51.64	50.23
1 year yield	56.93	62.74	155.01	49.16	72.31	48.71	48.43
2 year yield	45.43	46.14	143.62	43.17	61.20	42.86	42.79
3 year yield	44.18	47.61	134.73	44.20	56.94	42.06	39.84
4 year yield	43.50	45.69	128.38	45.25	51.06	41.75	38.96
5 year yield	42.45	43.86	123.57	42.46	47.55	40.63	36.30
10 year yield	43.98	45.03	129.22	53.41	69.57	51.22	28.49
20 year yield	46.70	46.82	116.71	55.92	71.77	55.41	25.85

Panel B: 3-Month Ahead Forecast

	Model 1	Model 1*	Model 2	Model 3	Model 4	Model 4*	Model 5
3 month yield	141.13	134.67	173.08	106.31	111.13	104.04	102.75
6 month yield	99.27	106.49	171.31	97.41	106.83	93.28	92.98
1 year yield	90.50	92.70	165.56	90.97	105.39	89.11	88.92
2 year yield	83.13	85.38	154.35	84.03	97.22	82.02	80.15
3 year yield	74.49	75.79	145.47	74.80	92.33	75.36	73.09
4 year yield	69.17	71.92	139.00	69.23	84.21	72.07	68.93
5 year yield	65.07	67.45	134.30	65.88	76.05	69.74	64.96
10 year yield	60.14	61.89	139.24	69.69	78.42	70.02	57.16
20 year yield	58.86	60.97	126.60	68.14	82.01	71.82	51.77

Panel C: 6-Month Ahead Forecast

	Model 1	Model 1*	Model 2	Model 3	Model 4	Model 4*	Model 5
3 month yield	190.11	181.12	184.33	176.64	196.94	179.86	139.64
6 month yield	180.62	190.75	183.13	179.43	199.36	176.73	136.51
1 year yield	176.91	180.21	178.21	171.37	187.89	169.78	127.28
2 year yield	155.79	159.02	166.62	153.73	165.00	157.97	114.11
3 year yield	139.86	141.69	157.64	138.82	147.71	134.24	103.56
4 year yield	129.20	134.24	150.86	130.20	136.57	130.57	97.68
5 year yield	120.29	124.29	146.30	123.88	128.13	121.42	92.00
10 year yield	116.24	116.52	152.09	120.90	120.93	117.71	85.10
20 year yield	107.43	109.47	139.46	109.53	112.28	107.81	77.07

Notes to Table: For each month t, we estimate the models using data for the previous 60 months, and forecast one month, three months, and six months ahead respectively. We provide the root mean square error (RMSE) of the forecast in basis points. Model 1: Component model. Model 1*: Unit root component model. Model 2: Benchmark macro model. Model 3: Model with three latent factors. Model 4: Model with macro variables and three latent factors. Model 4*: Model with macro variables and three latent factors estimated using two-step estimation method. Model 5: Random walk.

Table 3.A.1 Performance of Related Models (RMSE)

Panel A: In-Sample Model Fit

	DL*	DL	DLM*	DLM
3 month yield	187.34	172.12	112.57	109.02
6 month yield	128.03	105.59	78.67	76.24
1 year yield	92.16	76.12	72.16	65.02
2 year yield	76.12	62.34	57.46	52.90
3 year yield	73.90	60.26	46.02	43.12
4 year yield	71.13	46.35	43.46	40.40
5 year yield	65.12	48.46	46.10	42.24
10 year yield	61.24	54.36	46.31	43.13
20 year yield	69.24	59.29	50.12	45.12

Panel B: Out-of-Sample Model Fit

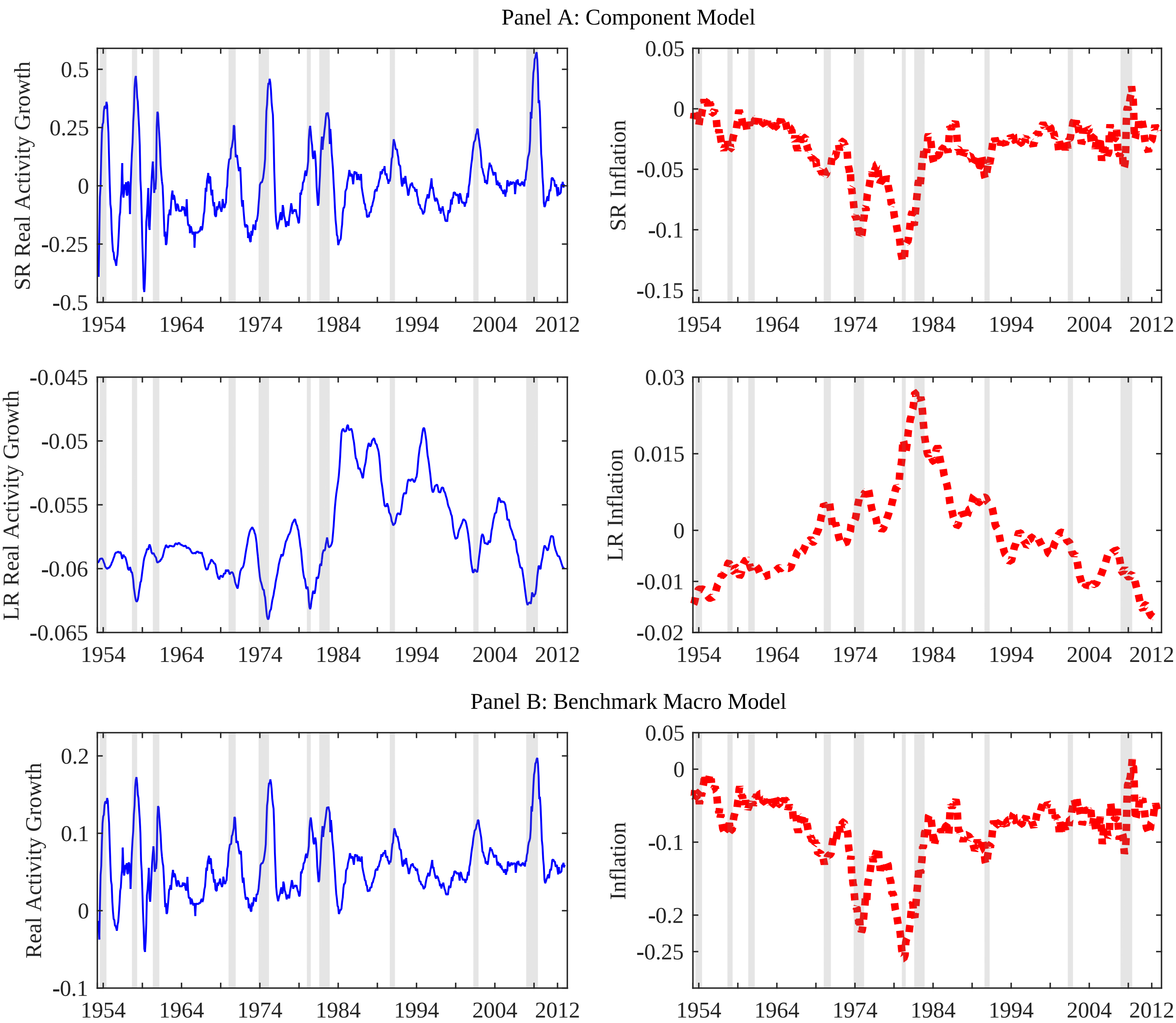
	1-Month Ahead Forecast				3-Month Ahead Forecast			
	DL*	DL	DLM*	DLM	DL*	DL	DLM*	DLM
3 month yield	380.53	324.12	249.71	237.12	504.73	407.88	387.72	356.04
6 month yield	259.84	206.98	128.73	118.31	306.20	239.34	196.68	186.39
1 year yield	182.11	160.44	109.36	101.60	235.85	201.43	147.38	129.94
2 year yield	151.38	117.12	101.72	98.35	207.99	152.43	138.68	121.16
3 year yield	157.09	136.01	102.42	88.25	213.75	169.81	129.64	112.94
4 year yield	152.11	123.05	100.94	90.04	207.95	170.37	124.72	116.29
5 year yield	129.02	115.15	100.47	89.31	184.04	152.89	130.93	113.78
10 year yield	124.86	111.87	101.89	90.39	162.93	137.80	124.23	110.65
20 year yield	128.29	103.75	100.73	93.86	168.93	140.63	121.85	109.03

Notes to Table: We provide the RMSE (in basis points) for several alternative models. DL refers to the model in Dewachter and Lyrio (2006). DL* refers to the same model without correlations between the real short rate and the macro variables. DLM refers to the model in Dewachter, Lyrio and Maes (2006). DLM* refers to the same model without correlations. For the out-of-sample fit, we estimate the models each month using data for the previous 60 months and forecast one month and three months ahead.

Figure 3.1 Model Fit

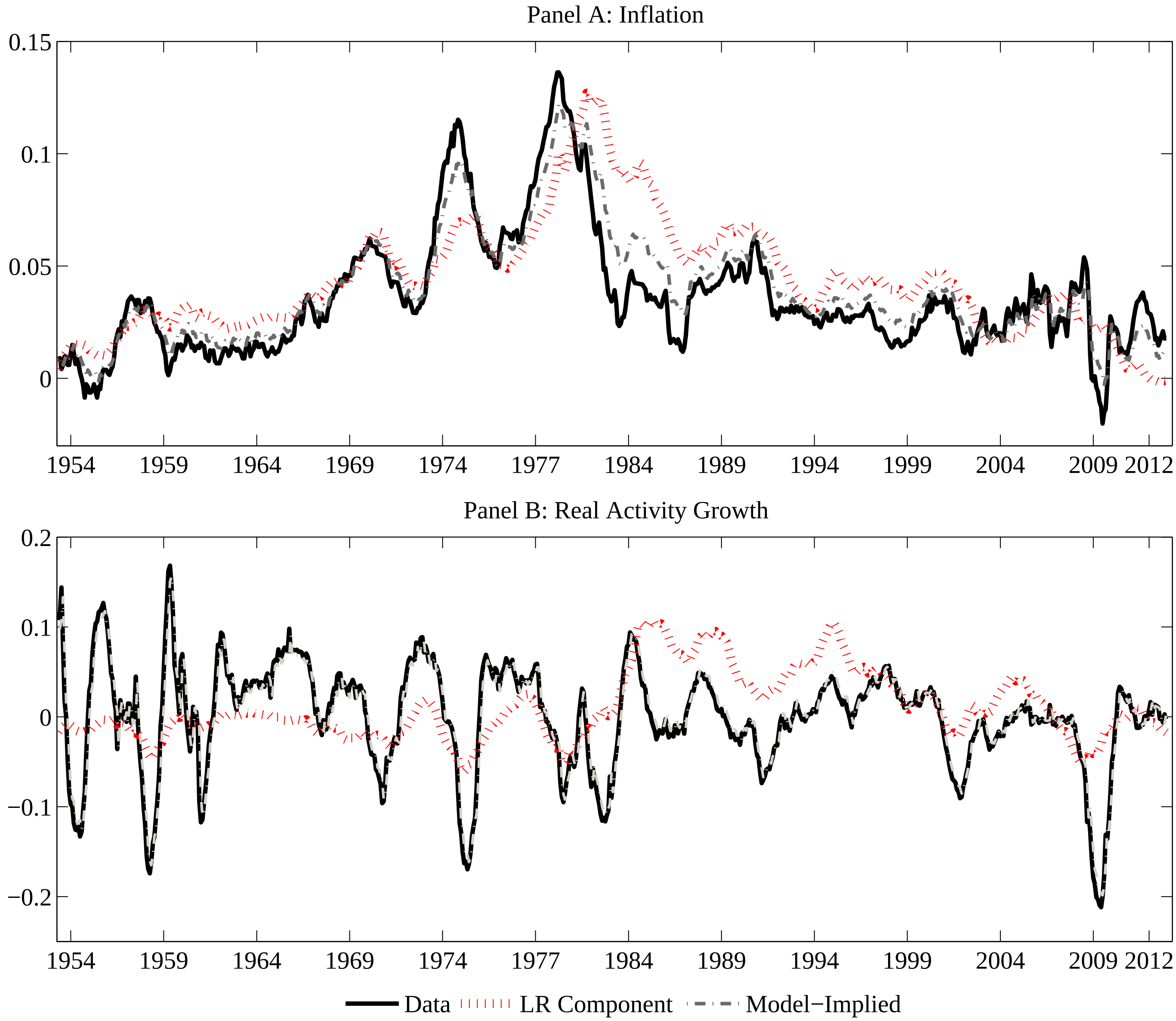
Notes to Figure: We plot data and fitted model yields for various maturities. The solid line shows the continuously compounded monthly zero-coupon bond yields. The dotted line shows the fitted yields from the benchmark macro model. The highlighted line shows the fitted yields from the component model.

Figure 3.2 The Market Price of Risk



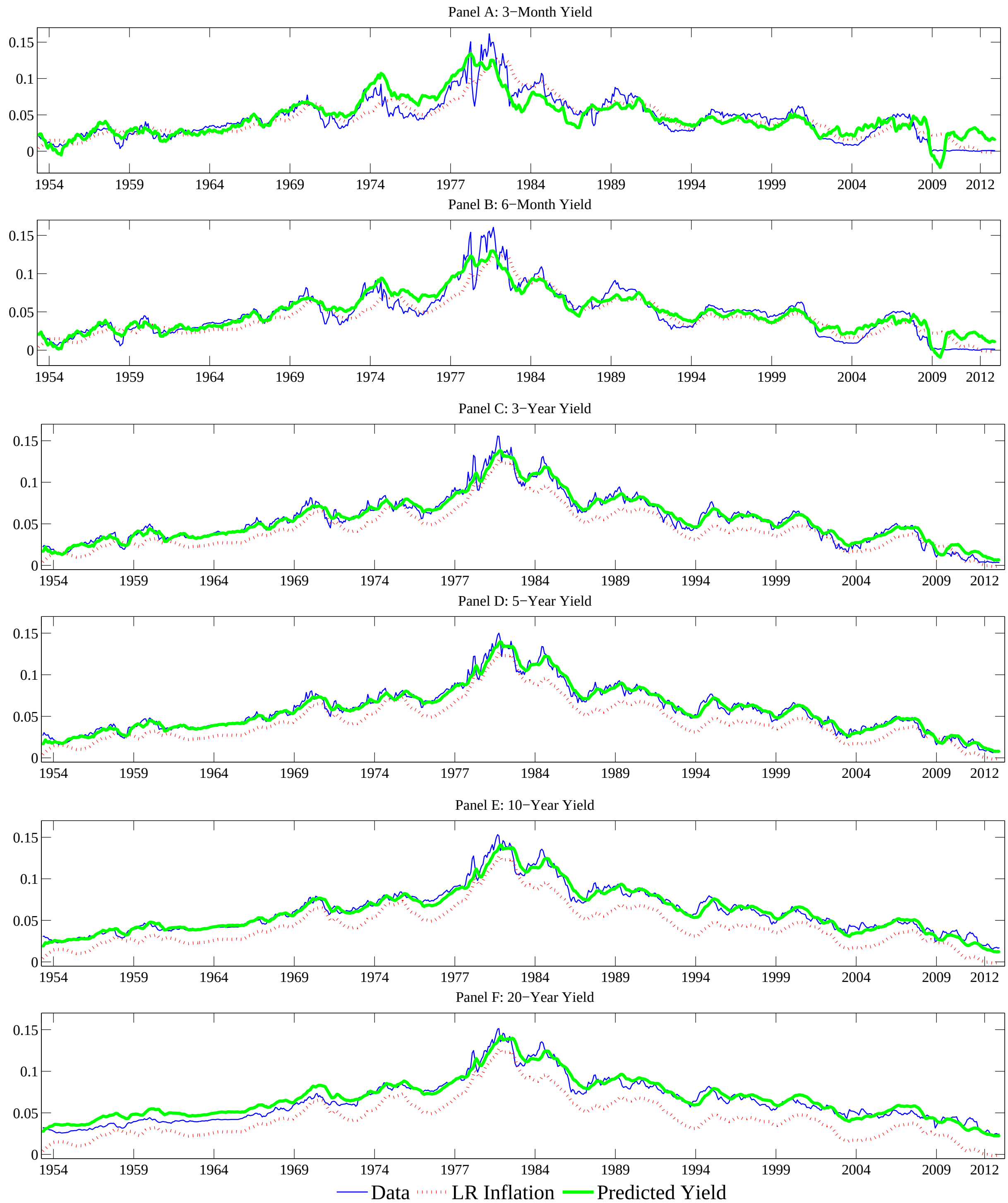
Notes to Figure: We plot the time-varying market price of risk $\lambda_0 + \lambda_1 F_t$, where F_t are the macro variables, real activity growth g_t and inflation π_t, and the time-varying intercepts of the macro variables, μ_{gt} and $\mu_{\pi t}$. The solid line shows the market price of risk for real activity growth, and the dotted line shows the market price of risk for inflation. Panel A presents the market price of risk for the short-run and long-run components from the component model. Panel B presents the market price of risk from the benchmark macro model. The shaded areas indicate recessions defined by the National Bureau of Economic Research (NBER).

Figure 3.3 Macro Variables and Long-Run Components



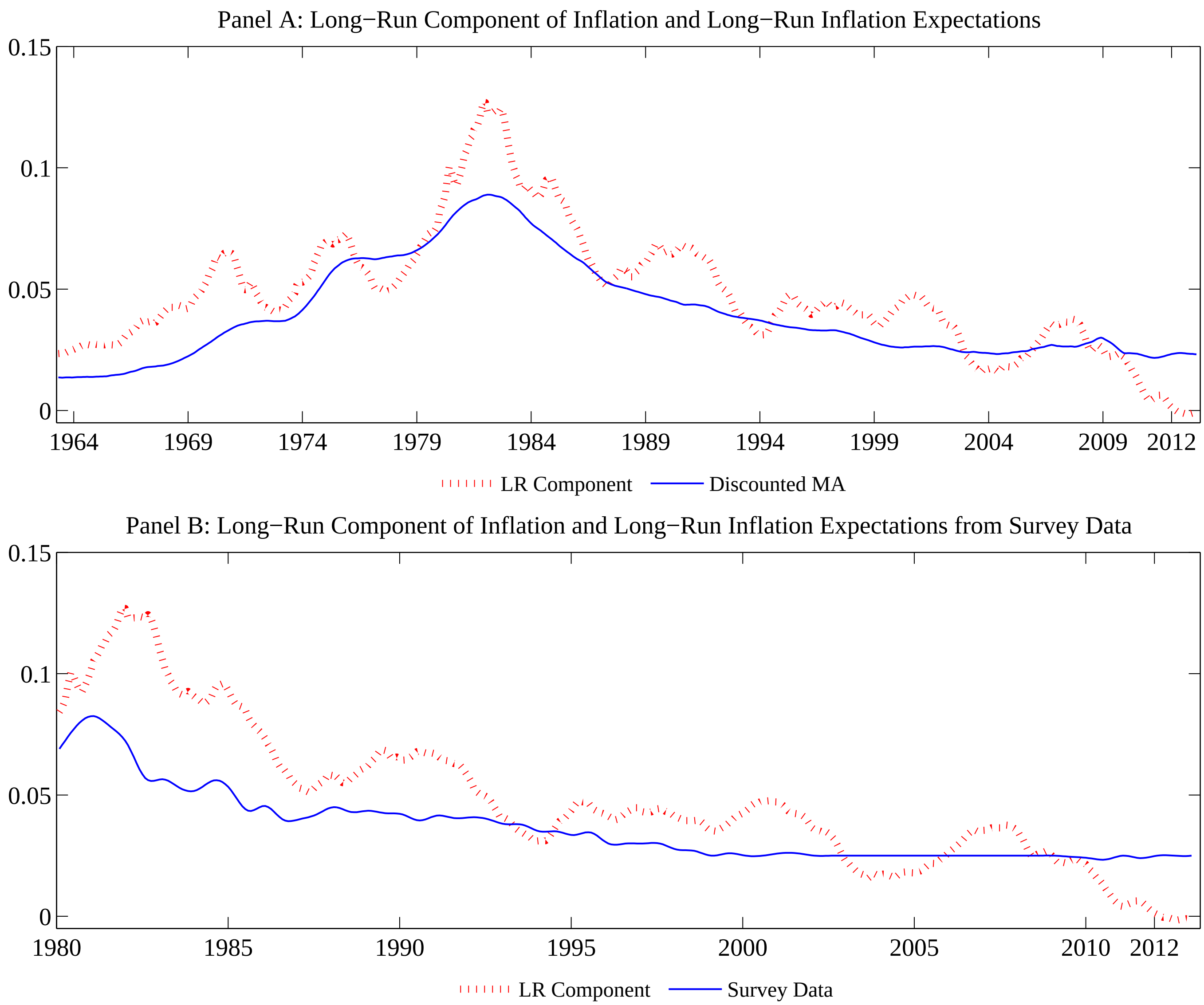
Notes to Figure: We show data and model-implied inflation and real activity growth together with the filtered long-run components. Panel A is for inflation and Panel B is for real activity growth. The solid line shows the macro variables, real activity growth g_t and inflation π_t. The dotted line shows the filtered long-run components, g_t^* and π_t^*, which are the long-run means of the macro variables. The dash-dot line shows the macro variables as fitted by the component model.

Figure 3.4 The Long-Run Component of Inflation and Yields



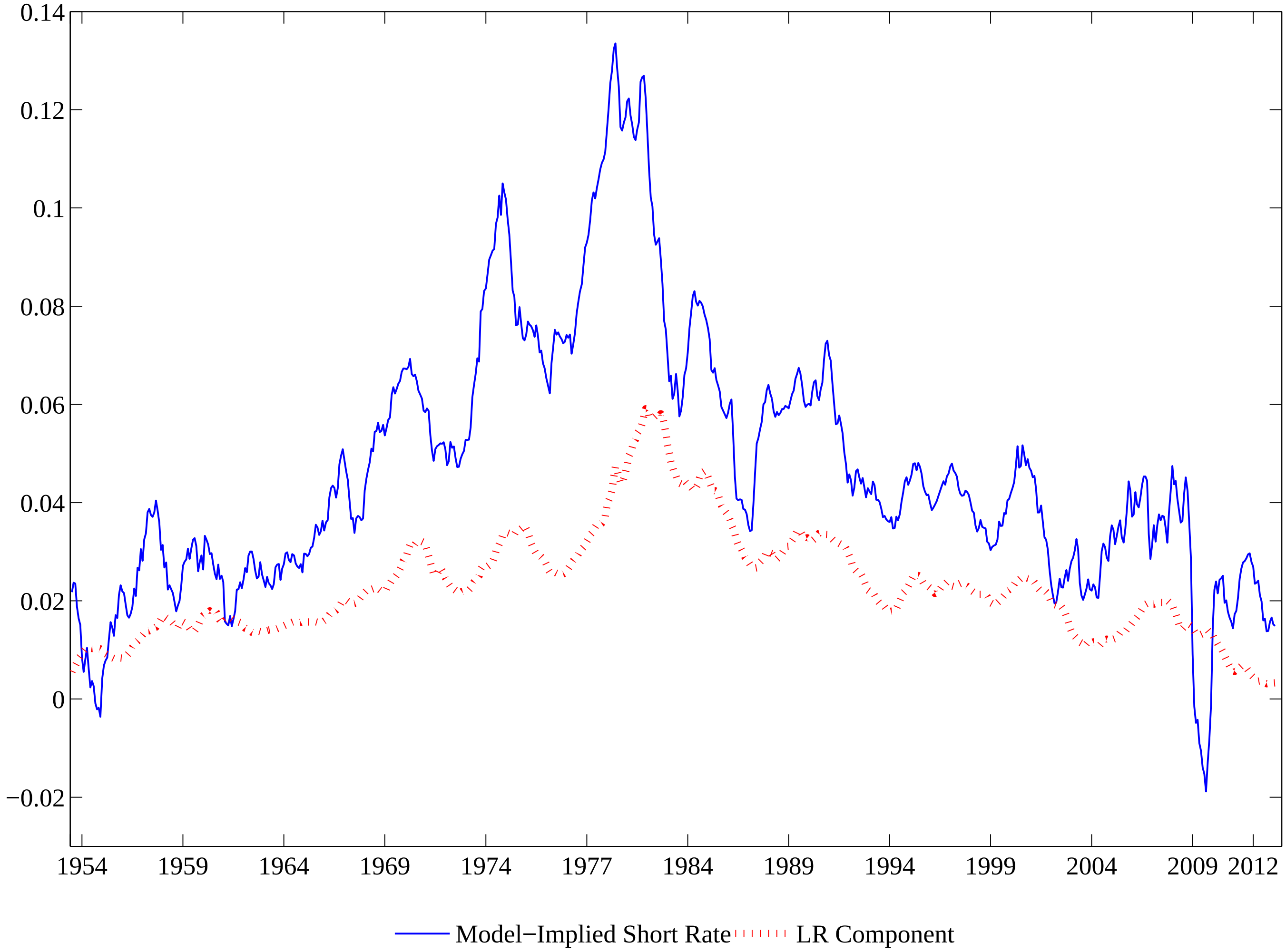
Notes to Figure: We show data and model-implied yields for the component model, as well as the long-run component of inflation. The solid line shows the continuously compounded monthly zero-coupon bond yields. The dotted line shows the long-run component of inflation π_t^*. The highlighted line shows the fitted yields from the component model.

Figure 3.5 The Long-Run Component of Inflation and Long-Run Inflation Expectations



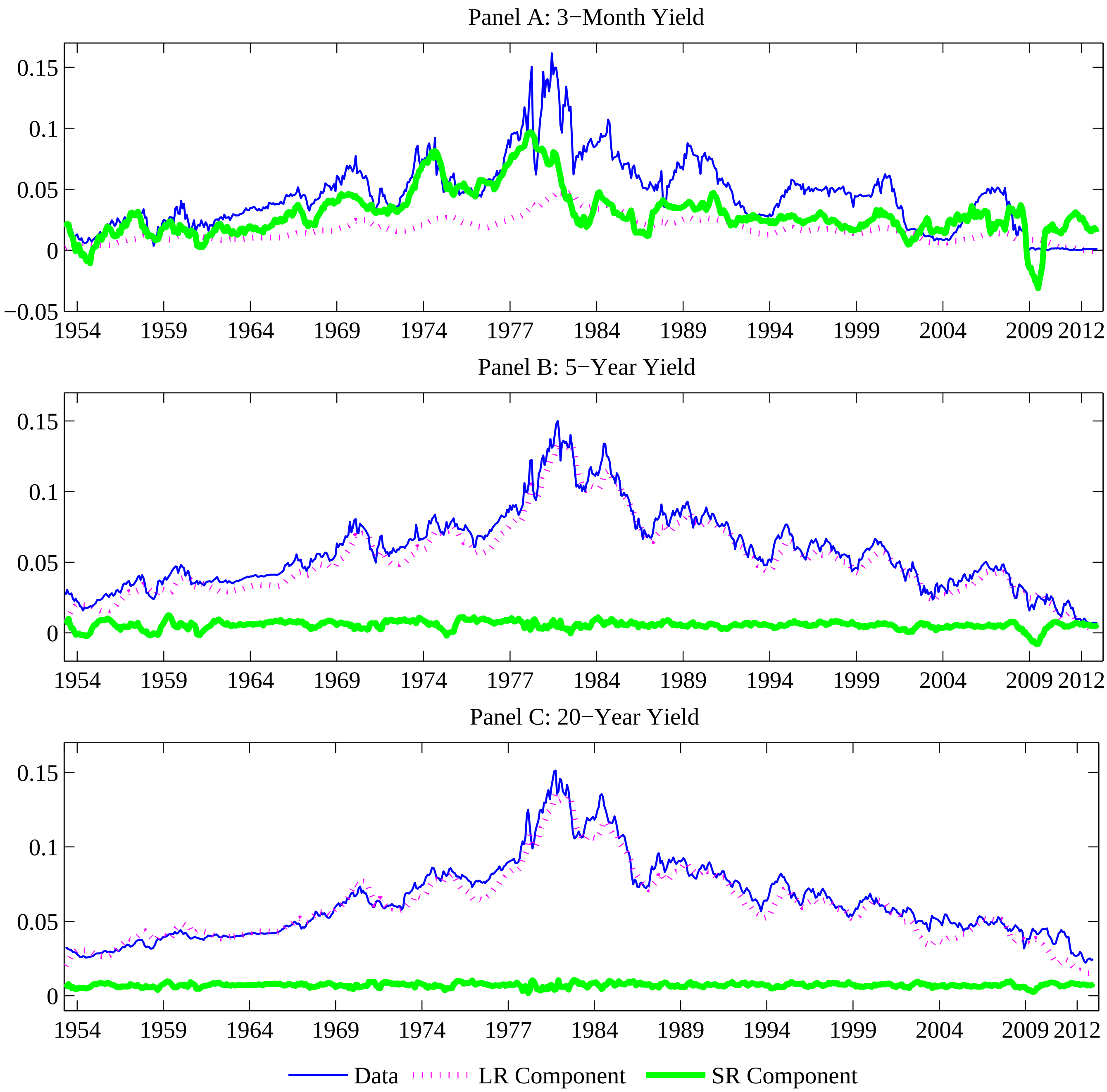
Notes to Figure: We compare the filtered long-run component of inflation π_t^* with long-run expectations of inflation. In Panel A, we construct a proxy for the long-run expectations of inflation as a discounted 10-year moving average of past CPI inflation. The sample period is from 1963:04 to 2012:12. In Panel B, we use an alternative measure of inflation expectations based on surveys. We use median ten-year CPI inflation expectations from the Livingston survey and from the Blue Chip Economic Indicators survey. The sample period is from 1979:12 to 2012:12. The dotted line shows the long-run component of inflation π_t^*, and the solid line shows long-run inflation expectations.

Figure 3.6 The Long-Run Component of the Short Rate



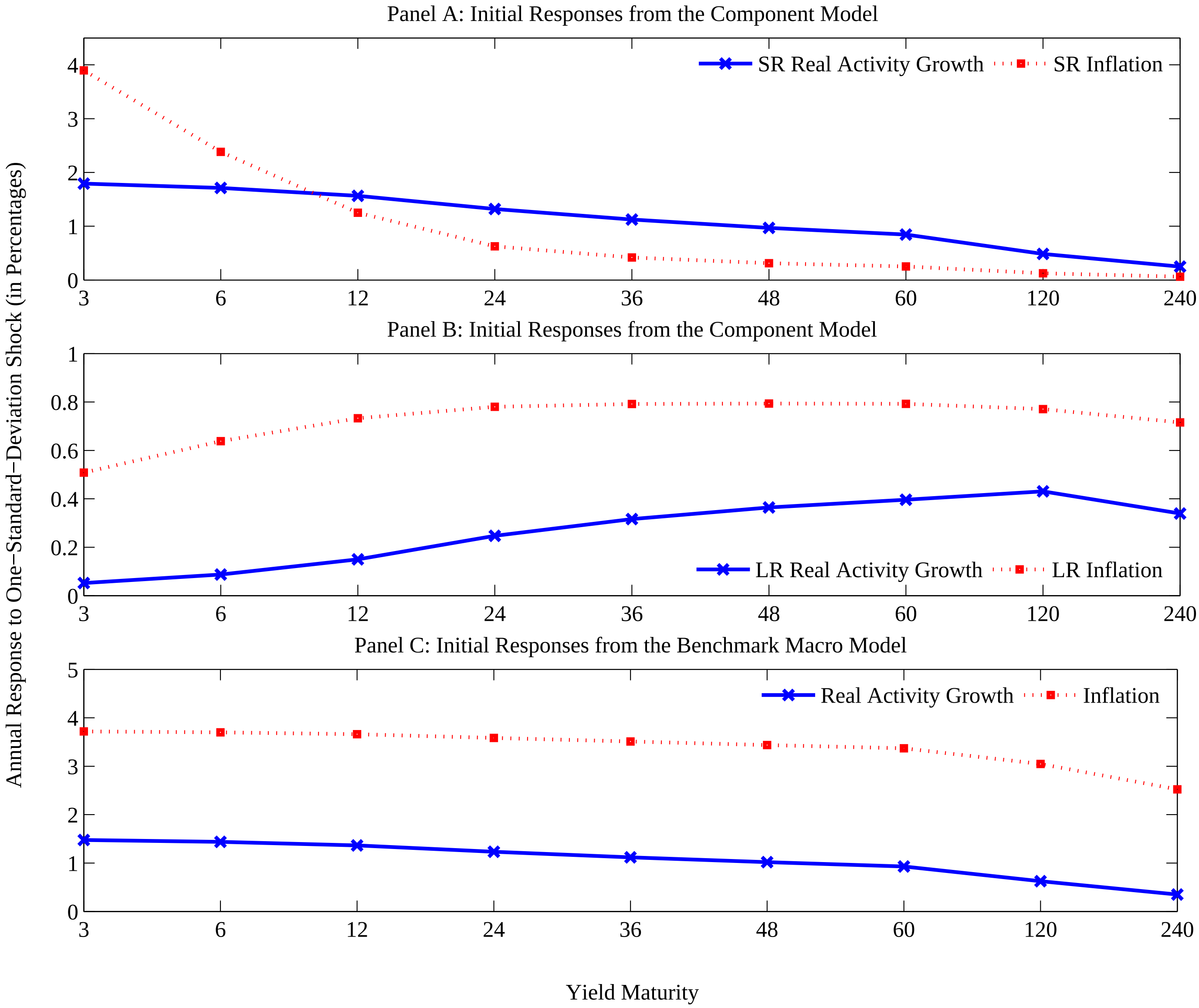
Notes to Figure: We show the model-implied short rate and its long-run component. The solid line shows the model-implied short rate, and the dotted line shows the long-run component of the short rate.

Figure 3.7 Long-Run Components of Yields



Notes to Figure: We plot yields and their model-implied long-run and short-run components for various maturities. The solid line shows the continuously compounded monthly zero-coupon bond yields. The dotted line shows the long-run components of the yields. The highlighted line shows the short-run components of the yields.

Figure 3.8 Initial Impulse Responses



Notes to Figure: We plot the initial responses to one-standard-deviation shocks to the factors, for various maturities. Panel A presents the initial response for different maturity yields for a one-standard-deviation shock to the short-run components of the macro variables. Panel B presents the initial responses for different maturity yields for a one-standard-deviation shock to the long-run components of the macro variables. The responses in Panels A and B are generated using the component model. Panel C presents the initial responses for a one-standard-deviation shock to inflation and real activity growth using the benchmark macro model. All responses are annualized and expressed in percentages.

Figure 3.9 Impulse Responses

Figure 3.9a: Component Model

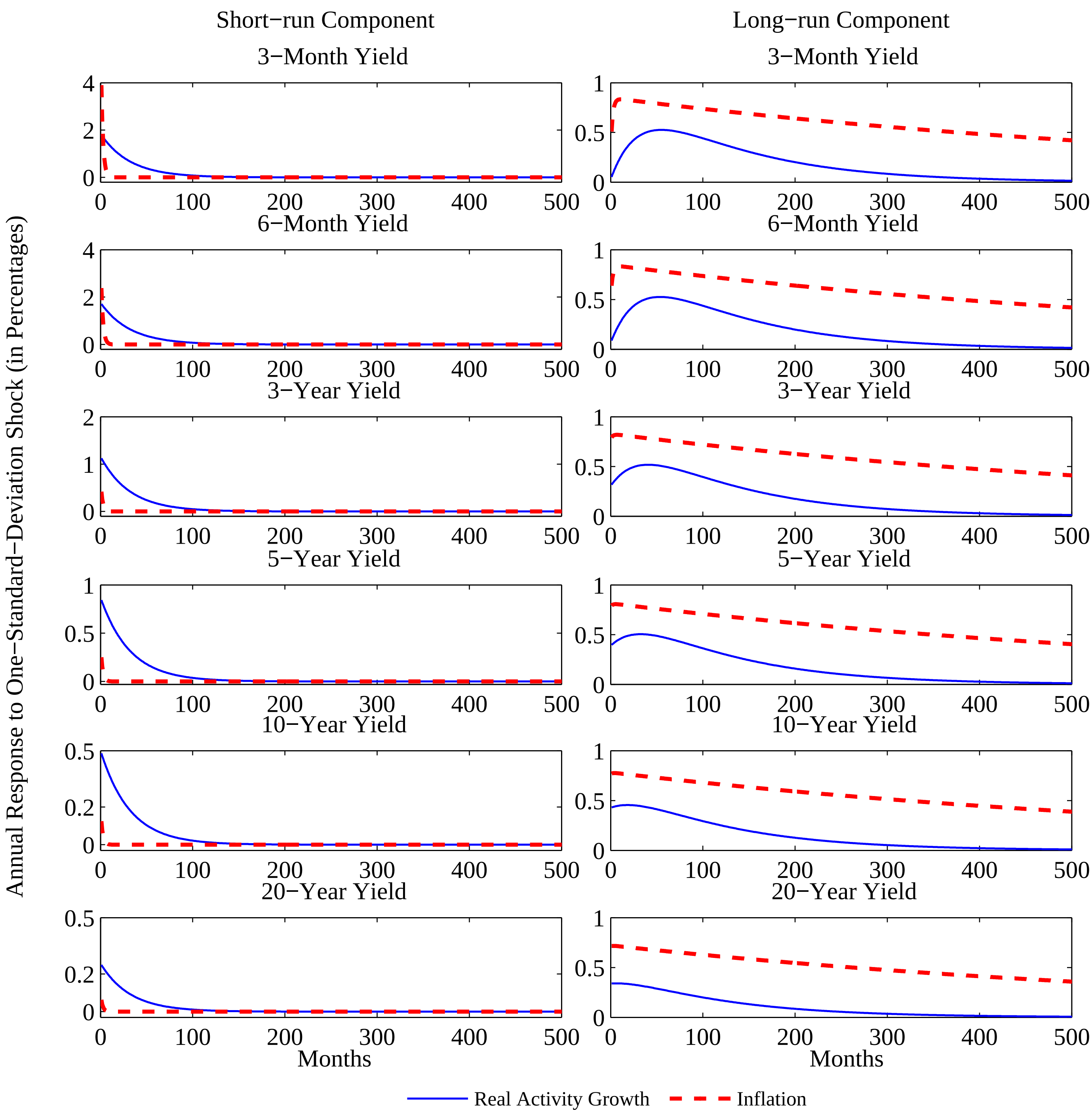
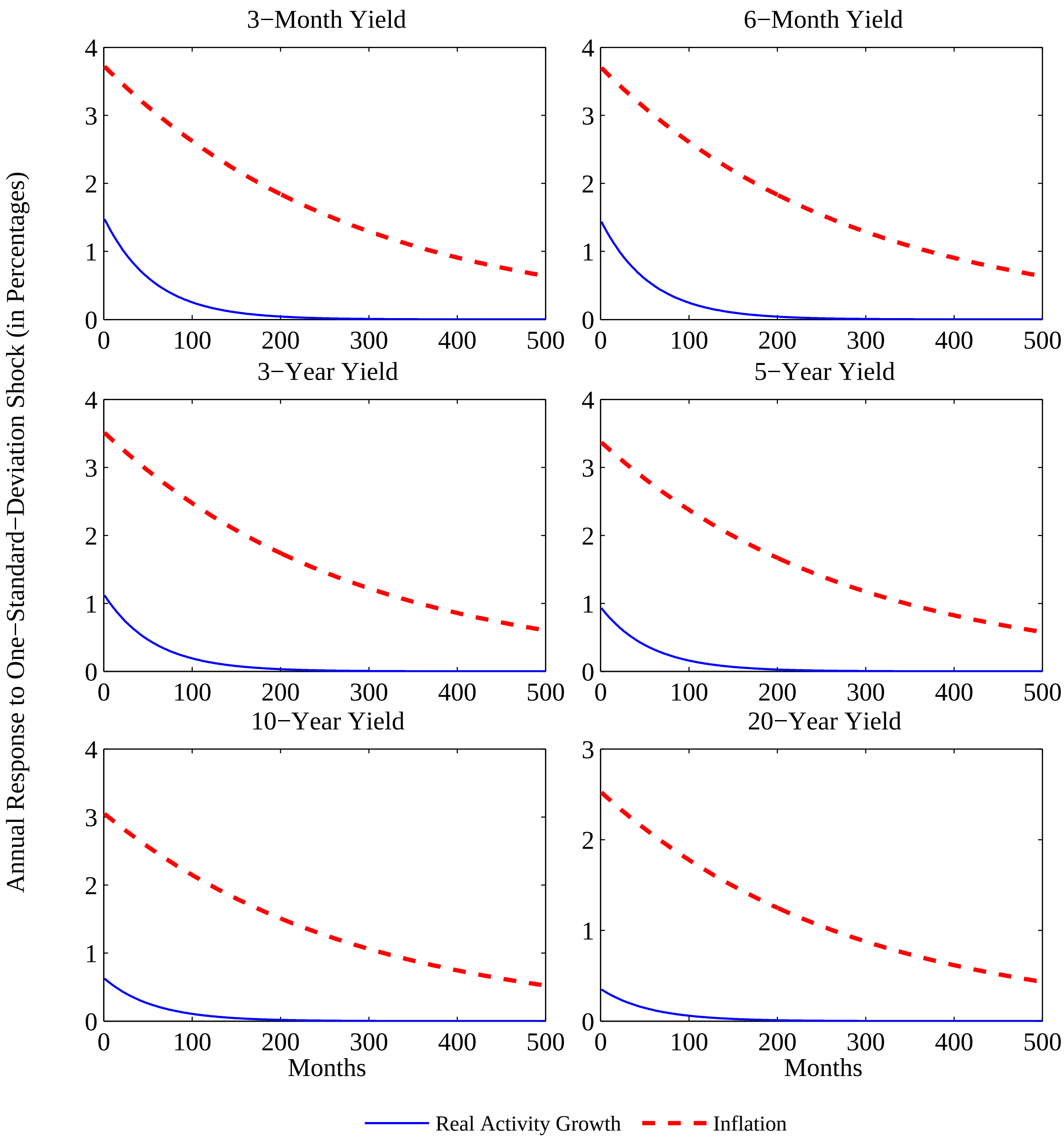
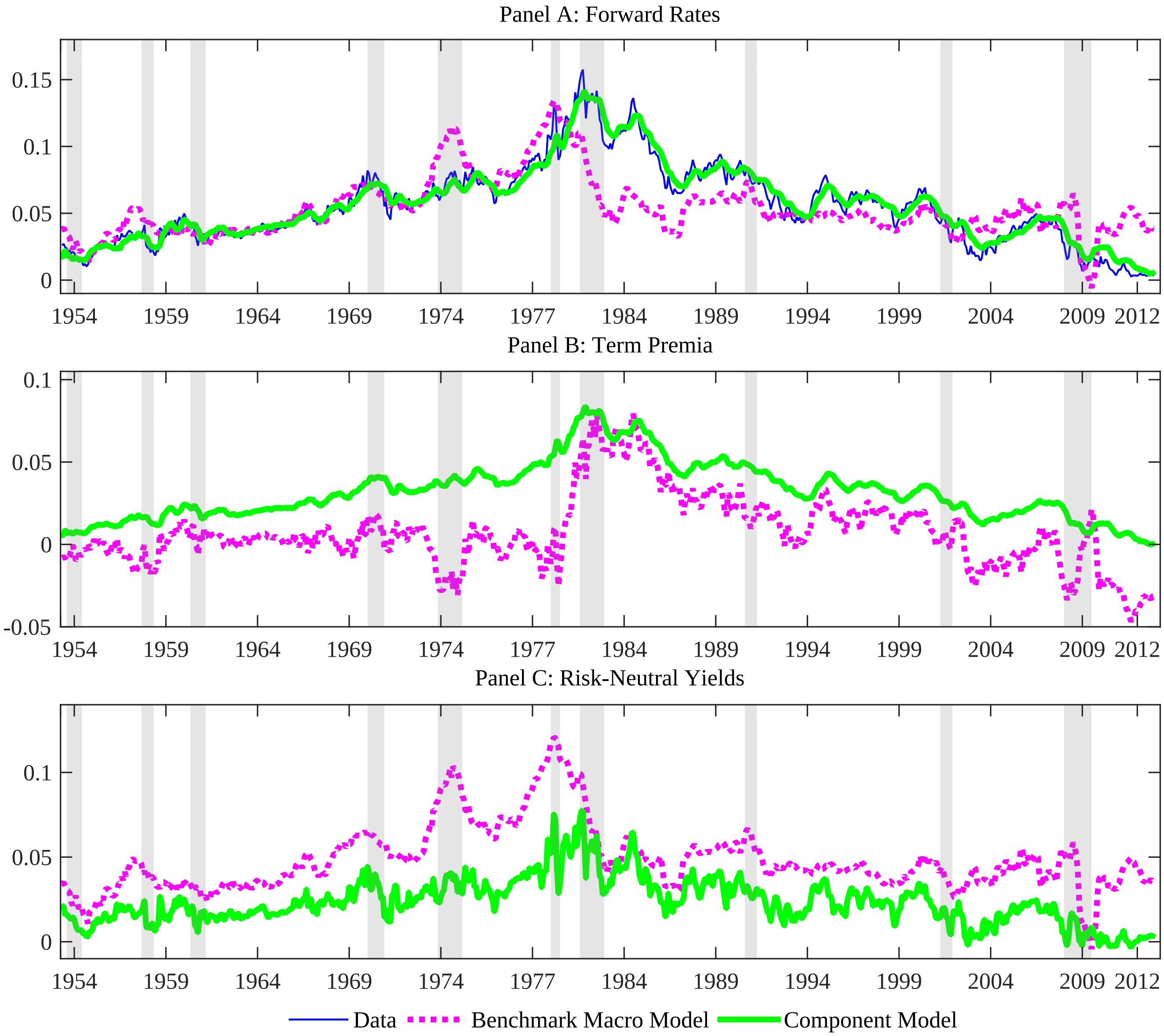


Figure 3.9b: Benchmark Macro Model



Notes to Figure: We plot the impulse responses of yields. Figure 9a presents the impulse responses of yields to long-run and short-run shocks obtained from the component model. The panels on the left in Figure 9a present the impulse response to short-run shocks for various maturities and different forecast horizons. The panels on the right present the impulse response to long-run shocks for various maturities and different forecast horizons. Figure 9b presents the impulse responses of yields with various maturities using the benchmark macro model. All impulse responses are scaled to correspond to a one-standard-deviation shock and are annualized and expressed in percentages.

Figure 3.10 Expectations of Forward Rates and Term Premia



Notes to Figure: Panel A plots data and model-implied five-to-ten-year forward rates. We decompose the forward rates into short-term interest rate expectations and term premium components. The forward term premium is defined as the difference between the forward rate one can lock in today for a five-year loan starting in five years, and the expected yield on a five-year bond purchased five years from now. Panel B plots the estimates of the term premia, and Panel C plots the estimates of the risk-neutral yields, i.e., the expectations component. The solid line shows the data. The dotted line shows the estimates from the benchmark macro model. The highlighted line shows the estimates from the component model. The shaded areas indicate recessions defined by the National Bureau of Economic Research (NBER).