Bayesian Sequential Inference and Machine Learning Applied to Yield Curve Forecasting

Tomasz Dubiel-Teleszynski

Department of Statistics

The London School of Economics and Political Science

This dissertation is submitted for the degree of

Doctor of Philosophy

December 2022

To my family

Declaration

I certify that the thesis I have presented for examination for the PhD degree of the London School of Economics and Political Science is my own work where I incorporated feedback from my supervisor Dr Konstantinos Kalogeropoulos and the external collaborator Dr Nikolaos Karouzakis. The copyright of this thesis rests with the author. Quotation from it is permitted, provided that full acknowledgement is made. This thesis may not be reproduced without my prior written consent. I warrant that this authorisation does not, to the best of my belief, infringe the rights of any third party. I declare that my thesis consists of about 50,000 words.

Tomasz Dubiel-Teleszynski

December 2022

Acknowledgements

First, I thank my supervisor Dr Kostas Kalogeropoulos without whom this thesis, and much more, would not become reality, and the external collaborator Dr Nikolas Karouzakis who led me through the realm of finance. I likewise thank my adviser Professor Qiwei Yao who has been always there for me when I sought additional guidance. I also thank the external examiners Professor Andreas Kaeck and Dr Nikolas Kantas for providing me with valuable feedback on this work.

Second, I thank everyone at the London School of Economics and Political Science, especially at the Department of Statistics, who supported me in the process. Professor Piotr Fryzlewicz and Professor Umut Cetin, Ian Marshall and Penny Montague, Sue Plater and Rose Harris, I will be forever in their debt.

Third, I thank my parents, grandparents, siblings and friends for endless motivation and continuous support. I also thank my mentors who turned friends, and vice versa, for the same and for their constructive critique. It has been invaluable having them all on my side throughout this exciting journey.

Last but not least, I thank Professor Tomasz Szapiro and Professor Aleksander Welfe who supervised me earlier at the Warsaw School of Economics and enabled me to pursue this route in the first place.

Abstract

We use machine learning, applied mathematics and techniques from modern statistics to refine Dynamic Term Structure Models. Specifically, we propose alternative, sequential Monte Carlo based solutions for the model selection problem, the extraction of unobserved factors from the yield curve and the identification of nonlinear associations between bond yields and the economy. The computational algorithms improve interest rate forecasts, significantly. In particular, they considerably facilitate the process of turning predictive performance into economic benefits to bond investors, verified within a dedicated portfolio optimization framework. Empirical results for the US Treasuries are especially important for companies and institutions navigating global fixed-income markets, including university endowments, pension funds and insurance corporations.

Table of Contents

List of Algorithms

List of Figures

List of Tables

Introduction

Motivations

The term structure of interest rates, or simply the yield curve, is an object of great importance for the financial markets. It is not an overstatement that, the ever changing dynamics of bond yields determines financial flows around the globe. This is particularly the case for the US Treasuries, where the underlying price developments are cautiously monitored by the participants in the fixed-income markets worldwide. However, equity investors, commodity traders and other major players in global finance also pay close attention to what is currently happening to bond yields, especially in the US. Yet, similar applies to international bond markets, such as Europe and Japan. On one hand, these are university endowments, pension funds, insurance corporations, investment banks, asset managers, sovereign wealth funds, hedge funds, family offices and finally retail investors, that rely on bonds in their portfolio strategies. For them, understanding interest rate developments is crucial for successful asset allocation. On the other hand, we have central banks and commercial banks where yield curve developments in relation to economy are of central importance for the effective transmission of monetary policy. Even more important is to identify the contribution of bond risk premia to longer term interest rates, what largely depends on the ability to accurately infer expectations of the future path for the short end of the yield curve. In what follows, we restrict our attention to understanding bond risk premia, specifically from the US term structure of interest

rates, and investigate how they react when changes in economic environment unfold. We then apply this knowledge to forecasting.

To successfully estimate and forecast bond risk premia it is crucial to account for no-arbitrage. It implies tight restrictions on the dynamics of risk compensation. Failure to impose such restrictions leads to absence of no-arbitrage and generation of artificially stable short rate expectations and highly volatile bond risk premia (Kim and Singleton, 2012; Bauer, 2018). There are two leading classes of econometric models which respect this limiting assumption. Namely, the arbitrage-free Dynamic Nelson-Siegel models (Diebold and Rudebusch, 2013) and Gaussian Affine Term Structure Models (ATSMs), or simply Dynamic Term Structure Models (DTSMs). In what comes next, we focus on the latter class. The early work on arbitrage-free affine yield-factor models of the term structure of interest rates in continuous time dates back to seminal papers by Duffie and Kan (1996) and later by Dai and Singleton (2000). Then, Duffee (2002) introduces the essentially affine class. Extensions and refinements follow (Ang and Piazzesi, 2003; Duffee, 2006, 2011). It is only until another seminal work, by Joslin et al. (2011), that DTSMs suffer from the ambiguity about their potential identification problems (Collin-Dufresne et al., 2008). We thus adopt their identification scheme throughout. Further extensions and refinements that ensue later include macro-finance DTSMs by Joslin et al. (2014) and related work by Bauer and Rudebusch (2016), as well as this by Creal and Wu (2015) on stochastic volatility. Up till then, the majority of results obtained with yields-only and macro-finance DTSMs are based on classical estimation methods, such as the Maximum Likelihood or even the Ordinary Least Squares (Adrian et al., 2013).

After Chib and Ergashev (2009) and Chib and Kang (2013), it takes until Bauer (2018) that DTSMs are again efficiently handled using Bayesian methods. Potential reason being that Bayesian estimation of DTSMs is often problematic and most MCMC samplers display very slow convergence rates and require a lot of tuning. The reason is that parameters enter the likelihood in a highly nonlinear fashion, what

requires various Metropolis-Hastings (MH) steps that are often inefficient (Bauer, 2018). In that latter work, Bauer specifically concentrates on risk parameters and shows that they can be sampled using a Gibbs step, what coupled with tailored MH steps, which require no tuning, for all the other parameters, leads to a fast MCMC algorithm with excellent convergence properties for a yields-only DTSM. In the following, we take this MCMC algorithm as a starting point and tailor it to form important building blocks in different sequential Monte Carlo (SMC) schemes we devise throughout. These draw from Chopin (2002, 2004) and Del Moral et al. (2006). Among the benefits provided by SMC is that, the IBIS algorithm, specifically, is an alternative choice when MCMC algorithms have poor mixing and convergence properties and, in general, proves more robust when the target posterior is challenging, for instance multimodal. The former issues are not necessarily material for yields-only DTSMs, however they may emerge when we consider the extensions.

Contributions

We provide competitive, SMC based solutions for the model selection problem, leveraging Stochastic Search Variable Selection (SSVS) of George and McCulloch (1993) and including also sparse priors, the extraction of unobserved factors from the yield curve, applying the well established Kalman filter, and the identification of nonlinear associations between bond yields and the economy, with Gaussian Processes (GP), see Rasmussen and Williams (2006).

The first chapter explores the statistical and economic importance of restrictions on the dynamics of risk compensation (Bauer, 2018), from the perspective of a real-time Bayesian learner that predicts bond excess returns using a DTSM. We propose a novel methodological framework, which successfully handles sequential model search and parameter estimation over the restriction space landscape in real time, allowing investors to revise their beliefs when new information arrives, thus informing their

asset allocation and maximizing their expected utility. Our setup provides the entire predictive density of returns, allowing us to revisit the evident puzzling behaviour between statistical predictability and meaningful out-of-sample economic benefits to bond investors. Empirical results reveal the importance of different sets of restrictions across market conditions and monetary policy actions. Furthermore, our results reinforce the argument of sparsity in the market price of risk specification, since we find strong evidence of out-of-sample predictability only for those models that allow for level risk to be priced. Most importantly, such statistical evidence is turned into economically significant utility gains, across prediction horizons. The sequential version of the SSVS scheme developed offers an effective diagnostic, as it monitors potential changes in the importance of different risk prices over time and provides further improvement during periods of macroeconomic uncertainty, where results are more pronounced.

There has been some interest, in the recent literature, in unspanned latent factors affecting bond yields (Duffee, 2011). Thus, in the second chapter, we present a generalized modelling framework for Gaussian ATSMs with unspanned unobserved components, that allows us to study their dynamics and use them for prediction purposes. We develop a suitable sequential Monte Carlo inferential and prediction scheme, that takes advantage of the linear state space structure of the model by incorporating the associated Kalman filter. The sequential scheme can also be used along with a dynamic portfolio optimization framework to assess the potential of predictions to generate economic value. The extracted unspanned latent factors are also contrasted with observed macroeconomic indices. Our regression results provide evidence that relationships between them are statistically significant. It makes the former an even more attractive alternative to the latter for ATSMs, given that data revisions do not apply in this case. Moreover, we show that the models are actually quite competitive in terms of predictive performance when compared to yields-only ATSMs.

The importance of unspanned macroeconomic information for DTSMs has also been discussed in the recent literature (Joslin et al., 2014). To our best knowledge earlier studies considered linear relationships between macroeconomic indices and the real-world measure in DTSMs. Only recently, Bianchi et al. (2021) provide evidence on the significance of nonlinearities for detecting information relevant to forecasting bond risk premia. Hence, in the third chapter, we propose a generalized modelling setup for DTSMs, which allows nonlinear associations between bond yields and the economy, and we apply it to forecasting. Specifically, we construct a custom SMC estimation and forecasting scheme, where we introduce GP priors to model nonlinearities. Both individual and composite macroeconomic variables are considered. These include the core inflation and the real economic activity. Sequential scheme we propose is then coupled with dynamic portfolio optimization framework, to assess the potential of excess return forecasts for generating economic benefits to bond investors. We contrast predictions from nonlinear models with those stemming from an application of a linear model which is conceptually closest to Joslin et al. (2014). Eventually, we find that the former lead to significant increases in economic value across maturities, when specific parts of unspanned nonlinear macros, which are hidden from the yield curve, are nonlinear in nature as well.

Chapter 1

Sequential Learning and Economic Benefits from Dynamic Term Structure Models

1.1 Introduction

1.1.1 Restrictions and Puzzling Behaviour

Accurately estimating and forecasting bond risk premia, in real time, is of central economic importance for the transmission mechanism of monetary policy as well as for investors' portfolio strategies. Even more important is understanding and identifying the contribution of risk premia to longer term interest rates, which largely depends on our ability to accurately infer expectations of the future path for the short end of the yield curve[1]. To successfully do so, it is essential to account for no-arbitrage, which implies restrictions on the cross-sectional and time series dynamics of the term structure (see, Joslin et al. (2011) and Bauer (2018)). The latter are largely exploited in related literature by Dynamic Term Structure Models

[1]See, Kim and Wright (2005) and Cochrane and Piazzesi (2009), among others, for studies that attempt to decompose forward rates into expectations of short rates and risk premia.

(DTSMs), which impose tight restrictions on the dynamics of risk compensation, an essential component of the models. Failure to impose such restrictions, as in the unrestricted maximally flexible model widely used by almost all existing studies, leads to absence of no-arbitrage and, as such, to the generation of artificially stable short rate expectations and highly volatile risk premia (Kim and Orphanides, 2012; Bauer, 2018).

From a statistical or machine learning viewpoint, the problem of restrictions may be thought of as overfit. If more parameters than needed are used to capture the signal of the market price of risk, it becomes more likely to capture noise rather than systematic patterns, thus leading to poor predictive performance. With this in mind, we introduce a general methodological framework that utilises Bayesian inference and forecasting simultaneously, while allowing for model and parameter uncertainty to be incorporated in a sequential manner. Setting up in the context of Bauer (2018), we construct a sequential learning scheme following the principles of Chopin (2002) and Del Moral et al. (2006). The modelling approach successfully handles sequential model searches over the space of all possible restrictions in real time, allowing investors to monitor changes in the importance of particular restrictions over certain time periods and monetary actions. The developed setup is then used to predict bond returns and explore the out-of-sample statistical and economic importance of restrictions, across maturities and prediction horizons.

The importance of the market price of risk specification, and the associated restrictions related to it, has been extensively studied in earlier research (see, Dai and Singleton (2000), Duffee (2002), Ang and Piazzesi (2003), Kim and Wright (2005)), which has mainly focused on imposing ad hoc[2] zero restrictions on the parameters governing the dynamics of the risk premia[3]. This practice, however, has been criticized

[2]Ad hoc restrictions, based on setting risk premia parameters to zero, are used in Dewachter and Lyrio (2006), and Rudebusch and Wu (2008), among others. Furthermore, the route of imposing prior restrictions is followed by Ang et al. (2007).

[3]A common practice used is to, first, estimate an unrestricted maximally flexible model, and at a second step, to re-estimate it by setting to zero those parameters that have large standard errors. According to Bauer (2018), such an approach often leads to the wrong model.

(see, Kim and Singleton (2012) and Bauer (2018)), since it raised concerns about, first, the joint significance of the constraints, second, the magnitude of the associated standard errors[4] and, third, the failure to provide meaningful economic justification for the estimated parameters and the resulting state variables. Only recently, a few studies have investigated more systematic approaches to imposing restrictions on the dynamics of risk compensation[5]. In particular, Cochrane and Piazzesi (2009) and Duffee (2011) introduce tight restrictions, driven by prior empirical analysis, while Joslin et al. (2014) select zero restrictions[6] based on different information criteria[7]. An alternative approach is followed by Chib and Ergashev (2009), who impose strong prior restrictions such that the yield curve is (on average) upward sloping[8], an assumption that is empirically and economically plausible. Bauer (2018) proposes a Bayesian econometric model choice framework to estimate and identify DTSMs with restrictions on risk prices, using draws from three alternative model selection samplers[9]. The framework is then used to assess the economic implications of restrictions, in particular, the persistence of interest rates, and its effect on short rate expectations and term premia components. Although the literature has noted the importance of restrictions, yet, no study has, so far, addressed and quantified their statistical and economic importance, out-of-sample[10]. Most importantly, there is no prior evidence as to how restrictions 'react' to changes in the monetary environment, considering that existing studies on monetary policy effects (see, Piazzesi et al. (2006),

[4] According to Kim and Singleton (2012), it is unclear how small the associated standard errors have to be in order to set a parameter to zero.

[5] See, Cochrane and Piazzesi (2009) for a 4-factor affine model, Joslin et al. (2014) for an unspanned macro-finance DTSM, and Duffee (2011) and Bauer (2018) for yields-only versions of DTSMs, among others.

[6] Together with zero restrictions, Joslin et al. (2014) impose an additional one on the largest eigenvalue of the feedback matrix, by assuming equality under the risk-neutral and historical measures.

[7] The three criteria selected are Akaike (1998), Hannan and Quinn (1979), and Schwarz (1978).

[8] Chib and Kang (2013) also use similar prior restrictions to reflect the meaningful assumption of a positive term premium.

[9] Samplers proposed are the Gibbs Variable Selection, the SSVS, and the Reversible-jump Markov Chain Monte Carlo.

[10] Empirical tests in Duffee (2011) suggest that the choice of no-arbitrage restrictions does not influence the out-of-sample performance of the models, given that they produce forecasts with indistinguishable differences.

Ang and Longstaff (2011), and Orphanides and Wei (2012)) suggest that restrictions selected based on the in-sample process may not be economically plausible around periods of monetary policy shifts, interventions, or under fragile economic conditions.

Our approach differs from previous studies, allowing us to overcome a number of important challenges and offers several advantages. First, in a similar style to Fulop et al. (2019) but tailored to the context of this chapter, it allows us to update the estimates and predictive density as new data arrive, without the need to rerun everything from scratch. Second, it allows for potentially more powerful prediction techniques, such as ensemble learning and Bayesian model averaging, to be implemented. In this chapter we develop a sequential version of the Stochastic Search Variable Selection (SSVS) scheme (can also be used for Gibbs Variable Selection) that allows incorporating model and parameter uncertainty in a sequential manner. This is of particular importance taking into account that investors often face model uncertainty, which highlights the need for a framework that is capable of monitoring, identifying and adjusting models in real time. Third, it provides a more robust alternative to the Markov Chain Monte Carlo (MCMC) sampler and model choice algorithms of Bauer (2018) and Gargano et al. (2019), as its sequential setup naturally provides inference in a parallel way that can potentially overcome issues such as poor mixing, slow convergence properties, and multi-modalities. For example, while we adopt the canonical setup of Joslin et al. (2011) where the vector of unobserved risk factors is rotated such that state variables are linear combinations of observed yields that allows for the construction of a very efficient MCMC sampler[11] as in Bauer (2018), there is room for more challenging setups, as in the case of Chib and Ergashev (2009). Also, the more global nature of sequential Monte Carlo can be of help in the model choice context, where the exploration of the model space may be challenging.

[11] In our setup, the Bayesian learner is learning and revising her beliefs about the unknown parameters from a directly observed state vector.

1.1.2 Bond Predictability and Economic Benefits

The modelling and evaluation framework developed in this chapter is also put in action to revisit two important and economically relevant empirical questions. First, are there any yields-only DTSMs that are capable of consistently predicting bond risk premia? Second, can the predictability of bond excess returns, if any, be turned into economic benefits for bond investors?

Some recent literature (e.g. Duffee (2011), Barillas (2011), Thornton and Valente (2012), Adrian et al. (2013), Joslin et al. (2014)), suggests that yields-only DTSMs cannot capture the predictability of bond risk premia, since the required information to predict premia is not spanned[12] by the cross section of yields, implying that more (mainly unspanned) factors are needed. Duffee (2011) implements a five-factor yields-only model which is capable of capturing the hidden information in the bonds market, while Wright (2011), Barillas (2011), Joslin et al. (2014) and Cieslak and Povala (2015) use measures of macroeconomic activity to predict bond excess returns. In a recent study however, Bauer and Hamilton (2018) cast doubt on prior conclusions, suggesting that the evidence on variables other than the three yield factors predicting excess returns is not convincing. Along the same direction are the results of Sarno et al. (2016) and Feunou and Fontaine (2018), who implement extended versions of yields-only DTSM and argue that their approaches help those models capture the required predictability of excess returns in the bonds market. A similar conclusion is reached by Bauer (2018), who estimates a DTSM under alternative risk price restrictions, and Bianchi et al. (2021), who utilize information inferred from different machine learning methodologies. However, those studies, with the exception of Bianchi et al. (2021), either do not consider the out-of-sample economic performance (as in Duffee (2011), Bauer (2018), and Feunou and Fontaine (2018), etc.) or do not

[12]The spanning hypothesis suggests that the yield curve contains (i.e. spans) all relevant information required to forecast future yields and excess returns. Unspanned (or hidden) factors are factors that are not explained by the yield curve, while at the same time, are useful for predicting risk premia (see, Cochrane and Piazzesi (2005), Duffee (2011), Joslin et al. (2014) and Cieslak (2018)).

fully explore potential economic benefits for bond investors when compared to the non-predictability (constant risk premia) Expectations Hypothesis (EH) benchmark, which is one of the empirical questions we target in this chapter. In fact, existing literature on economic value points towards a negative answer (see, Della Corte et al. (2008), Thornton and Valente (2012), Sarno et al. (2016), and Gargano et al. (2019)[13]). In particular, using a dynamic mean-variance allocation strategy, Della Corte et al. (2008), Thornton and Valente (2012), and Sarno et al. (2016) find that statistical predictability is not turned into superior portfolio performance when compared to the EH benchmark[14]. Only recently, however, Bianchi et al. (2021) report positive utility gains for models that utilize information in the yield curve only, when Neural Networks (NN) are used. In that respect, our empirical analysis complements several prior related studies and contributes to the debate on whether statistical and economic significance move in the same direction.

Our evaluation framework consists of two stages, initially targeting monthly excess bond returns across different prediction horizons, and evaluating the predictive performance using metrics such as the out-of-sample R^2 of Campbell and Thompson (2008) and the log scores (LS) as in Geweke and Amisano (2010). To investigate the economic value of the out-of-sample excess return forecasts generated by alternative models, we construct a dynamically rebalanced portfolio as in Della Corte et al. (2008) and Thornton and Valente (2012), for an investor with power utility preferences, and compute standard metrics such as certainty equivalence returns (CER) (see, Johannes et al. (2014) and Gargano et al. (2019), among others).

Our results help us infer a host of interesting conclusions in the context of the US market. We confirm existing literature (e.g. Thornton and Valente (2012) and

[13] Gargano et al. (2019) find some significant economic gains for investors at specific (mainly longer) maturities when macroeconomic factors are used. However, according to Ghysels et al. (2018) and Fulop et al. (2019), such economic benefits vanish when fully revised macroeconomic information is replaced by real-time data. Our approach is not reliant on macroeconomic data and, as such, our analysis is independent of the debate between 'fully-revised' vs. 'real-time' macros.

[14] Contradicting conclusions are reached by Fulop et al. (2019) and Gargano et al. (2019) who test models that utilize information from macroeconomic factors.

Sarno et al. (2016)) in that yields-only DTSMs with some or no restrictions on the risk premia are not capable of predicting well generally and naturally do not produce economic value either. Nevertheless, the situation is reversed when heavy restrictions are placed. More specifically, we find that only level risk is priced (in line with Cochrane and Piazzesi (2009)) since only those models which allocate one or two non-zero risk premia parameters solely on the level factor perform consistently well. We identify two such models and also confirm that they are able to improve out-of-sample portfolio performance and earn economically meaningful utility gains, when compared to the EH benchmark, for investors who dynamically rebalance their portfolio. These findings are in contrast to Thornton and Valente (2012) and Sarno et al. (2016), who find that statistical evidence of out-of-sample predictability is not turned into economic value, for bond investors who utilize information from the yield curve only, and in line with Bianchi et al. (2021), who argue that statistical significance is turned into economic gains when NN with two or three layers are used.

Comparing the performance of alternative model specifications over time reveals the importance of different sets of restrictions across monetary policy actions and market conditions. In particular, the importance of parameters that capture variation in the slope factor increases during periods of curve steepening and decreases when the curve flattens. This effect is reversed for risk price parameters capturing shocks due to the level factor. Those effects are reflected in bond return predictability which is also linked to the economic cycle, since investors demand compensation for carrying more slope risk in recessions rather than in expansions. In particular, we document that both statistical predictability and economic benefits are substantially higher following the post-crisis recession period, when the US market experienced high uncertainty and interest rates hit the zero-lower bound. This also reveals a stronger relationship between the slope of the yield curve and bond excess returns during periods of macroeconomic uncertainty. Finally, our results document a substantial improvement in the CER values, across the maturity spectrum, with a

clear tendency for larger gains at the short end of the curve, where short-maturity bonds offer higher economic gains. This is in contrast to the pre-crisis expansion period, where predictability is much lower and bond returns almost halve in value, with CER appearing to be higher at longer maturities. A clear message coming out of our analysis is that focusing on a single model could be quite problematic.

1.1.3 Outline

The remainder of this chapter is organized as follows. Section 1.2 describes the modelling framework. Section 1.3 presents the sequential learning and forecasting procedure along with the framework for assessing the predictive and economic performance of models. Section 1.4 discusses the data and the sample period used and presents the best models inferred through the sequential SSVS scheme. Section 1.5 discusses the results both in terms of predictive performance and economic value. Finally, Section 1.6 concludes the chapter by providing some relevant discussion.

1.2 Dynamic Term Structure Model, Likelihood, and Restrictions

In this section we briefly describe the adopted model and the associated likelihood function in order to set up the notation and formulate our research question explicitly. More details can be found in Joslin et al. (2011) where this framework was introduced.

1.2.1 Canonical Model

The model belongs to the no-arbitrage class of Affine Term Structure Models (ATSMs) (see, Ang and Piazzesi (2003) and Cochrane and Piazzesi (2005)), under which the one period risk-free interest rate r_t[15] is assumed to be an affine function of a $(N \times 1)$

[15]Working with monthly data implies that r_t is the 1-month yield.

vector of state variables X_t, namely

$$r_t = \delta_0 + \delta_1' X_t$$

where δ_0 is a scalar and δ_1 is a $(N \times 1)$ vector. In Gaussian ATSMs, the physical (real-world) probability measure $\mathbb{P}$ is assumed to be a first-order Gaussian Vector Autoregressive (VAR) process

$$X_t = \mu + \Phi X_{t-1} + \Sigma \varepsilon_t \tag{1.1}$$

where $\varepsilon_t \sim N(0,\ I_N)$, Σ is a $(N \times N)$ lower triangular matrix, μ is a $(N \times 1)$ vector and Φ is a $(N \times N)$ matrix. Lack of arbitrage implies the existence of a pricing kernel $\mathcal{M}_{t+1}$ defined as

$$\mathcal{M}_{t+1} = \exp(-r_t - \frac{1}{2}\lambda_t'\lambda_t - \lambda_t'\varepsilon_{t+1}) \tag{1.2}$$

with λ_t being the time-varying market price of risk which is assumed to be affine in the state X_t[16]

$$\lambda_t = \Sigma^{-1}\left(\lambda_0 + \lambda_1 X_t\right)$$

where λ_0 is a $(N \times 1)$ vector and λ_1 is a $(N \times N)$ matrix. If we assume that the pricing kernel $\mathcal{M}_{t+1}$ prices all bonds in the economy and we let P_t^n denote the time-t price of an n-period zero-coupon bond, then the price of the bond is computed from $P_t^n = E_t\left(\mathcal{M}_{t+1} P_{t+1}^{n-1}\right)$ where $E_t(\cdot)$ denotes expectation given the information available up to time t. Hence, it follows that bond prices are exponentially affine

[16]This is the 'essentially-affine' specification introduced in Duffee (2002). Existing studies have proposed alternative specifications for the market price of risk, such as the 'completely-affine' model of Dai and Singleton (2000), the 'semi-affine' model of Duarte (2004), and the 'extended-affine' model of Cheridito et al. (2007). See Feldhütter (2016) for a useful comparison of the models.

functions of the state vector (see, Duffie and Kan (1996))

$$P_t^n = \exp(A_n + B_n' X_t), \quad n = 1, \ldots, J$$

with loadings, A_n being a scalar and B_n a $(N \times 1)$ vector, satisfying the following recursions

$$\begin{aligned} A_{n+1} &= A_n + B_n'(\mu - \lambda_0) + \frac{1}{2} B_n' \Sigma \Sigma' B_n - \delta_0 \\ B_{n+1} &= (\Phi - \lambda_1)' B_n - \delta_1 \end{aligned}$$

with $A_0 = 0$ and $B_0 = 0_{N \times 1}$. Pricing kernel $\mathcal{M}_{t+1}$ in (1.2) lets define the change of probability measure (via Girsanov theorem) from real-world $\mathbb{P}$ in (1.1) to risk-neutral $\mathbb{Q}$. This implies that the pricing $\mathbb{Q}$-dynamics of the state vector are given by

$$X_t = \mu^{\mathbb{Q}} + \Phi^{\mathbb{Q}} X_{t-1} + \Sigma \varepsilon_t^{\mathbb{Q}} \tag{1.3}$$

where $\mu^{\mathbb{Q}} = \mu - \lambda_0$, $\Phi^{\mathbb{Q}} = \Phi - \lambda_1$ and $\varepsilon_t^{\mathbb{Q}} \sim N(0,\ I_N)$. The continuously compounded n-period yield y_t^n is also an affine function of the state vector

$$y_t^n = -\frac{\log P_t^n}{n} = A_{n,X} + B_{n,X}' X_t \tag{1.4}$$

where the loading scalar $A_{n,X}$ and the loading $(N \times 1)$ vector $B_{n,X}$ are calculated using the above recursions, as $A_{n,X} = -A_n/n$ and $B_{n,X} = -B_n/n$. Similarly, for the continuously compounded $(J \times 1)$ vector of yields y_t, we have that

$$y_t = A_X + B_X X_t \tag{1.5}$$

where $(J \times 1)$ vector A_X and $(J \times N)$ matrix B_X contain the model-implied loadings of yields on risk factors, specifically $A_{n,X}$ are elements of vector A_X and $B_{n,X}$ are transposed rows of matrix B_X.

In theory, it is possible to specify the likelihood function based on equations (1.1) and (1.4) but, in practice, estimation and identification of these formulations has proved to be challenging (see, Ang and Piazzesi (2003), Ang et al. (2007), Chib and Ergashev (2009), Duffee and Stanton (2012), Hamilton and Wu (2012), and Bauer (2018)), especially when ATSMs are expressed in terms of an unobserved X_t. Additional normalizing restrictions need to be imposed to ensure identifiability, such as the canonical setup of Joslin et al. (2011) that we adopt here. More specifically, X_t is rotated to be a linear combination of the observed yields, and as such, perfectly priced by the no-arbitrage restrictions. In particular, we rotate X_t to match the first N principal components (PCs) of observed yields

$$\mathcal{P}_t = W y_t = W A_X + W B_X X_t \tag{1.6}$$

with W being the $(N \times J)$ matrix that contains the PCs' loadings. Following common practice, we consider the case of $N = 3$, noting that the first three extracted PCs are typically sufficient to capture most of the variation in the yield curve and often correspond to its level, slope and curvature, respectively (Litterman and Scheinkman, 1991). We also adhere to the identification scheme in Proposition 1 by Joslin et al. (2011), where the short rate is the sum of the state variables, namely $r_t = iX_t$ with i being a vector of ones, and the parameters $\mu^{\mathbb{Q}}$ and $\Phi^{\mathbb{Q}}$ of the $\mathbb{Q}$-dynamics are given as $\mu^{\mathbb{Q}} = [k_\infty^{\mathbb{Q}}, 0, 0]$ and $\Phi^{\mathbb{Q}} = \text{diag}(g^{\mathbb{Q}})$, where $g^{\mathbb{Q}}$ denotes a $(N \times 1)$ vector containing the real and distinct eigenvalues of $\Phi^{\mathbb{Q}}$ [17].

Statistical inference can proceed using the observations $Y = \{y_t, \mathcal{P}_t : t = 0, 1, \ldots, T\}$. The likelihood factorizes into two parts stemming from the $\mathbb{P}$ and $\mathbb{Q}$ respectively. In order to specify the latter, henceforth denoted as $\mathbb{Q}$ likelihood, the affine transformation of (1.6) is applied to (1.3) to obtain the dynamics of $\mathcal{P}_t$

[17] Alternative specifications for the eigenvalues are considered in Joslin et al. (2011); however, real eigenvalues are found to be empirically adequate.

under $\mathbb{Q}$

$$\mathcal{P}_t = \mu_{\mathcal{P}}^{\mathbb{Q}} + \Phi_{\mathcal{P}}^{\mathbb{Q}} \mathcal{P}_{t-1} + \Sigma_{\mathcal{P}} \varepsilon_t^{\mathbb{Q}} \tag{1.7}$$

where the risk-neutral measure parameters $\mu_{\mathcal{P}}^{\mathbb{Q}}$, $\Phi_{\mathcal{P}}^{\mathbb{Q}}$ and $\Sigma_{\mathcal{P}}$ are following

$$\begin{aligned}
\mu_{\mathcal{P}}^{\mathbb{Q}} &= W B_X \mu^{\mathbb{Q}} + (I_N - \Phi_{\mathcal{P}}^{\mathbb{Q}}) W A_X \\
\Phi_{\mathcal{P}}^{\mathbb{Q}} &= W B_X \Phi^{\mathbb{Q}} (W B_X)^{-1} \\
\Sigma_{\mathcal{P}} &= W B_X \Sigma
\end{aligned}$$

In a similar manner, the yield equation (1.5) can be rewritten as a function of $\mathcal{P}_t$[18]

$$y_t = A_{\mathcal{P}} + B_{\mathcal{P}} \mathcal{P}_t \tag{1.8}$$

where the loadings $A_{\mathcal{P}}$ and $B_{\mathcal{P}}$ are derived accordingly as

$$A_{\mathcal{P}} = A_X - B_X (W B_X)^{-1} W A_X \tag{1.9}$$

$$B_{\mathcal{P}} = B_X (W B_X)^{-1} \tag{1.10}$$

and $A_{\mathcal{P}}$ is a $(J \times 1)$ vector with elements $A_{n,\mathcal{P}}$, whereas $B_{\mathcal{P}}$ is a $(J \times N)$ matrix which transposed rows are $(N \times 1)$ vectors $B_{n,\mathcal{P}}$.

Note that in (1.5) and (1.8), yields are assumed to be observed without any measurement error. Nevertheless, an N-dimensional observable state vector cannot perfectly price $J > N$ yields, and as such, we further assume that the $(J - N)$ bond yields used in the estimation are observed with independent $N(0,\ \sigma_e^2)$ measurement errors. An equivalent way to formulate this is to write

$$y_t = A_{\mathcal{P}} + B_{\mathcal{P}} \mathcal{P}_t + e_t \tag{1.11}$$

[18] According to Duffee (2011), outside of knife-edge cases, the matrix $(W B_X)$ is invertible, and as such, $\mathcal{P}_t$ contains the same information as X_t.

and to consider the dimension of e_t as effectively being $(J-N)\times 1$. Letting $W_\perp$ denote a basis of the null space of W, the measurement error assumption can also be expressed as

$$W_\perp e_t \sim N(0,\ \sigma_e^2 I_{J-N})$$

where $(W_\perp e_t)$ is a $(J-N)\times 1$ vector (Bauer, 2018).

Under the new observable state vector $\mathcal{P}_t$, the one period short rate r_t is also an affine function of $\mathcal{P}_t$ given as

$$r_t = \delta_{0\mathcal{P}} + \delta'_{1\mathcal{P}}\mathcal{P}_t \tag{1.12}$$

with

$$\delta_{0\mathcal{P}} = \delta_0 - \delta'_1(WB_X)^{-1}WA_X$$
$$\delta'_{1\mathcal{P}} = (WB_X)^{-1}\delta_1$$

and the market price of risk specification becomes accordingly

$$\lambda_t = \Sigma_{\mathcal{P}}^{-1}\left(\lambda_{0\mathcal{P}} + \lambda_{1\mathcal{P}}\mathcal{P}_t\right)$$

where

$$\lambda_{0\mathcal{P}} = WB_X\lambda_0 - WB_X\lambda_1(WB_X)^{-1}WA_X \tag{1.13}$$
$$\lambda_{1\mathcal{P}} = WB_X\lambda_1(WB_X)^{-1} \tag{1.14}$$

Finally, in order to specify the $\mathbb{P}$ likelihood we note that $\mathbb{P}$-dynamics of $\mathcal{P}_t$ are of equivalent form to (1.7) but with ε_t instead of $\varepsilon_t^{\mathbb{Q}}$ and

$$\mu_{\mathcal{P}}^{\mathbb{P}} = \mu_{\mathcal{P}}^{\mathbb{Q}} + \lambda_{0\mathcal{P}} \tag{1.15}$$
$$\Phi_{\mathcal{P}}^{\mathbb{P}} = \Phi_{\mathcal{P}}^{\mathbb{Q}} + \lambda_{1\mathcal{P}} \tag{1.16}$$

where $\lambda_{0\mathcal{P}}$ is a $(N \times 1)$ vector and $\lambda_{1\mathcal{P}}$ is a $(N \times N)$ matrix specified in (1.13) and (1.14), reflecting the market price of risk in $\mathcal{P}_t$ terms.

1.2.2 Likelihood and Risk Price Restrictions

The joint likelihood (conditional on the initial point $\mathcal{P}_0$) can now be written as

$$\begin{aligned} f(Y|\theta) = &\left\{\prod_{t=0}^{T} f^{\mathbb{Q}}(y_t|\mathcal{P}_t, k_\infty^{\mathbb{Q}}, g^{\mathbb{Q}}, \Sigma_{\mathcal{P}}, \sigma_e^2)\right\} \times \\ &\left\{\prod_{t=1}^{T} f^{\mathbb{P}}(\mathcal{P}_t|\mathcal{P}_{t-1}, k_\infty^{\mathbb{Q}}, g^{\mathbb{Q}}, \lambda_{0\mathcal{P}}, \lambda_{1\mathcal{P}}, \Sigma_{\mathcal{P}})\right\} \end{aligned} \tag{1.17}$$

where the $\mathbb{Q}$-likelihood components $f^{\mathbb{Q}}(\cdot)$ are given by (1.11) and capture the cross-sectional dynamics of the risk factors and the yields, whereas $\mathbb{P}$-likelihood components $f^{\mathbb{P}}(\cdot)$ capture the time-series dynamics of the observed risk factors. The parameter vector is set to $\theta = (\sigma_e^2, k_\infty^{\mathbb{Q}}, g^{\mathbb{Q}}, \lambda_{0\mathcal{P}}, \lambda_{1\mathcal{P}}, \Sigma_{\mathcal{P}})$.

Note that in the case of all entries in $\lambda_{0\mathcal{P}}$, $\lambda_{1\mathcal{P}}$ being non-zero, also known as the maximally flexible model, the mapping between θ and $\tilde{\theta} = (\sigma_e^2, k_\infty^{\mathbb{Q}}, g^{\mathbb{Q}}, \mu_{\mathcal{P}}^{\mathbb{P}}, \Phi_{\mathcal{P}}^{\mathbb{P}}, \Sigma_{\mathcal{P}})$ is 1-1. This allows for the following equivalent likelihood specification

$$f(Y|\tilde{\theta}) = \left\{\prod_{t=0}^{T} f^{\mathbb{Q}}(y_t|\mathcal{P}_t, k_\infty^{\mathbb{Q}}, g^{\mathbb{Q}}, \Sigma_{\mathcal{P}}, \sigma_e^2)\right\} \times \left\{\prod_{t=1}^{T} f^{\mathbb{P}}(\mathcal{P}_t|\mathcal{P}_{t-1}, \mu_{\mathcal{P}}^{\mathbb{P}}, \Phi_{\mathcal{P}}^{\mathbb{P}}, \Sigma_{\mathcal{P}})\right\} \tag{1.18}$$

The fact that both likelihoods (1.17) and (1.18) are equivalent suggests that, under the maximally flexible model, the parameters $k_\infty^{\mathbb{Q}}$ and $g^{\mathbb{Q}}$ are effectively estimated only from the $\mathbb{Q}$ likelihood, in other words based solely on cross-sectional information and without reference to the real-world dynamics[19]. This is no longer the case if one or more entries of the $\lambda_{0\mathcal{P}}$, $\lambda_{1\mathcal{P}}$ matrices are set to zero, in other words if restrictions are imposed. Since the mapping from θ to $\tilde{\theta}$ is no longer 1-1, with $\tilde{\theta}$ now being of higher dimension, we can only use the likelihood specification of (1.17), which allows utilizing time series information in the $\mathbb{Q}$ parameters estimation, and as such, limiting the effect of the no-arbitrage cross-sectional constraints.

[19] According to Joslin et al. (2011), the ordinary least squares estimates of parameters $\mu_{\mathcal{P}}^{\mathbb{P}}$ and $\Phi_{\mathcal{P}}^{\mathbb{P}}$ are almost identical to those estimated using maximum likelihood.

Nevertheless, this raises the issue of how to choose between the 2^{N+N^2} possible sets of restrictions in the $\lambda_{0\mathcal{P}}$, $\lambda_{1\mathcal{P}}$ matrices; e.g. in the case of $N = 3$ there are 4,096 distinct sets of restrictions. Bauer (2018) suggests using Bayesian model choice, aiming to maximize the model evidence of each restriction specification. In this chapter we propose choosing the restriction set with the optimal predictive performance among all possible restriction sets. Models that are optimal in the Bayesian sense, i.e. achieving the highest model evidence, are typically parsimonious and therefore exhibit good predictive performance. In a related argument, Fong and Holmes (2020) show that model evidence is formally equivalent with exhaustive leave-p-out cross-validation combined with the log posterior predictive scoring rule. Hence, it will not be surprising if the same set of restrictions was obtained from both approaches; in fact this is the case in data from the US market as we illustrate in Section 1.5. Nevertheless, this is not always guaranteed to be the case and, in situations where different answers are obtained, the predictive performance criterion may be more relevant in the context of DTSMs, e.g. it can be translated into economic value and give an investor point of view.

The methodology introduced in this chapter allows us to identify the optimal set of restrictions in the $\lambda_{0\mathcal{P}}$, $\lambda_{1\mathcal{P}}$ matrices, while offering useful by-products. First, the ability to monitor the restrictions in a sequential manner and detect changes over certain time periods of interest, second, combining multiple sets of restrictions with sequential Bayesian model averaging, and third, quantifying the associated predictive performance and translating it in terms of economic value.

1.3 Sequential Estimation, Model Choice, and Forecasting

In this section we develop a sequential Monte Carlo (SMC) framework for Gaussian ATSMs. We draw from the work of Chopin (2002, 2004) (see also Del Moral et al.

(2006)) and make the necessary adaptations to tailor the methodology to the data and models considered in this chapter. Furthermore, we extend the framework to allow for sequential Bayesian model choice by incorporating the SSVS algorithm that allows searching over 2^{N+N^2} models; see Schäfer and Chopin (2013) for some relevant work in the linear regression context. Overall, the developed framework allows the efficient performance of tasks such as sequential parameter estimation, model choice, and forecasting. We begin by providing the main skeleton of the scheme and then provide the details of its specific parts, such as the MCMC scheme for exploring the model space, and the framework for obtaining and evaluating the economic benefits of predictions.

1.3.1 Sequential Framework

Let $Y_{0:t} = (Y_0, Y_1 \ldots, Y_t)$ denote all the data available up to time t, such that $Y_{0:T} = Y$. Similarly, the likelihood based on data up to time t is $f(Y_{0:t}|\theta)$ and is defined in (1.17). Combined with a prior on the parameters $\pi(\theta)$, see Appendix 1.A for details, it yields the corresponding posterior

$$\pi(\theta|Y_{0:t}) = \frac{1}{m(Y_{0:t})} f(Y_{0:t}|\theta)\pi(\theta) \tag{1.19}$$

where $m(Y_{0:t})$ is the model evidence based on data up to time t. Moreover, the posterior predictive distribution, which is the main tool for Bayesian forecasting, is defined as

$$f(Y_{t+h}|Y_{0:t}) = \int f(Y_{t+h}|Y_t, \theta)\pi(\theta|Y_{0:t})d\theta \tag{1.20}$$

where h is the prediction horizon.

Note that the predictive distribution in (1.20) incorporates parameter uncertainty by integrating θ out according to the posterior in (1.19). Usually, prediction is carried out by expectations with respect to (1.20), e.g. $E(Y_{t+h}|Y_{0:t})$ but, since (1.20) is typically not available in closed form, Monte Carlo can be used in the presence of

samples from $\pi(\theta|Y_{0:t})$. This process may accommodate various forecasting tasks; for example forecasting several points, functions thereof, and potentially further ahead in the future. A typical forecasting evaluation exercise requires taking all the consecutive times t from the nearest integer of, say, $T/2$ to $T-1$. In each of these times, $Y_{0:t}$ serves as the training sample, and points of Y after t are used to evaluate the forecasts. Hence, carrying out such a task requires samples from (1.20), and therefore from $\pi(\theta|Y_{0:t})$, for several times t. Note that this procedure can be quite laborious and in some cases infeasible.

An alternative approach that can handle both model choice and forecasting assessment tasks is to use sequential Monte Carlo (see, Chopin (2002) and Del Moral et al. (2006)) to sample from the sequence of distributions $\pi(\theta|Y_{0:t})$ for $t = 0, 1, \ldots, T$. A general description of the Iterated Batch Importance Sampling (IBIS) scheme of Chopin (2002), see also Del Moral et al. (2006) for a more general framework, is provided in Algorithm 1.1. The degeneracy criterion is usually defined through the

Algorithm 1.1 IBIS algorithm for Gaussian Affine Term Structure Models

Initialize N_θ particles by drawing independently $\theta_i \sim \pi(\theta)$ with importance weights $\omega_i = 1$, $i = 1, \ldots, N_\theta$. For $t, \ldots, T$ and each time for all i:

(a) Calculate the incremental weights from

$$u_t(\theta_i) = f(Y_t|Y_{0:t-1}, \theta_i) = f(Y_t|Y_{t-1}, \theta_i)$$

(b) Update the importance weights ω_i to $\omega_i u_t(\theta_i)$.

(c) If some degeneracy criterion (e.g. ESS(ω)) is triggered, perform the following two sub-steps:

(i) Resampling: Sample with replacement N_θ times from the set of θ_is according to their weights ω_i. The weights are then reset to one.

(ii) Jittering: Replace θ_is with $\tilde{\theta}_i$s by running MCMC chains with each θ_i as input and $\tilde{\theta}_i$ as output. Set $\theta_i = \tilde{\theta}_i$.

Effective Sample Size (ESS) which is equal to

$$ESS(\omega) = \frac{(\sum_{i=1}^{N_\theta} \omega_i)^2}{\sum_{i=1}^{N_\theta} \omega_i^2} \tag{1.21}$$

and is of the form $ESS(\omega) < \alpha N_\theta$ for some $\alpha \in (0,1)$, where ω is the vector containing the weights.

The IBIS algorithm provides a set of weighted θ samples, or else particles, that can be used to compute expectations with respect to the posterior, $E[g(\theta)|Y_{0:t}]$, for all t using the estimator $\sum_i[\omega_i g(\theta_i)]/\sum_i \omega_i$. Chopin (2004) shows consistency and asymptotic normality of this estimator as $N_\theta \to \infty$ for all appropriately integrable $g(\cdot)$. The same holds for expectations with respect to the posterior predictive distribution, $f(Y_{t+h}|Y_t)$; the weighted θ samples can be transformed into weighted samples from $f(Y_{t+h}|Y_t)$ by simply applying $f(Y_{t+h}|Y_t, \theta)$. A very useful by-product of the IBIS algorithm is the ability to compute $m(Y_{0:t}) = f(Y_{0:t})$, which is the criterion for conducting formal Bayesian model choice. Computing the following quantity in step (a) in Algorithm 1.1 yields a consistent and asymptotically normal estimator of $f(Y_t|Y_{0:t-1})$, namely

$$m_t = \frac{1}{\sum_{i=1}^{N_\theta} \omega_i} \sum_{i=1}^{N_\theta} \omega_i u_t(\theta_i)$$

An additional benefit provided by sequential Monte Carlo is that it provides an alternative choice when MCMC algorithms have poor mixing and convergence properties and, in general, is more robust when the target posterior is challenging, e.g. multimodal. Finally, as we demonstrate in Section 1.5, the sequential nature of the algorithm allows it to produce informative descriptive output to monitor the evolution of key parameters in time.

In order to apply the IBIS output to models and data in this chapter, the following adaptations and extensions are needed. First, the choice of defining the incremental weights in step (a) in Algorithm 1.1, also known as data tempering, is suitable for getting access to sequences of predictive distributions, needed to assess forecasting performance, but at the same time it is quite prone to numerical stability issues and very low effective sample sizes, in particular early on, that is at the initial time points.

This is because the learning rate is typically higher at the beginning, especially when transitioning from a vague prior. An alternative approach that guarantees a pre-specified minimum effective sample size level, and therefore some control over the Monte Carlo error, is to use adaptive tempering; see, for example, Jasra et al. (2011). In order to combine the benefits of both approaches we use a hybrid adaptive tempering scheme which we present in Appendix 1.C. The idea of this scheme is to use adaptive tempering within each transition between the posteriors based on $Y_{0:t}$ and $Y_{0:t+1}$ for each t. Similar ideas have been applied in Schäfer and Chopin (2013) and Kantas et al. (2014).

Second, and quite crucially in this chapter, we extend the framework presented in Section 1.3.2 to handle sequential model searches over the space of all possible risk price restrictions. Third, we note that the MCMC sampler, used in sub-step (ii) of step (c) in Algorithm 1.1, needs to be automated as it will have to be rerun for each time point and particle without the luxury of having initial trial runs, as it is often the case when running a simple MCMC on all the data. The problem is intensified by the fact that the MCMC algorithm used here, developed in Bauer (2018), consists of independence samplers that are known to be unstable. To address this, we utilize the IBIS output and estimate posterior moments to obtain independence sampler proposals; see Appendix 1.B for details. Finally, we connect the IBIS output with the construction of a model-driven dynamically rebalanced portfolio of bond excess returns and calculate its economic value.

In empirical application, we work with $N_\theta = 2000$ particles and 5 MCMC steps when jittering, whereas in relation to minimum ESS we set $\alpha = 0.7$. We decide on 5 steps at the jittering stage because the mixing behaviour of the underlying MCMC is quite satisfactory. We observed the correlation between particles before and after that stage to find that performance was already reasonable with this number of iterations.

1.3.2 Sequential Model Choice Across Risk Price Restrictions

As mentioned in Section 1.2, the specification of the market price of risk is conducted via $\lambda_{0\mathcal{P}}$ and $\lambda_{1\mathcal{P}}$. For brevity of further exposition, we let $\lambda^{\mathcal{P}} = [\lambda_{0\mathcal{P}}, \lambda_{1\mathcal{P}}]$ and $\lambda = \lambda_{1\mathcal{P}}$. If all the entries in $\lambda^{\mathcal{P}}$ are free parameters we get the maximally flexible model. Alternative models have also been proposed in the existing studies, e.g. Cochrane and Piazzesi (2009) and Bauer (2018), where some of these entries are set to zero. More specifically, in most models the set of unrestricted parameters is usually a subset of $\lambda^{\mathcal{P}}$. A standard approach to facilitating Bayesian model choice is via assigning spike-and-slab priors (Mitchell and Beauchamp, 1988; George and McCulloch, 1993; Madigan and Raftery, 1994) on each of the $\lambda_{ij}^{\mathcal{P}}$s, $i = 1, \ldots, N$, $j = 1, \ldots, N+1$, by means of the following mixture

$$\lambda_{ij}^{\mathcal{P}} \sim (1 - \gamma_{ij}) N(0,\ \tau_{ij}^{(0)}) + \gamma_{ij} N(0,\ \tau_{ij}^{(1)})$$

where γ_{ij}s are Bernoulli random variables taking value of zero if the corresponding $\lambda_{ij}^{\mathcal{P}}$ is small (almost equal to zero), or one if it is large (significantly different from zero). Hence $\tau_{ij}^{(0)}$ is typically given a very small value, thus forcing the underlying parameter towards zero, while $\tau_{ij}^{(1)}$ is set to a larger value so that the data determine the value of the parameter in question; see Appendix 1.A for details. The γ_{ij}s are also estimated from the data using MCMC; see Appendix 1.B for details. The proportion of the MCMC draws in which each γ_{ij} is equal to one provides the posterior probability of the corresponding $\lambda_{ij}^{\mathcal{P}}$ being non-zero, also known as posterior inclusion probability.

We consider two approaches to Bayesian model choice in order to explore its links with predictive performance. The first approach is to implement the spike-and-slab approach on some data used for training purposes in order to select the top models. The sequential scheme in Algorithm 1.1 is then applied to each of them, without using spike-and-slab priors and thus omitting step (b) in Algorithm 1.3, and with

some $\lambda_{ij}^{\mathcal{P}}$ being exactly equal to zero, extracting their predictive distributions and contrasting them with the observed data. Under the second approach, sequential inference on both the models and the parameters is drawn. This is implemented by running a single instance of the sequential scheme in Algorithm 1.1, modified to incorporate the SSVS based on the spike-and-slab priors, exactly as it is outlined in Algorithm 1.3. In this case, the parameter vector includes the γ_{ij}s allowing us to calculate the inclusion probabilities, using the particle weights, at each time t based on all the data up to and including t.

This approach offers several advantages in exploring the landscape of the risk price restriction space as we can monitor potential changes in the importance of different $\lambda_{ij}^{\mathcal{P}}$s over time. Moreover, the global search nature of sequential Monte Carlo may be helpful in exploring this landscape across different models. Each θ particle contains a set of γ_{ij}s and corresponds to a particular model. The set of θ particles therefore contains instances of the leading models among the 2^{N+N^2} possible ones. Every time resampling and jittering take place, the list of models can be potentially updated giving more focus to the cases with higher weights, or else posterior probability, and potentially depleting the ones with lower weights. Hence it is now less likely to get trapped in local modes when exploring the model space. Finally, this scheme allows combining different models and incorporating model uncertainty into forecasting via model averaging in a sequential manner.

1.3.3 Assessing Predictive Performance and Economic Value

Failure of the EH implies that bond returns are strongly predictable (see, Fama and Bliss (1987), Campbell and Shiller (1991), Cochrane and Piazzesi (2005), and Ludvigson and Ng (2009), among others). Based on the evaluation framework presented in this section, we then attempt to revisit evidently conflicting results reported in the existing studies (e.g. Duffee (2011), Barillas (2011), Thornton and Valente (2012), Joslin et al. (2014), Sarno et al. (2016), Feunou and Fontaine

(2018), and Bianchi et al. (2021)) on the ability of yields-only ATSMs to capture the predictability of risk premia in the US Treasury market. In particular, we seek to understand whether information on the current yield curve is sufficient to predict excess returns, in light of the alternative restrictions imposed on the dynamics of risk compensation. Furthermore, we attempt to explore whether statistical predictability, if any, can be turned into economic benefits for bond investors.

1.3.3.1 Bond Excess Returns

Define the observed h-holding period return from buying an n-year bond at time t and selling it at time $(t+h)$ as

$$r^n_{t,t+h} = p^{n-h}_{t+h} - p^n_t$$

where p^{n-h}_{t+h} is the log price of the $(n-h)$-period bond at time $(t+h)$ and p^n_t is the log price of the n-period bond at time t. The latter translates to the corresponding yield in the following manner

$$y^n_t = -\frac{1}{n} p^n_t$$

Furthermore, define the observed continuously compounded excess return of an n-year bond as the difference between the holding period return of the n-year bond, expressed above in terms of log prices, and the h-period yield as

$$rx^n_{t,t+h} = -(n-h)y^{n-h}_{t+h} + ny^n_t - hy^h_t$$

If instead of taking the observed one, we take the model-implied continuously compounded yield y^n_t, calculated according to (1.8), we arrive at the predicted excess return $\widetilde{rx}^n_{t,t+h}$ which becomes

$$\widetilde{rx}^n_{t,t+h} = A_{n-h,\mathcal{P}} - A_{n,\mathcal{P}} + A_{h,\mathcal{P}} + B'_{n-h,\mathcal{P}}\widetilde{\mathcal{P}}_{t+h} - (B_{n,\mathcal{P}} - B_{h,\mathcal{P}})'\mathcal{P}_t \tag{1.22}$$

where $\mathcal{P}_t$ is observed and $\widetilde{\mathcal{P}}_{t+h}$ is a prediction from the model. Our developed framework, see Algorithm 1.1, allows drawing from the predictive distribution of $(\widetilde{\mathcal{P}}_{t+h}, \widetilde{rx}^n_{t,t+h})$ based on all information available up to time t. More specifically, for each θ_i particle the $\mathbb{P}$-dynamics of $\mathcal{P}_t$ can be used to obtain a particle of $\widetilde{\mathcal{P}}_{t+h}$, which then can be transformed into a particle of $\widetilde{rx}^n_{t,t+h}$ via equation (1.22). Detailed steps for the case of $h = 1$ are outlined in Algorithm 1.2.

Algorithm 1.2 Predictive distribution of excess returns for Gaussian Affine Term Structure Models

First, at time t, for some n and $h = 1$, using (ω_i, θ_i), $i = 1, ..., N_\theta$, from IBIS algorithm, iterate over i:

(a) Given θ_i, compute $A_{i_1,\mathcal{P}}$ and $B_{i_1,\mathcal{P}}$, for $i_1 \in \{1, n-1, n\}$, from (1.9) and (1.10).

(b) Given θ_i, obtain prediction of $\mathcal{P}_{t+1}$ by drawing from

$$\widetilde{\mathcal{P}}^{(i)}_{t+1} | \mathcal{P}_t \sim N\left(\mu^{\mathbb{P}}_{\mathcal{P}} + \Phi^{\mathbb{P}}_{\mathcal{P}} \mathcal{P}_t,\ \Sigma_{\mathcal{P}} \Sigma'_{\mathcal{P}}\right)$$

(c) Compute particle prediction of $rx^n_{t,t+1}$ as

$$\widetilde{rx}^{n(i)}_{t,t+1} = A_{n-1,\mathcal{P}} - A_{n,\mathcal{P}} + A_{1,\mathcal{P}} + B'_{n-1,\mathcal{P}} \widetilde{\mathcal{P}}^{(i)}_{t+1} - (B_{n,\mathcal{P}} - B_{1,\mathcal{P}})' \mathcal{P}_t$$

Second, since $(\omega_i, \widetilde{\mathcal{P}}^{(i)}_{t+1}, \widetilde{rx}^{n(i)}_{t,t+1})$, $i = 1, ..., N_\theta$, is a particle approximation to predictive distribution of $(\mathcal{P}_{t+1}, rx^n_{t,t+1})$, compute point prediction of $rx^n_{t,t+1}$ using particle weights ω_i as

$$\widetilde{rx}^n_{t,t+1} = \frac{1}{\sum_{i=1}^{N_\theta} \omega_i} \sum_{i=1}^{N_\theta} \omega_i \widetilde{rx}^{n(i)}_{t,t+1}$$

Third, repeat above two steps for different n and h. For $h > 1$, use $\widetilde{\mathcal{P}}^i_{t+h-1}$ in place of $\mathcal{P}_t$, and $i_h \in \{h, n-h, n\}$ in place of i_1.

The predictive accuracy of bond excess return forecasts is measured in relation to an empirical benchmark. We follow related literature and adopt the EH as this benchmark, which essentially uses historical averages as the optimal forecasts of bond excess returns. This empirical average is

$$\overline{rx}^n_{t+h} = \frac{1}{t-h} \sum_{j=1}^{t-h} rx^n_{j,j+h}$$

We consider two metrics to assess the predictive ability of models considered. First, following Campbell and Thompson (2008), we compute the out-of-sample R^2 (R^2_{os}) as

$$R^2_{os} = 1 - \frac{\sum_{s=t_0}^{t}(rx^n_{s,s+h} - \widetilde{rx}^n_{s,s+h})^2}{\sum_{s=t_0}^{t}(rx^n_{s,s+h} - \overline{rx}^n_{s+h})^2}$$

for $\widetilde{rx}^n_{s,s+h}$ being the mean of the predictive distribution. Positive values of this statistic mean that model-implied forecasts outperform the empirical averages and suggest evidence of time-varying return predictability. Second, in order to assess the entire predictive distribution offered by our scheme, rather than just point predictions, we use the log score; a standard choice among scoring rules with several desirable properties, such as being strictly proper, see for example Dawid and Musio (2014). The log score of the predictive distribution can be approximated numerically using kernel methods. For the EH case, this evaluation can be done analytically. These metrics are aggregated over all prediction times (t_0 to t) for each maturity. In order to get a feeling for how large the differences from the EH benchmark are, we report the p-values from one-sided Diebold-Mariano test (see, Gargano et al. (2019)), with Clark and West (2007) adjustment as in Fulop et al. (2019), this in case of R^2_{os} only, noting that these are viewed as indices rather than formal hypothesis tests. They are based on t-statistics computed taking into account potential serial correlations in the standard errors (see, Newey and West (1987)).

1.3.3.2 Economic Performance of Excess Return Forecasts

From a bond investor's point of view it is of paramount importance to establish whether the predictive ability of a model can generate economically significant portfolio benefits, out-of-sample. The portfolio performance may also serve as a metric to compare models that impose different sets of restrictions on the price of risk specification. In that respect, our approach is different from Thornton and Valente (2012) and Sarno et al. (2016), who test the economic significance for an investor

with mean-variance preferences[20] and conclude that statistical significance is not turned into better economic performance, when compared to the EH benchmark. It is more in line with Gargano et al. (2019) and Bianchi et al. (2021), who arrive at similar conclusions for models which utilize information coming solely from the yield curve (e.g. yields, forwards, etc.). Computationally, it is quite similar to the approach of Fulop et al. (2019), tailored to the context of this chapter.

We consider a Bayesian investor with power utility preferences

$$U(W_{t+h}) = U(w_t^n, rx_{t+h}^n) = \frac{W_{t+h}^{1-\gamma}}{1-\gamma}$$

where W_{t+h} is an h-period portfolio value and γ is the coefficient of relative risk aversion. If we let w_t^n be a portfolio weight on the risky n-period bond and $(1 - w_t^n)$ be a portfolio weight of the riskless h-period bond, then the portfolio value h periods ahead is given as

$$W_{t+h} = (1 - w_t^n)\exp(r_t^f) + w_t^n \exp(r_t^f + rx_{t,t+h}^n)$$

where r_t^f is the risk-free rate, here synonymous with the h-period yield. Such an investor maximizes her expected utility over h-periods in the future, based on $x_{1:t} = \{rx_{1:t}^n, \mathcal{P}_{0:t}\}$

$$\begin{aligned} E_t[U(W_{t+h})|x_{1:t}] &= \int U(W_{t+h}) f(W_{t+h}|x_{1:t}) dW_{t+h} \\ &= \int U(w_t^n, rx_{t+h}^n) f(rx_{t+h}^n|x_{1:t}) drx_{t+h}^n \end{aligned}$$

where $f(rx_{t+h}^n|x_{1:t})$ is the predictive density described earlier. At every time t, our Bayesian learner solves an asset allocation problem getting optimal portfolio weights

[20] In fact, Sarno et al. (2016), also use an approximation of the power utility solution. Furthermore, they allow for the variance to be constant (or rolling window) and in-sample, in line with Thornton and Valente (2012).

as result of numerically solving

$$\widetilde{w}_t^n = \arg\max \frac{1}{\sum_{i=1}^{N_\theta} \omega_i} \sum_{j=1}^{N_\theta} \omega_j \left\{ \frac{[(1 - w_t^n) \exp(r_t^f) + w_t^n \exp(r_t^f + \widetilde{rx}_{t,t+h}^{n,j})]^{1-\gamma}}{1-\gamma} \right\}$$

for example by means of Nelder-Mead simplex algorithm (Lagarias et al., 1998), where N_θ is the number of particles from the predictive density of excess returns, weighted using importance weights ω_i, $i = 1, ..., N_\theta$, which come from the IBIS algorithm.

To obtain the economic value generated by each model, we use the resulting optimum weights to compute the CER as in Johannes et al. (2014) and Gargano et al. (2019). In particular, for each model, we define the CER as the value that equates the average utility of each model against the average utility of the EH benchmark specification. Denoting realized utility from the predictive model as $\widetilde{U}_t = U\left(\widetilde{w}_t^n, \{\widetilde{rx}_{t,t+h}^{n,j}\}_{j=1}^{N_\theta}\right)$ and realized utility from the EH benchmark as $\overline{U}_t$, we get

$$CER = \left(\frac{\sum_{s=t_0}^{t} \widetilde{U}_s}{\sum_{s=t_0}^{t} \overline{U}_s} \right)^{\frac{1}{1-\gamma}} - 1$$

Similar to R_{os}^2 and log-score introduced in Section 1.3.3.1, statistical significance of CER results is based on one-sided Diebold-Mariano test where p-values are based on t-statistics calculated using Newey-West adjusted standard errors.

1.4 Data and Models

In this section we discuss data we use throughout in detail. Specifically, we elaborate on the US Treasury yields data we employ in this chapter, how we choose subsamples thereof, and why exactly we split these into training and testing periods in the manner described below. We then also explain what models we consider, as distinguished by different risk price restrictions, and how we arrive at such choice, in particular using our developed sequential SSVS scheme and rationale provided in the literature.

1.4.1 Yields Only

The data set contains monthly observations of zero-coupon US Treasury yields with maturities of 1-year, 2-year, 3-year, 4-year, 5-year, 7-year, and 10-year, spanning the period from January 1985 to the end of 2016. We consider two sub-samples, one ending at the end of 2007 and that, as such, precludes the recent financial crisis, and a second one which includes the period after the end of 2007, a period determined by, first, different monetary actions and establishment of unconventional policies and, second, interest rates hitting the zero-lower bound. The first sub-sample has been used by most of the existing studies (see, Joslin et al. (2011), Joslin et al. (2014), Bauer (2018), and Bauer and Hamilton (2018)). Following related literature, we choose the starting date avoiding the early 1980s, a period with evidence of the Fed changing its monetary policy. The post-2007 global financial crisis period is excluded from our first sub-sample due to concerns about the capability of Gaussian ATSMs to deal with the zero-lower bound (see, Kim and Singleton (2012) and Bauer and Rudebusch (2016)). These concerns are explored in the second sub-sample, spanning the period from January 1990 to end of 2016, as in Bauer and Hamilton (2018). Overall, the data contain different market conditions and monetary policy actions.

In the analyses of the two sub-samples the data are split into a training period, where the data are used only for estimation purposes, and a testing period, starting immediately afterwards, where we start evaluating the model predictions while incorporating additional information as the data become available. Specifically, the training periods are 1985-1996 and 1990-2005 respectively. The data processing involves extracting the first three principal components from the yield curve. This can be done either by calculating the principal component (PC) loadings from the training or the entire period; the difference in the resulting principal component time series are depicted in Figure 1.1. The correlations between these series reach levels above 0.99, suggesting negligible differences. Nevertheless, in order to prevent any

data leaking issues when assessing the predictive performance, the loadings from the training period only are used.

1.4.2 Models and Rationale Behind

In terms of models, as mentioned in Section 1.2, there are 4,096 possible distinct sets of risk price restrictions in the case of three factors driving the state variables. The first model considered, (M_0), in line with previous empirical studies (e.g. Duffee (2011), Sarno et al. (2016), and Bauer (2018), among others), allows the risk prices to be completely unrestricted and is therefore known as the maximally flexible model being only subject to normalizing restrictions, as in Joslin et al. (2011).

The next three models are the ones suggested by our developed sequential SSVS scheme as the models with the highest posterior probability (model evidence). Our results reinforce the argument of sparsity in the market price of risk specification, as in Bauer (2018), since in these models only one or two risk price parameters are free. Specifically, the model with the highest posterior probability is the one that allows a single free parameter, $\lambda_{1,2}$, which we denote as M_1. This is the parameter which drives variation in the price of level risk due to changes in the slope of the yield curve. The next model, denoted by M_2, has two free risk price parameters, $\lambda_{1,2}$ as M_1 and also $\lambda_{1,1}$. Interestingly enough, in the next best model in terms of posterior probability, denoted by M_3, only $\lambda_{1,1}$ is free. This suggests that the variation in the price of level risk is driven by changes in the level factor.

Finally, we consider Bayesian model averaging with model weights obtained from our developed sequential SSVS scheme. We explore three formulations: the first one (M_4) assumes uniform prior distribution over models, where each element of $\lambda^{\mathcal{P}}$ is independently Bernoulli distributed with success probability 0.5, as in Bauer (2018). The second one (M_5) uses a hierarchical prior, namely Beta$(1,1)$-Binomial; see for example Wilson et al. (2010) and Consonni et al. (2018). The list of models

is completed with M_6, a filtered version of M_4 with only $\lambda_{1,1}$ and $\lambda_{1,2}$ allowed to be non-zero.

1.5 Empirical Results

This section presents the main results on the statistical and economic performance of excess return forecasts resulting from the models we develop in this chapter. In particular, we assess the models based on different sets of restrictions and explore the evident puzzling behaviour between statistical predictability and meaningful out-of-sample economic benefits for bond investors. Furthermore, we monitor how the optimum set of restrictions behaves around periods of monetary policy shifts, interventions, and fragile economic conditions.

1.5.1 Yield Curve and Risk Price Dynamics

The data contain several important events and it is interesting to examine the trajectories of the principal components during these periods, provided by Figure 1.1. More specifically, during the period between 2004 and 2006, referred to by former Fed Chairman Greenspan as the 'conundrum period', the Federal Reserve applied a tight monetary policy by substantially increasing its target federal funds rate by 4%, reaching a value of 5.25% by mid-2006. At the same time, long-term yields actually declined, directly affecting the shape of the yield curve, which flattened profoundly. This episode is captured by the first two implied PCs, as shown in Figure 1.1. In particular, the first PC increased substantially during the aforementioned period, reflecting the increase in the level of the term structure, while the second decreased to very low levels, reflecting the flattening of the curve. The data also covers the two recession periods of 2001-2002 and 2008-2009, where the yield curve shared qualitatively similar characteristics. Both episodes started from a flat yield curve during the pre-recession periods, following an inversion of the curve that reflects an

increase in short rates due to expectations about the Fed tightening its policy, and then a steepening as a reaction to policy adjustments by the Fed reflecting strong growth and inflation expectations. This is reflected in the two principal components, with the first PC decreasing and the second increasing during the recession periods.

The above are captured from the sequential setup developed in this chapter, which is capable of monitoring variations across time in the estimates of parameters, restrictions, and the importance thereof, in contrast to previous studies (e.g. Sarno et al. (2016), Bauer (2018), and Gargano et al. (2019)). Figures 1.2 and 1.3 contain the results obtained from fitting the model M_6 on the two previously mentioned sub-samples, focusing on the testing periods. The plots depict information for the parameters $\lambda_{1,1}$, $\lambda_{1,2}$ and their posterior inclusion probabilities, as well as the highest eigenvalue of $\Phi_{\mathcal{P}}^{\mathbb{P}}$ and the parameter $k_{\infty}^{\mathbb{Q}}$ which is linked to the long-run mean of the short rate r_t under $\mathbb{Q}$. From the restricted parameters, we chose to report only $\lambda_{1,1}$ and $\lambda_{1,2}$ as these were the only parameters with posterior inclusion probabilities not close to zero. In fact, based on the median probability principle that recommends keeping only variables with inclusion probabilities above 0.5, perhaps only the $\lambda_{1,2}$ parameter should be allowed to be free, thus pointing to model M_1. Nevertheless, as we discuss in the remainder of the chapter, it may be helpful to consider freeing $\lambda_{1,1}$.

The posterior inclusion probabilities of the risk premia parameters seem to closely follow the changes in policy actions. During the conundrum period, the inclusion probability for $\lambda_{1,1}$ increases while at the same time that for $\lambda_{1,2}$ deteriorates. This suggests that the parameter, which links compensation for level risk to the slope factor, becomes more important during periods of yield curve steepening than periods when the curve flattens. Similarly, the parameter $\lambda_{1,1}$ is more likely to be important during periods where the level of the term structure increases. The situation is reversed during the recession periods in a similar manner, with changes in the principal components. Note that the developed sequential scheme uses an expanding window approach, under which the detection of changes towards the end of the

time period is becoming increasingly difficult; such changes would have been more pronounced had a rolling window been adopted. On the other hand, there seems to be enough information in the training samples to estimate these particular parameters well and additional information from the testing samples contributes little to their final estimates.

It is also interesting to look at $k_\infty^{\mathbb{Q}}$ over time, the posterior trajectory of which follows interest rate expectations and yield curve fluctuations. In the first sub-sample, it starts at a high level and progressively moves down to zero until 2005. This is followed by an increase during and after the conundrum period of 2004-2006, due to the substantial increase of the federal funds rate. Qualitatively similar conclusions are made when looking at the second sub-sample, where $k_\infty^{\mathbb{Q}}$ remains close to zero until the period after 2008, when it starts slightly increasing reflecting the steepening of the curve, due to the Fed's policies. The increase is more pronounced during and after the 2013 'taper tantrum' events, reflecting and capturing the sharp increase in medium-to-long maturity yields. Finally, we monitor the posterior trajectories of the largest eigenvalue of the feedback matrix $\Phi_{\mathcal{P}}^{\mathbb{P}}$. Its posterior mean remains nearly constant and very close to unity over the entire sample period, indicating a generally high $\mathbb{P}$-persistence, implied by the restrictions imposed on the risk-price parameters. This also reflects an enhanced time variation of short rate expectations and more stable risk premiums, implying a larger role of the expectation component over the risk premium component, in line with policy making and interventions.

1.5.2 Bond Return Predictability

This section presents results on the predictive performance exercise described in Section 1.3.3.1. Table 1.1 reports out-of-sample R^2 values for all models across bond maturities and prediction horizons. The in-sample period is January 1985 to the end of 1996 and the out-of-sample period is January 1997 to the end of 2007, which thus precludes the post-2007 financial crisis. Results suggest that models with some

or no restrictions on the dynamics of risk compensation, such as the maximally flexible model M_0, widely used in the vast majority of prior studies (e.g. Sarno et al. (2016)), fail to predict well and perform poorly out-of-sample compared to the EH benchmark, as showcased by predictive R^2_{os} that are mostly negative, especially at the longer maturities (beyond 5-year).

Results, however, are qualitatively opposite when heavy restrictions are imposed on the risk price dynamics. We find that restrictions add considerable improvement to the out-of-sample predictive performance compared to models with some or no restrictions at all. The former also generate more accurate forecasts, compared to risk premia implied by the EH, as showcased by positive R^2_{os}, suggesting strong evidence of out-of-sample bond return predictability. This is of particular importance if we take into account the low-dimensional setting, where only information in the yield curve is utilized. These observations are in contrast to the prior literature (see, Duffee (2011), Barillas (2011), Thornton and Valente (2012), Adrian et al. (2013), and Joslin et al. (2014)), which suggests that yields-only DTSMs are not capable of capturing the predictability of bond risk premia, and are more in line with the results of Sarno et al. (2016), Feunou and Fontaine (2018) and Bianchi et al. (2021). Additionally, we identify two such models, M_1 and M_3, where only shocks that affect the market price of level risk are priced, thus suggesting that investors seek compensation for level risk only, which is in line with Cochrane and Piazzesi (2009). Finally, SSVS schemes (e.g. models M_4 and M_6) attain good performance with significantly positive R^2_{os}, when compared to the EH benchmark. Such improvement occurs across the maturity spectrum and investment horizons.

Furthermore, our results suggest that, during the sample period, R^2_{os} decrease with maturity and increase with the prediction horizon, reaching their peak values at horizons of 6-months and 9-months. In particular, for model M_1, for the 2-year (5-year) maturity bond, R^2_{os} increase from 0.05 (0.03) at the 1-month horizon to 0.10 (0.09) at the 6-month horizon. Qualitatively similar increases are observed for other

models. For example, when we take model M_3 and sequential SSVS model M_6, we observe R^2_{os} for a 2-year maturity bond increasing from 0.03 to 0.13 and from 0.04 to 0.10, respectively, when moving from the 1-month to the 9-month horizons.

Similar conclusions, yet more pronounced, are drawn from Table 1.2 that reports R^2_{os} results for the post-crisis period of January 2007 to end of 2016, suggesting that predictability is linked to the economic cycle. In particular, we confirm that predictability is substantially higher following the post-crisis recession period, when the US market experienced high uncertainty and low interest rates, compared to the pre-crisis low volatility and high yield period, where R^2_{os} are substantially lower. This finding is robust across models tested and methodologies applied. A very interesting observation comes from looking at models M_1 and M_3. In particular, M_3 seems to perform really well during the pre-crisis period, generating accurate forecasts, and substantially high R^2_{os}. Such performance, however, completely vanishes during the post-2007 period, where R^2_{os} turn negative across maturities and investment horizons. This reveals that shocks to the level of the yield curve, captured through $\lambda_{1,1}$, are important components of time-varying risk premia only during high yield and low uncertainty periods. Simultaneously, R^2_{os} for model M_1 almost triples in magnitude following the post-crisis low yield and high uncertainty sub-period. This outperformance is more pronounced at longer maturities and horizons. In particular, at the 12-month investment horizon, R^2_{os} increases from 0.08 up to 0.34 for the 10-year maturity bond. Comparably substantial improvements are also experienced in the case of sequential SSVS models which provide quantitatively similar R^2_{os} across the maturity spectrum.

The out-of-sample R^2 measure, used to assess the predictive ability of the models, only focused on a point summary of the predictive distribution and may ignore important information. To account for full information associated with the distribution of excess returns, we evaluate the accuracy of their density forecasts using the log score. At each time step t, we obtain the LS by taking the log of the predictive densities of

excess returns for each one of the tested models against the EH benchmark. Tables 1.3 and 1.4 report out-of-sample LS across models, maturities, and prediction horizons. The main conclusions remain unchanged since results are qualitatively similar to those reported for the R^2_{os}, showing evidence in favour of statistical predictability only for those models that allow for heavy restrictions. In particular, in most cases, corresponding LS are positive and non-negligible, indicating statistical evidence of out-of-sample predictability. The only model that performs poorly out-of-sample compared to the EH benchmark is the maximally flexible model, M_0, where LS are mostly negative, especially at longer maturities. M_1 is performing very well in the second sub-sample but its superiority is challenged by M_3 in the first sub-sample. The sequential SSVS model that combines these two, M_6, does very well on both occasions although the differences between these models are quite small.

Overall, the results highlight the importance of different sets of restrictions across market conditions and monetary policy actions. A message coming out of the analysis, despite the very good performance of M_1, is that sequential model averaging can provide improvements in predictive performance, and capture model uncertainty as well as model instability.

1.5.3 Economic Performance

In this section we concentrate on performance in terms of economic value, as described in Section 1.3.3.2. Table 1.5 reports results for the annualized CERs, generated using out-of-sample forecasts of bond excess returns across maturities and prediction horizons. The coefficient of relative risk aversion is $\gamma = 3$ and no portfolio constraints are imposed[21]. Results show clear evidence of positive out-of-sample economic benefits for bond investors. We find that, in most cases, corresponding CERs are positive and non-negligible, indicating that yields-only models, with heavy restrictions on the

[21] These non-conservative choices are motivated by early exploratory character of the analysis conducted herein. The goal is to find economic value first and in future research examine if it can be exploited when stricter conditions apply.

dynamics of risk compensation, not only provide statistical evidence of out-of-sample predictability, but also generate valuable economic gains for bond investors relative to the EH benchmark. Concurrently, models with some or no restrictions on the risk price dynamics, such as the maximally flexible model M_0, fail to offer any positive out-of-sample economic benefits compared to the EH benchmark, generating CER values which are consistently negative across the maturity spectrum and investment horizons. Furthermore, consistent with the conclusions coming from the predictability analysis in Section 1.5.2, such models are identified to be the exact same ones, suggesting that only models which allocate one or two non-zero risk price parameters solely to the level factor are able to generate meaningful economic gains for investors who dynamically rebalance their portfolio when new information arrives. Finally, the developed sequential SSVS scheme, which only searches among the best models available, attains very good performance, generating CER values which are quantitatively similar, though marginally better, compared to model M_1. Those findings are in contrast to the conclusions of Thornton and Valente (2012) and Sarno et al. (2016), who argue that bond investors utilizing information from the yield curve only are not able to systematically earn any economic premium out-of-sample.

Furthermore, our results suggest larger gains from predictability at longer maturities, reflecting substantially higher CER values and consequently more profitable investments. For example, for model M_1 at the 9-month prediction horizon, CER increases from 0.84% for a 2-year maturity bond, up to 3.38% for the 10-year maturity. Qualitatively similar increases are generated for most of the models. In particular, for M_3 at the 9-month horizon CER increases from 0.72% for the 2-year maturity bond to 1.22% for a long-term bond with 10-year maturity. In turn, for the sequential SSVS model M_6, which searches only among the best models available, CER increases from 1.07% for the 2-year maturity up to 2.96%, and highly significant, for a 10-year maturity bond. In fact, during this sample period, the latter model appears to outperform all other models tested.

Comparing the performance of alternative model specifications over time sheds light on the importance of particular restrictions across monetary policy actions and market conditions. Table 1.6 displays CER values for the period 2007-2016, which covers the aftermath of the recession of 2007-2009 as well as the most interesting phases of the unfolding of the Fed's policy responses to it. Results reveal that economic benefits are even more pronounced compared to the pre-crisis period. Such an upturn in CER values occurs across the maturity spectrum with a tendency for substantially larger gains at the short end of the curve, where investments on the 2-year maturity bond are now the most profitable. This is in contrast to the results from the first sub-period, where the most profitable investments are those in long maturity bonds. In particular, looking at all models, other than M_0 and M_3, CER values for a 2-year maturity bond almost triple when compared to the pre-crisis period. In fact, CER values increase with bond maturity during the expansion period and decrease with maturity during the recession phase of the economic cycle. More specifically, for model M_1 at the 6-month investment horizon, CER value declines from highly significant 5.18% for a 2-year maturity bond down to 2.34% for bonds with 10-year maturity. Qualitatively similar results are observed for other models tested, as displayed in Table 1.6.

An interesting conclusion is inferred when comparing CER values for model M_3 over time. M_3 appears to perform really well during the pre-crisis period, generating high gains to bond investors, as showcased by the consistently positive CER values, especially at longer maturities. However, the situation is reversed in the aftermath of the global financial crisis where CER values turn negative across the maturity spectrum. Model M_3, as well as the maximally flexible model M_0, fail to produce any benefits to investors out-of-sample, when compared to the EH benchmark. This result, which is in line with the statistical analysis in the previous section, suggests that the effect on time-varying risk premia, through parameter $\lambda_{1,1}$, completely vanishes in the second sub-sample. Sequential SSVS models provide very good

performance, consistently generating large gains to bond investors, across maturities and investment horizons.

1.6 Conclusions

We propose a novel methodological framework, which utilizes Bayesian inference and allows us to identify the optimal set of restrictions on the dynamics of risk compensation, in real time. The sequential version of the SSVS scheme developed successfully handles sequential model searches over the restriction space landscape, thus offering an important diagnostic tool capable of monitoring and identifying the optimum set of restrictions across market conditions and monetary policy actions. The modelling setup takes into account model and parameter uncertainty and provides the entire predictive distribution of bond returns, allowing investors to revise their beliefs when new information becomes available and thus informing their asset allocation in real time. The framework is then put into action to revisit the evident puzzling behaviour between statistical predictability and out-of-sample economic benefits in light of different model specifications.

Empirical results reveal the need for heavy restrictions on the dynamics of risk compensation. In particular, we find strong evidence of predictability only for those models that allow for level risk to be priced. Most importantly, such statistical evidence is turned into economically meaningful utility gains, out-of-sample. The sequential SSVS scheme provides further improvements during turbulent periods. Furthermore, our results highlight the importance of different sets of restrictions over time. In particular, the importance of parameters that capture variation in the slope factor increases during periods of curve steepening and decreases when the curve flattens. Finally, we document that both statistical predictability and economic benefits are substantially higher following the post-crisis high uncertainty

and low yield period, suggesting a strong relationship between the slope of the yield curve and bond excess returns in the US market.

Going forward, we note the data typically contain yield curve fluctuations that could be attributed to changes in market conditions and policy actions. One message coming out of our analysis is that this has an impact on the estimates of some of the parameters of the adopted ATSM formulation, for example the mean under the pricing measure and the volatility that appear to be time varying. It may therefore be sensible to consider such extensions. Our approach can capture that to some extent, but it is a bit limited since, due to its expanding data window nature, it is hard to capture changes as the data accumulate. Another promising future direction is to incorporate spanned or unspanned macroeconomic variables in the models; however, unlike Joslin et al. (2014), in a nonlinear rather than linear manner.

Appendix 1.A Specification of Priors

In the following, we provide the prior distributions that were not mentioned in the main body of the chapter. Low informative priors on each θ component can be assigned although some relevant information is available in this context (Chib and Ergashev, 2009).

We first transform all restricted range parameters so that they have unrestricted range. We also scale parameters which typically take very small values. Specifically, we consider a Cholesky factorization of $\Sigma_{\mathcal{P}}$ where the diagonal elements are transformed to the real line and off-diagonal elements are scaled by 10^6. In order to preserve the ordering of the eigenvalues $g^{\mathbb{Q}}$ we apply a reparametrization and work with their increments that are again transformed to the real line. Finally, we scale $k_{\infty}^{\mathbb{Q}}$ by 10^6 as well.

Next, independent normal distributions with zero means and large variances are assigned to each component of θ, except $\lambda^{\mathcal{P}}$, as explained below, and σ_e^2, where we

assign a conjugate Inverse-Gamma prior as in Bauer (2018) with parameters $\alpha/2$ and $\beta/2$. We take $\alpha = \beta = 0$ for an entirely diffuse case.

In general, prior for $\lambda^{\mathcal{P}}$ cannot be uninformative because it would cause Bayes factors indeterminate (Kass and Raftery, 1995). Although immaterial in large samples, its choice plays a significant role for inference in small samples that one is dealing with in DTSM setting. Following Bauer (2018) and independent of whether in a model with or without SSVS, prior specification for the elements of $\lambda^{\mathcal{P}}$ included in the model, we employ here, consists of independent normal distributions centered around zero, that is $\lambda_{ij}^{\mathcal{P}}|\gamma_{ij} = 1 \sim N(0,\ \tau_{ij}^{(1)})$. Similarly, we assume conditional prior independence between elements of $\lambda^{\mathcal{P}}$. To that end, we use orthogonalized g-prior, where g is equal to the number of observations. Specifically, the covariance matrix of such a g-prior distribution, which is a normal, is proportional to that of the least-squares estimator but with off-diagonal elements equal to zeros. To obtain it, all model parameters other than $\lambda^{\mathcal{P}}$ are set to their maximum likelihood estimates and all γ_{ij}s are set to ones. Hence, we are specifying a maximally flexible DTSM. Then least-squares estimates of $\lambda^{\mathcal{P}}$ are calculated. Eventually, we have that $\tau_{ij}^{(0)} = \frac{1}{g}\hat{\sigma}^2_{\lambda_{ij}^{\mathcal{P}}}$ and $\tau_{ij}^{(1)} = g\hat{\sigma}^2_{\lambda_{ij}^{\mathcal{P}}}$, where $\hat{\sigma}^2_{\lambda_{ij}^{\mathcal{P}}}$ are diagonal elements of the covariance matrix for the resulting least-squares estimates of $\lambda^{\mathcal{P}}$. In the model without SSVS only $\tau_{ij}^{(1)}$ is used.

Appendix 1.B Markov Chain Monte Carlo Scheme

Following from (1.17) and (1.19), and given a prior $\pi(\theta)$ as described in Appendix 1.A, the posterior can be written in a more detailed manner as

$$\begin{aligned}\pi(\theta|Y) = &\left\{\prod_{t=0}^{T} f^{\mathbb{Q}}(y_t|\mathcal{P}_t, k_\infty^{\mathbb{Q}}, g^{\mathbb{Q}}, \Sigma_{\mathcal{P}}, \sigma_e^2)\right\} \times \\ &\left\{\prod_{t=1}^{T} f^{\mathbb{P}}(\mathcal{P}_t|\mathcal{P}_{t-1}, k_\infty^{\mathbb{Q}}, g^{\mathbb{Q}}, \lambda_{0\mathcal{P}}, \lambda_{1\mathcal{P}}, \Sigma_{\mathcal{P}})\right\} \times \pi(\theta)\end{aligned}$$

Since the above posterior is not available in closed form, methods such as MCMC can be used to draw samples from it to calculate expectations of interest, $E[g(\theta)|Y]$,

provided that they exist, using Monte Carlo. However, note that the MCMC output is not guaranteed to lead to accurate Monte Carlo calculations since the corresponding Markov chain may have poor mixing and convergence properties, thus leading to highly autocorrelated samples.

It is therefore essential to construct a suitable MCMC algorithm that does not exhibit such unfavourable characteristics. Various studies (see, for example, Chib and Ergashev (2009) and Bauer (2018)), note a substantial improvement to the quality of the MCMC output if a Gibbs scheme is adopted, where the parameters are updated in blocks, some of them being full Gibbs steps. For the remaining blocks, independence samplers may be constructed, using the maximum likelihood estimate as the mean in respective proposal density and negative inverse of the corresponding Hessian as its covariance. Such an MCMC scheme is described in Algorithm 1.3.

Algorithm 1.3 MCMC scheme for Gaussian Affine Term Structure Models

Initialize all values of θ. Then at each iteration of the algorithm:

(a) Update σ_e^2 from its full conditional distribution that can be shown to be an Inverse Gamma distribution with parameters $\tilde{\alpha}/2$ and $\tilde{\beta}/2$, such that $\tilde{\alpha}$ is $\alpha + T(J-N)$ and $\tilde{\beta}$ is $\beta + \sum_{t=0}^{T} \|\hat{e}_t\|^2$, where $\alpha = \beta = 0$, since prior is assumed diffuse, $\hat{e}_t$ is a time-t residual from (1.11), and $\|\cdot\|^2$ is Euclidean norm squared.

(b) Update $\gamma_{i,j}$, $i = 1, ..., N$, $j = 1, ..., N+1$, successively, in random order, from their full conditional distributions, which are Bernoulli and depend only on respective $\lambda_{i,j}^{\mathcal{P}}$s and own hyper-parameters, see for example Bauer (2018) for details on equal prior model probabilities and Consonni et al. (2018) for details on sparse model priors.

(c) Update $\lambda^{\mathcal{P}}$ from its full conditional distribution that can be shown to be the normal distribution with mean and variance obtained from the restricted VAR framework, see for example Bauer (2018) for details.

(d) Update $\Sigma_{\mathcal{P}}$ using an independence sampler based on the MLE and the Hessian obtained before running the MCMC, using multivariate t-distribution with 5 degrees of freedom as proposal distribution.

(e) Update $(k_\infty^{\mathbb{Q}}, g^{\mathbb{Q}})$ in a similar manner to (d).

Appendix 1.C Adaptive Tempering

Adaptive tempering serves the purpose of smoothing peaked likelihoods. It is achieved by bridging two successive targets via an intermediate target sequence. The idea is to modify the sequence of target distributions so that is evolves from the prior to the posterior more smoothly. See Jasra et al. (2011) and Kantas et al. (2014) for details.

Implementation of the IBIS scheme with hybrid adaptive tempering steps is outlined in Algorithm 1.4. It is important to note that, unlike it is shown in Algorithm 1.1 for the general IBIS case, in the specific case we are dealing here with, we initialize the particles by drawing from the posterior $\pi(\theta|Y_{0:t-1})$ instead of the prior $\pi(\theta)$. This is done in-sample based on training data, as it is described in detail in Section 1.4. While it is straightforward to implement step 4(b)iv in Algorithm 1.4 for an independence sampler, adjustments are necessary for a full Gibbs step. In particular, this is the case for σ_e^2 in step (a) and $\lambda^{\mathcal{P}}$ in step (c) in Algorithm 1.3, see Appendix 1.B.

We begin with the former by observing that, assuming tempering parameter ϕ, the following holds

$$\left[f(\hat{e}_T|\sigma_e^2)\right]^{\phi} = \frac{\exp\left(-\frac{\phi}{2\sigma_e^2}\|\hat{e}_T\|^2\right)}{\left[(2\pi)^{J-N}\sigma_e^{2(J-N)}(J-N)\right]^{\frac{\phi}{2}}}$$

for $\hat{e}_T$ which is a time-T residual from (1.11), $f(\hat{e}_T|\sigma_e^2)$ is the associated likelihood and $\|\cdot\|^2$ denotes Euclidean norm squared. Then for $\hat{e}_{0:T}$ we let

$$f(\hat{e}_{0:T}|\sigma_e^2,\phi) = f(\hat{e}_{0:T-1}|\sigma_e^2) \times \left[f(\hat{e}_T|\sigma_e^2)\right]^{\phi}$$

be the corresponding tempered likelihood. Combined with an $IG\left(\alpha/2,\ \beta/2\right)$ prior that is proportional to

$$p(\sigma_e^2|\alpha,\beta) \propto \left(\sigma_e^2\right)^{-\frac{\alpha}{2}-1} \times \exp\left(-\frac{\beta}{2\sigma_e^2}\right)$$

Algorithm 1.4 IBIS algorithm with hybrid adaptive tempering for Gaussian Affine Term Structure Models

Initialize N_θ particles by drawing independently $\theta_i \sim \pi(\theta|Y_{0:t-1})$ with importance weights $\omega_i = 1$, $i = 1, \ldots, N_\theta$. For $t, \ldots, T$ and each time for all i:

1 *Set $\omega_i' = \omega_i$.*

2 *Calculate the incremental weights from*

$$u_t(\theta_i) = f(Y_t|Y_{t-1}, \theta_i)$$

3 *Update the importance weights ω_i to $\omega_i u_t(\theta_i)$.*

4 *If degeneracy criterion ESS(ω) is triggered, perform the following sub-steps:*

(a) *Set $\phi = 0$ and $\phi' = 0$.*

(b) *While $\phi < 1$*

i. *If degeneracy criterion ESS(ω'') is not triggered, where $\omega_i'' = \omega_i'[u_t(\theta_i)]^{1-\phi'}$, set $\phi = 1$, otherwise find $\phi \in [\phi', 1]$ such that ESS(ω''') is greater than or equal to the trigger, where $\omega_i''' = \omega_i'[u_t(\theta_i)]^{\phi-\phi'}$, for example using bisection method, see Kantas et al. (2014).*

ii. *Update the importance weights ω_i to $\omega_i'[u_t(\theta_i)]^{\phi-\phi'}$.*

iii. *Resample: Sample with replacement N_θ times from the set of θ_is according to their weights ω_i. The weights are then reset to one.*

iv. *Jitter: Replace θ_is with $\tilde{\theta}_i$s by running MCMC chains with each θ_i as input and $\tilde{\theta}_i$ as output, using likelihood given by $f(Y_{0:t-1}|\theta_i)[f(Y_t|\theta_i)]^\phi$. Set $\theta_i = \tilde{\theta}_i$.*

v. *Calculate the incremental weights from*

$$u_t(\theta_i) = f(Y_t|Y_{t-1}, \theta_i)$$

vi. *Set $\omega_i' = \omega_i$ and $\phi' = \phi$.*

we get tempered posterior as

$$\pi(\sigma_e^2|\hat{e}_{0:T}, \alpha, \beta, \phi) \propto f(\hat{e}_{0:T}|\sigma_e^2, \phi) \times p(\sigma_e^2|\alpha, \beta)$$

and eventually we arrive at

$$\pi(\sigma_e^2|\hat{e}_{0:T}, \alpha, \beta, \phi) \propto \left(\sigma_e^2\right)^{-\frac{1}{2}[\alpha+(T-1+\phi)(J-N)]-1} \times \exp\left(-\frac{\beta + \phi\|\hat{e}_T\|^2 + \sum_{t=0}^{T-1}\|\hat{e}_t\|^2}{2\sigma_e^2}\right)$$

what assuming

$$\tilde{\alpha} = \alpha + (T - 1 + \phi)(J - N)$$
$$\tilde{\beta} = \beta + \phi\|\hat{e}_T\|^2 + \sum_{t=0}^{T-1} \|\hat{e}_t\|^2$$

leads us to an $IG(\tilde{\alpha}/2,\ \tilde{\beta}/2)$ posterior for σ_e^2.

Following sections C.1 and C.2 in Online Appendix to Bauer (2018), and details from Lütkepohl (2005), to obtain tempered posterior distribution for $\lambda^{\mathcal{P}}$, it suffices to replace the Z and the x, according to notation in Online Appendix to Bauer (2018), with Z_ϕ and x_ϕ, which are defined as

$$Z_\phi = \left[Z_{0:T-2}, \sqrt{\phi} Z_{T-1}\right]$$

and

$$x_\phi = \begin{bmatrix} x_{1:T-1} \\ \sqrt{\phi} x_T \end{bmatrix}$$

where ϕ is the tempering parameter. In our notation, $Z = [Z_0, \ldots, Z_{T-1}]$ where $Z_t = [1, \mathcal{P}_t']'$, $t = 0, \ldots, T-1$, and $x = vec(X)$ where $X = [\mathcal{P}_1, \ldots, \mathcal{P}_T]$. Rest is straightforward to conclude from Bauer (2018).

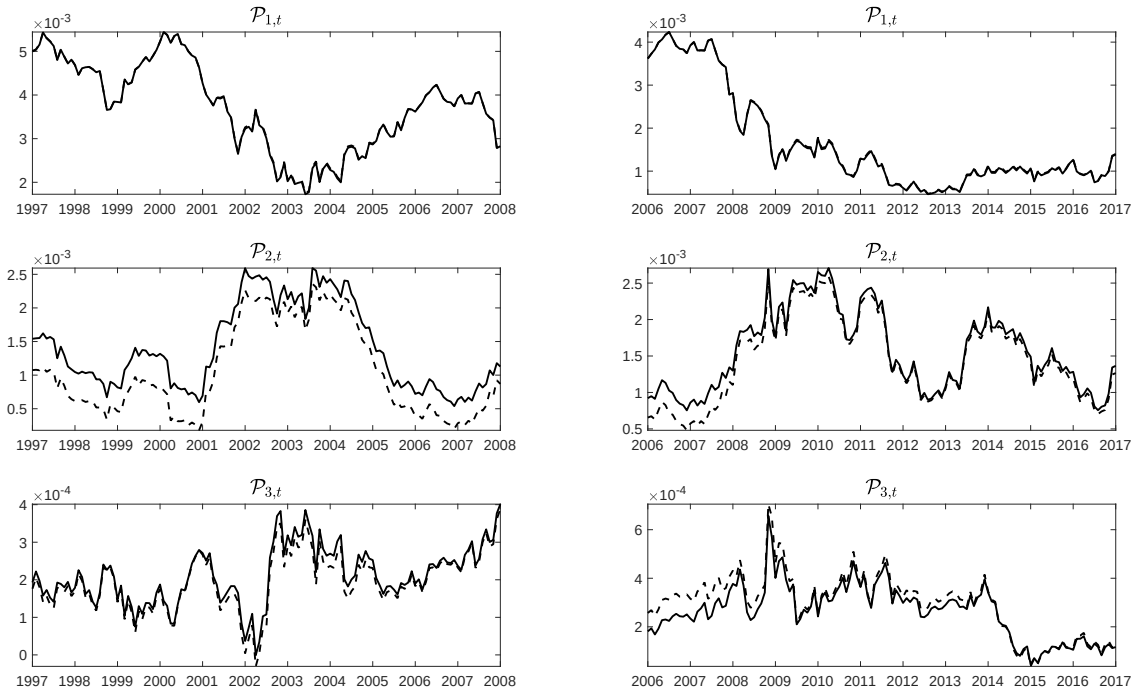

Figure 1.1: Principal components extracted from the yield curve for the two periods where the predictions were evaluated (January 1997 - end of 2007 and January 2006 - end of 2016). The dashed lines correspond to principal components based on loadings calculated from the entire sub-samples (January 1985 - end of 2007 and January 1990 - end of 2016). The solid lines, which are the ones used for the data analysis, were based on loadings calculated from the training periods only (January 1985 - end of 1996 and January 1990 - end of 2005).

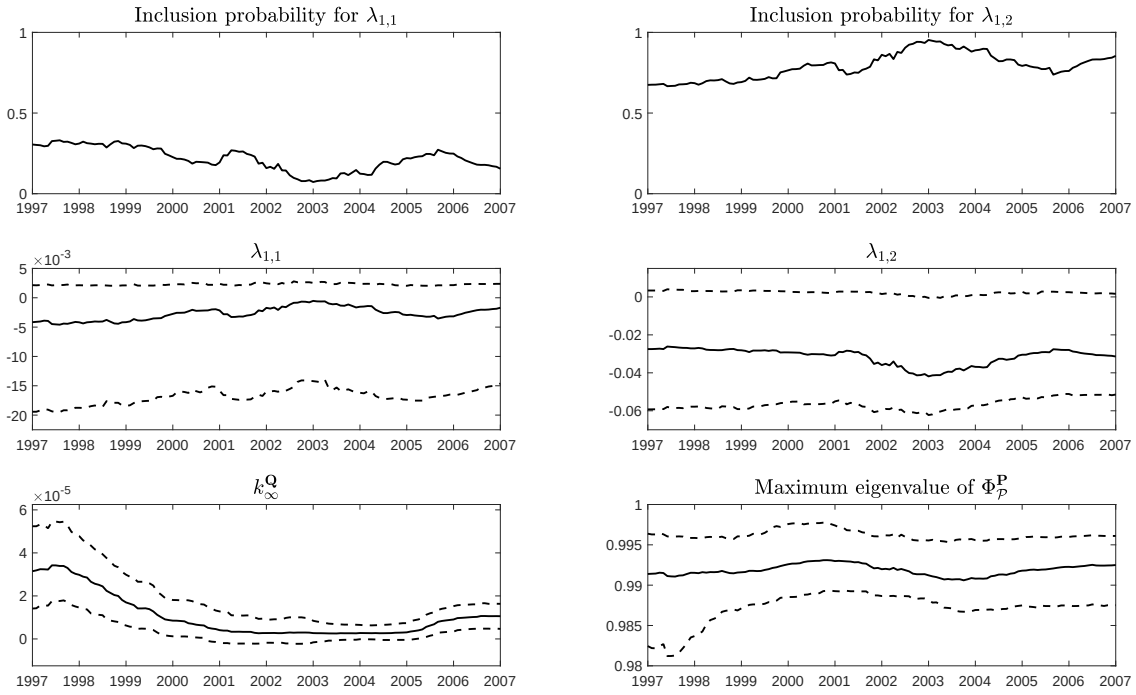

Figure 1.2: Output from model M_6 fitted on the first sub-sample (January 1985 to end of 2007). Here focus is given on the period between January 1997 and the end of 2006 for which predictions of the model were also evaluated, but the estimates are based on all the data from January 1985. The top row contains posterior probabilities of corresponding $\lambda_{i,j}$ being non-zero (i.e. inclusion probabilities). The medium row presents posterior means (solid line) and 95% credible intervals (dashed line) for the market price of risk parameters $\lambda_{1,1}$ and $\lambda_{1,2}$. The lower row plots real-time estimates of $k_\infty^{\mathbb{Q}}$, which is linked to the long-run mean of the short rate r_t under $\mathbb{Q}$, and of the largest eigenvalue of the feedback matrix $\Phi_{\mathcal{P}}^{\mathbb{P}}$, which is a measure of persistence.

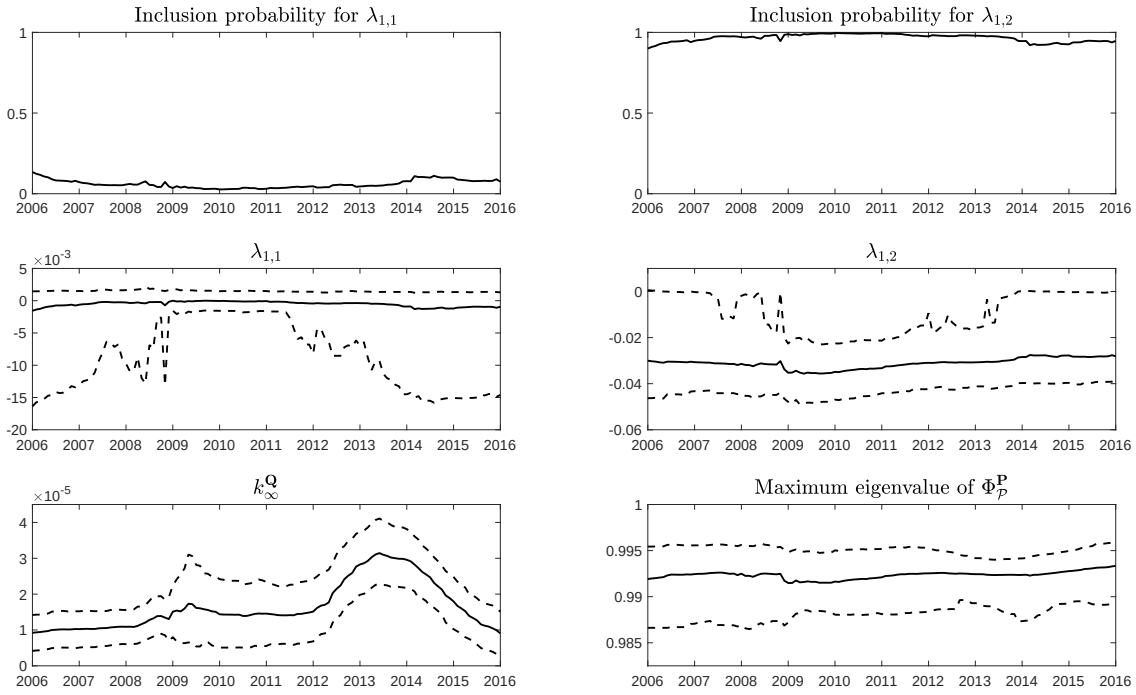

Figure 1.3: Output from model M_6 fitted on the second sub-sample (January 1990 to end of 2016). Here focus is given on the period between January 2006 and the end of 2015 for which predictions of the model were also evaluated, but the estimates are based on all the data from January 1990. The top row contains posterior probabilities of corresponding $\lambda_{i,j}$ being non-zero (i.e. inclusion probabilities). The medium row presents posterior means (solid line) and 95% credible intervals (dashed line) for the market price of risk parameters $\lambda_{1,1}$ and $\lambda_{1,2}$. The lower row plots real-time estimates of $k_\infty^{\mathbb{Q}}$, which is linked to the long-run mean of the short rate r_t under $\mathbb{Q}$, and of the largest eigenvalue of the feedback matrix $\Phi_{\mathcal{P}}^{\mathbb{P}}$, which is a measure of persistence.

Table 1.1: Out-of-sample statistical performance of bond excess return forecasts measured via R^2_{os} over multiple prediction horizons - period: January 1985 - end of 2007.

h\n	**2Y**	**3Y**	**4Y**	**5Y**	**7Y**	**10Y**
			M_0			
1m	0.03***	0.01**	0.00**	-0.02*	-0.03	-0.08
3m	0.06**	0.02**	0.01**	-0.04**	-0.06**	-0.22
6m	0.09**	0.04**	0.01**	-0.04**	-0.14*	-0.35
9m	0.07*	0.02*	-0.02*	-0.07*	-0.20	-0.43
12m	0.08*	0.02	-0.03	-0.06	-0.22	-0.43
			M_1			
1m	0.05***	0.04***	0.03***	0.03**	0.02**	0.03**
3m	0.07**	0.06**	0.06**	0.04*	0.06**	0.05**
6m	0.10**	0.09*	0.10**	0.09*	0.09*	0.08**
9m	0.08	0.08	0.10*	0.10*	0.11*	0.09*
12m	0.07	0.06	0.08	0.08	0.09*	0.08**
			M_2			
1m	0.04***	0.03**	0.03**	0.02*	0.02*	0.03**
3m	0.07**	0.06**	0.06**	0.04*	0.05*	0.05**
6m	0.09**	0.08*	0.09*	0.08*	0.08*	0.07*
9m	0.07	0.07	0.09	0.08	0.08*	0.06*
12m	0.05	0.04	0.05	0.05	0.06*	0.03*
			M_3			
1m	0.03**	0.02*	0.02*	0.02*	0.02*	0.02*
3m	0.05**	0.05**	0.05**	0.03**	0.05**	0.04**
6m	0.09**	0.07**	0.07**	0.06**	0.06***	0.05***
9m	0.13**	0.12**	0.11**	0.10***	0.10***	0.07***
12m	0.16**	0.13**	0.13**	0.11***	0.11***	0.08***
			M_4			
1m	0.05**	0.03***	0.03**	0.02**	0.02**	0.02**
3m	0.08**	0.07**	0.07**	0.05**	0.07**	0.06**
6m	0.10**	0.09**	0.10**	0.09**	0.09**	0.08*
9m	0.09	0.09*	0.10*	0.10*	0.11*	0.11*
12m	0.08	0.06	0.08	0.08*	0.10*	0.10**
			M_5			
1m	0.05***	0.04***	0.03***	0.03**	0.03**	0.02**
3m	0.07**	0.06**	0.06**	0.04*	0.06**	0.05*
6m	0.08*	0.07*	0.08*	0.07*	0.08*	0.07*
9m	0.08	0.08	0.09	0.09	0.09	0.10
12m	0.08	0.05	0.06	0.07	0.08	0.09
			M_6			
1m	0.04***	0.04***	0.03***	0.03**	0.02**	0.03***
3m	0.08**	0.07**	0.07**	0.05*	0.06**	0.06**
6m	0.10**	0.10**	0.10**	0.09*	0.09**	0.09**
9m	0.10*	0.10*	0.11*	0.11*	0.11*	0.10*
12m	0.09	0.08	0.09	0.09*	0.10*	0.08**

This table reports out-of-sample R^2 across alternative models and at different prediction horizons of h = 1-month, 3-month, 6-month, 9-month and 12-month. The seven forecasting models used are ATSM with alternative risk price restrictions. R^2 values are generated using the out-of-sample R^2 measure of Campbell and Thompson (2008). In particular, out-of-sample R^2 measures the predictive accuracy of bond excess return forecasts relative to the EH benchmark. The EH implies the historical mean being the optimal forecast of excess returns. Positive values of this statistic imply that the forecast outperforms the historical mean forecast and suggests evidence of time-varying return predictability. Statistical significance is measured using a one-sided Diebold-Mariano statistic with Clark-West adjustment, based on Newey-West standard errors. * denotes significance at 10%, ** significance at 5% and *** significance at 1% level. The in-sample period is January 1985 to end of 1996, and the out-of-sample period starts in January 1997 and ends in end of 2007.

Table 1.2: Out-of-sample statistical performance of bond excess return forecasts measured via R^2_{os} over multiple prediction horizons - period: January 1990 - end of 2016.

h\n	**2Y**	**3Y**	**4Y**	**5Y**	**7Y**	**10Y**
			M_0			
1m	-0.08	-0.10	-0.09	-0.06	-0.05	-0.05
3m	-0.16	-0.18	-0.17	-0.12	-0.12	-0.13
6m	-0.26	-0.31	-0.28	-0.21	-0.19	-0.17
9m	-0.46	-0.62	-0.60	-0.44	-0.36	-0.27
12m	-0.57	-0.81	-0.81	-0.59	-0.45	-0.27
			M_1			
1m	0.04***	0.04***	0.04***	0.04**	0.03**	0.03***
3m	0.06**	0.06***	0.06***	0.07***	0.07***	0.07***
6m	0.07**	0.09**	0.10**	0.11***	0.12***	0.15***
9m	0.07*	0.11**	0.14**	0.18***	0.18***	0.25***
12m	0.01	0.12**	0.17**	0.25***	0.26***	0.34***
			M_2			
1m	0.03**	0.03***	0.03***	0.03***	0.02**	0.02**
3m	0.06**	0.06***	0.05**	0.06**	0.06**	0.06**
6m	0.07**	0.09**	0.08**	0.10**	0.10**	0.13***
9m	0.09*	0.12**	0.13**	0.18**	0.16**	0.23***
12m	0.02	0.13**	0.17**	0.25***	0.25***	0.32***
			M_3			
1m	0.01**	-0.01	-0.01	-0.01	-0.01	-0.03
3m	0.01**	-0.04	-0.04	-0.04	-0.04	-0.10
6m	-0.01**	-0.09	-0.10	-0.09	-0.08	-0.13
9m	-0.01**	-0.19	-0.25	-0.23	-0.20	-0.25
12m	0.07**	-0.19	-0.32	-0.29	-0.27	-0.29
			M_4			
1m	0.03**	0.03**	0.03**	0.03**	0.02**	0.02**
3m	0.03	0.03*	0.03*	0.05**	0.05**	0.05**
6m	0.00	0.01	0.03	0.06*	0.07*	0.10**
9m	-0.03	-0.02	0.01	0.09*	0.10*	0.19***
12m	-0.11	-0.05	0.01	0.13*	0.16**	0.28***
			M_5			
1m	0.02**	0.02**	0.02**	0.02**	0.02**	0.02**
3m	0.05**	0.05**	0.05**	0.06*	0.05**	0.06**
6m	0.05*	0.06*	0.07**	0.09**	0.09**	0.12***
9m	0.04	0.08*	0.10*	0.15**	0.14**	0.22***
12m	-0.01	0.08*	0.13**	0.22**	0.23***	0.32***
			M_6			
1m	0.04**	0.03**	0.03**	0.03**	0.03**	0.02**
3m	0.06***	0.06***	0.06***	0.07***	0.07***	0.07***
6m	0.08**	0.09**	0.10**	0.12***	0.11***	0.14***
9m	0.08*	0.12**	0.14**	0.18***	0.17***	0.24***
12m	0.03	0.13**	0.17**	0.25***	0.26***	0.33***

This table reports out-of-sample R^2 across alternative models and at different prediction horizons of h = 1-month, 3-month, 6-month, 9-month and 12-month. The seven forecasting models used are ATSM with alternative risk price restrictions. R^2 values are generated using the out-of-sample R^2 measure of Campbell and Thompson (2008). In particular, out-of-sample R^2 measures the predictive accuracy of bond excess return forecasts relative to the EH benchmark. The EH implies the historical mean being the optimal forecast of excess returns. Positive values of this statistic imply that the forecast outperforms the historical mean forecast and suggests evidence of time-varying return predictability. Statistical significance is measured using a one-sided Diebold-Mariano statistic with Clark-West adjustment, based on Newey-West standard errors. * denotes significance at 10%, ** significance at 5% and *** significance at 1% level. The in-sample period is January 1990 to end of 2005, and the out-of-sample period starts in January 2006 and ends in end of 2016.

Table 1.3: Out-of-sample statistical performance of bond excess return forecasts measured via log predictive score over multiple prediction horizons - period: January 1985 - end of 2007.

h\n	**2Y**	**3Y**	**4Y**	**5Y**	**7Y**	**10Y**
			M_0			
1m	-0.01	0.00	0.00	-0.01	-0.02	-0.03
3m	0.05	0.04	0.04	0.01	-0.01	-0.07
6m	0.06	0.03	0.04	0.02	-0.04	-0.09
9m	0.01	0.00	0.01	0.00	-0.04	-0.07
12m	-0.05	-0.05	-0.03	-0.02	-0.06	-0.08
			M_1			
1m	0.01	0.00	0.02**	0.00	0.00	0.01
3m	0.04**	0.04**	0.04***	0.03*	0.02*	0.01
6m	0.05	0.05*	0.06**	0.05*	0.04*	0.03*
9m	0.03	0.05	0.06*	0.05	0.04	0.05*
12m	0.01	0.03	0.05	0.05	0.05	0.05*
			M_2			
1m	-0.01	0.01	0.02*	0.00	-0.01	-0.01
3m	0.04**	0.04**	0.05**	0.02	0.02	0.01
6m	0.06**	0.06**	0.06**	0.03*	0.03	0.03
9m	0.04	0.06	0.06*	0.03	0.03	0.03
12m	-0.01	0.03	0.04	0.05	0.03	0.03
			M_3			
1m	-0.01	0.00	0.01	0.00	-0.01	0.00
3m	0.05**	0.04*	0.04**	0.03**	0.01	0.01
6m	0.05	0.05	0.05**	0.05***	0.03**	0.03***
9m	0.06	0.07*	0.08**	0.06**	0.05**	0.04***
12m	0.05	0.06	0.08*	0.07**	0.06***	0.06***
			M_4			
1m	0.00	0.01	0.00	0.00	-0.01	0.00
3m	0.05**	0.03*	0.05***	0.04***	0.03**	0.00
6m	0.06*	0.07**	0.06***	0.04*	0.05**	0.04**
9m	0.05	0.06	0.07**	0.06**	0.05*	0.04**
12m	0.03	0.03	0.05	0.05	0.05	0.05*
			M_5			
1m	0.00	0.00	0.01	-0.01	-0.01	0.00
3m	0.04*	0.03*	0.04**	0.03***	0.02*	0.01
6m	0.05	0.04	0.04	0.02	0.03	0.03
9m	0.03	0.04	0.05	0.05	0.04	0.03
12m	0.00	0.02	0.03	0.05	0.03	0.04*
			M_6			
1m	0.01	0.02**	0.01	0.00	0.00	0.00
3m	0.04**	0.04**	0.04***	0.02*	0.03*	0.01
6m	0.04	0.05**	0.06***	0.05**	0.04**	0.04**
9m	0.05	0.06*	0.07**	0.06**	0.06**	0.05**
12m	0.02	0.04	0.06*	0.06*	0.05*	0.05*

This table reports out-of-sample log predictive score (LS) across alternative models and at different prediction horizons of h = 1-month, 3-month, 6-month, 9-month and 12-month. The seven forecasting models used are ATSM with alternative risk price restrictions. The EH implies the historical mean being the optimal forecast of excess returns. Positive values of this statistic imply that the forecast outperforms the historical mean forecast and suggests evidence of time-varying return predictability. Statistical significance is measured using a one-sided Diebold-Mariano statistic computed with Newey-West standard errors. * denotes significance at 10%, ** significance at 5% and *** significance at 1% level. The in-sample period is January 1985 to end of 1996, and the out-of-sample period starts in January 1997 and ends in end of 2007.

Table 1.4: Out-of-sample statistical performance of bond excess return forecasts measured via log predictive score over multiple prediction horizons - period: January 1990 - end of 2016.

h\n	**2Y**	**3Y**	**4Y**	**5Y**	**7Y**	**10Y**
			M_0			
1m	-0.01	-0.03	-0.04	-0.04	-0.03	-0.07
3m	0.01	-0.04	-0.06	-0.05	-0.10	-0.05
6m	-0.23	-0.22	-0.14	-0.12	-0.10	-0.09
9m	-0.27	-0.26	-0.23	-0.17	-0.16	-0.13
12m	-0.16	-0.19	-0.21	-0.19	-0.18	-0.13
			M_1			
1m	-0.01	-0.01	0.01*	0.00	0.00	-0.04
3m	0.07**	0.05**	0.04***	0.04***	-0.01	0.02
6m	0.10**	0.07*	0.05**	0.05**	0.05**	0.06**
9m	0.15***	0.10**	0.07***	0.05**	0.06**	0.06*
12m	0.18***	0.13***	0.08**	0.07***	0.06**	0.09**
			M_2			
1m	0.01	0.00	0.00	0.00	0.00	-0.05
3m	0.08***	0.07***	0.05***	0.03**	0.03**	0.03*
6m	0.08	0.07*	0.05*	0.05**	0.06***	0.06*
9m	0.13*	0.09**	0.05*	0.05*	0.03	0.06*
12m	0.17***	0.12***	0.08**	0.07***	0.05*	0.08*
			M_3			
1m	-0.01	-0.02	-0.02	-0.02	-0.01	-0.01
3m	0.06*	0.03	0.02	-0.01	-0.04	-0.08
6m	0.08	0.03	-0.01	-0.02	-0.02	-0.06
9m	0.14**	0.04	-0.03	-0.06	-0.07	-0.09
12m	0.25***	0.11**	0.00	-0.05	-0.08	-0.10
			M_4			
1m	0.00	0.00	0.01	0.00	0.00	-0.03
3m	0.07**	0.05**	0.03*	0.02	0.02*	0.01
6m	0.07	0.05	0.02	0.03	0.02	0.03
9m	0.12*	0.06	0.03	0.02	0.02	0.05
12m	0.13*	0.08	0.04	0.04	0.03	0.06
			M_5			
1m	0.02**	0.00	0.00	0.00	0.00	-0.05
3m	0.07**	0.05**	0.03**	0.03***	0.04***	0.02
6m	0.10**	0.07**	0.04*	0.05**	0.04*	0.05*
9m	0.14**	0.09**	0.05*	0.04	0.03	0.06*
12m	0.17***	0.12***	0.08**	0.07**	0.06*	0.08**
			M_6			
1m	0.00	0.00	0.00	-0.01	-0.01	-0.04
3m	0.06	0.05**	0.03**	0.04***	0.02	0.03*
6m	0.10*	0.08**	0.04	0.05**	0.05**	0.05*
9m	0.14**	0.09**	0.06**	0.05**	0.05**	0.07*
12m	0.18***	0.12***	0.07***	0.06**	0.05**	0.08**

This table reports out-of-sample log predictive score (LS) across alternative models and at different prediction horizons of h = 1-month, 3-month, 6-month, 9-month and 12-month. The seven forecasting models used are ATSM with alternative risk price restrictions. The EH implies the historical mean being the optimal forecast of excess returns. Positive values of this statistic imply that the forecast outperforms the historical mean forecast and suggests evidence of time-varying return predictability. Statistical significance is measured using a one-sided Diebold-Mariano statistic computed with Newey-West standard errors. * denotes significance at 10%, ** significance at 5% and *** significance at 1% level. The in-sample period is January 1990 to end of 2005, and the out-of-sample period starts in January 2006 and ends in end of 2016.

Table 1.5: Out-of-sample economic performance of bond excess return forecasts measured via certainty equivalent returns (%) over multiple prediction horizons - period: January 1985 - end of 2007.

h\n	**2Y**	**3Y**	**4Y**	**5Y**	**7Y**	**10Y**
			M_0			
1m	-1.07	-2.95	-2.52	-3.76	-4.66	-12.10
3m	-1.77	-2.79	-2.44	-3.05	-4.97	-10.63
6m	-2.01	-2.64	-2.57	-2.98	-4.93	-7.72
9m	-3.92	-3.87	-3.35	-3.38	-4.77	-6.25
12m	-4.69	-3.97	-3.18	-2.77	-4.19	-5.34
			M_1			
1m	1.64	0.27	0.58	1.30	1.63	1.54
3m	1.70	1.23	1.43	1.81	1.53	1.48
6m	2.00	2.28*	2.58**	2.70**	2.90**	3.07**
9m	0.84	1.83	2.54**	2.71**	3.21***	3.38***
12m	-0.22	1.05	1.75	2.17**	2.76***	2.97***
			M_2			
1m	1.26	-0.34	-0.14	0.70	0.80	0.88
3m	2.00	1.61	1.70	2.02	1.45	1.36
6m	1.52	1.96*	2.29**	2.38**	2.53**	2.78**
9m	0.57	1.54	2.21**	2.29**	2.74***	2.97***
12m	-0.36	0.92	1.52	1.84*	2.38**	2.59**
			M_3			
1m	1.21	1.82	2.34	3.11	3.00*	1.52
3m	1.40	1.74*	2.08**	2.42**	1.52*	0.55
6m	0.88	1.33	1.42**	1.34**	1.20**	1.22**
9m	0.72	1.26	1.43	1.12	1.15*	1.22*
12m	0.20	1.22	1.41	1.21	1.24	1.16
			M_4			
1m	1.04	-0.18	-0.14	0.28	0.21	-1.17
3m	1.84	1.46	1.69	2.09	1.51	0.47
6m	1.28	1.42	1.68*	1.79*	1.80*	1.70*
9m	0.69	1.25	1.69**	1.73**	1.97**	2.01***
12m	0.01	0.76	1.11	1.36*	1.61**	1.68***
			M_5			
1m	0.79	0.17	0.58	1.19	1.73	0.08
3m	0.82	0.47	0.65	1.03	0.68	-0.25
6m	0.48	0.64	0.78	0.92	0.94	0.97
9m	0.23	0.61	0.82	0.80	0.98	1.18
12m	0.10	0.60	0.62	0.76	0.87	1.02*
			M_6			
1m	1.85	0.92	1.24	1.97	2.01	1.85
3m	2.27	1.77	1.86	2.18	1.66	1.41
6m	1.99*	2.17*	2.37**	2.43**	2.54**	2.75**
9m	1.07	1.86*	2.42**	2.46**	2.83***	2.96***
12m	-0.04	1.09	1.68*	1.91**	2.37***	2.50**

This table reports annualized certainty equivalent returns (CERs) across alternative models and at different prediction horizons of h = 1-month, 3-month, 6-month, 9-month and 12-month. The coefficient of risk aversion is $\gamma = 3$. No portfolio constraints are imposed. CERs are generated by out-of-sample forecasts of bond excess returns and are reported in %. At every time step t, an investor with power utility preferences evaluates the entire predictive density of bond excess returns and solves the asset allocation problem, thus optimally allocating her wealth between a riskless bond and risky bonds with maturities 2, 3, 4, 5, 7 and 10-years. CER is then defined as the value that equates the average utility of each alternative model against the average utility of the EH benchmark. The seven forecasting models used are ATSM with alternative risk price restrictions. Positive values indicate that the models perform better than the EH benchmark. Statistical significance is measured using a one-sided Diebold-Mariano statistic computed with Newey-West standard errors. * denotes significance at 10%, ** significance at 5% and *** significance at 1% level. The in-sample period is January 1985 to end of 1996, and the out-of-sample period starts in January 1997 and ends in end of 2007.

Table 1.6: Out-of-sample economic performance of bond excess return forecasts measured via certainty equivalent returns (%) over multiple prediction horizons - period: January 1990 - end of 2016.

h\n	**2Y**	**3Y**	**4Y**	**5Y**	**7Y**	**10Y**
			M_0			
1m	-13.38	-15.71	-14.85	-11.84	-11.51	-16.85
3m	-10.55	-11.77	-10.64	-7.34	-7.73	-9.61
6m	-12.93	-12.74	-11.25	-8.58	-7.78	-7.37
9m	-11.10	-11.45	-10.81	-9.18	-8.32	-6.93
12m	-9.13	-9.62	-9.48	-8.09	-7.34	-5.63
			M_1			
1m	4.66**	3.41**	3.77**	3.80**	5.37**	3.59
3m	4.72***	3.65**	3.04**	3.40**	3.04**	2.39
6m	5.18***	3.94***	2.64***	2.67***	2.35**	2.34**
9m	5.82***	4.16***	2.75***	2.36***	1.92**	2.36**
12m	6.37***	5.32***	3.76***	3.22***	2.61***	2.87***
			M_2			
1m	2.23	1.48	1.41	1.89	3.46	0.64
3m	3.72**	2.90*	2.30	2.80*	2.20	1.70
6m	4.67***	3.43***	2.03**	2.05**	1.58	1.80
9m	5.23***	3.58***	2.24**	1.93**	1.51	2.01
12m	5.82***	4.64***	3.05***	2.65***	2.04**	2.43**
			M_3			
1m	-4.17	-6.04	-4.34	-3.18	0.29	-6.08
3m	-3.29	-4.27	-3.76	-3.02	-3.01	-7.59
6m	-2.86	-4.06	-4.39	-3.76	-3.25	-5.35
9m	-1.73	-3.63	-4.66	-4.77	-4.56	-5.42
12m	0.41	-1.37	-3.05	-3.43	-3.78	-4.37
			M_4			
1m	3.33	1.97	2.28	2.67	3.96*	0.39
3m	3.20*	2.15	1.50	2.18	1.60	0.79
6m	3.16**	1.91	0.69	0.95	0.46	0.73
9m	3.65**	2.08	0.85	0.65	0.23	0.80
12m	4.21***	3.15*	1.72	1.41	0.91	1.39
			M_5			
1m	2.07	0.73	0.82	1.08	2.35	0.27
3m	4.38***	3.50**	2.77**	3.05*	2.36*	1.62
6m	4.67***	3.42***	2.15**	2.18**	1.72*	1.64
9m	5.26***	3.57***	2.19**	1.88**	1.47	1.88
12m	5.67***	4.64***	3.16***	2.72***	2.21**	2.51**
			M_6			
1m	3.70**	2.01	2.05	1.90	3.48*	2.21
3m	4.81***	3.76**	3.14**	3.49**	2.99**	2.21
6m	5.16***	3.94***	2.65***	2.67***	2.27**	2.19*
9m	5.88***	4.19***	2.90***	2.51***	2.08**	2.41**
12m	6.35***	5.36***	3.83***	3.32***	2.77***	2.89**

This table reports annualized certainty equivalent returns (CERs) across alternative models and at different prediction horizons of h = 1-month, 3-month, 6-month, 9-month and 12-month. The coefficient of risk aversion is $\gamma = 3$. No portfolio constraints are imposed. CERs are generated by out-of-sample forecasts of bond excess returns and are reported in %. At every time step t, an investor with power utility preferences evaluates the entire predictive density of bond excess returns and solves the asset allocation problem, thus optimally allocating her wealth between a riskless bond and risky bonds with maturities 2, 3, 4, 5, 7 and 10-years. CER is then defined as the value that equates the average utility of each alternative model against the average utility of the EH benchmark. The seven forecasting models used are ATSM with alternative risk price restrictions. Positive values indicate that the models perform better than the EH benchmark. Statistical significance is measured using a one-sided Diebold-Mariano statistic computed with Newey-West standard errors. * denotes significance at 10%, ** significance at 5% and *** significance at 1% level. The in-sample period is January 1990 to end of 2005, and the out-of-sample period starts in January 2006 and ends in end of 2016.

Chapter 2

On Unspanned Latent Factors in Dynamic Term Structure Models

2.1 Introduction

2.1.1 Unobserved Factors or Macroeconomic Variables?

The presence of information hidden from the yield curve, yet highly relevant for predicting bond excess returns has attracted particular attention in recent literature. The reason being that several empirical studies (see, Ludvigson and Ng (2009), Cooper and Priestley (2009), Duffee (2011), Wright (2011), Joslin et al. (2014), Cieslak and Povala (2015), Gargano et al. (2019), Bianchi et al. (2021) and Li et al. (2021)) cast doubt on the fundamental and powerful stylized fact of the term structure of interest rates, which suggests that current yields contain all relevant information for forecasting future yields, returns and bond risk premia. In particular, the aforementioned studies have documented evidence of various variables, containing significant predictive power above and beyond the yield curve. While the literature

has mainly focused on factors extracted from macroeconomic time-series[1], yet the list of unspanned risk factors can be long and diverse as argued by Joslin et al. (2014).

More recently, however, the capability of unspanned macroeconomic risks, offering evidence of statistical and economic benefits to bond investors, has come under scrutiny. First, Bauer and Hamilton (2018) question earlier results, concluding that, in many cases, prior evidence of excess return predictability is weaker than initially documented and far from statistically significant. Most importantly, in contrast to the in-sample analysis, they conclude that the addition of extra macroeconomic factors leads to less stable and accurate predictions and, ultimately, to the deterioration of the out-of-sample performance[2]. A similar observation is made by Barillas (2011) who argue that an unspanned factor adds little predictability to the term structure and Giacoletti et al. (2021) who use Bayesian learning to evaluate the real-time out-of-sample performance of Joslin et al. (2014) macro-finance dynamic model and conclude that it actually underperforms the simple, nested, yields-only model.

Second, related literature has recently noted the importance of a real-time investor having access to real-time macroeconomic information as opposed to a fully revised information set. Recent evidence highlights the discrepancy on the out-of-sample predictive performance of bond excess returns when real-time macro variables are used as predictors vis-à-vis fully revised factors. In particular, Ghysels et al. (2018), argue that the evident predictive power of macroeconomic variables is largely due to data revisions and as such, it diminishes when real-time macro data is used, a conclusion that is also reached by Fulop et al. (2019). Most importantly, statistical predictability due to macroeconomic factors is not turned into utility gains for bond investors, in contrast to what was initially argued by Gargano et al. (2019) and Bianchi et al. (2021).

[1] e.g. the output gap of Cooper and Priestley (2009), the 'real' and 'inflation' factors of Ludvigson and Ng (2009), the measures of economic activity and inflation of Joslin et al. (2014), the long-run inflation expectation of Cieslak and Povala (2015), etc.

[2] A notable exception is the case of Ludvigson and Ng (2009) factors which slightly improve the predictive performance of the models both in-sample and out-of-sample.

Third, a notable exception to the analysis of unspanned macroeconomic risks is the study of Duffee (2011) (and Barillas (2011)) which introduces an unspanned latent factor framework, that is not reliant on macroeconomic data. According to Duffee (2011), estimation of term structure models using directly observable macroeconomic data might be a rather powerful approach, yet it comes with a great risk of misspecification.

2.1.2 Unspanned Latent Approach

The above-mentioned issues raise important questions on the validity of prior empirical and theoretical results. With this in mind, we propose a novel class of arbitrage-free unspanned DTSM, that embeds a stochastic market price of risk specification. Our approach resembles the macro-finance framework of Joslin et al. (2014) and the unspanned latent factor framework of Duffee (2011), in that the model is factorised into a 'spanned' component, where risk factors can be retrieved from the information provided by current yield curve, as well as an 'unspanned' component, that could include factors extracted from macroeconomic variables. It is assumed that the latter is not determined by the yield curve data, yet remains highly relevant for inference and, more importantly, prediction purposes. The developed setup is then used to explore how valuable is the unspanned information to a real-time Bayesian investor seeking to forecast future excess bond returns and generate systematic economic gains. Our approach differs from previous studies and offers several advantages.

First, we depart from Joslin et al. (2014) in that the unspanned component is regarded as a latent stochastic process, rather than consisting of directly observable variables and as such, our analysis is independent from the debate between 'fully-revised' vis-à-vis 'real-time' macros. In line with Duffee (2011)[3], we estimate this latent factor using Kalman filter, which allows us to infer any hidden information

[3]Duffee (2011) does not explore the econometric identification of such a model, nor does he empirically implement a dynamic term structure model with unspanned risks. We further depart from Duffee (2011) by allowing a time-varying price of risk not only on the level factor but on the slope and curvature as well, whereas Duffee (2011) assumes compensation for slope risk to be fixed.

from the term structure dynamics. In particular, we develop a suitable sequential Monte Carlo inferential and prediction scheme that draws from Chopin (2002, 2004) and takes advantage of the linear state space structure of the model by incorporating the associated Kalman filter. Doing so, the unobserved component may absorb model misspecification or be regarded as this element in the model which is capturing the dynamics of wider market environment. There are three main aims to our formulation. One, dimension reduction is pursued, when it comes to risk prices, in order to consider parsimonious models that contain almost all the information in the yield curve data. Two, we aim to incorporate in the model additional latent factors, avoiding their direct involvement in the bond pricing procedure, so that they cannot be fully recovered by the cross section of yields. Hence, such latent factors are meant to capture information which is hidden from the yield curve and may be impossible to summarise using information coming solely from observable macroeconomic time-series (see, Ludvigson and Ng (2009)). Three, we seek to investigate whether there is a direct link between information incorporated into the hidden component and macroeconomic activity.

Second, we take the perspective of a real-time Bayesian learner, who takes into account all relevant uncertainties and updates her beliefs using a DTSM with unspanned latent factors. Our setup directly produces the entire forecasting distribution, through predictive densities, as new information arrives. Thus, it can naturally assess the predictive performance of any model, across maturities and prediction horizons.

Third, while prior studies have investigated the predictability of excess returns in-sample, there is less empirical evidence on the out-of-sample performance of the unspanned (macro) DTSM. We revisit the out-of-sample performance of our model under different restrictions for the market price of risk specification. Most importantly, we attempt to assess whether a bond investor can actually exploit the evident statistical predictability when making investment decisions. Is investors' economic utility improved by making full use of unspanned latent factors?

2.1.3 Outline

The remainder of this chapter is organized as follows. Section 2.2 describes the modelling framework. Section 2.3 presents the sequential learning with latent processes and forecasting procedure along with the framework for assessing the predictive and economic performance of models. Section 2.4 discusses the data and the sample period used and presents the family of models considered in this chapter. Section 2.5 discusses the results both in terms of predictive performance and economic value, including the associated explanatory power, as well as reveals the links between the latent factors and the economy. Finally, Section 2.6 concludes the chapter by providing some relevant discussion.

2.2 Dynamic Term Structure Model, Likelihood, and Latent Factors

2.2.1 Incorporating Unspanned Latent Components

In what follows, we consider an extension of the model presented in Section 1.2.1. Our approach resembles the framework of Joslin et al. (2014) in that the model is factorised into a 'spanned' component, i.e. risk factors that can be retrieved by the information provided in historical yield curve data, as well as an 'unspanned' component that could include background factors such as macroeconomic variables. It is assumed that the latter is not determined by the yield curve data but it remains highly relevant for the inference and, more importantly, prediction purposes.

The point where our approach differs from Joslin et al. (2014) is that the unspanned component is regarded as a latent stochastic process, rather than consisting of directly observable macroeconomic variables. In this way the unobserved component may absorb model misspecification or be regarded as market environment specific factor. There are two main aims to this alternative formulation. First, di-

mension reduction is pursued in order to consider parsimonious models that contain almost all the information in the yield curve data. This is achieved by the PCA approach as in (1.8) but the projection is now made to R observable factors, where $R < N$. Second, we would like to incorporate additional latent components in the model avoiding their direct involvement in the bond pricing procedure, so that they cannot be fully recovered by the yield data.

The extended model can be regarded as the following transformation of X_t that decomposes into the spanned component $\mathcal{P}_t$ and the unspanned latent Z_t:

$$\begin{bmatrix} \mathcal{P}_t \\ Z_t \end{bmatrix} = \begin{bmatrix} W_R A_X \\ \gamma_0 \end{bmatrix} + \begin{bmatrix} W_R B_X \\ \gamma_1 \end{bmatrix} X_t \tag{2.1}$$

where $\mathcal{P}_t$ are the first R principal components extracted from the yield data, W_R consists of the first R rows of the matrix W containing the PCA weights, A_X and B_X are the loadings matrices appearing in (1.5) (and are of dimension $(J \times 1)$ and $(J \times N)$ respectively). The vector γ_0 and matrix γ_1 (of dimension $(N-R) \times 1$ and $(N-R) \times N$) may contain arbitrary constants subject to the restrictions in Appendix B of Joslin et al. (2014), mainly ensuring that the matrix $[W_R B_X;\ \gamma_1]$ is invertible.

Given the linear nature of (2.1) the $\mathbb{P}$-dynamics of $[\mathcal{P}_t;\ Z_t]$ are of the following form

$$\begin{bmatrix} \mathcal{P}_t \\ Z_t \end{bmatrix} = \begin{bmatrix} \mu_{\mathcal{P}}^{\mathbb{P}} \\ \mu_Z^{\mathbb{P}} \end{bmatrix} + \begin{bmatrix} \Phi_{\mathcal{P}}^{\mathbb{P}} & \Phi_{\mathcal{P}Z}^{\mathbb{P}} \\ \Phi_{Z\mathcal{P}}^{\mathbb{P}} & \Phi_Z^{\mathbb{P}} \end{bmatrix} \begin{bmatrix} \mathcal{P}_{t-1} \\ Z_{t-1} \end{bmatrix} + \Sigma_{\mathcal{P}Z} \varepsilon_t \tag{2.2}$$

where $\mu_{\mathcal{P}}^{\mathbb{P}}$ and $\Phi_{\mathcal{P}}^{\mathbb{P}}$ are defined in (1.15) and (1.16), respectively. The covariance matrix $\Sigma_{\mathcal{P}Z}\Sigma'_{\mathcal{P}Z}$ is assumed to be block diagonal with components $\Sigma_{\mathcal{P}}\Sigma'_{\mathcal{P}}$ and $\Sigma_Z\Sigma'_Z$, where Σ_Z is a $(N-R) \times (N-R)$ diagonal matrix with positive entries. For identification purposes, as well as practical considerations, we proceed by setting $\mu_Z^{\mathbb{P}}$ and $\Phi_{Z\mathcal{P}}^{\mathbb{P}}$ to a vector and a matrix of zeros, whereas $\Phi_{\mathcal{P}Z}^{\mathbb{P}}$ is a $(N-R)$-dimensional identity matrix. Finally, $\Phi_Z^{\mathbb{P}}$ is a $(N-R) \times (N-R)$ diagonal matrix with non-zero entries taking

values in $(-1, 1)$, to guarantee stability. This lets write the model as

$$\mathcal{P}_t = \mu_{\mathcal{P}}^{\mathbb{P}} + \Phi_{\mathcal{P}}^{\mathbb{P}} \mathcal{P}_{t-1} + Z_{t-1} + \Sigma_{\mathcal{P}} \varepsilon_t^{\mathcal{P}} \tag{2.3}$$

$$Z_t = \Phi_Z^{\mathbb{P}} Z_{t-1} + \Sigma_Z \varepsilon_t^Z \tag{2.4}$$

where

$$\varepsilon_t^{\mathcal{P}} \sim N(0,\ I_R)$$
$$\varepsilon_t^Z \sim N(0,\ I_{N-R})$$

Compared to the standard case in Section 1.2.1, the model of (2.3) and (2.4) above has $\mu_{\mathcal{P}}^{\mathbb{P}} + Z_{t-1}$, instead of just $\mu_{\mathcal{P}}^{\mathbb{P}}$, in the drift. We can therefore view the Z_ts as random effects operating at each time interval $[t, t+1)$ and being correlated. To ensure that Z_t is not spanned by $\mathcal{P}_t$ we set r_t as in (1.12) and define the pricing kernel as

$$\mathcal{M}_{t+1} = \exp(-r_t - \frac{1}{2}\lambda_t'\lambda_t - \lambda_t'\varepsilon_{t+1}^{\mathcal{P}})$$

where $\varepsilon_{t+1}^{\mathcal{P}}$ consists of the first R components of ε_t in (2.2) and the market price of risk λ_t is defined conditional on Z_t as

$$\begin{aligned}\lambda_t &= \Sigma_{\mathcal{P}}^{-1}\left[\mu_{\mathcal{P}}^{\mathbb{P}} - \mu_{\mathcal{P}}^{\mathbb{Q}} + (\Phi_{\mathcal{P}}^{\mathbb{P}} - \Phi_{\mathcal{P}}^{\mathbb{Q}})\mathcal{P}_t + \Phi_{\mathcal{P}Z}^{\mathbb{P}} Z_t\right] \\ &= \Sigma_{\mathcal{P}}^{-1}\left(\lambda_{0\mathcal{P}} + \lambda_{1\mathcal{P}}\mathcal{P}_t + Z_t\right)\end{aligned}$$

Appropriate choice of restrictions on $\lambda_{0\mathcal{P}}$ and $\lambda_{1\mathcal{P}}$ in λ_t is assumed, see Section 2.2.3 for related discussion.

The model is completed by setting the $\mathbb{Q}$-dynamics for $\mathcal{P}_t$ as in (1.7). Note that since Z_t is not part of the pricing procedure its risk-neutral dynamics are not required.

2.2.2 Kalman Filtering the Unobserved

The model of (2.3) and (2.4), introduced in the previous section, can be slightly modified to match the general linear Gaussian state-space model notation, see for example Durbin and Koopman (2012), by working with $\alpha_t = Z_{t-1}$, instead of Z_{t-1}, for convenience, and setting the initial conditions as

$$\alpha_0 \sim N(a_{0|0},\ P_{0|0})$$

with $a_{0|0} = 0$ and $P_{0|0} = \Sigma_Z \Sigma_Z'$. This allows for the Kalman filter to be applied to marginalise out Z_t and evaluate the densities $f^{\mathbb{P}}(\mathcal{P}_t|\mathcal{P}_{t-1}, \theta)$, as in (2.10) below, with θ denoting the relevant parameters specified in the next section. More specifically, we let

$$\mathcal{P}_t - \mu_{\mathcal{P}}^{\mathbb{P}} - \Phi_{\mathcal{P}}^{\mathbb{P}} \mathcal{P}_{t-1} = s_t \tag{2.5}$$

and rewrite (2.3) and (2.4) as

$$s_t = \alpha_t + \Sigma_{\mathcal{P}} \varepsilon_t^{\mathcal{P}} \tag{2.6}$$

$$\alpha_{t+1} = \Phi_Z^{\mathbb{P}} \alpha_t + \widehat{\Sigma}_Z \varepsilon_t^Z \tag{2.7}$$

Then, as in Durbin and Koopman (2012), starting with $a_{0|0}$ and $P_{0|0}$ defined above for $t = 0$, the Kalman filter prediction step is

$$a_{t+1} = \Phi_Z^{\mathbb{P}} a_{t|t}$$

$$P_{t+1} = \Phi_Z^{\mathbb{P}} P_{t|t} \Phi_Z^{\mathbb{P}\prime} + \widehat{\Sigma}_Z \widehat{\Sigma}_Z'$$

for each $t = 1, ..., T-1$, whereas the Kalman filter update step is

$$
\begin{aligned}
v_t &= s_t - a_t \\
F_t &= P_t + \Sigma_{\mathcal{P}} \Sigma_{\mathcal{P}}' \\
K_t &= P_t F_t^{-1} \\
a_{t|t} &= a_t + K_t v_t \qquad (2.8) \\
P_{t|t} &= P_t - K_t P_t \qquad (2.9)
\end{aligned}
$$

In the above, a_{t+1} is state prediction, P_{t+1} is its covariance, v_t is prediction error, F_t is its covariance, K_t is the Kalman gain, $a_{t|t}$ is state update and $P_{t|t}$ is its covariance. Consequently, latent α_t (or Z_{t-1}) is distributed as $N(a_{t|t},\ P_{t|t})$, where the moments are obtained from (2.8) and (2.9). We eventually arrive at the log-likelihood representation of $f^{\mathbb{P}}(\mathcal{P}_t|\mathcal{P}_{t-1}, \theta)$, first due to Schweppe (1965), which is following

$$
\log f^{\mathbb{P}}(\mathcal{P}_t|\mathcal{P}_{t-1}, \theta) = -\frac{R}{2} \log 2\pi - \frac{1}{2} \left(\log |F_t| + v_t' F_t^{-1} v_t \right) \qquad (2.10)
$$

where $|\cdot|$ is matrix determinant. The predictive distribution of $\mathcal{P}_{T+1}$, based on observations up to time T, is also straightforward to obtain from (2.5), (2.6) and (2.7) as

$$
\mathcal{P}_{T+1}|\mathcal{P}_T, \alpha_T \sim N\left(\mu_{\mathcal{P}}^{\mathbb{P}} + \Phi_{\mathcal{P}}^{\mathbb{P}} \mathcal{P}_T + \Phi_Z^{\mathbb{P}} \alpha_T,\ \Sigma_Z \Sigma_Z' + \Sigma_{\mathcal{P}} \Sigma_{\mathcal{P}}' \right)
$$

where α_T is distributed as $N(a_{T|T},\ P_{T|T})$, with $a_{T|T}$ and $P_{T|T}$ obtained from the Kalman filter.

2.2.3 Likelihood and Risk Price Restrictions

Statistical inference can proceed using the observations $Y = \{y_t, \mathcal{P}_t : t = 0, 1, \ldots, T\}$. The likelihood factorizes into two parts stemming from the $\mathbb{P}$ and $\mathbb{Q}$ respectively. For R observable factors, and not N as in the standard case in Section 1.2.1, the

joint likelihood (conditional on the initial point $\mathcal{P}_0$) can be written as

$$f(Y|\theta,\widehat{\Sigma}_Z) = \left\{\prod_{t=0}^{T} f^{\mathbb{Q}}(y_t|\mathcal{P}_t, k_\infty^{\mathbb{Q}}, g^{\mathbb{Q}}, \Sigma_{\mathcal{P}}, \sigma_e^2)\right\} \times \left\{\prod_{t=1}^{T} f^{\mathbb{P}}(\mathcal{P}_t|\mathcal{P}_{t-1}, k_\infty^{\mathbb{Q}}, g^{\mathbb{Q}}, \lambda_{0\mathcal{P}}, \lambda_{1\mathcal{P}}, \Sigma_{\mathcal{P}}, \Phi_Z^{\mathbb{P}}, \widehat{\Sigma}_Z)\right\} \tag{2.11}$$

where the $\mathbb{Q}$-likelihood components $f^{\mathbb{Q}}(\cdot)$ are given by (1.11) and capture the cross-sectional dynamics of the risk factors and the yields, whereas $\mathbb{P}$-likelihood components $f^{\mathbb{P}}(\cdot)$, where the Z_ts have been marginalised out by the Kalman filter, are obtained from (2.10) and capture the time-series dynamics of the observed risk factors. The parameter vector is set to $\theta = (\sigma_e^2, k_\infty^{\mathbb{Q}}, g^{\mathbb{Q}}, \lambda_{0\mathcal{P}}, \lambda_{1\mathcal{P}}, \Sigma_{\mathcal{P}}, \Phi_Z^{\mathbb{P}})$. To assure identification, we tune Σ_Z in-sample and fix it out-of-sample at $\widehat{\Sigma}_Z$. Details of the tuning procedure are in Appendix 2.D.

For brevity of further exposition, we let $\lambda^{\mathcal{P}} = [\lambda_{0\mathcal{P}}, \lambda_{1\mathcal{P}}]$ and $\lambda = \lambda_{1\mathcal{P}}$. If all the entries in $\lambda^{\mathcal{P}}$ are free parameters we get the maximally flexible model (model M_0 in this chapter). Alternative models, where some of these entries are set to zero, have also been proposed in the existing literature. Early studies focused mainly on imposing zero ad-hoc restrictions on the parameters governing the dynamics of the risk premia. However, the process of imposing such restrictions in an ad-hoc way has been recently criticised and a few studies have investigated more systematic approaches on how to impose restrictions on the market price of risk parameters. For details, see the associated discussion with related references in Section 1.1.1. Generally, in most models the set of unrestricted parameters is usually a subset of $\lambda^{\mathcal{P}}$.

In this chapter we adopt restriction set with optimal predictive performance, understood predominantly as economic value and less as statistical predictability, among all possible restriction sets, as evidenced in Section 1.5.3, based on out-of-sample period of data closely overlapping this here, see Section 2.4. Therein we develop a sequential version of stochastic search variable selection scheme and arrive at a conclusion that restriction set offering such an optimal predictive performance

based on out-of-sample period from 2007 to 2016 is to only leave $\lambda_{1,2}$ unrestricted (model M_1 there and in this chapter). The same model has been recommended earlier by Bauer (2018), however, in a MCMC framework. This choice is also to some extent in accordance with the early work on risk dynamics, such as Duffee (2002), which finds that variation in the price of level risk is necessary to capture the failure of the Expectations Hypothesis (EH). This variation was usually captured by linking the price of level risk to the slope of the term structure (Duffee, 2011).

Consequently, we can conveniently restate the likelihood specification of (2.11) as

$$\begin{aligned} f(Y|\theta, \widehat{\Sigma}_Z) = &\left\{\prod_{t=0}^{T} f^{\mathbb{Q}}(y_t|\mathcal{P}_t, k_\infty^{\mathbb{Q}}, g^{\mathbb{Q}}, \Sigma_{\mathcal{P}}, \sigma_e^2)\right\} \times \\ &\left\{\prod_{t=1}^{T} f^{\mathbb{P}}(\mathcal{P}_t|\mathcal{P}_{t-1}, k_\infty^{\mathbb{Q}}, g^{\mathbb{Q}}, \Sigma_{\mathcal{P}}, \Phi_Z^{\mathbb{P}}, \lambda_{1,2}, \widehat{\Sigma}_Z)\right\} \end{aligned} \tag{2.12}$$

where $\theta = (\sigma_e^2, k_\infty^{\mathbb{Q}}, g^{\mathbb{Q}}, \Sigma_{\mathcal{P}}, \Phi_Z^{\mathbb{P}}, \lambda_{1,2})$ is revised accordingly.

2.3 Sequential Estimation, Filtering, and Forecasting

In this section we develop a sequential Monte Carlo (SMC) framework for Gaussian DTSMs with unspanned latent factors. We draw from the work of Chopin (2002, 2004) (see also Del Moral et al. (2006)) and make the necessary adaptations to tailor the methodology to the data and models considered in this chapter. Furthermore, we extend the framework to allow for sequential Bayesian treatment of filtering unspanned latent information from the yield curve. We achieve this via the associated Kalman filter. Overall, the developed framework allows the efficient performance of tasks such as sequential parameter estimation and forecasting. We begin by providing the main skeleton of the scheme and then elaborate the details of its specific parts, such as the MCMC algorithm including the Kalman filter, and the framework for obtaining and evaluating the economic benefits of predictions.

2.3.1 Sequential Framework with Latent Processes

Let $Y_{0:t} = (Y_0, Y_1 \ldots, Y_t)$ denote all the data available up to time t, such that $Y_{0:T} = Y$. Similarly, the likelihood based on data up to time t is $f(Y_{0:t}|\theta, \widehat{\Sigma}_Z)$ and is defined in (2.11). Combined with a prior on the parameters $\pi(\theta)$, see Appendix 2.A for details, it yields the corresponding posterior

$$\pi(\theta|Y_{0:t}, \widehat{\Sigma}_Z) = \frac{1}{m(Y_{0:t}|\widehat{\Sigma}_Z)} f(Y_{0:t}|\theta, \widehat{\Sigma}_Z)\pi(\theta) \tag{2.13}$$

where $m(Y_{0:t}|\widehat{\Sigma}_Z)$ is the model evidence based on data up to time t. Moreover, the posterior predictive distribution, which is the main tool for Bayesian forecasting, is defined as

$$f(Y_{t+h}|Y_{0:t}, \widehat{\Sigma}_Z) = \int f(Y_{t+h}|Y_t, \alpha_t, \theta, \widehat{\Sigma}_Z)\pi(\theta|Y_{0:t}, \widehat{\Sigma}_Z)d\theta \tag{2.14}$$

where h is the prediction horizon and $\alpha_t \sim N(a_{t|t},\ P_{t|t})$ is obtained from the Kalman filter.

Note that the predictive distribution in (2.14) incorporates parameter uncertainty by integrating θ out according to the posterior in (2.13). Usually, prediction is carried out by expectations with respect to (2.14), e.g. $E(Y_{t+h}|Y_{0:t}, \widehat{\Sigma}_Z)$ but, since (2.14) is typically not available in closed form, Monte Carlo can be used in the presence of samples from $\pi(\theta|Y_{0:t}, \widehat{\Sigma}_Z)$. This process may accommodate various forecasting tasks but the procedures can be quite laborious and in some cases impracticable, see Section 1.3.1 for further details.

An alternative approach that can also handle forecasting assessment tasks is to use sequential Monte Carlo (see, Chopin (2002) and Del Moral et al. (2006)) to sample from the sequence of distributions $\pi(\theta|Y_{0:t}, \widehat{\Sigma}_Z)$ for $t = 0, 1, \ldots, T$. A general description of the Iterated Batch Importance Sampling (IBIS) scheme of Chopin (2002), see also Del Moral et al. (2006) for a more general framework, is provided in

Algorithm 2.1. The degeneracy criterion is usually taken to be the Effective Sample

Algorithm 2.1 IBIS algorithm for Gaussian Affine Term Structure Models with unspanned latent factors

Initialize N_θ particles by drawing independently $[\theta_i, \alpha_0^{(i)}] \sim [\pi(\theta), N(a_{0|0}, P_{0|0})]$ with importance weights $\omega_i = 1$, $i = 1, \ldots, N_\theta$. For $t, \ldots, T$ and each time for all i:

(a) Calculate the incremental weights from

$$u_t([\theta_i, \alpha_0^{(i)}]) = f(Y_t | Y_{0:t-1}, [\theta_i, \alpha_0^{(i)}], \widehat{\Sigma}_Z)$$

and practically from

$$u_t([\theta_i, \alpha_{t-1}^{(i)}]) = f(Y_t | Y_{t-1}, [\theta_i, \alpha_{t-1}^{(i)}], \widehat{\Sigma}_Z)$$

where instead of conditioning on $\alpha_0^{(i)}$ we condition on $\alpha_{t-1}^{(i)}$, which is distributed as $N\left(a_{t-1|t-1}^{(i)}, P_{t-1|t-1}^{(i)}\right)$, an thus we reduce computational cost of the Kalman filter that is then iterated only once for each particle and we additionally update $\alpha_{t-1}^{(i)}$ to $\alpha_t^{(i)} \sim N\left(a_{t|t}^{(i)}, P_{t|t}^{(i)}\right)$ from the Kalman filter.

(b) Update the importance weights ω_i to $\omega_i u_t([\theta_i, \alpha_{t-1}^{(i)}])$.

(c) If some degeneracy criterion (e.g. ESS(ω)) is triggered, perform the following two sub-steps:

(i) Resampling: Sample with replacement N_θ times from the set of θ_is according to their weights ω_i. The weights are then reset to one.

(ii) Jittering: Replace θ_is with $\tilde{\theta}_i$s by running MCMC chains, and thus all iterations of the Kalman filter, with each θ_i as input and $\tilde{\theta}_i$ as output. Set $[\theta_i, \alpha_t^{(i)}] = [\tilde{\theta}_i, \tilde{\alpha}_t^{(i)}]$.

Size (ESS) which is defined in (1.21).

The IBIS algorithm provides a set of weighted θ samples, or else particles, that can be used to compute expectations with respect to the posterior, $E[g(\theta)|Y_{0:t}, \widehat{\Sigma}_Z]$, for all t using the estimator $\sum_i [\omega_i g(\theta_i)] / \sum_i \omega_i$. The same holds for expectations with respect to the posterior predictive distribution, $f(Y_{t+h}|Y_t, \alpha_t, \widehat{\Sigma}_Z)$; the weighted θ samples can be transformed into weighted samples from $f(Y_{t+h}|Y_t, \alpha_t, \widehat{\Sigma}_Z)$ by simply applying $f(Y_{t+h}|Y_t, \alpha_t, \theta, \widehat{\Sigma}_Z)$. A very useful by-product of the IBIS algorithm is the ability to compute

$$m(Y_{0:t}|\widehat{\Sigma}_Z) = f(Y_{0:t}|\widehat{\Sigma}_Z)$$

which is the criterion for conducting formal Bayesian model choice. Computing the following quantity in step (a) in Algorithm 2.1 yields a consistent and asymptotically normal estimator of $f(Y_t|Y_{0:t-1}, \widehat{\Sigma}_Z)$, namely

$$m_t = \frac{1}{\sum_{i=1}^{N_\theta} \omega_i} \sum_{i=1}^{N_\theta} \omega_i u_t([\theta_i, \alpha_{t-1}^{(i)}])$$

where $\alpha_{t-1}^{(i)} \sim N\left(a_{t-1|t-1}^{(i)}, P_{t-1|t-1}^{(i)}\right)$ is obtained from the Kalman filter for each $i = 1, \ldots, N_\theta$. Further qualities behind IBIS can be found in Section 1.3.1.

To apply IBIS output to models and data in this chapter, the following adaptations and extensions are made. In a manner similar to Section 1.3.1, and for reasons mentioned therein, we combine the benefits of data tempering and adaptive tempering (Jasra et al., 2011; Schäfer and Chopin, 2013; Kantas et al., 2014) in a hybrid adaptive tempering scheme which we present in Appendix 2.C. Since the MCMC algorithm used here is an extension of Bauer (2018) and thus it consists of independence samplers that are known to be unstable, we utilize the IBIS output and estimate posterior moments to obtain independence sampler proposals; see Appendix 2.B for details, and Section 1.3.1 for further motivation. Within such adapted framework, we extend the methodology presented in Sections 2.2.1 and 2.2.3 to handle Kalman filtering sequentially. Eventually, we use IBIS output in the construction of a model-driven dynamically rebalanced portfolio of bond excess returns and measure its economic performance, like in Section 1.3.3.2.

In applied work, we resort to $N_\theta = 2000$ particles and 5 MCMC steps when jittering, while for minimum ESS we set $\alpha = 0.7$. We choose 5 steps at the jittering stage because the mixing behaviour of the underlying MCMC is satisfactory enough. We tracked the correlation between particles before and after that stage to find that performance was already acceptable with this number of iterations.

2.3.2 Assessing Predictive Performance and Economic Value

Based on the evaluation framework outlined in this section, we then seek to understand whether unspanned latent information extracted from yields-only Gaussian ATSMs, using Kalman filter, lets better predict excess returns, compared to corresponding yields-only models. Furthermore, we attempt to explore whether such statistical predictability, if any, can be turned into consistent economic benefits for bond investors. Finally, we explore whether making full use (all factors) of such unspanned latent factors benefits them most, or to the contrary (selected factors).

2.3.2.1 Bond Excess Returns

Following from Section 1.3.3.1, we define the observed continuously compounded excess return of an n-year bond as the difference between the holding period return of the n-year bond and the h-period yield as

$$rx_{t,t+h}^{n} = -(n-h)y_{t+h}^{n-h} + ny_{t}^{n} - hy_{t}^{h}$$

If, instead of taking the observed one, we take the model-implied continuously compounded yield y_t^n, calculated according to (1.8), we arrive at the predicted excess return $\widetilde{rx}_{t,t+h}^{n}$ which becomes

$$\widetilde{rx}_{t,t+h}^{n} = A_{n-h,\mathcal{P}} - A_{n,\mathcal{P}} + A_{h,\mathcal{P}} + B_{n-h,\mathcal{P}}'\widetilde{\mathcal{P}}_{t+h} - (B_{n,\mathcal{P}} - B_{h,\mathcal{P}})'\mathcal{P}_t \tag{2.15}$$

where $\mathcal{P}_t$ is observed and $\widetilde{\mathcal{P}}_{t+h}$ is a prediction from the model. Our developed framework, see Algorithm 2.1, allows drawing from the predictive distribution of $(\widetilde{\mathcal{P}}_{t+h}, \widetilde{rx}_{t,t+h}^{n})$ based on all information available up to time t. More specifically, for each θ_i particle the $\mathbb{P}$-dynamics of $\mathcal{P}_t$ can be used to obtain a particle of $\widetilde{\mathcal{P}}_{t+h}$, which then can be transformed into a particle of $\widetilde{rx}_{t,t+h}^{n}$ via equation (2.15). Detailed steps for the case of $h = 1$ are outlined in Algorithm 2.2.

Algorithm 2.2 Predictive distribution of excess returns for Gaussian Affine Term Structure Models with unspanned latent factors

First, at time t, for some n and $h = 1$, using (ω_i, θ_i), $i = 1, ..., N_\theta$, from IBIS algorithm, iterate over i:

(a) Given θ_i, compute $A_{i_1,\mathcal{P}}$ and $B_{i_1,\mathcal{P}}$, for $i_1 \in \{1, n-1, n\}$, from (1.9) and (1.10).

(b) Given θ_i, obtain prediction of $\mathcal{P}_{t+1}$ by drawing, first from

$$\widetilde{\alpha}_{t+1}^{(i)} | \alpha_t^{(i)} \sim N\left(\Phi_Z^{\mathbb{P}} \alpha_t^{(i)},\ \widehat{\Sigma}_Z \widehat{\Sigma}_Z'\right)$$

and then from

$$\widetilde{\mathcal{P}}_{t+1}^{(i)} | \mathcal{P}_t, \widetilde{\alpha}_{t+1}^{(i)} \sim N\left(\mu_{\mathcal{P}}^{\mathbb{P}} + \Phi_{\mathcal{P}}^{\mathbb{P}} \mathcal{P}_t + \widetilde{\alpha}_{t+1}^{(i)},\ \Sigma_{\mathcal{P}} \Sigma_{\mathcal{P}}'\right)$$

where distribution of $\alpha_t^{(i)}$ is obtained from the Kalman filter as $N\left(a_{t|t}^{(i)},\ P_{t|t}^{(i)}\right)$.

(c) Compute particle prediction of $rx_{t,t+1}^n$ as

$$\widetilde{rx}_{t,t+1}^{n(i)} = A_{n-1,\mathcal{P}} - A_{n,\mathcal{P}} + A_{1,\mathcal{P}} + B_{n-1,\mathcal{P}}' \widetilde{\mathcal{P}}_{t+1}^{(i)} - (B_{n,\mathcal{P}} - B_{1,\mathcal{P}})' \mathcal{P}_t$$

Second, since $(\omega_i, \widetilde{\mathcal{P}}_{t+1}^{(i)}, \widetilde{rx}_{t,t+1}^{n(i)})$, $i = 1, ..., N_\theta$, is a particle approximation to predictive distribution of $(\mathcal{P}_{t+1}, rx_{t,t+1}^n)$, compute point prediction of $rx_{t,t+1}^n$ using particle weights ω_i as

$$\widetilde{rx}_{t,t+1}^n = \frac{1}{\sum_{i=1}^{N_\theta} \omega_i} \sum_{i=1}^{N_\theta} \omega_i \widetilde{rx}_{t,t+1}^{n(i)}$$

Third, repeat above two steps for different n and h. For $h > 1$, use $\widetilde{\mathcal{P}}_{t+h-1}^i$ and $\widetilde{\alpha}_{t+h-1}^{(i)}$ in place of $\mathcal{P}_t$ and $\alpha_t^{(i)}$, and $i_h \in \{h, n-h, n\}$ in place of i_1.

The predictive accuracy of bond excess return forecasts is measured in relation to an empirical benchmark. We follow related literature and first adopt the EH as this benchmark, which essentially uses historical averages as the optimal forecasts of bond excess returns. This empirical average is

$$\overline{rx}_{t+h}^n = \frac{1}{t-h} \sum_{j=1}^{t-h} rx_{j,j+h}^n$$

Then we also measure the predictive accuracy of bond excess return forecasts in relation to model M_1 (see, Chapter 1).

To assess the predictive ability of the models considered, we follow in part the methodology outlined in Section 1.3.3.1 and compute the associated out-of-sample R^2 (R^2_{os}), due to Campbell and Thompson (2008).

2.3.2.2 Economic Performance of Excess Return Forecasts

From a bond investor's point of view it is of paramount importance to establish whether the predictive ability of a model can generate economically significant portfolio benefits, out-of-sample. The portfolio performance may also serve as a metric to compare models that make more or less use of unspanned latent information, that is include more or fewer unspanned latent factors.

To that end, we consider a Bayesian investor with power utility preferences

$$U(W_{t+h}) = U(w_t^n, rx_{t+h}^n) = \frac{W_{t+h}^{1-\gamma}}{1-\gamma}$$

where W_{t+h} is an h-period portfolio value, γ is the coefficient of relative risk aversion and w_t^n is a portfolio weight on the risky n-period bond. If we let $(1 - w_t^n)$ be a portfolio weight of the riskless h-period bond, then the portfolio value h periods ahead is given as

$$W_{t+h} = (1 - w_t^n)\exp(r_t^f) + w_t^n \exp(r_t^f + rx_{t,t+h}^n)$$

where r_t^f is the risk-free rate. At every time t, our Bayesian learner solves an asset allocation problem getting optimal portfolio weights. For more details, see Section 1.3.3.2.

To obtain the economic value generated by each model, we use the resulting optimum weights to compute the CER as in Johannes et al. (2014) and Gargano et al. (2019). In this respect, we closely follow the exposition in Section 1.3.3.2 and we compare portfolio performance of each model we develop in this chapter against the EH benchmark and model M_1 (see, Chapter 1).

2.4 Data and Models

In this section we discuss data we use throughout in detail. Specifically, we elaborate on the US Treasury yields data we employ in this chapter, and how we split them into a training and a testing subsample. Furthermore, we describe the macroeconomic variables we use to identify the nature of the hidden components extracted from the latent factors we obtain. We then also explain what models we consider, as distinguished by different positions unspanned latent factors take in the given model.

2.4.1 Yields and Macros

The bond data set we use is an updated version of this described in Section 1.4.1 and contains monthly observations of zero-coupon US Treasury yields with maturities of 1-year, 2-year, 3-year, 4-year, 5-year, 7-year, and 10-year, spanning the period from January 1985 to the end of 2018. We consider two sub-samples, a training one ending at the end of 2007 which, as such, precludes the recent financial crisis, and a testing one which includes the period after the end of 2007, a period determined by, first, different monetary actions and establishment of unconventional policies and, second, interest rates hitting the zero-lower bound. The post-2007 global financial crisis period is excluded from our training sub-sample. Doing so we let the models learn about it merely online and we expect the use of unspanned latent information to facilitate that. As such, we also verify the concerns about the capability of Gaussian ATSMs to deal with the zero-lower bound (see, Kim and Singleton (2012) and Bauer and Rudebusch (2016)). These concerns are explicitly explored in the testing sub-sample, spanning the period from January 2008 to the end of 2018. Data processing, which involves extracting the first three principal components from the yield curve, follows from Section 1.4.1.

In terms of macroeconomic variables for the US economy which we use here specifically to identify the nature of the hidden part in the risk premium factor, as

described in the next section, we consider several indices, some of which are well covered in the literature. These include, core[4] inflation (CPI) as in Cieslak and Povala (2015), three-month moving average of the Chicago Fed National Activity Index (GRO) from Joslin et al. (2014), which is a measure of current economic conditions, as well as three variables from Ludvigson and Ng (2009). Namely, real activity ($F1$), its cube ($F1^3$), and stock market ($F8$) factors. Moreover, we complement these with unemployment rate (UNR) and manufacturing capacity utilization (MNF) we take from St. Louis FRED (https://fred.stlouisfed.org). All these macroeconomic variables are at a monthly frequency, they are seasonally adjusted, revised, and except for GRO and the factors from Ludvigson and Ng (2009), they are in percent changes from year ago. Underlying data period is such that it corresponds to this for yields, as discussed above.

2.4.2 Models and Rationale Behind

In terms of models, we differentiate them by positions the unspanned latent factors take in the given model. It proceeds according to a general scheme which is following

$$LF_{ijk} \quad i, j, k \in \{0, 1\} \tag{2.16}$$

where, for example for $i = 0$ and $j = k = 1$, LF_{011} means that we allow for unspanned latent information in the second and third equation in (2.3), under the assumption that $R = 3$. We only investigate a subset of the available alternative models in greater detail.

As result of risk price restrictions (on $\lambda_{\mathcal{P}}$) we adopt in this chapter from Section 1.4.2, without loss of generality for $i = j = k = 1$ in (2.16), we define the stochastic

[4]In contrast to Cieslak and Povala (2015), who devise trend inflation by appropriately smoothing core inflation, we use the latter as is. Such choice is better suited for comparisons with unspanned latent factors, which are *a-priori* independent, first order autoregressive processes, not necessarily smooth.

risk premium factor as

$$RP_t^Z = \lambda \mathcal{P}_t + Z_t = \begin{bmatrix} \lambda_{1,2}\mathcal{P}_{2,t} + Z_{1,t} \\ Z_{2,t} \\ Z_{3,t} \end{bmatrix} \tag{2.17}$$

where the first element is similar to Duffee (2011), yet in our case its time-varying component, $RP_t = \lambda_{1,2}\mathcal{P}_{2,t}$, is a restricted version (only $\lambda_{1,2}$ is free) of the risk premium factor defined in Duffee's paper. Its stochastic component $Z_{1,t}$ is in that particular case the first element of the unspanned latent variable Z_t. If instead $i = 0$ and $j = k = 1$, then

$$RP_t^Z = \begin{bmatrix} \lambda_{1,2}\mathcal{P}_{2,t} \\ Z_{2,t} \\ Z_{3,t} \end{bmatrix}$$

where the first element, which is related to level risk and is not stochastic but time-varying only, equals RP_t and is also equivalent to the corresponding factor in model M_1 (see, Chapter 1) we adopt risk price restrictions from.

According to Duffee (2011), time-varying risk premium factor RP_t, which in Duffee's case is a linear combination of the state vector obtained from model shocks to yields, determines the compensation investors require to face fixed-income risk from t to $t+1$, and it contains all information relevant to predicting one-step-ahead, yet not h-step-ahead for $h > 1$, excess returns. We instead assume that investors require compensation for level risk stemming from changes to slope only. However, stochastic term in (2.17), which is *a priori* unspanned and latent, affects this compensation. It is important to understand whether this impact is not distorting information relevant to predicting excess returns, already available from RP_t alone. One way forward towards achieving this goal, is to conduct predictability and economic value exercises for the models considered and make appropriate comparisons, as done in Section 2.5.

Different from Duffee (2011), we obtain the state vector from observed yields directly. Nevertheless, for $i, j, k \in \{0, 1\}$ in (2.16), we are able to similarly define the hidden part of the stochastic risk premium factor, as the part unspanned by $\mathcal{P}_t$, in the following way

$$\begin{aligned} \widetilde{RP}_t^Z &= RP_t^Z - E\left[RP_t^Z|\mathcal{P}_t\right] \\ &= Z_t - E[Z_t|\mathcal{P}_t] \end{aligned} \tag{2.18}$$

where $E[Z_t|\mathcal{P}_t]$ is the projection of unspanned latent factors on principal components obtained from observed yields which is thus spanned and consequently not hidden part of this factor. After Duffee (2011), $\widetilde{RP}_t^Z$ can be estimated as residual from a regression of the former on the latter

$$\widetilde{RP}_t^Z \equiv Z_t - a - b'\mathcal{P}_t$$

where we take Z_t as the mean from its posterior distribution and a and b are the underlying Ordinary Least Squares parameter estimates.

We can then examine *a posteriori* to what extent Z_t assumed unspanned and latent *a priori* is actually hidden from the yield curve, and whether there is any link between such hidden information and the predictability and economic value results we obtain for the models we consider. Furthermore, by means of regression analyses we can inspect whether these unspanned latent factors we obtain and their hidden, as well as not hidden, parts are spanned by information contained in macroeconomic variables we investigate in this chapter.

2.5 Empirical Results

The focus of this section is to present the main results on the statistical and economic performance of excess return forecasts resulting from the models we develop in this

chapter. These, despite unspanned latent factors, are in fact still yields-only when it comes to data involved in estimation. Specifically, we assess the models based on different number of unspanned latent factors, which we filter out from the yield curve, and more importantly their position in the underlying $\mathbb{P}$-dynamics.

Initially, we inspect the unobserved factors we estimate, in particular how their dynamics unfolds under different economic conditions. Next, we briefly explore the explanatory power behind these latent factors. Only then, we investigate whether puzzling behavior between statistical predictability and out-of-sample economic gains for bond investors, as it is often evidenced in case of yields-only models, can still be observed when unspanned latent information is incorporated in the model.

Eventually, we decompose risk premium factor into a part which is hidden from the yield curve and a part which is spanned by the underlying principal components extracted from yields. Then, by regressing these on macroeconomic variables broadly considered in the related literature, we uncover a link between such macro spanning and performance results we obtain for the models with unspanned latent factors.

2.5.1 Observing the Unobserved

The data contain an important economic event and therefore it is interesting to examine the trajectories of unobserved factors, especially during the testing period. In particular, data cover the financial crisis of 2008-2009. This episode started from a flat yield curve during the pre-recession period, following an inversion of the curve that reflects an increase in short rates due to expectations about the Fed tightening its policy, and then a steepening as a reaction to policy adjustments by the Fed reflecting strong growth and inflation expectations. For related details, yet focused on principal components rather than unobserved factors, see Section 1.5.1.

Recession of 2008-2009 is differently picked up by each of the unspanned latent factors, see Figure 2.1. Economic uncertainty arising immediately in 2008 and lasting until about 2012, a year approximately marking the end of the crisis, is well reflected

by a decrease, from positive to negative, in the posterior mean of the latent factor $Z_{2,t}$ (models LF_{111} and LF_{011}), affecting the slope of the yield curve, following a rebound to zero thereafter. This picture is quite different when we look at the posterior mean of the unobserved factor $Z_{1,t}$ (model LF_{111}), which has an impact on the level of the yield curve. In this case, we only observe a steep decrease from zero in the beginning of 2008, followed by a W-shaped recovery lasting until early 2009. A look at the unobserved dynamics behind the curvature of the yield curve, namely the posterior mean of $Z_{3,t}$ (models LF_{111}, LF_{011} and LF_{001}), reveals increased oscillations around zero starting in 2008 and lasting approximately until 2012, which are definite shocks, though nontrivial to interpret.

2.5.2 Explanatory Power of Unspanned Latent Factors

Large explanatory power of the unspanned latent factor $Z_{2,t}$ can be revealed from results in Table 2.1. The table presents adjusted R^2 for the projection of the three unspanned latent factors on the three principal components. Quite interestingly, regardless the modelling specification, a very large fraction (ranging between 0.45 and 0.54) of the variation of $Z_{2,t}$ is explained by the information spanned by the yield curve. Nevertheless, an important question that arises is whether there is also information hidden from the yield curve and, as such, cannot be captured by the three PCs, while it is useful for prediction purposes. Furthermore, of particular importance is whether this information, if any, is linked to macroeconomic forces. We seek to answer the aforementioned questions in what follows below.

Table 2.2 reports adjusted R^2 for model LF_{001}, which features a latent factor $Z_{3,t}$ in addition to the three yield factors. When adding an unspanned latent factor on the third PC, the explanatory power across the maturity spectrum remains, in general, intact. There are a few small, albeit statistically significant, increases mainly at longer maturities and short prediction horizons.

The situation is reversed when adding $Z_{2,t}$ to the information set. More specifically, the latent factor substantially improves explanatory power on excess returns, especially at shorter maturities and longer horizons, as shown in Table 2.3. In particular, at 2-year maturity, adjusted R^2 increases from 13.54%, in the yields-only case, at the 6-month horizon, up to 16.15%. At 12-month horizon, the associated R^2 jumps from 17.3% up to 23.58%, a statistically significant increase of 7.59 percentage points. Those increases do not hold at longer maturities, implying that the predictive power of $Z_{2,t}$ is stronger at short maturities only, what reveals that unspanned $Z_{2,t}$ captures information associated with the short end of the yield curve. This stands in line with prior evidence that different maturities do not move on a single factor.

Beside the fact that adding a latent factor on the first PC, on top of the other latent factors in model LF_{111}, has no impact on the explanatory power, as measured by $\bar{R}^2$, results for this model and also for LF_{011} are quantitatively and qualitatively similar, thus not shown. Moving forward, it is of paramount importance to access whether the evident explanatory power of the unspanned factors on excess returns is enough to offer any statistical and economic evidence of return predictability.

2.5.3 Bond Return Predictability and Economic Performance

In this section we seek to investigate whether (statistical) predictability of excess bond returns, as well as economic benefits to bond investors, can be enhanced by integrating information hidden from the yield curve.

Only recently, Fulop et al. (2019) suggest that no matter the maturity, the model or the methodology considered, statistical predictability of excess returns is hardly translated into economic gains for bond investors, when real-time macroeconomic factors replace a fully-revised dataset. This is in direct contrast to the earlier results of Gargano et al. (2019) and, more recently, to those in Bianchi et al. (2021), where fully revised information set is used. Our modelling approach attempts to stay away from the aforementioned debate, as our unspanned component is regarded as a latent

stochastic process, rather than consisting of directly observable (macroeconomic) variables. Nevertheless, risk premia are conditioned on information available to bond investors in real time. As such, our results are directly comparable to those studies that seek to predict excess returns based on real-time data rather than revised datasets.

2.5.3.1 Bond Return Predictability

Table 2.4 reports the out-of-sample R^2 values for all models with unspanned latent factors, across maturities and investment horizons. The in-sample period is 1985-2007 and the out-of-sample analysis spans the period 2008-2018. All models attain good predictive performance compared to the EH benchmark, with positive R^2_{os} across maturities. Results reveal evidence of out-of-sample bond return predictability across model specifications. This is important, taking into account that all models utilise information coming solely from the cross section of yields (either 'yield-only' or 'yield-plus' in Duffee (2011) terminology).

Quite interestingly, for maximally flexible model M_0, R^2_{os} decrease with maturity and increase with prediction horizon. Results, however, are qualitatively opposite for restricted model M_1 and for models with unspanned latent factors, which generate more accurate forecasts and substantially higher R^2_{os} at longer maturities. In particular, at 12-month investment horizon, for model LF_{010} (LF_{011}), which embeds a single unspanned latent factor on the second PC, R^2_{os} increases from 0.15 (0.16) up to 0.33 (0.32) for the 10-year maturity bond.

Of particular importance, however, is whether the unspanned latent factors share information to predict returns that goes beyond the cross section of yields. Table 2.5 displays the out-of-sample R^2 values for all models with unspanned latent factors. Comparisons are in relation to model M_1, which imposes heavy restrictions on the dynamics of risk compensation and has been proved to offer the best predictive performance among all restricted models, as evidenced in Section 1.5.3. Results

provide clear evidence of significant improvement to the out-of-sample predictive performance, suggesting that the unspanned latent factors share information to predict returns that goes beyond the yield curve. Such improvements are particularly pronounced at shorter maturities and longer investment horizons.

Comparing across models, it seems apparent that model LF_{111}, which embeds three latent factors, offers the best performance in terms of statistical predictability. In particular, for model LF_{111}, for the 2-year maturity bond, R^2_{os} is 0.04 at 1-month horizon, moving up to 0.30 at the 12-month horizon. Qualitatively similar (albeit weaker) predictive performance is observed for models with one or two unspanned latent factors (e.g. LF_{010} and LF_{011}). Common to the three model specifications is latent factor $Z_{2,t}$, which seems to add considerable predictive power to the models. The only model that fails to predict well and performs poorly compared to M_1 is model LF_{001}, which generates R^2_{os} mostly close to zero. Such model features a single latent factor on the third PC, suggesting that $Z_{3,t}$ is not capable of capturing information relevant to predict returns. Finally, comparing across bond maturities, predictability decreases with increasing maturity, suggesting that unspanned latent factors carry information, associated with the short end of the maturity spectrum, which is not contained in the yield curve.

2.5.3.2 Economic Performance

Turning to economic performance, we ask whether models with unspanned latent factors offer any economic evidence of return predictability in the US bond market. The coefficient of relative risk aversion is chosen to be $\gamma = 3$ and we do not impose any portfolio constraints[5]. Table 2.6 reports results for the out-of-sample economic value exercise described in Section 2.3.2.2. In particular, it displays annualised CER values, relative to the EH benchmark, for different models, generated using

[5]These non-conservative choices are motivated by early exploratory character of the analysis conducted herein. The goal is to find economic value first and in future research examine if it can be exploited when stricter conditions apply.

forecasts of bond excess returns. As expected, the only model that fails to translate predictability into economic value is the maximally flexible model, M_0, which is not capable of producing any positive out-of-sample economic benefits, mostly generating CERs with negative sign across maturities and investment horizons. The situation is reversed, however, when heavy restrictions are placed on the risk price dynamics, as evidenced by CER values generated by model M_1, in line with Section 1.5.3.

Moving to models with unspanned latent factors, our results tell a consistent story. No matter the model specification considered, corresponding CER values are positive and statistically significant, revealing strong economic performance across maturities and investment horizons. Importantly, models are capable of translating the evident statistical predictability into economic gains for bond investors seeking to maximise their utility. This further suggests that latent factors share information relevant not only to predict returns but also to offer value and generate significant portfolio benefits, out-of-sample. This is quite an important finding, taking into account that investors only have access to yield curve data, and no other (spanned or unspanned) information is utilised, in contrast to studies which seek to detect further information from macroeconomic sources (see, Gargano et al. (2019), Bianchi et al. (2021), among others).

Comparing the economic performance of models across maturities, sheds light on an interesting observation. CERs increase with maturity for the heavily constrained model M_1, suggesting more profitable investments for long maturity bonds. The situation is reversed, however, for models with unspanned factors, which generate CERs that tend to decrease or remain flat as maturity increases, revealing higher gains for short maturity bonds. This further suggests evidence of economic improvement offered by latent factors at shorter maturities.

Next, we move to investigate deeper, whether models with unspanned latent factors offer any economic benefits to bond investors, when compared to model M_1, which has been shown to generate the largest portfolio benefits among all 'yields-only'

models (see, Chapter 1). We therefore seek to understand whether the improved statistical performance, evidenced in the previous section, is indeed translated into better economic performance. As such, we repeat the economic value exercise, by computing the corresponding CERs, for all models with unspanned latent factors, relative to model M_1. Table 2.7 summarizes the results which offer interesting insights.

In line with the existing literature (see, Gargano et al. (2019), Fulop et al. (2019), etc.), and in order to overcome the inference issues evidenced by Bauer and Hamilton (2018), our analysis is mainly concentrated on the 1-month investment horizon. Those are the only CER values that are generated using forecasts from non-overlapping bond returns and as such, are directly comparable to studies that examine economic evidence of predictability generated through those forecasts (see, Gargano et al. (2019), Bianchi et al. (2021), Fulop et al. (2019)). We find that at 1-month horizon, CER values are positive and statistically significant for all models other than model LF_{111}, which has an inferior performance and generates CERs with negative sign. In particular, for model LF_{001}, CER values range between 1.65% and 2.74%, with the most profitable investment being the 7-year bond, while for model LF_{010}, statistically significant CERs range between 1.60% and 2.23%, with the 5-year maturity bond being the most profitable. Results provide evidence that non-overlapping excess bond returns, not only are predictable, but also offer significant economic benefits to investors, when compared to the M_1 benchmark. This is not entirely surprising, however, given that the latent factors evolve according to an autoregressive process of order one, thus improvements, if any, are expected at such a short investment horizon. It further supports the evident ability of the models to exploit information hidden from the yield curve. The gains are not substantial at longer horizons, however, with CER values that mostly have negative signs or are either close to zero or marginally positive but still statistically insignificant, indicating that the relevant information is short lived, nevertheless still of value to investors. Those findings are

in direct contrast to the conclusions of Fulop et al. (2019) who argue that statistical predictability is not turned into economic benefits and in line with the studies of Gargano et al. (2019) and Bianchi et al. (2021), which find support of economically significant predictability in bond returns. The latter, however, exploit information from a fully revised macroeconomic dataset, which has proved to enhance predictive power (see, Ghysels et al. (2018)). Our results are not reliant on macroeconomic data.

2.5.4 Linking Unspanned Latent Factors with Macroeconomy

An intuitively interesting and economically important question that arises is, whether information incorporated in the unspanned latent factors link with macroeconomic forces. In particular, we are interested in investigating further the relationship, if any, between the hidden component of the latent factors, as defined in (2.18) and macroeconomic (rather than pure financial) variables proven to forecast variation in bond excess returns. Table 2.8 presents adjusted R^2 from regressing the hidden component of risk premia on different macroeconomic variables. Looking at model LF_{010} (i.e. second panel), our results reveal that particular macros are capable of explaining substantial amounts of variation in the hidden component of the latent factor $Z_{2,t}$, which is related to slope risk. In particular, GRO explains 0.11 while MNF (UNR) captures up to 0.19 (0.15) of the variation in $\widetilde{RP}^{Z}_{2,t}$. Even more importantly, the hidden component of bond risk premia associated with the second latent factor, seems to be countercyclical, as revealed by the negative sign on the (statistically significant) coefficients of the real activity indicators, GRO and $F1$[6], displayed in Table 2.9. This further implies that shocks to GRO (and $F1$) induce counter-cyclical movements in the risk premium associated with slope risk. On

[6]We follow Duffee (2011) and normalise the real activity factor of Ludvigson and Ng (2009), $F1$ (and consequently its cube $F1^3$), to positively co-vary with GRO.

the other side, the loading on unemployment (UNR) is positive and statistically significant.

Explanatory power is even more pronounced for models LF_{011} and LF_{111}. In the former case, 0.13 and 0.11 of the variation in $\widetilde{RP}^{Z}_{2,t}$ is captured by measures of economic activity (i.e. GRO and $F1$), whereas MNF leads to a large R^2 of 0.21. In the latter model, the fraction of the hidden component of $Z_{2,t}$ that is explained by macros (MNF and UNR) is even higher (with adjusted R^2 being 0.23 and 0.27 respectively). Corresponding results on the cyclicality of the hidden component for this latent factor in these two models are qualitatively similar as evidenced by the negative (statistically significant) coefficients, see Table 2.9.

Overall, we document that the hidden component of bond risk premia associated with the slope factor is counter-cyclical and mostly related to variables that proxy real activity. As such, our results uncover a direct link between macroeconomic activity and excess return predictability in the US bond market. Although, from the theoretical viewpoint, the unspanned latent factors might only correct for model misspecification, what can be attributed to autocorrelated residuals in the yields-only model M_1, and perhaps a second order vector autoregressive model is more appropriate for the $\mathbb{P}$-dynamics, the fact that they are statistically related to several key macroeconomic variables supports the argument that we may interpret them as market environment factors.

2.6 Conclusions

This chapter explores the importance of information hidden from the yield curve and assesses the capability of unspanned risks to offer evidence of statistical and economic benefits to bond investors. We propose a novel class of arbitrage-free unspanned DTSM, that embeds a stochastic price of risk specification. The model is factorised into a 'spanned' component, which exploits information provided by the yield curve,

and an 'unspanned' component, which integrates information hidden from the cross section of yields, yet relevant for prediction purposes. The developed setup is then used to explore how valuable the unspanned information is to a real-time Bayesian investor seeking to forecast future excess bond returns and generate systematic economic gains.

Empirical results provide clear evidence of out-of-sample bond return predictability compared to the EH benchmark. Comparisons relative to the restricted yields-only model M_1, reveal improvement to the out-of-sample predictive performance of models that allow for unspanned latent factors, suggesting that those factors contain significant predictive power above and beyond the yield curve. Such improvements are particularly pronounced at shorter maturities revealing that latent factors carry information associated with the short end of the maturity spectrum.

Most importantly, these models generate CER values which are positive and statistically significant, and as such, they are capable of translating the evident statistical predictability into economic gains for bond investors seeking to maximise their utility. Comparisons relative to yields-only model M_1, help us infer further important conclusions for the 1-month non-overlapping bond excess returns. We find that CER values are positive and statistically significant revealing that non-overlapping returns, not only are predictable, but also offer significant economic gains to investors, out-of-sample. These results further support the ability of the models to exploit information hidden from the yield curve, in particular, information which is short lived. This finding is in contrast to the existing literature (see, Fulop et al. (2019), etc.).

Finally, we examine whether movements in excess returns bear any relationship with the economy. As such, we explore the linkages between the hidden component of the unspanned latent factors and macroeconomic variables. We find that particular macros explain substantial variation of the hidden component, revealing a direct line between macroeconomic activity and excess return predictability in the US

bond market. In particular, the hidden component associated with slope risk is countercyclical and relates to real activity.

Appendix 2.A Specification of Priors

In what follows, we concisely explain the prior distributions that were not mentioned in the main body of the chapter. For parameters in $g^{\mathbb{Q}}$, $k_{\infty}^{\mathbb{Q}}$, σ_e^2 and $\lambda^{\mathcal{P}}$, or effectively $\lambda_{1,2}$, priors are constructed in the same manner as in the related Appendix 1.A. The only exceptions are parameters in Φ_Z and $\Sigma_{\mathcal{P}}$.

In both remaining cases, we first transform their restricted range components so that they have unrestricted range. We also scale these of their elements which typically take very small values. Specifically, we work with *logit* transformation to constrain diagonal elements of Φ_Z in $(-1, 1)$ and consider a Cholesky factorization of $\Sigma_{\mathcal{P}}$ where the diagonal elements are transformed to the real line and off-diagonal elements are scaled by 10^4. Next, independent normal distributions with zero means are assigned to each of their components. Then large variances are assigned to each element of $\Sigma_{\mathcal{P}}$ and for the diagonal elements of Φ_Z respective variances are set to 2.

Appendix 2.B Markov Chain Monte Carlo Scheme

Following from (2.12) and (2.13), and given a prior $\pi(\theta)$ as described in Appendix 2.A, the posterior can be written in a more detailed manner as

$$\begin{aligned}\pi(\theta|Y, \widehat{\Sigma}_Z) = &\left\{\prod_{t=0}^{T} f^{\mathbb{Q}}(y_t|\mathcal{P}_t, k_{\infty}^{\mathbb{Q}}, g^{\mathbb{Q}}, \Sigma_{\mathcal{P}}, \sigma_e^2)\right\} \times \\ &\left\{\prod_{t=1}^{T} f^{\mathbb{P}}(\mathcal{P}_t|\mathcal{P}_{t-1}, k_{\infty}^{\mathbb{Q}}, g^{\mathbb{Q}}, \Sigma_{\mathcal{P}}, \Phi_Z^{\mathbb{P}}, \lambda_{1,2}, \widehat{\Sigma}_Z)\right\} \times \pi(\theta)\end{aligned}$$

The above posterior is not available in closed form. Thus, methods such as MCMC can be used to draw samples from it using Monte Carlo. Nevertheless, the MCMC output is not guaranteed to lead to accurate Monte Carlo calculations due to poor

mixing and convergence properties the corresponding Markov chain may have, what consequently leads to highly autocorrelated samples.

It is therefore recommended to construct a suitable MCMC algorithm that does not exhibit such unfavourable characteristics. For further details regarding its design, see the related Appendix 1.B. Such an MCMC scheme is outlined in Algorithm 2.3.

Algorithm 2.3 MCMC scheme for Gaussian Affine Term Structure Models with unspanned latent factors

Initialize all values of θ. Then at each iteration of the algorithm:

(a) Update σ_e^2 from its full conditional distribution that can be shown to be an Inverse Gamma distribution with parameters $\tilde{\alpha}/2$ and $\tilde{\beta}/2$, such that $\tilde{\alpha}$ is $\alpha + T(J-R)$ and $\tilde{\beta}$ is $\beta + \sum_{t=0}^{T} \|\hat{e}_t\|^2$, where $\alpha = \beta = 0$, since prior is assumed diffuse, $\hat{e}_t$ is a time-t residual from (1.11), and $\|\cdot\|^2$ is Euclidean norm squared.

(b) Update $\Sigma_{\mathcal{P}}$ using an independence sampler based on the MLE and the Hessian obtained before running the MCMC, using multivariate t-distribution with 5 degrees of freedom as proposal distribution.

(c) Update $(k_\infty^{\mathbb{Q}}, g^{\mathbb{Q}})$ in a similar manner to (b).

(d) Update $(\Phi_Z^{\mathbb{P}}, \lambda_{1,2})$ in a similar manner to (b).

Appendix 2.C Adaptive Tempering

Adaptive tempering serves the goal of smoothing peaked likelihoods. For background information, see the related Appendix 1.C. Implementation of the IBIS scheme with hybrid adaptive tempering steps is presented in Algorithm 2.4. It is important to note that, unlike it is shown in Algorithm 2.1 for the general IBIS case, in the specific case we are dealing here with we initialize the particles by drawing from the posterior $\pi(\theta|Y_{0:t-1}, \widehat{\Sigma}_Z)$ instead of the prior $\pi(\theta)$, and $N(a_{0|0}, P_{0|0})$ for the latent part. This is done in-sample based on training data, as detailed in Section 2.4.1.

It is straightforward to implement step 4(b)iv in Algorithm 2.4 for an independence sampler, however adjustments are necessary for a full Gibbs step. This is the case

for σ_e^2 in step (a) in Algorithm 2.3, see Appendix 2.B. Technical details are the same as in the corresponding appendix to Chapter 1 we refer to above.

Algorithm 2.4 IBIS algorithm with hybrid adaptive tempering for Gaussian Affine Term Structure Models with unspanned latent factors

Initialize N_θ particles by drawing independently $[\theta_i, \alpha_{t-1}^{(i)}] \sim \pi(\theta|Y_{0:t-1}, \widehat{\Sigma}_Z)$ with importance weights $\omega_i = 1$, $i = 1, \ldots, N_\theta$. For $t, \ldots, T$ and each time for all i:

1 *Set $\omega_i' = \omega_i$.*

2 *Calculate the incremental weights from*

$$u_t([\theta_i, \alpha_{t-1}^{(i)}]) = f(Y_t|Y_{t-1}, \alpha_{t-1}^{(i)}, \theta_i, \widehat{\Sigma}_Z)$$

where $\alpha_{t-1}^{(i)} \sim N\left(a_{t-1|t-1}^{(i)},\ P_{t-1|t-1}^{(i)}\right)$, and update $\alpha_{t-1}^{(i)}$ to $\alpha_t^{(i)} \sim N\left(a_{t|t}^{(i)},\ P_{t|t}^{(i)}\right)$ from the Kalman filter.

3 *Update the importance weights ω_i to $\omega_i u_t([\theta_i, \alpha_{t-1}^{(i)}])$.*

4 *If degeneracy criterion ESS(ω) is triggered, perform the following sub-steps:*

(a) *Set $\phi = 0$ and $\phi' = 0$.*

(b) *While $\phi < 1$*

i. *If degeneracy criterion ESS(ω'') is not triggered, where $\omega_i'' = \omega_i'[u_t([\theta_i, \alpha_{t-1}^{(i)}])]^{1-\phi'}$, set $\phi = 1$, otherwise find $\phi \in [\phi', 1]$ such that ESS(ω''') is greater than or equal to the trigger, where $\omega_i''' = \omega_i'[u_t([\theta_i, \alpha_{t-1}^{(i)}])]^{\phi-\phi'}$, for example using bisection method, see Kantas et al. (2014).*

ii. *Update the importance weights ω_i to $\omega_i'[u_t([\theta_i, \alpha_{t-1}^{(i)}])]^{\phi-\phi'}$.*

iii. *Resample: Sample with replacement N_θ times from the set of θ_is according to their weights ω_i. The weights are then reset to one.*

iv. *Jitter: Replace θ_is with $\tilde{\theta}_i$s by running MCMC chains with each θ_i as input and $\tilde{\theta}_i$ as output, using likelihood given by $f(Y_{0:t-1}|\theta_i, \widehat{\Sigma}_Z)[f(Y_t|\alpha_t^{(i)}, \theta_i, \widehat{\Sigma}_Z)]^\phi$. Set $[\theta_i, \alpha_{t-1}^{(i)}, \alpha_t^{(i)}] = [\tilde{\theta}_i, \tilde{\alpha}_{t-1}^{(i)}, \tilde{\alpha}_t^{(i)}]$.*

v. *Calculate the incremental weights from*

$$u_t([\theta_i, \alpha_{t-1}^{(i)}]) = f(Y_t|Y_{t-1}, \alpha_{t-1}^{(i)}, \theta_i, \widehat{\Sigma}_Z)$$

where $\alpha_{t-1}^{(i)} \sim N\left(a_{t-1|t-1}^{(i)},\ P_{t-1|t-1}^{(i)}\right)$.

vi. *Set $\omega_i' = \omega_i$ and $\phi' = \phi$.*

Appendix 2.D Tuning the Latent Process

For identification purposes, we tune Σ_Z in-sample and fix it out-of-sample at $\widehat{\Sigma}_Z$. Details about the underlying data are in Section 2.4.1. To that end, we follow a multi-step process which is entirely based on in-sample data. First, as in the standard case in Section 1.2.1, we estimate by maximum likelihood a yields-only DTSM where $N = 3$ and, out of $\lambda_{\mathcal{P}}$, only $\lambda_{1,2}$ is unrestricted, to match risk price restrictions we adopt in this chapter. Resulting MLEs let us then obtain $\hat{s}_t$, $t = 1, ..., \widetilde{T}$, where $\widetilde{T}$ refers to in-sample period, from (2.5).

Second, we formulate an amended version of the likelihood in (2.12), using only its $\mathbb{P}$-likelihood components $f^{\mathbb{P}}(\cdot)$ modified in the following way

$$\tilde{f}(\widetilde{Y}|\theta, \hat{k}_{\infty}^{\mathbb{Q}}, \hat{g}^{\mathbb{Q}}, \widehat{\Sigma}_{\mathcal{P}}, \hat{\lambda}_{1,2}) = \left\{ \prod_{t=1}^{\widetilde{T}} \tilde{f}^{\mathbb{P}}(\mathcal{P}_t|\mathcal{P}_{t-1}, \hat{k}_{\infty}^{\mathbb{Q}}, \hat{g}^{\mathbb{Q}}, \widehat{\Sigma}_{\mathcal{P}}, \Phi_Z^{\mathbb{P}}, \hat{\lambda}_{1,2}, c) \right\} \tag{2.19}$$

where $\widetilde{Y}$ refers to in-sample data, $\hat{k}_{\infty}^{\mathbb{Q}}$, $\hat{g}^{\mathbb{Q}}$, $\widehat{\Sigma}_{\mathcal{P}}$ and $\hat{\lambda}_{1,2}$ are the MLEs from the first step, and $\theta = (\Phi_Z^{\mathbb{P}}, c)$, with scalar $c > 0$, consists of parameters we estimate by maximum likelihood next. However, before that we parametrize Σ_Z in (2.19) as

$$vec(\Sigma_Z) = c \sqrt{\left(I_{(N-R)^2} - \Phi_Z^{\mathbb{P}} \otimes \Phi_Z^{\mathbb{P}}\right) vec\left\{diagm\left[Var(\hat{s})\right]\right\}} \tag{2.20}$$

where here $diagm\left[Var(\hat{s})\right]$ is a $(N - R) \times (N - R)$ matrix including diagonal elements of the covariance matrix $Var(\hat{s})$ on its diagonal and zeros elsewhere. This for $\hat{s} = \left[\hat{s}_1, \ldots, \hat{s}_{\widetilde{T}}\right]$, where $\hat{s}_t$, $t = 1, \ldots, \widetilde{T}$, are practically as in the first step.

Third, we proceed with the Kalman filter, as in Section 2.2.3, to arrive at the log-likelihood representation of (2.19), similar to (2.10), which we maximize. Consequently, it lets us fix Σ_Z out-of-sample at $\widehat{\Sigma}_Z$, which we calculate as in (2.20) with $(\widehat{\Phi}_Z^{\mathbb{P}}, \hat{c})$ being the MLEs $\hat{\theta}$ of θ from this last step and $Var(\hat{s})$ remains unchanged.

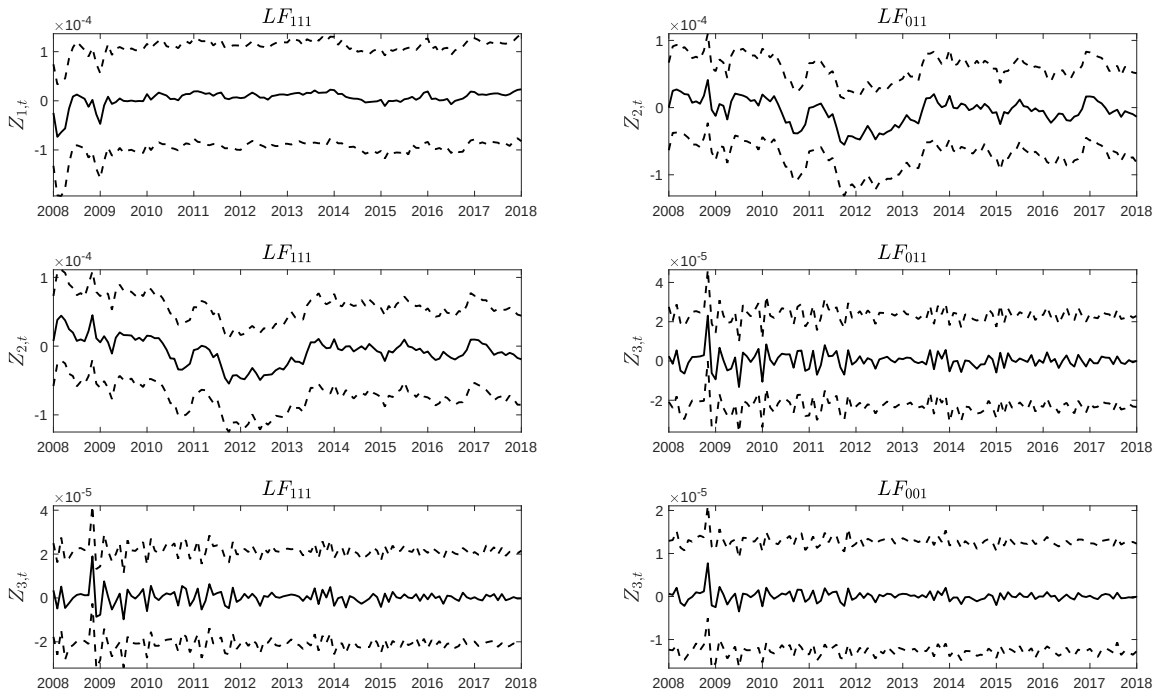

Figure 2.1: Unspanned latent factors filtered from the yield curve. Focus is given on the period between January 2008 and the end of 2017 for which predictions of the model were also evaluated, but the factors are based on all the data from January 1985. Left column presents the factors from model LF_{111}. Right column shows the factors from models LF_{011}, first and second row, and LF_{001}, third row. Throughout, solid lines represent posterior means and dashed lines are 95% credible intervals.

Table 2.1: Explanatory power of principal components when fitting latent factors, measured via $\bar{R}^2$ - period: January 1985 - end of 2017.

$\bar{R}^2 : Z_t = a + b'\mathcal{P}_t + e_t$			
	$Z_{1,t}$	$Z_{2,t}$	$Z_{3,t}$
LF_{001}			0.09
LF_{010}		0.45	
LF_{011}		0.54	0.08
LF_{111}	0.05	0.52	0.06

This table reports in-sample $\bar{R}^2$ across alternative regression specifications. The explained variables are individual latent factors $Z_{j,t}$, $j \in \{1, 2, 3\}$, in Z_t, from models LF_{001}, LF_{010}, LF_{011} and LF_{111}. The explanatory variables are the principal components $\mathcal{P}_t$. The sample period is January 1985 to end of 2017.

Table 2.2: Explanatory power gains from latent factor estimated using model LF_{001}, when fitting excess bond returns, measured via $\bar{R}^2$ over multiple prediction horizons - period: January 1985 - end of 2017.

h\n	**2Y**	**3Y**	**4Y**	**5Y**	**7Y**	**10Y**
–	$\bar{R}^2(\%) : rx^n_{t,t+h} = a + b'\mathcal{P}_t + e_t$					
1	4.26	3.29	3.08	2.74	2.77	3.55
3	9.50	7.93	8.69	7.45	8.60	8.53
6	13.54	12.70	14.10	13.75	14.42	13.70
9	14.97	14.16	15.76	16.59	18.35	19.03
12	17.30	14.77	16.37	17.83	20.47	22.90
LF_{001}	$\Delta_{(3)}\bar{R}^2(\%) : rx^n_{t,t+h} = a + b'\mathcal{P}_t + cZ_{3,t} + e_t$					
1	0.30	0.72*	0.81*	0.99**	1.62**	1.98**
3	0.13*	0.29*	0.42*	0.86**	1.02**	0.77**
6	-0.05	-0.13	-0.21	-0.25	-0.11	-0.04
9	-0.12	-0.19	-0.22	-0.24	-0.24	-0.23
12	-0.17	-0.24	-0.25	-0.25	-0.21	-0.19
LF_{001}	$\bar{R}^2(\%) : rx^n_{t,t+h} = a + b'\mathcal{P}_t + cZ_{3,t} + e_t$					
1	4.55	3.99	3.86	3.71	4.34	5.45
3	9.62	8.19	9.07	8.24	9.53	9.23
6	13.49	12.58	13.91	13.53	14.33	13.67
9	14.87	13.99	15.56	16.39	18.15	18.84
12	17.16	14.57	16.16	17.62	20.30	22.74

This table reports in-sample $\bar{R}^2$ in % across alternative regression specifications and at different prediction horizons of h = 1-month, 3-month, 6-month, 9-month and 12-month. The explained variables are different (by maturities) excess bond returns. The explanatory variables are the principal components $\mathcal{P}_t$ and the estimated (via posterior mean) latent factor $Z_{3,t}$ from model LF_{001}. In-sample $\bar{R}^2$ values are obtained in a similar manner to the out-of-sample R^2 measure of Campbell and Thompson (2008) but using in-sample fit instead of out-of-sample forecasts, and incorporating penalty adjustment. In particular, $\bar{R}^2$ in the top panel (all highly statistically significant hence not denoted) measures explanatory power gains from using principal components on top of the in-sample average to fit excess bond returns, whereas $\bar{R}^2$ in the bottom panel (all highly statistically significant hence not denoted) measures explanatory power gains from using principal components and $Z_{3,t}$ on top of the in-sample average to fit excess bond returns. Further, $\Delta_{(3)}$ next to $\bar{R}^2$ in the mid panel means that the latter measures explanatory power gains from using latent factor $Z_{3,t}$ estimated using model LF_{001} on top of the in-sample average and the principal components to do the same. Positive values of this statistic imply that there is explanatory power gain from adding extra variables. Statistical significance is measured using a one-sided Diebold-Mariano statistic with Clark-West adjustment, based on Newey-West standard errors. * denotes significance at 10%, ** significance at 5% and *** significance at 1% level. The sample period is January 1985 to end of 2017.

Table 2.3: Explanatory power gains from latent factor estimated using model LF_{010}, when fitting excess bond returns, measured via $\bar{R}^2$ over multiple prediction horizons - period: January 1985 - end of 2017.

h\n	**2Y**	**3Y**	**4Y**	**5Y**	**7Y**	**10Y**
–	$\bar{R}^2(\%) : rx^n_{t,t+h} = a + b'\mathcal{P}_t + e_t$					
1	4.26	3.29	3.08	2.74	2.77	3.55
3	9.50	7.93	8.69	7.45	8.60	8.53
6	13.54	12.70	14.10	13.75	14.42	13.70
9	14.97	14.16	15.76	16.59	18.35	19.03
12	17.30	14.77	16.37	17.83	20.47	22.90
LF_{010}	$\Delta_{(2)}\bar{R}^2(\%) : rx^n_{t,t+h} = a + b'\mathcal{P}_t + cZ_{2,t} + e_t$					
1	0.12	0.00	-0.18	-0.21	-0.2	-0.25
3	0.65	0.12	-0.21	-0.25	-0.22	0.04
6	3.02**	1.59*	0.41	-0.02	-0.24	-0.07
9	5.66**	3.45*	1.51	0.54	-0.18	-0.14
12	7.59**	5.90**	3.54*	2.00	0.62	-0.21
LF_{010}	$\bar{R}^2(\%) : rx^n_{t,t+h} = a + b'\mathcal{P}_t + cZ_{2,t} + e_t$					
1	4.38	3.29	2.90	2.54	2.57	3.30
3	10.08	8.04	8.50	7.22	8.40	8.56
6	16.15	14.09	14.45	13.74	14.22	13.64
9	19.78	17.12	17.03	17.03	18.20	18.91
12	23.58	19.80	19.33	19.47	20.96	22.73

This table reports in-sample $\bar{R}^2$ in % across alternative regression specifications and at different prediction horizons of h = 1-month, 3-month, 6-month, 9-month and 12-month. The explained variables are different (by maturities) excess bond returns. The explanatory variables are the principal components $\mathcal{P}_t$ and the estimated (via posterior mean) latent factor $Z_{2,t}$ from model LF_{010}. In-sample $\bar{R}^2$ values are obtained in a similar manner to the out-of-sample R^2 measure of Campbell and Thompson (2008) but using in-sample fit instead of out-of-sample forecasts, and incorporating penalty adjustment. In particular, $\bar{R}^2$ in the top panel (all highly statistically significant hence not denoted) measures statistical performance gains from using principal components on top of the in-sample average to fit excess bond returns, whereas $\bar{R}^2$ in the bottom panel (all highly statistically significant hence not denoted) measures statistical performance gains from using principal components and $Z_{2,t}$ on top of the in-sample average to fit excess bond returns. Further, $\Delta_{(2)}$ next to $\bar{R}^2$ in the mid panel means that the latter measures explanatory power gains from using latent factor $Z_{2,t}$ estimated using model LF_{010} on top of the in-sample average and the principal components to do the same. Positive values of this statistic imply that there is explanatory power gain from adding extra variables. Statistical significance is measured using a one-sided Diebold-Mariano statistic with Clark-West adjustment, based on Newey-West standard errors. * denotes significance at 10%, ** significance at 5% and *** significance at 1% level. The sample period is January 1985 to end of 2017.

Table 2.4: Out-of-sample statistical performance of bond excess return forecasts against the EH, measured via R^2_{os} over multiple prediction horizons - period: January 1985 - end of 2018.

h\n	**2Y**	**3Y**	**4Y**	**5Y**	**7Y**	**10Y**
			M_0			
1m	-0.04	-0.06	-0.05	-0.04	-0.02	-0.01**
3m	0.08**	0.02**	0.01*	0.00*	-0.01*	-0.03**
6m	0.25***	0.16***	0.11**	0.08**	0.06**	0.01**
9m	0.34***	0.21**	0.15**	0.12**	0.12**	0.07**
12m	0.44***	0.32**	0.22**	0.19**	0.19**	0.13**
			M_1			
1m	0.01	0.03**	0.03*	0.02	0.02*	0.04**
3m	0.05**	0.08***	0.07**	0.07**	0.08**	0.13***
6m	0.10***	0.17***	0.16***	0.15**	0.14**	0.20***
9m	0.12**	0.23***	0.25***	0.25***	0.23***	0.31***
12m	0.00*	0.18***	0.26***	0.28***	0.29***	0.37***
			LF_{001}			
1m	0.03**	0.04**	0.04**	0.03*	0.03*	0.05***
3m	0.05**	0.09***	0.07**	0.07**	0.08**	0.13***
6m	0.09**	0.16***	0.15***	0.15**	0.14**	0.21***
9m	0.12**	0.23***	0.25***	0.25***	0.23***	0.31***
12m	0.00*	0.18***	0.26***	0.28***	0.29***	0.37***
			LF_{010}			
1m	0.06***	0.06***	0.05***	0.04**	0.03**	0.03***
3m	0.09***	0.10***	0.08***	0.07**	0.06**	0.09***
6m	0.17***	0.21***	0.17***	0.15***	0.13**	0.16***
9m	0.25***	0.31***	0.29***	0.27***	0.22***	0.27***
12m	0.15**	0.28***	0.32***	0.32***	0.29***	0.33***
			LF_{011}			
1m	0.06***	0.05***	0.04***	0.03**	0.02*	0.03**
3m	0.10***	0.11***	0.08***	0.08**	0.06**	0.10***
6m	0.18***	0.21***	0.17***	0.15**	0.12**	0.16***
9m	0.25***	0.31***	0.28***	0.25***	0.20***	0.25***
12m	0.16**	0.29***	0.32***	0.31***	0.28***	0.32***
			LF_{111}			
1m	0.05***	0.05***	0.03**	0.02*	0.02*	0.02**
3m	0.09**	0.09**	0.06**	0.06**	0.05**	0.08***
6m	0.22***	0.22***	0.17***	0.14**	0.10**	0.13***
9m	0.36***	0.39***	0.33***	0.29***	0.21***	0.24***
12m	0.30***	0.40***	0.39***	0.36***	0.30***	0.31***

This table reports out-of-sample R^2 across alternative models and at different prediction horizons of h = 1-month, 3-month, 6-month, 9-month and 12-month. The six forecasting models used are ATSM with alternative risk price restrictions or different number of latent factors. R^2 values are generated using the out-of-sample R^2 measure of Campbell and Thompson (2008). In particular, out-of-sample R^2 measures the predictive accuracy of bond excess return forecasts relative to the EH benchmark. The EH implies the historical mean being the optimal forecast of excess returns. Positive values of this statistic imply that the forecast outperforms the historical mean forecast and suggests evidence of time-varying return predictability. Statistical significance is measured using a one-sided Diebold-Mariano statistic with Clark-West adjustment, based on Newey-West standard errors. * denotes significance at 10%, ** significance at 5% and *** significance at 1% level. The in-sample period is January 1985 to end of 2007, and the out-of-sample period starts in January 2008 and ends in end of 2018.

Table 2.5: Out-of-sample statistical performance of bond excess return forecasts against model M_1, measured via R^2_{os} over multiple prediction horizons - period: January 1985 - end of 2018.

h\n	**2Y**	**3Y**	**4Y**	**5Y**	**7Y**	**10Y**
			LF_{001}			
1m	0.02**	0.01***	0.01**	0.01**	0.01**	0.01*
3m	0.01	0.00	0.00	0.00	0.00	0.00
6m	-0.01	-0.01	-0.01	0.00	0.00	0.01*
9m	0.00	0.00	0.00	0.00	0.00	0.00
12m	0.00	0.00	0.00	0.00	0.00	0.00
			LF_{010}			
1m	0.05***	0.03***	0.02***	0.02***	0.01	0.00
3m	0.05***	0.02***	0.01	0.00	-0.02	-0.05
6m	0.08***	0.05***	0.02**	0.00	-0.02	-0.05
9m	0.14***	0.10***	0.05***	0.02	-0.01	-0.07
12m	0.14***	0.13***	0.09***	0.05***	0.01	-0.06
			LF_{011}			
1m	0.04***	0.02***	0.01**	0.01	0.00	-0.01
3m	0.06***	0.03***	0.01*	0.00	-0.02	-0.04
6m	0.09***	0.05***	0.02*	0.00	-0.02	-0.05
9m	0.15***	0.10***	0.04***	0.00	-0.03	-0.09
12m	0.16***	0.14***	0.08***	0.04***	-0.01	-0.08
			LF_{111}			
1m	0.04**	0.02	0.00	0.00	-0.01	-0.02
3m	0.04*	0.01	-0.01	-0.02	-0.03	-0.06
6m	0.13***	0.06*	0.01	-0.02	-0.05	-0.09
9m	0.28***	0.20***	0.11**	0.04	-0.03	-0.10
12m	0.30***	0.27***	0.18***	0.11***	0.01	-0.09

This table reports out-of-sample R^2 across alternative models and at different prediction horizons of h = 1-month, 3-month, 6-month, 9-month and 12-month. The four forecasting models used are ATSM with different number of latent factors. R^2 values are generated using the out-of-sample R^2 measure of Campbell and Thompson (2008). In particular, out-of-sample R^2 measures the predictive accuracy of bond excess return forecasts relative to model M_1 from Chapter 1. Positive values of this statistic imply that the forecast outperforms forecast from the benchmark model M_1. Statistical significance is measured using a one-sided Diebold-Mariano statistic with Clark-West adjustment, based on Newey-West standard errors. * denotes significance at 10%, ** significance at 5% and *** significance at 1% level. The in-sample period is January 1985 to end of 2007, and the out-of-sample period starts in January 2008 and ends in end of 2018.

Table 2.6: Out-of-sample economic performance of bond excess return forecasts against the EH, measured via certainty equivalent returns (%) over multiple prediction horizons - period: January 1985 - end of 2018.

h\n	**2Y**	**3Y**	**4Y**	**5Y**	**7Y**	**10Y**
			M_0			
1m	-10.77	-11.81	-10.27	-8.05	-3.62	-7.49
3m	-3.20	-3.84	-3.08	-1.85	-1.63	-3.61
6m	-0.25	-0.68	-0.83	-0.64	-0.45	-1.10
9m	-0.07	-0.78	-1.08	-1.04	-0.50	-0.16
12m	0.84	0.17	-0.58	-0.51	-0.30	0.28
			M_1			
1m	2.31	1.88	1.38	0.80	2.30	2.55
3m	3.76**	3.84*	3.77**	4.01*	4.19**	5.95***
6m	4.65***	4.83***	4.34***	4.11***	3.87***	5.21***
9m	4.52***	4.76***	4.67***	4.51***	4.35***	5.76***
12m	3.83***	4.30***	4.15***	4.18***	4.02***	5.31***
			LF_{001}			
1m	3.96**	3.78**	3.50*	3.25	5.04*	4.84**
3m	4.38***	4.20**	3.96**	3.99*	4.06**	5.97***
6m	4.96***	5.05***	4.51***	4.27***	3.98***	5.51***
9m	4.76***	4.90***	4.69***	4.55***	4.40***	5.88***
12m	4.08***	4.48***	4.36***	4.26***	4.15***	5.48***
			LF_{010}			
1m	3.91**	3.79**	3.56*	3.03	4.20	3.10*
3m	3.25*	3.56**	3.61**	3.66**	3.03*	3.81**
6m	4.35***	4.55***	4.00***	3.71***	3.16**	4.08***
9m	4.62***	4.71***	4.47***	4.16***	3.71***	4.61***
12m	3.95***	4.48***	4.23***	4.12***	3.73***	4.39***
			LF_{011}			
1m	3.70**	3.30*	3.15*	2.40	3.51	2.09
3m	3.65**	3.73**	3.60**	3.45*	2.69	3.75**
6m	4.40***	4.50***	3.78***	3.42***	2.73**	3.89***
9m	4.42***	4.50***	4.23***	3.95***	3.51***	4.45***
12m	3.92***	4.21***	3.98***	3.95***	3.57***	4.35***
			LF_{111}			
1m	-0.05	0.41	0.40	0.17	1.82	0.23
3m	0.44	0.89	1.71	2.43	2.50*	3.00*
6m	3.45***	3.46***	2.87***	2.69**	2.25*	3.07***
9m	3.74***	3.62***	3.31***	3.11***	2.76***	3.49***
12m	3.18***	3.45***	2.91***	2.96***	2.64***	3.23***

This table reports annualized certainty equivalent returns (CERs) across alternative models and at different prediction horizons of h = 1-month, 3-month, 6-month, 9-month and 12-month. The coefficient of risk aversion is $\gamma = 3$. No portfolio constraints are imposed. CERs are generated by out-of-sample forecasts of bond excess returns and are reported in %. At every time step t, an investor with power utility preferences evaluates the entire predictive density of bond excess returns and solves the asset allocation problem, thus optimally allocating her wealth between a riskless bond and risky bonds with maturities 2, 3, 4, 5, 7 and 10-years. CER is then defined as the value that equates the average utility of each alternative model against the average utility of the EH benchmark. The six forecasting models used are ATSM with alternative risk price restrictions or different number of latent factors. Positive values indicate that the models perform better than the EH benchmark. Statistical significance is measured using a one-sided Diebold-Mariano statistic computed with Newey-West standard errors. * denotes significance at 10%, ** significance at 5% and *** significance at 1% level. The in-sample period is January 1985 to end of 2007, and the out-of-sample period starts in January 2008 and ends in end of 2018.

Table 2.7: Out-of-sample economic performance of bond excess return forecasts against model M_1, measured via certainty equivalent returns (%) over multiple prediction horizons - period: January 1985 - end of 2018.

h\n	**2Y**	**3Y**	**4Y**	**5Y**	**7Y**	**10Y**
			LF_{001}			
1m	1.65***	1.90***	2.12**	2.45***	2.74***	2.29**
3m	0.61	0.36	0.18	-0.01	-0.13	0.02
6m	0.30*	0.21	0.17	0.16	0.11	0.28*
9m	0.23*	0.13	0.02	0.04	0.04	0.11
12m	0.24**	0.17*	0.20**	0.08	0.12*	0.17**
			LF_{010}			
1m	1.60*	1.90**	2.17***	2.23**	1.90**	0.55
3m	-0.51	-0.27	-0.16	-0.35	-1.14	-2.11
6m	-0.30	-0.28	-0.33	-0.39	-0.69	-1.10
9m	0.10	-0.05	-0.19	-0.34	-0.62	-1.11
12m	0.11	0.17	0.08	-0.05	-0.28	-0.87
			LF_{011}			
1m	1.39*	1.41*	1.77**	1.60*	1.21	-0.46
3m	-0.11	-0.10	-0.18	-0.55	-1.48	-2.16
6m	-0.24	-0.33	-0.54	-0.68	-1.11	-1.29
9m	-0.09	-0.26	-0.42	-0.54	-0.82	-1.26
12m	0.08	-0.09	-0.16	-0.22	-0.43	-0.90
			LF_{111}			
1m	-2.35	-1.47	-0.98	-0.63	-0.48	-2.31
3m	-3.29	-2.92	-2.04	-1.56	-1.67	-2.90
6m	-1.17	-1.34	-1.44	-1.38	-1.58	-2.09
9m	-0.75	-1.10	-1.31	-1.36	-1.55	-2.18
12m	-0.63	-0.82	-1.19	-1.17	-1.33	-1.97

This table reports annualized certainty equivalent returns (CERs) across alternative models and at different prediction horizons of h = 1-month, 3-month, 6-month, 9-month and 12-month. The coefficient of risk aversion is $\gamma = 3$. No portfolio constraints are imposed. CERs are generated by out-of-sample forecasts of bond excess returns and are reported in %. At every time step t, an investor with power utility preferences evaluates the entire predictive density of bond excess returns and solves the asset allocation problem, thus optimally allocating her wealth between a riskless bond and risky bonds with maturities 2, 3, 4, 5, 7 and 10-years. CER is then defined as the value that equates the average utility of each alternative model against the average utility of model M_1 from Chapter 1. The four forecasting models used are ATSM with different number of latent factors. Positive values indicate that the models perform better than the benchmark model M_1. Statistical significance is measured using a one-sided Diebold-Mariano statistic computed with Newey-West standard errors. * denotes significance at 10%, ** significance at 5% and *** significance at 1% level. The in-sample period is January 1985 to end of 2007, and the out-of-sample period starts in January 2008 and ends in end of 2018.

Table 2.8: Explanatory power of macroeconomic variables when fitting latent factors and their components, measured via $\bar{R}^2$ - period: January 1985 - end of 2017.

$\bar{R}^2 : Z_{j,t}/E[Z_{j,t}\|\mathcal{P}_t]/\widetilde{RP}^Z_{j,t} = a_j + b_j' M_t + e_{j,t},\ j \in \{1,2,3\}$										
LF_{001}	CPI	GRO	$F1$	$F1^3$	$F8$	UNR	MNF	M^I	M^{II}	M^{III}
$Z_{3,t}$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$E[Z_{3,t}\|\mathcal{P}_t]$	0.00	0.17	0.17	0.14	0.00	0.04	0.00	0.17	0.18	0.06
$\widetilde{RP}^Z_{3,t}$	0.00	0.01	0.01	0.01	0.00	0.00	0.00	0.00	0.01	0.00
LF_{010}	CPI	GRO	$F1$	$F1^3$	$F8$	UNR	MNF	M^I	M^{II}	M^{III}
$Z_{2,t}$	0.05	0.03	0.03	0.00	0.02	0.20	0.08	0.08	0.05	0.21
$E[Z_{2,t}\|\mathcal{P}_t]$	0.09	0.01	0.00	0.00	0.01	0.06	0.00	0.11	0.01	0.17
$\widetilde{RP}^Z_{2,t}$	0.00	0.11	0.09	0.02	0.01	0.15	0.19	0.11	0.10	0.20
LF_{011}	CPI	GRO	$F1$	$F1^3$	$F8$	UNR	MNF	M^I	M^{II}	M^{III}
$Z_{2,t}$	0.06	0.01	0.02	0.00	0.02	0.19	0.07	0.07	0.04	0.20
$E[Z_{2,t}\|\mathcal{P}_t]$	0.09	0.02	0.01	0.01	0.01	0.04	0.00	0.12	0.02	0.13
$\widetilde{RP}^Z_{2,t}$	0.00	0.13	0.11	0.03	0.01	0.18	0.21	0.13	0.12	0.23
$Z_{3,t}$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-0.01	0.00
$E[Z_{3,t}\|\mathcal{P}_t]$	0.00	0.17	0.16	0.14	0.00	0.03	0.00	0.17	0.18	0.03
$\widetilde{RP}^Z_{3,t}$	0.00	0.01	0.01	0.01	0.00	0.00	0.01	0.01	0.01	0.01
LF_{111}	CPI	GRO	$F1$	$F1^3$	$F8$	UNR	MNF	M^I	M^{II}	M^{III}
$Z_{1,t}$	0.05	0.12	0.17	0.01	0.00	0.16	0.15	0.18	0.21	0.18
$E[Z_{1,t}\|\mathcal{P}_t]$	0.10	0.00	0.00	0.00	0.02	0.09	0.00	0.10	0.01	0.24
$\widetilde{RP}^Z_{1,t}$	0.02	0.14	0.17	0.01	0.00	0.11	0.17	0.16	0.22	0.18
$Z_{2,t}$	0.07	0.04	0.05	0.01	0.01	0.26	0.09	0.11	0.07	0.26
$E[Z_{2,t}\|\mathcal{P}_t]$	0.09	0.01	0.00	0.00	0.01	0.06	0.00	0.10	0.01	0.17
$\widetilde{RP}^Z_{2,t}$	0.00	0.18	0.17	0.04	0.00	0.23	0.27	0.18	0.18	0.29
$Z_{3,t}$	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.01
$E[Z_{3,t}\|\mathcal{P}_t]$	0.00	0.18	0.17	0.14	0.00	0.04	0.00	0.18	0.19	0.05
$\widetilde{RP}^Z_{3,t}$	0.00	0.02	0.03	0.02	0.00	0.01	0.01	0.02	0.03	0.01

This table reports in-sample $\bar{R}^2$ across alternative regression specifications. The explained variables are individual latent factors $Z_{j,t}$ and their components $E[Z_{j,t}|\mathcal{P}_t]$ and $\widetilde{RP}^Z_{j,t}$, $j \in \{1,2,3\}$, see (2.18), from models LF_{001}, LF_{010}, LF_{011} and LF_{111}. The explanatory variables are individual macroeconomic variables or groups thereof, namely $M^I = [CPI, GRO]'$, $M^{II} = [F1, F1^3, F8]'$, $M^{III} = [UNR, MNF]'$. The sample period is January 1985 to end of 2017.

Table 2.9: Signs and significance of coefficients from explanatory power regressions of latent factors and their components on macroeconomic variables - period: January 1985 - end of 2017.

$sign(b_j) : Z_{j,t}/E[Z_{j,t}\|\mathcal{P}_t]/\widetilde{RP}^{Z}_{j,t} = a_j + b_j M_t + e_{j,t},\ j \in \{1,2,3\}$							
LF_{001}	CPI	GRO	$F1$	$F1^3$	$F8$	UNR	MNF
$Z_{3,t}$	$-$	$-$	$-$	$-$	$+$	$+$	$+$
$E[Z_{3,t}\|\mathcal{P}_t]$	$-$	$-^{***}$	$-^{***}$	$-^{***}$	$+$	$+^{*}$	$-$
$\widetilde{RP}^{Z}_{3,t}$	$-$	$+$	$+$	$+^{*}$	$+$	$-$	$+$
LF_{010}	CPI	GRO	$F1$	$F1^3$	$F8$	UNR	MNF
$Z_{2,t}$	$+^{**}$	$-^{**}$	$-^{**}$	$-^{*}$	$+^{***}$	$+^{***}$	$-^{***}$
$E[Z_{2,t}\|\mathcal{P}_t]$	$+^{***}$	$+^{**}$	$+$	$+^{***}$	$+^{**}$	$+^{**}$	$+$
$\widetilde{RP}^{Z}_{2,t}$	$+$	$-^{***}$	$-^{***}$	$-^{***}$	$+^{*}$	$+^{***}$	$-^{***}$
LF_{011}	CPI	GRO	$F1$	$F1^3$	$F8$	UNR	MNF
$Z_{2,t}$	$+^{***}$	$-^{*}$	$-^{**}$	$-$	$+^{**}$	$+^{***}$	$-^{***}$
$E[Z_{2,t}\|\mathcal{P}_t]$	$+^{***}$	$+^{***}$	$+^{**}$	$+^{***}$	$+^{**}$	$+^{**}$	$+$
$\widetilde{RP}^{Z}_{2,t}$	$+$	$-^{***}$	$-^{***}$	$-^{***}$	$+^{*}$	$+^{***}$	$-^{***}$
$Z_{3,t}$	$-$	$-$	$-$	$-$	$+$	$-$	$+^{*}$
$E[Z_{3,t}\|\mathcal{P}_t]$	$-$	$-^{***}$	$-^{***}$	$-^{***}$	$-$	$+$	$-$
$\widetilde{RP}^{Z}_{3,t}$	$-$	$+^{**}$	$+^{*}$	$+^{**}$	$+$	$-^{*}$	$+^{**}$
LF_{111}	CPI	GRO	$F1$	$F1^3$	$F8$	UNR	MNF
$Z_{1,t}$	$-^{***}$	$+^{***}$	$+^{***}$	$+^{***}$	$+$	$-^{***}$	$+^{***}$
$E[Z_{1,t}\|\mathcal{P}_t]$	$-^{***}$	$-$	$-$	$-$	$-^{**}$	$-^{***}$	$-$
$\widetilde{RP}^{Z}_{1,t}$	$-^{**}$	$+^{***}$	$+^{***}$	$+^{***}$	$+^{**}$	$-^{***}$	$+^{***}$
$Z_{2,t}$	$+^{***}$	$-^{***}$	$-^{***}$	$-^{**}$	$+^{**}$	$+^{***}$	$-^{***}$
$E[Z_{2,t}\|\mathcal{P}_t]$	$+^{***}$	$+^{**}$	$+$	$+^{***}$	$+^{**}$	$+^{**}$	$+$
$\widetilde{RP}^{Z}_{2,t}$	$+$	$-^{***}$	$-^{***}$	$-^{***}$	$+$	$+^{***}$	$-^{***}$
$Z_{3,t}$	$-^{**}$	$+$	$+$	$+$	$+$	$-$	$+^{**}$
$E[Z_{3,t}\|\mathcal{P}_t]$	$-$	$-^{***}$	$-^{***}$	$-^{***}$	$+$	$+$	$-$
$\widetilde{RP}^{Z}_{3,t}$	$-$	$+^{***}$	$+^{***}$	$+^{**}$	$+$	$-^{**}$	$+^{**}$

This table reports signs and statistical significance of coefficients b_j, $j \in \{1,2,3\}$, across alternative regression specifications. The explained variables are individual latent factors $Z_{j,t}$ and their components $E[Z_{j,t}|\mathcal{P}_t]$ and $\widetilde{RP}^{Z}_{j,t}$, $j \in \{1,2,3\}$, see (2.18), from models LF_{001}, LF_{010}, LF_{011} and LF_{111}. The explanatory variables are macroeconomic variables. Statistical significance is measured using t-statistic computed with Newey-West standard errors. * denotes significance at 10%, ** significance at 5% and *** significance at 1% level. The sample period is January 1985 to end of 2017.

Chapter 3

Dynamic Term Structure Models with Nonlinear Information

3.1 Introduction

3.1.1 Term Structure and Macroeconomic Information

A fundamental and economically important argument in macro-finance literature suggests that the current yield curve spans all relevant information for forecasting future yields, returns and bond risk premia. Although the argument is implied by most macro-finance models, recent evidence raises important questions on its validity. In particular, several empirical studies (see, Cooper and Priestley (2009), Ludvigson and Ng (2009), Duffee (2011), Joslin et al. (2014), Cieslak and Povala (2015), Gargano et al. (2019), Bianchi et al. (2021)) provide support for unspanned macroeconomic risks[1] possessing considerable predictive power above and beyond the yield curve. The vast majority of the extant literature, however, assumes that the relationship between those unspanned macroeconomic factors and the cross section of yields is mostly, if not purely, linear in nature.

[1] e.g.: the output gap of Cooper and Priestley (2009), the 'real' and 'inflation' factors of Ludvigson and Ng (2009), the measures of economic activity and inflation of Joslin et al. (2014) and the long-run inflation expectation of Cieslak and Povala (2015), among others.

Duffee (2011) is the first to comment on the possibility of a nonlinear (albeit, economically weak) relationship between bond risk premia and macroeconomic activity. Only recently, Bianchi et al. (2021) provide evidence on the importance of accounting for nonlinearities in order to detect further information important for forecasting bond risk premia. Due to the rapid increase in the use of machine learning methods, the presence of nonlinearities, relevant for predicting excess returns, has recently attracted attention in the literature. However, we are unaware of any prior research that explores the possibility of an asymmetric/nonlinear relationship between unspanned macroeconomic risks and return predictability in the US bond market within an arbitrage-free pricing model. In this chapter we seek to investigate this relationship, if any, further. With this in mind, we propose a novel class of arbitrage-free Dynamic Term Structure Models (DTSM) that embed nonlinear unspanned macroeconomic risks using Gaussian Processes (Bishop, 2006; Rasmussen and Williams, 2006). The latter have been applied in the context of nonlinear nonparametric state-space models, where a Gaussian process prior was placed over the state transition dynamics (Frigola et al., 2013). However, our setting is different, as described below.

This chapter contributes to the literature that studies the linkages between bond risk premia and the macroeconomy. Our first contribution is methodological. Following Joslin et al. (2014), we embed unspanned information coming from (directly observable) macroeconomic risks within a DTSM. The point where our approach deviates from prior studies is that this information is assumed exogenous and is entering the model in a nonlinear fashion, without making specific assumption on the manner it affects the yield curve. To do so, we introduce a novel methodological framework that utilises Bayesian nonparametrics within a DTSM, thus staying agnostic about the functional form between macroeconomic activity and bond risk premia. Our setup, also allows us to handle Gaussian Processes sequentially and effectively perform tasks such as sequential parameter estimation and forecasting.

Drawing from the work of Chopin (2002) and Del Moral et al. (2006), we allow for a sequential Bayesian treatment of exogeneous macroeconomic information in a nonlinear manner.

Our second contribution is empirical. From a fundamental and economic perspective, the proposed framework seeks to enhance our understanding behind the determinants of bond risk premia, in particular, whether movements in excess bond returns bear any relation to the macroeconomy. Are there important driving forces of bond return predictability that are nonlinear and hard, if not impossible, to summarize utilising information coming solely from linear transformations of the data? Furthermore, we seek to investigate deeper whether such nonlinearities, if any, allow investors to exploit the evident statistical predictability of the resulting models and generate economically significant portfolio benefits to bond investors, out-of-sample.

3.1.2 Economic Benefits from Nonlinearities

Our results reveal a direct line linking the nonlinear component of the unspanned measures of macroeconomic activity to model performance, especially when it comes to generating additional economic value. In order to scrutinize nonlinear models in a meaningful way, as benchmark we apply a version of the macro-finance DTSM proposed by Joslin et al. (2014) which incorporates unspanned macroeconomic variables in a linear manner. However, we proceed without modeling their underlying dynamics in order to remain closest possible to the nonlinear setup we propose.

First, we find that nonlinear models provide a competitive edge over linear models in cases where *ex-ante* nonlinear functions of macroeconomic variables preserve their nonlinear nature *ex-post* in the part which is hidden from the yield curve. In this chapter these are specifically the prices and not the economic activity which we also consider. We demonstrate that by resorting to risk premium factor decomposition from Duffee (2011).

Second, results in this chapter also indicate that for some models where macroeconomic risk directly affects the level of the yield curve, a nonlinear formulation like ours, which relies on Gaussian Processes, does not necessarily mean outstanding model performance in terms of predictability and economic value gains, however it lets avoid substantial deterioration on that front. In particular, this is the case for macroeconomic variables which in abnormal times, such as the 2008-2009 financial crisis or the more recent COVID-19 recession, are subject to outliers, like in this chapter the economic activity.

3.1.3 Outline

The remainder of this chapter is organized as follows. Section 3.2 is a brief introduction to Gaussian Processes. Section 3.3 describes the proposed modelling framework. Section 3.4 presents the procedure for sequential learning with Gaussian Processes and forecasting, along with the framework for assessing the predictive and economic performance of the resulting models. Section 3.5 discusses the data and the sample period used and presents the family of models considered in this chapter. Section 3.6 discusses the results both in terms of predictive performance and economic value, including the associated explanatory power where applicable, and reveals the links between these results and the hidden nonlinearities. Finally, Section 3.7 concludes the chapter by providing some relevant discussion.

3.2 Gaussian Processes

In Bayesian nonparametrics a Gaussian Process prior is used to estimate unknown function in supervised learning setting. One does not assume a specific function between some y and x and instead performs Bayesian inference on the function. Univariate exposition below follows from Bishop (2006), Rasmussen and Williams (2006) and Murphy (2012).

3.2.1 Gaussian Process Theory

Let $\mathscr{X}$ be a set and $\mathscr{V}$ be a set of functions over $\mathscr{X}$, for example smooth functions. We observe (x_1, y_1) ... (x_t, y_t) ... (x_T, y_T), where $x_t \in \mathscr{X}$, $y_t \in \mathscr{R}$, $t = 1, ..., T$, satisfying

$$y_t = v(x_t) + \epsilon_t$$

where $v \in \mathscr{V}$ and $\epsilon = (\epsilon_1, ..., \epsilon_T)$ are independent of $x = (x_1, ..., x_T)$. Errors ϵ have density $\pi(\epsilon)$ and usually it is $N(0_T,\ \sigma_\epsilon^2 I_T)$, 0_T being $(T \times 1)$ vector of zeros, what further implies that

$$y = (y_1, ..., y_T)|v, x, \sigma_\epsilon^2 \sim N(v,\ \sigma_\epsilon^2 I_T)$$

where $v = [v(x_1), ..., v(x_T)]'$. In a Bayes regression we assign prior $\pi(v)$ to v and compute the posterior

$$\pi(v|y) = \frac{\pi(v) f(\epsilon|v)}{\int_{\mathscr{V}} \pi(v) f(\epsilon|v) dv}$$

where $f(\epsilon|v)$ is the likelihood. Then v can be estimated by its posterior mean.

Definition. Let $\mathscr{X}$ be a set. A random function $v : \mathscr{X} \to \mathscr{R}$ is called a Gaussian Process (GP) if for any $x_1, ..., x_T \in \mathscr{X}$, $v = [v(x_1), ..., v(x_T]'$ has a multivariate normal distribution.

Gaussian Process is characterized by the mean $v_0(\cdot)$ and the covariance kernel $K(\cdot, \cdot)$. The latter is a positive definite function $K : \mathscr{X} \times \mathscr{X} \to \mathscr{R}$. If v is a GP with mean v_0 and covariance kernel K then

$$v = [v(x_1), ..., v(x_T]' \sim N(v_0,\ K)$$

where $v_0 = [v_0(x_1), ..., v_0(x_T]$ and K is $(T \times T)$ matrix with elements $K(s, t)$ such that $s, t = 1, ..., T$. There are numerous possibilities when it comes to covariance

kernels and their choice depends on the application. For instance, they can be linear kernels

$$K(x, x') = \langle x, x' \rangle$$

or Gaussian kernels

$$K(x, x') = e^{-\frac{\|x-x'\|^2}{2\sigma^2}}$$

among others. For details about these and other kernels, and how they can be combined, see Rasmussen and Williams (2006).

3.2.2 Gaussian Process Regression

For $t = 1, ..., T$ assume that

$$y_t = v(x_t) + \epsilon_t$$

with

$$\pi(v) = GP(v_0,\ K)$$
$$\epsilon_t \sim N(0,\ \sigma_\epsilon^2)$$

where v and $(\epsilon_1, ..., \epsilon_T)$ are independent and for simplicity one can assume that $v_0 = 0$, what is also a common assumption made in applications.

Marginal distribution of y is thus multivariate normal with means $E(y_t) = 0$ and covariances $Cov(y_s, y_t) = K(x_s, x_t) + \sigma_\epsilon^2 I(s = t)$, with $s, t = 1, ..., T$. In matrix notation, where K_T denotes the $(T \times T)$ matrix containing all the (x_s, x_t) pairs, we get

$$y \sim N(0_{T\times 1},\ K_T + \sigma_\epsilon^2 I_T)$$

Since for each $x_t \in \mathscr{X}$ and y_t we have that $Cov(v(x_s), y_t) = K(x_s, x_t)$, the joint distribution of v and y is

$$\begin{pmatrix} v \\ y \end{pmatrix} \sim N\left[\begin{pmatrix} 0_{T\times 1} \\ 0_{T\times 1} \end{pmatrix}, \begin{pmatrix} K_T & K_T \\ K_T & K_T + \sigma_\epsilon^2 I_T \end{pmatrix}\right]$$

where $v = [v(x_1), ..., v(x_T)]'$ and $0_{T\times 1}$ is a $(T \times 1)$ vector of zeros.

Under such assumptions, by the multivariate normal regression lemma (Durbin and Koopman, 2012), the distribution of $v|y$ is a Gaussian with mean μ_v and covariance kernel Σ_v, which are following

$$\mu_v = K_T[K_T + \sigma_\epsilon^2 I_T]^{-1} y$$
$$\Sigma_v = K_T - K_T[K_T + \sigma_\epsilon^2 I_T]^{-1} K_T$$

To project y_{T+1} based on x_{T+1} we need to find the joint distribution of y and y_{T+1}, given x and x_{T+1}. To that end, we note that $Cov(y_t, y_{T+1}) = K(x_t, x_{T+1})$ for $t = 1, ..., T$ and we let

$$k_{T+1} = [K(x_1, x_{T+1}), ..., K(x_T, x_{T+1})]'$$

to arrive at

$$\begin{pmatrix} y_{T+1} \\ y \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0_{T\times 1} \end{pmatrix}, \begin{pmatrix} K(x_{T+1}, x_{T+1}) + \sigma_\epsilon^2 & k'_{T+1} \\ k_{T+1} & K_T + \sigma_\epsilon^2 I_T \end{pmatrix}\right]$$

Then, by the multivariate normal regression lemma, $y_{T+1}|y$ is a Gaussian with mean μ_{T+1} and variance Σ_{T+1} such that

$$\mu_{T+1} = k'_{T+1}[K_T + \sigma_\epsilon^2 I_T]^{-1} y$$
$$\Sigma_{T+1} = K(x_{T+1}, x_{T+1}) + \sigma_\epsilon^2 - k'_{T+1}[K_T + \sigma_\epsilon^2 I_T]^{-1} k_{T+1}$$

Posterior distribution of v still depends on unknown vector of parameters θ consisting of σ_ϵ and any hyper-parameters of K, for example σ for squared exponential

kernel. In principle, two methods can be used to estimate θ. We can maximize the marginal likelihood $\pi(y|\theta)$ which is a multivariate normal density with mean $\mu_y = 0_{T\times 1}$ and covariance matrix $\Sigma_y = K_T + \sigma_\epsilon^2 I_T$. Doing so is denoted empirical Bayes. We can also put a prior on the hyper-parameters and estimate them by their posterior means. Doing this is known as hierarchical Bayes (Bishop, 2006; Murphy, 2012). In what follows, we pursue a pragmatic combination of the two approaches.

3.3 Dynamic Term Structure Model, Likelihood, and Nonlinear Macros

3.3.1 Incorporating Unspanned Nonlinear Macros

In this chapter we consider an extension of the model presented in Section 1.2.1. Our approach resembles the framework of Joslin et al. (2014) in that the model is factorised into a 'spanned' component, i.e. risk factors which can be retrieved by the information provided in historical yield curve data, as well as an 'unspanned' component that could include background factors such as macroeconomic variables. It is assumed that the latter is not determined by the yield curve, yet it remains highly relevant for the inference and, more importantly, prediction purposes. The points where our approach differs from Joslin et al. (2014) are as follows. First, the unspanned components are regarded as unknown, possibly nonlinear functions, which are to be estimated, of observable macroeconomic variables, rather than just macros entering the model in a linear manner. Having said that, the latter case is considered too, as benchmark in model comparisons, and is henceforth referred to as the linear model. Second, the unspanned components are not assumed endogenous but exogenous.

We thus introduce the following nonlinear model for the $\mathbb{P}$-dynamics of the state vector

$$\mathcal{P}_t = \mu_{\mathcal{P}}^{\mathbb{P}} + \Phi_{\mathcal{P}}^{\mathbb{P}} \mathcal{P}_{t-1} + v(M_{t-1}) + \Sigma_{\mathcal{P}} \varepsilon_t \tag{3.1}$$

where $\mu_{\mathcal{P}}^{\mathbb{P}}$ and $\Phi_{\mathcal{P}}^{\mathbb{P}}$ are defined in (1.15) and (1.16), respectively. The matrix $\Sigma_{\mathcal{P}}$ represents the Cholesky decomposition (i.e. lower triangular with positive diagonal elements) of the covariance matrix $\Sigma_{\mathcal{P}}\Sigma_{\mathcal{P}}'$, whereas ε_t is a vector of N error terms that are assumed to be normally distributed with zero mean and I_N, the identity matrix of dimension N, as the covariance matrix. The term $v(M_{t-1})$ is of dimension $(N \times 1)$ and reflects the impact of R (lagged) macros in M_{t-1}, a $(R \times 1)$ vector, on $\mathcal{P}_t$, with the function $v(\cdot)$ being such that $v : \mathbb{R}^R \to \mathbb{R}^N$. More details are provided in the next section.

The model is completed by setting the $\mathbb{Q}$-dynamics for $\mathcal{P}_t$ as in (1.7). To ensure that macros are not spanned by $\mathcal{P}_t$ we set r_t as in (1.12) and the pricing kernel as in (1.2), however the time-varying market prices of risk λ_t are now given by

$$\begin{aligned}
\lambda_t &= \Sigma_{\mathcal{P}}^{-1} \left[\mu_{\mathcal{P}}^{\mathbb{P}} - \mu_{\mathcal{P}}^{\mathbb{Q}} + (\Phi_{\mathcal{P}}^{\mathbb{P}} - \Phi_{\mathcal{P}}^{\mathbb{Q}}) \mathcal{P}_t + v(M_{t-1}) \right] \\
&= \Sigma_{\mathcal{P}}^{-1} \left[\lambda_{0\mathcal{P}} + \lambda_{1\mathcal{P}} \mathcal{P}_t + v(M_{t-1}) \right]
\end{aligned}$$

Final model specification assumes appropriate choice of restrictions on $\lambda_{0\mathcal{P}}$ and $\lambda_{1\mathcal{P}}$ in λ_t, as discussed in Section 3.3.3, and Gaussian Process outputs in $v(M_{t-1})$, as explained in the next section, where the latter depend on observed macroeconomic variables described in Section 3.5.

3.3.2 Gaussian Process Mean Ornstein-Uhlenbeck Model

In what follows, we define in more detail the function $v(\cdot)$ in (3.1), and the underlying hyper-parameters. To that end, we follow a Bayesian nonparametric approach using a Gaussian Process with multiple outputs. We focus on a special case of (3.1) with $N = 3$ and $R = 1$, noting that the first three extracted PCs are typically sufficient

to capture most of the variation in the yield curve and often correspond to its level, slope and curvature, respectively (Litterman and Scheinkman, 1991). For interpretation reasons, we work with models containing $R = 1$ macro and explore their submodels. We believe it is a natural step to gain the understanding as to how these models behave before moving to more complex specifications. Nevertheless, the computational scheme we develop in this chapter is capable of handling $R > 1$ macros in a single model.

Our starting point is the $\mathbb{P}$-dynamics for $\mathcal{P}_t$ as in (3.1) which, for convenience, we simplify to

$$s_t = v(M_{t-1}) + \Sigma_{\mathcal{P}}\varepsilon_t \tag{3.2}$$

by setting

$$\mathcal{P}_t - \mu_{\mathcal{P}}^{\mathbb{P}} - \Phi_{\mathcal{P}}^{\mathbb{P}}\mathcal{P}_{t-1} = s_t \tag{3.3}$$

Then we assign a Gaussian Process prior on $v(\cdot)$ in (3.3) and denote it as

$$\pi(v) = GP(v_0,\ K) \tag{3.4}$$

In the above we assume $v_0 = 0$, what is commonly done in applications (Rasmussen and Williams, 2006). Such a prior assumption about v_0 is also reasonable in our context where we are effectively filtering out GPs from residuals of a VAR model. Specific choice of kernel K is discussed later.

At this point it is worth noting that under these assumptions the term $\mu_{\mathcal{P}}^{\mathbb{P}} + v(M_{t-1})$, as backed out from (3.2) and (3.3), essentially determines (up to rescaling) the long term mean of the corresponding Ornstein-Uhlenbeck (OU) process that governs $\mathcal{P}_t$. Given that $v(M_{t-1})$ is now a Gaussian Process, the model can be viewed as N-dimensional GP mean OU process. Although according to our notation dimension of $v(M_{t-1})$ in (3.2) is $(N \times 1)$, we are able to include $G = 1, \ldots, N$ Gaussian Process outputs in a single model. For example, if $G = 2$ and we introduce two GP outputs, one in the second and one in the third equation of (3.2), then it

means that the first element of $v(M_{t-1})$ is effectively zero. It is implemented by setting K_1, the first diagonal block of matrix K defined in (3.7) below, to zeros. In the linear model, this would be equivalent to restricting the first row of matrix $\Phi^{\mathbb{P}}_{\mathcal{P}M}$ in (3.24) to zeros, what is explained in detail in Appendix 3.E.

Without loss of generality for $G = 3$, we adopt the following vector notation

$$s_t = \begin{bmatrix} s_t^{(1)} \\ s_t^{(2)} \\ s_t^{(3)} \end{bmatrix}, \quad s_{jt} = s_t^{(j)}, \quad S_j = [s_{j1}, \ldots, s_{jT}]', \quad j = 1, 2, 3$$

$$v(M_t) = \begin{bmatrix} v^{(1)}(M_t) \\ v^{(2)}(M_t) \\ v^{(3)}(M_t) \end{bmatrix}, \quad v_{jt} = v^{(j)}(M_t), \quad V_j = [v_{j0}, \ldots, v_{jT-1}]', \quad j = 1, 2, 3$$

Eventually we consider the concatenated vectors

$$S = \begin{bmatrix} S_1 \\ S_2 \\ S_3 \end{bmatrix}, \quad V = \begin{bmatrix} V_1 \\ V_2 \\ V_3 \end{bmatrix} \tag{3.5}$$

The distribution of S conditional on V is then

$$S|V \sim N\left(V,\ \Sigma_{\mathcal{P}}\Sigma'_{\mathcal{P}} \otimes I_T\right) \tag{3.6}$$

The next step is to choose a specific kernel for each GP $v^{(j)}(M_t)$, $j = 1, 2, 3$. We use the same type of kernel for all the GPs in the model. Namely, the squared exponential kernel, which is stationary and considered the choice most-widely made to start with in applications (Bishop, 2006; Rasmussen and Williams, 2006; Murphy, 2012). It can be specified as

$$k_{\ell,\sigma^2}(x_a, x_b) = \sigma^2 \exp\left(-\frac{\|x_a - x_b\|^2}{2\ell^2}\right)$$

for some scalars (or vectors) x_a and x_b, where ℓ is called the characteristic length-scale and σ is denoted as signal standard deviation. Squared exponential kernel is infinitely differentiable and thus very smooth, what we welcome here since it is the smooth signal within the rough noise that we are after.

To each V_j we thus assign separate yet the same type of kernel $k_{\ell_j,\sigma_j^2}(\cdot,\cdot)$, $j = 1, 2, 3$, what can be expressed as

$$K_j = \begin{bmatrix} k_{\ell_j,\sigma_j^2}(M_0, M_0) & \dots & k_{\ell_j,\sigma_j^2}(M_0, M_{T-1}) \\ \vdots & \ddots & \vdots \\ k_{\ell_j,\sigma_j^2}(M_{T-1}, M_0) & \dots & k_{\ell_j,\sigma_j^2}(M_{T-1}, M_{T-1}) \end{bmatrix}, \quad j = 1, 2, 3$$

in covariance terms. Following the assumption made about v_0 in (3.4), it lets us then determine the distribution for each of the V_js as

$$V_j \sim N(0_{T\times 1},\ K_j), \quad j = 1, 2, 3$$

where $0_{T\times 1}$ is a $(T \times 1)$ vector of zeros and K_j is the $(T \times T)$ matrix shown above.

Finally, in order to specify the distribution of V, which in our case is a single input GP with multiple outputs, we need to make an assumption about the dependence between V_js. The simplest way, and the one chosen here, is to assume independence between V_js, what implies the following distribution

$$V \sim N\left(0_{3T\times 1},\ K\right)$$

with

$$K = \begin{bmatrix} K_1 & 0_T & 0_T \\ 0_T & K_2 & 0_T \\ 0_T & 0_T & K_3 \end{bmatrix} \tag{3.7}$$

which is a $(3T \times 3T)$ matrix, and $0_{3T\times 1}$ and 0_T are a $(3T \times 1)$ and $(T \times T)$ vector and matrix of zeros, respectively. We collect the hyper-parameters governing covariance matrix K, which are σ_j and ℓ_j, $j = 1, 2, 3$, in $\sigma_K = [\sigma_1, \sigma_2, \sigma_3]'$ and $\ell_K = [\ell_1, \ell_2, \ell_3]'$.

There are also more flexible alternatives for tying up together the V_js and specifying the covariance of V, that is K, see for example Alvarez et al. (2012) for a survey of several methods. These include the intrinsic coregionalisation model, the semiparametric latent factor model and the linear model of coregionalisation, among others. Choosing between such models involves considering trade-offs between flexibility, computational cost and interpretability. Here we adopt (3.7) as a convenient starting point for DTSMs. Consequently, such a single input GP with multiple inputs is then equivalent to multiple independent single input GPs with a single output.

Next, using standard properties of Gaussian Processes we can proceed to obtain the conditional distribution of V given S, the marginal distribution of S, by integrating out V from (3.6), and the predictive distribution of s_{T+1}, and consequently that of $\mathcal{P}_{T+1}$, given all the data up to time T including M_T. For reasons mentioned in Section 3.4.2, in this chapter we focus on one-step-ahead predictions. To start with, the marginal distribution of S, which *nota bene* stands behind the $\mathbb{P}$-dynamics of $\mathcal{P}_t$, is following

$$S \sim N\left(0_{3T\times 1},\ K + \Sigma_{\mathcal{P}}\Sigma_{\mathcal{P}}' \otimes I_T\right) \tag{3.8}$$

To obtain conditional distribution of V given S we observe that joint distribution of V and S is

$$\begin{pmatrix} V \\ S \end{pmatrix} \sim N\left(\begin{bmatrix} 0_{3T\times 1} \\ 0_{3T\times 1} \end{bmatrix}, \begin{bmatrix} K & K \\ K & K_{\mathcal{P}} \end{bmatrix}\right)$$

where for clarity of further exposition we defined

$$K_{\mathcal{P}} = K + \Sigma_{\mathcal{P}}\Sigma_{\mathcal{P}}' \otimes I_T \tag{3.9}$$

Then, by the multivariate normal regression lemma, we get

$$V|S \sim N\left(KK_{\mathcal{P}}^{-1}S,\ K - KK_{\mathcal{P}}^{-1}K\right)$$

To arrive at the predictive distribution of s_{T+1}, we first derive the joint distribution of s_{T+1} and S, using all available macro data, that is $M_{0:T}$. Given the assumption about dependence between V_js, as it is made in (3.7), we first note that for $t = 1, \ldots, T$

$$Cov(s_t, s_{T+1}) = \begin{bmatrix} k_{\ell_1,\sigma_1^2}(M_{t-1}, M_T) & 0 & 0 \\ 0 & k_{\ell_2,\sigma_2^2}(M_{t-1}, M_T) & 0 \\ 0 & 0 & k_{\ell_3,\sigma_3^2}(M_{t-1}, M_T) \end{bmatrix}$$

Further, for $j = 1, 2, 3$, we let

$$k_{T+1}^j = [k_{\ell_j,\sigma_j^2}(M_0, M_T), ..., k_{\ell_j,\sigma_j^2}(M_{T-1}, M_T)]'$$

to eventually observe that joint distribution of s_{T+1} and S is as follows

$$\begin{pmatrix} s_{T+1} \\ S \end{pmatrix} \sim N\left(\begin{bmatrix} 0_{3\times 1} \\ 0_{3T\times 1} \end{bmatrix}, \begin{bmatrix} k_{T+1}^0 + \Sigma_{\mathcal{P}}\Sigma_{\mathcal{P}}' & k_{T+1}' \\ k_{T+1} & K_{\mathcal{P}} \end{bmatrix}\right)$$

where k_{T+1}^0 and k_{T+1} which are (3×3) and $(3T \times 3)$ matrices, respectively, are specified as

$$k_{T+1}^0 = \begin{bmatrix} k_{\ell_1,\sigma_1^2}(M_T, M_T) & 0 & 0 \\ 0 & k_{\ell_2,\sigma_2^2}(M_T, M_T) & 0 \\ 0 & 0 & k_{\ell_3,\sigma_3^2}(M_T, M_T) \end{bmatrix} \tag{3.10}$$

and

$$k_{T+1} = \begin{bmatrix} k^1_{T+1} & 0_{T\times 1} & 0_{T\times 1} \\ 0_{T\times 1} & k^2_{T+1} & 0_{T\times 1} \\ 0_{T\times 1} & 0_{T\times 1} & k^3_{T+1} \end{bmatrix} \tag{3.11}$$

Then, again by the multivariate normal regression lemma, we get the predictive distribution of s_{T+1} given all the data up to time T including M_T

$$s_{T+1}|S, M_{0:T} \sim N\left(k'_{T+1}K_{\mathcal{P}}^{-1}S,\ k^0_{T+1} + \Sigma_{\mathcal{P}}\Sigma'_{\mathcal{P}} - k'_{T+1}K_{\mathcal{P}}^{-1}k_{T+1}\right)$$

what translates into the corresponding predictive distribution of $\mathcal{P}_{T+1}$ as follows

$$\mathcal{P}_{T+1}|\mathcal{P}_T, S, M_{0:T} \sim N\left(\mu^{\mathbb{P}}_{\mathcal{P}} + \Phi^{\mathbb{P}}_{\mathcal{P}}\mathcal{P}_T + k'_{T+1}K_{\mathcal{P}}^{-1}S,\ k^0_{T+1} + \Sigma_{\mathcal{P}}\Sigma'_{\mathcal{P}} - k'_{T+1}K_{\mathcal{P}}^{-1}k_{T+1}\right)$$

3.3.3 Likelihood and Risk Price Restrictions

Statistical inference can be performed using the observations $Y = \{y_t, \mathcal{P}_t : t = 0, 1, \ldots, T\}$ and $M = \{M_t : t = 0, 1, \ldots, T-1\}$. The likelihood factorizes into two parts stemming from the $\mathbb{P}$ and $\mathbb{Q}$ respectively. For N observable factors, the joint likelihood (conditional on the initial point $\mathcal{P}_0$) can now be written as

$$\begin{aligned} f(Y|M, \theta, \widehat{\sigma}_K) = &\left\{\prod_{t=0}^{T} f^{\mathbb{Q}}(y_t|\mathcal{P}_t, k^{\mathbb{Q}}_\infty, g^{\mathbb{Q}}, \Sigma_{\mathcal{P}}, \sigma^2_e)\right\} \times \\ &\left\{\prod_{t=1}^{T} f^{\mathbb{P}}(\mathcal{P}_t|\mathcal{P}_{t-1}, M_{t-1}, k^{\mathbb{Q}}_\infty, g^{\mathbb{Q}}, \lambda_{0\mathcal{P}}, \lambda_{1\mathcal{P}}, \Sigma_{\mathcal{P}}, \ell_K, \widehat{\sigma}_K)\right\} \end{aligned} \tag{3.12}$$

where the $\mathbb{Q}$-likelihood components $f^{\mathbb{Q}}(\cdot)$ are given by (1.11) and capture the cross-sectional dynamics of the risk factors and the yields, whereas $\mathbb{P}$-likelihood components $f^{\mathbb{P}}(\cdot)$ are obtained from (3.14) below and capture the time-series dynamics of the observed risk factors. The parameter vector is set to $\theta = (\sigma^2_e, k^{\mathbb{Q}}_\infty, g^{\mathbb{Q}}, \lambda_{0\mathcal{P}}, \lambda_{1\mathcal{P}}, \Sigma_{\mathcal{P}}, \ell_K)$ and we tune σ_K in-sample to fix it out-of-sample at $\widehat{\sigma}_K$. Details of the tuning procedure are in Appendix 3.D.

For brevity of further exposition, we let $\lambda^{\mathcal{P}} = [\lambda_{0\mathcal{P}}, \lambda_{1\mathcal{P}}]$ and $\lambda = \lambda_{1\mathcal{P}}$. If all the entries in $\lambda^{\mathcal{P}}$ are free parameters we get the maximally flexible model (model M_0 in this chapter). Alternative specifications, with some of these entries set to zero, have been proposed in the literature. For details, see the associated discussion with related references in Section 1.1.1. Overall, in most models the set of unrestricted parameters is usually a subset of $\lambda^{\mathcal{P}}$. In this chapter we use restriction set with optimal model performance, in particular with respect to economic value, as evidenced in Section 1.5.3. It is also the same restriction set as in Section 2.2.3, for reasons mentioned therein. Namely, we only leave $\lambda_{1,2}$ unrestricted, as it is in model M_1, here and in the other chapters.

Consequently, the likelihood specification of (3.12) can now be restated as

$$\begin{aligned} f(Y|M,\theta,\widehat{\sigma}_K) = &\left\{\prod_{t=0}^{T} f^{\mathbb{Q}}(y_t|\mathcal{P}_t, k_\infty^{\mathbb{Q}}, g^{\mathbb{Q}}, \Sigma_{\mathcal{P}}, \sigma_e^2)\right\} \times \\ &\left\{\prod_{t=1}^{T} f^{\mathbb{P}}(\mathcal{P}_t|\mathcal{P}_{t-1}, M_{t-1}, k_\infty^{\mathbb{Q}}, g^{\mathbb{Q}}, \Sigma_{\mathcal{P}}, \ell_K, \lambda_{1,2}, \widehat{\sigma}_K)\right\} \end{aligned} \tag{3.13}$$

where $\theta = (\sigma_e^2, k_\infty^{\mathbb{Q}}, g^{\mathbb{Q}}, \Sigma_{\mathcal{P}}, \ell_K, \lambda_{1,2})$ is revised accordingly. To obtain $\mathbb{P}$-likelihood components $f^{\mathbb{P}}(\cdot)$ above, we resort to Gaussian Process formulation established so far, and exploit the associated marginal distribution of S in (3.8), together with (3.9), to eventually arrive at the standard log-likelihood representation of $\left\{\prod_{t=1}^{T} f^{\mathbb{P}}(\cdot)\right\}$ in (3.13), namely

$$\log\left\{\prod_{t=1}^{T} f^{\mathbb{P}}(\cdot)\right\} = -\frac{TN}{2}\log 2\pi - \frac{1}{2}\log|K_{\mathcal{P}}| - \frac{1}{2}\|K_{\mathcal{P}}^{-\frac{1}{2}}S\|^2 \tag{3.14}$$

where $|\cdot|$ is matrix determinant and $\|\cdot\|^2$ denotes Euclidean norm squared.

3.4 Sequential Estimation, Learning, and Forecasting

In this section we develop a sequential Monte Carlo (SMC) framework for Gaussian DTSMs with unspanned nonlinear macros. We draw from the work of Chopin (2002, 2004) (see also Del Moral et al. (2006)) and make the necessary adaptations to tailor the methodology to the data and models considered in this chapter. Furthermore, we extend the framework to allow for sequential Bayesian treatment of exogenous macroeconomic information both in a nonlinear and in a linear manner. The former using Gaussian Processes. Overall, the developed framework allows the efficient performance of tasks such as sequential parameter estimation and forecasting. We begin by providing the main skeleton of the scheme and then explain the details of its specific parts, such as the MCMC scheme including Gaussian Processes, and the framework for obtaining and evaluating the economic value of forecasts.

3.4.1 Sequential Framework with Gaussian Processes

Let $Y_{0:t} = (Y_0, Y_1 \ldots, Y_t)$ denote all bond related data available up to time t, such that $Y_{0:T} = Y$, and $M_{0:t-1} = (M_0, M_1 \ldots, M_{t-1})$ denote all macro data available up to time $t-1$, such that $M_{0:T-1} = M$. Similarly, the likelihood based on data up to time t is $f(Y_{0:t}|M_{0:t-1}, \theta, \widehat{\sigma}_K)$ and is defined in (3.12). Combined with a prior on the parameters $\pi(\theta)$, see the Appendix 3.A for details, it yields the corresponding posterior

$$\pi(\theta|Y_{0:t}, M_{0:t-1}, \widehat{\sigma}_K) = \frac{1}{m(Y_{0:t}|M_{0:t-1}, \widehat{\sigma}_K)} f(Y_{0:t}|M_{0:t-1}, \theta, \widehat{\sigma}_K)\pi(\theta) \tag{3.15}$$

where $m(Y_{0:t}|M_{0:t-1}, \widehat{\sigma}_K)$ is the model evidence based on data up to time t. Moreover, the posterior predictive distribution, which is the main tool for Bayesian forecasting,

is defined as

$$f(Y_{t+h}|Y_{0:t}, M_{0:t}, \widehat{\sigma}_K) = \int f(Y_{t+h}|Y_{0:t}, M_{0:t}, \theta, \widehat{\sigma}_K)\pi(\theta|Y_{0:t}, M_{0:t-1}, \widehat{\sigma}_K)d\theta \qquad (3.16)$$

where h is the prediction horizon. As explained further, in Section 3.4.2, in this chapter we focus on $h = 1$. Due to the Gaussian Process formulation, predictions depend computationally on all available data, what is reflected by conditioning on $Y_{0:t}$ and $M_{0:t}$ in the first term under the integral in (3.16). In theory, it is a potential issue as computational cost of deriving the predictive distribution grows fast. In particular, as new data arrives it becomes more costly to invert covariance matrix in (3.9). However, given that time series we consider are relatively short, computational time is far from prohibitive. Therefore, it is not a practical concern.

Note also that the predictive distribution in (3.16) incorporates parameter uncertainty by integrating θ out according to the posterior in (3.15). Usually, prediction is carried out by expectations with respect to (3.16), e.g. $E(Y_{t+h}|Y_{0:t}, M_{0:t}, \widehat{\sigma}_K)$ but, since (3.16) is usually not available in closed form, Monte Carlo can be applied in the presence of samples from $\pi(\theta|Y_{0:t}, M_{0:t-1}, \widehat{\sigma}_K)$. This process may facilitate various forecasting tasks, however the procedures can be quite intensive and in some cases not doable, see Section 1.3.1 for details.

Another approach that can also handle forecasting assessment tasks is to use sequential Monte Carlo (see, Chopin (2002) and Del Moral et al. (2006)) to sample from the sequence of distributions $\pi(\theta|Y_{0:t}, M_{0:t-1}, \widehat{\sigma}_K)$ for $t = 0, 1, \ldots, T$. A general overview of the Iterated Batch Importance Sampling (IBIS) scheme of Chopin (2002), see also Del Moral et al. (2006) for a more general framework, is provided in Algorithm 3.1. The degeneracy criterion is typically defined through the Effective Sample Size (ESS) which is specified in (1.21).

The IBIS algorithm provides a set of weighted θ samples, called also particles, that can be used to compute expectations with respect to the posterior distribu-

Algorithm 3.1 IBIS algorithm for Gaussian Affine Term Structure Models with unspanned nonlinear macros

Initialize N_θ particles by drawing independently $\theta_i \sim \pi(\theta)$ with importance weights $\omega_i = 1$, $i = 1, \ldots, N_\theta$. For $t, \ldots, T$ and each time for all i:

(a) Calculate the incremental weights from

$$u_t(\theta_i, Y_{0:t-1}, M_{0:t-1}) = f(Y_t | Y_{0:t-1}, M_{0:t-1}, \theta_i, \widehat{\sigma}_K)$$

where the Gaussian Process formulation makes them computationally dependant on all available data, what is reflected by conditioning on $Y_{0:t-1}$ and $M_{0:t-1}$.

(b) Update the importance weights ω_i to $\omega_i u_t(\theta_i, Y_{0:t-1}, M_{0:t-1})$.

(c) If some degeneracy criterion (e.g. ESS(ω)) is triggered, perform the following two sub-steps:

(i) Resampling: Sample with replacement N_θ times from the set of θ_is according to their weights ω_i. The weights are then reset to one.

(ii) Jittering: Replace θ_is with $\tilde{\theta}_i$s by running MCMC chains with each θ_i as input and $\tilde{\theta}_i$ as output. Set $\theta_i = \tilde{\theta}_i$.

tion, $E[g(\theta)|Y_{0:t}, M_{0:t-1}, \widehat{\sigma}_K]$, for all t using the estimator $\sum_i [\omega_i g(\theta_i)] / \sum_i \omega_i$. The same holds for expectations with respect to the posterior predictive distribution, $f(Y_{t+h}|Y_{0:t}, M_{0:t}, \widehat{\sigma}_K)$; the weighted θ samples can be conveniently transformed into weighted samples from $f(Y_{t+h}|Y_{0:t}, M_{0:t}, \widehat{\sigma}_K)$ by just applying $f(Y_{t+h}|Y_{0:t}, M_{0:t}, \theta, \widehat{\sigma}_K)$. A very useful by-product of the IBIS algorithm is the ability to compute

$$m(Y_{0:t}|M_{0:t-1}, \widehat{\sigma}_K) = f(Y_{0:t}|M_{0:t-1}, \widehat{\sigma}_K)$$

which is the criterion for conducting formal Bayesian model choice. Computing the following quantity in step (a) in Algorithm 3.1 yields a consistent and asymptotically normal estimator of $f(Y_t|Y_{0:t-1}, M_{0:t-1}, \widehat{\sigma}_K)$, namely

$$m_t = \frac{1}{\sum_{i=1}^{N_\theta} \omega_i} \sum_{i=1}^{N_\theta} \omega_i u_t(\theta_i, Y_{0:t-1}, M_{0:t-1})$$

Additional characteristics of IBIS are mentioned in Section 1.3.1.

So as to apply the IBIS output to models and data in this chapter, the following adaptations and extensions are carried out. Similar to Section 1.3.1, and pursuing motivation therein, we pool together the advantages of data tempering and adaptive tempering (Jasra et al., 2011; Schäfer and Chopin, 2013; Kantas et al., 2014) in a hybrid adaptive tempering scheme which we outline in Appendix 3.C. Because the MCMC algorithm used here is an extended version of Bauer (2018) and so it consists of independence samplers that are known to be unstable, we exploit the IBIS output and estimate posterior moments to arrive at independence sampler proposals; see Appendix 3.B for details, and Section 1.3.1 for additional rationale. Under such amended framework, and quite crucially in this chapter, we extend the framework presented in Sections 3.3.2 and 3.3.3 to handle Gaussian Processes sequentially. In the end, we apply IBIS output in the optimization of a model-driven dynamically rebalanced portfolio of bond excess returns and inspect its economic value, following Section 1.3.3.2.

In empirical work, we use $N_\theta = 2000$ particles, 5 MCMC steps when jittering, and with regards to minimum ESS we set $\alpha = 0.7$. The choice of 5 steps at the jittering stage is led by quite well mixing behaviour of the underlying MCMC. We monitored the correlation between particles before and after that stage to realize that performance was already reasonable with this number of iterations.

3.4.2 Assessing Predictive Performance and Economic Value

Leveraging the evaluation framework summarized in this section, we seek to understand whether macroeconomic information introduced into Gaussian ATSMs in a nonlinear manner using Gaussian Processes lets us predict excess returns better than when macros enter these models linearly. Furthermore, we attempt to explore whether such statistical predictability, if any, can be turned into consistent economic benefits for bond investors.

As explained further in Section 3.5.1, we use lagged macro data at monthly frequency. Given the limitations to availability of macroeconomic information going forward and its exogenous and not endogenous nature we assume in this chapter, which prevents us from associated forecasting further than a month ahead, we concentrate on 1-month prediction horizon. Thus, we refrain from making any assumptions about forward paths of the underlying macros. Consequently, returns we consider in our analysis are non-overlapping.

Along the lines of Section 1.3.3.1, yet with $h = 1$ in mind, we define the observed continuously compounded excess return of an n-year bond as the difference between the holding period return of the n-year bond and the h-period yield as

$$rx^n_{t,t+h} = -(n-h)y^{n-h}_{t+h} + ny^n_t - hy^h_t$$

If, instead of taking the observed one, we take the model-implied continuously compounded yield y^n_t, calculated according to (1.8), we arrive at the predicted excess return $\widetilde{rx}^n_{t,t+h}$ which becomes

$$\widetilde{rx}^n_{t,t+h} = A_{n-h,\mathcal{P}} - A_{n,\mathcal{P}} + A_{h,\mathcal{P}} + B'_{n-h,\mathcal{P}}\widetilde{\mathcal{P}}_{t+h} - (B_{n,\mathcal{P}} - B_{h,\mathcal{P}})'\mathcal{P}_t \tag{3.17}$$

where $\mathcal{P}_t$ is observed and $\widetilde{\mathcal{P}}_{t+h}$ is a prediction from the model. Our developed framework, see Algorithm 3.1, allows drawing from the predictive distribution of $(\widetilde{\mathcal{P}}_{t+h}, \widetilde{rx}^n_{t,t+h})$ based on all information available up to time t. More specifically, for each θ_i particle the $\mathbb{P}$-dynamics of $\mathcal{P}_t$ can be used to obtain a particle of $\widetilde{\mathcal{P}}_{t+h}$, which then can be transformed into a particle of $\widetilde{rx}^n_{t,t+h}$ via equation (3.17). Detailed steps are outlined in Algorithm 3.2.

To assess the predictive ability of the models considered, we compute the associated out-of-sample R^2 (R^2_{os}), due to Campbell and Thompson (2008). To measure the resulting economic value generated by each model, we consider a Bayesian investor with power utility preferences. Our Bayesian learner solves an asset allocation

Algorithm 3.2 Predictive distribution of excess returns for Gaussian Affine Term Structure Models with unspanned nonlinear macros

First, at time t, for some n and $h = 1$, using (ω_i, θ_i), $i = 1, ..., N_\theta$, from IBIS algorithm, iterate over i:

(a) Given θ_i, compute $A_{i_1,\mathcal{P}}$ and $B_{i_1,\mathcal{P}}$, for $i_1 \in \{1, n-1, n\}$, from (1.9) and (1.10).

(b) Given θ_i, obtain prediction of $\mathcal{P}_{t+1}$ by drawing from

$$\widetilde{\mathcal{P}}_{t+1}^{(i)} | \mathcal{P}_t, S, M_{0:t} \sim N\left(\mu_{\mathcal{P}}^{\mathbb{P}} + \Phi_{\mathcal{P}}^{\mathbb{P}} \mathcal{P}_t + k_{t+1}' K_{\mathcal{P}}^{-1} S,\ k_{t+1}^0 + \Sigma_{\mathcal{P}} \Sigma_{\mathcal{P}}' - k_{t+1}' K_{\mathcal{P}}^{-1} k_{t+1}\right)$$

where S, $K_{\mathcal{P}}$, k_{t+1}^0 and k_{t+1} are defined in (3.5), (3.9), (3.10) and (3.11), respectively, and except for S they all depend on $\widehat{\sigma}_K$.

(c) Compute particle prediction of $rx_{t,t+1}^n$ as

$$\widetilde{rx}_{t,t+1}^{n(i)} = A_{n-1,\mathcal{P}} - A_{n,\mathcal{P}} + A_{1,\mathcal{P}} + B_{n-1,\mathcal{P}}' \widetilde{\mathcal{P}}_{t+1}^{(i)} - (B_{n,\mathcal{P}} - B_{1,\mathcal{P}})' \mathcal{P}_t$$

Second, since $(\omega_i, \widetilde{\mathcal{P}}_{t+1}^{(i)}, \widetilde{rx}_{t,t+1}^{n(i)})$, $i = 1, ..., N_\theta$, is a particle approximation to predictive distribution of $(\mathcal{P}_{t+1}, rx_{t,t+1}^n)$, compute point prediction of $rx_{t,t+1}^n$ using particle weights ω_i as

$$\widetilde{rx}_{t,t+1}^n = \frac{1}{\sum_{i=1}^{N_\theta} \omega_i} \sum_{i=1}^{N_\theta} \omega_i \widetilde{rx}_{t,t+1}^{n(i)}$$

Third, repeat above two steps for different n.

problem getting optimal portfolio weights which we use to compute the CER as in Johannes et al. (2014) and Gargano et al. (2019). To that end, we follow the exposition in Section 2.3.2, where we refer the reader for further details. Similar to the latter chapter, we adopt the Expectations Hypothesis (EH) as initial empirical benchmark, however instead of then looking at the performance relative to model M_1 (see, Chapter 1), we compare the nonlinear with the corresponding linear models. Again, we only focus on a 1-month prediction horizon ($h = 1$).

3.5 Data and Models

In this section we discuss the data used in the chapter. We also explain what models we consider, as distinguished by different positions unspanned, nonlinear or linear, macros take in the given model. Eventually, we describe how we use risk premium

factor, rationale behind which is provided in Duffee (2011), to inspect whether information coming from macroeconomic variables we consider, in particular their components which are hidden from the yield curve, is *de facto* nonlinear in nature.

3.5.1 Yields and Macros

In terms of bond yields, we analyze the same data as these described in Section 2.4.1, splitting them in the same training (in-sample) and testing (out-of-sample) periods, thus for further details we refer the reader therein.

In terms of macroeconomic information about the US, we consider two variables which are well covered in the literature. These include, core[2] inflation (CPI) as in Cieslak and Povala (2015), as well as the three-month moving average of the Chicago Fed National Activity Index (GRO) from Joslin et al. (2014), which is a measure of current economic conditions. Together they offer parsimonious yet comprehensive enough, for our purposes, description of the US economy. Both macro variables are at a monthly frequency. They are also seasonally adjusted and revised. Unlike GRO, which is a smoothed index in levels, CPI is in percent changes from year ago. The underlying data period is such that it corresponds to this for yields, as discussed above, and also reflects the lagged character assumed for the macros in all models we consider in this chapter. Importantly, we standardize the macros only in the nonlinear case and, to that end, we use individual means and standard deviations calculated in-sample for the out-of-sample operations, to avoid look-ahead bias.

3.5.2 Models and Rationale Behind

In terms of models, we consider several alternatives when it comes to positions the unspanned, nonlinear or linear, macros take in the given model. More specifically, in

[2]In contrast to Cieslak and Povala (2015), who devise trend inflation by appropriately smoothing core inflation, we use the latter as is to arrive at our own, nonlinear function thereof.

the nonlinear case these are

$$GP_{ijk}(M) \quad i,j,k \in \{0,1\} \tag{3.18}$$

and for the linear model we replace GP with LM in this notation. In the above, M refers to a specific macroeconomic variable, CPI or GRO, introduced in the particular model. For example, for $i = 0$ and $j = k = 1$, that is $GP_{011}(M)$, index 011 means that macroeconomic impact is allowed in the second and third equation in (3.1), under the assumption that $N = 3$, $G = 2$ and $R = 1$. We only investigate a subset of the available alternative models in greater detail. Investigating and determining which macros, beyond the two included in our analysis, represent best match for specific PCs of the US yield curve, when interacting with the latter in a nonlinear or linear manner, is left for future research.

As result of risk price restrictions (on $\lambda_{\mathcal{P}}$) we adopt in this chapter from Section 1.4.2, without loss of generality for $i = j = k = 1$ in (3.18), we define the risk premium factor as

$$RP_t^V(M_{t-1}) = \lambda \mathcal{P}_t + v(M_{t-1}) = \begin{bmatrix} \lambda_{1,2}\mathcal{P}_{2,t} + v^{(1)}(M_{t-1}) \\ v^{(2)}(M_{t-1}) \\ v^{(3)}(M_{t-1}) \end{bmatrix} \tag{3.19}$$

where the first element is similar to Duffee (2011), however in our case its time-varying component, $RP_t = \lambda_{1,2}\mathcal{P}_{2,t}$, is a restricted version (only $\lambda_{1,2}$ is free) of the risk premium factor defined in Duffee's paper. The component $v^{(1)}(M_{t-1})$ is in that particular case the first element of the unspanned nonlinear macro $v(M_{t-1})$. If

instead $i = 0$ and $j = k = 1$ in (3.18), then

$$RP_t^V(M_{t-1}) = \begin{bmatrix} \lambda_{1,2}\mathcal{P}_{2,t} \\ v^{(2)}(M_{t-1}) \\ v^{(3)}(M_{t-1}) \end{bmatrix}$$

where the first element, which is related to level risk, equals RP_t and is also equivalent to the corresponding factor in model M_1 (see, Chapter 1) we adopt risk price restrictions from.

Following from Duffee (2011), the time-varying risk premium factor RP_t, which in the case therein is a linear combination of the state vector obtained from model shocks to yields, determines the compensation investors require to face fixed-income risk from t to $t+1$ and it contains all information relevant to predicting one-step-ahead, yet not h-step-ahead for $h > 1$, excess returns. We instead assume that investors require compensation for level risk stemming from changes to slope only. However, the term $v(M_{t-1})$ in (3.19), which is *a priori* unspanned and potentially nonlinear function (v) of the underlying macroeconomic variable (M), affects this compensation. It is important to understand if this impact is not distorting information relevant to predicting excess returns, already available from RP_t alone. One way to achieve this goal is to perform predictability and economic value exercises for the models considered and make appropriate comparisons. Thus, we do so in Section 3.6.

Different from Duffee (2011), we obtain the state vector from observed yields directly. Nevertheless, for $i, j, k \in \{0, 1\}$ in (3.18), we are able to similarly define the hidden part of the risk premium factor, as the part unspanned by $\mathcal{P}_t$, in the following way

$$\begin{aligned} \widetilde{RP}_t^V(M_{t-1}) &= RP_t^V(M_{t-1}) - E\left[RP_t^V(M_{t-1})|\mathcal{P}_t\right] \\ &= v(M_{t-1}) - E[v(M_{t-1})|\mathcal{P}_t] \end{aligned}$$

where $E[v(M_{t-1})|\mathcal{P}_t]$ is the projection of unspanned nonlinear macros on principal components obtained from observed yields, which is thus spanned by these PCs and as such not hidden from the yield curve. After Duffee (2011), the hidden $\widetilde{RP}_t^V(M_{t-1})$ can be estimated as residual from a regression of the former on the latter and expressed as

$$\widetilde{RP}_t^V(M_{t-1}) \equiv v(M_{t-1}) - a - b'\mathcal{P}_t \tag{3.20}$$

where we take $v(M_{t-1})$ as the mean from its posterior distribution and a and b are the underlying Ordinary Least Squares parameter estimates. We also note that

$$E[v(M_{t-1})|\mathcal{P}_t] = a + b'\mathcal{P}_t$$

and, by rearranging terms in (3.20), arrive at the following decomposition

$$v(M_{t-1}) \equiv E[v(M_{t-1})|\mathcal{P}_t] + \widetilde{RP}_t^V(M_{t-1}) \tag{3.21}$$

of unspanned nonlinear macros $v(M_{t-1})$ into two orthogonal components. While the first, $E[v(M_{t-1})|\mathcal{P}_t]$, which is an affine transformation of the PCs, is by definition not hidden from the yield curve, the second, $\widetilde{RP}_t^V(M_{t-1})$, duly is.

Then we can examine to what extent $v(M_{t-1})$ assumed nonlinear *a priori* is actually such *a posteriori*. More importantly, however, we investigate whether the same can be stated about its hidden component $\widetilde{RP}_t^V(M_{t-1})$. To that end, we separately regress these, as well as $E[v(M_{t-1})|\mathcal{P}_t]$, on the underlying macroeconomic variable M_{t-1} and inspect the resulting explanatory powers behind such linear associations, see Section 3.6.3 for related discussion. By doing so, we can determine whether there is any connection especially between the hidden component of the risk premium factor being nonlinear and the predictability of and economic benefits from nonlinear models we consider in this chapter. In what follows, we eventually

demonstrate that, when the latter outperform the linear ones on these two important fronts, this is indeed the case.

3.6 Empirical Results

The focus of this section is to present the main results on the statistical and economic performance of excess return forecasts resulting from the models we develop in this chapter. In particular, we assess the models based on different number of GP outputs involving various macros and, more importantly, their position in this part of the model which is governing the real-world dynamics of yields' PCs. We also investigate whether puzzling behavior between statistical predictability and out-of-sample economic gains for bond investors, as it is usually observed in case of yields-only models, still emerges when unspanned macroeconomic information is incorporated, especially in a nonlinear manner. To that end, we inspect the associated explanatory power, where applicable.

In order to better understand the role nonlinearities play when it comes to model performance, we also compare these nonlinear models with their linear counterparts, and as such provide statistical evidence that they matter. In what follows, we refrain from evaluating the relevance of particular macroeconomic variables, let alone the importance of their position in the $\mathbb{P}$-dynamics, for these performance results. We consider that as an interesting path for future research, especially when wider spectrum of information about the economy is taken into account.

Finally, we decompose the risk premium factor into a part which is hidden from the yield curve and a part which is spanned by the underlying principal components. Then, leveraging decomposition in (3.21), we individually regress elements thereof on the corresponding macroeconomic variable to show that, when especially the hidden component of the risk premium factor is nonlinear, using DTSMs with Gaussian

Processes, as in this chapter, is more beneficial in terms of predictability and economic value than relying on the linear benchmarks.

3.6.1 Uncovering Nonlinearities

The data contain an important economic event and thus it is interesting to have a look at the plots of individual elements in $v(M_{t-1})$ against M_{t-1}, for CPI and GRO, to see how it affects our analysis. These plots are generated from the IBIS output, as described in Section 3.4.1 and further in Appendix 3.C. In particular, data cover the financial crisis of 2008-2009. For related discussion, yet focused on principal components as opposed to nonlinear macros, see Section 1.5.1. In what follows, we only concentrate on models $GP_{110}(CPI)$ and $GP_{011}(GRO)$, which we also focus on in Section 3.6.3, for specific reasons mentioned therein.

By inspecting the plots for model $GP_{110}(CPI)$, which are shown in Figure 3.1, we observe that CPI interacts with the yield curve in a nonlinear rather than linear manner. This is particularly the case when we look at panels on the right hand side of this figure. In there, distributions of $v^{(1)}(CPI_{t-1})$ and $v^{(2)}(CPI_{t-1})$ plotted against CPI_{t-1} are estimated using the entire sample of data, including the recession period. Comparing these two panels with their counterparts on the left hand side in the same figure, which are based on data from the training period only and thus preclude the recession, we notice that the nonlinear relationship between CPI and the yield curve strengthens in time.

Next, we have a careful look at GRO, where things are quite different. In Figure 3.2, with associated plots for model $GP_{011}(GRO)$, we observe that GRO interacts with the yield curve more in a linear than in a nonlinear manner. It is particularly evident when inspecting the top panel in this figure, which shows the distribution of $v^{(2)}(GRO_{t-1})$ plotted against GRO_{t-1}. The vast majority of data points in the training period, that is on the left hand side, are located in the part of the graph where the association is linear. The remaining data points constitute the outliers. On

the right hand side, that is when the entire data sample is used, the part of the graph which is linear is accompanied by even more outliers. With hindsight, economic activity is heavily affected by outliers due to recession, what can also be noticed by comparing the horizontal axes in these plots between the left and the right hand side panels. Although the functional association of lagged GRO with the yield curve shown in the top panel deteriorates only slightly in time, this presented in the bottom panel for $v^{(3)}(GRO_{t-1})$ dilutes completely when the entire data sample is involved. Likely, the latter changes in the post-recession period, while the model assumes it is the same, what makes the forecasting exercise we conduct quite challenging but at the same time more realistic.

3.6.2 Bond Return Predictability and Economic Performance

In this section we present and discuss the results on bond return predictability and the associated economic performance of our nonlinear, and also linear, models, as it is showcased in Section 3.4.2. Where insightful, we also link these outcomes to the explanatory power behind the observed (linear) and the estimated (nonlinear) macros, where selected results are shown in Tables 3.1 and 3.2.

3.6.2.1 Predictive Performance

In what follows, we focus on statistical performance. Table 3.3 reports out-of sample R^2 values for all linear (including M_0, the maximally flexible model, and M_1, see Chapter 1) and nonlinear models across different bond maturities and at 1-month prediction horizon. The latter means that we only consider non-overlapping excess bond returns. The in-sample (training) period is from January 1985 to the end of 2007 and the out-of-sample (testing) period is from January 2008 to the end of 2018. The latter practically begins with the 2008-2009 financial crisis. Overall, results indicate, as expected and in line with the existing literature (Duffee, 2011; Joslin et al., 2014; Fulop et al., 2019), that models which incorporate macroeconomic

information, irrespective of whether in a linear or nonlinear manner, predict well and perform better out-of-sample compared to the EH benchmark, as well as when set against yields-only models. This is confirmed by predictive R^2_{os} that are exclusively positive and statistically significant for CPI, across all maturities, and mostly positive and statistically significant for GRO, especially for the short and medium term maturities. Across maturities, they range from 0.02 to 0.06 for CPI, whereas for GRO the positive and statistically significant R^2_{os} are between 0.02 to 0.05.

Generally, models with macroeconomic variables perform better, in terms of out-of-sample statistical performance, than the yields-only models we consider, namely M_0 and M_1. However, there are four notable exceptions in case of GRO, where R^2_{os} are practically close to zero. These are linear and nonlinear models with indices 111 and 110, which either perform as unsatisfactorily as M_1 (GP_{111}) or worse, and as such similar to M_0 (LM_{111}, LM_{110} and GP_{110}). What they interestingly have in common is that, in these models macroeconomic information interacts with the yield curve through the observed level, meaning that it is included in the equation governing the real-world dynamics of the first PC. This deterioration in statistical performance translates further into correspondingly poor economic value results, especially for linear models, what we discuss in the next section.

Selected results on explanatory power for the nonlinear models with GRO specifically listed above, which are presented in Table 3.2 (see fourth panel) for model GP_{111}, reveal that, contrary to the out-of-sample statistical performance results, there are statistically significant benefits from including $v^{(1)}(GRO_{t-1})$ when explaining the variability of excess bond returns. It is particularly evident when compared to what we observe for the corresponding nonlinear function of lagged CPI. For GRO, such gains decrease at longer maturities and range from 2.91% at 2-year, through 2.17% at 5-year, to 0.34% at 10-year maturity. For CPI, these gains are considerably smaller and amount to about 0.60% throughout 2- to 5-year maturities. Related results in Table 3.1 confirm that similar is true in the linear case. Namely, explanatory power

gains for GRO, albeit smaller than in the nonlinear case, range from 1.95% at 2-year to 0.56% at 4-year maturity, yet occur only at the short end of the yield curve. Those for CPI are in this case practically negligible. Corresponding results on explanatory power for model GP_{110} are quantitatively and qualitatively alike, thus not shown. The latter also applies to such results for all models with indices 011, 010 and 001.

One possible rationale behind the above argument is that GRO, the economic activity, varies more profoundly than CPI, the prices, in the aftermath of the 2008-2009 financial crisis, that is throughout the testing sample. Let us compare the horizontal axes on the left with those on the right hand side in the plots shown in Figures 3.1 and 3.2. In Figure 3.1, CPI_{t-1} is approximately between -2 and 2.5 over 1985-2007 and from -2.5 to 2.5 over 1985-2017. In Figure 3.2, GRO_{t-1} is approximately between -4 and 2 over 1985-2007 and from -10 to 2 over 1985-2017. Importantly, both variables are standardized over 1985-2017 based on their respective means and standard deviations computed over 1985-2007. Thus, throughout 1985-2017 the variation of CPI hardly moves relative to what happens for GRO, which becomes evidently negatively skewed as result of such adverse economic conditions. On the other hand, these abnormal changes in GRO serve explaining the variability of excess bond returns in these turbulent times well, what is also a sign of possible overfitting.

Nevertheless, if the underlying index does begin with 1 not 0 then these nonlinear models with GRO we specifically mention above, with indices 111 and 110, tend to significantly improve out-of-sample statistical performance results when compared to linear benchmarks, see Table 3.4. As such, they reduce the risk associated with out-of-sample predictions based on macroeconomic variables which become negatively skewed in times of economic/financial crises. Interestingly, similar improvements occur for CPI in the specific case of nonlinear model with index 110.

3.6.2.2 Economic Value

In what comes next, we concentrate on model performance in terms of economic value. Table 3.5 reports results for the annualized CERs, calculated using out-of-sample forecasts of bond excess returns across maturities and at 1-month prediction horizon. The in-sample and out-of-sample periods are the same as in the case of predictive performance. The coefficient of relative risk aversion is chosen to be $\gamma = 3$ and we do not impose any portfolio constraints[3]. We find that, in most cases, corresponding CERs are positive and non-negligible, indicating that DTSMs with unspanned macroeconomic information, introduced in the model in a linear or nonlinear way, not only perform well and in a statistically significant manner when it comes to out-of-sample predictability but also generate economic gains for bond investors relative to the EH benchmark.

For the linear case, this is in line with the existing literature, as numerous studies show (Duffee, 2011; Joslin et al., 2014; Fulop et al., 2019), whereas in case of nonlinear models only recently Bianchi et al. (2021), albeit in a regression and not in a no-arbitrage setting, show that forecasts based on neural networks fed with macroeconomic and yield information jointly, translate into economic gains that are larger than those obtained using yields alone. In line with all that, yields-only DTSMs we are concerned with in this chapter, that is M_0 and M_1, perform evidently worse in terms of economic value than models with macroeconomic information (see, Table 3.5). This is also consistent with their relatively inferior out-of-sample statistical performance, which we discuss in the previous section.

In case of CPI results show clear evidence of positive out-of-sample economic benefits for bond investors from introducing information about prices in the model. This is the case across most maturities (especially at 2-, 3- and 7-year) and in particular when CPI is introduced in a nonlinear manner. CERs for linear models

[3]These non-conservative choices are motivated by early exploratory character of the analysis conducted herein. The goal is to find economic value first and in future research examine if it can be exploited when stricter conditions apply.

with CPI are not even comparably as statistically significant as the former. For example, in case of GP_{010} they are statistically significant at all maturities and range from 3.67% (at 5-year) to 4.94% (at 7-year). In relation to GRO such evidence is less pronounced and results are only statistically significant for selected maturities and for linear and nonlinear models with indices 011 and 010 (at 3-, 4- and 7-year, in both cases), as well as 001 (all maturities but 5-year). The outcome is most positive in this last case and, when statistically significant, CERs for linear and nonlinear models with index 001 range from 3.37% (at 3-year, for the linear setup) to 4.55% (at 7-year, for the nonlinear setup).

When comparing nonlinear models with their linear counterparts, only in case of CPI the former perform significantly better than the latter (see, Table 3.6). For GP_{110}, GP_{010} and GP_{001} we observe positive and non-negligible CER values across most maturities, in particular at 2- to 5-year. If statistically significant, they span between 1.88% and 3.78% relative to the linear benchmarks. For GRO, where results are overall worse compared to those for CPI, we are not able to identify meaningful, statistically significant cases where nonlinear models outperform their linear benchmarks in terms for economic value. In the next section we provide some rationale why and when it might be so.

Overall, out-of-sample economic value results are consistent with those for predictive performance which we discuss in the previous section. As such, based on results presented in this chapter the puzzling behavior between statistical predictability and out-of-sample economic gains for bond investors cannot be unequivocally confirmed for DTSMs utilizing macroeconomic information, especially in a nonlinear manner. It is also important to note that considerable and statistically significant explanatory power gains (see, Tables 3.1 and 3.2), especially from macroeconomic variables which are prone to outliers like GRO, not necessarily translate into corresponding economic performance out-of-sample. Even to the contrary, they may lead to significant losses, especially when economic conditions are changing. It is particularly the case for

linear models with indices 111 and 110 (see, Table 3.5). However, nonlinear models may help cut such losses, e.g. by handling outliers better. In some cases, like for CPI (GP_{110}, GP_{010} and GP_{001}), they perform consistently better out-of-sample than their linear counterparts in terms of economic value (see, Table 3.6).

3.6.3 Benefiting from Hidden Nonlinearities

In this section we shed some light on why and when certain nonlinear models we consider in this chapter (with macroeconomic variables like CPI) outperform their linear counterparts in terms of economic value, whereas other (with macros such as GRO) do not. To that end, we resort to the results obtained from risk premium factor decomposition, as described in Section 3.5.2, which are presented in Tables 3.7 and 3.8. For pragmatic reasons, related to the tuning procedure (see, Appendix 3.D) which is not always optimal and as such affects the way certain aspects of this decomposition can be effectively demonstrated graphically, in what follows we focus on two particular models. Namely, $GP_{110}(CPI)$ and $GP_{011}(GRO)$.

We choose these two models for several reasons. First, both linear and nonlinear models with CPI and index 110 perform well in terms of out-of-sample predictability, however the nonlinear model performs significantly better across maturities than its linear counterpart - with R^2_{os} ranging from 0.04 to 0.06 against 0.02 and 0.03 for the linear model (see, Table 3.3). Second, both linear and nonlinear models with GRO and index 011 perform fairly well across most maturities when it comes to out-of-sample predictability - with statistically significant R^2_{os} ranging from 0.02 and 0.05 for both cases. Third, while nonlinear model with CPI does significantly better in terms of out-of-sample economic performance than its linear counterpart - with relative CER ranging from 2.55% to 3.35% across maturities when statistically significant - it is almost entirely the opposite case for the GRO. There, relative CER are not statistically significant and vary between -3.05% and 1.59% (see, Table 3.6). This is precisely what we are after to facilitate the argumentation which follows.

Knowing what we summarised above, we now resort to decomposition in (3.20) of nonlinear macro $v(M_{t-1})$ into a part about which complete information is contained in the yield curve, that is $E[v(M_{t-1})|\mathcal{P}_t]$, and a part which is entirely hidden, that is $\widetilde{RP}_t^V(M_{t-1})$. We regress individually each of these components on the underlying macro M_{t-1} in order to determine the degree to which the resulting relationship is linear, as measured by the associated adjusted R^2. Tables 3.7 and 3.8 contain such results for $GP_{110}(CPI)$ and $GP_{011}(GRO)$, respectively. They include related results for all the other models as well, however our focus on these two remains unchanged.

Beginning with CPI, in the third panel of Table 3.7 we notice that for model $GP_{110}(CPI)$ both $v^{(1)}(CPI_{t-1})$ and its part which is not hidden from the yield curve are to large extent linear functions of CPI_{t-1}, as measured by $\bar{R}^2$ that are equal to 0.73 and 0.70, respectively. At the same time the hidden part, that is $\widetilde{RP}_{1,t}^V(CPI_{t-1})$, is hardly in a linear relationship with CPI_{t-1}, as indicated by a very low value of $\bar{R}^2$ equal to 0.09 only. It is similar in case of $v^{(2)}(CPI_{t-1})$ where these metrics amount to 0.32, 0.67 and practically 0.00, in that order. Hidden parts of $v(CPI_{t-1})$ in model $GP_{110}(CPI)$ are thus nonlinear functions of the underlying macroeconomic variable, namely CPI_{t-1}. To realize that it is indeed so, it is sufficient to have a brief look at Figure 3.3. Therein, we see clearly that functional associations between these hidden parts and lagged CPI are highly nonlinear, stable or become even more pronounced (left versus right hand side). At this point it is safe to state that in this particular case there is a fair reason behind superiority of the nonlinear model, which is implemented using GPs in this chapter, over its linear counterpart. Namely, unspanned information coming from CPI, which is hidden from the yield curve yet it affects its $\mathbb{P}$-dynamics, is evidently nonlinear in nature.

Ending with GRO, in the first panel of Table 3.8 we observe that, for model $GP_{011}(GRO)$, both $v^{(2)}(GRO_{t-1})$ and its part which is not hidden from the yield curve are to considerable and to moderate extent, respectively, linear functions of GRO_{t-1}, as measured by $\bar{R}^2$ that are equal to 0.50 and 0.17. What is however

different from the case of CPI is that the hidden part, namely $\widetilde{RP}^{V}_{2,t}(GRO_{t-1})$, is to larger extent linear in GRO_{t-1} than the part which is not hidden from the yield curve, that is $E[v^{(2)}(GRO_{t-1})|\mathcal{P}_t]$, as indicated by $\bar{R}^2$ equal to 0.37 for the former. In case of $v^{(3)}(GRO_{t-1})$, these metrics amount to 0.00, 0.15 and 0.03, in that order. However, for reasons mentioned in the last paragraph of Section 3.6.1 and in relation to the tuning procedure (see, Appendix 3.D) which is not always optimal, we also look at the corresponding numbers for model $GP_{001}(GRO_{t-1})$ in the first panel of Table 3.7. They amount to 0.39, 0.11 and 0.33, respectively. In this case, the focus of the tuning procedure lies solely on the curvature, not slope and curvature together as in the former model. These latter results correspond more closely to what we observe for $v^{(2)}(GRO_{t-1})$ in model $GP_{011}(GRO_{t-1})$.

To consolidate the discussion, a careful look at Figure 3.4 is required. In the top panel, we notice immediately that what adjusted R^2 are telling us about $v^{(2)}(GRO_{t-1})$ in the model of interest is indeed so. Namely, the association between its hidden part and the lagged GRO is visibly more linear than nonlinear. The picture is similar, albeit less pronounced, to this in the corresponding panel in Figure 3.2, which we discuss in the last paragraph of Section 3.6.1. On both sides the majority of data points are located in the part of the graph where the association is linear, whereas the remaining data points are the outliers. What happens to $\widetilde{RP}^{V}_{3,t}(GRO_{t-1})$ in the bottom panel can also be linked to related arguments in the same section. It is thus not surprising that in this case the nonlinear model leads to inferior economic value results, let alone adds such a value, in comparison to the linear benchmark (see, Table 3.6). In this case and on such front it is evidently hard to beat the linear, more parsimonious model based merely on outliers.

3.7 Conclusions

We propose a novel methodological framework which combines Bayesian sequential inference with machine learning techniques, in particular Gaussian Processes. It allows us to incorporate unspanned macroeconomic information into Dynamic Term Structure Models in a potentially nonlinear manner. Sequential setup we develop successfully handles real-time adjustments to parameters governing such asymmetric/nonlinear associations. The methodology takes into account parameter uncertainty and provides entire predictive distribution of bond returns, allowing investors to review their beliefs when new information arrives and thus informing their asset allocation in an online manner. The framework is then tested against the Expectations Hypothesis, as well as against the linear benchmark, in a comprehensive out-of-sample exercise involving statistical predictability and economic value, where we assume availability of information about prices and economic activity. To that end, we scrutinize nonlinear models by developing a version of the macro-finance DTSM proposed by Joslin et al. (2014) that incorporates unspanned macroeconomic variables in a linear manner. However, to align with our nonlinear setup we consider them exogenous.

Empirical results confirm that in such an exercise the models with unspanned macroeconomic information, irrespective if it is introduced in a linear or nonlinear manner, perform overall better than the models which are yields-only by construction. Specifically, they also reveal that nonlinear models provide a competitive edge over linear models in cases when *a-priori* possibly nonlinear functions of macroeconomic variables admit nonlinear character *a-posteriori* in the parts which are hidden from the yield curve. In this chapter it is the case for the prices and not for the economic activity. To demonstrate that, we apply the risk premium factor decomposition from Duffee (2011).

However, results in this chapter also indicate that for certain models where macroeconomic information directly affects the yield curve level, using nonlinear

formulation might not necessarily lead to superb predictability and economic value gains but it might still mitigate substantial mistakes on those fronts. In particular, it is the case for models with macroeconomic variables which in abnormal times, such as the 2008-2009 financial crisis or the more recent COVID-19 recession, are subject to outliers like the economic activity. According to our results, Gaussian Processes handle such cases quite well.

Thinking ahead, we acknowledge the limitations of our study where we only consider a set of macroeconomic variables limited to two and Gaussian Processes we apply are of a single-input type and based on one commonly applied nonlinear kernel that we select to use. Verifying what we infer in this chapter on a broader set of macroeconomic information, while also using other nonlinear kernels such as Matern kernel or combined kernels, is a possible way going forward. Letting macroeconomic variables interact within a multi-input Gaussian Process framework, for example using ARD (Automatic Relevance Determination) kernels (Rasmussen and Williams, 2006), is also a potentially interesting and straightforward research avenue to pursue. Another intriguing, future direction is to incorporate spanned instead of unspanned nonlinear macros and embark on policy oriented applications, what such an extension facilitates.

Appendix 3.A Specification of Priors

In what comes next, we succinctly explain the prior distributions that were not specified in the main body of the chapter. For parameters in $\Sigma_{\mathcal{P}}$, $g^{\mathbb{Q}}$, $k_{\infty}^{\mathbb{Q}}$, σ_e^2 and $\lambda^{\mathcal{P}}$, or effectively $\lambda_{1,2}$, priors are constructed in the same manner as in the related Appendix 2.A.

The only exception are parameters in ℓ_K, which have range restricted to positive values. We thus transform them first, so that they have unrestricted range. Specifi-

cally, we work in log-scale of ℓ_K. Next, independent normal distributions with zero means and large variances are assigned to each of its components.

Appendix 3.B Markov Chain Monte Carlo Scheme

Following from (3.13) and (3.15), and given a prior $\pi(\theta)$ as described in Appendix 3.A, the posterior can be written in a more detailed manner as

$$\begin{aligned}\pi(\theta|Y, M, \widehat{\sigma}_K) = &\left\{\prod_{t=0}^{T} f^{\mathbb{Q}}(y_t|\mathcal{P}_t, k_\infty^{\mathbb{Q}}, g^{\mathbb{Q}}, \Sigma_{\mathcal{P}}, \sigma_e^2)\right\} \times \\ &\left\{\prod_{t=1}^{T} f^{\mathbb{P}}(\mathcal{P}_t|\mathcal{P}_{t-1}, M_{t-1}, k_\infty^{\mathbb{Q}}, g^{\mathbb{Q}}, \Sigma_{\mathcal{P}}, \ell_K, \lambda_{1,2}, \widehat{\sigma}_K)\right\} \times \pi(\theta)\end{aligned}$$

As the above posterior is not available in closed form, methods such as MCMC can be used to draw samples from it using Monte Carlo. Yet, the MCMC output is not assured to lead to precise Monte Carlo calculations since the corresponding Markov chain may have poor mixing and convergence properties, what leads to highly autocorrelated samples.

It is thus necessary to devise a suitable MCMC algorithm that does not exhibit such unfavourable traits. For further details regarding its construction, see the related Appendix 1.B. Such an MCMC scheme is shown in Algorithm 3.3.

Algorithm 3.3 MCMC scheme for Gaussian Affine Term Structure Models with unspanned nonlinear macros

Initialize all values of θ. Then at each iteration of the algorithm:

(a) Update σ_e^2 from its full conditional distribution that can be shown to be an Inverse Gamma distribution with parameters $\tilde{\alpha}/2$ and $\tilde{\beta}/2$, such that $\tilde{\alpha}$ is $\alpha + T(J-R)$ and $\tilde{\beta}$ is $\beta + \sum_{t=0}^{T} \|\hat{e}_t\|^2$, where $\alpha = \beta = 0$, since prior is assumed diffuse, $\hat{e}_t$ is a time-t residual from (1.11), and $\|\cdot\|^2$ is Euclidean norm squared.

(b) Update $\Sigma_{\mathcal{P}}$ using an independence sampler based on the MLE and the Hessian obtained before running the MCMC, using multivariate t-distribution with 5 degrees of freedom as proposal distribution.

(c) Update $(k_\infty^{\mathbb{Q}}, g^{\mathbb{Q}})$ in a similar manner to (b).

(d) Update $(\ell_K, \lambda_{1,2})$ in a similar manner to (b).

Appendix 3.C Adaptive Tempering

The aim of adaptive tempering is to smooth peaked likelihoods. For additional information, see the corresponding Appendix 1.C. Implementation of the IBIS scheme with hybrid adaptive tempering steps is outlined in Algorithm 3.4. It is important to note that, unlike it is presented in Algorithm 3.1 for the general IBIS case, in the specific case we are dealing here with we initialize the particles by drawing from the posterior $\pi(\theta|Y_{0:t-1}, M_{0:t-2}, \widehat{\sigma}_K)$ instead of the prior $\pi(\theta)$. This is done in-sample based on training data, as detailed in Section 3.5.1.

Although it is straightforward to implement step 4(b)iv in Algorithm 3.4 for an independence sampler, adjustments are necessary for a full Gibbs step. It is especially the case for σ_e^2 in step (a) in Algorithm 3.3, see Appendix 3.B. Implementation details are the same as in the corresponding appendix to Chapter 1 we mention above.

Appendix 3.D Tuning the Gaussian Process

Since we view σ_K as tuning parameter we tune it in-sample and fix out-of-sample at $\widehat{\sigma}_K$. Details about the underlying data are in Section 3.5.1. To that end, in a manner similar to this in the corresponding Appendix 2.D, we follow a multi-step process which is entirely based on in-sample data. First, as in the standard case in Section 1.2.1, we estimate by maximum likelihood a yields-only DTSM where $N = 3$ and, out of $\lambda_{\mathcal{P}}$, only $\lambda_{1,2}$ is left unrestricted to match the risk price restrictions we adopt in this chapter. Resulting MLEs let us then obtain $\hat{s}_t$, $t = 1, ..., \widetilde{T}$, where $\widetilde{T}$ refers to in-sample period, from (3.3).

Second, we formulate an amended version of the likelihood in (3.13), using only its $\mathbb{P}$-likelihood components $f^{\mathbb{P}}(\cdot)$ modified in the following way

$$\tilde{f}(\widetilde{Y}|\widetilde{M}, \theta, \hat{k}_{\infty}^{\mathbb{Q}}, \hat{g}^{\mathbb{Q}}, \widehat{\Sigma}_{\mathcal{P}}, \hat{\lambda}_{1,2}) = \left\{ \prod_{t=1}^{\widetilde{T}} \tilde{f}^{\mathbb{P}}(\mathcal{P}_t|\mathcal{P}_{t-1}, M_{t-1}, \hat{k}_{\infty}^{\mathbb{Q}}, \hat{g}^{\mathbb{Q}}, \widehat{\Sigma}_{\mathcal{P}}, \ell_K, \hat{\lambda}_{1,2}, c) \right\} \tag{3.22}$$

Algorithm 3.4 IBIS algorithm with hybrid adaptive tempering for Gaussian Affine Term Structure Models with unspanned nonlinear macros

Initialize N_θ particles by drawing independently $\theta_i \sim \pi(\theta|Y_{0:t-1}, M_{0:t-2}, \widehat{\sigma}_K)$ with importance weights $\omega_i = 1$, $i = 1, \ldots, N_\theta$. For $t, \ldots, T$ and each time for all i:

1 *Set $\omega_i' = \omega_i$.*

2 *Calculate the incremental weights from*

$$u_t(\theta_i, Y_{0:t-1}, M_{0:t-1}) = f(Y_t|Y_{0:t-1}, M_{0:t-1}, \theta_i, \widehat{\sigma}_K)$$

3 *Update the importance weights ω_i to $\omega_i u_t(\theta_i, Y_{0:t-1}, M_{0:t-1})$.*

4 *If degeneracy criterion ESS(ω) is triggered, perform the following sub-steps:*

(a) *Set $\phi = 0$ and $\phi' = 0$.*

(b) *While $\phi < 1$*

i. *If degeneracy criterion ESS(ω'') is not triggered, where $\omega_i'' = \omega_i'[u_t(\theta_i, Y_{0:t-1}, M_{0:t-1})]^{1-\phi'}$, set $\phi = 1$, otherwise find $\phi \in [\phi', 1]$ such that ESS(ω''') is greater than or equal to the trigger, where $\omega_i''' = \omega_i'[u_t(\theta_i, Y_{0:t-1}, M_{0:t-1})]^{\phi-\phi'}$, for example using bisection method, see Kantas et al. (2014).*

ii. *Update the importance weights ω_i to $\omega_i'[u_t(\theta_i, Y_{0:t-1}, M_{0:t-1})]^{\phi-\phi'}$.*

iii. *Resample: Sample with replacement N_θ times from the set of θ_is according to their weights ω_i. The weights are then reset to one.*

iv. *Jitter: Replace θ_is with $\tilde{\theta}_i$s by running MCMC chains with each θ_i as input and $\tilde{\theta}_i$ as output, using likelihood given by $f(Y_{0:t-1}|M_{0:t-2}, \theta_i, \widehat{\sigma}_K)[f(Y_t|M_{t-1}, \theta_i, \widehat{\sigma}_K)]^\phi$. Set $\theta_i = \tilde{\theta}_i$.*

v. *Calculate the incremental weights from*

$$u_t(\theta_i) = f(Y_t|Y_{0:t-1}, M_{0:t-1}, \theta_i, \widehat{\sigma}_K)$$

vi. *Set $\omega_i' = \omega_i$ and $\phi' = \phi$.*

---

where $\widetilde{Y}$ and $\widetilde{M}$ refer to in-sample data, $\hat{k}_\infty^{\mathbb{Q}}$, $\hat{g}^{\mathbb{Q}}$, $\widehat{\Sigma}_{\mathcal{P}}$ and $\hat{\lambda}_{1,2}$ are the MLEs from the first step, and $\theta = (\ell_K, c)$, with scalar $c > 0$, consists of parameters we estimate by maximum likelihood next. However, before that we parametrize σ_K in (3.22) as

$$\sigma_K = c\sqrt{diagv\left[Var(\hat{s})\right]} \tag{3.23}$$

where $diagv\left[Var(\hat{s})\right]$ is a 3×1 (in accordance with exposition in this chapter where we choose $G = 3$) vector including diagonal elements of the covariance matrix $Var(\hat{s})$ for $\hat{s} = \left[\hat{s}_1, \ldots, \hat{s}_{\widetilde{T}}\right]$, and $\hat{s}_t$, $t = 1, \ldots, \widetilde{T}$, are practically as in the first step.

Third, as in Section 3.3.3 we proceed with the the log-likelihood representation of (3.22), similar to (3.14), which we maximize. Consequently, it lets us fix σ_K out-of-sample at $\widehat{\sigma}_K$, which we calculate from (3.23) with $\hat{c}$ being the MLE of c from this last step and $Var(\hat{s})$ remains unchanged.

Appendix 3.E Linear Model with Macros

It is straightforward to extend the estimation framework of Bauer (2018), and consequently this in Chapter 1, to incorporate in a linear manner unspanned macros which are assumed exogenous. It is only the way we handle the $\mathbb{P}$-dynamics of $\mathcal{P}_t$ in (3.1) what needs to be adjusted to

$$\mathcal{P}_t = \mu_{\mathcal{P}}^{\mathbb{P}} + \Phi_{\mathcal{P}}^{\mathbb{P}} \mathcal{P}_{t-1} + \Phi_{\mathcal{P}M}^{\mathbb{P}} M_{t-1} + \Sigma_{\mathcal{P}} \varepsilon_t^{\mathcal{P}} \tag{3.24}$$

where $\Phi_{\mathcal{P}M}^{\mathbb{P}}$ is $(N \times R)$ matrix, which represents the feedback from M_{t-1} to $\mathcal{P}_t$.

Following sections C.1 and C.2 in Online Appendix to Bauer (2018) and details from Lütkepohl (2005), to consider $\Phi_{\mathcal{P}M}^{\mathbb{P}}$ coefficients next to M_{t-1} in (3.24) in estimation, it suffices to tackle them jointly with $\lambda_{\mathcal{P}}$, or in our case with $\lambda_{1,2}$ only, to match the risk price restrictions chosen for the nonlinear case in Section 3.3.3. Adopting notation from Online Appendix to Bauer (2018) to ours where necessary, we can rewrite (3.24) in vector form as

$$X = BZ + U$$

where $X = [\mathcal{P}_1, \ldots, \mathcal{P}_T]$, $U = [u_1, \ldots, u_T]$, with adapted $u_t = \Sigma_{\mathcal{P}} \varepsilon_t^{\mathcal{P}}$, $t = 1, \ldots, T$, $Z = \left[Z_0, \ldots, Z_{(T-1)}\right]$, with modified $Z_t = [1, \mathcal{P}_t', M_t']'$, $t = 0, \ldots, T-1$, and amended $B = \left(\mu_{\mathcal{P}}^{\mathbb{P}}, \Phi_{\mathcal{P}}^{\mathbb{P}}, \Phi_{\mathcal{P}M}^{\mathbb{P}}\right)$.

Then linear constraints after Bauer (2018) are following

$$\beta = vec(B) = \lambda + r = S\lambda_\gamma + r$$

where in our case S is a $[N(N+R+1)] \times [NR+1]$ selection matrix of zeros and ones, λ_γ is a $[NR+1] \times 1$ vector with $\lambda_{1,2}$ and those elements of $\Phi_{\mathcal{P}M}^{\mathbb{P}}$ we decide to leave unrestricted. For example, if our goal it to compare a linear model with a corresponding nonlinear case where $G = 2$ and there is no GP included in the first equation of (3.2), to allow for a meaningful comparison of results we would restrict the first row in $\Phi_{\mathcal{P}M}^{\mathbb{P}}$ to zeros, in a similar fashion to restricting risk prices in $\lambda_{\mathcal{P}}$. Finally, $r = vec\left[\mu_{\mathcal{P}}^{\mathbb{Q}}, \Phi_{\mathcal{P}}^{\mathbb{Q}}, 0_{NR\times 1}\right]$. For clarity, λ contains all elements of λ_γ, as well as $N(N+R+1) - [NR+1]$ zeros, and in our case $N = 3$ and $R = 1$.

After the above modifications the rest is straightforward to conclude and one can easily follow in the footsteps of Bauer (2018) for a Bayesian framework, as well as Chapter 1 for a corresponding sequential implementation thereof, to eventually arrive at a an almost complete estimation framework for a linear model with macros which is comparable to the setup we develop in this chapter for the nonlinear case. What is still missing though is the prior specification for $\Phi_{\mathcal{P}M}^{\mathbb{P}}$. This we choose to be non-informative and thus assign independent normal distribution with zero mean and large variance to each element thereof.

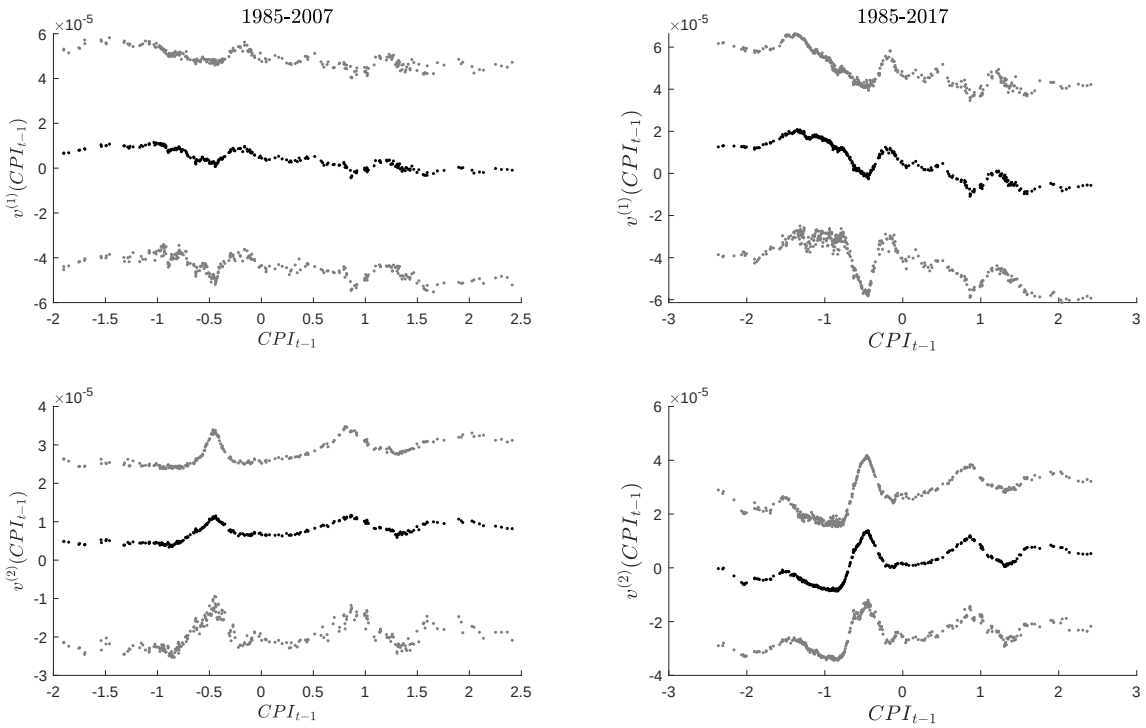

Figure 3.1: Plots of lagged CPI against its nonlinear function v, obtained from model GP_{110}. Plots in the left column are based on model parameters estimated using data from the training period only (January 1985 - end of 2007). Plots in the right column result from model parameters estimated using the entire sample of data (January 1985 - end of 2017). Throughout, points in black correspond to posterior mean of v and those in grey to its 95% credible intervals, all calculated from the IBIS output.

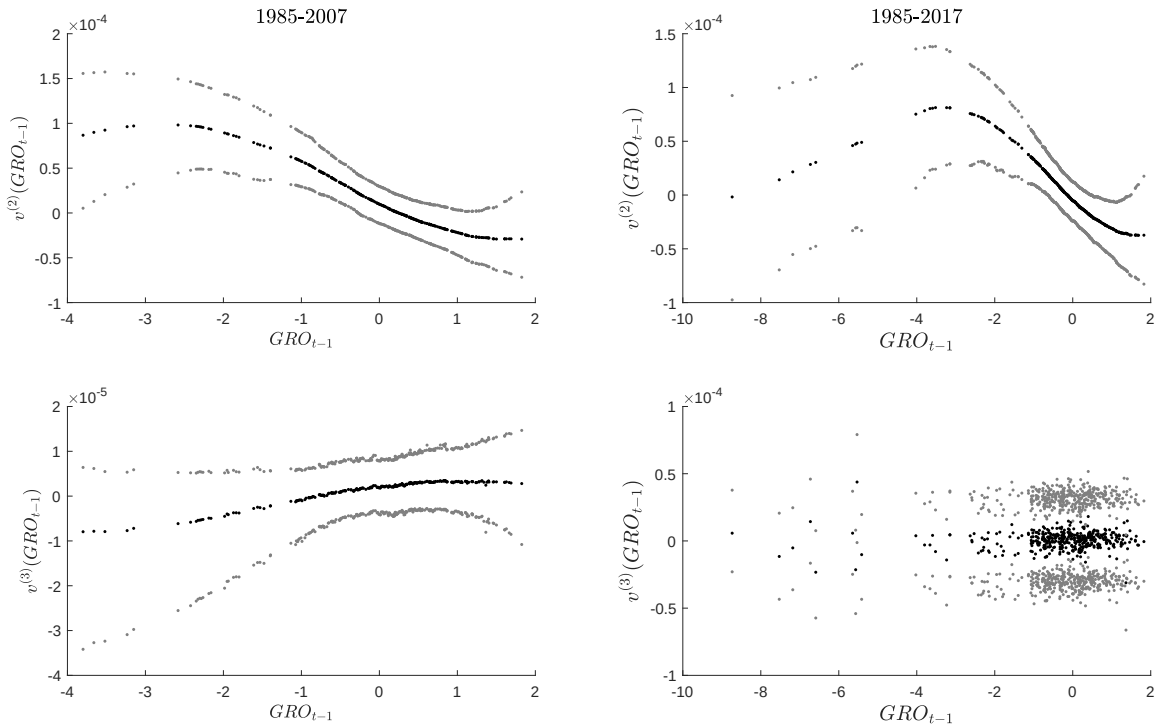

Figure 3.2: Plots of lagged GRO against its nonlinear function v, obtained from model GP_{011}. Plots in the left column are based on model parameters estimated using data from the training period only (January 1985 - end of 2007). Plots in the right column result from model parameters estimated using the entire sample of data (January 1985 - end of 2017). Throughout, points in black correspond to posterior mean of v and those in grey to its 95% credible intervals, all calculated from the IBIS output.

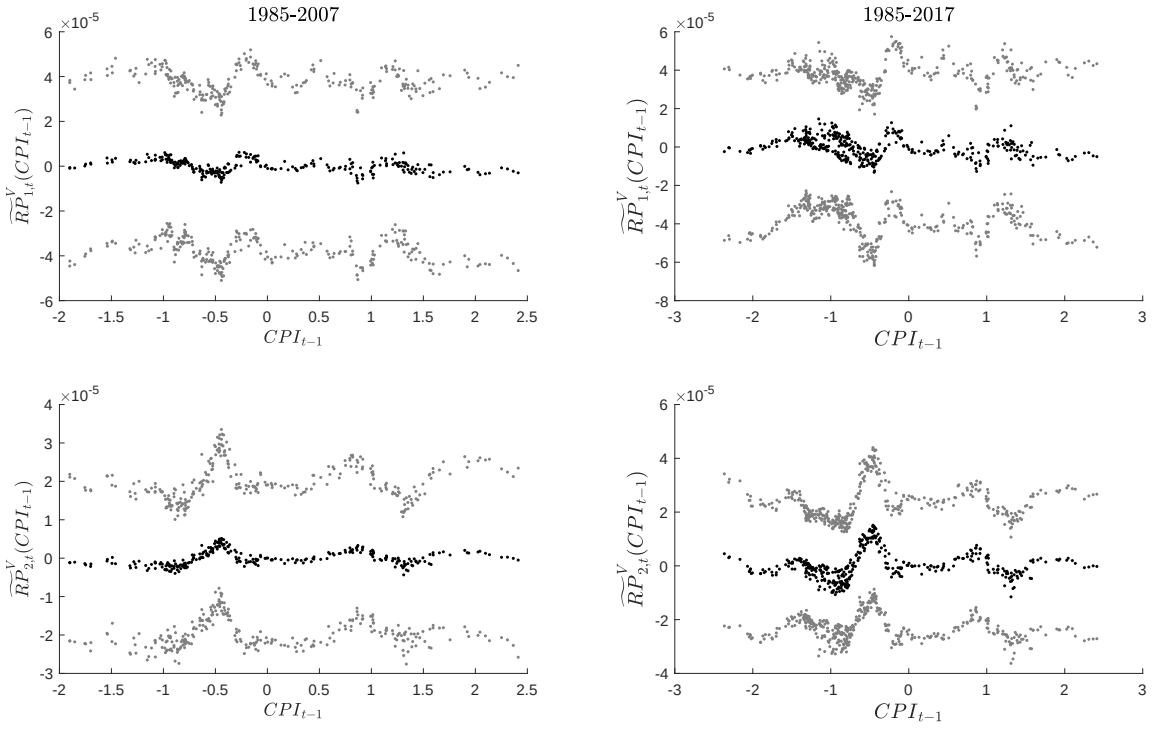

Figure 3.3: Plots of lagged CPI against the hidden part $\widetilde{RP}^V$ of its nonlinear function v, obtained from model GP_{110}. Plots in the left column are based on model parameters estimated using data from the training period only (January 1985 - end of 2007). Plots in the right column result from model parameters estimated using the entire sample of data (January 1985 - end of 2017). Throughout, points in black correspond to posterior mean of $\widetilde{RP}^V$ and those in grey to its 95% credible intervals, all calculated from the IBIS output.

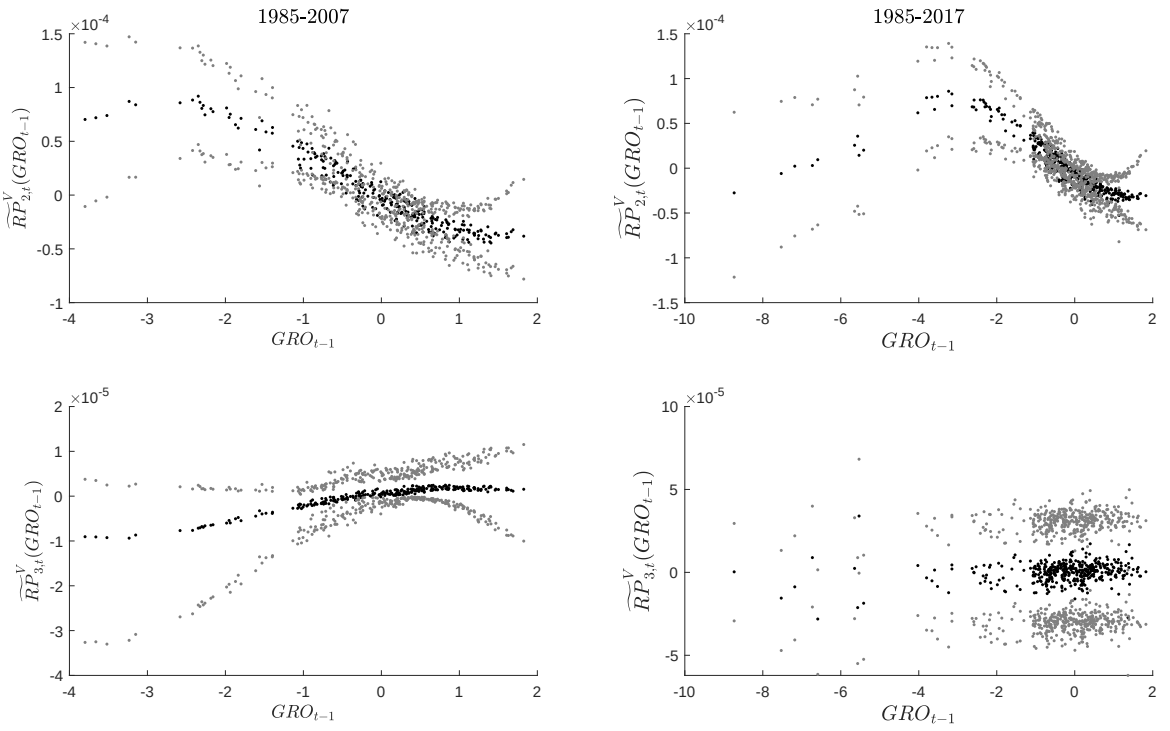

Figure 3.4: Plots of lagged GRO against the hidden part $\widetilde{RP}^V$ of its nonlinear function v, obtained from model GP_{011}. Plots in the left column are based on model parameters estimated using data from the training period only (January 1985 - end of 2007). Plots in the right column result from model parameters estimated using the entire sample of data (January 1985 - end of 2017). Throughout, points in black correspond to posterior mean of $\widetilde{RP}^V$ and those in grey to its 95% credible intervals, all calculated from the IBIS output.

Table 3.1: Explanatory power gains from macroeconomic variables, CPI and GRO, when fitting excess bond returns, measured via $\bar{R}^2$ at 1-month prediction horizon - period: January 1985 - end of 2017.

n	**2Y**	**3Y**	**4Y**	**5Y**	**7Y**	**10Y**
	$\bar{R}^2(\%) : rx^n_{t,t+1} = a + b'\mathcal{P}_t + e_t$					
	4.26	3.29	3.08	2.74	2.77	3.55
	$\Delta_{(0)}\bar{R}^2(\%) : rx^n_{t,t+1} = a + b'\mathcal{P}_t + cM_{t-1} + e_t$					
CPI	-0.24	-0.24	-0.22	-0.16	0.05	0.10*
GRO	1.95***	1.47***	0.56*	0.08	-0.21	-0.17
	$\bar{R}^2(\%) : rx^n_{t,t+1} = a + b'\mathcal{P}_t + cM_{t-1} + e_t$					
CPI	4.03	3.06	2.86	2.58	2.81	3.65
GRO	6.13	4.71	3.61	2.82	2.56	3.38

This table reports in-sample $\bar{R}^2$ in % across alternative regression specifications at h = 1-month prediction horizon. The explained variables are different (by maturities) excess bond returns. The explanatory variables are the principal components $\mathcal{P}_t$ and observed (lagged) macros M_{t-1}, in particular CPI and GRO. In-sample $\bar{R}^2$ values are obtained in a similar manner to the out-of-sample R^2 measure of Campbell and Thompson (2008) but using in-sample fit instead of out-of-sample forecasts, and incorporating penalty adjustment. In particular, $\bar{R}^2$ in the top panel (all highly statistically significant hence not denoted) measures explanatory power gains from using principal components on top of the in-sample average to fit excess bond returns, whereas $\bar{R}^2$ in the bottom panel (all highly statistically significant hence not denoted) measures explanatory power gains from using principal components and M_{t-1} on top of the in-sample average to fit excess bond returns. Further, $\Delta_{(0)}$ next to $\bar{R}^2$ in the mid panel means that the latter measures explanatory power gains from using macro M_{t-1} on top of the in-sample average and the principal components to do the same. Positive values of this statistic imply that there is explanatory power gain from adding extra variables. Statistical significance is measured using a one-sided Diebold-Mariano statistic with Clark-West adjustment, based on Newey-West standard errors. * denotes significance at 10%, ** significance at 5% and *** significance at 1% level. The sample period is January 1985 to end of 2017.

Table 3.2: Explanatory power gains from nonlinear macros estimated using model GP_{111}, when fitting excess bond returns, measured via $\bar{R}^2$ at 1-month prediction horizon - period: January 1985 - end of 2017.

n	**2Y**	**3Y**	**4Y**	**5Y**	**7Y**	**10Y**
	$\bar{R}^2(\%) : rx^n_{t,t+1} = a + b'\mathcal{P}_t + e_t$					
	4.26	3.29	3.08	2.74	2.77	3.55
GP_{111}	$\Delta_{(3)}\bar{R}^2(\%) : rx^n_{t,t+1} = a + b'\mathcal{P}_t + c\, v^{(3)}(M_{t-1}) + e_t$					
CPI	-0.14	0.17	0.44*	0.71**	1.48**	2.29***
GRO	-0.24	-0.11	0.04	0.26	1.19*	2.19**
GP_{111}	$\Delta_{(2)3}\bar{R}^2(\%) : rx^n_{t,t+1} = a + b'\mathcal{P}_t + c'[v^{(2)}(M_{t-1}), v^{(3)}(M_{t-1})]' + e_t$					
CPI	1.01**	0.42*	0.17	-0.12	-0.20	-0.25
GRO	1.20**	0.52*	0.11	-0.15	-0.25	-0.25
GP_{111}	$\Delta_{(1)23}\bar{R}^2(\%) : rx^n_{t,t+1} = a + b'\mathcal{P}_t + c'[v^{(1)}(M_{t-1}), v^{(2)}(M_{t-1}), v^{(3)}(M_{t-1})]' + e_t$					
CPI	0.66**	0.58**	0.59**	0.64**	0.07	-0.13
GRO	2.91***	2.86***	2.74***	2.17***	1.27***	0.34*
GP_{111}	$\Delta_{(123)}\bar{R}^2(\%) : rx^n_{t,t+1} = a + b'\mathcal{P}_t + c'[v^{(1)}(M_{t-1}), v^{(2)}(M_{t-1}), v^{(3)}(M_{t-1})]' + e_t$					
CPI	1.51***	1.16***	1.20**	1.23***	1.34**	1.91***
GRO	3.85***	3.25***	2.88***	2.28***	2.20***	2.27***
GP_{111}	$\bar{R}^2(\%) : rx^n_{t,t+1} = a + b'\mathcal{P}_t + c'[v^{(1)}(M_{t-1}), v^{(2)}(M_{t-1}), v^{(3)}(M_{t-1})]' + e_t$					
CPI	5.71	4.40	4.23	3.93	4.07	5.39
GRO	7.94	6.43	5.86	4.95	4.90	5.73

This table reports in-sample $\bar{R}^2$ in % across alternative regression specifications at h = 1-month prediction horizon. The explained variables are different (by maturities) excess bond returns. The explanatory variables are the principal components $\mathcal{P}_t$ and the estimated (via posterior mean) nonlinear macros $v(M_{t-1})$ from model GP_{111}, for CPI and GRO. In-sample $\bar{R}^2$ values are obtained in a similar manner to the out-of-sample R^2 measure of Campbell and Thompson (2008) but using in-sample fit instead of out-of-sample forecasts, and incorporating penalty adjustment. In particular, $\bar{R}^2$ in the top panel (all highly statistically significant hence not denoted) measures explanatory power gains from using principal components on top of the in-sample average to fit excess bond returns, whereas $\bar{R}^2$ in the bottom panel (all highly statistically significant hence not denoted) measures explanatory power gains from using principal components and $v^{(1)}(M_{t-1})$, $v^{(2)}(M_{t-1})$ and $v^{(3)}(M_{t-1})$ on top of the in-sample average to fit excess bond returns. Further, $\Delta_{(3)}$ next to $\bar{R}^2$ in the second panel means that the latter measures explanatory power gains from using nonlinear macro $v^{(3)}(M_{t-1})$ estimated using model $GP_{111}(M)$ on top of the in-sample average and the principal components to do the same. Next, $\Delta_{(2)3}$ in the third panel means that the associated $\bar{R}^2$ measures explanatory power gains from using to that end the nonlinear macro $v^{(2)}(M_{t-1})$ on top of the in-sample average, the principal components and nonlinear macro $v^{(3)}(M_{t-1})$. Then, $\Delta_{(1)23}$ in the fourth panel means that the associated $\bar{R}^2$ measures explanatory power gains from using to that end the nonlinear macro $v^{(1)}(M_{t-1})$ on top of the in-sample average, the principal components and nonlinear macros $v^{(2)}(M_{t-1})$ and $v^{(3)}(M_{t-1})$. Finally, $\Delta_{(123)}$ next to $\bar{R}^2$ in the fifth panel means that is measures the joint effect of including all latent factors on top of the remaining explanatory variables to fit excess bond returns. Positive values of this statistic imply that there is explanatory power gain from adding extra variables. Statistical significance is measured using a one-sided Diebold-Mariano statistic with Clark-West adjustment, based on Newey-West standard errors. * denotes significance at 10%, ** significance at 5% and *** significance at 1% level. The sample period is January 1985 to end of 2017.

Table 3.3: Out-of-sample statistical performance of bond excess return forecasts against the EH, measured via R^2_{os} at 1-month prediction horizon - period: January 1985 - end of 2018.

n	**2Y**	**3Y**	**4Y**	**5Y**	**7Y**	**10Y**
M_0	-0.04	-0.06	-0.05	-0.04	-0.02	-0.01**
M_1	0.01	0.03**	0.03*	0.02	0.02*	0.04**
			CPI			
LM_{111}	0.03**	0.04***	0.03**	0.03**	0.03**	0.03***
GP_{111}	0.02*	0.04**	0.03**	0.03*	0.03*	0.05***
LM_{011}	0.03**	0.05***	0.04**	0.03**	0.03**	0.04***
GP_{011}	0.03**	0.04***	0.03**	0.03**	0.03*	0.04***
LM_{110}	0.03**	0.03***	0.03**	0.02**	0.02**	0.03***
GP_{110}	0.06***	0.06***	0.05***	0.04**	0.04**	0.04***
LM_{010}	0.03*	0.04***	0.04**	0.03**	0.03**	0.04***
GP_{010}	0.05***	0.05***	0.04***	0.04**	0.03**	0.04***
LM_{001}	0.03*	0.04**	0.03**	0.03*	0.03*	0.04***
GP_{001}	0.03**	0.04***	0.04**	0.03**	0.03*	0.04***
			GRO			
LM_{111}	-0.19	-0.03	-0.03	-0.04	-0.01	0.01
GP_{111}	-0.03	0.01*	0.01*	0.01	0.01	0.01*
LM_{011}	0.02	0.05**	0.04***	0.03**	0.02*	-0.02
GP_{011}	0.03*	0.05**	0.03**	0.02**	0.01	0.00
LM_{110}	-0.22	-0.05	-0.05	-0.07	-0.04	0.00
GP_{110}	-0.04	0.00	0.00	0.00	0.00	0.00
LM_{010}	0.01	0.05**	0.03**	0.02**	0.02*	-0.01
GP_{010}	0.03*	0.05**	0.03**	0.02**	0.01*	0.00
LM_{001}	0.03*	0.04**	0.04**	0.04**	0.03*	0.03**
GP_{001}	0.03**	0.04***	0.04**	0.03**	0.03**	0.04***

This table reports out-of-sample R^2 across alternative models at h = 1-month prediction horizon. The forecasting models used are DTSM with either alternative risk price restrictions or different number of GP outputs with various macros or just other macros. R^2 values are generated using the out-of-sample R^2 measure of Campbell and Thompson (2008). In particular, out-of-sample R^2 measures the predictive accuracy of bond excess return forecasts relative to the EH benchmark. The EH implies the historical mean being the optimal forecast of excess returns. Positive values of this statistic imply that the forecast outperforms the historical mean forecast and suggests evidence of time-varying return predictability. Statistical significance is measured using a one-sided Diebold-Mariano statistic with Clark-West adjustment, based on Newey-West standard errors. * denotes significance at 10%, ** significance at 5% and *** significance at 1% level. The in-sample period is January 1985 to end of 2007, and the out-of-sample period starts in January 2008 and ends in end of 2018.

Table 3.4: Out-of-sample statistical performance of bond excess return forecasts against the corresponding linear model, measured via R^2_{os} at 1-month prediction horizon - period: January 1985 - end of 2018.

n	**2Y**	**3Y**	**4Y**	**5Y**	**7Y**	**10Y**
			CPI			
GP_{111}	-0.01	0.00	0.00	0.00	0.01	0.02*
GP_{011}	0.00	0.00	0.00	0.00	-0.01	0.00
GP_{110}	0.03**	0.03**	0.02*	0.01	0.01	0.02*
GP_{010}	0.02**	0.01	0.01	0.00	0.00	0.01
GP_{001}	0.01	0.01*	0.01	0.01	0.00	-0.01
			GRO			
GP_{111}	0.13*	0.04	0.04	0.05*	0.02	0.01
GP_{011}	0.01	-0.01	-0.01	-0.01	-0.01	0.02
GP_{110}	0.15**	0.05*	0.05	0.06*	0.03	0.00
GP_{010}	0.02	0.00	0.00	0.00	-0.01	0.01
GP_{001}	0.00	0.00	0.00	-0.01	0.00	0.01

This table reports out-of-sample R^2 across alternative models at h = 1-month prediction horizon. The forecasting models used are DTSM with different number of GP outputs and various macros or just other macros. R^2 values are generated using the out-of-sample R^2 measure of Campbell and Thompson (2008). In particular, out-of-sample R^2 measures the predictive accuracy of bond excess return forecasts relative to the corresponding linear model. Positive values of this statistic imply that the forecast outperforms forecast from the benchmark model. Statistical significance is measured using a one-sided Diebold-Mariano statistic with Clark-West adjustment, based on Newey-West standard errors. * denotes significance at 10%, ** significance at 5% and *** significance at 1% level. The in-sample period is January 1985 to end of 2007, and the out-of-sample period starts in January 2008 and ends in end of 2018.

Table 3.5: Out-of-sample economic performance of bond excess return forecasts against the EH, measured via certainty equivalent returns (%) at 1-month prediction horizon - period: January 1985 - end of 2018.

n	**2Y**	**3Y**	**4Y**	**5Y**	**7Y**	**10Y**
M_0	-10.77	-11.81	-10.27	-8.05	-3.62	-7.49
M_1	2.31	1.88	1.38	0.80	2.30	2.55
			CPI			
LM_{111}	1.68	1.22	1.57	1.36	3.61**	1.45
GP_{111}	3.00*	2.79*	2.73	2.56	4.30*	3.97*
LM_{011}	2.79*	2.54*	2.31	1.70	3.59	2.39
GP_{011}	3.69**	3.30**	2.91*	2.50	4.01*	3.69*
LM_{110}	0.74	0.02	0.07	0.12	2.56*	0.06
GP_{110}	4.10**	3.79**	3.27	2.68	3.98	2.68
LM_{010}	2.22*	2.07	2.13	2.09	4.33**	2.28
GP_{010}	4.39***	4.29**	4.09**	3.67*	4.94*	3.69**
LM_{001}	2.40*	2.29	2.01	1.71	3.25	2.67
GP_{001}	4.14**	4.43***	3.99**	3.60*	4.10	3.33*
			GRO			
LM_{111}	-7.42	-1.44	-5.44	-5.89	-10.11	-10.29
GP_{111}	2.28	3.34	2.85	1.96	1.88	-0.52
LM_{011}	3.33	3.93**	3.48*	3.16	5.65*	-3.89
GP_{011}	3.51	3.20*	2.80	1.80	2.59	-2.30
LM_{110}	-12.12	-4.54	-9.08	-8.39	-13.46	-12.49
GP_{110}	0.33	1.44	1.23	0.60	0.84	-1.46
LM_{010}	2.59	3.41	2.62	2.12	5.28*	-2.63
GP_{010}	3.45	3.11*	2.50*	1.96	2.73	-2.86
LM_{001}	3.61**	3.37**	3.75**	3.50	4.29	1.60
GP_{001}	3.86**	3.91**	3.84*	3.17	4.55*	4.24**

This table reports annualized certainty equivalent returns (CERs) across alternative models at h = 1-month prediction horizon. The coefficient of risk aversion is $\gamma = 3$. No portfolio constraints are imposed. CERs are generated by out-of-sample forecasts of bond excess returns and are reported in %. At every time step t, an investor with power utility preferences evaluates the entire predictive density of bond excess returns and solves the asset allocation problem, thus optimally allocating her wealth between a riskless bond and risky bonds with maturities 2, 3, 4, 5, 7 and 10-years. CER is then defined as the value that equates the average utility of each alternative model against the average utility of the EH benchmark. The forecasting models used are DTSM with either alternative risk price restrictions or different number of GP outputs with various macros or just other macros. Positive values indicate that the models perform better than the EH benchmark. Statistical significance is measured using a one-sided Diebold-Mariano statistic computed with Newey-West standard errors. * denotes significance at 10%, ** significance at 5% and *** significance at 1% level. The in-sample period is January 1985 to end of 2007, and the out-of-sample period starts in January 2008 and ends in end of 2018.

Table 3.6: Out-of-sample economic performance of bond excess return forecasts against the corresponding linear model, measured via certainty equivalent returns (%) at 1-month prediction horizon - period: January 1985 - end of 2018.

n	**2Y**	**3Y**	**4Y**	**5Y**	**7Y**	**10Y**
			CPI			
GP_{111}	1.32	1.56	1.16	1.20	0.69	2.52
GP_{011}	0.89*	0.76	0.60	0.79	0.42	1.30
GP_{110}	3.35***	3.78***	3.19**	2.55*	1.41	2.62*
GP_{010}	2.16**	2.21***	1.95**	1.58	0.61	1.41
GP_{001}	1.88**	2.54**	2.34**	2.17**	0.61	-0.36
			GRO			
GP_{111}	9.76	4.78	8.32	7.89	12.09*	9.85
GP_{011}	0.17	-0.73	-0.68	-1.35	-3.05	1.59
GP_{110}	12.58	6.00	10.39	9.06	14.46*	11.15
GP_{010}	0.86	-0.30	-0.12	-0.17	-2.54	-0.23
GP_{001}	0.24	0.54	0.09	-0.33	0.25	2.63

This table reports annualized certainty equivalent returns (CERs) across alternative models at h = 1-month prediction horizon. The coefficient of risk aversion is $\gamma = 3$. No portfolio constraints are imposed. CERs are generated by out-of-sample forecasts of bond excess returns and are reported in %. At every time step t, an investor with power utility preferences evaluates the entire predictive density of bond excess returns and solves the asset allocation problem, thus optimally allocating her wealth between a riskless bond and risky bonds with maturities 2, 3, 4, 5, 7 and 10-years. CER is then defined as the value that equates the average utility of each alternative model against the average utility of the corresponding linear model. The forecasting models used are DTSM with different number of GP outputs and various macros or just other macros. Positive values indicate that the models perform better than the benchmark model. Statistical significance is measured using a one-sided Diebold-Mariano statistic computed with Newey-West standard errors. * denotes significance at 10%, ** significance at 5% and *** significance at 1% level. The in-sample period is January 1985 to end of 2007, and the out-of-sample period starts in January 2008 and ends in end of 2018.

Table 3.7: Explanatory power of macroeconomic variables when fitting the corresponding nonlinear macros and their components from models GP_{001}, GP_{010} and GP_{110}, measured via $\bar{R}^2$ - period: January 1985 - end of 2017.

$\bar{R}^2 : C_{j,t} = a_j + b_j M_{t-1} + e_{j,t},\ j \in \{1,2,3\}$		
GP_{001}	CPI	GRO
$v^{(3)}(M_{t-1})$	0.06	0.39
$E[v^{(3)}(M_{t-1})\|\mathcal{P}_t]$	0.70	0.11
$\widetilde{RP}^{V}_{3,t}(M_{t-1})$	0.00	0.33
GP_{010}	CPI	GRO
$v^{(2)}(M_{t-1})$	0.23	0.53
$E[v^{(2)}(M_{t-1})\|\mathcal{P}_t]$	0.66	0.18
$\widetilde{RP}^{V}_{2,t}(M_{t-1})$	0.00	0.39
GP_{110}	CPI	GRO
$v^{(1)}(M_{t-1})$	0.73	0.25
$E[v^{(1)}(M_{t-1})\|\mathcal{P}_t]$	0.70	0.17
$\widetilde{RP}^{V}_{1,t}(M_{t-1})$	0.09	0.17
$v^{(2)}(M_{t-1})$	0.32	0.45
$E[v^{(2)}(M_{t-1})\|\mathcal{P}_t]$	0.67	0.17
$\widetilde{RP}^{V}_{2,t}(M_{t-1})$	0.00	0.34

This table reports in-sample $\bar{R}^2$ across alternative regression specifications. The explained variables $C_{j,t}$ are individual nonlinear macros $v^{(j)}(M_{t-1})$ and their components $E[v^{(j)}(M_{t-1})|\mathcal{P}_t]$ and $\widetilde{RP}^{V}_{j,t}(M_{t-1})$, $j \in \{1,2,3\}$, see (3.21), from models GP_{001}, GP_{010} and GP_{110}. The explanatory variables are the corresponding lagged macroeconomic variables. The sample period is January 1985 to end of 2017.

Table 3.8: Explanatory power of macroeconomic variables when fitting the corresponding nonlinear macros and their components from models GP_{011} and GP_{111}, measured via $\bar{R}^2$ - period: January 1985 - end of 2017.

$\bar{R}^2 : C_{j,t} = a_j + b_j M_{t-1} + e_{j,t},\ j \in \{1,2,3\}$		
GP_{011}	CPI	GRO
$v^{(2)}(M_{t-1})$	0.44	0.50
$E[v^{(2)}(M_{t-1})\|\mathcal{P}_t]$	0.68	0.17
$\widetilde{RP}^{V}_{2,t}(M_{t-1})$	0.01	0.37
$v^{(3)}(M_{t-1})$	0.60	0.00
$E[v^{(3)}(M_{t-1})\|\mathcal{P}_t]$	0.66	0.15
$\widetilde{RP}^{V}_{3,t}(M_{t-1})$	0.07	0.03
GP_{111}	CPI	GRO
$v^{(1)}(M_{t-1})$	0.75	0.25
$E[v^{(1)}(M_{t-1})\|\mathcal{P}_t]$	0.70	0.16
$\widetilde{RP}^{V}_{1,t}(M_{t-1})$	0.09	0.17
$v^{(2)}(M_{t-1})$	0.50	0.44
$E[v^{(2)}(M_{t-1})\|\mathcal{P}_t]$	0.69	0.17
$\widetilde{RP}^{V}_{2,t}(M_{t-1})$	0.02	0.32
$v^{(3)}(M_{t-1})$	0.63	0.00
$E[v^{(3)}(M_{t-1})\|\mathcal{P}_t]$	0.65	0.14
$\widetilde{RP}^{V}_{3,t}(M_{t-1})$	0.08	0.04

This table reports in-sample $\bar{R}^2$ across alternative regression specifications. The explained variables $C_{j,t}$ are individual nonlinear macros $v^{(j)}(M_{t-1})$ and their components $E[v^{(j)}(M_{t-1})|\mathcal{P}_t]$ and $\widetilde{RP}^{V}_{j,t}(M_{t-1})$, $j \in \{1,2,3\}$, see (3.21), from models GP_{011} and GP_{111}. The explanatory variables are the corresponding lagged macroeconomic variables. The sample period is January 1985 to end of 2017.

References

Adrian, T., Crump, R. K. and Moench, E. (2013), 'Pricing the term structure with linear regressions', *Journal of Financial Economics* **110**(1), 110–138.

Akaike, H. (1998), Information theory and an extension of the maximum likelihood principle, *in* E. Parzen, K. Tanabe and G. Kitagawa, eds, 'Selected Papers of Hirotugu Akaike', Springer, New York, pp. 199–213.

Alvarez, M. A., Rosasco, L. and Lawrence, N. D. (2012), 'Kernels for vector-valued functions: A review', *Foundations and Trends® in Machine Learning* **4**(3), 195–266.

Ang, A., Dong, S. and Piazzesi, M. (2007), No-arbitrage taylor rules, Working Paper 13448, National Bureau of Economic Research.

Ang, A. and Longstaff, F. A. (2011), 'Systemic sovereign credit risk: Lessons from the us and europe', *Journal of Monetary Economics* **60**(5), 493–510.

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.

Barillas, F. (2011), Can we exploit predictability in bond markets?, Working paper, Emory University.

Bauer, M. D. (2018), 'Restrictions on risk prices in dynamic term structure models', *Journal of Business & Economic Statistics* **36**(2), 196–211.

Bauer, M. D. and Hamilton, J. D. (2018), 'Robust bond risk premia', *The Review of Financial Studies* **31**(2), 399–448.

Bauer, M. D. and Rudebusch, G. D. (2016), 'Resolving the spanning puzzle in macro-finance term structure models', *Review of Finance* **21**(2), 511–553.

Bianchi, D., Büchner, M. and Tamoni, A. (2021), 'Bond risk premiums with machine learning', *The Review of Financial Studies* **34**(2), 1046–1089.

Bishop, C. M. (2006), *Pattern Recognition and Machine Learning*, Springer.

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**(3), 495–514.

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

Cheridito, P., Filipović, D. and Kimmel, R. L. (2007), 'Market price of risk specifications for affine models: Theory and evidence', *Journal of Financial Economics* **83**(1), 123–170.

Chib, S. and Ergashev, B. (2009), 'Analysis of multifactor affine yield curve models.', *Journal of the American Statistical Association* **104**(488), 1324–1337.

Chib, S. and Kang, K. H. (2013), 'Change-points in affine arbitrage-free term structure models', *Journal of Financial Econometrics* **11**(2), 302–334.

Chopin, N. (2002), 'A sequential particle filter method for static models', *Biometrika* **89**(3), 539–552.

Chopin, N. (2004), 'Central limit theorem for sequential monte carlo methods and its application to bayesian inference', *The Annals of Statistics* **32**(6), 2385–2411.

Cieslak, A. (2018), 'Short-rate expectations and unexpected returns in treasury bonds', *The Review of Financial Studies* **31**(9), 3265–3306.

Cieslak, A. and Povala, P. (2015), 'Expected returns in treasury bonds', *The Review of Financial Studies* **28**(10), 2859–2901.

Clark, T. E. and West, K. D. (2007), 'Approximately normal tests for equal predictive accuracy in nested models', *Journal of Econometrics* **138**(1), 291–311.

Cochrane, J. H. and Piazzesi, M. (2005), 'Bond risk premia', *The American Economic Review* **95**(1), 138–160.

Cochrane, J. H. and Piazzesi, M. (2009), Decomposing the yield curve, Working paper, AFA 2010 Atlanta Meetings.

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.

Consonni, G., Fouskakis, D., Liseo, B. and Ntzoufras, I. (2018), 'Prior Distributions for Objective Bayesian Analysis', *Bayesian Analysis* **13**(2), 627 – 679.

Cooper, I. and Priestley, R. (2009), 'Time-varying risk premiums and the output gap', *The Review of Financial Studies* **22**(7), 2801–2833.

Creal, D. D. and Wu, J. C. (2015), 'Estimation of affine term structure models with spanned or unspanned stochastic volatility', *Journal of Econometrics* **185**, 60–81.

Dai, Q. and Singleton, K. J. (2000), 'Specification analysis of affine term structure models', *The Journal of Finance* **55**(5), 1943–1978.

Dawid, A. P. and Musio, M. (2014), 'Theory and applications of proper scoring rules', *METRON* **72**(2), 169–183.

Del Moral, P., Doucet, A. and Jasra, A. (2006), 'Sequential monte carlo samplers', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **68**(3), 411–436.

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.

Dewachter, H. and Lyrio, M. (2006), 'Macro factors and the term structure of interest rates', *Journal of Money, Credit and Banking* **38**(1), 119–140.

Diebold, F. X. and Rudebusch, G. D. (2013), *Yield Curve Modeling and Forecasting: The Dynamic Nelson-Siegel Approach*, Princeton University Press.

Duarte, J. (2004), 'Evaluating an alternative risk preference in affine term structure models', *Review of Financial Studies* **17**(2), 379–404.

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

Duffee, G. R. (2006), 'Term structure estimation without using latent factors', *Journal of Financial Economics* **79**, 507–536.

Duffee, G. R. (2011), 'Information in (and not in) the term structure', *The Review of Financial Studies* **24**(9), 2895–2934.

Duffee, G. R. and Stanton, R. H. (2012), 'Estimation of dynamic term structure models', *The Quarterly Journal of Finance* **2**(02), 1250008.

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

Durbin, J. and Koopman, S. J. (2012), *Time Series Analysis by State-Space Methods*, Oxford University Press.

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

Feldhütter, P. (2016), 'Can affine models match the moments in bond yields?', *Quarterly Journal of Finance* **6**(2), 1650009.

Feunou, B. and Fontaine, J.-S. (2018), 'Bond risk premia and Gaussian term structure models', *Management Science* **64**(3), 1413–1439.

Fong, E. and Holmes, C. C. (2020), 'On the marginal likelihood and cross-validation', *Biometrika* **107**(2), 489–496.

Frigola, R., Lindsten, F., Schoen, T. B. and Rasmussen, C. E. (2013), 'Bayesian inference and learning in gaussian process state-space models with particle mcmc', *NIPS'13: Proceedings of the 26th International Conference on Neural Information Processing Systems* **2**, 3156–3164.

Fulop, A., Li, J. and Wan, R. (2019), 'Real-time bayesian learning and bond return predictability', *Journal of Econometrics* . Forthcoming.

Gargano, A., Pettenuzzo, D. and Timmermann, A. (2019), 'Bond return predictability: Economic value and links to the macroeconomy', *Management Science* **65**(2), 508–540.

George, E. I. and McCulloch, R. E. (1993), 'Variable selection via gibbs sampling', *Journal of the American Statistical Association* **88**(423), 881–889.

Geweke, J. and Amisano, G. (2010), 'Comparing and evaluating bayesian predictive distributions of asset returns', *International Journal of Forecasting* **26**(2), 216–230.

Ghysels, E., Horan, C. and Moench, E. (2018), 'Forecasting through the rearview mirror: Data revisions and bond return predictability', *The Review of Financial Studies* **31**(2), 678–714.

Giacoletti, M., Laursen, K. T. and Singleton, K. J. (2021), 'Learning from disagreement in the us treasury bond market', *The Journal of Finance* **76**(1), 395–441.

Hamilton, J. D. and Wu, J. C. (2012), 'Identification and estimation of Gaussian affine term structure models', *Journal of Econometrics* **168**(2), 315–331.

Hannan, E. J. and Quinn, B. G. (1979), 'The determination of the order of an autoregression', *Journal of the Royal Statistical Society. Series B (Methodological)* **41**(2), 190–195.

Jasra, A., Stephens, D. A., Doucet, A. and Tsagaris, T. (2011), 'Inference for lévy-driven stochastic volatility models via adaptive sequential monte carlo', *Scandinavian Journal of Statistics* **38**(1), 1–22.

Johannes, M., Korteweg, A. and Polson, N. (2014), 'Sequential learning, predictability, and optimal portfolio returns', *The Journal of Finance* **69**(2), 611–644.

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.

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.

Kantas, N., Beskos, A. and Jasra, A. (2014), 'Sequential monte carlo methods for high-dimensional inverse problems: a case study for the navier-stokes equations', *SIAM/ASA Journal on Uncertainty Quantification* **2**, 464–489.

Kass, R. E. and Raftery, A. E. (1995), 'Bayes factors', *Journal of the American Statistical Association* **90**, 81–94.

Kim, D. H. and Orphanides, A. (2012), 'Term structure estimation with survey data on interest rate forecasts', *Journal of Financial and Quantitative Analysis* pp. 241–272.

Kim, D. H. and Singleton, K. J. (2012), 'Term structure models and the zero bound: an empirical investigation of japanese yields', *Journal of Econometrics* **170**(1), 32–49.

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, Working paper, Finance and Economics Discussion Series 2005-33, Federal Reserve Board of Governors.

Lagarias, J. C., Reeds, J. A., Wright, M. H. and Wright, P. E. (1998), 'Convergence properties of the nelder-mead simplex method in low dimensions', *SIAM Journal of Optimization* **9**(1), 112–147.

Li, J., Sarno, L. and Zinna, G. (2021), 'Risks and risk premia in the us treasury market', *Available at SSRN 3640341* .

Litterman, R. B. and Scheinkman, J. (1991), 'Common factors affecting bond returns', *The Journal of Fixed Income* **1**(1), 54–61.

Lütkepohl, H. (2005), *New introduction to multiple time series analysis*, Springer.

Ludvigson, S. C. and Ng, S. (2009), 'Macro factors in bond risk premia', *The Review of Financial Studies* **22**(12), 5027–5067.

Madigan, D. and Raftery, A. E. (1994), 'Model selection and accounting for model uncertainty in graphical models using occam's window', *Journal of the American Statistical Association* **89**(428), 1535–1546.

Mitchell, T. J. and Beauchamp, J. J. (1988), 'Bayesian variable selection in linear regression', *Journal of the American Statistical Association* **83**(404), 1023–1032.

Murphy, K. P. (2012), *Machine Learning: A Probabilistic Perspective*, MIT Press.

Newey, W. K. and West, K. D. (1987), 'A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix', *Econometrica* **55**(3), 703–708.

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.

Piazzesi, M., Schneider, M., Benigno, P. and Campbell, J. Y. (2006), 'Equilibrium yield curves [with comments and discussion]', *NBER Macroeconomics Annual* **21**, 389–472.

Rasmussen, C. E. and Williams, C. K. I. (2006), *Gaussian processes for machine learning*, MIT Press.

Rudebusch, G. D. and Wu, T. (2008), 'A macro-finance model of the term structure, monetary policy and the economy', *The Economic Journal* **118**(530), 906–926.

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

Schäfer, C. and Chopin, N. (2013), 'Sequential Monte Carlo on large binary sampling spaces', *Statistics and Computing* **23**(2), 163–184.

Schwarz, G. (1978), 'Estimating the dimension of a model', *The Annals of Statistics* **6**(2), 461–464.

Schweppe, F. C. (1965), 'Evaluation of likelihood functions for gaussian signals', *IEEE Transactions on Information Theory* **11**, 61–70.

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

Wilson, M. A., Iversen, E. S., Clyde, M. A., Schmidler, S. C. and Schildkraut, J. M. (2010), 'Bayesian model search and multilevel inference for SNP association studies', *The Annals of Applied Statistics* **4**(3), 1342 – 1364.

Wright, J. H. (2011), 'Term premia and inflation uncertainty: Empirical evidence from an international panel dataset', *The American Economic Review* **101**(4), 1514–34.