Learning about Term Structure Predictability under Uncertainty

Shuo Cao*

Version: January 2016†

Abstract

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

Keywords: Affine Term Structure Models; Learning; Parameter Uncertainty; Model Uncertainty; Ambiguity Aversion; Bayesian Methods.
JEL Classification Codes: C1, C3, C5, D8, E4, G1.

*Adam Smith Business School, University of Glasgow, Glasgow, UK. Email: shuo.cao@outlook.com
†Any comments are welcome. Please do not circulate without permission.

1 Introduction

Modeling the interest rate term structure is essential in understanding expectations of risk compensation and the future path of monetary policy. For instance, the affine class of arbitrage-free term structure models has gained great popularity in both pricing and predicting future movements of bonds, because of its parsimonious factor structure and tractability. As a stylized fact, the predictability of bond returns is widely recognized in Fama and Bliss (1987), Cochrane and Piazzesi (2005), Sarno, Thornton and Valente (2007) and Ludvigson and Ng (2009). The traditional Expectations Hypothesis (EH) has been strongly rejected statistically and therefore, the term premia should be time-varying.[1] Accurate term premium predictions should be useful for portfolio optimization, as it guides a mean-variance investor to making a tradeoff between expected returns and the volatility of the portfolio.

Surprisingly, significant predictability in expected bond returns cannot be translated into large economic gains, as suggested by Della Corte, Sarno and Thornton (2008), Thornton and Valente (2012) and Sarno, Schneider and Wagner (2014). They reach the conclusion that when compared with a EH investor, the investor using alternative prediction models with statistical significance cannot improve economic utility. The seemingly contradictory evidence in this literature is indeed puzzling. Seeking to resolve this puzzle with those findings, Gargano, Pettenuzzo and Timmermann (2014) allow for parameter and model uncertainty. However, the resolution to the puzzle is far from perfect and further research is required. To resolve the *economic value puzzle*, it is necessary to understand the uncertainty in the predictability and, moreover, to consider various sources of uncertainty when making the optimal portfolio choice.

This issue is revisited by taking account of both parameter and model uncertainty. We propose a flexible term structure model which includes time-varying coefficients, stochastic volatility and dynamic model selection. These features are sensible in an agent's pricing and forecasting problem because parameters and models are uncertain without sufficient data. Like an econometrician, the agent needs to learn about the evolution of the state of the economy (Cagetti et al. (2002) and Hansen (2007)). These features can be formalized in an affine model to accommodate structural changes, where the learning speed or 'gain' is specified.[2]

[1] A weak form of EH requires the term premia to be a constant, which implies that expected excess bond returns should not be predictable.

[2] As shown in Evans and Honkapohja (2001), this framework provides an expectationally stable solution as long as the gain parameter is sufficiently small.

The seminal contribution of Timmermann (1993, 1996) studies the implications of learning in explaining the volatility and predictability of asset returns. There is substantial literature employing learning to explain a range of financial market anomalies. Specifically, Piazzesi and Schneider (2007) and Collin-Dufresne, Johannes and Lochstoer (2013) examine the implications of learning in a preference-based asset pricing framework, and they show that learning can explain standard puzzles in bond yields. Using a reduced-form pricing kernel, Kozicki and Tinsley (2001) and Dewachter and Lyrio (2008) study the learning problem in which agents continuously update their beliefs regarding the central bank's policy targets, but they only allow for time variations of the drift parameter. Laubach, Tetlow and Williams (2007), Orphanides and Wei (2012) and Cieslak and Povala (2014) relax the assumptions about the potential sources of structural instability and allow for updating beliefs of all model parameters. A common practice is to use macro variables as pricing factors, which may cause undesired mispricing as indicated by Anh and Joslin (2013). To avoid the potential mispricing problem and increase predictive power, in this paper we consider portfolios as risk factors, similar to the learning model of Giacoletti, Laursen and Singleton (2014).[3]

However, learning does not guarantee the convergence of agents' heterogeneous beliefs. Bond yields are highly persistent, and Kurz (1994) suggests that if the economic system is close to nonstationary, limited data would make it difficult for rational investors to identify the correct model from alternative ones. Model uncertainty can arise from the imposition of restrictions related to model identification. In order to increase forecast performance, researchers impose over-identifying restrictions motivated statistically or economically, see for example, Christensen, Diebold and Rudebusch (2011), Duffee and Stanton (2012) and Joslin, Priebsch and Singleton (2014). The economic dynamics are ambiguous with undetermined restrictions and therefore, statistical methods are employed to determine the optimal specification. Joslin, Priebsch and Singleton (2014) and Jotikasthira, Le and Lundblad (2015) choose restrictions based on the Bayesian Information Criterion (BIC), while Bauer (2015) uses Bayesian model averaging method to calculate the weighted average across specifications for robust inference. The above model selections are conducted with full-sample data, and a real-time dynamic model selection is desirable in the topic of interest rate forecasting. In stock return predictions, Cremers (2002) and Avramov (2002) have shown that allowing investors to dynamically

[3] Giacoletti, Laursen and Singleton (2014) incorporate macro information in the setup of unspanned macro risks proposed by Joslin, Priebsch and Singleton (2014), where the prices of risk can depend on both latent factors (portfolios) and macro variables.

select between different models is useful to control for the data snooping problem and can increase out-of-sample predictability. More recent studies, for instance, Dangl and Halling (2012), Johannes, Korteweg and Polson (2013) and Gargano, Pettenuzzo and Timmermann (2014) adopt Bayesian approaches to accommodate model uncertainty in stock and bond return forecasts. In this paper, we adopt the same framework as Dangl and Halling (2012) to conduct a real-time Bayesian model selection, explicitly exploring model uncertainty in term structure modeling.

Uncertainty in parameters and models is not explicitly priced in classical term structure models, but a sophisticated investor should be aware of the uncertainty when making investment decisions. Pástor and Stambaugh (1999, 2000) investigate how the uncertainty in parameters or models changes the way we make portfolio decisions.[4] Investors prefer known risks over unknown risks, so Uppal and Wang (2003) introduce an important extension to allow for ambiguity aversion. While most of the related research focuses on the portfolio allocation of stocks, very few recent papers have approached the topic of bond returns.[5] In order to close this gap, we consider a generalized framework with ambiguity aversion that nests the case of ordinary risk-averse investors, following and extending Garlappi, Uppal and Wang (2007).

This paper builds upon the work of Giacoletti, Laursen and Singleton (2014), who construct a learning framework of arbitrage-free affine term structure models and who investigate different learning rules in term structure forecasts. We further introduce model uncertainty in the learning problem, which provides flexibility in selecting the best restrictions imposed on factor dynamics and the optimal learning gain/speed. More importantly, this extension allows the analysis of the uncertainty in the predictive performance in order to reveal the sources of prediction uncertainty. Our work is also related to Gargano, Pettenuzzo and Timmermann (2014), who evaluate the economic gains of models with parameter and model uncertainty, but differs in a way that we consider a more generalized portfolio allocation problem with ambiguity aversion. In this framework, we explore to what extent investors benefit from ambiguity aversion in addition to the traditional risk aversion.

[4]Recent contributions on portfolio choice under uncertainty include Brandt et al. (2005), Avramov and Chordia (2006), and Rapach, Strauss and Zhou (2009).

[5]Gagliardini, Porchia and Trojani (2009) and Ulrich (2013) show that considering ambiguity aversion helps explain the term premia dynamics but they do not explicitly explore the portfolio allocation problem. Della Corte, Sarno and Thornton (2008), Thornton and Valente (2012) and Sarno, Schneider and Wagner (2014) study the bond portfolio selection problem for risk-averse investors. Gargano, Pettenuzzo and Timmermann (2014) evaluate the economic gains by considering parameter and model uncertainty, but also for risk-averse investors.

In particular, the proposed learning model nests most of the affine term structure models with learning, and is flexible in selecting the optimal specification from different learning speeds and model restrictions. Utilizing our approach we make several contributions to understanding the US bond market from 1961:06 to 2014:10. The first finding is that the pricing dynamics have not varied much since the 1960s, which is consistent with Giacoletti, Laursen and Singleton (2014), but we observe large variability in factor dynamics under the physical measure. The proposed model is promising in forecasting, as its predictive performance using conditional information is similar to the benchmark model using full information. By analyzing the sources of predictive uncertainty, it can be seen that, apart from observational variance, parameter instability is the main driver of predictive variance. Uncertainty in learning speed or model specification, *vis-à-vis* parameter instability, does not generally play an important role.

With respect to asset allocation, we consider both parameter and model uncertainty by extending the mean-variance framework proposed by Garlappi, Uppal and Wang (2007). Our ambiguity-averse investor successfully turns the predictability of excess returns implied by the learning model into substantial economic gains, when compared with the Expectations Hypothesis benchmark. In addition to parameter uncertainty, the consideration of model uncertainty is the key to ensuring success. This finding is robust compared to different subsample periods, despite that the economic gains can be eroded during the financial crisis. Therefore, this framework resolves the *economic value puzzle* in bond return predictions with the evidence in the previous term structure literature.

The rest of the paper is structured as follows. Section 2 describes the methodology, the term structure models considered, and the framework with ambiguity aversion for evaluating the out-of-sample predictability of excess returns. Section 3 outlines the empirical results of the learning model and its out-of-sample portfolio performance, including discussion about pricing dynamics, physical dynamics and term structure predictability. Section 4 concludes.

2 Methodology

2.1 A Canonical Gaussian Dynamic Term Structure Model (GDTSM)

We firstly consider an economic environment in which agents value nominal bonds using the stochastic discount factor or pricing kernel. The one-period pricing kernel or stochastic discount factor of an asset is given by

$$\mathcal{M}_{Z,t+1} = e^{-r_t - \frac{1}{2}\Lambda'_{Zt}\Lambda_{Zt} - \Lambda'_{Zt}\epsilon^{\mathbb{P}}_{t+1}}, \tag{2.1}$$

where the $N \times 1$ state vector Z_t encompasses all risks in the economy, Λ_{Zt} is the vector collecting market prices of risk, and r_t is the one-period bond yield. In the absence of macro risks, Z_t is a linear rotation of $N \times 1$ vector of portfolio risk factors $\mathcal{P}_t$.[6]

Following Joslin, Singleton and Zhu (2011), we specify the pricing kernel in the bond market. The bond-market-specific $\mathcal{M}_{P,t+1}$, conditional on the information of the priced risks in the bond market $\mathcal{P}_t$,[7] is now given by

$$\mathcal{M}_{\mathcal{P},t+1} = e^{-r_t - \frac{1}{2}\Lambda'_{\mathcal{P}t}\Lambda_{\mathcal{P}t} - \Lambda'_{\mathcal{P}t}\epsilon^{\mathbb{P}}_{\mathcal{P},t+1}}, \tag{2.2}$$

where the short rate r_t is an affine function of $\mathcal{P}_t$,

$$r_t = \rho_{0\mathcal{P}} + \rho_{1\mathcal{P}} \cdot \mathcal{P}_t, \tag{2.3}$$

and the risks $\epsilon^{\mathbb{P}}_{\mathcal{P},t+1}$ are the N innovations from the unconstrained first-order vector-autoregressive (VAR) model under the physical or historical measure $\mathbb{P}$[8]

$$\mathcal{P}_t = K^{\mathbb{P}}_{0\mathcal{P}} + K^{\mathbb{P}}_{\mathcal{P}\mathcal{P}}\mathcal{P}_{t-1} + \sqrt{\Sigma_{\mathcal{P}\mathcal{P}}}\epsilon^{\mathbb{P}}_{\mathcal{P}t}, \tag{2.4}$$

where $\epsilon^{\mathbb{P}}_{\mathcal{P}t} \sim N(0, I_N)$ and $\Sigma_{\mathcal{P}\mathcal{P}}$ is an $N \times N$ nonsingular matrix. We close the model by specifying the dynamics of $\mathcal{P}_t$ under the pricing (risk-neutral) distribution $\mathbb{Q}_t$

$$\mathcal{P}_t = K^{\mathbb{Q}}_{0\mathcal{P}} + K^{\mathbb{Q}}_{\mathcal{P}\mathcal{P}}\mathcal{P}_{t-1} + \sqrt{\Sigma_{\mathcal{P}\mathcal{P}}}\epsilon^{\mathbb{Q}}_{\mathcal{P}t}. \tag{2.5}$$

Under the above assumptions and the absence of arbitrage opportunities, the yield

---

[6] Our model can be easily extended to a setup with unspanned macro risks where Z_t includes the information of macroeconomic risk factors M_t in addition to portfolio risk factors $\mathcal{P}_t$.

[7] Without loss of generality, we rotate the N risk factors to make normalization. Accordingly, $\mathcal{P}_t$ corresponds to the N portfolios of bond yields; for example, $\mathcal{P}$ can be the first N principal components (PCs) of bond yields. Joslin, Singleton and Zhu (2011) show that the rotation is normalized so that the parameters governing the $\mathbb{Q}$ distribution of yields, i.e. $(\rho_{0\mathcal{P}}, \rho_{1\mathcal{P}}, K^{\mathbb{Q}}_{0\mathcal{P}}, K^{\mathbb{Q}}_{\mathcal{P}\mathcal{P}})$ are fully determined by the parameter set $(\Sigma_{\mathcal{P}\mathcal{P}}, \lambda^{\mathbb{Q}}, r^{\mathbb{Q}}_{\infty})$, where $\lambda^{\mathbb{Q}}$ denotes the N-vector of ordered nonzero eigenvalues of $K^{\mathbb{Q}}_{\mathcal{P}\mathcal{P}}$ and $r^{\mathbb{Q}}_{\infty}$ denotes the long-run mean of r_t under $\mathbb{Q}$.

[8] This representation, which characterizes the dynamics of the full set of N risk factor, can be viewed as the companion form of a higher-order VAR of the state vector of risk factors.

on an m-period bond, for any $m > 0$, is an affine function of $\mathcal{P}_t$,

$$y_t^m = A_{\mathcal{P}}(m) + B_{\mathcal{P}}(m)\mathcal{P}_t, \tag{2.6}$$

where the loadings $A_{\mathcal{P}}(m)$ and $B_{\mathcal{P}}(m)$ govern the $\mathbb{Q}$ distribution of yields.[9] The detailed expressions of the loadings can be found in Appendix A.

The scaled market prices of risk are also affine functions of $\mathcal{P}_t$,

$$\Sigma_{\mathcal{P}\mathcal{P}}^{1/2}\Lambda_{\mathcal{P}}(\mathcal{P}_t) = \Lambda_0 + \Lambda_1\mathcal{P}_t, \tag{2.7}$$

where $\Lambda_1 = K_{\mathcal{P}\mathcal{P}}^{\mathbb{P}} - K_{\mathcal{P}\mathcal{P}}^{\mathbb{Q}}$ is an $N \times N$ matrix and $\Lambda_0 = K_{0\mathcal{P}}^{\mathbb{P}} - K_{0\mathcal{P}}^{\mathbb{Q}}$ is an $N \times 1$ vector.

2.2 Learning and Model Uncertainty

From the last section, we can see that the bond-market-specific pricing kernel $\mathcal{M}_{\mathcal{P},t+1}$ is a function of priced risks $\mathcal{P}$ and a set of parameters $\Theta \equiv (\Theta^{\mathbb{P}}, \Theta^{\mathbb{Q}})$ that govern the dynamics under the physical measure $\mathbb{P}$ and risk-neutral measure $\mathbb{Q}$. To be more specific, the parameter set $\Theta^{\mathbb{P}} \equiv (K_{0\mathcal{P}}^{\mathbb{P}}, K_{\mathcal{P}\mathcal{P}}^{\mathbb{P}})$ governs the drift of $\mathcal{P}$ under the physical measure, whereas the set $\Theta^{\mathbb{Q}} \equiv (\Sigma_{\mathcal{P}\mathcal{P}}, \lambda^{\mathbb{Q}}, r_\infty^{\mathbb{Q}})$ determines the risk-neutral dynamics, i.e. the pricing distribution. Note that the variance matrix $\Sigma_{\mathcal{P}\mathcal{P}}$ in fact enters both the physical and risk-neutral dynamics, which can be estimated from Equation (2.4) that describes the physical dynamics of pricing factor $\mathcal{P}$.

We consider a representative agent who can adaptively learn about the evolution of the state of the economy. He or she may have different perceptions of the pricing kernel $\mathcal{M}_{\mathcal{P},t+1}$ at different points in time.[10] As mentioned in Evans and Honkapohja (2001), the concept of *adaptive learning* (AL) introduces a specific form of bounded rationality, and provides a means of approximating agents' expectations that incorporates learning as well as a rationale for rational expectations. Based on the learning concept, we rewrite the evolution of the one-period pricing kernel under the physical measure $\mathbb{P}$, conditional

[9] As we will see in the next section, the loadings are known functions of parameters $\Theta^{\mathbb{Q}} \equiv (\Sigma_{\mathcal{P}\mathcal{P}}, \lambda^{\mathbb{Q}}, r_\infty^{\mathbb{Q}})$.

[10] Similar to Giacoletti, Laursen and Singleton (2014), we consider a model with a reduced-form pricing kernel, which does not clearly specify agents' preferences when compared with preference-based models such as Piazzesi and Schneider (2007) and Collin-Dufresne, Johannes and Lochstoer (2013). Nevertheless, as shown in Duffie (2001) and Piazzesi (2010), we can link the pricing equation to fundamentals within a representative agent endowment economy where preference parameters are specified.

on the information at time t, as

$$\mathcal{M}_{t,t+1} = E_t^{\mathbb{P}}[f_{\mathcal{M}}(\Theta_t, \mathcal{P}_{t+1})|\mathcal{P}_t] = F_{\mathcal{M}}(\Theta_t, \mathcal{P}_t). \tag{2.8}$$

Therefore, the price D_t^m of a zero-coupon bond issued at date t and maturing at date m, is also a function of $(\Theta_t, \mathcal{P}_t)$ under the physical measure $\mathbb{P}$

$$D_t^m = E_t^{\mathbb{P}}[\prod_{s=1}^{m} \mathcal{M}_{t+s-1,t+s}] = F_{D_m}(\Theta_t, \mathcal{P}_t). \tag{2.9}$$

To simplify the estimation problem in our learning system, we have the following assumptions:

- The portfolio risk factors $\mathcal{P}_t$ are measured without errors.
- The parameters Θ, which may evolve over time, are unknown to the agent, and hence, need to be estimated statistically at each point in time t.
- The risk of unknown parameters is not priced.

These assumptions are standard in the literature of term structure pricing or learning, see Joslin, Singleton and Zhu (2011), Joslin, Priebsch and Singleton (2014) and Giacoletti, Laursen and Singleton (2014). With these mild assumptions, we can partition the parameter set Θ_t into subsets $\Theta^{\mathbb{P}}$ and $\Theta^{\mathbb{Q}}$ and estimate them respectively.

For the physical dynamics, we consider a case where the agent believes the law of motion (*perceived law of motion*) of parameters $\Theta^{\mathbb{P}}$ is a random walk process, and then Equation (2.4) becomes

$$\mathcal{P}_t = K_{t,0\mathcal{P}}^{\mathbb{P}} + K_{t,\mathcal{P}\mathcal{P}}^{\mathbb{P}} \mathcal{P}_{t-1} + \sqrt{\Sigma_{t,\mathcal{P}\mathcal{P}}} \epsilon_{\mathcal{P}t}^{\mathbb{P}}, \tag{2.10}$$

$$\begin{bmatrix} K_{t,0\mathcal{P}}^{\mathbb{P}} \\ \mathrm{vec}(K_{t,\mathcal{P}\mathcal{P}}^{\mathbb{P}}) \end{bmatrix} = \begin{bmatrix} K_{t-1,0\mathcal{P}}^{\mathbb{P}} \\ \mathrm{vec}(K_{t-1,\mathcal{P}\mathcal{P}}^{\mathbb{P}}) \end{bmatrix} + u_t, \tag{2.11}$$

where $\mathrm{vec}(\cdot)$ means the vectorization of a matrix and u_t is a vector of transition errors. The above system can be estimated using a (Bayesian) Kalman filter.

For the pricing dynamics, the *perceived law of motion* of parameters $(r_\infty^{\mathbb{Q}}, \lambda^{\mathbb{Q}})$ in $\Theta^{\mathbb{Q}}$

is also a random walk process. We rewrite Equation (2.6) as

$$y_t^m = A_{\mathcal{P}}^m(\Sigma_{t,\mathcal{P}\mathcal{P}}, \lambda_t^{\mathbb{Q}}, r_{t,\infty}^{\mathbb{Q}}) + B_{\mathcal{P}}^m(\lambda_t^{\mathbb{Q}})\mathcal{P}_t, \tag{2.12}$$

$$\begin{bmatrix} r_{t,\infty}^{\mathbb{Q}} \\ \lambda_t^{\mathbb{Q}} \end{bmatrix} = \begin{bmatrix} r_{t-1,\infty}^{\mathbb{Q}} \\ \lambda_{t-1}^{\mathbb{Q}} \end{bmatrix} + \mathrm{u}_t, \tag{2.13}$$

where $\Sigma_{t,\mathcal{P}\mathcal{P}}$ is estimated from Equation (2.10) and u_t is a vector of transition errors, see Joslin, Singleton and Zhu (2011) for technical details. We estimate the above nonlinear system with the unscented Kalman filter.[11]

2.2.1 Learning Rules

Let us start with the physical dynamics. For a more convenient description, we rewrite the learning dynamics (2.10) and (2.11) under the physical measure as a form of p-lag time-varying parameter vector autoregression (TVP-VAR)[12]

$$z_t = X_t\beta_t + v_t^{\mathbb{P}}, \tag{2.14}$$

$$\beta_{t+1} = \beta_t + u_t, \tag{2.15}$$

where $z_t = \mathcal{P}_t$, $X_t = I_N \otimes \left[z'_{t-1}, ..., z'_{t-p}\right]$, $\beta_t = \left[c_t, vec\,(B_{1t})', \cdots, vec\,(B_{pt})'\right]'$ is a vector summarizing all VAR coefficients, $v_t^{\mathbb{P}} \sim N\,(0, \Sigma_t)$ with Σ_t an $n\times n$ measurement covariance matrix, and $u_t \sim N\,(0, Q_t)$ with an $n \times n$ transition covariance matrix.

As we have mentioned, the system can be solved by means of the Kalman filter, see Appendix B.1 for details. The solution for this system follows a recursive rule given by

$$\beta_t|D_t \sim N\,(m_t, \Phi_t)\,, \tag{2.16}$$

where D_t is the information set at time t. The solution is equivalent to a special case of the class of *adaptive least squares* (ALS) learning proposed by McCulloch (2007), which also nests the *ordinary least squares* (OLS) and *constant gain least squares* (CGLS) algorithm introduced by Sargent (2002) and Evans and Honkapohja (2001). The ALS formulas are

[11] For identification one can fix $r_\infty^{\mathbb{Q}} = 0$, see for example Dai and Singleton (2000), Christensen, Diebold and Rudebusch (2011) and Joslin, Singleton and Zhu (2011).

[12] Note that p is usually set to 1 in most of the no-arbitrage affine term structure models.

given by

$$m_t = m_{t|t-1} + R_t^{-1} X_t' \Sigma_t^{-1} \widetilde{v}_t, \tag{2.17}$$
$$R_t = (1 - \gamma_t) R_{t-1} + X_t' \Sigma_t^{-1} X_t, \tag{2.18}$$

where $\widetilde{v}_t = z_t - X_t m_{t|t-1}$ is the prediction error and γ_t is the gain parameter which belongs to interval $[0, 1)$. Note that apart from the learning gain, stochastic volatility also plays a role in controlling the informativeness of incoming information flows, which parallels the finding of Cieslak and Povala (2015*b*) that stochastic volatility has a non-trivial effect on the conditional distribution of interest rates.

By setting the gain parameter to different values, we have different learning algorithms or rules:

- **Learning Rule 1**: When $\gamma_t = Q_t(\Phi_{t-1} + Q_t)^{-1}$ and $R_t = \Phi_t^{-1}$, the learning algorithm is the most general case of the ALS, i.e. the standard Kalman filter solution.[13]

- **Learning Rule 2**: When γ_t is replaced by a sufficiently small constant, as in Sargent (2002) and Evans and Honkapohja (2001), the learning rule becomes the *constant gain least squares* (CGLS) algorithm. This case is also consistent with the 'forgetting factor' algorithm proposed by Koop and Korobilis (2012, 2013), see Appendix B.1.

- **Learning Rule 3**: When $\gamma_t = 0$, the learning algorithm becomes the *recursive least squares* (RLS), i.e. a recursive form of *ordinary least squares* (OLS).

We will focus on the last two learning rules. In learning rule 3, with $\gamma = 0$ we immediately get $Q_t = 0$ (Appendix B.2), so the ALS degenerates to a constant parameter case, or a *decreasing gain* case in Evans and Honkapohja (2001). When there are no structural changes, m_t will converge to the true value when $t \to \infty$. However, when compared with the constant gain case, the decreasing gain case has a lower convergence speed. Moreover, gain sequences decrease to zero in the constant gain case when $t \to \infty$, so this model cannot sufficiently deal with structural changes. Therefore, we need to consider learning rule 2 with constant gains and face a trade-off: A larger constant gain is better at tracking changes but at the cost of larger variance. Hence, similar to Sargent

[13] The derivation of this result is provided in Appendix B.2

(2002), we only consider small gains to avoid instability.[14]

For the nonlinear system of the pricing dynamics, we can still write the rules of *adaptive learning* similar to the formulas of the Kalman filter, see Appendix B.3.

2.2.2 Model Uncertainty and Dynamic Model Selection

On top of adaptive learning, an agent may also have a set of possible models because of insufficient histories of data. A robust model needs to take this model uncertainty into account. In this paper, we consider model uncertainty regarding *physical dynamics* from two perspectives, both of which are closely related the predictability.[15] The first issue is the speed of learning. We can specify different values for the gain parameter γ, which controls the time variability of regression coefficients. A model with a small gain parameter would not be sensitive to new information, which means the agent slowly learns about structural changes. In an extreme case when γ is set to zero, the model boils down to a constant-coefficient case so the agent assumes there would not be structural breaks. The second issue we are concerned with regards the restrictions on the physical dynamics, which corresponds to the persistence of pricing factors. As Duffee (2011) and Joslin, Priebsch and Singleton (2014) suggest high persistence may boost the predictive performance, we incorporate this point in a time-varying manner.

In a time-varying framework, when implementing joint estimation of coefficients and model probabilities for $k = 1, ..., K$ models, it means that we need to estimate the following sum:

$$p(\beta_{t-1}|D_{t-1}) = \sum_{k=1}^{K} p(\beta_{t-1}^{(k)}|L_{t-1} = k, D_{t-1})\mathbf{Pr}(L_{t-1} = k|D_{t-1}), \tag{2.19}$$

where $L_{t-1} = k$ means the k_{th} model is selected at time $t-1$ and $p(\beta_{t-1}^{(k)}|L_{t-1} = k, D_{t-1})$ is given by the Kalman filter. Technical details regarding the computation of the above quantities are left to Appendix B.4 and are explained in detail in Koop and Korobilis (2012, 2013). We implement a dynamic model selection (DMS) approach that chooses the best model with the highest probability at any point in time, in order to obtain the optimal restrictions the representative agent imposes in a time-varying manner.

[14]We also need the gain to be sufficiently small in order to ensure convergence, see Evans and Honkapohja (2001) or more technically, Benveniste, Métivier and Priouret (1990).

[15]In light of the argument in Joslin, Singleton and Zhu (2011), we focus on physical dynamics only as the predictive performance of pricing factors is unrelated to pricing dynamics in our setup.

2.2.3 Decomposition of the Sources of Uncertainty

Following Dangl and Halling (2012), we conduct the following variance decomposition from the law of total variance. Through the decomposition, we aim to understand all possible sources of uncertainty with respect to the prediction of our pricing factor $\mathcal{P}$.

Firstly, we can decompose the variance with respect to different choices of learning gain parameter γ:

$$Var(\mathcal{P}) = E_\gamma\Big(Var(\mathcal{P}|\gamma)\Big) + Var_\gamma\Big(E(\mathcal{P}|\gamma)\Big), \tag{2.20}$$

where the operators $E_\gamma(\cdot)$ and $Var_\gamma(\cdot)$ are the expectation and variance with regards to γ, respectively. The former term in the above equation can be further decomposed with respect to different choices of forecasting model L:

$$E_\gamma\Big(Var(\mathcal{P}|\gamma)\Big) = E_L\Big(Var(\mathcal{P}|L,\gamma)\Big) + Var_L\Big(E(\mathcal{P}|L,\gamma)\Big). \tag{2.21}$$

After some algebra and using the expressions detailed in previous sections and Appendix B, we have

$$\begin{aligned} Var(\mathcal{P}_{t+1}) = & \underbrace{\sum_j \Big[\sum_k (\Sigma_t|L_k,\gamma_j,D_t)P(L_k|\gamma_j,D_t)\Big]P(\gamma_j|D_t)}_{\text{Observational variance}} \\ & + \underbrace{\sum_j \Big[\sum_k (X_t\Phi_{t|t-1}X_t'|L_k,\gamma_j,D_t)P(L_k|\gamma_j,D_t)\Big]P(\gamma_j|D_t)}_{\text{Parameter uncertainty}} \\ & + \underbrace{\sum_j \Big[\sum_k (\hat{\mathcal{P}}^j_{t+1,k} - \hat{\mathcal{P}}^j_{t+1})^2 P(L_k|\gamma_j,D_t)\Big]P(\gamma_j|D_t)}_{\text{Model restriction uncertainty}} \\ & + \underbrace{\sum_j (\hat{\mathcal{P}}^j_{t+1} - \hat{\mathcal{P}}_{t+1})^2 P(\gamma_j|D_t)}_{\text{Learning speed uncertainty}}, \end{aligned} \tag{2.22}$$

where Σ_t denotes the variance of the disturbance term in the observation equation, $\Phi_{t|t-1}$ denotes the unconditional variance of the time-t prior of the coefficient vector β_t, $\hat{\mathcal{P}}^j_{t+1}$ is the weighted average conditional on γ_j and $\hat{\mathcal{P}}_{t+1}$ is the weighted average over all candidate models.

The individual terms of Equation (2.22) state the sources of prediction uncertainty and have intuitive interpretations. The first term measures the expected observational variance, calculated over different choices of learning gain γ and forecast model L. This term in fact captures the random fluctuations or risks in the pricing factors, relative to the predictable drift component. The second term is the expected variance from errors in the estimation of the coefficient vector, which can be interpreted as the source of estimation or parameter uncertainty. The third term captures model uncertainty with respect to model restrictions. The last term measures the uncertainty with respect to the learning speed, which can also be considered as the time variability of the model coefficients.

2.3 Portfolio Allocation under Uncertainty

In the last section we describe the term structure pricing model allowing for parameter and model uncertainty, but the uncertainty is not priced for the representative agent. That is to say, no matter how many models are available, provided a model estimated and selected by the agent ex post, there is no uncertainty but only interest-rate or inflation risk.[16] Investors may rebalance the portfolio because of speculation or hedging demand, but it is hard to tell whether the term premia accounts for the uncertainty or not, and to what degree. In the case where the representative agent truly requires compensation for the uncertainty, the market prices of risk may be overestimated.[17] Therefore, the model that does not explicitly take uncertainty premia into account can cause some anomalies, for example, high Sharpe ratios suggested by Duffee (2010). This can be explained by the inability to separate the uncertainty premia from the risk premia, see Knight (1921). We do not intend to decompose the term premia into risk premia and uncertainty premia in this paper, but we are interested in whether allowing for uncertainty aversion can increase economic value for a small short-term investor with no market impact.

The aversion to uncertainty is essential when we consider a short-term investment by holding a long-term bond for a relatively short period say one year, as Sangvinatsos and Wachter (2005) and Johannes, Korteweg and Polson (2013) suggest that failing to hedge out the uncertainty carries a high utility cost. A classical, or maybe naive, short-term investor who is given only one pricing model and who does not consider parameter uncertainty, can end up with an investment strategy with high volatility and has little

[16]This implies long-term investors do not perceive high uncertainty, because once an investment decision is made, they do not rebalance the portfolio frequently. Sangvinatsos and Wachter (2005) show that investors with long investment horizons indeed take extreme long positions in long-term bonds because of hedging demands. If a bond is held to maturity, the expected return is fixed and irrelevant to the model, given that the U.S. treasury bonds are usually considered non-defaultable. The long-term institutional investors hold major share of the U.S. bond market and hence has high market power.

[17]This in fact is a small-sample problem, which can be resolved with very long histories of data, as we can recover the true model with learning and dynamic model selection.

economic value. In contrast, a sophisticated mean-variance investor will consider a robust strategy because he or she is averse to parameter and model uncertainty.[18]

2.3.1 A Mean-Variance Portfolio Model with Parameter Uncertainty Aversion

To begin with, we consider the classical mean-variance model proposed by Markowitz (1952) and Sharpe (1970), where the optimal portfolio weight of M^r risky assets, w, is given by the solution of the following optimization problem:[19]

$$\max_{w} w^{\intercal}\mu - \frac{\boldsymbol{\gamma}}{2} w^{\intercal}\Sigma w, \tag{2.23}$$

where μ is the M^r-vector of the true expected excess returns over the risk-free asset, Σ is the $M^r \times M^r$ covariance matrix of excess returns, and the scalar $\boldsymbol{\gamma}$ is the risk aversion parameter. The solution to this problem is

$$w = \frac{1}{\boldsymbol{\gamma}} \Sigma^{-1}\mu. \tag{2.24}$$

However, an investor knows that the expected excess returns are from a model which may generate imprecise estimates of expected excess returns $\hat{\mu}$, and therefore, pursues robustness when determining the portfolio weights. The demand for robustness is equivalent to investors' aversion to the uncertainty associated with the parameters estimated, see Gilboa and Schmeidler (1989) and Chen and Epstein (2002). To explicitly account for the uncertainty aversion, we introduce two elements to the above optimization problem following Garlappi, Uppal and Wang (2007). Firstly, the investor recognizes that the expected excess return for each asset can lie within a specified interval of its estimated value. This implies that the point estimate of the expected excess return is not the only possible value considered by the investor. Secondly, we introduce an additional optimization: The investor minimizes over the choice of expected returns, subject to the constraint of the specified interval.

The max-min problem above originates from the model of Gilboa and Schmeidler

[18] Gargano, Pettenuzzo and Timmermann (2014) analyze the portfolio selection problem under uncertainty with power utility, but they do not consider robust control. Johannes, Korteweg and Polson (2013) suggest mean-variance utility is similar to power utility in absence of fat tails, so in this paper we only consider investors with mean-variance utility for simplicity.

[19] In order to keep the classical representation, the following equations in this section have abuse of notation $\boldsymbol{\gamma}$. Note that the notation $\boldsymbol{\gamma}$ in bold in this section means the risk aversion parameter, which is different from the learning gain parameter γ in previous sections.

(1989), which is given by the form

$$\max_{w} \min_{\mu} w^{\intercal}\mu - \frac{\gamma}{2} w^{\intercal}\Sigma w, \tag{2.25}$$

subject to

$$f_C(\mu, \hat{\mu}, \Sigma) \leq \epsilon. \tag{2.26}$$

To clarify the constraint (2.26), consider a case where the excess returns follow a multivariate Gaussian distribution with the true mean μ and the expected returns $\hat{\mu}$ are estimated by the sample mean with T observations. Then the quantity

$$T(\hat{\mu} - \mu)^{\intercal}\Sigma^{-1}(\hat{\mu} - \mu)$$

has a χ^2 distribution with M^r degree of freedom, where M^r is the dimension of the vector of returns.[20] Let $f_C = T(\hat{\mu} - \mu)^{\intercal}\Sigma^{-1}(\hat{\mu} - \mu)$ and ϵ be a chosen quantile for the χ^2 distribution. The constraint (2.26) can be expressed as

$$T(\hat{\mu} - \mu)^{\intercal}\Sigma^{-1}(\hat{\mu} - \mu) \leq \epsilon.$$

It means the constraint corresponds to a confidence interval in the probabilistic statement

$$P[T(\hat{\mu} - \mu)^{\intercal}\Sigma^{-1}(\hat{\mu} - \mu) \leq \epsilon] = 1 - p,$$

for an appropriate level p, say 5%.

Now we parameterize the constraint (2.26) as

$$(\hat{\mu} - \mu)^{\intercal}\Sigma^{-1}(\hat{\mu} - \mu) \leq \varepsilon. \tag{2.27}$$

Then the max-min problem (2.25) subject to constraint (2.27) can be simplified into a maximization problem and we can obtain an intuitive expression of the optimal portfolio weights. Garlappi, Uppal and Wang (2007) have the following proposition:

Proposition 1. *The max-min problem (2.25) subject to constraint (2.27) is equivalent*

[20] If Σ is not known, then the quantity $\frac{T(T-M^r)}{(T-1)M^r}(\hat{\mu} - \mu)^{\intercal}\hat{\Sigma}^{-1}(\hat{\mu} - \mu)$ follows an F distribution with M^r and $T - M^r$ degrees of freedom, see Garlappi, Uppal and Wang (2007).

to the following maximization problem

$$\max_{w} w^{\intercal}\hat{\mu} - \frac{\gamma}{2} w^{\intercal}\Sigma w - \sqrt{\varepsilon w^{\intercal}\Sigma w}. \tag{2.28}$$

The optimal portfolio weights for this problem can be expressed as

$$w^* = \frac{1}{\gamma}\Sigma^{-1}\Big(\frac{1}{1+\frac{\sqrt{\varepsilon}}{\gamma\sigma_p^*}}\Big)\hat{\mu}, \tag{2.29}$$

where σ_p^* is the variance of the optimal portfolio that can be obtained from solving a second degree polynomial equation, see Appendix C.1.

This framework nests the classical mean-variance portfolio. When $\varepsilon \to 0$, we immediately obtain Equation (2.24). Without loss of generality, the framework of parameter uncertainty aversion measures the effect of uncertainty aversion with respect to rare events, as indicated in Liu, Pan and Wang (2005). This means investors have robust control for rare events and allow for the worst-cast scenario that rare disasters actually happen.[21] With higher value of ε investors become more pessimistic. When $\varepsilon \to \infty$, investors become so pessimistic that they would not invest on any risky assets, which in turn gives a minimum-variance portfolio.

2.3.2 An Extension with Model Uncertainty

In this section, we extend the optimization problem (2.25) to characterize model uncertainty. In Equation (2.25) we only use one model to generate expected excess returns $\hat{\mu}$, without the freedom of selecting alternative models. Suppose we have a set of candidate models, then the max-min problem becomes

$$\max_{w,\hat{\mu}} \min_{\mu} w^{\intercal}\mu - \frac{\gamma}{2} w^{\intercal}\Sigma w, \tag{2.30}$$

subject to

$$(\hat{\mu}-\mu)^{\intercal}\Sigma^{-1}(\hat{\mu}-\mu) \leq \varepsilon, \tag{2.31}$$

$$\hat{\mu} \in \{\hat{\mu}_k : k = 1, ..., K\}, \tag{2.32}$$

[21] To see this, recall that ε has a probabilistic interpretation. Our max-min problem mimics investors' 'pessimism' that they assume the worst-case scenario will always happen when making investment decisions.

where $\hat{\mu}$ can be chosen from a set of K candidate models. The above max-min problem can also be simplified into a maximization problem which is easier to solve. Extending the results of Garlappi, Uppal and Wang (2007), we have the following proposition:

Proposition 2. *The max-min problem (2.30) subject to constraint (2.31) and (2.32) is equivalent to the following maximization problem*

$$\max_{w,\hat{\mu}} w^{\intercal}\hat{\mu} - \frac{\gamma}{2}w^{\intercal}\Sigma w - \sqrt{\varepsilon w^{\intercal}\Sigma w}, \tag{2.33}$$

subject to $\hat{\mu} \in \{\hat{\mu}_k : k = 1, ..., K\}$*. The optimal portfolio weights for this problem can be expressed as*

$$w^{*}* = \frac{1}{\gamma}\Sigma^{-1}\Big(\frac{1}{1+\frac{\sqrt{\varepsilon}}{\gamma\sigma_p^*}}\Big)\hat{\mu}^*, \tag{2.34}$$

where $\hat{\mu}^*$ *is the optimal expected excess return selected from* K *candidate models, and* σ_p^* *is the variance of the optimal portfolio that can be obtained from solving a second degree polynomial equation, see Appendix C.2.*

We can see how this extension contributes to investors' portfolio allocation in an intuitive way. Equation (2.31) shows we expand the feasible region in our minimum optimization problem, which is the same as the case of parameter uncertainty aversion. However, with condition (2.32), we then shrink the admissible region to the area associated with the optimal forecasts generated from candidate models. Indeed the region we search is expanded for the minimum optimization, but we only select the weights that give us higher utility in the maximum optimization. This refinement helps investors avoid unrealistic pessimistic investment, especially during the period when the minimum-variance strategy performs poorly. Even in a conservative situation where ε is large, investors still intend to hold some risky assets. It is indeed more realistic: An investor can hedge out risks by diversification according to a selected forecasting model, instead of being extremely pessimistic and only invest in risk-free assets.

3 Results

In this paper, we use the smoothed US bond yields provided from the US Federal Reserve by Gürkaynak, Sack and Wright (2007). We also include the 3- and 6-month Treasury Bills (Secondary Market Rate) from St. Louis Federal Reserve Economic Data (FRED).

The empirical analysis focuses on yields with 12 maturities of 3, 6, 12, 24, 36, 48, 60, 72, 84, 96, 108 and 120 months. The full sample of end of the month yield data is from June 1961 to October 2014. Our training sample has 121 observations from the beginning, up to and including June 1971. We do not introduce any other variables in addition to bond yield data, because we aim to understand the predictability that purely comes from the information in the bond market. From a finance viewpoint, we aim to explore all sources of the prediction uncertainty and how investors can benefit from a robust model which takes these sources of uncertainty into account.[22]

3.1 Pricing Dynamics and Market Prices of Risk

In our pricing setup, we specify a parsimonious factor structure so that a few portfolio risk factors can effectively model the term structure. Three risk factors – *Level*, *Slope* and *Curvature* – can capture most of the variance of bond yields, see Nelson and Siegel (1987) and Litterman and Scheinkman (1991). In line with most of the literature, we also use these three pricing factors or portfolios to price bonds when specifying our model. The portfolio weights are fixed for consistency and tractability.[23] Following Joslin, Singleton and Zhu (2011), we assume our portfolio risk factors $\mathcal{P}$ are measured without errors. Given this assumption, the cross-sectional arbitrage-free restrictions are irrelevant to the conditional point forecasts of $\mathcal{P}$ under $\mathbb{P}$. Therefore, we can separately estimate pricing dynamics, provided that the covariance of the innovations of $\mathcal{P}$ is estimated from physical dynamics.

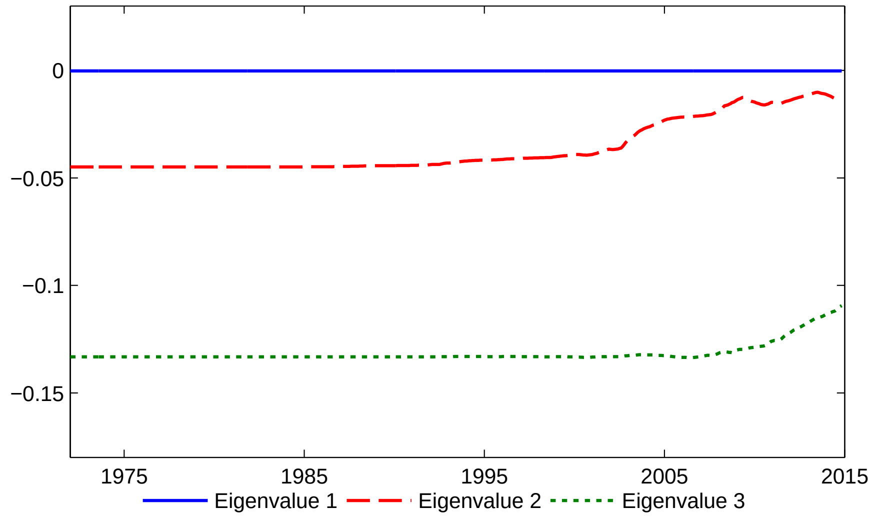
Figure 1 displays the evolution of three real eigenvalues $\lambda^{\mathbb{Q}}$ associated with $K^{\mathbb{Q}}_{\mathcal{PP}}$ over time. The first eigenvalue is slightly below zero, which implies that the Level factor is very persistent and close to a unit root process. The second eigenvalue is smaller but it still implies a highly persistent process of the Slope factor. The third eigenvalue is the smallest among the three, suggesting a less persistent evolution of the Curvature factor. We can see that the eigenvalues are stable over time, and therefore the arbitrage-free relation that specifies the link between market prices of risk and risk factors is unlikely

[22] Ludvigson and Ng (2009) and Joslin, Priebsch and Singleton (2014) suggest that unspanned macro information has predictive power for future movements of bond yields, whilst Bauer and Rudebusch (2015) provide evidence that some key macroeconomic variables are indeed spanned by bond yields. It is interesting to develop hybrid models with both spanned and unspanned macroeconomic risks, and explore the prediction uncertainty from different choices of predictors. We do not pursue this direction in this paper, although our framework can be easily extended to allow for the unspanned macro risks or hybrid models.

[23] Our findings are robust to different approximation methods of the portfolio weights and pricing factors.

to have a significant change. The above findings are consistent with Joslin, Priebsch and Singleton (2014) and Giacoletti, Laursen and Singleton (2014). Our new finding is that the factor process in pricing dynamics becomes more persistent in the recent decade, which implies a relatively more flat forward curve. The second eigenvalue is gradually growing up since the middle of the 2000s, while the third eigenvalue is rising from the start of the financial crisis.

Figure 1: Eigenvalues $\lambda^{\mathbb{Q}}$



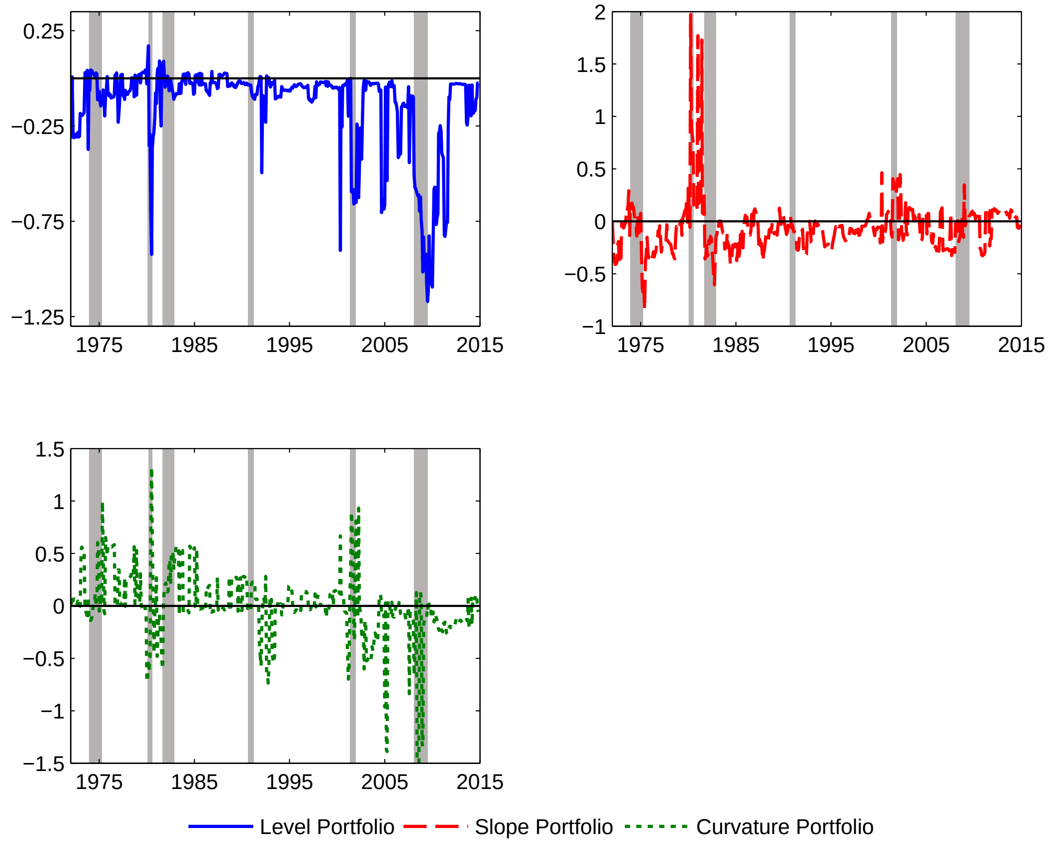
Notes: This graph shows the estimates of time-varying parameters $\lambda^{\mathbb{Q}}$ associated with $K^{\mathbb{Q}}_{\mathcal{PP}}$, which govern the loading matrix $B^m_{\mathcal{P}}(\lambda^{\mathbb{Q}}_t)$ in Pricing Equation (2.13). Sample period is from 1971:07 to 2014:10.

In this framework, we can gain knowledge about the priced risk by looking into Equation (2.7). Joslin, Priebsch and Singleton (2014) show that to a first-order approximation, the three elements of the scaled prices of risk represent the expected excess returns of three factor-mimicking portfolios, respectively. To be more specific, the excess return on a factor-mimicking portfolio, say Level, changes locally one-to-one with changes in the corresponding factor, but whose value is unresponsive to changes in other factors. Figure 2 depicts the one-period expected excess returns of Level, Slope and Curvature factor-mimicking portfolios.[24] Exposures to Level lose money if rates are expected to fall, which is usually when monetary policy is eased, for example, the recession periods.

[24] We relax the zero restrictions imposed on the price of the Curvature risk by Joslin, Priebsch and Singleton (2014) and Giacoletti, Laursen and Singleton (2014). This relaxation does not affect the power of in-sample fitting or out-of-sample forecasting in our framework. Further discussion is followed in the next section.

Exposures to Slope lose (gain) money if the curve is going to steepen (flatten), which is connected with the changes in the stance of monetary policy or investors' expectations, e.g. the monetarist experiment during the early 1980s. Exposures to Curvature may be difficult to interpreted, but Litterman, Scheinkman and Weiss (1991) link curvature to the volatility of the Level factor via the argument of yield curve convexity. We also find that the expected returns of Level portfolio heavily drop to historical low in the global financial crisis, which potentially reflects the 'flight-to-quality' demand suggested by Christensen, Lopez and Rudebusch (2010) and Bauer, Rudebusch and Wu (2014).[25]

Figure 2: One-Period Expected Excess Returns of Factor-Mimicking Portfolios



Notes:
1. This figure displays the one-period expected excess returns of Level, Slope and Curvature factor-mimicking portfolios from 1971:07 to 2014:10, which can be obtained from Equation (2.7).
2. Shaded areas are recession periods based on the NBER Recession Indicators. The unit is percentage.

Additionally, we can assess the economic significance of three pricing factors by cal-

[25]Specifically, Bauer and Rudebusch (2015) indicate inflation measures are mainly correlated with the Level, 'measures of slack' are most closely correlated with the Slope, and growth measures are correlated most strongly with the Curvature. Similar evidence can also be found in Diebold and Rudebusch (2013).

culating their contribution to the variability of the pricing kernel. From Equation (2.2) we have

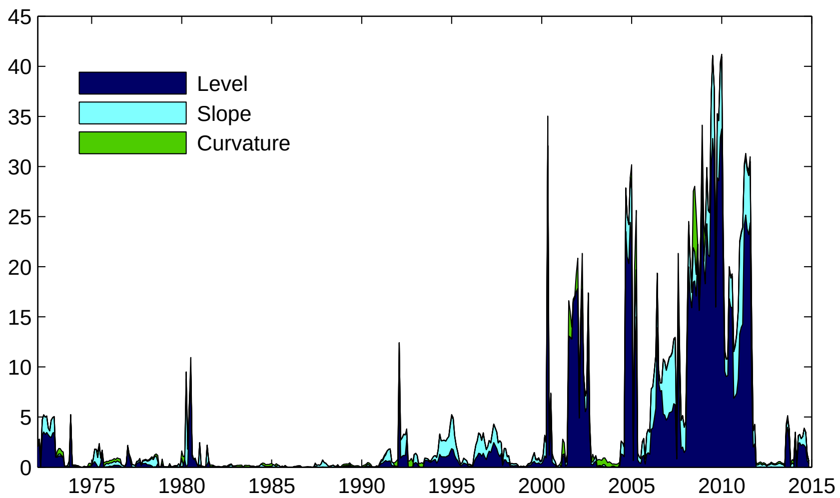
$$\ln \mathcal{M}_{\mathcal{P},t+1} = -r_t - \frac{1}{2}\Lambda'_{\mathcal{P}t}\Lambda_{\mathcal{P}t} - \Lambda'_{\mathcal{P}t}\epsilon^{\mathbb{P}}_{\mathcal{P},t+1}. \tag{3.1}$$

Following Adrian, Crump and Moench (2013), we decompose the conditional volatility of the pricing kernel into the contributions due to each price of risk according to

$$Var(\ln \mathcal{M}_{\mathcal{P},t+1}) = \Lambda'_{\mathcal{P}t}\Lambda_{\mathcal{P}t} = \sum_{j=1}^{N} \Lambda^2_{j,\mathcal{P}t}. \tag{3.2}$$

Figure 3 sets out the contribution of risk prices of all three factors to the time variation of the pricing kernel. We find that there are several peaking periods of the time variance after 2000, and the time variance is extremely high around the financial crisis, which means agents' expectations of excess returns are very uncertain at that time. Consistent with Adrian, Crump and Moench (2013), pricing kernel time variance is largely, though not exclusively, driven by the Level risk, which implies that the expected excess returns on the long-term bonds are also largely driven by the Level risk.

Figure 3: Pricing Kernel Variance Decomposition



Notes: This graph shows the decomposition of the conditional volatility of the pricing kernel by decomposing $\Lambda'_{\mathcal{P}t}\Lambda_{\mathcal{P}t}$ into three components corresponding to Level, Slope and Curvature risks. Sample period is from 1971:07 to 2014:10.

3.2 Physical Dynamics and Out-of-Sample Predictability

As mentioned above, the physical dynamics are crucial for term structure predictability. In the model setup, the agent is able to learn about the evolution of parameters over time. Specifically, we allow for both time-varying volatility and coefficients, as Johannes, Korteweg and Polson (2013) and Gargano, Pettenuzzo and Timmermann (2014) suggest these extensions are useful in capturing time-varying features and improving out-of-sample predictability. In addition to parameter uncertainty, our method also consider model uncertainty and therefore is robust to potential structural breaks, see Gürkaynak and Wright (2012).

We introduce two layers of model uncertainty. The first layer is the learning speed, which is controlled by the learning gain parameter γ discussed in Section 2.2.1. We define a wide grid of values for γ: $[0, 0.01, 0.02, 0.03, 0.04]$, which covers the last two learning rules. While $\gamma = 0$ is equivalent to the constant parameter case, $\gamma = 0.04$ gives us a very flexible model as the observation two years ago only receive 45% as much weight as the newly incoming observation. The learning speed characterizes how agents adjust to structural changes and form their expectations, and hence is related to the out-of-sample predictability.

The next layer of model uncertainty is about the restrictions we impose on the physical dynamics. The restrictions are motivated by the finding of Joslin, Singleton and Zhu (2011) and Duffee (2011) that cross-sectional restrictions are unrelated to the predictive performance.[26] Diebold and Li (2006) and Diebold, Rudebusch and Aruoba (2006) indicate that mixed evidence is found concerning the usefulness of various restrictions, where they consider both cases of related factors and unrelated factors. In our specification, we have in total 20 models, which nests the cases of related and unrelated factors in Diebold and Li (2006), as well as the random walk restrictions in Duffee (2011).[27] Combining two layers of model uncertainty, we have in total $5 \cdot 20 = 100$ models for selection at each

[26] Note that Joslin, Priebsch and Singleton (2014) impose restrictions on the market prices of risk which increase the persistence of the physical dynamics, and hence the out-of sample forecasting performance can be improved. Our flexible model selection framework in fact implicitly nests the restrictions of the same purpose, and therefore it is not necessary to explicitly impose zero restrictions on the market prices of risk.

[27] In addition to the unrestricted model, we restrict that the Level factor is unaffected by other two factors, so we have two zero restrictions in the first row; we further have $2^4 = 16$ combinations of zero restrictions imposed on off-diagonal elements of the remaining two rows of the coefficient matrix. We then have additionally three more models that ensure the first one, two and three factors follow random walks, respectively. Intuitively, these restrictions can enforce a high degree of persistence under $\mathbb{P}$ and hence may increase the forecast performance as suggested by Joslin, Priebsch and Singleton (2014).

point in time.[28] Actually, our method is robust to model specification and can mitigate the small sample bias indicated in Duffee and Stanton (2012) and Bauer, Rudebusch and Wu (2012), as the one 'true' model will always be selected with a long history of data.

3.2.1 Learning about the (un)predictability in the term structure

As mentioned in previous sections, our proposed term structure model with learning nests most of the current term structure models using conditional information at each point in time. In terms of the predictive performance of bond yields, we can safely focus on the conditional forecasts of pricing factors only in our framework. Joslin, Singleton and Zhu (2011) have shown that the cross-sectional no-arbitrage restriction is irrelevant for the conditional forecast of $\mathcal{P}$ under measure $\mathbb{P}$.

In this section, we compare the out-of-sample performance of the proposed model with two challenging benchmark models: random walk and the full-sample estimation following Joslin, Singleton and Zhu (2011).[29] Duffee (2002) remarks that it is hard for term structure models to beat the random walk, though the random walk cannot provide informative economic implications in terms of the dynamics of risk premia. The full-sample estimation of Joslin, Singleton and Zhu (2011) in fact is an in-sample forecasting exercise, which gains huge advantages as it incorporate the information of realized values. However, the full-sample estimation may be contaminated by the realized expectations of interest rates, which therefore do not perfectly reflect real-time expectations using conditional information.

Table 1 shows the predictive performance of the proposed model relative to benchmarks. The performance of the learning model is similar to the benchmark models, at least at shorter forecast horizons; the learning model even outperforms the benchmark models for some maturities. This is not surprising as conditional information should be helpful for short-horizon forecasts. Moreover, the short-rate forecasts, which is most related to future short rate expectations and term premium estimates, are relatively accurate even for longer horizons. Therefore, we can consider the term premium estimates of the learning model a plausible alternative to that of the model using full-sample information. However, a rather surprising result is that the performance of either the learning model or the benchmark using full-sample information, is close to that of ran-

[28] To speed up the estimation process, we employ the algorithm proposed by Koop and Korobilis (2013), see Appendix B for technical details.

[29] The model of Joslin, Singleton and Zhu (2011) is in fact nested within our framework if we focus on out-of-sample performance.

dom walk. This observation urges us to have a deeper understanding of the interest rate (un)predictability.

We explore the sources of prediction uncertainty by the variance decomposition noted in Section 2.2.3. The total prediction variance can be decomposed into observational variance, variance due to errors in the estimation of the coefficients (parameter uncertainty), variance due to model uncertainty in terms of the choice of the restrictions (restriction uncertainty), and variance due to model uncertainty regarding the choice of learning speed (learning speed uncertainty). Figure 4 displays the variance decomposition for three pricing factors, where Panel A shows that the predominant source of uncertainty is observational variance. This finding is consistent with the findings of Dangl and Halling (2012), as the asset prices frequently fluctuate randomly over their expected values. These fluctuations serve as the source of risk premia, and dominate the drift components in the term structure model. Therefore the fluctuations in fact contaminate the predictability of term structure models, especially during the periods when pricing factors are highly persistent.[30]

In Panel B of Figure 4, by excluding the observational variance we can focus upon the relative weights of the remaining sources of prediction uncertainty. The parameter uncertainty turns out to be the dominant source of prediction uncertainty, which implies parameter instability is another main source causing interest rate unpredictability. Therefore, a success forecasting model should at least consider the feature of time-varying parameters. The restriction uncertainty is less important but can be meaningful during certain periods. For example, restriction uncertainty rises steeply around the year 2003 and in the beginning of the financial crisis for the Slope factor. The uncertainty in learning speed is detectable but not of essence for most of the time. To highlight the importance of parameter uncertainty, Figure 5 sets out the persistence of the physical factor dynamics over time, which is examined by considering the behavior of the eigenvalues. There is a rising trend for the third eigenvalue since 1980s. We also detect significant drops in the persistence during recession periods, when the restrictions aiming to increase the persistence may not be valid.

Therefore, the large observational variance together with the high persistence in the data-generating process of bond pricing factors, gives rise to the similarity in the predictive performance between valid term structure models and the random walk. The

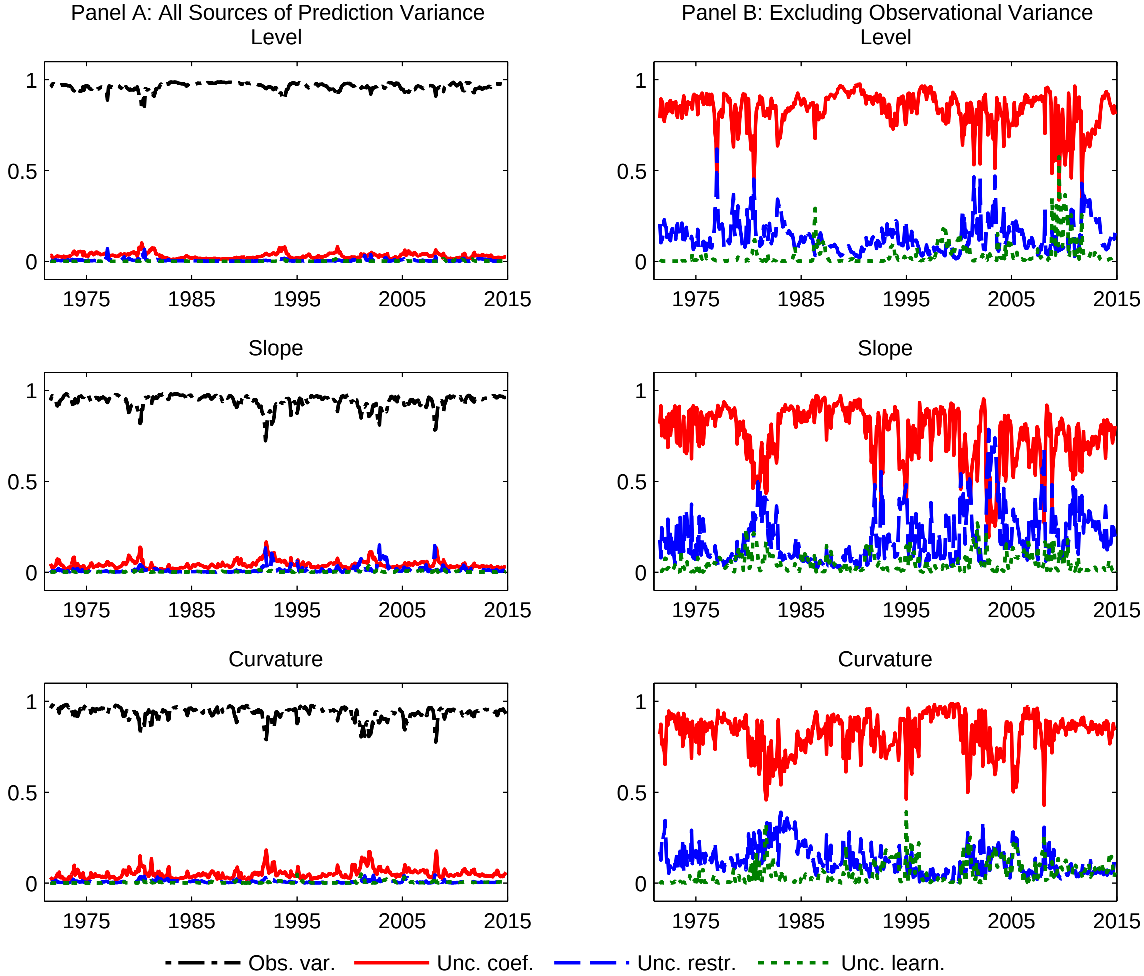
[30]This does not at all mean term structure models are not useful. For instance, term structure models can reveal informative dynamics of market prices of risks and have reliable term premia of long-term bonds, which can not be offered by the random walk model.

Table 1: Predictive Performance of the Learning Model Relative to Benchmarks

		FH \MA	3	6	12	24	36	48	60	72	84	96	108	120
MSPE	Learning versus RW	h=1	1.00	1.02	1.01	0.99	1.00	1.02	1.03	1.02	1.00	0.98	0.99	1.03
		h=3	0.99	1.01	0.98	1.00	1.01	1.02	1.02	1.02	1.01	1.01	1.01	1.02
		h=6	0.98	1.01	0.98	1.01	1.02	1.03	1.03	1.03	1.02	1.02	1.02	1.02
	Learning versus Full Sample	h=1	0.99	1.00	0.98	0.99	0.99	0.99	1.00	1.00	1.00	1.00	1.00	1.02
		h=3	1.00	1.01	0.99	1.01	1.01	1.02	1.02	1.03	1.03	1.03	1.04	1.05
		h=6	1.02	1.02	0.99	1.02	1.03	1.04	1.05	1.05	1.06	1.06	1.06	1.07
MAPE	Learning versus RW	h=1	1.01	1.03	1.06	1.01	1.01	1.02	1.02	1.02	1.01	1.01	1.01	1.02
		h=3	0.99	1.02	1.02	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.02	1.02
		h=6	1.00	1.02	1.01	1.01	1.00	1.00	1.01	1.01	1.01	1.01	1.01	1.01
	Learning versus Full Sample	h=1	1.00	1.01	1.00	1.00	1.00	1.01	1.01	1.02	1.01	1.01	1.01	1.02
		h=3	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.01	1.02	1.02	1.02	1.03
		h=6	1.02	1.02	1.01	1.01	1.01	1.01	1.01	1.02	1.02	1.02	1.03	1.03

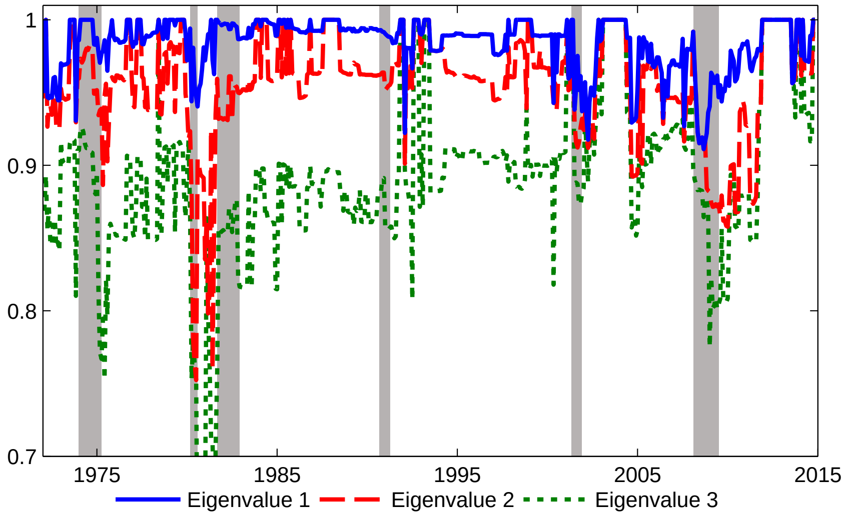
Notes:
1. This table shows 1-, 3- and 6-month ahead forecasts of bond yields with maturities ranging from 3 months to 120 months. The predictive duration is from 1971:07 to 2014:10.
2. The MSPE-based and MAPE-based statistics relative to the random walk and full-sample estimation are reported.
3. In this table, we use following abbreviations. MSPE: mean squared prediction error; MAPE: mean absolute prediction error; RW: random walk; Full Sample: the full-sample (1961:06-2014:10) estimation following Joslin, Singleton and Zhu (2011); MA: maturity; FH: forecast horizon.

Figure 4: Sources of Prediction Variance



Notes:
1. This figure displays the decomposition of the prediction variance with respect to different sources.
2. In Panel A, the prediction variance is split into observational variance (Obs. var.), variance caused by errors in the estimation of coefficients (Unv. coef.), variance caused by the uncertainty with respect to the choice of restrictions (Unc. restr.) and variance caused by the uncertainty with respect to the learning speed (Unc. learn.). The illustration shows the relative weights of these components.
3. Panel B masks out observational variance and shows relative weights of the remaining variance.

Figure 5: Eigenvalues $\lambda^{\mathbb{P}}$



Notes: This graph shows the time-varying eigenvalues of estimated $K^{\mathbb{P}}_{t,\mathcal{PP}}$ in Eq (2.10). Sample period is from 1971:07 to 2014:10. Shaded areas are recession periods based on the NBER Recession Indicators.

inability to beat the random walk does not mean no predictability in excess returns, as excess return predictability is about whether excess returns can be explained by any pricing factors. The random walk model can be viewed as a special case of term structure models in which pricing factors are extremely persistent, and in that case the excess returns can be predicted by the pricing factors.

Campbell and Shiller (1991) indicate that no predictability in excess bond returns is equivalent to the Expectations Hypothesis (EH), and Adrian, Crump and Moench (2013) show the realized excess returns can be decomposed into innovations and a predictable element. If innovations are at a reasonable level, we should be able to detect predictability in excess returns by capturing the factor dynamics. Actually, Cieslak and Povala (2015*a*) show term premia is spanned by pricing factors and excess returns are predictable when compared with the EH benchmark.[31] However, in the previous literature, it seems difficult to translate the predictability of term premia into significant economic value, see for example, Della Corte, Sarno and Thornton (2008). In the following sections, we will evaluate that whether allowing for different sources of uncertainty can contribute to economic gains over the EH when investors make portfolio allocations.

[31] The violation of the Expectations Hypothesis (EH) does not depend on the persistence of pricing factors, and hence the random walk model can also generate predictable excess returns if the loadings for short rates are not consistent with that for long rates.

3.3 Portfolio Selection

3.3.1 Predictability of excess returns

A simple approach to modeling the term structure is the Expectations Hypothesis (EH) that expected future short rates explain long rates. Campbell and Shiller (1991) indicates the empirical evidence fails to justify the strong form of Expectations Hypothesis and the idea that long-term interest rate are simply determined by the average of current and future expected short-term rates. However, EH could be resuscitated in weak form allowing for a constant term premia, consistent with an upward sloping yield curve. Based on the weak form of the Expectations Hypothesis, the long-term yield is average of expected future short term rates $y_t(\tau)^{EH}$ plus a constant risk premium C^{EH}:

$$y_t(\tau) = y_t(\tau)^{EH} + C^{EH}, \tag{3.3}$$

where the Expectations Hypothesis (EH) consistent bond yield $y_t(\tau)^{EH}$ is given by:

$$y_t(\tau)^{EH} = \frac{1}{\tau} \sum_{i=0}^{\tau-1} E_t y_{t+i}(1), \tag{3.4}$$

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

The belief in Expectations Hypothesis is closely related to investors' behavior. If the weak form of the Expectations Hypothesis holds, then risk premia is constant. In other words, we should not be able to predict the short-term returns in the future. If an investor believes that the Expectations Hypothesis does not hold and the term premium should be time-varying, then the investor can rely on a specific prediction model to guide his/her portfolio choice.

To show the above argument, we follow Cieslak and Povala (2015*a*) to decompose the *excess holding period return*. First, we define the *holding period return* as the return on buying a τ-year zero coupon bond at time t and then selling it, as a $(\tau - m)$-year zero coupon bond, at time $t + m$. This holding period return is given by:

$$HPR_{t+m}(\tau, m) = \frac{1}{m}[p_{t+m}(\tau - m) - p_t(\tau)] \tag{3.5}$$

where $p_t(\tau)$ is the log price of τ-year zero coupon bond at time t and $p_{t+m}(\tau - m)$ is the log price of $(\tau - m)$-year zero coupon bond at time $t + m$. The difference between holding period return and the m-year continuously compounded short yield is the *excess holding period return*:

$$EXR_{t+m}(\tau, m) = HPR_{t+m}(\tau, m) - y_t(m). \tag{3.6}$$

Note that a general form of term premium is given by

$$TP_t(\tau) = y_t(\tau) - y_t(\tau)^{EH}. \tag{3.7}$$

where $TP_t(\tau) \equiv C^{EH}$ if the EH holds. We can rewrite Equation (3.7) by relating the term premium to the *excess holding period return*:

$$TP_t(\tau) = \frac{1}{\tau} E_t[\sum_{i=0}^{\tau-2} EXR_{t+1+i}(\tau, 1)]. \tag{3.8}$$

where $E_t[\cdot]$ is the expectation operator. By the linearity of expectation, we can write the 1-period ahead expected *excess holding period return* as

$$E_t[EXR_{t+1}(\tau, 1)] = -(\tau - 1)E_t[TP_{t+1}(\tau - 1)] + \tau\, TP_t(\tau). \tag{3.9}$$

Therefore, it is not difficult to see that under the weak form of the Expectations Hypothesis, the m-period ahead expected *excess holding period return* is a constant:

$$E_t[EXR^{EH}_{t+m}(\tau, m)] = C^{EH}_m, \tag{3.10}$$

where C^{EH}_m is associated with the constant risk premium C^{EH} and often approximated by the historical average, see Rapach, Strauss and Zhou (2009) and Thornton and Valente (2012). But risk, and hence the term premia, is unlikely to be constant while underlying variables are changing. Cochrane and Piazzesi (2008) construct a test by regressing the excess bond returns on the forward rates and show that the forward rates have significant predictive power. Similar evidence can be found in Duffee (2002), Cochrane and Piazzesi (2005), Sarno, Thornton and Valente (2007), Tang and Xia (2007) and Gürkaynak and Wright (2012). In the case where the term premia is time-varying, the m-period *excess holding period return* is denoted by $xp_{t,m}$:

$$xp_{t,m} = E_t[EXR_{t+m}(\tau, m)] = -(\tau - m)E_t[TP_{t+m}(\tau - m)] + \tau\, TP_t(\tau). \tag{3.11}$$

It is straightforward to obtain $xp_{t,m}$ using a prediction model.

Although the EH is rejected by strong statistical evidence, it is puzzling that such predictability could not bring an improvement in economic utility of mean-variance investors, see Della Corte, Sarno and Thornton (2008), Thornton and Valente (2012) and Sarno, Schneider and Wagner (2014). Gargano, Pettenuzzo and Timmermann (2014) show that for investors with power utility and accounting for estimation error and model uncertainty, the predictability can be translated into higher economic value. Building upon previous literature, we consider a general framework that considers ambiguity aversion and nests ordinary risk aversion. It allows us to see if investors have significant improvement in their realized utility when considering potential sources of uncertainty.

3.3.2 Economic value

The evaluation of out-of-sample predictability does not consider the risk borne by an investor. It raises the issue of economic value of a forecasting model, as statistical significance does not measure its economic significance. Here we evaluate whether the model predictability is sufficiently large to be of economic value to risk-averse, or more generally, ambiguity-averse investors. Following Campbell and Thompson (2008), Welch and Goyal (2008), and Rapach, Strauss and Zhou (2009), we assume each investor, who is small and hence with no market impact, chooses portfolio weights based on the return forecasts. In this paper, we allow the investor to be able to invest in 10 fixed-income assets: 1- to 10-year US bonds. We then calculate realized utility gains for a mean-variance investor on a real-time basis.

To demonstrate the evaluation of the above strategies, we firstly discuss the case of an Expectations Hypothesis (EH) investor. We can compute the average utility for the mean-variance investor with relative risk aversion parameter γ_R who allocates his or her portfolio monthly among all assets using forecasts of the excess returns based on the historical average. This exercise requires the investor to forecast the variance of excess returns. Following Campbell and Thompson (2008), we assume that the investor estimates the variance matrix $\hat{\Sigma}_{t+1}^{-1}$ using a 5-year rolling window using monthly data of excess annually returns. A mean-variance investor who forecasts the excess bond returns using the historical average $\bar{r}_{t+1}$ will decide at the end of period t to allocate the following share of his or her portfolio to securities in period $t+1$:

$$w_{0,t} = \frac{1}{\gamma_R}\hat{\Sigma}_{t+1}^{-1}\bar{r}_{t+1} \tag{3.12}$$

where $\hat{\sigma}^2_{t+1}$ is the 5-year rolling-window estimate of the variance of excess returns.[32]

Over the out-of-sample period, the average of the realized utility of the investor is given by

$$\hat{v}_0 = \hat{\mu}_0 - (\frac{1}{2})\gamma_R \hat{\sigma}^2_0 \tag{3.13}$$

where $\hat{\mu}_0$ and $\hat{\sigma}^2_0$ are respectively the sample mean and variance of the excess holding period returns on the benchmark portfolio of the EH investor, which is constructed using forecasts of the excess returns based on the historical average.

Similarly, we can calculate the average utility for the same investor, when his or her decision is made by using a model j to forecast the excess bond returns. As it is noted in Section 2.3, we can construct the share $w_{j,t}$ based on the forecasts of model j.

The resulting realized average utility level is

$$\hat{v}_j = \hat{\mu}_j - (\frac{1}{2})\gamma_R \hat{\sigma}^2_j \tag{3.14}$$

where $\hat{\mu}_j$ and $\hat{\sigma}^2_j$ are the sample mean and variance of the excess holding period returns on the portfolio indexed by j. The investor forms the portfolio j using forecasts of the excess returns of bonds according to the jth forecasting model.

We can compute the utility gain, or certainty equivalent return, as the difference between $\hat{v}_j$ in Eq. (3.14) and $\hat{v}_0$ Eq. (3.13)

$$\Delta = \hat{v}_j - \hat{v}_0. \tag{3.15}$$

The utility gain that is expressed in average annualized percentage return can be interpreted as the portfolio management fee that an investor would be willing to pay to have access to the additional information available in a predictive model relative to the information in the historical term premia alone. We report results for risk aversion parameters $\gamma_R = 1, 3, 6$; the results are qualitatively similar for other reasonably values (ranging from 1 to 10).

[32]Following Campbell and Thompson (2008), Rapach, Strauss and Zhou (2009) and Thornton and Valente (2012), we constrain the portfolio weight on bonds to lie between -100% and 200% each month, so in Eq. (3.12) $w_{0,t} = -1$ ($w_{0,t} = 2$) if $w_{0,t} < -1$ ($w_{0,t} > 2$).

3.3.3 Performance of portfolio choice with uncertainty aversion

In this paper, we evaluate economic gains of 5 strategies holding for one year, over the mean-variance portfolio based on the Expectations Hypothesis (EH). The strategies reported in Table 2 include minimum-variance portfolio, mean-variance portfolio based on the proposed learning model (learning portfolio), EH portfolio with parameter uncertainty aversion (PUA), learning portfolio with PUA, and learning portfolio considering parameter uncertainty aversion and model uncertainty (PUA & MU). In Table 2 we also report different degrees of parameter uncertainty level, representing different views of rare events.[33]

The results are rather surprising. Our proposed model has very significant economic value in contrast to the EH benchmark. The utility gain based on the proposed model is ranging from 4% to a remarkably high value 27%. The economic value of the learning portfolios with uncertainty aversion and minimum-variance portfolio peak when $\gamma_R = 6$. The minimum-variance strategy naturally performs well at high risk-averse level, so learning portfolios with uncertainty aversion also have favorable performance as they are shrunk toward the minimum-variance strategy. Panel B of Table 2 shows when short sales are not allowed, the EH strategy seems to perform much worse, so the economic gains of strategies allowing for uncertainty aversion become extremely high.

It is worth noting that the strategies we proposed have very consistent performance. At low risk-averse level, whilst the minimum-variance portfolio have a relatively small economic gain (0.46%), the proposed strategies still have $4-6\%$ utility gains. This is because we search the portfolio weights in the admissible region based on reliable forecasts, so we do not fall into the 'pessimism trap' that no investment in risky assets is made.

Figure 6 gives the cumulative sum of log returns generated by the above strategies over time, so we can have an intuitive impression. The minimum-variance portfolio has the smallest cumulative sum of log returns, but it is the most stable strategy which therefore can provide high economic value for a risk-averse investor. The learning portfolio considering parameter uncertainty aversion and model uncertainty (PUA&MU) is slightly less than the learning portfolio with parameter uncertainty only, but is more stable so it has a higher economic gain. The EH-based strategies perform less favorably owing to the drops in the late 1970s and early 1980s, probably because the economy was undergoing

[33] As we have discussed in Section 2.3.1, the quantity ϵ has a probability interpretation corresponding to an F distribution with degrees of freedoms N and $T-N$.

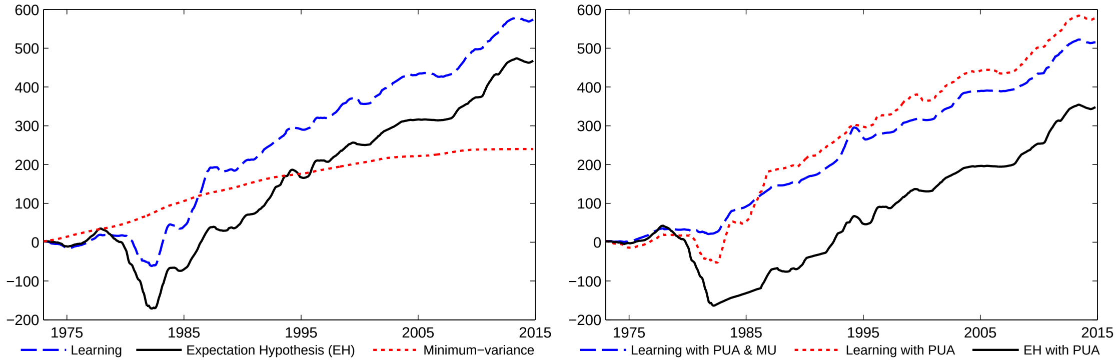
Table 2: Economic Gains of Different Strategies

	Strategy	Utility gain (Δ)		
		$\gamma_R = 6$	$\gamma_R = 3$	$\gamma_R = 1$
Panel A: Short sales allowed				
	Minimum-variance	29.69	12.15	0.46
	Learning	11.76	7.14	4.08
	EH with PUA	3.86	0.52	-1.72
$\epsilon = 2.78$ (99%)	Learning with PUA	13.93	8.28	4.50
	Learning with PUA&MU	27.22	13.12	6.18
	EH with PUA	0.63	0.07	-0.28
$\epsilon = 2.07$ (95%)	Learning with PUA	12.77	7.51	4.01
	Learning with PUA&MU	27.05	13.26	6.23
Panel B: Short sales not allowed				
	Minimum-variance	94.43	39.87	3.49
	Learning	15.08	8.39	3.92
	EH with PUA	18.58	6.59	-1.40
$\epsilon = 2.78$ (99%)	Learning with PUA	22.04	11.94	5.21
	Learning with PUA&MU	42.38	17.59	5.88
	EH with PUA	5.63	2.29	0.13
$\epsilon = 2.07$ (95%)	Learning with PUA	18.40	9.77	4.02
	Learning with PUA&MU	42.12	17.29	5.67

Notes: 1. The table reports the out-of-sample 12-month holding utility gain (Δ) on different portfolio strategies, over the evaluation period from 1971:07 to 2014:10.
2. Utility gain (Δ) is the portfolio management fee (in annualized percentage return) that an investor with mean-variance preferences would be willing to pay, in order to switch from the Expectations Hypothesis (EH) strategy to another strategy. The utility gain is computed at three risk aversion levels, i.e., $\gamma_R = 1, 3, 6$. Higher utility gain is preferred.
3. We report the performance of strategies relative to the mean-variance portfolio based on the EH. The strategies reported include minimum-variance portfolio, mean-variance portfolio based on the proposed learning model (learning portfolio), EH portfolio with parameter uncertainty aversion (PUA), learning portfolio with PUA, and learning portfolio considering parameter uncertainty aversion and model uncertainty (PUA&MU). In parenthesis, we report the percentage size of the confidence interval for an $F_{9,51}$ implied by the values of ϵ.

a structural change at that time.

Figure 6: Cumulative Sum of Log Returns



Notes:
1. This figure displays the cumulative sum of log returns generated by respective strategies when short sales are allowed and when we set $\gamma_R = 3$, $\epsilon = 2.78$. The strategies include minimum-variance portfolio, mean-variance portfolio based on the proposed learning model (learning portfolio), EH portfolio with parameter uncertainty aversion (PUA), learning portfolio with PUA, and learning portfolio considering parameter uncertainty aversion and model uncertainty (PUA&MU).
2. The evaluation period is from 1971:07 to 2014:10. The unit is percentage.

3.3.4 Robustness

In Figure 6 of the last section, we detect a notable fall in returns for the EH portfolio in and around 1980, while Federal Reserve's famous 'monetarist experiment' was conducted. We consider robustness checks by excluding this period. Table 3, 4 and 5 display the performance of the same strategies from 1990, 2000 and 2010 onward, respectively.

The resulting economic gains for our proposed portfolios are weakened, but still significant. Note that whilst the minimum-variance portfolios tend to have distinct performance at different risk-aver levels, the learning portfolios with uncertainty aversion perform stably and have significant positive gains (2% or more). Even when the minimum-variance portfolios have significantly negative gains, the performance of learning portfolios does not fall along the same way. Our proposed portfolios seem to have relatively weaker performance from 2000 onward, which we find is mainly due to highly uncertain estimates of pricing kernel in and around the financial crisis, recall Figure 3. This characteristic is potentially related to the decrease in persistence under the physical measure, see Figure 5. Moreover, the *zero lower bound* problem during the financial crisis may induce some unfeasible return forecasts which should be excluded when we construct the optimal portfolio. After 2010, we see that the proposed portfolios get back on track and have economic value around 2%. Note that from these robustness checks, forecasts implied by our learning model alone does not guarantee substantial and consistent economic value, so ambiguity aversion is imperative in generating satisfactory economic utility. Therefore, by considering parameter and model uncertainty, investors truly benefit from the predictability of excess returns, and hence the *economic value puzzle* in bond returns can be resolved.

Moreover, by a simple utility gain decomposition from our results, we can reveal different degrees of utility gains because of aversion to various sources of uncertainty. The utility gain decomposition is done by computing the difference in gains among portfolios, i.e. learning, learning with PUA, and PUA&MU. The difference between learning portfolio and learning with PUA is the gain (if any) from the aversion to parameter uncertainty. The difference between learning portfolio and learning with PUA&MU is the gain from the ambiguity aversion allowing for both parameter and model uncertainty, which is generally much larger than the former one, except in very few cases. This finding is informative and confirms the necessity of incorporating model uncertainty. Although we have mentioned in Section 3.2.1 that the parameter uncertainty is the main source of prediction uncertainty when compared with model uncertainty, we find that *allowing for the smaller fraction of prediction variance originated from model uncertainty is of essence to generate significant and consistent economic value.* This finding further assures the robustness of our learning framework with model uncertainty.

Table 3: Economic Gains of Different Strategies (from 1990 onward)

	Strategy	Utility gain (Δ)		
		$\gamma_R = 6$	$\gamma_R = 3$	$\gamma_R = 1$
Panel A: Short sales allowed				
	Minimum-variance	1.21	-5.88	-10.60
	Learning	-0.93	-0.19	0.30
$\epsilon = 2.78$ (99%)	EH with PUA	-0.56	-0.53	-0.51
	Learning with PUA	-0.89	-0.32	0.06
	Learning with PUA&MU	3.35	0.12	-0.44
$\epsilon = 2.07$ (95%)	EH with PUA	-0.42	-0.37	-0.36
	Learning with PUA	-0.81	-0.21	0.18
	Learning with PUA&MU	3.22	0.25	-0.34
Panel B: Short sales not allowed				
	Minimum-variance	22.36	-2.86	-19.67
	Learning	-0.94	-0.29	0.15
$\epsilon = 2.78$ (99%)	EH with PUA	-1.47	-1.27	-1.14
	Learning with PUA	-0.41	-0.29	-0.21
	Learning with PUA&MU	8.19	2.71	0.06
$\epsilon = 2.07$ (95%)	EH with PUA	-1.09	-0.89	-0.76
	Learning with PUA	-0.35	-0.12	0.04
	Learning with PUA&MU	8.07	2.66	-0.06

Notes: 1. The table reports the out-of-sample 12-month holding utility gain (Δ) on different portfolio strategies, over the evaluation period from 1990:01 to 2014:10.
2. Utility gain (Δ) is the portfolio management fee (in annualized percentage return) that an investor with mean-variance preferences would be willing to pay, in order to switch from the Expectations Hypothesis (EH) strategy to another strategy. The utility gain is computed at three risk aversion levels, i.e., $\gamma_R = 1, 3, 6$. Higher utility gain is preferred.
3. We report the performance of strategies relative to the mean-variance portfolio based on the EH. The strategies reported include minimum-variance portfolio, mean-variance portfolio based on the proposed learning model (learning portfolio), EH portfolio with parameter uncertainty aversion (PUA), learning portfolio with PUA, and learning portfolio considering parameter uncertainty aversion and model uncertainty (PUA&MU). In parenthesis, we report the percentage size of the confidence interval for an $F_{9,51}$ implied by the values of ϵ.

Table 4: Economic Gains of Different Strategies (from 2000 onward)

	Strategy	Utility gain (Δ)		
		$\gamma_R = 6$	$\gamma_R = 3$	$\gamma_R = 1$
Panel A: Short sales allowed				
	Minimum-variance	-3.07	-8.03	-11.34
	Learning	-1.42	0.68	2.09
$\epsilon = 2.78$ (99%)	EH with PUA	0.01	0.01	0.01
	Learning with PUA	-1.10	0.86	2.17
	Learning with PUA&MU	-0.50	-1.36	-1.41
$\epsilon = 2.07$ (95%)	EH with PUA	0.01	0.01	0.00
	Learning with PUA	-0.99	0.95	2.25
	Learning with PUA&MU	-0.14	-1.40	-1.52
Panel B: Short sales not allowed				
	Minimum-variance	4.88	-12.70	-24.42
	Learning	-4.49	-1.31	0.80
$\epsilon = 2.78$ (99%)	EH with PUA	0.00	0.00	0.00
	Learning with PUA	-3.84	-1.00	0.89
	Learning with PUA&MU	0.14	-0.43	0.28
$\epsilon = 2.07$ (95%)	EH with PUA	0.00	0.00	0.00
	Learning with PUA	-3.47	-0.73	1.09
	Learning with PUA&MU	0.71	-0.38	0.28

Notes: 1. The table reports the out-of-sample 12-month holding utility gain (Δ) on different portfolio strategies, over the evaluation period from 2000:01 to 2014:10.
2. Utility gain (Δ) is the portfolio management fee (in annualized percentage return) that an investor with mean-variance preferences would be willing to pay, in order to switch from the Expectations Hypothesis (EH) strategy to another strategy. The utility gain is computed at three risk aversion levels, i.e., $\gamma_R = 1, 3, 6$. Higher utility gain is preferred.
3. We report the performance of strategies relative to the mean-variance portfolio based on the EH. The strategies reported include minimum-variance portfolio, mean-variance portfolio based on the proposed learning model (learning portfolio), EH portfolio with parameter uncertainty aversion (PUA), learning portfolio with PUA, and learning portfolio considering parameter uncertainty aversion and model uncertainty (PUA&MU). In parenthesis, we report the percentage size of the confidence interval for an $F_{9,51}$ implied by the values of ϵ.

Table 5: Economic Gains of Different Strategies (from 2010 onward)

	Strategy	Utility gain (Δ)		
		$\gamma_R = 6$	$\gamma_R = 3$	$\gamma_R = 1$
Panel A: Short sales allowed				
	Minimum-variance	-1.08	-7.74	-12.17
	Learning	1.06	0.61	0.30
	EH with PUA	0.00	0.00	0.00
$\epsilon = 2.78$ (99%)	Learning with PUA	1.06	0.61	0.30
	Learning with PUA&MU	2.84	2.36	1.74
	EH with PUA	0.00	0.00	0.00
$\epsilon = 2.07$ (95%)	Learning with PUA	1.06	0.61	0.30
	Learning with PUA&MU	2.35	2.10	1.74
Panel B: Short sales not allowed				
	Minimum-variance	14.85	-7.08	-21.70
	Learning	-0.06	0.02	0.07
	EH with PUA	0.00	0.00	0.00
$\epsilon = 2.78$ (99%)	Learning with PUA	-0.06	0.02	0.07
	Learning with PUA&MU	1.82	0.75	0.33
	EH with PUA	0.00	0.00	0.00
$\epsilon = 2.07$ (95%)	Learning with PUA	-0.06	0.02	0.07
	Learning with PUA&MU	1.82	0.75	0.79

Notes: 1. The table reports the out-of-sample 12-month holding utility gain (Δ) on different portfolio strategies, over the evaluation period from 2010:01 to 2014:10.
2. Utility gain (Δ) is the portfolio management fee (in annualized percentage return) that an investor with mean-variance preferences would be willing to pay, in order to switch from the Expectations Hypothesis (EH) strategy to another strategy. The utility gain is computed at three risk aversion levels, i.e., $\gamma_R = 1, 3, 6$. Higher utility gain is preferred.
3. We report the performance of strategies relative to the mean-variance portfolio based on the EH. The strategies reported include minimum-variance portfolio, mean-variance portfolio based on the proposed learning model (learning portfolio), EH portfolio with parameter uncertainty aversion (PUA), learning portfolio with PUA, and learning portfolio considering parameter uncertainty aversion and model uncertainty (PUA&MU). In parenthesis, we report the percentage size of the confidence interval for an $F_{9,51}$ implied by the values of ϵ.

4 Conclusion

This paper studies the problem of a representative agent who learns about the information in the bond market over time, with the consideration of parameter uncertainty and model uncertainty. In addition to adaptive learning about parameters as considered in Giacoletti, Laursen and Singleton (2014), this proposed framework provides flexibility in specifying different learning speeds and model restrictions. The optimal specification can be selected according to predictive performance over time, and, therefore, reduce the risk of data snooping. This method is robust in the sense that it reveals the agent's expectations in real time by using conditional information. We find that apart from observational variance, parameter instability is the dominant driving force of predictive uncertainty, when compared with uncertainty in learning speed or model restrictions. It suggests that a successful term structure model should at least consider time-varying parameters when making conditional forecasts.

The problem of asset allocation for an investor with ambiguity aversion building upon Garlappi, Uppal and Wang (2007) is studied. After learning the parameters, the ambiguity-averse investor forms optimal portfolios by maximizing mean-variance expected utility. The ensemble of all salient features offered by our framework is essential in producing significant and consistent economic value over the Expectations Hypothesis benchmark. Ambiguity aversion with model uncertainty ensures that the search for portfolio weights is in a reliable region, which in turn not only boosts but also stabilizes the gains. Therefore, ambiguity aversion is a key to salvaging the models with significant predictability but little economic value used in the previous literature, and the *economic value puzzle* in bond returns can be resolved following this direction.

There are various important directions in which this approach can be extended. By allowing for more general model specifications, such as incorporating more information from macro-finance predictor variables or economic constraints as in Pettenuzzo, Timmermann and Valkanov (2014), it is possible to further improve model performance and provide meaningful economic rationales. It would also be interesting to develop hybrid models with both spanned and unspanned macroeconomic risks and explore the prediction uncertainty from different choices of predictors, as suggested by Bauer and Rudebusch (2015). Finally, our results suggest that the zero lower bound problem could hinder the performance of our portfolio strategy. We leave these directions for further research.

References

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Appendix A Bond Pricing in GDTSMs

Under the assumptions in Section 2.1, the price of an m-period zero-coupon bond is given by

$$D_t^m = E_t^{\mathbb{Q}}[e^{-\sum_{i=1}^{m-1} r_{t+i}}] = e^{\mathcal{A}_m + \mathcal{B}_m \cdot \mathcal{P}_t}, \tag{A.1}$$

where $(\mathcal{A}_m, \mathcal{B}_m)$ solve the first-order difference equations

$$\mathcal{A}_{m+1} - \mathcal{A}_m = (K_0^{\mathbb{Q}})'\mathcal{B}_m + \frac{1}{2}\mathcal{B}_m' \Sigma_{\mathcal{P}\mathcal{P}} \mathcal{B}_m - \rho_0, \tag{A.2}$$

$$\mathcal{B}_{m+1} - \mathcal{B}_m = (K_1^{\mathbb{Q}})'\mathcal{B}_m - \rho_1, \tag{A.3}$$

subject to the initial conditions $\mathcal{A}_0 = 0$, $\mathcal{B}_0 = 0$. The loadings for the corresponding bond yield are $A_m = -\mathcal{A}_m/m$ and $B_m = -\mathcal{B}_m/m$. See Dai and Singleton (2003) for details.

Appendix B Estimation Methods

B.1 Bayesian Kalman Filter with Forgetting Factor

We conduct the Kalman filter estimation for the state space model with Equation (2.14) and Equation (2.15):

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

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

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

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

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

where $\wp$ is the support of β_{t-1}. The solution to this problem can be defined using a Bayesian generalization of the typical Kalman filter recursions. Given an initial condition $\beta_0 \sim N(m_0, \Phi_0)$ we can define (cf. West and Harrison (1997))[34]:

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

1. Posterior at time $t-1$

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

2. Prior at time t

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

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

3. Posterior at time t

$$\beta_t|D_t \sim N\left(m_t, \Phi_t\right), \tag{B.1}$$

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

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

$$\Sigma_t|D_{t-1} \sim iW\left(S_{t-1}, \delta n_{t-1}\right), \tag{B.2}$$

while the posterior takes the form

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

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

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

$$Q_t = \left(\lambda^{-1} - 1\right)\Phi_{t-1}, \tag{B.3}$$

for a given forgetting factor λ. Note that λ is mathematically equivalent to the quantity $1-\gamma$ in the *constant gain least squares* (CGLS) algorithm, see Appendix B.2 and McCulloch (2007). Therefore, the forgetting factor λ and the gain parameter γ are two sides to the same coin. As λ becomes larger, the γ becomes smaller, so the model would adjust more slowly if a structural break happens.

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

B.2 The Link between the Kalman Filter and Adaptive Least Squares

From the Kalman filter described in last section, we have the following formulas

$$m_t = m_{t|t-1} + \Phi_{t|t-1} X_t' (V_t^{-1})' \tilde{v}_t, \tag{B.4}$$

$$\Phi_t = \Phi_{t|t-1} - \Phi_{t|t-1} X_t' (V_t^{-1})' X_t \Phi_{t|t-1}', \tag{B.5}$$

$$V_t = X_t \Phi_{t|t-1} X_t' + \Sigma_t, \tag{B.6}$$

where $\tilde{v}_t = z_t - X_t m_{t|t-1}$ is the prediction error. Post-multiply (B.5) by X_t' and combine with (B.6) we obtain

$$\begin{aligned}\Phi_t X_t' &= \Phi_{t|t-1}\Big(X_t' - X_t'(V_t^{-1})' X_t \Phi_{t|t-1}' X_t'\Big) \\ &= \Phi_{t|t-1}\Big(X_t' - X_t'(V_t^{-1})'(V_t - \Sigma_t)\Big) \\ &= \Phi_{t|t-1} X_t' (V_t^{-1})' \Sigma_t. \end{aligned}\tag{B.7}$$

We can get the expressions of ALS by post-multiplying (B.7) by Σ_t^{-1} and substituting it back to (B.4) and (B.5), respectively.

For (B.4) we have

$$m_t = m_{t|t-1} + R_t^{-1} X_t' \Sigma_t^{-1} \tilde{v}_t, \tag{B.8}$$

where we set $R_t = \Phi_t^{-1}$. So we obtain the evolution of the drift in ALS.

We continue the previous substitution in (B.5) with $\Phi_{t|t-1} = \Phi_{t-1} + Q_t$ in hand, which gives

$$\Phi_t = \Phi_{t-1} + Q_t - \Phi_t X_t' \Sigma_t^{-1} X_t \Phi_{t-1} - \Phi_t X_t' \Sigma_t^{-1} X_t Q_t. \tag{B.9}$$

Setting $R_t = \Phi_t^{-1}$, we can get the final equation after some manipulation

$$R_t = (I + Q_t \Phi_t^{-1})^{-1} R_{t-1} + X_t' \Sigma_t^{-1} X_t. \tag{B.10}$$

If we set $Q_t = \frac{\gamma}{1-\gamma} \Phi_{t-1}$, then we have the *constant gain least squares* (CGLS) algorithm.

B.3 Brief Introduction of the Unscented Kalman Filter

Consider the following nonlinear discrete-time stochastic system represented by:

$$z_t = f(\beta_t) + v_t, \tag{B.11}$$
$$\beta_{t+1} = \beta_t + u_t, \tag{B.12}$$

where z_t is an $n \times 1$ vector of variables, β_t are coefficients that govern the pricing equation $f(\cdot)$, $v_t \sim N(0, \Sigma_t)$ with Σ_t an $n \times n$ covariance matrix, and $u_t \sim N(0, Q_t)$.

As we mentioned before, the solution for this system follows a recursive rule given by

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

where D_t is the information set at time t. Similar to the Kalman filter, the unscented Kalman filter has the same recursive estimation process as in Appendix B.1, except the update equations (B.4) and (B.5) are replaced by:

$$m_t = m_{t|t-1} + \mathcal{K}_t \widetilde{v}_t, \tag{B.14}$$
$$\Phi_t = \Phi_{t|t-1} - \mathcal{K}_t P_{z_t} \mathcal{K}_t', \tag{B.15}$$

where $\mathcal{K}_t$ is the Kalman gain of the filter and P_{z_t} is the prior variance of z_t. The above updating equations are intuitively similar to the ones in the Kalman filter, except we use different formulas to obtain the Kalman gain and the prior variance. To be more specific, these quantities can be calculated by simulating sigma points around the mean of state variables, see Wan and Van Der Merwe (2000) and Appendix D for details.

B.4 Probabilities for Dynamic Model Selection

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

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

where $L_{t-1} = k$ means the k_{th} model[35] is selected and $p(\beta_{t-1}^{(k)}|L_{t-1} = k, D_{t-1})$ is given by the Kalman filter (Equation B.1). To simplify notation, let $\pi_{t|s,l} = \mathbf{Pr}(L_t = l|D_s)$.

[35] For example, it can be the k_{th} model in a pool of possible models with different restrictions or gain parameter γ.

Raftery, Kárnỳ and Ettler (2010) used an empirically sensible simplification that

$$\pi_{t|t-1,k} = \frac{\pi^{\alpha}_{t-1|t-1,k}}{\sum_{l=1}^{K} \pi^{\alpha}_{t-1|t-1,l}}, \tag{B.17}$$

where $0 < \alpha \leq 1$. A forgetting factor is also employed here, of which the meaning is discussed in the last section.[36] The huge advantage of using the forgetting factor α is that it does not require an MCMC algorithm to draw transitions between models or a simulation algorithm over model space.

The model updating equation is

$$\pi_{t|t,k} = \frac{\pi_{t|t-1,k} p_k(z_t|D_{t-1})}{\sum_{l=1}^{K} \pi_{t|t-1,l} p_l(z_t|D_{t-1})}, \tag{B.18}$$

where $p_k(z_t|D_{t-1})$ is the predictive likelihood. When proceeding with dynamic model selection (DMS), the model with the highest probability is the best model we would like to select. Alternatively, we can conduct dynamic model averaging (DMA), which average the predictions of all models with respective probabilities.

[36] In this paper, we set $\alpha = 1$ to put equal weights to previous information.

Appendix C Proof of Propositions

C.1 Heuristics of Proposition 1

Following Garlappi, Uppal and Wang (2007), we start with the inner minimization

$$\min_{\mu} w^\intercal \mu - \frac{\gamma}{2} w^\intercal \Sigma w, \tag{C.1}$$

subject to

$$(\hat{\mu} - \mu)^\intercal \Sigma^{-1} (\hat{\mu} - \mu) \leq \varepsilon. \tag{C.2}$$

The Lagrangian is given by

$$\mathcal{L}(\mu, \lambda^{\mathcal{L}}) = w^\intercal \mu - \frac{\gamma}{2} w^\intercal \Sigma w - \lambda^{\mathcal{L}} [\varepsilon - (\hat{\mu} - \mu)^\intercal \Sigma^{-1} (\hat{\mu} - \mu)]. \tag{C.3}$$

μ^* is a solution of the constrained problem above if and only if there exists a scalar $\lambda^{\mathcal{L}*} \geq 0$, such that $(\mu^*, \lambda^{\mathcal{L}*})$ is a solution of the following unconstrained problem

$$\min_{\mu} \max_{\lambda^{\mathcal{L}}} \mathcal{L}(\mu, \lambda^{\mathcal{L}}). \tag{C.4}$$

From the first order conditions with respect to μ in Equation (C.3), we have

$$\mu^* = \hat{\mu} - \frac{1}{2\lambda^{\mathcal{L}}} \Sigma w. \tag{C.5}$$

Substituting this in the Lagrangian (C.3) we obtain

$$\mathcal{L}(\mu^*, \lambda^{\mathcal{L}}) = w^\intercal \hat{\mu} - (\frac{1}{4\lambda^{\mathcal{L}}} + \frac{\gamma}{2}) w^\intercal \Sigma w - \lambda^{\mathcal{L}} \varepsilon. \tag{C.6}$$

Therefore, the original max-min problem with constraints is equivalent to the maximization problem below

$$\max_{w, \lambda^{\mathcal{L}}} w^\intercal \hat{\mu} - (\frac{1}{4\lambda^{\mathcal{L}}} + \frac{\gamma}{2}) w^\intercal \Sigma w - \lambda^{\mathcal{L}} \varepsilon. \tag{C.7}$$

Solving for $\lambda^{\mathcal{L}}$, we get $\lambda^{\mathcal{L}} = \frac{1}{2}\sqrt{\frac{w^\intercal \Sigma w}{\varepsilon}} > 0$. Then we can rewrite the maximization problem as

$$\max_{w} w^\intercal \hat{\mu} - \frac{\gamma}{2} w^\intercal \Sigma w (1 + \frac{2\sqrt{\varepsilon}}{\gamma \sqrt{w^\intercal \Sigma w}}). \tag{C.8}$$

It is easy to show, the first-order condition with respect to w gives

$$w = \Big(\frac{\sigma_p}{\gamma \sigma_p + \sqrt{\varepsilon}}\Big) \Sigma^{-1} \hat{\mu}. \tag{C.9}$$

With Equation (C.9), we can post-multiply $w^\intercal$ by Σw and obtain

$$\sigma_p^2 = \Big(\frac{\sigma_p}{\gamma\sigma_p + \sqrt{\varepsilon}}\Big)^2 \hat{\mu}^\intercal \Sigma^{-1} \hat{\mu}, \tag{C.10}$$

where $\sigma_p = \sqrt{w^\intercal \Sigma w}$.

After some manipulation, the optimal portfolio weight w^* is given by the positive real solution σ_p^* of the following polynomial

$$\gamma^2\sigma_p^2 + 2\sqrt{\varepsilon}\gamma\sigma_p + \varepsilon - \hat{\mu}^\intercal\Sigma^{-1}\hat{\mu} = 0. \tag{C.11}$$

If $\hat{\mu}^\intercal\Sigma^{-1}\hat{\mu}$ is sufficiently large, we have a unique positive real solution σ_p^*. Otherwise, we have a non-negative solution σ_p^*, i.e. $w^* = \mathbf{0}$. Therefore, using Equation (C.9) we have Equation (2.29) in **Proposition 1**.

C.2 Heuristics of Proposition 2

To solve the max-min problem

$$\max_{w,\hat{\mu}} \min_{\mu} w^\intercal \mu - \frac{\gamma}{2} w^\intercal \Sigma w, \tag{C.12}$$

subject to

$$(\hat{\mu} - \mu)^\intercal \Sigma^{-1} (\hat{\mu} - \mu) \leq \varepsilon, \tag{C.13}$$
$$\hat{\mu} \in \{\hat{\mu}_k : k = 1, ..., K\}, \tag{C.14}$$

we follow the same procedures as in Appendix C.1. The difference lies in the outer maximization:

$$\max_{w,\hat{\mu}} w^\intercal \hat{\mu} - \frac{\gamma}{2} w^\intercal \Sigma w (1 + \frac{2\sqrt{\varepsilon}}{\gamma\sqrt{w^\intercal\Sigma w}}), \tag{C.15}$$

where we need to consider first-order conditions with respect to $\hat{\mu}$ as well as w. We have the same formula (C.9) for w. However, in addition to w, we need to search $\hat{\mu}$ over a set of possible models at each point in time, and use the optimal forecasts $\hat{\mu}^*$ that give the largest value in the above maximization problem.

Appendix D Technical Details of the Unscented Kalman Filter

D.1 Unscented Transformation

The UKF is based on the *unscented transformation* (UT) in order to form a Gaussian approximation to the target distribution. The advantage of UT over the Taylor series based approximation in other nonlinear filters (for example, the extended Kalman filter) is that Jacobian and Hessian matrices are not need, so the estimation procedure is more convenient in a system where closed-form expressions are not available.

The follows show the procedure of unscented transformation:

1. We simulate a set of $2n+1$ sigma points $\mathcal{X}$ of the state variables x, where n is the dimension of the state, from the mean m and covariance matrix Φ:

$$
\begin{aligned}
\mathcal{X}^{(0)} &= m, \\
\mathcal{X}^{(i)} &= m + \sqrt{(n+\lambda)\Phi}, \quad i = 1, ..., n, \\
\mathcal{X}^{(i)} &= m - \sqrt{(n+\lambda)\Phi}, \quad i = n+1, ..., 2n,
\end{aligned} \tag{D.1}
$$

   with the associated weights W_m of the state variables x and W_c of the observations z:

$$
\begin{aligned}
W_m^{(0)} &= \frac{\lambda}{n+\lambda}, \\
W_c^{(0)} &= \frac{\lambda}{n+\lambda} + (1 - \alpha^2 + \beta), \\
W_m^{(i)} &= W_c^{(i)} = \frac{1}{2(n+\lambda)}, \quad i = 1, ..., 2n,
\end{aligned} \tag{D.2}
$$

   where $\lambda = \alpha^2(n+\kappa) - n$ is a scaling parameter. α and κ determine the spread of the sigma points around the state, and β is used to incorporate prior knowledge of the distribution of the state.[37]

2. The sigma points are propagated though non-linearity as

$$
\mathcal{Z}^{(i)} = f(\mathcal{X}^{(i)}), \quad i = 0, ..., 2n. \tag{D.3}
$$

[37] As suggested by Wan and Van Der Merwe (2000), normal values are $\alpha = 10^{-3}$, $\kappa = 0$ and $\beta = 2$. If the true distribution of x is Gaussian, $\beta = 2$ is optimal. Note that the simple approximation approach taken with the UT are accurate to the third order for all nonlinearities with Gaussian innovations, which has an advantage over Monte-Carlo methods which require (orders of magnitude) more sample points to provide an accurate distribution of the state.

3. We can compute the mean and covariance estimates for z:

$$\bar{z} \approx \sum_{i=0}^{2n} W_m^{(i)} \mathcal{Z}^{(i)},$$
$$P_z \approx \sum_{i=0}^{2n} W_c^{(i)} (\mathcal{Z}^{(i)} - \bar{z})(\mathcal{Z}^{(i)} - \bar{z})^\intercal. \tag{D.4}$$

4. Estimation of the cross-covariance between z and x is given by

$$P_{x,z} \approx \sum_{i=0}^{2n} W_c^{(i)} (\mathcal{X}^{(i)} - m)(\mathcal{Z}^{(i)} - \bar{z})^\intercal. \tag{D.5}$$

D.2 Estimation Procedure using UKF

Based on the discussion of UT above, we describe the following *prediction* and *update* steps of the UKF.

- *Prediction*: Compute the predicted state mean $m_{t|t-1}$ and covariance $\Phi_{t|t-1}$, the predicted observation mean $\hat{Z}_t$ and covariance P_{z_t}, and the cross-variance of the state and measurement P_{x_t,z_t}:

$$m_{t|t-1} = m_{t-1|t-1},$$
$$\Phi_{t|t-1} = \Phi_{t-1|t-1} + Q_t,$$
$$\hat{z}_t = \sum_{i=0}^{2n} W_m^{(i)} \mathcal{Z}_{t|t-1}^{(i)},$$
$$P_{z_t} = \sum_{i=0}^{2n} W_c^{(i)} (\mathcal{Z}_{t|t-1}^{(i)} - \hat{z}_t)(\mathcal{Z}_{t|t-1}^{(i)} - \hat{z}_t)^\intercal + \Sigma_t,$$
$$P_{x_t,z_t} = \sum_{i=0}^{2n} W_c^{(i)} (\mathcal{X}_{t|t-1}^{(i)} - m_{t|t-1})(\mathcal{Z}_{t|t-1}^{(i)} - \hat{z}_t)^\intercal. \tag{D.6}$$

- *Update*: Compute the filter gain $\mathcal{K}_t$ and the updated state mean $m_{t|t}$ and covariance $\Phi_{t|t}$ in order to get Equations (B.14) and (B.15):

$$\mathcal{K}_t = P_{x_t,z_t} P_{z_t}^{-1},$$
$$m_{t|t} = m_{t|t-1} + \mathcal{K}_t \widetilde{v}_t,$$
$$\Phi_{t|t} = \Phi_{t|t-1} - \mathcal{K}_t P_{x_t} \mathcal{K}_t'. \tag{D.7}$$

Following Koop and Korobilis (2012) and Koop and Korobilis (2013), we specify

$$Q_t = \left(\lambda_f^{-1} - 1\right) \Phi_{t-1}, \tag{D.8}$$

where λ_f is the 'forgetting factor'. We have an intuitive interpretation for the forgetting factor: the smaller the λ_f, the more weights UKF puts on the new information, and hence the system is more sensitive to structural changes.[38] To fix the idea, we set the value to 0.99 to ensure the stability of the system.

D.3 Detailed Specification of the ATSM

We adopt a specific parametric form of the class of Affine Term Structure Models (ATSMs) with arbitrage-free restrictions under the Duffie and Kan (1996) framework, which is similar to the majority of current related literature, see for example, Duffee (2002), Dai and Singleton (2003), Joslin, Singleton and Zhu (2011) and Joslin, Priebsch and Singleton (2014).[39] In this setup, the measurement equation in the nonlinear system is governed by parameter set $(\Sigma_{\mathcal{PP}}, \lambda^{\mathbb{Q}}, r^{\mathbb{Q}}_{\infty})$.[40] Joslin, Singleton and Zhu (2011) prove that every canonical GDTSM is observationally equivalent to the canonical GDTSM:

$$X_t = J(\lambda^{\mathbb{Q}}) + \sqrt{\Sigma_X}\epsilon^{\mathbb{Q}}_t, \tag{D.9}$$
$$r_t = r^{\mathbb{Q}}_{\infty} + \mathbf{1} \cdot X_t, \tag{D.10}$$

where X_t are normalized risk factors, $r^{\mathbb{Q}}_{\infty}$ denotes the unconditional mean of r_t under $\mathbb{Q}$, and $J(\lambda^{\mathbb{Q}})$ is a real Jordan form matrix associated with eigenvalues $\lambda^{\mathbb{Q}}$. We can conveniently apply invariant transformation of X_t and then replace the risk factors with preferred portfolio combinations, see Dai and Singleton (2000) and Joslin, Singleton and Zhu (2011) for details.

Solving for the bond prices of m-period zero-coupon bond D^m_t using the recursion given by

$$D^m_t = E^{\mathbb{Q}}_t[e^{-\sum_{i=1}^{m-1} r_{t+i}}], \tag{D.11}$$

we can obtain the following pricing equation for m-period bond yields as the measurement equation:

$$y^m_t = A^m_X(\Sigma_{t,\mathcal{PP}}, \lambda^{\mathbb{Q}}_t, r^{\mathbb{Q}}_{t,\infty}) + B^m_X(\lambda^{\mathbb{Q}}_t)X_t, \tag{D.12}$$

where

$$\begin{aligned} A^{m+1}_X - A^m_X &= \frac{1}{2}B^{m\prime}_X\,\Sigma_X B^m_X - r^{\mathbb{Q}}_{t,\infty}, \\ B^{m+1}_X - B^m_X &= J(\lambda^{\mathbb{Q}}_t)' B^m_X. \end{aligned} \tag{D.13}$$

[38]To see this, use Taylor series expansion around the observation mean. See Koop and Korobilis (2012) and Koop and Korobilis (2013) for detailed discussion about the ‘forgetting factor’.

[39]Joslin, Singleton and Zhu (2011) denote their proposed model as a canonical Gaussian dynamic term structure model (GDTSM). We use ATSM and GDTSM interchangeably.

[40]Following the notations in Joslin, Singleton and Zhu (2011) and Joslin, Priebsch and Singleton (2014), $\lambda^{\mathbb{Q}}$ denotes the N-vector of ordered nonzero eigenvalues of $K^{\mathbb{Q}}_{\mathcal{PP}}$ and $r^{\mathbb{Q}}_{\infty}$ denotes the long-run mean of r_t under $\mathbb{Q}$.