Predicting Bond Risk Premia via Sequential Learning

Tomasz Dubiel-Teleszynski* Konstantinos Kalogeropoulos† Nikolaos Karouzakis‡

January 15, 2019

Abstract

This paper revisits the evident puzzling behaviour between the statistical predictability of bond excess returns and the incapability to use such predictability to generate meaningful out-of-sample economic benefits for bond investors. We resolve this puzzle by implementing a Bayesian learning framework, under a dynamic term structure model (DTSM). We develop a sequential process for investors, who learn the entire predictive density of bond excess returns, when new information arrives, thus accounting for parameter and model uncertainty. We find strong evidence of out-of-sample predictability for bond investors who utilize information coming solely from the yield curve. Most importantly, such statistical evidence is turned into economically significant utility gains for investors, across prediction horizons. Furthermore, we find that the predictive power and economic gains decrease with maturity and increase with the prediction horizon. Our results are more pronounced when no portfolio constraints are imposed. Finally, we assess the efficiency and economic importance of risk price restrictions from a forecasting perspective. We find that maximally flexible model outperforms restricted models both statistically, as well as, economically.

Keywords: Bond Risk Premia, Bond Return Predictability, Economic Value, Affine Term Structure Models, Sequential Learning, Parameter Uncertainty

JEL classification: C11, C52, E43, G12, G17

---

*Department of Statistics, London School of Economics, Houghton Street, London, WC2A 2AE, United Kingdom. Email: T.Dubiel-Teleszynski1@lse.ac.uk

†Department of Statistics, London School of Economics, Houghton Street, London, WC2A 2AE, United Kingdom. Email: k.kalogeropoulos@lse.ac.uk

‡Corresponding author. ALBA Graduate Business School, The American College of Greece & University of Sussex Business School, UK. Address: 6-8 Xenias Str, 115 28 Athens, Greece. Email: nkarouzakis@alba.acg.edu.

1 Introduction

Failure of the expectation hypothesis (EH) implies that bond returns are strongly predictable[1] (see, Fama and Bliss (1987), Campbell and Shiller (1991), Cochrane and Piazzesi (2005), Ludvigson and Ng (2009), among others). In particular, Fama and Bliss (1987) and Campbell and Shiller (1991) propose forward and yield spreads as predictors and suggest that spreads have predictive power on excess returns, while Cochrane and Piazzesi (2005) use a linear combination of five forward rates as predictors and show that they can explain over 30% of bond return variation. More recently, Ludvigson and Ng (2009) find that predictors extracted from macroeconomic variables are also important to capture the predictability of bond excess returns. Such evidence, however is purely statistical.

An even more important and economically justifiable question, is whether such evident statistical predictability can be successfully turned into positive and statistically significant economic benefits for bond investors. At the moment, the answer to this vital question is negative (see, Della Corte et al. (2008), Thornton and Valente (2012), Sarno et al. (2016), Gargano et al. (2017)[2], Fulop et al. (2018)). Existing studies argue almost unequivocally, that models which utilize information coming solely from the yield curve (e.g. yields, forwards, etc.), are not capable to generate economic value to investors, out-of-sample. In particular, using a dynamic mean-variance allocation strategy, Della Corte et al. (2008), Thornton and Valente (2012) and Sarno et al. (2016) find that statistical significance is not turned into positive portfolio performance, implying that the evident statistical predictability is not translated into economic gains, when compared to the EH benchmark. Qualitatively similar conclusions are made by Gargano et al. (2017) and Fulop et al. (2018) who use alternative estimation approaches for an investor with power utility preferences. In this paper, we revisit this evident puzzling behaviour.

Dynamic Term Structure Models[3] (DTSM), are potentially powerful tools that have shown

[1] Under the EH bond excess returns are not predictable. Furthermore, for the EH to hold, the market price of risk should either be zero or constant over time (i.e. weak EH), and as such, time-varying risk premia is an indication of the EH being violated.

[2] Gargano et al. (2017) find some statistically significant economic gains for investors at specific (mainly longer) maturities when macroeconomic factors are used. However, according to Ghysels et al. (2017) and Fulop et al. (2018), such economic benefits vanish when fully revised macroeconomic information is replaced by real-time data.

[3] One of the most important classes of DTSMs is the multifactor Affine Term Structure Models (ATSM), characterized by the model of Duffie and Kan (1996), and have been extensively studied by Dai and Singleton (2000). Other important DTSMs are the affine model with regime shifts of Bansal and Zhou (2002) and the Gaussian quadratic model of Ahn et al. (2002).

considerable promise for capturing the dynamic behaviour of the term structure and as such, have become the main framework in the macroeconomics and empirical finance literature, when it comes to study a variety of questions about the interactions of asset prices, risk premia and economic variables[4].

Recent research (see, Duffee (2011), Barillas (2011), Thornton and Valente (2012), Adrian et al. (2013), Joslin et al. (2014), among others), however, suggests that yield only models and DTSMs, in particular, cannot capture the predictability of bond risk premia, since required information to predict premia is not spanned[5] by the cross section of yields, implying that more (mainly unspanned) factors are needed. In that respect, Duffee (2011) implements a five-factor yield only model which is capable to capture the hidden information in the bonds market, while Wright (2011), Barillas (2011) and Joslin et al. (2014) use measures of macroeconomic activity to predict bond excess returns.

More recently, Sarno et al. (2016) and Feunou and Fontaine (2017) implement extended versions of yield only ATSM and argue that their approaches, help ATSM capture the required predictability of excess returns in the bonds market. A similar conclusion is reached by Bauer (2017) who estimate an ATSM under alternative risk price restrictions. This is evidence that the spanning hypothesis holds true, implying that the yield curve itself spans all relevant information and as such, other risk factors (e.g. macroeconomic variables) are not essential, when it comes to forecasting future returns. However, these studies, either do not consider out-of-sample economic performance (as in Duffee (2011), Bauer (2017) and Feunou and Fontaine (2017), etc.), or, when they consider, they cannot confirm (as in Sarno et al. (2016)), whether statistical predictability is capable to offer economic benefits to bond investors, compared to the non-predictability (constant risk premia) EH benchmark.

In this paper, we estimate multifactor ATSMs and attempt to revisit, first, their ability to capture the predictability of excess returns in the bonds market and, second, their capability to turn such predictability into meaningful economic value for bond investors, thus addressing the

[4]See, for example, Duffee (2002) and Cochrane and Piazzesi (2009) for studies on the role of risk premia in interest rates, or Ang and Piazzesi (2003), and Rudebusch and Wu (2008) for studies on macroeconomic developments and monetary policy effects on the term structure of interest rates, among many others.

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

above mentioned puzzling behaviour between predictability and economic benefits.

We consider bond investors who assess investment opportunities based on historical evidence. Observing only a sample of such evidence, precludes them from knowing the true parameter values, thus, making their investment decisions in the presence of parameter uncertainty. Successfully addressing the above mentioned questions, implies the need to account for parameter uncertainty as well as model uncertainty. To do so, we propose a Bayesian learning framework[6] within a DTSM. In particular, we develop a sequential process for investors, who revise their beliefs about parameters and models, when new information arrives. This provides them the entire forecasting predictive density of bond excess returns, when making portfolio decisions, thus 'informing' their asset allocation and maximising their expected utility.

Our framework, allows us to overcome a number of important challenges reported in previous studies. First, it allows us to successfully handle model choice, which is of paramount importance when we want to effectively estimate risk premia. Bayesian model choice is carried out by calculating the marginal likelihood of the posterior distribution. This process usually involves different models of similar type, such as ATSMs with tight restrictions on the market price of risk parameters[7]. Such restrictions are imposed in order to take (full) advantage of the no-arbitrage condition, meaning, to try to link the cross-sectional and time-series variation of interest rates[8] (see, Cochrane and Piazzesi (2009), Duffee (2011), Bauer (2017)). This implies that the unrestricted, maximally flexible, specification used extensively in exiting studies, provides the weakest link between the two sets of dynamics and, as such, does not make any use of the information in the cross-section of interest rates, creating an extra layer of estimation difficulties. Our framework successfully overcomes this issue by testing the efficiency and economic importance of alternative risk price restrictions.

Second, it allows us to integrate out parameter uncertainty[9], which is now incorporated in the

[6]In particular, we adopt the Sequential Monte Carlo (SMC) framework of Chopin (2002) and Del Moral et al. (2006).

[7]See, Dai and Singleton (2000), Duffee (2002) and Ang and Piazzesi (2003) for an ad hoc way of imposing restrictions and Cochrane and Piazzesi (2009), Joslin et al. (2011), Joslin et al. (2014) and Bauer (2017) for a more systematic way.

[8]According to Cochrane and Piazzesi (2009), Duffee (2011) and Bauer (2017), this is the only way that allows us to use information from the cross-section of yields, and be able to effectively identify the time- series dynamics. If we do not impose no-arbitrage, through tight restrictions, then no information can be inferred from the cross section.

[9]Thornton and Valente (2012) report considerable time variation in parameter estimates, which causes a deteri-

posterior predictive distribution. This allows us to depart from the constant (or rolling window) conditional variance specification adopted by existing studies (see, Thornton and Valente (2012), Sarno et al. (2016)), and accommodate uncertainty on both moments, which are now time-varying and are updated on every time step.

Third, our framework directly produces the entire forecasting distribution, through predictive densities, when new information arrives. It is capable to accommodate various forecasting tasks such as forecasting several points and functions, thus assessing the predictive performance of any model, across predictive horizons. This allows us to analyze the economic benefits of excess return forecasts, for an investor with power utility preferences and assess the economic importance of the chosen restrictions, from a forecasting perspective.

Fourth, the sequential learning setup naturally provides sequential inference in a parallel way, allowing us to overcome the poor mixing and slow convergence properties as well as the inaccurate Monte Carlo calculations (that could potentially lead to highly autocorrelated samples) that conventional Bayesian estimation approaches experience[10] (see, Ang et al. (2007), Chib and Ergashev (2009), Ang and Longstaff (2011) and Bauer (2017), among others), making it fast and convenient to use.

Fifth, it is more robust and allows us to overcome the undesirable features of challenging target posteriors (e.g. multimodality and flat surface[11] which leads to locally[12] identifiable[13] models) and the estimation challenges, that have been reported by existing studies (see, Ang and Piazzesi (2003), Ang et al. (2007), Chib and Ergashev (2009), Joslin et al. (2011), Duffee and Stanton

oration on the model's out-of-sample performance.

[10]Chib and Ergashev (2009) and Bauer (2017) note that the quality of the MCMC output can by substantially improved if a Gibbs scheme is adopted, where the parameters are updated into blocks, some of them being full Gibbs steps. However, this framework cannot be used directly to address additional tasks such as model choice and forecasting.

[11]Ang and Piazzesi (2003) report that the likelihood surface is very flat in λ_0, which determines the mean of long yields.

[12]In a locally identifiable model, the likelihood function has a number of different maxima. In a globally identifiable model, the likelihood function possesses only one global maximum.

[13]In particular, the likelihood function experiences a large number of inequivalent local maxima with similar likelihood, and the function around these maxima can be quite flat among many directions in the parameter space. This is why it is common for numerical search methods, used for the optimization of the likelihood, to end up in regions of the parameter space that cannot be locally identified. This creates the need of repeatedly generating a random vector of starting values to maximize the likelihood function, in order to achieve convergence in this highly non-linear system; and even then, there is no guarantee that the global maximum has been found. However, this identification issue has been, largely, solved by the canonical setup of Joslin et al. (2011).

(2012), Hamilton and Wu (2012), Bauer (2017)), due to the large number of parameters to be estimated, the non-linear relationship (and cross-sectional restrictions) between the parameters and the yields, the short sample period and the very high persistence of interest rates[14].

Our empirical analysis is conducted using a discrete-time Gaussian no-arbitrage ATSM. For identification and econometric tractability, we adopt the canonical setup of Joslin et al. (2011), where the vector of unobserved risk factors is rotated, such that state variables are linear combinations of observed yields. Following related literature (see, Cochrane and Piazzesi (2005), Joslin et al. (2011), Joslin et al. (2014), Bauer (2017)), we estimate different models, based on alternative, economically plausible, restrictions on the price of risk parameters. To assess the statistical performance of the models, we allow them to produce out-of-sample forecasts spanning the period 1990:2008. We run predictive regressions on non-overlapping (as in Bauer and Hamilton (2017)) monthly excess bond returns across different predictive horizons and evaluate the out-of-sample predictability using Campbell and Thompson (2008) out-of-sample R^2. To investigate the economic value of the out-of-sample excess return forecasts generated by alternative models, we construct a dynamically rebalanced portfolio as in Della Corte et al. (2008) and Thornton and Valente (2012), for an investor with power utility preferences, and compute standard metrics such as certainty equivalence returns (CER) (see, Johannes et al. (2014) and Gargano et al. (2017), among others). We consider three different scenarios for investors. In the first two, short positions and leveraging are precluded (see, Thornton and Valente (2012), Gargano et al. (2017)). In the third, no portfolio constraints are imposed (see, Fulop et al. (2018)).

Our results help us infer a host of interesting conclusions. First, we find strong evidence that investors, who use specific restricted models, can use predictability to improve portfolio performance and earn economically meaningful utility gains, out-of-sample. The corresponding CERs are positive and statistical significant, across risk preferences, indicating that such models, not only provide statistical evidence of out-of-sample predictability, but also provide statistically significant economic gains to bond investors, relative to the EH benchmark. These findings are in contrast to existing literature. In particular, our results are in contrast to those reported by Thornton and Valente (2012), Gargano et al. (2017) and Fulop et al. (2018), who argue that the statistical evidence of bond return predictability is not turned into economic value for investors,

[14]It is hard to estimate the unconditional mean and the speed of mean reversion since a very persistent time series does not revert to its mean very often.

who utilise information from the yield curve only and, also, in contrast to Sarno et al. (2016) who find that yield only ATSM models, are not capable to systematically earn any economic premium for bond investors, out-of-sample.

Second, we find that all models, restricted and unrestricted, produce positive out-of-sample R^2, when compared to the EH-implied risk premia model. This indicates that ATSM outperform the EH benchmark both in- and out-of-sample, which suggests strong evidence of out-of-sample bond return predictability, using models that utilise their information content from the yield curve only. Our results are in contrast to Duffee (2011), Barillas (2011) and Joslin et al. (2014), who argue that ATSM cannot capture the evident predictability of returns, and, consistent with the conclusions of Sarno et al. (2016), Bauer (2017) and Feunou and Fontaine (2017), who suggest that their approaches help ATSM models to capture such predictability.

Third, comparing across models, our results suggest that the maximally flexible model outperforms restricted models both statistically (highest R^2), as well as, economically (highest CER values). Furthermore, models with fewer restrictions clearly outperform highly restricted models, provided that the unrestricted parameters are the ones that reflect the fact of risk premia being earned as compensation to the level (and slope) factor(s).

Fourth, our results suggest that R^2 decrease with bond maturity, while CER values increase as the maturity of the bond increases, reaching their peak values at the five-year maturity. Comparing across prediction horizons, we find that both out-of-sample R^2 and CER values increase with the horizon, reaching their peak values at horizons of six- and nine-months, suggesting that evidence against the EH is stronger at longer horizons.

Fifth, allowing for portfolio weights to be unconstrained, our results are even more pronounced, since CER values increase substantially, especially at shorter maturities. The most profitable bond is now the two-year maturity bond. This result is in contrast to the results implied by the constrained weights scenario, where economic performance generally increases with maturity. This is also in contrast to the equities market[15] where constraints on the portfolio weights improve performance out-of-sample.

Our paper makes several contributions to the extant literature. First, we provide a Bayesian learning framework within DTSMs, which successfully handles parameter uncertainty and model

[15]See, Avramov (2003), among others, who suggest that imposing portfolio constraints, considerably reduces discrepancies among specifications, thus, improving portfolio performance and Johannes et al. (2014) who argue that out-of-sample returns are much worse when weights are constrained.

choice. While the use of particle filters in the economics, finance and macro-finance literature[16] has increased in recent years, we are unaware of prior research that uses Bayesian learning within a DTSM. Our framework enables us to work with the joint density of the state variables and model parameters, and as such, to jointly estimate the cross-section and time-series dynamics, thus accounting for no-arbitrage, while at the same time, it naturally provides sequential inference in a parallel way, making it fast and convenient to use. In that respect, compared to existing literature (see, Thornton and Valente (2012), Adrian et al. (2013) and Sarno et al. (2016)), our framework directly addresses forecasting since its output provides the entire forecasting distribution (and estimates of the expectations) at each time step, across prediction horizons.

Second, we revisit the predictive performance of ATSM and assess the statistical and economic significance that these models imply, in light of the alternative restrictions imposed on the risk prices. We attempt to overcome the estimation, numerical and uncertainty challenges associated with these models and explore the benefits of sequential learning when it comes to the predictability of bond excess returns. Thus, our paper is related to Cochrane and Piazzesi (2005), Duffee (2011), Joslin et al. (2014), Bauer (2017), Feunou and Fontaine (2017), however, it complements their statistical analysis, by, primarily, attempting to assess the financial implications of risk price restrictions, out of sample.

Third, we address the evident puzzling behaviour between the statistical predictability of bond excess returns and the failure of the (yield only) models to turn such predictability into economic benefits for bond investors. Compared to existing studies (see, Thornton and Valente (2012), Sarno et al. (2016)), we use different modelling assumptions, different estimation methodologies and different prediction horizons. Furthermore, our work is distinct from Gargano et al. (2017) and Fulop et al. (2018) in an important respect. These studies base their analysis on predictive regressions, rather than jointly modelling the cross-section and time-series dynamics, under a term structure model.

The remainder of this paper is organised as follows. Section 2 describes the modelling framework. Section 3 discusses the sequential learning estimation procedure. Section 4 discusses the data used and the sample period. Section 5 discusses the alternative restrictions imposed on the

[16] In economics literature, see, Fernández-Villaverde and Rubio-Ramírez (2007) for the first study in the estimation of non-linear DSGE models. In finance, see, Johannes et al. (2009), Christoffersen et al. (2010), Johannes et al. (2014) for the estimation of stochastic volatility models. In macro-finance, see, Gargano et al. (2017), Fulop et al. (2018) for predictability of bond excess returns.

risk prices. Section 6 discusses how we empirically evaluate the predictive performance of alternative models and presents the empirical results on bond return predictability. Section 7 assesses the economic performance of the alternative models across different predictive horizons and presents the empirical results. Finally, section 8 concludes the paper.

2 The Dynamic Term Structure Model

The model used in our setting belongs to the no-arbitrage class of discrete-time Affine Term Structure Models (ATSM) (see, Ang and Piazzesi (2003) and Cochrane and Piazzesi (2005)). In Gaussian ATSM, under the physical probability measure $\mathbb{P}$, a $(N \times 1)$ vector of state variables, X_t, evolves according to a first-order Gaussian Vector Autoregressive (VAR) process,

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

where, $\epsilon_t \sim N(0, I_N)$, Σ is an $(N \times N)$ lower triangular matrix, μ is a $(N \times 1)$ vector and Φ is a $(N \times N)$ matrix. Under this framework, the one period risk-free interest rate r_t[17], is assumed to be an affine function of the state variables,

$$r_t = \delta_0 + \delta_1' X_t \tag{2}$$

where, δ_0 is a scalar and δ_1 is a $(N \times 1)$ vector. Absence of arbitrage implies the existence of a pricing Kernel, M_{t+1}, specified as an exponentially affine function of the state vector,

$$M_{t+1} = \exp(-r_t - \frac{1}{2}\lambda_t' \lambda_t - \lambda_t' \varepsilon_{t+1}) \tag{3}$$

with, λ_t being the time-varying market prices of risk, which is also assumed to be affine in the state vector X_t[18],

$$\lambda_t = \lambda_0 + \lambda_1 X_t \tag{4}$$

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

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

where, λ_0 is a $(N \times 1)$ vector and λ_1 is a $(N \times N)$ matrix. If we assume that the pricing kernel, M_{t+1}, prices all bonds in the economy and we let, P_t^n, denote the time-t price of an n-period zero-coupon bond, then the price of the bond is computed from, $P_t^{n+1} = E_t\left(M_{t+1}P_{t+1}^n\right)$. Then, it follows that bond prices are exponential affine functions of the state vector (see, Duffie and Kan (1996)),

$$P_t^n = \exp(A_n + B_n' X_t) \tag{5}$$

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

$$A_{n+1} = A_n + B_n'(\mu - \Sigma\lambda_0) + \frac{1}{2}B_n'\Sigma\Sigma' B_n - \delta_0 \tag{6}$$

$$B_{n+1} = (\Phi - \Sigma\lambda_1)' B_n - \delta_1 \tag{7}$$

with $A_1 = -\delta_0$ and $B_1 = -\delta_1$. This implies that the $\mathbb{Q}$ dynamics of the state vector are given by,

$$X_t = \mu^{\mathbb{Q}} + \Phi^{\mathbb{Q}} X_{t-1} + \Sigma\epsilon_t^{\mathbb{Q}} \tag{8}$$

where, $\mu^{\mathbb{Q}} = \mu - \Sigma\lambda_0$, $\Phi^{\mathbb{Q}} = \Phi - \Sigma\lambda_1$ and $\epsilon_t^{\mathbb{Q}} \sim N(0, I_N)$. The continuously compounded $(J \times 1)$ vector of yields, y_t^n, is also an affine function of the state vector,

$$y_t^n = -\frac{\log P_t^n}{n} = A_{n,X} + B_{n,X}' X_t \tag{9}$$

where, the loading $(J \times 1)$ vector, $A_{n,X}$ and the loading $(J \times N)$ matrix, $B_{n,X}$, are calculated using the above recursions as, $A_{n,X} = -A_n/n$ and $B_{n,X} = -B_n/n$.

As discussed earlier, estimation and identification of these type of DTSM models has been proven to be challenging (see, Ang and Piazzesi (2003), Ang et al. (2007), Chib and Ergashev (2009), Duffee and Stanton (2012), Hamilton and Wu (2012), Bauer (2017)), especially when ATSM are expressed in terms of a 'latent' state vector, which leads to models that are only 'locally' identifiable.

To overcome this problem, such that all parameters are econometrically identified, additional normalizing restrictions need to be imposed. We follow the canonical setup of Joslin et al. (2011) and rotate the vector of unobserved state variables, X_t, such that they are linear combinations of the observed yields, and as such, are perfectly priced by the no-arbitrage restrictions. In particular, we rotate X_t to the first three principal components (PCs) of observed yields, as, $\mathcal{P}_t = W y_t = W A_{n,X} + W B_{n,X} X_t$, where, W denoting an $(N \times J)$ matrix that contains the loadings of the first three PCs of yields, and $A_{n,X}$ and $B_{n,X}$ are given by equations above. This allows us to re-write the yield equation (9) as a function of the observable risk factors, $\mathcal{P}_t$, as,

$$y_t^n = A_{n,\mathcal{P}} + B_{n,\mathcal{P}} \mathcal{P}_t \tag{10}$$

where, the loadings $A_{n,\mathcal{P}}$ and $B_{n,\mathcal{P}}$ are given below[19] (see, Appendix A for the derivation),

$$A_{n,\mathcal{P}} = A_{n,X} - B_{n,x}(W B_{n,x})^{-1}(W A_{n,x}) \tag{11}$$

$$B_{n,\mathcal{P}} = B_{n,x}(W B_{n,x})^{-1} \tag{12}$$

Furthermore, the risk-neutral dynamics of $\mathcal{P}_t$ are given as (see, Appendix A for the derivation),

$$\mathcal{P}_t = \mu_{\mathcal{P}}^{\mathbb{Q}} + \Phi_{\mathcal{P}}^{\mathbb{Q}} \mathcal{P}_{t-1} + \Sigma_{\mathcal{P}} \epsilon_t^{\mathbb{Q}} \tag{13}$$

where,

$$\mu_{\mathcal{P}}^{\mathbb{Q}} = (I - \Phi_{\mathcal{P}}^{\mathbb{Q}})(W A_{n,X}) + (W B_{n,x})\mu^{\mathbb{Q}} \tag{14}$$

$$\Phi_{\mathcal{P}}^{\mathbb{Q}} = (W B_{n,x})\Phi^{\mathbb{Q}}(W B_{n,x})^{-1} \tag{15}$$

$$\Sigma_{\mathcal{P}} = (W B_{n,x})\Sigma \tag{16}$$

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

$$r_t = \delta_{0\mathcal{P}} + \delta'_{1\mathcal{P}} \mathcal{P}_t \tag{17}$$

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

with,

$$\delta_{0\mathcal{P}} = \delta_0 - \delta_1'(WB_{n,x})^{-1}(WA_{n,X}) \tag{18}$$

$$\delta_{1\mathcal{P}}' = (WB_{n,x}')^{-1}\delta_1 \tag{19}$$

and the market price of risk specification becomes,

$$\lambda_t = \lambda_{0\mathcal{P}} + \lambda_{1\mathcal{P}}\mathcal{P}_t \tag{20}$$

where,

$$\lambda_{0\mathcal{P}} = \lambda_0 - \lambda_1(WB_{n,x})^{-1}(WA_{n,X}) \tag{21}$$

$$\lambda_{1\mathcal{P}} = \lambda_1(WB_{n,x})^{-1} \tag{22}$$

In our setting, we also follow one of the identification schemes proposed in Joslin et al. (2011) (see, proposition 1), where the short rate is the sum of the state variables, given as, $r_t = iX_t$, with i being a vector of ones, and the parameters $\mu^{\mathbb{Q}}$ and $\Phi^{\mathbb{Q}}$ of the state vector's $\mathbb{Q}$-dynamics, given as, $\mu^{\mathbb{Q}} = [k_{\infty}^{\mathbb{Q}}, 0, 0]$ and $\Phi^{\mathbb{Q}} = diag(g^{\mathbb{Q}})$, where, $g^{\mathbb{Q}}$ denotes a $(N \times 1)$ vector containing the real and distinct eigenvalues of $\Phi^{\mathbb{Q}}$ [20].

At this point, it is important to mention an assumption we make. In our setup, yields are observed without any measurement error. Doing so, allows us to separate the estimation of the parameters under the physical measure $\mathbb{P}$, which can be obtained through simple Ordinary Least Squares (OLS) estimation[21], from the ones under the risk-neutral measure $\mathbb{Q}$. However, an N-dimensional observable state vector, cannot perfectly price $J > N$ yields, and as such, we further assume that the $(J - N)$ bond yields used in the estimation are observed with i.i.d. measurement errors such that, $y_t^n = \widehat{y}_t^n + e_{t,n}$, with $e_{t,n} \sim N(0, \sigma_e^2 I_{J-N})$.

One of the main advantages of Joslin et al. (2011) methodology, also adopted in this paper, is that the rotation to the observable state variables (i.e. PCs), allows for the separation of

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

[21] According to Joslin et al. (2011),the OLS estimates of parameters μ and Φ are almost identical to those estimated through the maximum likelihood (ML).

those parameters governing the risk neutral pricing of yields from those driving the observed risk factors. This allows the $\mathbb{Q}$ parameters to be estimated from time series data only and as such, been unaffected by the no-arbitrage cross-sectional constraints. This is also directly mapped on the joint likelihood. Denoting $Y = \{y_t, \mathcal{P}_t : t = 0, 1, \dots, T\}$, we get the conditional on Y_0 that the joint likelihood is given by

$$\begin{aligned} f(Y|\theta) &= \prod_{t=1}^{n} f(y_t|y_{t-1};\theta) \text{ where} \\ f(y_t|y_{t-1};\theta) &= f(y_t|\mathcal{P}_t; k_\infty^{\mathbb{Q}}, g^{\mathbb{Q}}, \Sigma, \sigma_e^2) \times f(\mathcal{P}_t|\mathcal{P}_{t-1}; \mu, \Phi, \Sigma) \end{aligned} \tag{23}$$

where, $\theta = \{\sigma_e^2, k_\infty^{\mathbb{Q}}, g^{\mathbb{Q}}, \Sigma, \mu, \Phi\}$. The first component captures the cross-sectional dynamics of the risk factors and the yields ($\mathbb{Q}$-likelihood) and the second component captures the time-series dynamics of the observed risk factors ($\mathbb{P}$-likelihood). The logarithms of the two terms are expressed as,

$$\log f(y_t|\mathcal{P}_t; k_\infty^{\mathbb{Q}}, g^{\mathbb{Q}}, \Sigma, \sigma_e^2) = const - (J - N)\log(\sigma_e^2) - \frac{1}{2\sigma_e^2}\|y_t - A_{n,\mathcal{P}} - B_{n,\mathcal{P}}\mathcal{P}_t\|^2 \tag{24}$$

$$\log f(\mathcal{P}_t|\mathcal{P}_{t-1}; \mu, \Phi, \Sigma) = const - \frac{1}{2}\log|\Sigma\Sigma'| - \frac{1}{2}\|\Sigma^{-1}\left(\mathcal{P}_t - \mu - \Phi\mathcal{P}_{t-1}\right)\|^2 \tag{25}$$

3 Estimation

Recent existing studies (e.g. Joslin et al. (2011), Joslin et al. (2014), among others) estimate the parameters using a two step Maximum Likelihood Estimation (MLE) process, where they first estimate the $\mathbb{P}$ parameters (i.e. μ and Φ) using standard OLS regressions and at a second stage, they estimate the remaining $\mathbb{Q}$ parameters. The first step also provides some initial estimates for the sample conditional variance of $\mathcal{P}_t$, Σ, which are used in the second step as starting values for the population variance in a suitable optimisation routine.

3.1 Specification of priors

In this paper we consider an alternative approach adopting the Bayesian framework. Low informative priors on each θ component can be assigned although some relevant information is available in this context (see, Chib and Ergashev (2009)). We first transform all restricted range parameters so that they have unrestricted range. Specifically, we work in the log-scale of $k_\infty^{\mathbb{Q}}$ and consider a

Cholesky factorisation of Σ where again the diagonal elements are transformed to the real line. In order to preserve the ordering of the eigenvalues $g^{\mathbb{Q}}$ we apply a reparameterisation and work with their increments that are again transformed to the real line. Next, independent Normal distributions with zero means and large variance are assigned to each component of θ with the exception of σ_e^2 where we assign the conjugate Inverse-Gamma prior as in Bauer (2017) with parameters $\alpha_0/2$ and $\beta_0/2$.

3.2 Markov Chain Monte Carlo scheme

Combining the likelihood provided by (23) and priors , denoted by $\pi(\theta)$, provide the posterior up to proportionality,

$$\pi(\theta|Y) = \pi(\theta) f(y_0|\mathcal{P}_0; k_\infty^{\mathbb{Q}}, g^{\mathbb{Q}}, \Sigma, \sigma_e^2) \prod_{t=1}^{n} f(y_t|\mathcal{P}_t; k_\infty^{\mathbb{Q}}, g^{\mathbb{Q}}, \Sigma, \sigma_e^2) \times f(\mathcal{P}_t|\mathcal{P}_{t-1}; \mu, \Phi, \Sigma). \tag{26}$$

As the above posterior is not available in closed form, methods such as MCMC can be used to draw samples from it. The posterior samples that can then be used to calculate expectations of interest $E[g(\theta)|Y]$ (provided that they exist) such as the posterior mean, variance and percentiles using Monte Carlo. Note however that the MCMC output is not guaranteed to lead to accurate Monte Carlo calculations since the corresponding Markov chain may have poor mixing and convergence properties thus leading to highly autocorrelated samples. It is therefore essential to construct a suitable MCMC algorithm that does not exhibit such issues. Various studies (see, for example Chib and Ergashev (2009) and Bauer (2017)), note the substantial improvement to the quality of the MCMC output if a Gibbs scheme is adopted, where the parameters are updated into blocks, some of them being full Gibbs steps. For the remaining block, independence samplers maybe constructed using the MLE and the Hessian. Such an MCMC algorithm is described in Table 3.1.

3.3 Sequential framework

The MCMC scheme in Table 3.1 performs reasonably well and can be used to obtain posterior summaries from (26). Nevertheless, it cannot be used directly to address additional tasks such as model choice and forecasting. Bayesian model choice is carried out by calculating Bayes factor or, equivalently, the model evidence or else marginal likelihood $\pi(Y) = \int_\theta f(Y|\theta)\pi(\theta)d\theta$. Numerous methods are available to estimate $\pi(Y)$, mainly via Monte Carlo, but they typically require additional work. In some cases, usually involving different models of similar type such as ATSMs

Initialise all values of θ. Then at time each iteration of the algorithm

(a) *Update σ_e^2 from its full conditional distribution that can be shown to be the Inverse Gamma distribution with parameters $\alpha + n$ and $\beta + \sum_{t,n}(y_t^n - \hat{y}_t^n)^2$.*

(b) *Update λ_t from its full conditional distribution that can be shown to be the Normal distribution with mean and variance provided from the restricted VAR framework, see for example Bauer (2017).*

(c) *Update Σ using an independence sampler based on the MLE and its Hessian obtained before running the MCMC.*

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

Table 3.1: Markov Chain Monte Carlo scheme for Gaussian Affine Term Structure Models

with restrictions, an alternative approach is to extend this algorithm to a reversible jump MCMC scheme by constructing moves between models with potentially different number of parameters (see, Bauer (2017)).

Bayesian forecasting is performed via the posterior predictive distribution.

$$f(Y_{t+1}|Y_{0:t}) = \int f(Y_{t+1}|\theta)\pi(\theta|Y_{0:t})d\theta, \tag{27}$$

where $Y_{0:t} = (Y_0, Y_1 \dots, Y_t)$ such that $Y_{0:T} = Y$. Note that (27) incorporates parameter uncertainty by integrating θ out according to $\pi(\theta|Y_{0:t})$. Usually prediction is carried out by expectations with respect to (27), e.g. $E(Y_{t+1}|T_{0:t})$. Since (27) is is typically not available in closed form, Monte Carlo can be used in the presence of samples from $\pi(\theta|Y_{0:t})$. This process may accommodate various forecasting tasks; for example forecasting several points, functions thereof and potentially further ahead in the future. A typical forecasting evaluation exercise requires taking all the consecutive times t from the nearest integer of -say $T/2$- to $T-1$. In each of these times, $Y_{0:t}$ serves as the training sample, and points of Y after t are used to evaluate the forecasts. Hence, carrying out such a task requires sample samples from (27), and therefore from $\pi(\theta|Y_{0:t})$, for several times t. Note that this procedure can be quite laborious, in some cases infeasible, if MCMC alone is to be used; this implies that algorithms such as the one it Table 3.1 will have to be run for each t separately.

An alternative approach that can handle both model choice and forecasting assessment tasks is to use sequential Monte Carlo (see, Chopin (2002) and Del Moral et al. (2006)) to sample from the sequence of distributions $\pi(\theta|Y_{0:t})$ for $t = 0, 1, \ldots, T$. The Iterated Batch Importance Sampling (IBIS) of Chopin (2002) algorithm, see also Del Moral et al. (2006) for a more general framework, is provided in Table 3.2. The degeneracy criterion is usually defined through the Effective Sample

---

Sample with an initial value $\theta^i \sim \pi(\theta)$ and set $\omega^i \leftarrow 1$ for $i = 1, \ldots, N_\theta$. Then at time each time t ($t = 1, \ldots, T$)

(a) Compute the incremental weights and the weighted average

$$u_t(\theta^i) = f(Y_t|Y_{0:t-1}, \theta^i) = f\big(Y_t|Y_{t-1}, \theta^i\big)$$

$$L_t = \frac{1}{\sum_{i=1}^{N_\theta} \omega^i} \sum_{i=1}^{N_\theta} \omega^i u_t(\theta^i).$$

(b) Update the importance weights $\omega^i \leftarrow \omega^i u_t(\theta^i)$,

(c) If some degeneracy criterion is fulfilled, sample $\tilde{\theta}^i$ from the mixture

$$\frac{1}{\sum_{i=1}^{N_\theta} \omega^i} \sum_{i=1}^{N_\theta} \omega^i K_t(\theta^i, \cdot)$$

where K_t is a MCMC kernel. Finally set $\big(\theta^i, 1\big) \leftarrow \big(\tilde{\theta}^i, 1\big)$.

Table 3.2: IBIS algorithm

Size (ESS) which is equal to $(\sum_i \omega_1)^2 / \sum_i \omega_i^2$, and is of the form ESS $< \alpha N_\theta$ for some $\alpha \in (0, 1)$. Step (c) of the algorithm can be performed in two sub-steps:

1. *Resampling:* Resample the set of θ^i's according to their weights ω_i's, e.g. using multinomial resampling. This process results in a new set of θ^i's where the old θ^i's with high weights tend to appear multiple times, whereas the ones with low weights are likely to be excluded.

2. *Jittering:* Further replace the set of θ_i's by running a MCMC algorithm for a few iterations starting at each of θ_i and setting the last sample it provides as $\tilde{\theta}_i$.

The output of the IBIS algorithm can provide estimates of the expectations $E[g(\theta)|Y_{0:t}]$ for all t using the estimator $\sum_i [\omega^i g(\theta^i)] / \sum_i \omega^i$. Chopin (2004) shows consistency and asymptotic normality of this estimator as $N_\theta \to \infty$ for all appropriately integrable $g(\cdot)$. In the case forecasting the values of Y at time $t+1$ using information up to time t, note that propagating the samples θ^i using the likelihood (23) will provide samples of Y_{t+1}^i according to the posterior predictive distribution of (27). Expectations with respect to (27) of the form $E[g(Y_{t+1})|Y_{0:t}]$ can therefore be calculated by $\sum_i [\omega^i g(Y_{t+1}^i)] / \sum_i \omega^i$ for all t. Moreover, each L_t, computed in Step (a), is a consistent and asymptotically normal estimator of the model evidence $f(Y_t|Y_{0:t-1})$. Hence, the IBIS algorithm for tasks of model choice and assessment of predictive performance. An additional benefit provided by sequential Monte Carlo is that it provides an alternative choice when MCMC algorithms have poor mixing and convergence properties and is general more robust when the target posterior is challenging, e.g. multimodal. Note that despite the fact that this approach also uses MCMC, it is substantially more efficient than running the MCMC for each t.

4 Data

Our empirical study is based on a data set that comprises monthly observations of zero-coupon U.S. Treasury yields with maturities of 1-year, 2-year, 3-year, 4-year, 5-year, 7-year and 10-year, spanning the period January 1990 to December 2007. Following the related literature (see, Joslin et al. (2011) and Bauer (2017)), we do not include in our sample the period of the recent financial crisis. The reason is that after 2008 the interest rates market has experienced a quick downward move at the short end of the yield curve (zero-lower bound) which has also lasted for a long period of time. Kim and Singleton (2012) and Bauer and Rudebusch (2016) argue that Gaussian affine models are not capable to deal with almost zero rates since they generate large negative risk premiums.

5 Market Price of Risk Restrictions

The market price of risk specification and the associated restrictions related to it, is an important component of DTSM, especially when it comes to estimation of risk premia. This section discusses the alternative restrictions used on the market price of risk parameters. Earlier studies (see, Dai and Singleton (2000), Duffee (2002), Ang and Piazzesi (2003), Kim and Wright (2005)) have, mainly, focused on imposing zero ad-hoc restrictions on the parameters governing the dynamics of the risk premia. A common practice used, is to first estimate a maximally flexible model (i.e. a model with unrestricted risk prices), and in a second step, to re-estimate it by setting to zero those parameters that have large standard errors. However, the process of imposing such restrictions in an ad-hoc way has been recently criticised (see, Kim and Singleton (2012), Bauer (2017)), since it has raised concerns about, first, the joint significance of the constraints, second, the magnitude of the associated standard errors[22] and, third, the economic meaning (i.e. not motivated by the economic theory) of the estimated parameters and the resulting state variables.

Only recently, a few studies have investigated more systematic approaches on how to impose restrictions on the market price of risk parameters. These approaches are mainly based on alternative information criteria. In particular, Bauer (2017) propose a Bayesian econometric approach to select restrictions, by using draws from three alternative model selection samplers[23]. Empirical evidence suggests that only one (or one to three) out of twelve parameters of the market price of risk (essentially affine) specification is non-zero (i.e. unrestricted), which implies that the data call for tightly restricted models. In particular, this parameter is the one that determines how the slope factor influences the price of level risk, which shows evidence of level risk been priced (as well as being time-varying), consistent with Fama and Bliss (1987) and Campbell and Shiller (1991).

Similar tight restrictions, introduced, partly, by prior empirical analysis, are also imposed by Cochrane and Piazzesi (2009), whose risk premia specification is the most restricted one, given that only one parameter in λ_0 and one parameter in λ_1 are allowed to be non-zero. Furthermore, the restriction on λ_1 reflects the fact of risk premia being earned as compensation to the level factor only, in line with Duffee (2011) and Bauer (2017). An alternative approach is followed by Chib and Ergashev (2009), who impose strong prior restrictions such that the yield curve is (on average) upward sloping[24], an assumption that is empirically and economically plausible.

Finally, similar systematic approaches have been used in the macro-finance DTSM literature[25]. Joslin et al. (2014) use model selection methodologies and select all possible zero restrictions[26] based on different Bayesian information criteria[27]. Similar restrictions are also tested by Bauer

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

[23]Samplers proposed by Bauer (2017) are the Gibbs Variable Selection, the Stochastic Search Variable Selection and the Reversible-Jump MCMC.

[24]Chib and Kang (2009) and Chib and Kang (2012) also use similar prior restrictions to reflect the meaningful assumption of a positive term premium

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

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

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

and Rudebusch (2016), however their results are not affected by the choice of restriction and, as such, they work with the maximally flexible specification only.

In this paper, we follow related literature and try alternative, economically plausible, zero restrictions for the market price of risk parameters. On our initial specification we allow the risk prices to be completely unrestricted (i.e. maximally flexible model), in line with Duffee (2011) and Bauer (2017), and we denote this model as M_0. Then, such model is tested against alternative zero restriction models, proposed in existing studies (see, Cochrane and Piazzesi (2009), Bauer (2017)), i.e. models with specific factors driving the variation of risk premia[28]. In particular, these models are the model with parameters $\lambda_{0,1}$, $\lambda_{1,1}$ and $\lambda_{1,2}$ being unrestricted, denoted by M_1, the model with parameters $\lambda_{0,1}$ and $\lambda_{1,2}$ being unrestricted, denoted by M_2, the model with parameter $\lambda_{1,1}$ and $\lambda_{1,2}$ being unrestricted, denoted by M_3, the model with only parameter $\lambda_{1,2}$ being unrestricted, denoted by M_4, the model with only parameter $\lambda_{1,1}$ being unrestricted, denoted by M_5 and the model with parameters $\lambda_{1,1}$, $\lambda_{1,2}$ and $\lambda_{2,3}$ being unrestricted, denoted by M_6. Note that the latter model, M_6, allows both the level and slope risks to be priced, instead of just level risk, as in models, M_1 to M_5.

A valid question arises on whether the same restrictions are also effective for the forecasting performance of the model, meaning, whether or not the models selected based on the in-sample procedure should be used on the out-of-sample forecasting exercise. A thorough analysis is presented in Duffee (2011) where a number of Gaussian DTSM models, both restricted and unrestricted, are compared. Empirical tests suggest that the choice of the no-arbitrage restrictions does not influence the out-of-sample performance of the models, given that they produce forecasts with indistinguishable differences[29].

Despite the importance of restrictions on the forecasting proccess, empirical studies are still inconclusive. The methodology applied in this paper allows us to test the efficiency and economic importance of the chosen restrictions, from a forecasting perspective. In particular, we select the in-sample zero restrictions used in existing studies and attempt to verify which one, if any, can be effectively used for the out-of-sample exercise.

---

[28] Duffee (2010) who impose constraints on Sharpe ratios and Joslin et al. (2014) who study unspanned macro-finance models, find evidence of two priced factors, i.e. the level and the slope

[29] This, however, is not in line with evidence of existing studies on monetary policy effects (see, Piazzesi et al. (2006), Ang and Longstaff (2011), Orphanides and Wei (2012), which suggest that restrictions selected on the in-sample process may not be economically plausible, around periods of monetary policy shifts, interventions or fragile economic conditions.

5.1 Restrictions and Persistence of Interest Rates

Comparison between models is important when it comes to the degree of persistence of risk factors, since it gives us an indication of the short rate expectations and time-varying risk premia. Table 8.1 displays the persistence among models under the risk neutral ($\mathbb{Q}$) and physical dynamics ($\mathbb{P}$), measured by the largest eigenvalue of matrices $\Phi_{\mathcal{P}}^{\mathbb{Q}}$ and $\Phi_{\mathcal{P}}^{\mathbb{P}}$.

A direct comparison between models suggests that the degree of persistence under $\mathbb{Q}$ is higher compared to the one under the physical measure $\mathbb{P}$. This difference is informative of time-varying risk premia (or market price of risk being time-varying) embedded in bond yields and suggests that market price of risk is potentially an important factor to explain yield variation.

A second important observation comes from the risk-neutral estimates of the largest eigenvalues, which are very similar across models and indifferent of any restrictions imposed. This implies that $\mathbb{Q}$-persistence is not affected by the no-arbitrage restrictions, which suggests that $\mathbb{Q}$-distribution parameters are mainly determined by the cross-sectional restrictions on bond yields rather than the time-series ones implied by the physical measure. This means that any difference in risk premiums are due to either the loadings of yields onto the pricing factors or to feedback matrices under $\mathbb{P}$. However, loadings are largely indifferent across models since they are only affected/determined by the risk neutral parameters. Thus, we expect that the difference is mainly driven by the feedback matrices under $\mathbb{P}$. In contrast to the $\mathbb{Q}$-persistence, the $\mathbb{P}$ eigenvalues possess important differences among models, which implies that $\mathbb{P}$-persistence is largely influenced by restrictions imposed. In particular, restricted models exhibit higher persistence than the maximally-flexible model, in all cases but model M_4, suggesting that the plausible and empirically valid restrictions imposed on the market price of risk parameters increase the persistence of interest rates and bring the $\mathbb{Q}$ and $\mathbb{P}$ dynamics closer together, thus tightening the gap between the cross-section and time series and, as such, links their variation.

Comparing across models sheds light to the effect of the selected restrictions when it comes to the true ($\mathbb{P}$) persistence of the interest rates. The $\mathbb{P}$-persistence of the maximally flexible model (M_0) is low (0.9846), suggesting that short rate expectations are nearly constant, which implies that any variation in long term rates is due to variation in risk premia. This is an empirically puzzling observation which indicates that DTSM are not capable to properly predict variation of long-term rates. However, imposing the zero restrictions as in highly restricted models M_3, M_4 and M_5 increases the associated largest eigenvalue to 0.9929, 0.9923 and 0.9919 respectively and, thus, bringing the persistence of the interest rates close to their true values. In fact, low restricted models, meaning models with one or two free parameters exhibit higher persistence than the maximally flexible one, while low constrained models (e.g. M_1 and M_6) exhibit lower persistence than model M_0. This further suggests that restrictions enhance the time variation of short rate expectations and help to deteriorate the role of risk premia on long rate movements. This is in line with policy making, given that constant expectation of short rates is economically implausible since short rates are also driven by central banks' intervation. This suggests that restrictions imply a larger role to the expectation component and a smaller role to the risk premia component, compared to the unrestricted models and as such, the puzzling implausible disadvantage of DTSM is partly corrected.

Furthermore, table 8.1 reports confidence intervals (CI) among models. We observe that unrestricted models exhibit larger CIs, which suggests time-series information alone is not enough to estimate interest rates. Restricted models have tight CIs, since no-arbitrage, through zero restrictions, makes cross-section information important to explain time series properties of interest rates.

6 Bond Return Predictability

Failure of the expectation hypothesis implies that bond returns are strongly predictable (see, Fama and Bliss (1987), Campbell and Shiller (1991), Cochrane and Piazzesi (2005), Ludvigson and Ng (2009), among others). This is an important stylised fact of the bonds market, and as such, it should be captured by our modelling framework. A valid and important question however, is whether all alternative (yield-only) ATSMs tested can capture this return predictability; in other words whether tight restrictions influence the predictability performance of ATSMs, given the less flexible specification of the market prices of risk.

In order to investigate the predictability of bond returns across models, we follow existing literature and run predictive regressions on non-overlapping (as in Bauer and Hamilton (2017)) monthly excess bond returns across different predictive horizons, ranging from one-month to twelve-months. The main difference among approaches used by previous studies is down to the predictors used in the regression. In particular, Fama and Bliss (1987) and Campbell and Shiller (1991) propose forward and yield spreads as predictors and suggest that spreads have predictive power on excess returns, while Cochrane and Piazzesi (2005) use a linear combination of five forward rates as predictors and show that they can explain over 30% of bond return variation. Furthermore, Joslin et al. (2011) and Adrian et al. (2013) use principal components of yields in order to predict bond excess returns. Given that we work under a three factor yield only model, we mainly follow Joslin et al. (2011), Adrian et al. (2013) and Bauer (2017) and use as predictors the first three PCs of the yield curve, corresponding to the three risk factors.

In line with existing literature, we define the holding period return from buying an n-year bond at time-t and selling is at time $(t+h)$ as,

$$r^n_{t,t+h} = p^{n-h}_{t+h} - p^n_t \tag{28}$$

where, p^{n-h}_{t+h} is the price of the $(n-h)$-period bond at time $(t+h)$ and p^n_t is the price of the n-period bond at time t. Furthermore, we define the continuously compounded excess return of an n-year bond as the difference between the holding period return of the n-year bond and the h-period yield as,

$$rx^n_{t,t+h} = -(n-h)y^{n-h}_{t+h} + ny^n_t - hy^h_t \tag{29}$$

where, the continuously compounded yield y^n_t has been defined in equation (10). On our empirical tests, we assume h to range from 1-month to 12-months, and n to be 2, 3, 4, 5, 7 and 10 years. Similarly, fitted excess returns $\widehat{rx}^n_{t,t+h}$ are calculated using model implied yields as,

$$\widehat{rx}^n_{t,t+h} = -(n-h)\widehat{y}^{n-h}_{t+h} + n\widehat{y}^n_t - h\widehat{y}^h_t \tag{30}$$

$$= A_{n-h} - A_n + A_h + B'_{n-h}\mathcal{P}_{t+h} - (B_n - B_h)'\mathcal{P}_t \tag{31}$$

The regression specification takes the form,

$$rx^n_{t,t+h} = \alpha^n + \beta^n \mathcal{P}_t + e^n_t \tag{32}$$

where, $rx^n_{t,t+h}$ are the h-month holding period returns, in excess of the h-period yield, n is the maturity of the bond and $\mathcal{P}_t$ is the set of the predetermined predictors, chosen to be the first three PCs of the yield curve. Finally, e^n_t is the error term and α^n and β^n are the regression coefficients.

What is of high importance is to access the predictability of our modeling framework under the alterntive restrictions. To do so we calculate model implied R^2, using the following predictive regression,

$$rx^n_{t,t+h} = \alpha^n + \beta^n \mathbb{E}_t[\widehat{rx}^n_{t,t+h}] + e^n_t \tag{33}$$

where, in this case realized bond excess returns are regressed on model implied expected excess returns given by,

$$\mathbb{E}_t[\widehat{rx}^n_{t,t+h}] = -(n-h)\widehat{y}^{n-h}_{t+h|t} + n\widehat{y}^n_t - h\widehat{y}^h_t \tag{34}$$

$$= A_{n-h} - A_n + A_h + B'_{n-h}\mathcal{P}_{t+h|t} - (B_n - B_h)'\mathcal{P}_t \tag{35}$$

It is important to mention at this point that most existing studies have tested predictability of bond excess returns using a horizon of one year. However, according to Bauer and Hamilton (2017) overlapping bond returns create further estimation issues when predictors are persistent. Therefore, in our predictability analysis we consider a monthly horizon and as such excess returns are non-overlapping monthly returns.

We measure the predictive accuracy of the bond excess return forecasts relative to an empirical benchmark. We follow related literature and assume the benchmark to be the EH which implies the historical mean being the optimal forecast of bond excess returns, given as,

$$\overline{rx}^n_{t+1} = \frac{1}{t}\sum_{j=1}^{t} rx^n_t \tag{36}$$

In particular we forecast from the EH, that projects excess returns on a constant. Following Campbell and Thompson (2008), we compute the out-of-sample R^2 of our models relative to the EH model, as below,

$$R^2_{os} = 1 - \frac{\sum_{s=t_0}^{t}(rx^n_{s+1} - \widehat{rx}^n_{s+1})^2}{\sum_{s=t_0}^{t}(rx^n_{s+1} - \overline{rx}^n_{s+1})^2} \tag{37}$$

Positive values of this statistic implies that the forecast outperforms the historical average forecast and suggests evidence of time-varying return predictability.

Table 8.2 shows results for the R^2_{os} across models, at the 1-month prediction horizon. We find that in all cases R^2_{os} are positive when compared to the EH benchmark and therefore, all models, restricted and unrestricted, outperform the EH-implied risk premia in- and out-of-sample. These results suggest strong evidence of out-of-sample bond return predictability using information from the yield curve only.

Our findings are stronger than previous studies (see, Sarno et al. (2016), Fulop et al. (2018)) who find some statistical evidence (mainly depending on maturity and prediction horizon) of out-of-sample bond return predictability, for models that only utilize information from the yield curve. This suggests that our methodology through the joint modelling of the $\mathbb{P}$ and $\mathbb{Q}$ dynamics together with the sequential learning estimation procedure, help models, which utilize information from the yield curve only, to generate accurate forecasts, improving their predictive accuracy across maturities and prediction horizons.

Furthermore, our results show that R^2_{os} decrease with bond maturity. Looking across models, the highest R^2_{os} are found at the 2-year maturity bond, while the lowest at the 10-year maturity. In particular, for the maximally flexible model M_0, R^2_{os} decreases from 0.28 (0.20), and highly significant, at 2-year (3-year) maturity to 0.09 (0.03), and insignificant, at 7-year (10-year) maturity. Similarly, for risk price restricted model M_1, R^2_{os} substantially deteriorates in value, from 0.26 (0.18), and highly significant, at 2-year (3-year) maturity, down to 0.09 (0.05) for the longer maturity bonds, 7-year (10-year). Qualitatively similar results are observed for all models, regardless the restrictions imposed.

Comparing R^2_{os} across models sheds light to the effect of risk price restrictions on the predictive accuracy of bond excess return forecasts relative to the EH benchmark. The highest R^2_{os} are found for the maximally flexible model M_0 and for restricted models M_1 and M_6. In particular, for the 2-year (5-year) bond, the R^2_{os} is 0.28 (0.16) for model M_0 and 0.26 (0.14) and 0.26 (0.16) for models M_1 and M_6 respectively. This suggests that models with fewer restrictions (or more free parameters) outperform highly restricted models with only one free parameter (e.g. M_4, M_5). Furthermore, small differences among models M_1 and M_6, when compared to maximally flexible model M_0 indicate that two or three free parameters are enough to explain the predictability of excess return forecasts in the bonds market, provided that these parameters are the ones which reflect the fact of risk premia being earned as compensation to the level factor only, as in M_1, or to both the level and slope factors as in model M_6.

We now extend our analysis on the direction of alternative prediction horizons, departing from the case of one-month horizon, followed by Gargano et al. (2017) and Fulop et al. (2018). Table 8.3 displays the results for the R^2_{os}, for horizons ranging from 1-month to 12-months, across different model specifications. All forecasts across models are more accurate than the EH benchmark at all different horizons tested. Our results also suggest that in all cases, R^2_{os} increase with the prediction horizon reaching their peak values at horizons of 6-months and 9-months. In particular, for the maximally flexible model M_1, for a 2-year (5-year) maturity bond, R^2_{os} increases from 0.28 (0.16) at 1-month horizon to 0.63 (0.62) and 0.59 (0.66) at 6-month and 9-month horizons respectively. Furthermore, for restricted models M_1 and M_6 the increase is qualitatively similar, with R^2_{os} for a 2-year (5-year) bond increasing from 0.26 (0.14) to 0.60 (0.59) at 1-month and 6-month horizons and from 0.26 (0.16) to 0.61 (0.60) respectively.

Comparing across alternative models, sheds light to the effect of price restrictions when it comes to the forecasting ability of the models across predictive horizons. Similar to the 1-month horizon case, our results indicate that maximally flexible model M_0 and restricted models M_1 and M_6 outperform alternative models. In particular for a 2-year maturity bond, the corresponding R^2_{os} is 0.63 (0.59) at 6-month (9-month) horizon for model M_0 and 0.61 (0.59) for model M_6, while for a 5-year maturity bond, R^2_{os} for model M_0 is 0.62 (0.66) at 6-month (9-month) horizon and for model M_6 is 0.60 (0.65). This indicates that model M_6 outperforms all models (and is similar to model M_0), across horizons and maturities, suggesting that slope factor is vital when it comes to predictability of excess bond returns across maturities, especially at longer horizons.

7 Economic Performance of Excess Return Forecasts

While statistical significance is important when it comes to predictability of excess returns in the bonds market, what is arguably of paramount importance for bond investors, is to be able to use such predictability, to generate economically significant portfolio benefits, out-of-sample. In this section, we concentrate on the economic value exercise and attempt to revisit the evident puzzling behaviour between statistical predictability and significant economic benefits for investors. Furthermore, we attempt to assess and compare the performance of asset allocations based on the maximally flexible model versus the ones that impose restrictions on the price of risk specification.

In that respect, our approach is different than Thornton and Valente (2012) and Sarno et al. (2016), who test the economic significance for an investor with mean-variance preferences[30], and conclude that statistical significance is not turned into better economic performance when com-

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

pared to the EH benchmark. It is more in line with Gargano et al. (2017) and Fulop et al. (2018) who arrive to similar conclusions for models which utilize information coming solely from the yield curve (e.g. yields, forwards, etc.).

We consider a Bayesian investor with power utility preferences,

$$U(W_{t+h}) = U(w_t^n, rx_{t+h}^n) = \frac{W_{p,t+h}^{1-\gamma}}{1-\gamma} \tag{38}$$

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

$$W_{p,t+h} = (1 - w_t^n)\exp(r_t^f) + w_t^n(\exp(r_t^f) + rx_{t,t+h}^n) \tag{39}$$

with, r_t^f being the risk-free rate. Such an investor maximises her expected utility over h-periods in the future,

$$\max E_t[U(W_{t+h})|x_{1:t}] \tag{40}$$

with, $x_{1:t} = \{rx_{1:t}^n, \mathcal{P}_{1:t|t}\}$ being a time series of returns and predictors. Then, the expected utility is computed as,

$$E_t[U(W_{t+h})|x_{1:t}] = \int U(W_{t+h})p(W_{t+h}|x_{1:t})dW_{t+h} = \int U(w_t^n, rx_{t+h}^n)p(rx_{t+h}^n|x_{1:t})drx_{t+h}^n \tag{41}$$

where, the predictive density is calculated as in equation (27).

At every time step, t, the investor evaluates the entire predictive density of bond excess returns and solves the asset allocation problem by computing the optimum portfolio weights by,

$$\widehat{w}_t^n = \arg\max \frac{1}{M}\sum_{j=1}^{M}\left\{\frac{[(1 - w_t^n)\exp(r_t^f) + w_t^n(\exp(r_t^f) + rx_{t,t+h}^n)]^{1-\gamma}}{1-\gamma}\right\} \tag{42}$$

with M being the number of particles from the predictive density of excess returns.

To compute the economic value generated by alternative models, we use the resulting optimum weights to compute the certainty equivalence returns (CER) as in Johannes et al. (2014) and Gargano et al. (2017). In particular, for each model, we define the CER as the value that equates the average utility of each model against the average utility of the EH benchmark.

We consider, three different scenarios for investors. First, we follow Thornton and Valente (2012) and restrict portfolio weights to range in the interval $[-1, 2]$, imposing a maximum leverage of 100%, such that extreme investments are prevented. Second, we restrict portfolio weights to the interval $[0, 1]$, such that short selling is not allowed, as in Gargano et al. (2017). At a third scenario, we do not impose any restrictions to investors and allow for portfolio weights to be free of constraints[31], thus allowing extreme investments, as in Fulop et al. (2018). In line with the values used in existing studies, we set the coefficient of risk aversion to be either 3, 5 or 10, across scenarios. The statistical significance of the results is based on the one-sided Diebold-Mariano test applied to the out-of-sample period.

Table 8.4 shows evidence in favour of positive out-of-sample economic benefits for bond investors. In particular, it shows the results for the annualized certainty equivalence returns (CER), across different coefficients of relative risk aversion (e.g. $\gamma = 3$, 5 and 10). Panel A presents the case of the first scenario, where weights are restricted to $[-1, 2]$. Positive values indicate that the models perform better than the EH benchmark. We find that in many cases, and in specific models, corresponding CERs are positive and statistically significant, indicating that these models not only provide statistical evidence of out-of-sample predictability, but also provide statistically significant economic gains to bond investors, relative to the EH benchmark. This result is evident across risk aversion coefficients. These findings are in contrast to previous literature (Thornton and Valente (2012), Sarno et al. (2016), Fulop et al. (2018)) which suggests that regardless of the model used, or the maturity considered, bond investors, who utilise information from the yield curve only, are not able to systematically earn any economic premium out-of-sample, when compared to the EH benchmark.

Furthermore, our results show that CERs increase with bond maturity, reaching their peak values at the 5-year maturity, which suggests that as maturity increases the economic performance improves. In particular, for model M_0, when the coefficient of risk aversion is 3, CER value is 9 basis points (bps) and significant at 5-year maturity deteriorating to one third (3 bps) of its value for a 2-year maturity bond. This evidence is consistent across models, since results are qualitatively similar.

[31](Footnote: As mentioned in Gargano et al. (2017), Johannes et al. (2014) argue that under this scenario the expected utility is not allowed to become unbounded)

Comparing across different models sheds light to the effect of the price risk restrictions on the ability of the models to produce positive and statistically significant economic gains for bond investors. The highest CER values are found for the maximally flexible model (M_0) which seems to outperform models with risk price restrictions. In particular, CERs for M_0 are positive and statistically significant and range from 3 bps to 9 bps. The maximally flexible model is then followed by restricted models M_1 (CERs ranging between 2 and 5 bps and only significant at 2-year maturity), M_2 (ranging between 1 and 4 bps and marginally significant) and M_6 (ranging from 2 bps to 6 bps and marginally significant). The results suggest that models with fewer restrictions (e.g. M_1, M_2, M_6) or with unrestricted constant risk premia components, clearly outperform models with only one or two free parameters (e.g. M_3, M_4, M_5). The later fail to provide positive economic value to investors, which also suggests that the closest the link between the $\mathbb{P}$ and $\mathbb{Q}$ dynamics, or in other words, the closest we are to the no-arbitrage condition, the less capable the model is to provide statistically significant economic gains, compared to the EH model.

Finally, an interesting conclusion is made when comparing across different models. In particular, when looking at short maturity bonds (mainly at 2-year maturity), model M_6 seems to produce quantitatively similar CER values with the maximally flexible model, and higher values compared to restricted models M_1 and M_2. Given that model M_6 is the only restricted model that allows both the level and the slope risks to be priced (models M_1 and M_2 allow for the level risk only), this might be an indication that slope risk is an important, probably vital, factor when it comes to economic benefits for bond investors. This conclusion (about the importance of slope risk) is along the direction of the, statistically evident, predictability of excess returns in the bonds market, as suggested by Fama and Bliss (1987), Campbell and Shiller (1991) and more recently by Joslin et al. (2014).

We now analyse asset allocation using the scenario of no portfolio constraints. Panel B presents CER values from the second scenario, where portfolio weights are allowed to be unconstrained. Our results in this case, are even more pronounced compared to the constrained scenario of panel A, since the CER values increase substantially, especially for the short maturity bonds. In most cases and across models and risk preferences, CERs are positive and statistical significant, suggesting economic gains to bond investors, relative to the EH benchmark. In particular, when the risk aversion coefficient is 3 for the maximally flexible model, M_0, the CER value for the

2-year maturity bond substantially increases from 3 bps to 42 bps, while for the 5-year bond, the increase is smaller but still noticeable, moving from 9 bps to 27 bps. Similar improvements on the magnitude of CERs are observed for models with risk price restrictions. For example, for model M_1, CER increases from 2 to 34 bps for the 2-year maturity bond, and from 5 bps (and not significant) to 20 bps (and significant) for the 5-year maturity. Furthermore, risk price restricted model, M_6, which allows both the level and slope risks to be priced, also experiences comparable improvements, with CERs increasing from 2 (6) bps to 34 (23) bps for a 2-year (5-year) maturity bond. These results suggest that models, which offer positive and statistically significant economic gains out-of-sample, do even more so, when portfolio weights are not subject to constraints.

Comparing across models, our results are qualitatively similar to the constrained scenario case, meaning that the maximally flexible model, still outperforms the models with risk price restrictions. Furthermore, models with fewer restrictions (e.g. M_1, M_6) clearly outperform models, where all but one, out of the risk price parameters are restricted (e.g. M_4, M_5). Furthermore, similar to the first scenario, models M_1 and M_6 provide quantitatively similar CER values, suggesting that slope is an important factor on enhancing the economic performance of bond excess return forecasts.

Finally, under the unconstrained scenario, CERs experience higher values at shorter maturities (e.g. 2-year bond) compared to longer maturities (5-year bond), which suggests that the most profitable maturity is now the 2-year maturity of the bond. This result is in contrast to the results implied by the constrained weights scenario, where economic performance generally increases with maturity. This is also in contrast to the equities market where constraints on the portfolio weights improve performance out-of-sample, in line with Gargano et al. (2017) conclusions.

We now extend our analysis on the direction of alternative prediction horizons. In particular, we depart from the 1-month prediction horizon, adopted by Gargano et al. (2017) and Fulop et al. (2018) and repeat the out-of-sample economic performance analysis for horizons ranging from 1-month to 12-months.

Table 8.5 displays the results for the annualized CER values, for coefficient of relative risk aversion, $\gamma = 3$, across different prediction horizons. Panel A presents the case of the first scenario, where portfolio weights are restricted to [-1,2] and panel B presents CER values for the case where weights are unrestricted. We find that, in all cases (models), CER values increase with the prediction horizon, suggesting that evidence against the EH is stronger over longer horizons.

In particular, as prediction horizon increases, the models not only continue to perform well and produce positive and statistically significant CERs, but also CER values are comparably higher in magnitude than in the 1-month horizon case. Furthermore, CERs reach their peak values at horizons of 6-months and 9-months, deteriorating in magnitude at the 12-month horizon. This result is consistent across models. For example, for the maximally flexible model, the CER value for a 2-year maturity bond increases (actually triples) from 3 bps at 1-month horizon to 9 bps at 6- and 9-month horizons and for a 7-year maturity bond, CERs increase from 9 bps (and not significant) at 1-month prediction horizon to 27 bps (and significant) at 9-month horizon. Furthermore, for the restricted models M_1 (M_6), CER values for a short-maturity bond increase from 2 (3) bps at 1-month horizon to 9 (9) bps at 9-month horizon and for 7-year maturity bonds, CERs increase from 6 (6) bps at 1-month horizon to 27 (28) bps at 9-month horizon.

These findings are in contrast to Sarno et al. (2016), who find that their economic value measure, theta, decreases with the prediction horizon and, somehow, in line with Gargano et al. (2017) who suggest that at the 3-month horizon, CER values continue to be positive (and significant) for some models (mainly the macro ones), but are substantially smaller at the 12-month horizon[32].

Comparing across models sheds light to the effect of risk-price restrictions when it comes to economic gains performance across different prediction horizons. Our results suggest that as horizon increases, the outperformance (especially at longer maturities) of the maximally flexible model compared to the restricted models vanishes, since differences in CER values are comparably, almost, zero in magnitude. For example, for a 5-year (7-year) maturity bond, CER values are 21 (24) and 21 (25) bps for models M_0 and M_6 respectively, at the 6-month horizon and 21 (27) and 21 (28) bps at 9-month horizon.

Turning into the second scenario, where no portfolio constraints are imposed, panel B displays CER values across all alternative models. Our results are in line with the analysis of previous section, meaning that, if we do not impose any restrictions on portfolio weights, CERs substantially increase, especially for short maturity bonds. Not only that, but the differences at longer horizons are even larger, indicating that as horizon increases, positive CER values become higher and statistical significant, suggesting higher (out-of-sample) economic gains for bond investors compared to the EH model. In particular, for the maximally flexible model, M_0, the CER value

[32] According to Gargano et al. (2017), this is an indication of the presence of a fast moving predictable component that deteriorates in magnitude at long enough (i.e. 12-month) horizons.

for the 2-year maturity bond substantially increases from 9 (9) bps to 90 (74) bps at 6-month (9-month) horizon, while for the 5-year maturity bond increases from 21 (21) bps to 57 (58) bps. Similar improvements, are observed across models with risk price restrictions, such as for model M_1, where CER values for a 2-year bond increase from 10 (9) to 85 (81) bps and for model M_6, where CER values increase from 10 (9) to 85 (76) bps at prediction horizon of 6-month (9-month). Qualitatively similar results are found for alternative risk aversion coefficients (γ=5, 10), as shown in tables 8.6 and 8.7.

Comparing across alternative models, the maximally flexible model seems to slightly outperform restricted model M_1, and it also possess similar quantitative characteristics with model M_6, which allows both the level and slope factors to be priced, suggesting that slope risk is vital when it comes to the economic value of bond excess return forecasts, especially at longer prediction horizons. In particular, for a 2-year maturity bond the corresponding CER values are 90 (74) bps at the 6-month (9-month) horizon for model M_0 and 85 (76) bps for model M_6, while for a 5-year bond, CERs are respectively 57 (58) and 56 (60) bps. This indicates that the investment on the 2-year bond is the most profitable across models and maturities, and in particular at the 6-month prediction horizon.

8 Conclusion

In this paper we revisit the puzzling behaviour between the statistical predictability of bond excess returns and the failure to translate such predictability into economic gains for bond investors. Furthermore, we revisit the ability of yield-only ATSM to capture the predictability of excess returns in the bonds market. We implement a generic, efficient and computationally fast sequential learning framework under DTSM. We account for model uncertainty, parameter uncertainty and forecasting performance, which are important issues when it comes to predictability of bond excess returns. We compare the performance of the alternative ATSM restricted models against the expectation hypothesis benchmark and evaluate the out-of-sample statistical and economic accuracy of bond excess return forecasts.

We find strong evidence that investors, who use specific restricted models, can use predictability to improve portfolio performance and earn economically meaningful utility gains, out-of-sample. Furthermore, we find that all ATSM models, restricted and unrestricted, produce positive out-of-sample R^2, when compared to the EH-implied risk premia model, which suggests strong evidence of out-of-sample bond return predictability, using models that utilise their information content from the yield curve only. Finally, comparing across risk price specifications, we find that maximally flexible model outperforms restricted models both statistically (highest R^2), as well as, economically (highest CER values).

Appendix A

The rotation of the state vector X_t to the first three PCs of yileds, given by, $\mathcal{P}_t = Wy_t = WA_{n,X} + WB_{n,X}X_t$, implies that the state vector can be expressed as,

$$X_t = (WB_{n,x})^{-1}(\mathcal{P}_t - WA_{n,X}) = (WB_{n,x})^{-1}\mathcal{P}_t - (WB_{n,x})^{-1}(WA_{n,X}) \tag{43}$$

So, the $\mathbb{Q}$ dynamics of the state vector, $X_t = \mu^{\mathbb{Q}} + \Phi^{\mathbb{Q}}X_{t-1} + \Sigma\epsilon_t^{\mathbb{Q}}$ can now be written as,

$$(WB_{n,x})^{-1}\mathcal{P}_t - (WB_{n,x})^{-1}(WA_{n,X}) = \mu^{\mathbb{Q}} + \Phi^{\mathbb{Q}}\Big((WB_{n,x})^{-1}\mathcal{P}_{t-1} - (WB_{n,x})^{-1}(WA_{n,X})\Big) + \Sigma\epsilon_t^{\mathbb{Q}} \tag{44}$$

$$\Rightarrow (WB_{n,x})^{-1}\mathcal{P}_t = \mu^{\mathbb{Q}} + (WB_{n,x})^{-1}(WA_{n,X}) - \Phi^{\mathbb{Q}}(WB_{n,x})^{-1}(WA_{n,X}) + \Phi^{\mathbb{Q}}(WB_{n,x})^{-1}\mathcal{P}_{t-1} + \Sigma\epsilon_t^{\mathbb{Q}} \tag{45}$$

which means that,

$$\Rightarrow \mathcal{P}_t = \Big((WB_{n,x})\mu^{\mathbb{Q}} + (WA_{n,X}) - (WB_{n,x})\Phi^{\mathbb{Q}}(WB_{n,x})^{-1}(WA_{n,X})\Big) \tag{46}$$

$$+ (WB_{n,x})\Phi^{\mathbb{Q}}(WB_{n,x})^{-1}\mathcal{P}_{t-1} \tag{47}$$

$$+ (WB_{n,x})\Sigma\epsilon_t^{\mathbb{Q}} \tag{48}$$

and therefore,

$$\mathcal{P}_t = \mu_{\mathcal{P}}^{\mathbb{Q}} + \Phi_{\mathcal{P}}^{\mathbb{Q}}\mathcal{P}_{t-1} + \Sigma_{\mathcal{P}}\epsilon_t^{\mathbb{Q}} \tag{49}$$

with,

$$\mu_{\mathcal{P}}^{\mathbb{Q}} = (I - \Phi_{\mathcal{P}}^{\mathbb{Q}})(WA_{n,X}) + (WB_{n,x})\mu^{\mathbb{Q}} \tag{50}$$

$$\Phi_{\mathcal{P}}^{\mathbb{Q}} = (WB_{n,x})\Phi^{\mathbb{Q}}(WB_{n,x})^{-1} \tag{51}$$

$$\Sigma_{\mathcal{P}} = (WB_{n,x})\Sigma \tag{52}$$

In the special case of $\mu^{\mathbb{Q}} = [k^{\mathbb{Q}}_{\infty}, 0, 0]$ and $\Phi^{\mathbb{Q}} = diag(g)$, with g denoting a $(N \times 1)$ vector consisting of real and distinct eigenvalues, results are similar to Joslin et al. (2011) (see, proposition 2).

Given the state vector of equation (43), the continuously compounded equation of yields (see, equation (9)) is written as,

$$y_t = A_{n,X} + B_{n,x}\left((WB_{n,x})^{-1}\mathcal{P}_t - (WB_{n,x})^{-1}(WA_{n,X})\right) \tag{53}$$

$$= A_{n,X} + B_{n,x}(WB_{n,x})^{-1}\mathcal{P}_t - B_{n,x}(WB_{n,x})^{-1}(WA_{n,X}) \tag{54}$$

$$= \left(A_{n,X} - B_{n,x}(WB_{n,x})^{-1}(WA_{n,X})\right) + B_{n,x}(WB_{n,x})^{-1}\mathcal{P}_t \tag{55}$$

which gives,

$$y_t^n = A_{n,\mathcal{P}} + B_{n,\mathcal{P}}\mathcal{P}_t \tag{56}$$

with, loadings $A_{n,\mathcal{P}}$ and $B_{n,\mathcal{P}}$ given as,

$$A_{n,\mathcal{P}} = A_{n,X} - B_{n,x}(WB_{n,x})^{-1}(WA_{n,x}) \tag{57}$$

$$B_{n,\mathcal{P}} = B_{n,x}(WB_{n,x})^{-1} \tag{58}$$

Similarly, for the short rate equation r_t,

$$r_t = \delta_0 + \delta_1\left((WB_{n,x})^{-1}\mathcal{P}_t - (WB_{n,x})^{-1}(WA_{n,X})\right) \tag{59}$$

$$= \left(\delta_0 - \delta_1(WB_{n,x})^{-1}(WA_{n,X})\right) + \delta_1(WB_{n,x})^{-1}\mathcal{P}_t \tag{60}$$

which gives,

$$r_t = \delta_{0\mathcal{P}} + \delta'_{1\mathcal{P}}\mathcal{P}_t \tag{61}$$

with,

$$\delta_{0\mathcal{P}} = \delta_0 - \delta'_1(WB_{n,x})^{-1}(WA_{n,X}) \tag{62}$$

$$\delta'_{1\mathcal{P}} = (WB'_{n,x})^{-1}\delta_1 \tag{63}$$

References

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

Ahn, Dong-Hyun, Robert F Dittmar, and A Ronald Gallant, 2002, Quadratic term structure models: Theory and evidence, *The Review of financial studies* 15, 243–288.

Akaike, Hirotogu, 1998, Information theory and an extension of the maximum likelihood principle, in *Selected Papers of Hirotugu Akaike*, 199–213 (Springer).

Ang, Andrew, Sen Dong, and Monika Piazzesi, 2007, No-arbitrage taylor rules, Technical report, National Bureau of Economic Research.

Ang, Andrew, and Francis A Longstaff, 2011, Systemic sovereign credit risk: Lessons from the us and europe, *Journal of Monetary Economics* 60, 493–510.

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

Avramov, Doron, 2003, Stock return predictability and asset pricing models, *The Review of Financial Studies* 17, 699–738.

Bansal, Ravi, and Hao Zhou, 2002, Term structure of interest rates with regime shifts, *The Journal of Finance* 57, 1997–2043.

Barillas, Francisco, 2011, Can we exploit predictability in bond markets? .

Bauer, Michael D, 2017, Restrictions on risk prices in dynamic term structure models, *Journal of Business & Economic Statistics* 1–16.

Bauer, Michael D, and James D Hamilton, 2017, Robust bond risk premia, *The Review of Financial Studies* 31, 399–448.

Bauer, Michael D, and Glenn D Rudebusch, 2016, Resolving the spanning puzzle in macro-finance term structure models, *Review of Finance* 21, 511–553.

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, 495–514.

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

Cheridito, Patrick, Damir Filipović, and Robert L Kimmel, 2007, Market price of risk specifications for affine models: Theory and evidence, *Journal of Financial Economics* 83, 123–170.

Chib, Siddhartha, and Bakhodir Ergashev, 2009, Analysis of multifactor affine yield curve models., *Journal of the American Statistical Association* 104, 1324–1337.

Chib, Siddhartha, and Kyu Ho Kang, 2009, Change points in affine term-structure models: pricing, estimation and forecasting, *Journal of Business and Economic Statistics, forthcoming* .

Chib, Siddhartha, and Kyu Ho Kang, 2012, Change-points in affine arbitrage-free term structure models, *Journal of Financial Econometrics* 11, 302–334.

Chopin, Nicolas, 2002, A sequential particle filter method for static models, *Biometrika* 89, 539–552.

Chopin, Nicolas, 2004, Central limit theorem for sequential monte carlo methods and its application to bayesian inference, *Ann. Statist.* 32, 2385–2411.

Christoffersen, Peter, Kris Jacobs, and Karim Mimouni, 2010, Volatility dynamics for the s&p500: evidence from realized volatility, daily returns, and option prices, *The Review of Financial Studies* 23, 3141–3189.

Cochrane, John H, and Monika Piazzesi, 2005, Bond risk premia, *The American Economic Review* 95, 138–160.

Cochrane, John H, and Monika Piazzesi, 2009, Decomposing the yield curve, *AFA 2010* .

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

Del Moral, Pierre, Arnaud Doucet, and Ajay Jasra, 2006, Sequential monte carlo samplers, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 68, 411–436.

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, 158–174.

Dewachter, Hans, and Marco Lyrio, 2006, Macro factors and the term structure of interest rates, *Journal of Money, Credit, and Banking* 38, 119–140.

Duarte, Jefferson, 2004, Evaluating an alternative risk preference in affine term structure models, *Review of Financial Studies* 17, 379–404.

Duffee, Gregory R, 2002, Term premia and interest rate forecasts in affine models, *The Journal of Finance* 57, 405–443.

Duffee, Gregory R, 2010, Sharpe ratios in term structure models, Technical report, Working papers//the Johns Hopkins University, Department of Economics.

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

Duffee, Gregory R, and Richard H Stanton, 2012, Estimation of dynamic term structure models, *The Quarterly Journal of Finance* 2.

Duffie, Darrell, and Rui Kan, 1996, A yield-factor model of interest rates, *Mathematical finance* 6, 379–406.

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

Feldhütter, Peter, 2016, Can affine models match the moments in bond yields?, *Quarterly Journal of Finance* 6, 1650009.

Fernández-Villaverde, Jesús, and Juan F Rubio-Ramírez, 2007, Estimating macroeconomic models: A likelihood approach, *The Review of Economic Studies* 74, 1059–1087.

Feunou, Bruno, and Jean-Sébastien Fontaine, 2017, Bond risk premia and gaussian term structure models, *Management Science* 64, 1413–1439.

Fulop, Andras, Junye Li, and Runqing Wan, 2018, Real-time learning and bond return predictability .

Gargano, Antonio, Davide Pettenuzzo, and Allan Timmermann, 2017, Bond return predictability: Economic value and links to the macroeconomy, *Management Science* .

Ghysels, Eric, Casidhe Horan, and Emanuel Moench, 2017, Forecasting through the rearview mirror: Data revisions and bond return predictability, *The Review of Financial Studies* 31, 678–714.

Hamilton, James D, and Jing Cynthia Wu, 2012, Identification and estimation of gaussian affine term structure models, *Journal of Econometrics* 168, 315–331.

Hannan, Edward J, and Barry G Quinn, 1979, The determination of the order of an autoregression, *Journal of the Royal Statistical Society. Series B (Methodological)* 190–195.

Johannes, Michael, Arthur Korteweg, and Nicholas Polson, 2014, Sequential learning, predictability, and optimal portfolio returns, *The Journal of Finance* 69, 611–644.

Johannes, Michael S, Nicholas G Polson, and Jonathan R Stroud, 2009, Optimal filtering of jump diffusions: Extracting latent states from asset prices, *The Review of Financial Studies* 22, 2759–2799.

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, 1197–1233.

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

Kim, Don H, and Kenneth J Singleton, 2012, Term structure models and the zero bound: an empirical investigation of japanese yields, *Journal of Econometrics* 170, 32–49.

Kim, Don H, and Jonathan H Wright, 2005, An arbitrage-free three-factor term structure model and the recent behavior of long-term yields and distant-horizon forward rates, *Finance and Economics Discussion Series 2005-33, Federal Reserve Board of Governors.* .

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

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

Piazzesi, Monika, Martin Schneider, Pierpaolo Benigno, and John Y Campbell, 2006, Equilibrium yield curves [with comments and discussion], *NBER macroeconomics Annual* 21, 389–472.

Rudebusch, Glenn D, and Tao Wu, 2008, A macro-finance model of the term structure, monetary policy and the economy, *The Economic Journal* 118, 906–926.

Sarno, Lucio, Paul Schneider, and Christian Wagner, 2016, The economic value of predicting bond risk premia, *Journal of Empirical Finance* 37, 247–267.

Schwarz, Gideon, et al., 1978, Estimating the dimension of a model, *The annals of statistics* 6, 461–464.

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

Wright, Jonathan H, 2011, Term premia and inflation uncertainty: Empirical evidence from an international panel dataset, *American Economic Review* 101, 1514–34.

Table 8.1: Persistence of interest rates across models.

Model \	$\mathbb{Q}$	$\mathbb{P}$
M0	0.9988 [0.9981,0.9994]	0.9846 [0.9630,0.9988]
M1	0.9987 [0.9980,0.9994]	0.9781 [0.9719,0.9899]
M2	0.9987 [0.9977,0.9995]	0.9893 [0.9881,0.9944]
M3	0.9987 [0.9981,0.9995]	0.9929 [0.9873,0.9987]
M4	0.9987 [0.9980,0.9994]	0.9923 [0.9871,0.9981]
M5	0.9987 [0.9979,0.9994]	0.9919 [0.9868,0.9983]
M6	0.9988 [0.9982,0.9995]	0.979 [0.9678,0.9961]

This table reports the persistence of risk factors under the risk neutral ($\mathbb{Q}$) and physical ($\mathbb{P}$) dynamics. Persistence is measured by the largest eigenvalue of the feedback matrices $\Phi^{\mathbb{Q}}_{\mathcal{P}}$ and $\Phi^{\mathbb{P}}_{\mathcal{P}}$ respectively. The seven models used are ATSM with alternative risk price restrictions. In particular, the maximally flexible model, M_0 (no restrictions) and restricted models with some unrestricted parameters, as in, M_1 ($\lambda_{0,1}$, $\lambda_{1,1}$ and $\lambda_{1,2}$ unrestricted), M_2 ($\lambda_{0,1}$ and $\lambda_{1,2}$ unrestricted), M_3 ($\lambda_{1,1}$ and $\lambda_{1,2}$ unrestricted), M_4 ($\lambda_{1,2}$ unrestricted), M_5 ($\lambda_{1,1}$ unrestricted) and M_6 ($\lambda_{1,1}$, $\lambda_{1,2}$ and $\lambda_{2,3}$ unrestricted). Confidence intervals are reported in brackets.

Table 8.2: Out-of-sample statistical performance of Bond excess return forecasts: 1-month prediction horizon.

**Model **	**2Y**	**3Y**	**4Y**	**5Y**	**7Y**	**10Y**
M0	0.28***	0.20**	0.16**	0.16**	0.09	0.03
	(-2.99)	(2.18)	(1.74)	(1.71)	(1.05)	(0.23)
M1	0.26***	0.18**	0.14**	0.14**	0.09	0.05
	(3.13)	(2.38)	(1.90)	(1.91)	(1.30)	(0.60)
M2	0.16***	0.10**	0.07	0.09*	0.05	0.04
	(2.54)	(1.80)	(1.26)	(1.51)	(0.78)	(0.55)
M3	0.10***	0.07***	0.04*	0.05**	0.02	0.04
	(3.49)	(2.62)	(1.69)	(2.09)	(0.80)	(1.17)
M4	0.11***	0.08**	0.06*	0.07**	0.04	0.06*
	(2.75)	(2.21)	(1.60)	(2.00)	(1.08)	(1.41)
M5	0.10**	0.06	0.03	0.04	0.00	0.01
	(1.87)	(1.24)	(0.83)	(0.94)	(0.14)	(0.32)
M6	0.26***	0.19**	0.15**	0.16**	0.11	0.05
	(3.07)	(2.35)	(1.88)	(1.87)	(1.30)	(0.49)

This table reports out-of-sample R^2 across alternative models, at 1-month prediction horizon. The seven forecasting models used are ATSM with alternative risk price restrictions. R^2 values are generated using the out-of-sample R^2 measure of Campbell and Thompson (2008). In particular, out-of-sample R^2 measures the predictive accuracy of bond excess return forecasts relative to the EH benchmark. The EH implies the historical mean being the optimal forecast of excess returns. Positive values of this statistic imply that the forecast outperforms the historical mean forecast and suggests evidence of time-varying return predictability. Statistical significance is measured using a one-sided Diebold-Mariano statistic. * denotes significance at 10%, ** significance at 5% and *** significance at 1% level. The in-sample period is January 1990 to December 2003, and the out-of-sample period is January 2004 to December 2007.

Table 8.3: Out-of-sample statistical performance of Bond excess return forecasts: multiple prediction horizons.

(h\n − h)	2Y	3Y	4Y	5Y	7Y	10Y
			M0			
1m	0.28***	0.20**	0.16**	0.16**	0.09	0.03
3m	0.49***	0.42***	0.37***	0.35***	0.26**	0.10
6m	0.63***	0.62***	0.63***	0.62***	0.54***	0.32*
9m	0.59***	0.60***	0.64***	0.66***	0.63***	0.46**
12m	0.50***	0.46***	0.48***	0.51***	0.53***	0.42**
			M1			
1m	0.26***	0.18**	0.14**	0.14**	0.09	0.05
3m	0.44***	0.35***	0.29***	0.27***	0.21**	0.13
6m	0.60***	0.56***	0.55***	0.54***	0.50***	0.39**
9m	0.60***	0.57***	0.59***	0.59***	0.60***	0.56
12m	0.53***	0.45***	0.44***	0.44***	0.47**	0.49**
			M2			
1m	0.16***	0.10**	0.07	0.09*	0.05	0.04
3m	0.23***	0.17**	0.13*	0.12	0.10	0.09
6m	0.34***	0.28***	0.26**	0.26**	0.27**	0.26*
9m	0.32***	0.26***	0.25**	0.23**	0.28**	0.31**
12m	0.26***	0.13	0.11	0.07	0.12	0.17
			M3			
1m	0.10***	0.07***	0.04*	0.05**	0.02	0.04
3m	0.12***	0.08**	0.05	0.04	0.04	0.07
6m	0.18***	0.13***	0.10**	0.07	0.07	0.08
9m	0.20***	0.13**	0.11**	0.06	0.08	0.07
12m	0.16***	0.06	0.04	-0.03	0.00	-0.04
			M4			
1m	0.11***	0.08**	0.06*	0.07**	0.04	0.06*
3m	0.14***	0.10**	0.07	0.06	0.06	0.10*
6m	0.18***	0.13**	0.11*	0.08	0.10	0.13*
9m	0.19***	0.12**	0.10*	0.06	0.10	0.12
12m	0.15***	0.04	0.02	-0.05	-0.01	-0.02
			M5			
1m	0.10**	0.06	0.03	0.04	0.00	0.01
3m	0.20***	0.16***	0.11**	0.0**	0.05*	0.05*
6m	0.29***	0.25***	0.20***	0.13***	0.07***	0.01
9m	0.34***	0.30***	0.27***	0.18***	0.11***	-0.03
12m	0.35***	0.31***	0.29***	0.22***	0.16***	-0.01
			M6			
1m	0.26***	0.19**	0.15**	0.16**	0.11	0.05
3m	0.43***	0.36***	0.31***	0.29***	0.2**	0.10
6m	0.61***	0.60***	0.60***	0.60***	0.54***	0.36*
9m	0.59***	0.60***	0.63***	0.65***	0.64***	0.51**
12m	0.52***	0.47***	0.48***	0.50***	0.51**	0.44**

This table reports out-of-sample R^2 across alternative models, at different prediction horizons, of h= 1-month, 3-month, 6-month, 9-month and 12-month. The seven forecasting models used are ATSM with alternative risk price restrictions. R^2 values are generated using the out-of-sample R^2 measure of Campbell and Thompson (2008). In particular, out-of-sample R^2 measures the predictive accuracy of bond excess return forecasts relative to the EH benchmark. The EH implies the historical mean being the optimal forecast of excess returns. Positive values of this statistic imply that the forecast outperforms the historical mean forecast and suggests evidence of time-varying return predictability. Statistical significance is measured using a one-sided Diebold-Mariano statistic. * denotes significance at 10%, ** significance at 5% and *** significance at 1% level. The in-sample period is January 1990 to December 2003, and the out-of-sample period is January 2004 to December 2007.

Table 8.4: Out-of-sample Economic performance of Bond excess return forecasts across investment scenarios: 1-month prediction horizon.

Panel A: w= [-1,2]							Panel B: w= [-inf,+inf]						
Model	**2Y**	**3Y**	**4Y**	**5Y**	**7Y**	**10Y**	**Model**	**2Y**	**3Y**	**4Y**	**5Y**	**7Y**	**10Y**
							CER: ($\gamma = 3$)						
M0	3**	4**	7**	9**	9	2	**M0**	42**	33**	28*	27*	13	5
M1	2*	3	3	5	6	3	**M1**	34***	24**	20**	20**	10	8
M2	1*	2*	3*	4*	5	3	**M2**	15*	7	6	8	1	5
M3	0	0	0	0	0	1	**M3**	11***	7*	7**	7*	0	5
M4	0	0	0	0	1	2	**M4**	7	3	3	4	-2	4
M5	0	0	0	0	0	0	**M5**	15	12*	10	8	0	4
M6	2**	3	4	6*	6	2	**M6**	34***	26**	22**	23**	12	7
							CER: ($\gamma = 5$)						
M0	3**	5**	7**	9**	8	3	**M0**	25**	19**	16*	16*	8	3
M1	2*	3	4	6*	5	4	**M1**	20***	14**	12**	12**	6	5
M2	1*	2*	3*	4*	4	5	**M2**	9*	4	4	5	1	3
M3	0	0	0	0	1	2	**M3**	6***	4*	4**	4**	0	3
M4	0	0	0	1*	2*	4	**M4**	5	2	2	3	-1	3
M5	0	0	0	0	0	2	**M5**	9	7*	6	5	0	2
M6	2**	3*	4	6*	6	4	**M6**	20***	15**	13**	14**	7	4
							CER: ($\gamma = 10$)						
M0	3**	6**	7*	8*	4	1	**M0**	12**	10**	8*	8*	4	1
M1	2**	3*	4*	6**	3	2	**M1**	10***	7**	6**	6**	3	2
M2	2**	2*	3*	4**	0	2	**M2**	5*	2	2	3	0	2
M3	0	0	1	2*	0	2	**M3**	3***	2*	2**	2**	0	2
M4	0	1*	2**	3**	0	1	**M4**	2	1	1	1	0	1
M5	0	0	2	3	0	1	**M5**	4	3*	3	3	0	1
M6	3**	4*	5*	7*	4	2	**M6**	10***	8**	7**	7**	4	2

This table reports annualized certainty equivalent returns (CERs) across alternative models, at 1-month prediction horizon. CERs are generated by out-of-sample forecasts of bond excess returns and are reported in basis points. At every time step, t, an investor with power utility preferences, evaluates the entire predictive density of bond excess returns and solves the asset allocation problem, thus optimally allocating her wealth between a riskless bond and risky bonds with maturities 2, 3, 4, 5, 7 and 10-years. CER is, then, defined as the value that equates the average utility of each alternative model against the average utility of the EH benchmark. The seven forecasting models used are ATSM with alternative risk price restrictions. The coefficient of risk aversion is either $\gamma = 3$, 5 or 10. Positive values indicate that the models perform better than the EH benchmark. Panel A presents CERs under the first scenario, where, portfolio weights are restricted to range in the interval [-1, 2], such that investors are prevented from extreme investments. Panel B, reports CER values under the second scenario, where extreme investments are allowed and, as such, portfolio weights are free of constraints. Statistical significance is measured using a one-sided Diebold-Mariano statistic. * denotes significance at 10%, ** significance at 5% and *** significance at 1% level. The in-sample period is January 1990 to December 2003, and the out-of-sample period is January 2004 to December 2007.

Table 8.5: Out-of-sample Economic performance of Bond excess return forecasts across investment scenarios: multiple prediction horizons ($\gamma = 3$).

CER ($\gamma = 3$)													
	Panel A: w= [-1,2]							**Panel B: w= [-inf,+inf]**					
h	**2Y**	**3Y**	**4Y**	**5Y**	**7Y**	**10Y**	**h**	**2Y**	**3Y**	**4Y**	**5Y**	**7Y**	**10Y**
M0													
1m	3**	4**	7**	9**	9	2	**1m**	42**	33**	28*	27*	13	5
3m	5***	8**	11**	15**	17*	3	**3m**	69***	52***	41***	42***	27*	6
6m	9***	12***	16***	21***	24**	12	**6m**	90***	67***	62***	57***	41**	14
9m	9***	11**	14**	21**	27**	16	**9m**	74***	65***	58***	58***	44**	18
12m	5*	2	2	6	13	8	**12m**	-1	0	12	21	23	8
M1													
1m	2*	3	3	5	6	3	**1m**	34***	24**	20**	20**	10	8
3m	6***	7**	8**	11**	13*	7	**3m**	54***	36***	27**	29**	22*	13
6m	10***	13***	16***	21***	26**	23*	**6m**	85***	58***	54***	52***	46***	33**
9m	9***	12**	15**	20**	28**	30*	**9m**	81***	64***	55***	57***	52**	43**
12m	6**	4	3	7	11	16	**12m**	27	14	16	20	27	27
M2													
1m	1*	2*	3*	4*	5	3	**1m**	15*	7	6	8	1	5
3m	2*	2	3	6*	8	7	**3m**	9	2	1	5	6	10
6m	3**	4*	5*	9**	13**	17*	**6m**	18	13	22*	26**	30**	30**
9m	1	0	1	4	8	16*	**9m**	5	15	20*	28**	35**	37***
12m	-2	-7	-10	-9	-8	1	**12m**	-66	-46	-22	-9	10	21*
M3													
1m	0	0	0	0	0	1	**1m**	11***	7*	7**	7*	0	5
3m	0	0	0	0	0	3*	**3m**	-5	-4	-5	-2	-1	7
6m	0	0	0	0	0	3**	**6m**	-15	-8	2	5	10**	15***
9m	0	0	0	0	0	1	**9m**	-46	-12	-1	6	12***	18***
12m	0	0	0	0	0	0	**12m**	-101	-48	-20	-11	5	15***
M4													
1m	0	0	0	0	1	2	**1m**	7	3	3	4	-2	4
3m	0	0	0	0	1	4*	**3m**	-9	-10	-9	-6	-3	7
6m	0	0	0	0	1*	5**	**6m**	-12	-5	5	9*	14***	19***
9m	0	0	0	0	0	5**	**9m**	-34	-4	6	13**	19***	25***
12m	0	0	0	0	0	1	**12m**	-98	-44	-16	-5	12*	20***
M5													
1m	0	0	0	0	0	0	**1m**	15	12*	10	8	0	4
3m	0	0	0	0	0	2	**3m**	14*	12*	8	7*	3	4
6m	0	0	0	0	0	-1	**6m**	-4	0	4*	1	-2	-3
9m	0	0	0	0	0	-1	**9m**	-16	6*	5*	3	-2	-4
12m	0	0	0	0	0	0	**12m**	-51	-11	2	0	1	-1
M6													
1m	2**	3	4	6*	6	2	**1m**	34***	26**	22**	23**	12	7
3m	6***	7**	8*	11**	12	1	**3m**	49***	35***	28**	30**	22	6
6m	10***	13***	16***	21***	25**	17	**6m**	85***	62***	59***	56***	46**	24
9m	9***	12**	16**	21**	28**	25	**9m**	76***	67***	59***	60***	51**	32*
12m	5*	4	5	9	14	15	**12m**	29	22	24	26	26	17

This table reports annualized certainty equivalent returns (CERs) across alternative models, at different prediction horizons, of h= 1-month, 3-month, 6-month, 9-month and 12-month. The coefficient of risk aversion is $\gamma = 3$. CERs are generated by out-of-sample forecasts of bond excess returns and are reported in basis points. At every time step, t, an investor with power utility preferences, evaluates the entire predictive density of bond excess returns and solves the asset allocation problem, thus optimally allocating her wealth between a riskless bond and risky bonds with maturities 2, 3, 4, 5, 7 and 10-years. CER is, then, defined as the value that equates the average utility of each alternative model against the average utility of the EH benchmark. The seven forecasting models used are ATSM with alternative risk price restrictions. Positive values indicate that the models perform better than the EH benchmark. Panel A presents CERs under the first scenario, where, portfolio weights are restricted to range in the interval [-1, 2], such that investors are prevented from extreme investments. Panel B, reports CER values under the second scenario, where extreme investments are allowed and, as such, portfolio weights are free of constraints. Statistical significance is measured using a one-sided Diebold-Mariano statistic. * denotes significance at 10%, ** significance at 5% and *** significance at 1% level. The in-sample period is January 1990 to December 2003, and the out-of-sample period is January 2004 to December 2007.

Table 8.6: Out-of-sample Economic performance of Bond excess return forecasts across investment scenarios: multiple prediction horizons ($\gamma = 5$).

CER ($\gamma = 5$)													
Panel A: w= [-1,2]							**Panel B: w= [-inf,+inf]**						
h	**2Y**	**3Y**	**4Y**	**5Y**	**7Y**	**10Y**	**h**	**2Y**	**3Y**	**4Y**	**5Y**	**7Y**	**10Y**
M0													
1m	3**	5**	7**	9**	8	3	**1m**	25**	19**	16*	16*	8	3
3m	6***	9**	12**	17**	16*	4	**3m**	41***	31***	25***	25***	16*	4
6m	10***	13***	17***	22***	23**	9	**6m**	54***	40***	37***	34***	25**	9
9m	9***	11**	16**	23**	27**	11	**9m**	44***	39***	35***	35***	27**	11
12m	5*	3	3	10	17	5	**12m**	-1	1	8	13	14	5
M1													
1m	2*	3	4	6	5	4	**1m**	20***	14**	12**	12**	6	5
3m	6***	7**	9**	12**	13*	8	**3m**	32***	22***	16**	17**	13*	8
6m	10***	13***	17***	21***	24**	19*	**6m**	51***	35***	32***	31***	28***	20**
9m	9***	12**	16**	22**	29**	25*	**9m**	49***	38***	34***	34***	32**	26**
12m	6**	5	5	9	15	16	**12m**	16	9	10	13	16	17
M2													
1m	1*	2*	3*	4*	4	5	**1m**	9*	4	4	5	1	3
3m	2*	3	4	7*	8	7	**3m**	5	1	0	3	3	6
6m	4**	5*	8*	12**	15**	16*	**6m**	10	8	13*	16**	18**	18**
9m	1	2	4	9	15	19*	**9m**	0	7	12	17**	22**	24***
12m	-2	-5	-6	-3	4	10	**12m**	-46	-30	-14	-6	7	14**
M3													
1m	0	0	0	0	1	2	**1m**	6***	4*	4**	4**	0	3
3m	0	0	0	1	3	5*	**3m**	-4	-4	-4	-2	-1	4
6m	0	0	0	1	4	8**	**6m**	-10	-5	1	3	6**	9***
9m	0	0	0	0	4	9	**9m**	-28	-8	0	4	8***	12***
12m	0	0	0	0	2	7	**12m**	-67	-32	-14	-7	4	10***
M4													
1m	0	0	0	1	2	4	**1m**	5	2	2	3	-1	3
3m	0	0	0	2	4	6*	**3m**	-6	-6	-6	-4	-2	4
6m	0	0	0	2	6*	10**	**6m**	-9	-4	3	5	9***	12***
9m	0	0	0	1	7	12**	**9m**	-22	-3	3	8**	13***	16***
12m	0	0	0	0	4	9	**12m**	-63	-30	-12	-5	7*	13***
M5													
1m	0	0	0	0	0	2	**1m**	9	7*	6	5	0	2
3m	0	0	0	1	2	2	**3m**	9*	7*	5	4*	2	2
6m	0	0	0	0	-2	-2	**6m**	-2	1	3*	0	-1	-2
9m	0	0	0	0	-1	-2	**9m**	-8	4*	3*	2	-1	-2
12m	0	0	0	0	1	0	**12m**	-30	-6	2	0	1	0
M6													
1m	2**	3	4	6*	6	4	**1m**	20***	15**	13**	14**	7	4
3m	6***	7**	9*	12**	12	4	**3m**	29***	21***	17**	18**	13*	4
6m	10***	13***	17***	23***	25**	14	**6m**	50***	37***	35***	33***	28**	15
9m	9***	13**	17**	24**	30**	19	**9m**	45***	40***	36***	36***	31**	20*
12m	5*	5	7	12	18	12	**12m**	15	13	14	16	16	11

This table reports annualized certainty equivalent returns (CERs) across alternative models, at different prediction horizons, of h= 1-month, 3-month, 6-month, 9-month and 12-month. The coefficient of risk aversion is $\gamma = 5$. CERs are generated by out-of-sample forecasts of bond excess returns and are reported in basis points. At every time step, t, an investor with power utility preferences, evaluates the entire predictive density of bond excess returns and solves the asset allocation problem, thus optimally allocating her wealth between a riskless bond and risky bonds with maturities 2, 3, 4, 5, 7 and 10-years. CER is, then, defined as the value that equates the average utility of each alternative model against the average utility of the EH benchmark. The seven forecasting models used are ATSM with alternative risk price restrictions. Positive values indicate that the models perform better than the EH benchmark. Panel A presents CERs under the first scenario, where, portfolio weights are restricted to range in the interval [-1, 2], such that investors are prevented from extreme investments. Panel B, reports CER values under the second scenario, where extreme investments are allowed and, as such, portfolio weights are free of constraints. Statistical significance is measured using a one-sided Diebold-Mariano statistic. * denotes significance at 10%, ** significance at 5% and *** significance at 1% level. The in-sample period is January 1990 to December 2003, and the out-of-sample period is January 2004 to December 2007.

Table 8.7: Out-of-sample Economic performance of Bond excess return forecasts across investment scenarios: multiple prediction horizons ($\gamma = 10$).

CER ($\gamma = 10$)												
Panel A: w= [-1,2]							**Panel B: w= [-inf,+inf]**					
h	**2Y**	**3Y**	**4Y**	**5Y**	**7Y**	**10Y**	**h**	**2Y**	**3Y**	**4Y**	**5Y**	**7Y**	**10Y**
M0												
1m	3**	6**	7*	8*	4	1	**1m**	12**	10**	8*	8*	4	1
3m	6***	10***	12***	13***	8*	2	**3m**	20***	16***	12***	12***	8*	2
6m	10***	14***	17***	17***	12**	4	**6m**	27***	20***	18***	17***	12**	4
9m	9***	13***	17***	17***	13**	5	**9m**	22***	19***	18***	18***	13**	5
12m	5**	5	7	8	7	2	**12m**	-1	1	4	7	7	2
M1												
1m	2**	3*	4*	6**	3	2	**1m**	10***	7**	6**	6**	3	2
3m	6***	8**	8**	8**	7*	4	**3m**	16***	11***	8**	9**	7*	4
6m	10***	14***	16***	15***	14***	10**	**6m**	25***	17***	16***	15***	14***	10**
9m	10***	13***	17***	17***	16**	13**	**9m**	24***	19***	17***	17***	16**	13**
12m	6**	7	8	6	8	8	**12m**	8	5	5	6	8	8
M2												
1m	2**	2*	3*	4**	0	2	**1m**	5*	2	2	3	0	2
3m	3**	4**	4*	3	2	3	**3m**	2	0	0	1	2	3
6m	5***	7***	8***	8**	9***	9**	**6m**	5	4	6*	8**	9***	9**
9m	3**	5**	8**	8**	11**	12***	**9m**	-1	3	6	8**	11**	12***
12m	-1	-1	0	-1	4	7**	**12m**	-23	-15	-7	-3	4	7**
M3												
1m	0	0	1	2*	0	2	**1m**	3***	2*	2**	2**	0	2
3m	0	1**	1	0	-1	2	**3m**	-2	-2	-2	-1	-1	2
6m	0	1**	2**	2	3**	5***	**6m**	-5	-3	1	2	3**	5***
9m	0	0	2	2**	4***	6***	**9m**	-14	-4	0	2	4***	6***
12m	0	0	-1	-2	2	5***	**12m**	-33	-16	-7	-4	2	5***
M4												
1m	0	1*	2**	3**	0	1	**1m**	2	1	1	1	0	1
3m	0	1***	1*	1	-1	2	**3m**	-3	-3	-3	-2	-1	2
6m	0	1***	3***	3*	4***	6***	**6m**	-5	-2	1	2	4***	6***
9m	0	1**	3**	4**	6***	8***	**9m**	-11	-2	2	4**	6***	8***
12m	0	0	1	-1	4*	7***	**12m**	-31	-15	-6	-3	4*	7***
M5												
1m	0	0	2	3	0	1	**1m**	4	3*	3	3	0	1
3m	0	1**	2	2*	1	1	**3m**	4*	4*	2	2*	1	1
6m	0	0	1**	0	-1	-1	**6m**	-1	0	1*	0	-1	-1
9m	0	0	2*	1	-1	-1	**9m**	-4	2*	2*	1	-1	-1
12m	0	0	1	0	0	0	**12m**	-15	-3	1	0	0	0
M6												
1m	3**	4*	5*	7*	4	2	**1m**	10***	8**	7**	7**	4	2
3m	6***	8**	8**	9**	7*	2	**3m**	14***	10***	8**	9**	7*	2
6m	10***	15***	17***	17***	14**	7	**6m**	25***	18***	17***	17***	14**	7
9m	10***	14***	18***	18***	15**	10*	**9m**	22***	20***	18***	18***	15**	10*
12m	6**	7	9	8	8	5	**12m**	7	6	7	8	8	5

This table reports annualized certainty equivalent returns (CERs) across alternative models, at different prediction horizons, of h= 1-month, 3-month, 6-month, 9-month and 12-month. The coefficient of risk aversion is $\gamma = 10$. CERs are generated by out-of-sample forecasts of bond excess returns and are reported in basis points. At every time step, t, an investor with power utility preferences, evaluates the entire predictive density of bond excess returns and solves the asset allocation problem, thus optimally allocating her wealth between a riskless bond and risky bonds with maturities 2, 3, 4, 5, 7 and 10-years. CER is, then, defined as the value that equates the average utility of each alternative model against the average utility of the EH benchmark. The seven forecasting models used are ATSM with alternative risk price restrictions. Positive values indicate that the models perform better than the EH benchmark. Panel A presents CERs under the first scenario, where, portfolio weights are restricted to range in the interval [-1, 2], such that investors are prevented from extreme investments. Panel B, reports CER values under the second scenario, where extreme investments are allowed and, as such, portfolio weights are free of constraints. Statistical significance is measured using a one-sided Diebold-Mariano statistic. * denotes significance at 10%, ** significance at 5% and *** significance at 1% level. The in-sample period is January 1990 to December 2003, and the out-of-sample period is January 2004 to December 2007.