DEPARTMENT OF ECONOMICS AND BUSINESS ECONOMICS
AARHUS UNIVERSITY

CREATES
Center for Research in Econometric Analysis of Time Series

Long and short memory in dynamic term structure models

Salman Huseynov

CREATES Research Paper 2021-15

Department of Economics and Business Economics
Aarhus University
Fuglesangs Allé 4
DK-8210 Aarhus V
Denmark

Email: oekonomi@au.dk
Tel: +45 9352 1936

Long and short memory in dynamic term structure models

Salman Huseynov*

This version: 3 December 2021

First draft: 16 July 2020

Abstract

I provide a unified theoretical framework for long memory term structure models and show that the recent state-space approach suffers from a parameter identification problem. I propose a different framework to estimate long memory models in a state-space setup, which addresses the shortcomings of the existing approach. The proposed framework allows asymmetrically treating the physical and risk-neutral dynamics, which simplifies estimation considerably and helps to conduct an extensive comparison with standard term structure models. Relying on a battery of tests, I find that standard term structure models perform just as well as the more complicated long memory models and produce plausible term premium estimates.

JEL Classification: C32, E43, G12

Keywords: Dynamic term structure models, Long memory, Affine model, Shadow rate model

*shuseynov@econ.au.dk, Aarhus University and CREATES. I would like to especially thank Martin Møller Andreasen and Juan Carlos Parra-Alvarez for careful reading, detailed comments and valuable suggestions, which have considerably improved the quality of the paper. I am also grateful to an anonymous referee, Leopoldo Catania, Bent Jesper Christensen, Peter Feldhütter, Burçin Kisacikoglu, Morten Ørregaard Nielsen, and seminar participants at Aarhus University, the DGPE workshop at Kobæk Strand Center, the Nordic Finance workshop and Bilkent University for their comments and suggestions. I would also like to thank Martin Møller Andreasen, Adam Golinski and Marcel Priebsch for generously sharing their Matlab codes. All remaining errors are mine. The numerical results presented in this work were obtained at the Centre for Scientific Computing, Aarhus University `http://www.cscaa.dk/`

1 Introduction

Understanding term premium in the bond markets is important for financial market participants and policy institutions. Because term premium cannot be directly observed, it must be estimated from data on the yield curve. Short memory models[1] generally draw on a simple first-order vector autoregressive VAR(1) process to estimate term premium. Long memory term structure models, in turn, are based on a more general process and usually nest short memory specification. They can accommodate the strong persistence observed in nominal yields and even non-stationarity but still ensure dynamic mean reversion. Thus, several studies argue that accounting for long memory can improve the performance of term structure models (see among others, Abritti, Gil-Alana, Lovcha, and Moreno, 2016; Golinski and Zaffaroni, 2016; Osterrieder and Schotman, 2017).

Although both long and short memory term structure models can be employed to generate term premium estimates, it is not clear which model performs best. Admittedly, there are important studies in the long memory literature that investigate both models along several dimensions. However, to the best of my knowledge, no single study compares term premium estimates obtained from these models against a battery of tests and formally shows a superior one. For instance, Abritti et al (2016) claim that long memory models can generate more realistic term premium estimates as these models enjoy a much slower mean-reversion of their long-horizon forecasts than do short memory models. But they only visually demonstrate both term premium estimates without running any formal tests, such as predictability tests in the spirit of Dai and Singleton (2002). Similar remarks can be made about other studies. Thus, this study aims to fill this gap in the literature.

To carry out an extensive comparison between these models, I address two challenges in this study. First, I improve the theoretical framework to estimate long memory term structure models and generalize the state-space approach proposed in Golinski and Zaffaroni (2016). Second, I propose a factor specification that considerably simplifies estimation and allows me to conduct a wide range of comparisons with standard term structure models. Using this framework, I investigate whether the standard and parsimonious short memory model a la Joslin, Singleton, and Zhu (2011) can improve on long memory term structure models.

[1] Technically speaking, a process with an order of integration $d \leq 0$ is called a short memory process. However, in this paper, a short memory process is meant to be an integrated order of zero, $I(0)$, process. A long memory process is an integrated order of d process, $I(d)$, where $d > 0$ and d can take fractional values.

There are several contributions of the paper. First, I generalize the theoretical framework proposed by Osterrieder and Schotman (2017) that allows defining different orders of integration for pricing factors under different probability measures in long memory term structure models and develop it further. I also establish a theoretical relation between orders of integration of factors under different probability measures and market price of risk. Although it is well-known in the literature how drifts or autoregressive terms are adjusted when we change probability measures, a similar adjustment, in general, is not straightforward in the case of orders of integration. This theoretical result can be applied, for instance, to find the order of integration of the short-term policy rate r_t under the Q-measure in Osterrieder and Schotman (2017). The authors estimate different orders of integration for the market price of risk λ_t and r_t under the P-measure. Based on the results, I find that r_t under the Q-measure in Osterrieder and Schotman (2017) accordingly follows a long memory process with the order of integration $d^Q = d^P - d^\lambda$, where d^P and d^λ are the orders of integration of r_t under the P-measure and λ_t, respectively.

Second, I show that the state-space representation proposed in Golinski and Zaffaroni (2016) to estimate long memory term structure models suffers from an identification problem. Their infinite order vector moving average VMA(∞) representation only allows to identify one out of the three level parameters, i.e., the intercept of pricing factors under the physical probability measure, the constant market price of risk, or the level of the short rate. Although the authors estimate two of these parameters together with the latent state variable by implicitly setting the intercept of the pricing factors under the physical probability measure to zero, I show that these parameters and the latent variable are unidentified in their framework. While a refinement of their procedure is available, it does not allow to identify two out of the three level parameters together with the latent state vector as in the case of a short memory model.

Another shortcoming of the VMA(∞) representation is that the mean in the VMA(∞) process is not well-defined when pricing factors are non-stationary, i.e., when the order of integration $d > 0.5$ (Johansen and Nielsen, 2016). In addition, the VMA(∞) framework is not suitable for asymmetrically treating the physical and risk-neutral dynamics of pricing factors and defining different orders of integration under the two measures.

To address these problems, I propose to estimate long memory term structure models using an infinite order vector autoregressive, VAR(∞), representation. One advantage of the new approach is that its coefficients decay much faster than the corresponding coefficients of the VMA(∞) process (especially when $d > 0.5$), which makes the selection of a large truncation lag unnecessary. [2] As shown by Grassi and de Magistris (2014) a long truncation lag may lead to biased estimators in finite samples.

After addressing the above methodological issues in long memory models, I estimate several dynamic term structure models (DTSMs) with short and long memory, and perform various tests in a rigorous setting. Using a battery of tests[3], I find that short memory term structure models perform just as well as long memory models. I obtain plausible term premium estimates, which are very similar to estimates in the literature. Only the magnitudes of the term premium estimates of the long and short memory models somewhat differ, although the term premium dynamics is similar. In general, the long memory models have relatively larger term premium estimates. Overall, I show that standard term structure models with short memory produce similar term premium estimates as more complicated long memory models.

While this framework shares some features with Abritti et al (2015) and Golinski and Zaffaroni (2016), it differs from these studies along the following lines. Abritti et al (2015) estimate a vector autoregressive fractionally integrated moving average VARFIMA(p,d,q) model using only observed state variables, whereas the framework proposed in this paper allows for including both observed and unobserved factors. Another novelty of my framework is its asymmetric treatment of factors under different probability measures. Following Joslin, Le and Singleton (2013), I propose modeling the factors under the risk-neutral probability measure Q as a short-memory process (i.e., VAR(1)) while specifying a long-memory process under the physical probability measure P. As Joslin et al (2013) argue, setting a lag order of one for the pricing factors under the risk-neutral measure is sufficiently comprehensive.

Although the proposed framework is fairly general and allows me to estimate long memory under both probability measures as in Golinski and Zaffaroni (2016), I employ the asymmetric specification due to its clear advantages. First, the asymmetric approach enables me to apply standard pricing formulas and substantially reduce computational burden, which is another limitation of the Golinski and Zaffaroni (2016) approach. Second, it also facilitates the application of the efficient identification restrictions proposed in Joslin et al (2011). These extra

[2] I would like to thank an anonymous referee and Morten Ørregaard Nielsen for the last two comments.

[3] The results of a forecasting exercise can be found in the appendix

benefits generally impose no additional costs.

2 Factor specifications

Since the employed short memory specification is fairly standard, I only discuss the long memory specification (see the appendix for the short memory case). In particular, I assume that a vector of the pricing factors $x_t = (x_{1t}\, x_{2t} \ldots x_{Kt})'$ follows a vector autoregressive fractionally integrated moving average VARFIMA(p,d,q) process under the Q-measure

$$\Phi^Q(L)\, D^Q(L)(x_t - \mu^Q) = \Theta^Q(L)\Sigma\varepsilon_t^Q, \tag{1}$$

where $\varepsilon_t^Q \sim N(0, I)$, ε_t^Q is a $K \times 1$ vector, $\Phi^Q(L) = I - \Phi_1^Q L - \Phi_2^Q L^2 - \ldots - \Phi_p^Q L^p$, Φ_i^Q is a $K \times K$ matrix for $i = 1, 2, \ldots, p$ and L is the usual lag operator. In addition, $\Theta^Q(L) = I + \Theta_1^Q L + \ldots + \Theta_q^Q L^q$, Θ_j^Q is a $K \times K$ matrix for $j = 1, 2, .., q$ and $D^Q(L)$ is a diagonal matrix defined as follows

$$D^Q(L) = \begin{pmatrix} (1-L)^{d_1^Q} & 0 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & (1-L)^{d_K^Q} \end{pmatrix}.$$

This VARFIMA(p,d,q) specification is sufficiently general and nests several existing specifications in the literature. For instance, if the lag polynomial term $D^Q(L)$ is an identity matrix, then we obtain a VARMA(p,q) specification. In addition, if the lag polynomial $\Theta^Q(L)$ is also an identity, then we obtain a familiar VAR(p) specification.

If all eigenvalues of the convolution $\Phi^Q(L)\, D^Q(L)$ fall inside the unit circle, then x_t can be expressed as an infinite order vector moving average VMA(∞) process as well:

$$x_t = \mu^Q + \sum_{i=0}^{\infty} \Psi_i^Q \varepsilon_{t-i}^Q, \tag{2}$$

where Ψ_i^Q is a $K \times K$ matrix.

Furthermore, I assume that the pricing factors x_t follows a VARFIMA(p,d,q) process under the P-measure as well

$$\Phi(L)\, D(L)\, (x_t - \mu) = \Theta(L)\Sigma\varepsilon_t, \tag{3}$$

where $\varepsilon_t \sim N(0, I)$ and $D(L)$ is a diagonal matrix as before with possibly different orders of integration in the main diagonal. Under the assumption that all eigenvalues of the convolution $\Phi(L)\, D(L)$ lie inside the unit circle, x_t can be represented as an infinite order VMA(∞) process

$$x_t = \mu + \sum_{i=0}^{\infty} \Psi_i \varepsilon_{t-i}, \tag{4}$$

where Ψ_i is a $K \times K$ matrix. The short term interest rate r_t is defined as a linear function of the pricing factors

$$r_t = \delta_0 + \delta_1' x_t, \tag{5}$$

where δ_0 is a scalar and δ_1 is a $K \times 1$ vector. The (logarithmic) stochastic discount factor (SDF) is a quadratic function of the K risk factors

$$-m_{t+1} = r_t + \frac{1}{2}\lambda_t' \lambda_t + \lambda_t' \varepsilon_{t+1}, \tag{6}$$

where λ_t is a vector of the market price of risks, which is defined as a linear function of the shocks as follows

$$\lambda_t = \lambda_0 + \sum_{i=0}^{\infty} \zeta_i \varepsilon_{t-i}, \tag{7}$$

where λ_0 is a $K \times 1$ vector and ζ_i is a $K \times K$ matrix. Defining market price of risk as an infinite order VMA(∞) process is in the spirit of Christoffersen, Elkamhi, Feunou, and Jacobs (2009) and Osterrider and Schotman (2017). In addition to encompassing the widely used affine specification, such definition of λ_t enables us to specify a more general affine process in the factors as clearly exhibited in the following example.

Example 1. *Let x_t be a single factor. Define x_t under the P-measure as an $ARFIMA(1, d, 0)$ process*

$$(1-L)^d x_t = \mu + \phi(1-L)^d x_{t-1} + \Sigma\varepsilon_t^P, \tag{8}$$

and the market price of risk, λ_t, as

$$\lambda_t = \lambda_0 + \lambda_1(1-L)^d x_t. \tag{9}$$

Assuming the (logarithmic) stochastic discount factor as in (6) and using $\varepsilon_t^P = \varepsilon_t^Q - \lambda_{t-1}$, x_t under the Q-measure becomes

$$(1-L)^d x_t = (\mu - \Sigma\lambda_0) + (\phi - \Sigma\lambda_1)(1-L)^d x_{t-1} + \Sigma\varepsilon_t^Q.$$

Therefore, although x_t is a long memory process with the same order of integration d under both measures, the market price of risk λ_t evolves as an I(0) process. Also note that despite being an $I(0)$ process, λ_t is specified as a linear function of the infinite lags of the factor, x_t.

Contrary to *Example 1* or extensively employed standard $I(0)/I(1)$ process, generally it is not straightforward to determine the orders of integration for factors under different measures and the market price of risk in long memory models. To clarify this point further, suppose that x_t under the P-measure remains as in (8) in Example 1 but now the market price of risk is defined as essentially affine in the factor

$$\lambda_t = \lambda_0 + \lambda_1 x_t. \tag{10}$$

Now, it is not immediately obvious how to pin down the order of integration for x_t under the Q-measure. Nevertheless, we can determine the orders of integration of factors under different measures by using the relations between the impulse-response functions of the factors and the market price of risk. With the specification for the market price of risk in (7), the impulse-response coefficients are related as follows (see Proof of Proposition 1):

$$\Psi_j = \Psi_j^Q + \sum_{k=0}^{j-1} \Psi_{j-k-1}^Q \zeta_k, \tag{11}$$

for $j = 1, 2, ...$ However, if the market price of risk is specified as essentially affine in the factors, then the relation becomes recursive:

$$\Psi_j = \Psi_j^Q + \sum_{k=0}^{j-1} \Psi_{j-k-1}^Q \lambda_1 \Psi_k. \tag{12}$$

The following proposition demonstrates how to determine the orders of integration if the factors x_t under different measures are defined as long memory processes with the market price of risk λ_t given in (7). In other words, it shows how the orders of integration change when we change probability measures.

Proposition 1: *Suppose that*

(A1) factor dynamics under the Q-measure as well as the P-measure are given as in (1) and (3),

(A2) orders of integration for factors under the Q-measure and P-measure are $d^Q \geq 0$ and $d^P \geq 0$, respectively,

(A3) (logarithmic) stochastic discount factor is defined as in (6)

(A4) impulse-response coefficients are not related recursively as in (12),

(A5) the market price of risk λ_t is defined as in (7) with orders of integration $d^\lambda \geq 0$.

If x_t is a single factor, then the relation between the orders of integration is as follows:

$$d^P = d^Q + d^\lambda. \tag{13}$$

Proof: See Appendix A.

Remark 1. *Osterrieder and Schotman (2017) estimate different orders of integration for the* *market price of risk λ_t and r_t under the P-measure. They assume that the order of integration $0 \leq d^\lambda < 0.5$ whereas $0 \leq d^r < 1$ under the P-measure. Thus, based on Proposition 1, we can say that r_t under the Q-measure is a long memory process with the order of integration $d^Q = d^P - d^\lambda$, which can be easily computed from their estimations.*

3 VMA(∞) representation

To estimate a DTSM with long memory in the latent factors, we can cast it in a state-space framework. Golinski and Zaffaroni (2016) estimate a DTSM with long memory process in a state-space framework using its VMA(∞) representation. This section illustrates the Golinski and Zaffaroni (2016) framework and discusses its shortcomings. In the next section, I explain some advantages of the VAR(∞) representation that I propose to estimate a long memory process in a state-space setting.

A $K \times 1$ dimension vector of pricing factors x_t with long memory has an infinite dimensional state-space representation under the P-measure (see Chan and Palma, 1998). Golinski and Zaffaroni (2016) differentiate between state factors x_t and latent variables C_t. They define an infinite dimensional state vector C_t as

$$C_t = \begin{bmatrix} E[x_t|x_t, x_{t-1}, ...] \\ E[x_{t+1}|x_t, x_{t-1}, ...] \\ \cdots \end{bmatrix}. \tag{14}$$

Thus, C_t captures all conditional expectations of x_t at time t under the P-measure. Then it is possible to represent the VARFIMA(p,d,q) process defined for x_t in the state-space setup relying on its infinite order VMA(∞) representation as follows

$$C_t = FC_{t-1} + H\varepsilon_t, \tag{15}$$

where an infinite dimensional H matrix is constructed using the impulse-response coefficients Ψ_i of the VARFIMA(p,d,q) process

$$H = \begin{pmatrix} \Psi_0 \\ \Psi_1 \\ \cdots \end{pmatrix}.$$

In addition, F is a double-infinite dimensional selection matrix with the following structure

$$F = \begin{pmatrix} 0 & \mathbb{I}_K & 0 & 0 & \cdots \\ 0 & 0 & \mathbb{I}_K & 0 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix},$$

where $\mathbb{I}_K$ is a $K \times K$ identity matrix. The relationship between the factors and the latent variables is defined as

$$x_t = GC_t, \tag{16}$$

where G is an infinite dimensional selection matrix

$$G = \begin{pmatrix} \mathbb{I}_K & 0 & 0 & \dots \end{pmatrix}. \tag{17}$$

Golinski and Zaffaroni (2016) assume an essentially affine specification (10) for the market price of risk. Given the stochastic discount factor in (6), it is possible to show that the zero-coupon bond prices are affine in the state vector C_t.

Proposition 2: *Suppose that the pricing factors $x_t = GC_t$ under the Q- and P-measures are defined in (1) and (3), respectively whereas the short rate is defined in (5). Given the stochastic factor defined in (6) and the market price of risk in (10), the no-arbitrage zero-coupon bond price P_t^n at time t for maturity n is*

$$P_t^n = exp(A_n + B_n' C_t), \tag{18}$$

where the state vector C_t is defined in (14) and (15), the loadings A_n and B_n satisfy the following Riccati recursions

$$\begin{aligned} A_{n+1} &= A_n - B_n' H\lambda_0 + 0.5 B_n' H H' B_n - \delta_0 \\ B_{n+1}' &= B_n'(F - H\lambda_1 G) - \delta_1' G \end{aligned} \tag{19}$$

Proof: See Appendix A.

Remark 2. *The pricing formulas for the zero-coupon bond prices are provided in Theorem 4.1 of Golinski and Zaffaroni (2016), where they assume an ARFIMA process for each factor. The pricing formulas in Proposition 2 represent a more general case where the factors follow a VARFIMA process. Note also the similarity between the formulas in (19) and the pricing formulas for the short memory case.*

Despite the infinite dimensionality of the state space, Chan and Palma (1998) show how to obtain the exact likelihood based on a sample of T observations. They also propose an approximate maximum likelihood approach using a truncated state-space system which delivers the same asymptotic properties, though in a fewer number of steps. Chan and Palma (1998) propose to set the truncation lag $m = T^{\beta}$, where $\beta \geq 1/2$. While setting a larger truncation lag may enhance convergence properties of the estimators, Grassi and de

Magistris (2014) claim that choosing a too large truncation lag may lead to biased estimators in finite samples.

For the truncated state-space representation, we can consider an approximate model for x_t of the form

$$x_t = \mu + \sum_{i=0}^{m} \Psi_i \varepsilon_{t-i}, \tag{20}$$

which corresponds to a VMA(m) instead of a VMA(∞) process. Then C_t is truncated at the lag m and the corresponding $Km \times 1$ vector $\tilde{C}_t$ will be defined as follows

$$\tilde{C}_t = \begin{bmatrix} E[x_t|x_t, x_{t-1}, ...] \\ E[x_{t+1}|x_t, x_{t-1}, ...] \\ \dots \\ E[x_{t+m-1}|x_t, x_{t-1}, ...] \end{bmatrix},$$

and the selection matrices are

$$\tilde{G} = \begin{pmatrix} \mathbb{I}_K & 0 & \dots & 0 \end{pmatrix}, \qquad \tilde{F} = \begin{pmatrix} 0 & \mathbb{I}_K & \dots & \dots \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \mathbb{I}_K \end{pmatrix},$$

where $\tilde{G}$ is a $K \times Km$ matrice and $\tilde{F}$ is a $Km \times Km$ matrice.

Golinski and Zaffaroni (2016) suggest to use the VMA(∞) representation of a long memory process to estimate DTSMs with latent variables in the state-space framework. By implicitly setting μ to zero, Golinski and Zaffaroni (2016) estimate δ_0 and λ_0 together with the state vector C_t (see Table 4 in their study). The following proposition shows that it is not possible to jointly identify (δ_0, λ_0) together with the state vector C_t.

Proposition 3: *Suppose that*

(i) the unobserved pricing factors $x_t = GC_t$ follows a VARFIMA(p,d,q) process in (3) where C_t and G are defined as in (14) and in (17).

(ii) the measurement equation is defined in (18) with its loadings A_n and B_n given in (19).

(iii) the transition equation is defined in (15).

(iv) μ in (3) is set to zero, i.e., $\mu = 0$.

(v) all parameters other than (δ_0, λ_0) are econometrically identified.

Then it is not possible to identify (δ_0, λ_0) together with the latent state vector C_t.

Proof: See Appendix A.

Remark 3. *In short memory case, when all factors are unobserved, it is only possible to identify two out of the three level parameters $(\delta_0, \lambda_0, \mu)$ whereas in the VMA(∞) representation* *it is only possible to identify δ_0 in a multivariable setting. However, if $K = 1$, it is possible to identify either δ_0 or λ_0 but not both.*

It can be shown that it is not possible to identify μ even when all parameters and pricing factors are identified as it does not appear neither in the transition equation nor in the measurement equation. For instance, Golinski and Zaffaroni (2016) estimate a five factor model with the first two factors being inflation and real activity. Although both inflation and real activity are observed, the authors cannot estimate μ for these factors as this is not possible in the VMA(∞) representation. An obvious refinement would be to make a minor modification to the definition of the state variable C_t in Golinski and Zaffaroni (2016) and to re-define it as de-meaned

$$C_t = \begin{bmatrix} E[x_t - \mu | x_t, x_{t-1}, ...] \\ E[x_{t+1} - \mu | x_t, x_{t-1}, ...] \\ \cdots \end{bmatrix}.$$

This allows the intercept under the physical measure to show up in the definition of the factors, i.e., $x_t = \mu + GC_t$. In this case, the recursion for the loading A_n becomes

$$A_{n+1} = A_n - \delta_1' \mu - B_n' H \lambda_0 + 0.5 B_n' H H' B_n - \delta_0.$$

Given that C_t and all parameters except μ and δ_0 are identified, it is not possible to econometrically identify μ and δ_0. To see this classical identification problem, choose an arbitrary value for $\mu^* \neq \mu$ and set the value of δ_0^* such that

$$(\delta_0^* - \delta_0) = -\delta_1' (\mu^* - \mu).$$

Then one can use the induction argument to prove that μ and δ_0 are not jointly identified.

Finally, it is interesting to explore a more general specification for the market price of risk given in (7) and its implications for pricing of the zero-coupon bonds and estimation of the term structure model. This specification allows to define different orders of integration for pricing factors, x_t under the two probability measures and for the market price of risk. To present the pricing formulas for the zero-coupon bond prices in this framework, one final step is to determine the law of motion for C_t under the risk neutral probability measure since it is not as straightforward as in the case of the essentially affine specification for the market price of risk.

Similar to the definition of C_t under the physical probability measure, we can define a state variable C_t^Q in which the conditional expectations are taken under the risk-neutral probability measure

$$C_t^Q = \begin{bmatrix} E^Q[x_t - \mu^Q | x_t, x_{t-1}, ...] \\ E^Q[x_{t+1} - \mu^Q | x_t, x_{t-1}, ...] \\ \cdots \end{bmatrix} = \begin{bmatrix} E^Q[\sum_{i=0}^{\infty} \Psi_i^Q \varepsilon_{t-i}^Q | x_t, x_{t-1}, ...] \\ E^Q[\sum_{i=0}^{\infty} \Psi_{i+1}^Q \varepsilon_{t-i}^Q | x_t, x_{t-1}, ...] \\ \cdots \end{bmatrix}.$$

It is possible to use the relation between the impulse-responses in (11) to establish that C_t and C_t^Q are the same state variables. Then the VARFIMA process under the risk-neutral probability can be represented as

$$C_t = FC_{t-1} + H^Q \varepsilon_t^Q, \tag{21}$$

where the H^Q matrix is now constructed using the impulse-response coefficients Ψ_i^Q of the VARFIMA process under the risk neutral probability measure. The following proposition presents the pricing formulas for the bonds.

Proposition 4: *Suppose that the pricing factors $x_t = \mu^Q + GC_t$ under the Q- and P-measures are defined in (1) and (3), respectively whereas the short rate is defined in (5). Given the stochastic factor defined in (6) and the market price of risk in (7), the no-arbitrage zero-coupon bond price P_t^n at time t for maturity n is*

$$P_t^n = exp(A_n + B_n' C_t), \tag{22}$$

where the state vector C_t is defined in (14) and (15), the loadings A_n and B_n satisfy the following Riccati recursions

$$\begin{aligned} A_{n+1} &= A_n - \delta_1' \mu^Q + 0.5 B_n' H^Q (H^Q)' B_n - \delta_0 \\ B_{n+1}' &= B_n' F - \delta_1' G \end{aligned} \tag{23}$$

Proof: Similar to Proof of Proposition 2.

Proposition 5: *Suppose that the unobserved factors $x_t = \mu^Q + GC_t$ under the Q- and P-measures are defined in (1) and (3), respectively. The transition equation for the state-space system is given in (15) and the measurement equation is defined by the price relation given in (22) . Then it is not possible to econometrically identify the intercept, μ of the factors under the P-measure. Besides, it is only possible to identify the level of the short rate, δ_0 together with the state vector C_t but not the intercept, μ^Q under the risk neutral probability measure.*

Proof: Similar to Proof of Proposition 3 and the induction argument above.

All these discussions demonstrate that the VMA(∞) representation is not suitable for a more general asymmetric specification of factor dynamics as it does not allow properly identifing the levels together with the state vector. On the contrary, in the next section we will see that the proposed VAR(∞) representation are robust to the identification issues inflicting the VMA(∞) representation and does not fall behind of short memory models in that respect.

4 VAR(∞) representation

The long memory process can also be represented as an infinite order VAR(∞) process as shown by Palma (2007, p 74). Given that $\Theta(L)$ is invertable, the VARFIMA process given in (1) or (3) can be written as

$$\Phi(L)\,D(L)\,\Theta^{-1}(L)\,x_{t+1} = \mu + \Sigma\varepsilon_{t+1}, \qquad \varepsilon_{t+1} \sim N(0, I).$$

Denoting $\Pi(L) = \Phi(L)\,D(L)\,\Theta^{-1}(L)$ and using the VAR(∞) representation, x_t can be expressed as

$$x_{t+1} = \mu + \pi_1 x_t + \pi_2 x_{t-1} + ... + \pi_m x_{t-m+1} + ... + \Sigma\varepsilon_{t+1}, \tag{24}$$

where π_i for $i = 1, 2, ...$ are obtained from the VAR(∞) representation as detailed in the appendix. It is worth noting that it is possible to use the truncated version of the state-space similar to the case of the VMA(∞) representation for estimations, i.e., it is possible to replace the VAR(∞) representation by a VAR(m) process.

Note that this representation of long memory imposes a high degree of structure on vector autoregressive coefficients π_i without compromising the degrees of freedom. Joslin et al (2013) estimate an asymmetric specification using a VAR(p) process with a maximum lag order $p = 4$ under the P-measure. Even specifying a smaller lag order for a VAR(p) process quickly wipes out the degrees of freedom when the number of variables included in a VAR(p) process increases gradually. On the contrary, the above specification allows us to choose as many lags as we wish while only using a small number of parameters to be estimated. For instance, in the following section I estimate a VAR(60) (i.e. setting the truncation lag $m = 60$) with 3 variables using only 18 parameters whereas in the standard unrestricted VAR framework, this would require around 190 parameters.

To estimate (24) in the state-space setting, let us define a latent variable C_t as follows

$$C_t = \begin{bmatrix} x_t \\ x_{t-1} \\ \dots \end{bmatrix}, \tag{25}$$

where C_t is a $Km \times 1$ vector. Then using its companion form, we can express the factor dynamics given in (24) more compactly. For that define $\tilde{\mu}$, H and Φ as

$$\tilde{\mu} = \begin{bmatrix} \mu \\ 0 \\ \dots \end{bmatrix}, \qquad H = \begin{pmatrix} 1 \\ 0 \\ \dots \end{pmatrix} \otimes \Sigma, \qquad \Phi = \begin{pmatrix} \pi_1 & \pi_2 & \dots & \pi_{m-1} & \dots \\ \mathbb{I}_K & 0 & \dots & 0 & \dots \\ \dots & \dots & \dots & \dots & \dots \end{pmatrix},$$

where $\otimes$ is the Kronecker product, $\tilde{\mu}$ is a $Km \times 1$ vector, H is a $Km \times K$ matrix and Φ is a $Km \times Km$ matrix. Thus, the latent variable dynamics are expressed as

$$C_{t+1} = \tilde{\mu} + \Phi C_t + H\varepsilon_{t+1}, \tag{26}$$

and the factors x_t are defined as

$$x_t = GC_t, \tag{27}$$

where $K \times Km$ selection matrix G is defined as before

$$G = \begin{pmatrix} \mathbb{I}_K & 0 & \dots \end{pmatrix}.$$

Using a similar notation, the state variable dynamics under the risk neutral measure can be written as

$$C_{t+1} = \tilde{\mu}^Q + \Phi^Q C_t + H\varepsilon_{t+1}. \tag{28}$$

The following proposition shows that the pricing formulas for the zero-coupon bonds are almost the same with the standard case.

Proposition 6: *Suppose that the pricing factors $x_t = GC_t$ under the Q- and P-measures are defined in (1) and (3), respectively whereas the short rate is defined in (5). Given the stochastic factor defined in (6) and the market price of risk in (7), the no-arbitrage zero-coupon bond price P_t^n at time t for maturity n is*

$$P_t^n = A_n + B_n' C_t, \tag{29}$$

where the state vector C_t is defined in (25) and (26), the loadings A_n and B_n satisfy the following Riccati recursions

$$A_{n+1} = A_n - B_n'\mu^Q + 0.5B_n'HH'B_n - \delta_0 \tag{30}$$
$$B_{n+1}' = B_n'\Phi^Q - \delta_1'G$$

Proof: Similar to the proofs above.

Apparently, in the VAR(∞) representation, it is possible to identify two out of the three level parameters similar to short memory models as the identification problem is essentially the same.

5 Data and estimation method

The estimation sample is started as in Bauer and Rudebusch (2016) and covers the period January, 1985 - December, 2018. I employ the 3-month and the 6-month T-bill yields, as well as the 1-, 2-, 3-, 5-, 7-, and 10-year bond yields in the estimations. The 3-month yield data is taken from Fama-Bliss estimations, whereas the 6-month comes from the St Louis Fed database. The remaining longer term yields are obtained from the periodically updated database of Gürkaynak, Sack and Wright (2007). All monthly yields are defined as the end of the month basis. For survey data, I draw on the Survey of Professional Forecasters (SPF) database of the Philadelphia Fed. For more information please see the appendix.

Potentially, it is possible to specify and estimate a long memory process under both measures in a dynamic term structure model. However, the proposed framework is quite flexible and also allows to define a different dynamics under the two measures. Here, I asymmetrically treat the risk-neutral and the physical dynamics. I specify a VAR(1) process under the risk neutral measure, which is a common specification in the literature but estimate a VARFIMA(1,d,0) process under the physical measure. It is assumed that the market price of risk λ_t follows a linear process given in (7). Although a long memory specification under both measures may have its advantages, the asymmetric specification speeds up the computations and allows me to carry out an extensive comparison with the long memory model. Furthermore, I specify a VAR(1) process for the factors under both measures in the short memory models.

In this study, I estimate both affine as well as shadow rate models (SRMs) with long and short memory. In the SRMs, the factor specifications and their representations remain the same as in the affine models. However, the short rate specification is modified slightly. Now, it is assumed that the shadow rate s_t is expressed as

$$s_t = \delta_0 + \delta_1^{'} x_t, \tag{31}$$

whereas the short rate is $r_t = max(0, s_t)$. Since SRMs do not have a closed form solution, I follow Priebsch (2013) to obtain the expressions for the zero-coupon bond yields. In my estimations, I rely on the first order approximation (see also Krippner, 2012) as it delivers adequate fit and sufficiently smaller errors with substantial reduction in computational time.

I apply the standard Kalman filter while estimating the affine models. To estimate the SRMs, I employ the widely used extended Kalman filter. In the case of the long memory SRM, the non-linear state-space representation becomes

$$y_t = Z(x_t) + v_t,$$
$$C_{t+1} = \tilde{\mu} + \Phi C_t + H\varepsilon_{t+1},$$

where the measurement error, $v_t \sim N(0, \sigma_v^2 I)$, y_t is a vector of yields, and $x_t = GC_t$. $Z(x_t)$ is a non-linear function in the pricing factors and obtained from Priebsch (2013) approximation under the Q-measure. Because I use the truncated state-space, I set the truncation lag $m = 60$ in the VAR(m) representation.

6 Empirical results

In this section, I present my results from the estimation of the term structure models. First of all, it should be mentioned that because the factors follow a short memory process under the Q-measure, the standard pricing formulas can be applied to the affine as well as the shadow rate models in the long memory framework. Second, as the factors are defined as a parsimonious VAR(1) process under the Q-measure, I apply the identification restrictions of Joslin et al (2011) in both short and long memory models. That is, I employ the restrictions that Φ^Q is an ordered Jordan matrix whereas Σ is a lower triangular matrix with positive diagonal entries. The δ_1 is restricted to the vector of ones and μ^Q is set to zero. In the long memory models, I additionally order the persistence estimates d in the descending order as in the Jordan form above.

Almost all models are estimated using three pricing latent factors, which is a common specification in the literature. However, in the forecasting exercise given in the appendix, I

Table 1: Short memory affine model

This table reports the empirical results from the standard short memory affine model over the period January, 1985-December, 2018. The standard errors in parentheses are calculated using the approach proposed by Harvey (1989), p 142 (see also p 938 of Andreasen, 2013). ϕ^Q represents the estimated eigenvalues. All estimated coefficients are statistically significant at the 5% level except the ones with the asterisk sign.

		Φ		ϕ^Q	μ		Σ	
latent 1	1.0055	0.0318	0.0448	0.9973	0.0008*	0.0041		
	(0.0045)	(0.0041)	(0.0042)	(0.0004)	(0.0009)	(0.0006)		
latent 2	−0.0176	0.9514	−0.0655	0.9684	−0.0021*	−0.0028	0.0053	
	(0.0051)	(0.0043)	(0.0038)	(0.0028)	(0.0013)	(0.0012)	(0.0011)	
latent 3	−0.0140	−0.0147	0.9195	0.8880	−0.0021	−0.0013*	−0.0046	0.0019
	(0.0023)	(0.0025)	(0.0035)	(0.0041)	(0.0009)	(0.0012)	(0.0012)	(0.0004)
δ_0	0.1455							
	(0.0010)							
σ_v	0.0007							
	(0.0000)							

also include inflation and real activity in addition to the three latent factors. Finally, I assume that all yields are measured with error.

Here, I only report the results for the affine models with short and long memory. The estimation results for the SRMs can be found in the appendix. The estimation results for the affine model with short memory as well as long memory are presented in Table 1 and Table 2, respectively. The parameters determining the risk-neutral dynamics (Φ^Q and δ_0) as well as Σ are similar in both specifications. In addition, the parameters are estimated very tightly in both cases. The eigenvalues from the Jordan form (ϕ^Q) in both models appear to be in line with the findings of the short memory literature (Joslin et al, 2011; Bauer et al, 2012). It is clear that the long memory assumption under the P-measure is innocuous and does not affect the risk-neutral dynamics.

Regarding the physical dynamics, it is clear that there are sizeable differences between the models. The parameters Φ and μ are estimated somewhat differently in both models. In the long memory case, the estimated orders of integration (d) for the factors are large and consistent with the findings in the literature (Abbritti et al, 2016; Osterrieder and Schotman, 2017). Since the estimated d coefficients are statistically larger than 0.5, all latent factors are non-stationary.

Table 2: Long memory affine model

This table reports the empirical results from the long memory affine model over the period January, 1985-December, 2018. The standard errors in parentheses are calculated using the approach proposed by Harvey (1989), p 142 (see also p 938 of Andreasen, 2013). ϕ^Q represents the estimated eigenvalues. All estimated coefficients are statistically significant at the 5% level except the ones with the asterisk sign.

	Φ			ϕ^Q	d	μ	Σ		
latent 1	0.0914	−0.1169	−0.0444	0.9973	0.8761	−0.0006	0.0039		
	(0.0042)	(0.0037)	(0.0054)	(0.0001)	(0.0033)	(0.0000)	(0.0001)		
latent 2	0.3805	0.7479	0.2572	0.9683	0.8761	0.0004	−0.0022	0.0046	
	(0.0034)	(0.0050)	(0.0042)	(0.0005)	(0.0034)	(0.0000)	(0.0002)	(0.0001)	
latent 3	−0.1807	−0.2832	0.3962	0.8892	0.5897	−0.0003	−0.0016	−0.0039	0.0018
	(0.0029)	(0.0033)	(0.0060)	(0.0021)	(0.0044)	(0.0000)	(0.0001)	(0.0001)	(0.0000)
δ_0	0.1450								
	(0.0014)								
σ_v	0.0007								
	(0.0000)								

7 Model properties

In this section, I discuss various properties of the estimated models, some of which may also be considered testing the ability of the models in matching conditional expectations of future yields. First, I present the in-sample fit of the models. Then I conduct the two "linear projections of yields" (LPY) tests proposed by Dai and Singleton (2002), which investigate whether the estimated models can replicate the desired slope coefficients from the standard and the risk-adjusted Campbell-Shiller regressions. I also report the results of the Mincer-Zarnowitz (1969) regressions for the models, which test whether the model-implied excess returns comply with the realized excess returns. Using another criteria, I explore whether the estimated models can predict the short rate expectations from the Survey of Professional Forecasters (SPF) for the 3-month bond yield over short and long horizons. In the last part, I discuss the estimates of term premium dynamics and the overall performance of the models. In the appendix, an interested reader can also find the results from a forecasting exercise for the affine models.

7.1 In-sample fit

The in-sample fit of all estimated models are very satisfactory (Table 3). Despite the 7 years of the zero lower bound (ZLB) period in the sample January, 1985 - December, 2018, the largest error in the affine models is not greater than 7.5 basis points. Long and short memory models fare comparably well in the estimation period. Once more, it is evident that the long

Table 3: In-sample fit

This table reports the in-sample fit of the estimated affine and shadow rate models with short "SM" and long "LM" memory over the period January, 1985-December, 2018. The fit is computed using the root mean squared error (RMSE). The errors are expressed in annualized basis points.

Model	3-month	6-month	1-year	2-year	3-year	5-year	7-year	10-year
Affine SM	7.03	6.52	7.44	3.28	4.81	5.58	3.21	6.32
Affine LM	7.17	6.48	7.45	3.35	4.82	5.56	3.25	6.35
SRM SM	7.51	6.57	6.88	3.59	4.79	5.07	3.22	5.51
SRM LM	6.88	6.41	7.01	3.54	4.78	5.26	3.18	5.89

memory specification under the physical measure does not influence the satisfactory fit of the models when compared to the short memory case. While I apply the first rather than second order approximation of Priebsch (2013) to find the solutions of the bond yields, the SRMs in both cases perform relatively well.

7.2 Campbell-Shiller regressions

Dai and Singleton (2002) propose two tests to check the ability of DTSMs to match the conditional mean of future yields. The first test (LPY I) explores whether population slope coefficients from the Campbell and Shiller (1991) regressions can match their counterparts from the data. It sets out and estimates the following regression

$$y_{t+m}^{j-m} - y_t^j = \delta_j + \phi_j \frac{m}{j-m}(y_t^j - y_t^m) + u_t^j, \tag{32}$$

where $u_t^j \sim IID(0, var(u_t^j))$ for maturities $j = m+1, m+2, ..., K$. A well-known result from the Campbell and Shiller (1991) study shows that the spread, $(y_t^j - y_t^m)$, can predict the movements in long maturity yields, $(y_{t+m}^{j-m} - y_t^j)$. Thus, in order for a DTSM to pass the LPY I test, the standard Campbell-Shiller regression on the simulated samples from the DTSM should replicate the slope coefficients from the data.

The second LPY II test investigates whether the yields in the sample comply with the expectations hypothesis once they are adjusted for risk using the term premium estimates from the DTSM. Specifically, it tests whether the loadings ϕ_j^Q are equal to 1 in the following regression

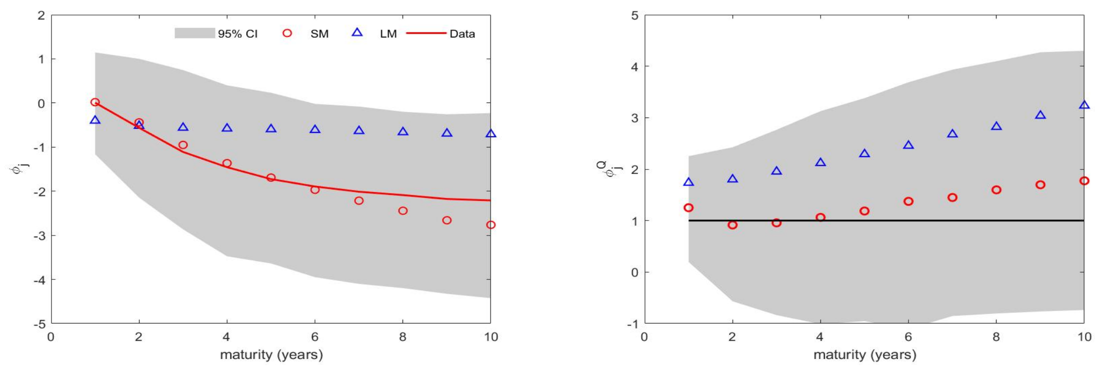
$$y_{t+m}^{j-m} - y_t^j - (c_{t+m}^{j-m} - c_t^{j-m}) + \frac{m}{j-m}\theta_t^{j-m} = \delta_j^Q + \phi_j^Q \frac{m}{j-m}(y_t^j - y_t^m) + u_t^{j,Q}, \tag{33}$$

where $c_t^j \equiv y_t^j - 1/j \sum_{i=0}^{j-1} E_t[r_{t+i}]$ is the term premium in the j^{th} maturity yield, $\theta_t^j \equiv f_t^j - E_t[r_{t+j}]$ is the term premium in the forward rate $f_t^j \equiv -log(P_t^{j+1}/P_t^j)$, and $u_t^{j,Q} \sim$

Figure 1: Affine models

This figure reports the Campbell-Shiller loadings implied by the model and the data. The "SM" denotes short memory model whereas the "LM" long memory. The 95% confidence interval (shaded area) for the LPY I test is based on the short memory model and computed using a block bootstrap of 5,000 replications with a window size of 12 months. The confidence bands for the LPY II tests are obtained using the term premium estimates from the short memory model.

(a) LPY I test **(b)** LPY II test



$IID(0, var(u_t^{j,Q}))$.

To carry out the suggested LPY I test, I generate 10,000 samples of the equal length as in the data from the term structure models. For each replication, I estimate (32) using $m = 6$ months and then take the mean of the slope estimates. For LPY II test, I use the term premium estimates from the models and adjust the corresponding yields accordingly. I compare these model-implied loadings with the estimates from the data covering the period January, 1985 - December, 2018. The 95% confidence intervals are calculated using the Campbell-Shiller regressions applied to the data. Here, I run 5,000 block bootstrap replications employing jointly the regressand and the regressor in the regressions with a block window of 12 months.

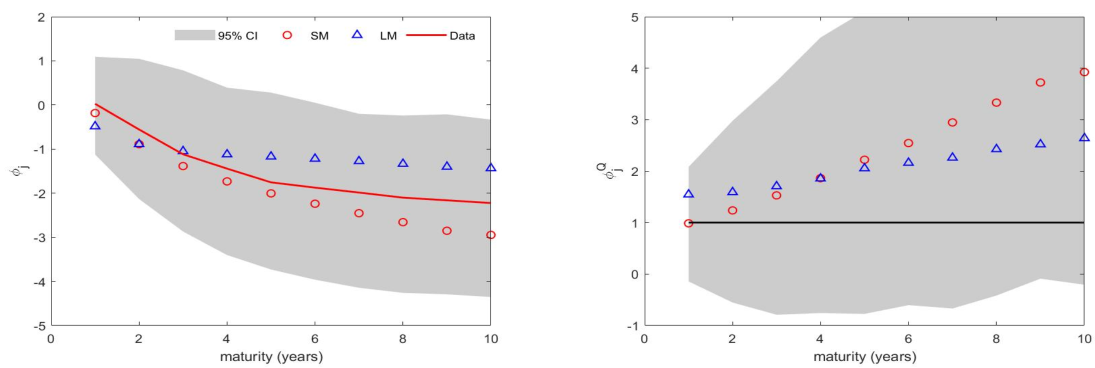
Figure 1 depicts the results of the Campbell-Shiller regressions for the affine models conducted for the yield maturities 1 to 10 years. As expected in the LPY I test, the simulated Campbell-Shiller loadings (ϕ_j) decline and remain in the 95% confidence band in both affine models although this decline is more pronounced in the short memory model. To pass the second test, the Campbell-Shiller loadings (ϕ_j^Q) from the models should not be statistically different from 1 (black line). Because the affine models meet this criterion, they pass the LPY II test as well.

Figure 2 shows the results of the two tests for the SRMs. Clearly, as the simulated Campbell-Shiller regression loadings remain in the confidence bands, the SRMs match the

Figure 2: Shadow rate models

This figure reports the Campbell-Shiller loadings implied by the model and the data. The "SM" denotes short memory model whereas the "LM" long memory. The 95% confidence interval (shaded area) for the LPY I test is based on the short memory model and computed using a block bootstrap of 5,000 replications with a window size of 12 months. The confidence bands for the LPY II tests are obtained using the term premium estimates from the short memory model.

(a) LPY I test **(b)** LPY II test



data and pass the two tests. Here as well, the short memory model performs relatively better in the first test as it generally aligns with the data (red solid line). However, these differences are small and generally not statistically significant.

7.3 Mincer-Zarnowitz test

The Mincer and Zarnowitz (1969) regression is another popular tool to test forecasting performance of a candidate model. In a DTSM context, I apply the Mincer and Zarnowitz (1969) regression to check the ability of the models in matching conditional moments. In particular, I run an m-period realized excess return on the model-implied expected excess return, i.e.

$$rx_{t+m}^{m,j} = \alpha_{0,j} + \alpha_{1,j} E_t[rx_{t+m}^{m,j}] + u_{t+m}^{m,j}, \tag{34}$$

where $rx_{t+m}^{m,j} \equiv -(j-m)y_{t+m}^{j-m} + jy_t^j - my_t^m$. For a satisfactory model, the intercept $\alpha_{0,j}$ should be equal to zero, whereas the slope $\alpha_{1,j}$ should not be statistically different from 1.

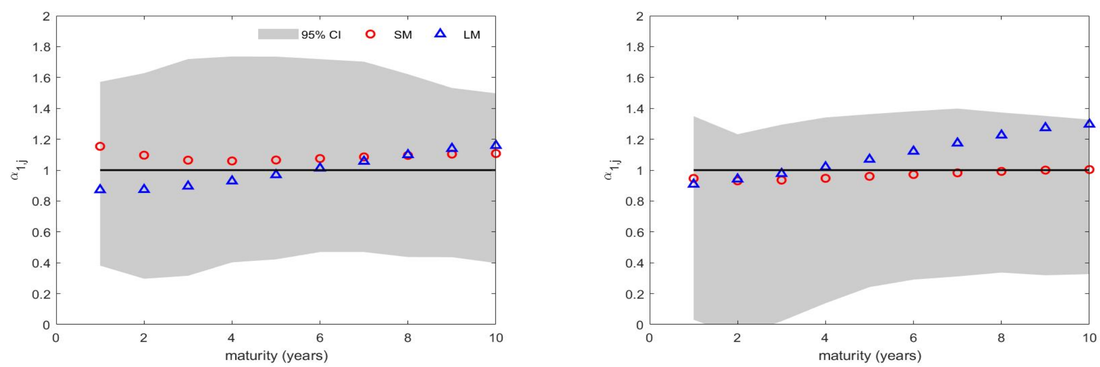
As in the case of the LPY tests, I generate 10,000 samples of equal length as in the data from each of the term structure models. For each replication, I estimate (34) using $m = 6$ months and then take the mean of the estimates. I compare these model-implied loadings with the estimates from the data covering the period January, 1985 - December, 2018.

Figure 3a depicts the estimation results obtained from the affine models for the slope coefficients $\alpha_{1,j}$. The horizontal axis shows the employed bond maturities $j = 1, 2, ..., 10$ in

Figure 3: Mincer-Zarnowitz test

This figure reports the Mincer-Zarnowitz loading implied by the model and the data. The "SM" denotes short memory model whereas the "LM" long memory. The 95% confidence intervals (shaded area) are obtained from the data and computed using a block bootstrap of 5,000 replications with a window size of 12 months.

(a) Affine models **(b)** Shadow rate models



years. It is clear that the slope coefficients $\alpha_{1,j}$ from the affine models are not statistically different from 1. Although not reported here, the intercept estimates $\alpha_{0,j}$ are not statistically different from zero as well. Thus, these models satisfy the Mincer-Zarnowitz test. Figure 3b displays the slope coefficients $\alpha_{1,j}$ for the SRMs from the Mincer-Zarnowitz regression. It is evident that they broadly satisfy the Mincer-Zarnowitz criterion while the short memory model performs marginally better.

7.4 Matching survey expectations

In this section, I discuss performances of the estimated models in fitting the 3 month short-rate expectations from the Survey of Professional Forecasters (SPF). First, I report the results of the models estimated with the yield data only in fitting 12-month- and 10-year (average)-ahead survey expectations for the 3-month bond yield. Then I also augment the yield data with the 6- and 12-month-ahead survey expectations from the SPF similar to Kim and Orphanides (2012) and examine whether the same models can improve upon their benchmark cases. Finally, I also augment the yield data with the survey expectations for the average 3-month bond yield over the next 10-year horizon and compare their performances with the rest of the models.

Table 4 presents the ability of the affine and the SRMs with short and long memory in fitting the short rate expectations from the SPF. Generally, the estimated models can fit the short horizon survey expectations comparably well even without the data augmentation. On average, the short memory models perform similarly with the long memory ones, except one case. Without data augmentation, the affine short memory model beats all the remaining models in fitting the 10-year average survey expectations. That is, all remaining models have larger than 200 basis points expectation errors over the next 10 years, whereas the short memory affine model has almost half of these errors.

Including only the short horizon expectations in the term structure estimation halves

Table 4: Matching short-rate expectations from surveys

This table reports the root mean squared errors (RMSE) in annualized basis points (bp) between the model-implied 3-month bond yield forecasts and the expected 3-month yield obtained from the Survey of Professional Forecasters (SPF) over the period January, 1985 - December, 2018. The "YDO" denotes the standard estimation of the models with the bond yields data only. The "SHS" denotes the models estimated with the yields data plus short horizon 6- and 12-month-ahead survey expectations from the SPF. The "LHS" denotes the models estimated with the yields data and the average 10-year expectations for the 3-month bond yield from the SPF.

	Affine		SRM	
	Short memory	Long memory	Short memory	Long memory
(a) 12-month				
Yields data only (YDO)	68.8	75.6	88.7	87.1
Short horizon survey data (SHS)	31.7	33.2	28.2	30.4
Long horizon survey data (LHS)	53.7	55.3	69.9	71.3
(b) 10-year average				
Yields data only (YDO)	109.9	217.3	214.8	252.3
Short horizon survey data (SHS)	64.8	85.3	65.3	77.0
Long horizon survey data (LHS)	27.5	23.9	28.9	27.8

the forecast errors for the period 12 months ahead and substantially reduces the errors for the next 10-year period (for more details see the appendix). Augmenting the models with only long horizon survey expectations helps to reduce the forecast errors for the long horizon substantially. To sum up, all the models perform satisfactorily well given the results from the literature (see, for instance, Andreasen and Meldrum, 2019).

7.5 Term premium

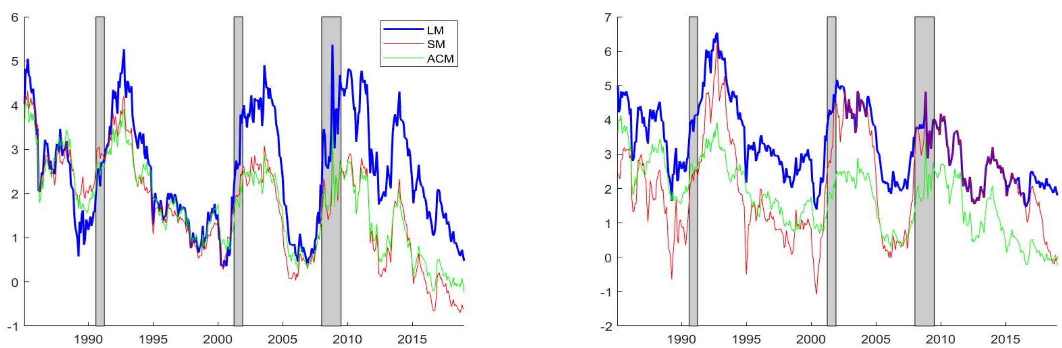
Term premium estimates of the long and short memory affine models for the 10-year bond yield are displayed in Figure 4a. It is striking that these two term premium estimates resemble the term premium estimates of Abbritti et al. (2016) during the period 1985-2011 (see p 347 in Abbritti et al, 2016). Not only their shapes, but also their scales are very similar. Abbritti et al. (2016) estimate a VARFIMA(1,d,0) model but employing the observed first three principal components (obtained from a panel of the bond yields) and a different estimation methodology in frequency domain. Nevertheless, model estimations using different methodologies and different sample periods produce similar term premium estimates. In addition, the term premium estimates from the affine models display a high correlation with the estimates of Adrian, Crump, and Moench (2013) study. This correlation is around 98% for the short memory model and 77% for the long memory model.

Figure 4a shows that the term premium estimates of the long memory affine model are

Figure 4: Term premium

Term premium estimates from long memory (LM) and short memory (SM) affine models. The green line (ACM) depicts Adrian, Crump, and Moench (2013) estimates. The shaded regions indicate NBER recession dates for the US. economy.

(a) Affine models **(b)** Shadow rate models



generally larger than the estimates of the short memory affine model during most of the sample period. Figure 4b displays the term premium estimates using the same maturity bond yield in the SRMs. They look like the estimates of the affine models despite the 7 years of the ZLB period in the sample. In the SRMs as well, the estimates of the term premium of the long memory model are generally larger than those of the short memory model.

All term premium estimates appear counter-cyclical as they rise before and during recessions and fall during expansions. However, they are generally lagging the cycle. The volatilities of the term premium estimates for the 10-year bond yield in affine models are similar and close to 1.2% whereas in the SRMs, the long memory model has a slightly lower volatility (1.2% vs 1.4%). Over the course of the sample, the 10-year bond yield exhibited a downward trend and declined by 8.3%. The long and short memory affine models attribute half of this decline, namely, 4.8% and 3.8%, to falling short-rate expectations. The SRMs explain a larger portion, i.e., around 5-6% of this decline by lower short-rate expectations. For comparison, the short-rate survey expectations attribute around half of this decline to falling short-rate expectations. Overall, the abilities of the models do not differ significantly in interpreting term premium dynamics.

8 Conclusion

In this study, I generalize the state-space approach proposed in Golinski and Zaffaroni (2016) to estimate long memory term structure models and address the shortcomings in their estimation framework. The computational efficiency of the new framework allows me to carry out an extensive comparison of long and short memory term structure models. Using a battery of tests, I find that the short memory term structure models perform just like the long memory models.

The estimated affine as well as the shadow rate models with short and long memory produce similar parameter estimates and term premium as in the literature. Moreover, the short and long memory models perform remarkably similar against a number of criteria. Even if there is a difference in their performance, this difference is statistically insignificant and does not provide strong support for any model type.

All estimated models achieve a similar in-sample fit, perform adequately in the Campbell-Schiller predictability tests as well as in the Mincer-Zarnowitz test. There is little to distinguish between the ability of these models in matching the short rate survey expectations as well. Only the magnitudes of the term premium estimates of the long and short memory models somewhat differ although the term premium dynamics is similar. In general, the long memory models have relatively larger term premium estimates. Overall, I show that standard term structure models with short memory produce similar term premium estimates as more complicated long memory models.

9 References

Abbritti, M., Gil-Alana, L. A., Lovcha, Y., and Moreno, A. (2016). Term structure persistence. Journal of Financial Econometrics, 14, 331-352.

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

Andreasen, M. M. and Meldrum, A. (2019). A shadow rate or a quadratic policy rule? The best way to enforce the zero lower bound in the United States. Journal of Financial and Quantitative Analysis, 54, 2261-2292.

Andreasen, M. M. and Christensen, B. J. (2015). The SR approach: A new estimation procedure for non-linear and non-Gaussian dynamic term structure models. Journal of Econometrics, 184, 420-451.

Andreasen, M. M. (2013). Non-linear DSGE models and the central Kalman filter. Journal of Applied Econometrics, 28, 929-955.

Banbura, M., Giannone, D., and Lenza, M., (2015). Conditional forecasts and scenario analysis with vector autoregessions for large cross sections. International Journal of Forecasting, 31, 739-756.

Bauer, M. D., and Rudebusch, G. D. (2019). Interest rates under the falling star. American Economic Review (forthcoming).

Bauer, M. D., and Rudebusch, G. D. (2016). Monetary policy expectations at the Zero Lower Bound. Journal of Money, Credit and Banking, 48, 439-1465.

Brockwell, P. J. and Davis, R. A. (1991). *Time series: theory and methods*, Springer.

Campbell, J.Y., and Shiller, R.J. (1991). Yield spreads and interest rate movements: a bird's eye view. Review of Financial Studies, 58, 495-514.

Campbell, J.Y., and Shiller, R.J. (1987). Cointegration and tests of present value models. Journal of Political Economy, 95, 1062-88.

Chan, N. H., and Palma, W. (1998). State space modelling of long-memory processes. Annals of Statistics, 26, 719-740.

Christoffersen, P., Elkamhi, R., Feunou, B., and Jacobs, K. (2010). Option valuation with conditional heteroskedasticity and nonnormality. Review of Financial Studies, 23, 2139-2183.

Dai, Q., and Singleton, K. J. (2002). Expectations puzzles, time-varying risk premia, and affine models of the term structure. Journal of Financial Economics, 63, 415-441.

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

Diebold, F. X., and Rudebusch, G. D. (1989). Long memory and persistence in aggregate output. Journal of Monetary Economics, 24, 189-209.

Durbin, J., and Koopman, S. J. (2012). *Time series analysis by state-space methods*, Second Edition. Oxford University Press, Oxford.

Engsted, T. and Tanggaard, C. (1994). Cointegration and the US term structure. Journal of Banking and Finance, 18, 167-181.

Golinski, A. and Zaffaroni, P. (2016). Long memory affine term structure models. Journal of Econometrics, 191, 33-56.

Grassi, S., and de Magistris, P. S. (2014). When long memory meets the Kalman filter: a comparative study. Computational Statistics and Data Analysis, 76, 301-319.

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

Hall, A.D., Anderson, H.M., and Granger, C.W.J. (1992). Cointegration analysis of treasury bill yields. Review of Economics and Statistics, 74, 116-126.

Harvey, A. C. (1989). *Forecasting, structural time series models and the Kalman filter*, Reprint Edition. Cambridge University Press, Cambridge.

Johansen, S., and Nielsen, M. (2012). Likelihood inference for a fractionally cointegrated vector autoregressive model, Econometrica 80, 2667-2732.

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

Joslin, S., Le, A., and Singleton, K. J. (2013). Gaussian macro-finance term structure models with lags. Journal of Financial Econometrics, 11, 581-609.

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

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

Knopp, K. (1990). *Theory and application of infinite series.* Dover Publications.

Kozicki, S., and Tinsley, P.A. (2001). Shifting endpoints in the term structure of interest rates. Journal of Monetary Economics, 47, 613-652.

Krippner, L. (2012). Modifying Gaussian term structure models when interest rates are near the zero lower bound. Reserve Bank of New Zealand Discussion Paper, 2012/02, available at SSRN: https://ssrn.com/abstract=2139510

Mincer, J.A. and Zarnowitz, V. (1969). *The evolution of economic forecasts.* NBER, Chapter 1, 3-46.

Osterrieder, D., and Schotman, P. C. (2017). The volatility of long-term bond returns: persistent interest shocks and time-varying risk premiums. Review of Economics and Statistics, 99, 884-895.

Palma, W. (2006). *Long-memory time series: theory and methods*, First Edition. Wiley-Interscience.

Priebsch, M. (2013). Computing arbitrage-free yields in multi-factor Gaussian shadow rate term structure models. Board of Governors of the Federal Reserve System (US) Finance and Economics Discussion Series, 2013-63, available at SSRN: https://ssrn.com/abstract=2350873

Wu, J. C., and Xia, F. D. (2016). Measuring the macroeconomic impact of monetary policy at the Zero Lower Bound. Journal of Money, Credit and Banking, 48, 253-291.

Appendices

A.1 Proofs

Lemma 1: *For finite $d_1 \geq 0$ and $d_2 \geq 0$ assume that*[4] *as $i \to \infty$*

$$\zeta_i \sim c\, i^{d_2-1}$$
$$\psi_i^Q \sim \begin{cases} c\, i^{d_1-1} \text{if } d_1 > 0 \\ r^i \text{ where } |r| < 1 \text{ if } d_1 = 0, \end{cases}$$

then as $n \to \infty$

$$\sum_{i=0}^{n} \psi_{n-i}^Q \zeta_i \sim c\, n^{d_1+d_2-1},$$

where c denotes an arbitrary constant, not always the same.

Proof: I only consider the case $d_1 > 0$ and $d_2 > 0$, as the proof of the other case follows a similar line of thoughts. Assume with no loss of generality that $\psi_i^Q \neq 0$ and $\zeta_i \neq 0$ for all $i < \infty$. Write

$$\sum_{i=0}^{n} \psi_{n-i}^Q \zeta_i \sim \sum_{i=0}^{n} c\,(n-i)^{d_1-1} i^{d_2-1} = \sum_{i=0}^{[n/2]} c\,(n-i)^{d_1-1} i^{d_2-1} + \sum_{i=[n/2]+1}^{n} c\,(n-i)^{d_1-1} i^{d_2-1},$$

where we express the above sum in two parts as we need to consider two cases: (i) when i is relatively smaller, i.e., $i \leq [n/2]$ and (ii) when i is relatively larger, i.e., $i > [n/2]$. For the first sum, as $n \to \infty$

$$\sum_{i=0}^{[n/2]} c\,(n-i)^{d_1-1} i^{d_2-1} \sim c\, n^{d_1-1} \sum_{i=0}^{[n/2]} i^{d_2-1} \sim c\, n^{d_1-1} n^{d_2} = c\, n^{d_1+d_2-1},$$

where the first relation $\sum_{i=0}^{[n/2]} c\,(n-i)^{d_1-1} i^{d_2-1} \sim c\, n^{d_1-1} \sum_{i=0}^{[n/2]} i^{d_2-1}$ follows from the intermediate results in Proof of Theorem 4.6 of Golinski and Zaffaroni (2016) and the second result, i.e., as $n \to \infty$, $\sum_{i=0}^{n} i^{d_2-1} \sim c\, n^{d_2}$ for $d_2 > 0$ follows from Lemma D.1 there (see also Knopp (1990), p 295). For the second sum, as $n \to \infty$

$$\sum_{i=[n/2]+1}^{n} c\,(n-i)^{d_1-1} i^{d_2-1} \sim c\, n^{d_2-1} \sum_{i=[n/2]+1}^{n} (n-i)^{d_1-1} =$$
$$n^{d_2-1} \sum_{i=0}^{[n/2]-1} i^{d_1-1} \sim c\, n^{d_2-1} n^{d_1} = c\, n^{d_1+d_2-1}.$$

[4] I say $a_n \sim b_n$, where $b_n \neq 0$, when $\frac{a_n}{b_n} \to 1$ as $n \to \infty$.

Thus, as $n \to \infty$

$$\sum_{i=0}^{n} \psi^Q_{n-i} \zeta_i \sim c\, n^{d_1+d_2-1}.$$

□

The result in Lemma 1 is used to prove Proposition 1.

Proof of Proposition 1: We first establish a relation between the impulse-response coefficients of the factors x_t under different measures and the market price of risk, λ_t and then apply Lemma 1 to this relation.

From the existence of the Radon-Nikodym derivative and Gaussian assumption of ε_t, we obtain $\varepsilon^Q_t = \varepsilon_t + \lambda_{t-1}$. Substituting λ_t given in (7) and $\varepsilon^Q_t = \varepsilon_t + \lambda_{t-1}$ in (2), we obtain

$$\begin{aligned} x_t &= \mu^Q + \sum_{j=0}^{\infty} \Psi^Q_j \varepsilon^Q_{t-j} = \mu^Q + \sum_{j=0}^{\infty} \Psi^Q_j (\varepsilon_{t-j} + (\lambda_0 + \sum_{k=0}^{\infty} \zeta_k \varepsilon_{t-j-k-1})) \\ &= \mu^Q + (\sum_{j=0}^{\infty} \Psi^Q_j)\lambda_0 + \sum_{j=0}^{\infty} \Psi^Q_j \varepsilon_{t-j} + \sum_{j=0}^{\infty}\sum_{k=0}^{\infty} \Psi^Q_j \zeta_k \varepsilon_{t-j-k-1} \\ &= \mu^Q + (\sum_{j=0}^{\infty} \Psi^Q_j)\lambda_0 + \sum_{j=0}^{\infty} \Psi^Q_j \varepsilon_{t-j} + \sum_{k=0}^{\infty}\sum_{j=k+1}^{\infty} \Psi^Q_{j-k-1} \zeta_k \varepsilon_{t-j} \\ &= \mu^Q + (\sum_{j=0}^{\infty} \Psi^Q_j)\lambda_0 + \sum_{j=0}^{\infty} \Psi^Q_j \varepsilon_{t-j} + \sum_{j=1}^{\infty}\sum_{k=0}^{j-1} \Psi^Q_{j-k-1} \zeta_k \varepsilon_{t-j}. \end{aligned}$$

Thus, using the VMA(∞) representation of x_t under the physical measure in (4) and equating the coefficients, we obtain

$$\mu = \mu^Q + (\sum_{j=0}^{\infty} \Psi^Q_j)\lambda_0, \tag{A.1}$$

for $j = 1, 2, ...$

$$\Psi_j = \Psi^Q_j + \sum_{k=0}^{j-1} \Psi^Q_{j-k-1} \zeta_k, \tag{11}$$

and $\Psi_0 = \Psi^Q_0$.

In the case of a single factor, as $j \to \infty$, the impulse-responses of x_t under the Q-measure and market price of risk, λ_t, are given as

$$\begin{aligned} &\zeta_j \sim c\, j^{d^\lambda - 1} \\ &\Psi^Q_j \sim \begin{cases} c\, j^{d^Q-1} \text{if } d^Q > 0 \\ r^j \text{ where } |r| < 1 \text{ if } d^Q = 0. \end{cases} \end{aligned}$$

Then as $j \to \infty$, by Lemma 1

$$\sum_{k=0}^{j-1} \Psi^Q_{j-k-1} \zeta_k \sim c\, j^{d^Q + d^\lambda - 1}.$$

Thus,

$$\Psi_j \sim c\, j^{d^P - 1},$$

where $d^P = max(d^Q, d^Q + d^\lambda) = d^Q + d^\lambda$ due to the assumptions of $d^Q \geq 0$ and $d^\lambda \geq 0$. $\square$

Proof of Proposition 2: To write the pricing formulas for the zero coupon bonds, I rely on the following relation

$$P_t^{n+1} = E_t^Q[exp(-r_t) P_{t+1}^n].$$

Before starting, note that by a mere change of measure the dynamics of C_t under the Q-measure is

$$C_t = -H\lambda_0 + (F - H\lambda_1 G) C_{t-1} + H\varepsilon_t^Q. \tag{A.2}$$

I guess that the pricing relation is as follows

$$P_t^n = exp(A_n + B_n' C_t).$$

Thus,

$$\begin{aligned}
P_t^{n+1} &= E_t^Q[exp(-r_t) exp(A_n + B_n' C_{t+1})] \\
&= exp(-\delta_0 - \delta_1' G C_t + A_n - B_n' H\lambda_0 + B_n'(F - H\lambda_1 G)) E_t^Q[exp(B_n' H \varepsilon_{t+1}^Q)] \\
&= exp\Big(A_n - B_n' H\lambda_0 + 0.5 B_n' H H' B_n - \delta_0 + (B_n'(F - H\lambda_1 G) - \delta_1' G) C_t\Big).
\end{aligned}$$

Thus,

$$\begin{aligned}
A_{n+1} &= A_n - B_n' H\lambda_0 + 0.5 B_n' H H' B_n - \delta_0 \\
B_{n+1}' &= B_n'(F - H\lambda_1 G) - \delta_1' G.
\end{aligned}$$

$\square$

Proof of Proposition 3: Let's fix some notations first. Define a new unobserved pricing vector $x_t^* = x_t + c$ and define $\tilde{c}$ as

$$\tilde{c} = \begin{pmatrix} c \\ c \\ \dots \end{pmatrix}.$$

Note that $c = G\,\tilde{c}$ where G is given in (17). Now, C_t^* is defined as

$$C_t^* = \begin{bmatrix} E[x_t^*|x_t^*, x_{t-1}^*, ...] \\ E[x_{t+1}^*|x_t^*, x_{t-1}^*, ...] \\ \cdots \end{bmatrix}, \tag{A.3}$$

where C_t^* follows a transition dynamics similar to C_t in (15)

$$C_t^* = FC_{t-1}^* + H\varepsilon_t. \tag{A.4}$$

In addition, since $C_t^* = C_t + \tilde{c}$, we can use the transition dynamics of C_t in (15) to write

$$C_t + \tilde{c} = (\tilde{c} - F\tilde{c}) + F(C_{t-1} + \tilde{c}) + H\varepsilon_t \Rightarrow$$

$$C_t^* = (\tilde{c} - F\tilde{c}) + FC_{t-1}^* + H\varepsilon_t. \tag{A.5}$$

Comparing (A.4) and (A.5), it is clear that

$$\tilde{c} - F\tilde{c} = 0, \tag{A.6}$$

holds in infinite dimension. Furthermore, choose $\lambda_0^* = \lambda_0 - \lambda_1 G\tilde{c}$ and $\delta_0^* = \delta_0 - \delta_1' G\tilde{c}$.
Here, I prove that $(\delta_0, \lambda_0, C_t)$ is observationally equivalent to $(\delta_0^*, \lambda_0^*, C_t^*)$. That is, if I show that for all m

$$P_t^{(m)} = exp(A_m^* + B_m' C_t^*) = exp(A_m + B_m' C_t), \tag{A.7}$$

then the proof is complete. Note that since all parameters except (δ_0, λ_0) are assumed to be identified, B_m is the same for both x_t and x_t^*.
I use the induction argument to prove (A7). First, to prove the base case $m = 1$, note that the price of the zero-coupon bond for the maturity $m = 1$ is

$$\begin{aligned} P_t^{(1)} &= exp(-\delta_0^* - \delta_1' G C_t^*) \\ &= exp(-\delta_0 + \delta_1' G\tilde{c} - \delta_1' G\tilde{c} - \delta_1' G C_t) \\ &= exp(-\delta_0 - \delta_1' G C_t), \end{aligned} \tag{A.8}$$

where I use the fact that $\delta_0^* = \delta_0 - \delta_1' G\tilde{c}$ and $C_t^* = C_t + \tilde{c}$.
Now, assume that for $m = n$, the relation in (A.7) holds, i.e.,

$$P_t^{(n)} = exp(A_n^* + B_n' C_t^*) = exp(A_n + B_n' C_t). \tag{A.9}$$

Since

$$A_n^* + B_n' C_t^* = A_n^* + B_n'\tilde{c} + B_n' C_t,$$

we obtain $A_n^* + B_n'\tilde{c} + B_n' C_t = A_n + B_n' C_t$ by relying on the relation in (A.9). In other words, our assumption for $m = n$ in (A.9) *implies* that

$$A_n^* + B_n'\tilde{c} = A_n. \tag{A.10}$$

Now, if I prove that the pricing relation in (A.7) holds for $m = n + 1$, then the proof is complete. To start, note that

$$\begin{aligned} P_t^{(n+1)} &= exp(A_{n+1}^* + B_{n+1}' C_t^*) \qquad \text{(A.11)} \\ &= exp(A_n^* - B_n' H \lambda_0^* + 0.5 B_n' H H' B_n - \delta_0^* + B_{n+1}'(C_t + \tilde{c})) \\ &= exp(A_n^* - B_n' H \lambda_0 + 0.5 B_n' H H' B_n - \delta_0 \\ &+ B_n' H \lambda_1 G \tilde{c} + \delta_1' G \tilde{c} + B_{n+1}' C_t + B_{n+1}' \tilde{c}) \end{aligned}$$

where in the second line I use the formula for A_{n+1}^* given in (19) and substitute $\lambda_0^* = \lambda_0 - \lambda_1 G\tilde{c}$ and $\delta_0^* = \delta_0 - \delta_1' G\tilde{c}$. Now, let's use the relation given in (A.6), i.e., $\tilde{c} - F\tilde{c} = 0$.

$$\begin{aligned} P_t^{(n+1)} &= exp(A_n^* - B_n' H \lambda_0 + 0.5 B_n' H H' B_n - \delta_0 \qquad \text{(A.11)} \\ &+ B_n' H \lambda_1 G \tilde{c} + \delta_1' G \tilde{c} + \underbrace{B_n'(\tilde{c} - F\tilde{c})}_{0} + B_{n+1}' C_t + B_{n+1}' \tilde{c}). \end{aligned}$$

Now, if we collect the terms, (A.11) can be re-written as

$$\begin{aligned} P_t^{(n+1)} &= exp(A_n^* - B_n' H \lambda_0 + 0.5 B_n' H H' B_n - \delta_0 \qquad \text{(A.11)} \\ &- \underbrace{(B_n'(F - H\lambda_1 G) - \delta_1' G)}_{B_{n+1}'} \tilde{c} + B_n'\tilde{c} + B_{n+1}' C_t + B_{n+1}' \tilde{c}). \end{aligned}$$

Note that $B_{n+1}' = B_n'(F - H\lambda_1 G) - \delta_1' G$ by (19). Now, if we use the *implied* assumption (A.10) (by assumption (A.9)), i.e., $A_n^* + B_n'\tilde{c} = A_n$ in (A.11) above, we obtain

$$\begin{aligned} P_t^{(n+1)} &= exp(A_{n+1}^* + B_{n+1}' C_t^*) \qquad \text{(A.11)} \\ &= exp(A_n^* - B_n' H \lambda_0 + 0.5 B_n' H H' B_n - \delta_0 - B_{n+1}'\tilde{c} + B_n'\tilde{c} + B_{n+1}' C_t + B_{n+1}' \tilde{c}) \\ &= exp(A_n - B_n' H \lambda_0 + 0.5 B_n' H H' B_n - \delta_0 + B_{n+1}' C_t) \\ &= exp(A_{n+1} + B_{n+1}' C_t), \end{aligned}$$

where the final line follows from the loading formula for A_{n+1} in (19). □

A.2 Short memory factor specifications

Although long memory specification is fairly general and nests short memory as well, I prefer to separately present short memory specification following the short memory literature. Thus, a vector of state factors $x_t = (x_{1t}\, x_{2t} \dots x_{Kt})'$ follows a parsimonious first order vector autoregressive VAR(1) process under the P-measure:

$$(x_t - \mu) = \Phi(x_{t-1} - \mu) + \Sigma\varepsilon_t, \tag{A.12}$$

where $\varepsilon_t \sim N(0, I)$, ε_t and μ are $K \times 1$ vectors, Φ is a $K \times K$ matrix and Σ is a $K \times K$ lower triangular matrix. The short term interest rate r_t is defined as a linear function of the pricing factors

$$r_t = \delta_0 + \delta_1^{'} x_t, \tag{5}$$

where δ_0 is a scalar and δ_1 is a $K \times 1$ vector. The (logarithmic) stochastic discount factor (SDF) is a quadratic function of the K risk factors

$$-m_{t+1} = r_t + \frac{1}{2}\lambda_t^{\prime}\lambda_t + \lambda_t^{\prime}\varepsilon_{t+1}, \tag{6}$$

where λ_t is a vector of the market price of risks, which are essentially affine in the pricing factors

$$\lambda_t = \lambda_0 + \lambda_1 x_t, \tag{10}$$

where λ_0 is a $K \times 1$ vector and λ_1 is a $K \times K$ matrix. Therefore, the pricing factors also follow a parsimonious VAR(1) process under the Q-measure

$$(x_t - \mu^Q) = \Phi^Q(x_{t-1} - \mu^Q) + \Sigma\varepsilon_t^Q, \tag{A.13}$$

where $\varepsilon_t^Q \sim N(0, I)$, ε_t^Q is a $K \times 1$ vector and

$$\mu^Q = (I - \Phi^Q)^{-1}(I - \Phi)\mu - \Sigma\lambda_0, \qquad \Phi^Q = \Phi - \Sigma\lambda_1.$$

A.3 Kalman Filter with Surveys

To estimate an affine, no-arbitrage DTSM, I apply the standard Kalman filter:

$$y_t = \mathcal{A} + \mathcal{B}\, x_t + v_t$$
$$C_{t+1} = \tilde{\mu} + \Phi C_t + H\varepsilon_{t+1},$$

where $x_t = GC_t$. The $J \times 1$ vector $\mathcal{A}$ and $J \times K$ dimension matrix $\mathcal{B}$ are defined as

$$\mathcal{A} = \begin{bmatrix} \mathcal{A}_{\tau_1} \\ \mathcal{A}_{\tau_2} \\ \dots \\ \mathcal{A}_{\tau_N} \end{bmatrix} \qquad \mathcal{B} = \begin{bmatrix} \mathcal{B}_{\tau_1} \\ \mathcal{B}_{\tau_2} \\ \dots \\ \mathcal{B}_{\tau_N} \end{bmatrix},$$

where $\mathcal{A}_{\tau_i} = -\frac{1}{\tau_i} A'_{\tau_i}$ and $\mathcal{B}_{\tau_i} = -\frac{1}{\tau_i} B'_{\tau_i}$. A_m and B_m are computed using the well-known Ricatti difference equations:

$$A_{m+1} = A_m + (\mu^Q)' B_m + \frac{1}{2} B'_m \Sigma \Sigma' B_m - \delta_0$$
$$B_{m+1} = (\Phi^Q)' B_m - \delta_1,$$

with starting values $A_0 = 0$ and $B_0 = 0$.

I augment the yields in the measurement equation with 3- and 6-month-ahead expectations as well as the 10-year-ahead (average) expectations of the 3-month T-bill rate. Now, the measurement equation becomes

$$\begin{bmatrix} y_t \\ y_{3,t}^{e,6} \\ y_{3,t}^{e,12} \\ \tilde{y}_{3,t}^{e,120} \end{bmatrix} = \begin{bmatrix} \mathcal{A} \\ \mathcal{A}_3 + \mathcal{B}_3 G (I - \Phi)^{-1} (I - \Phi^6) \tilde{\mu} \\ \mathcal{A}_3 + \mathcal{B}_3 G (I - \Phi)^{-1} (I - \Phi^{12}) \tilde{\mu} \\ \mathcal{A}_3 + 1/120 \mathcal{B}_3 G (I - \Phi)^{-1} \sum_{i=1}^{120} (I - \Phi^i) \tilde{\mu} \end{bmatrix} + \begin{bmatrix} \mathcal{B} G \\ \mathcal{B}_3 G \Phi^6 \\ \mathcal{B}_3 G \Phi^{12} \\ 1/120 \mathcal{B}_3 G \sum_{i=1}^{120} \Phi^i \end{bmatrix} C_t + \begin{bmatrix} v_t \\ v_t^e \\ v_t^e \\ \tilde{v}_t^e \end{bmatrix}.$$

The transition equation stays the same as in the case of the standard DTSM. For the VAR(∞) representation, C_t is defined as

$$C_t = \begin{bmatrix} x_t \\ x_{t-1} \\ x_{t-2} \\ \dots \\ x_{t-m+1} \end{bmatrix}.$$

To initialize the Kalman filter, I estimate C_1 (in fact, x_1). However, to set up P_1, I use the VMA(∞) representation of the VARFIMA(p,d,q) model. Since

$$x_{t+1} = \mu + \sum_{j=0}^{\infty} \Psi_j \varepsilon_{t-j},$$

then

$$E[(x_t - \mu)(x_{t-l} - \mu)'] = E[(\sum_{j=0}^{\infty} \Psi_j \varepsilon_{t-j})(\sum_{k=0}^{\infty} \Psi_k \varepsilon_{t-l-k})'] = \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \Psi_j E[\varepsilon_{t-j} \varepsilon'_{t-l-k}] \Psi'_k.$$

But since $\varepsilon_t \sim N(0, I)$ is i.i.d. shocks, then I only have non-zero terms when $k = j - l$

$$E[(x_t - \mu)(x_{t-l} - \mu)'] = \sum_{j=l}^{\infty} \Psi_j \Psi'_{j-l}.$$

Note that I truncate this sum at m. To compute covariance matrix for C_t, I set up an upper triangular Teoplitz matrix

$$T = \begin{pmatrix} \Psi_0 & \Psi_1 & \Psi_2 & \Psi_3 & \dots & \Psi_{m-1} \\ 0 & \Psi_0 & \Psi_1 & \Psi_2 & \dots & \Psi_{m-2} \\ 0 & 0 & \Psi_0 & \Psi_1 & \dots & \Psi_{m-3} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & \dots & \Psi_0 \end{pmatrix}.$$

Thus, to initialize P_1, I calculate the following expression:

$$P_1 = E[(C_1 - \tilde{\mu})\,(C_1 - \tilde{\mu}))'] = T\,T'.$$

A.4 Extended Kalman Filter

Following Durbin and Koopman (2012), I use the following recursions for the extended Kalman filter

$$v_t = y_t - Z(x_t), \qquad F_t = \dot{Z}_t P_t \dot{Z}_t' + \sigma_v^2\, I_N$$
$$C_{t+1} = \mu + \Phi C_t, \quad P_{t+1} = \Phi P_t (\Phi - K_t \dot{Z}_t)' + H\,H',$$

where $x_t = G\,C_t$, $K_t = \Phi P_t \dot{Z}_t' F_t^{-1}$ and $\dot{Z}_t = \frac{\partial Z}{\partial C_t}$. For updating I use the following relations

$$C_{t|t} = C_t + P_t \dot{Z}_t' F_t^{-1} v_t$$
$$P_{t|t} = P_t - P_t \dot{Z}_t' F_t^{-1} \dot{Z}_t P_t.$$

A.5 VMA(∞) representation

Brockwell and Davis (1990) show how to obtain impulse-response coefficients for an ARMA process (p 92). This result can be generalized to a multivariate setting as well. For that define $\Theta_0 = \Sigma$ and $\Theta_j = 0$ for $j > q$ as well as $\Phi_j = 0$ for $j > p$, then the impulse-response coefficients φ_j for a VARMA(p,q) process is obtained as

$$\varphi_j = \Theta_j + \sum_{0<k\leq j} \Phi_k\, \varphi_{j-k}, \quad 0 \leq j < max(p, q+1),$$

and

$$\varphi_j = \sum_{0<k\leq p} \Phi_k\, \varphi_{j-k}, \quad j \geq max(p, q+1).$$

Now, to derive the impulse-response coefficients Ψ_j for a VARFIMA(p,d,q) process, we need to multiply two lag polynomials, $\varphi(L)$ and $D(L)$. First, note that as shown by Diebold and Rudebusch (1989), a binomial expansion of the operator $(1-L)^d$ can be written as follows:

$$(1-L)^d = \sum_{j=0}^{\infty} \frac{\Gamma(j-d)L^j}{\Gamma(-d)\Gamma(j+1)} = 1 - dL + \frac{d(d-1)}{2!}L^2 - \frac{d(d-1)(d-2)}{3!}L^3 + ...$$

Since $D(L)$ is a diagonal matrix, its inverse is also diagonal. Denoting the inverse polynomial[5] as $\eta(L)$, the impulse-response coefficients Ψ for a VARFIMA(p,d,q) process can be obtained as

$$\begin{aligned}(\sum_{j=0}^{\infty}\eta_j L^j)(\sum_{k=0}^{\infty}\varphi_k L^k) &= \sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\eta_j\varphi_k L^{j+k} = \sum_{k=0}^{\infty}\sum_{j=k}^{\infty}\eta_{j-k}\varphi_k L^j \\ &= \sum_{j=0}^{\infty}\sum_{k=0}^{j}\eta_{j-k}\varphi_k L^j = \sum_{j=0}^{\infty}(\sum_{k=0}^{j}\eta_{j-k}\varphi_k)L^j\end{aligned}$$

Thus,

$$\Psi_j = \sum_{k=0}^{j}\eta_{j-k}\varphi_k.$$

A.6 VAR(∞) representation

Let us first compute the VAR(∞) representation of a VARMA(p,q) model, which is defined as

$$\Phi(L)x_t = \Theta(L)\varepsilon_t,$$

where $\Phi(L) = I - \Phi_1 L - \Phi_2 L^2 - ... - \Phi_p L^p$ and $\Theta(L) = I + \Theta_1 L + ... + \Theta_q L^q$. The VAR($\infty$) representation of the VARMA(p,q) model takes the following form

$$\varepsilon_t = \Theta(L)^{-1}\Phi(L)x_t = \Pi(L)x_t = \sum_{j=0}^{\infty}\Pi_j x_{t-j}.$$

The product of the two polynomials $\Theta(L)\Pi(L)$ are computed as

$$(\textstyle\sum_{k=0}^{q}\Theta_k L^k)(\sum_{j=0}^{\infty}\Pi_j L^j) = \sum_{k=0}^{q}\sum_{j=0}^{\infty}\Theta_k\Pi_j L^{j+k} = \sum_{k=0}^{q}\sum_{j=k}^{\infty}\Theta_k\Pi_{j-k}L^j = \\ \textstyle\sum_{j=0}^{\infty}\sum_{k=0}^{min(q,j)}\Theta_k\Pi_{j-k}L^j = \sum_{j=0}^{\infty}(\sum_{k=0}^{min(q,j)}\Theta_k\Pi_{j-k})L^j.$$

[5] Note that d will be replaced with $-d$ in the above binomial expansion

Using the relation that $\Theta(L)\Pi(L) = \Phi(L)$, we can recursively express the VAR(∞) representation coefficients as

$$-\Phi_j = \Pi_j + \sum_{k=1}^{min(q,j)} \Theta_k \Pi_{j-k},$$

where $\Phi_0 = -I$ and $\Phi_j = 0$ for $j > p$. Now, let us express the VAR(∞) representation of a VARFIMA(p,d,q) process. To find the respective coefficients, we need to multiply the two polynomials, $\Pi(L)D(L) = \Xi(L)$, where $D(L)$ is a diagonal matrix. That is, the VAR(∞) representation of the VARFIMA(p,d,q) process is defined as

$$\varepsilon_t = \Pi(L)D(L)x_t = \Xi(L)x_t = \sum_{j=0}^{\infty} \Xi_j x_{t-j}.$$

The product of the two polynomials, $\Pi(L)D(L)$, are computed as

$$\Pi(L)D(L) = (\sum_{k=0}^{\infty} \Pi_k L^k)(\sum_{j=0}^{\infty} D_j L^j) = \sum_{k=0}^{\infty}\sum_{j=0}^{\infty} \Pi_k D_j L^{j+k} =$$

$$\sum_{k=0}^{\infty}\sum_{j=k}^{\infty} \Pi_k D_{j-k} L^j = \sum_{j=0}^{\infty}\sum_{k=0}^{j} \Pi_k D_{j-k} L^j = \sum_{j=0}^{\infty}(\sum_{k=0}^{j} \Pi_k D_{j-k}) L^j.$$

Thus,

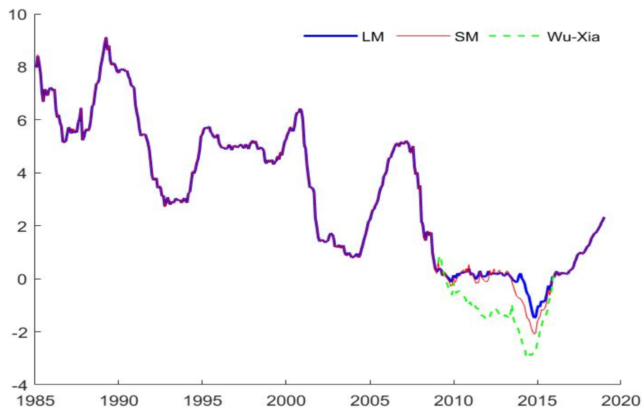
$$\Xi_j = \sum_{k=0}^{j} \Pi_k D_{j-k}.$$

A.7 Survey data

For survey data, I draw on the publicly available Survey of Professional Forecasters (SPF) database of the Fed Reserve Bank of Philadelphia. In my estimations, I employ 6- and 12-month-ahead forecasts of the 3-month T-bill rate as well as its 10-year average forecast. The 6- and 12-month-ahead forecasts are compiled on a quarterly basis and they are available since 1981. The Philadelphia Fed has started conducting these surveys in real time and adhering to a consistent timing since 1990:Q3. In the beginning of each quarter, the questionnaires are sent to the forecasters and the final deadline for reporting forecasts is the mid of the second month of the corresponding quarter. The Fed questionnaire asks to forecast for 6 quarters but only 4 quarters are relevant for forecasting ahead from the current quarter. Since these forecasts can be considered as the mid of month forecasts, I employ a piecewise linear interpolation method to obtain the end of month forecasts. For instance, to get the end of month forecast

Figure A.1: Shadow rate

This figure depicts the shadow rate estimates from long memory "LM" and short memory "SM" models along with the estimates of Wu and Xia (2016) which are obtained from the website of the Fed Reserve Bank of Atlanta.



for 6 months ahead in February I use the mean forecasts for 2 and 3 quarters ahead and interpolate these two points to get the end of month forecast for August. In the case of the long-horizon forecast (10-year-ahead average) for the 3-month T-bill rate, the survey is conducted on less frequently basis, i.e., once a year. For the long horizon forecast, I do not carry out any adjustments to the available forecast. In fact, in the case of the short horizon forecasts, my interpolations are very similar to the actual forecasts. In addition, the sample length of the long-horizon forecast (10-year-ahead average) is relatively shorter as it starts in 1992. Finally, it is worth noting that when I employ the survey forecasts in the Kalman filter, I treat unavailable dates as missing observations and adjust the filter accordingly.

A.8 Shadow rate models

The estimated results for the shadow rate models with short and long memory are presented in Table A.1 and Table A.2, respectively. The estimated coefficients of the long memory model are similar to the ones in the affine model with long memory whereas the coefficients of the short memory SRM resemble the estimates in the affine model with short memory. The coefficients determining the Q-dynamics in the short memory model are close to those of the long memory model too. All drifts of the factors in the long memory shadow rate model are statistically insignificant at the 5% significance level. Figure A1 depicts the estimated shadow rates from both models along with the estimated shadow rate of Wu and Xia (2016). When compared to the Wu and Xia (2016) estimates, the shadow rate estimates of both models are smaller and close to the zero in most of the period. This dynamics may explain the reason why affine models perform as adequate as the SRMs despite the presence of a long ZLB period in the sample.

Table A.1: Short memory SRM

This table reports the empirical results from the short memory shadow rate model. The standard errors in parentheses are calculated using the approach proposed by Harvey (1989), p 142 (see also p 938 of Andreasen, 2013). ϕ^Q represents the estimated eigenvalues. All estimated coefficients are statistically significant at the 5% level except the ones with the asterisk sign.

		Φ		ϕ^Q	μ		Σ	
latent 1	1.0034	0.0304	0.0464	0.9978	0.0005*	0.0036		
	(0.0105)	(0.0080)	(0.0047)	(0.0004)	(0.0003)	(0.0007)		
latent 2	−0.0025*	0.9725	−0.0323	0.9671	−0.0003*	−0.0020*	0.0060	
	(0.0027)	(0.0067)	(0.0061)	(0.0031)	(0.0003)	(0.0018)	(0.0014)	
latent 3	−0.0069	−0.0175	0.8910	0.8941	−0.0017*	−0.0013*	−0.0049	0.0031
	(0.0029)	(0.0031)	(0.0052)	(0.0062)	(0.0013)	(0.0017)	(0.0015)	(0.0006)
δ_0	0.1288							
	(0.0083)							
σ_v	0.0007							
	(0.0000)							

A.9 Macro-finance models

In this section, I estimate a 5-factor unspanned macro-finance term structure model (see, for instance, Joslin, Priebsch and Singleton, 2014) using both short and long memory over the same sample period. Similar to the latent factor approach in the previous sections, I treat the Q- and P-dynamics asymmetrically in the long memory framework. That is, the three pricing factors follow a VAR(1) process under the Q-measure, whereas they evolve as a VARFIMA(1,d,0) process under the P-measure together with two additional macro variables. Here, I assume that these macro variables affect physical dynamics only, i.e., they do not enter the pricing equations. In the short memory framework, the specification is standard, i.e., the pricing factors follow a VAR(1) process under both probability measures, but the same macro variables only affect the P-dynamics.

I assume that all factors are observed without any measurement errors. The two macro variables included are inflation expectations (INF) over the next one year obtained from the SPF and real economic activity indicator (GRO). In particular, GRO is the 3-month moving average of the Chicago Fed National Activity Index (CFNAI), a measure of current real sector

Table A.2: Long memory SRM

This table reports the empirical results from the long memory shadow rate model. The standard errors in parentheses are calculated using the approach proposed by Harvey (1989), p 142 (see also p 938 of Andreasen, 2013). ϕ^Q represents the estimated eigenvalues. All estimated coefficients are statistically significant at the 5% level except the ones with the asterisk sign.

	Φ			ϕ^Q	d	μ	Σ		
latent 1	0.1045	−0.1002	−0.0006*	0.9975	0.9317	−0.0005*	0.0036		
	(0.0115)	(0.0123)	(0.0112)	(0.0005)	(0.0108)	(0.0008)	(0.0007)		
latent 2	0.0815	0.4997	0.0619	0.9683	0.9317	−0.0000*	−0.0014*	0.0048	
	(0.0123)	(0.0119)	(0.0126)	(0.0037)	(0.0123)	(0.0011)	(0.0015)	(0.0012)	
latent 3	0.0995	−0.1270	0.6193	0.8800	0.4812	0.0000*	−0.0018*	−0.0037	0.0024
	(0.0117)	(0.0126)	(0.0120)	(0.0099)	(0.0122)	(0.0010)	(0.0013)	(0.0013)	(0.0005)
δ_0	0.1181								
	(0.0070)								
σ_v	0.0007								
	(0.0000)								

activity. The other three observed factors are the rotated counterparts of the latent factors determining the risk-neutral dynamics. Following Joslin et al (2011), I assume that the 0.5-, 2-, and 10-year zero-coupon yields are measured without errors.

In this setup, I can estimate the parameters governing the risk-neutral and physical dynamics separately.[6] Specifically, I estimate the 3×3 Jordan form Φ^Q and δ_0 employing the maximum likelihood approach for the 3 latent factors. For identification, the 3×1 vector μ^Q is set to zero while the 3×1 vector δ_1 to 1. I transform the estimated parameters to obtain the corresponding parameters of the rotated pricing factors $P_t = W y_t$, where W is a 3×8 weighting matrix.

In the short memory framework, I estimate the parameters of the physical dynamics by applying the OLS method as suggested by Joslin et al (2011). Contrary to the short memory model, it is not possible to estimate the parameters governing the physical dynamics in the long memory framework relying on the OLS method. This is because the physical dynamics is specified as a long memory process which requires estimation of the fractional orders of integration d for all factors. Therefore, I apply the Kalman filter to estimate the parameters of the physical dynamics in the long memory framework (see Table A3 and Table A4).

[6]Only the Σ parameter affects both dynamics. As literature shows (see, for instance, Joslin et al, 2011; Andreasen and Christensen, 2015; Abbritti et al 2016; Andreasen and Meldrum, 2019), Σ is better identified with the physical dynamics, so I include it in the estimation of the parameters of the physical dynamics only.

Table A.3: Short memory five factor affine model

This table reports the empirical results from the short memory affine model. The standard errors in parentheses are asymptotic standard errors. ϕ^Q represents the estimated eigenvalues. All estimated coefficients are statistically significant at the 5% level except the ones with the asterisk sign. The abbreviation "GRO" indicates the 3-month moving average of the Chicago Fed National Activity Index (CFNAI), the "Inf" denotes 1-year-ahead inflation expectations from the SPF, the latent factors numbered as 1-3 are the rotated counterparts of the 0.5-, 2-, and 10-year yields, respectively.

			Φ			ϕ^Q	μ
GRO	0.9360	−0.0112*	−0.1038	0.1141	−0.0135*		0.0001*
	(0.0182)	(0.0283)	(0.0400)	(0.0535)	(0.0223)		(0.0005)
Inf	0.0176	0.9940	−0.0237	0.0248*	−0.0035*		0.0002*
	(0.0053)	(0.0083)	(0.0117)	(0.0156)	(0.0065)		(0.0001)
latent 1	0.0727	0.0658	0.8048	0.2339	−0.0891	0.9999	0.0004*
	(0.0178)	(0.0276)	(0.0391)	(0.0522)	(0.0218)	(0.0000)	(0.0005)
latent 2	0.0592	0.0829	−0.0467*	1.0403	−0.0395*	0.9849	−0.0003*
	(0.0220)	(0.0341)	(0.0484)	(0.0646)	(0.0270)	(0.0005)	(0.0006)
latent 3	−0.0059*	0.0744	−0.0294*	0.0533*	0.9345	0.9009	0.0001*
	(0.0228)	(0.0353)	(0.0500)	(0.0668)	(0.0279)	(0.0040)	(0.0007)
δ_0	−0.0102*						
	(0.0036)						
σ_v	0.0011						
	(0.0000)						

Table A.4: Long memory five factor affine model

This table reports the empirical results from the long memory affine model. The standard errors in parentheses are asymptotic standard errors. ϕ^Q represents the estimated eigenvalues. All estimated coefficients are statistically significant at the 5% level except the ones with the asterisk sign. The abbreviation "GRO" indicates the 3-month moving average of the Chicago Fed National Activity Index (CFNAI), the "Inf" denotes 1-year-ahead inflation expectations from the SPF, the latent factors numbered as 1-3 are the rotated counterparts of the 0.5-, 2-, and 10-year yields, respectively.

			Φ			ϕ^Q	d	μ
GRO	0.8781	−0.0013*	−0.0671	0.0694	−0.0041		0.1735	−0.0001
	(0.0021)	(0.0047)	(0.0026)	(0.0028)	(0.0005)		(0.0034)	(0.0000)
Inf	0.0049	0.9080	0.0053	−0.0103	0.0033		0.6060	0.0000
	(0.0003)	(0.0021)	(0.0006)	(0.0007)	(0.0001)		(0.0030)	(0.0000)
latent 1	0.0752	0.0826	0.6998	0.2775	−0.0356	0.9999	0.2749	0.0005
	(0.0012)	(0.0040)	(0.0027)	(0.0026)	(0.0007)	(0.0000)	(0.0021)	(0.0000)
latent 2	0.0604	0.1147	0.0451	0.8837	0.0254	0.9851	0.2254	−0.0001
	(0.0014)	(0.0044)	(0.0029)	(0.0034)	(0.0006)	(0.0005)	(0.0017)	(0.0000)
latent 3	0.0154	0.0660	−0.0735	0.0638	0.9919	0.9006	0.0402	0.0001
	(0.0014)	(0.0049)	(0.0033)	(0.0038)	(0.0007)	(0.0040)	(0.0016)	(0.0000)
δ_0	0.0155							
	(0.0018)							
σ_v	0.0011							
	(0.0000)							

A.9.1 *Conditional projections*

Conditional projections (forecasts) are frequently used at policy making institutions such as central banks, IMF, etc. Most of the time, forecasts made at these institutions are conditioned on future paths of certain variables of interest such as policy rate and commodity prices. I call such forecasts conditional forecasts or conditional projections to distinguish them from the forecasts whose all future paths are obtained without conditioning on paths of any model variables. Conditional forecasting exercise is also employed in the forecasting literature for model comparison and validation although it is not a popular tool in the no-arbitrage DTSM literature. Here, I show that conditional projections can be effectively utilized for model comparison and validation in the DTSM literature as well.

As mentioned before, in conditional projections we provide future paths for some variables of interest and generate forecasts based on these conditioned paths. If conditioned variables are endogenous, then we need to take into account possible feedbacks originating from the other model variables. If we do not take these feedbacks into consideration, then our forecasts will not be model consistent. Therefore, to generate model consistent conditional forecasts, I apply the approach of Banbura, Giannone and Lenza (2015). This method, which is based on the Kalman filter and smoothing, is easy to implement. I estimate the models for the whole sample (January, 1985 - December, 2018) and then use the parameters from these estimations in the conditional projections covering the period January, 1990 - December, 2018, where the first five years are used as initial conditions.

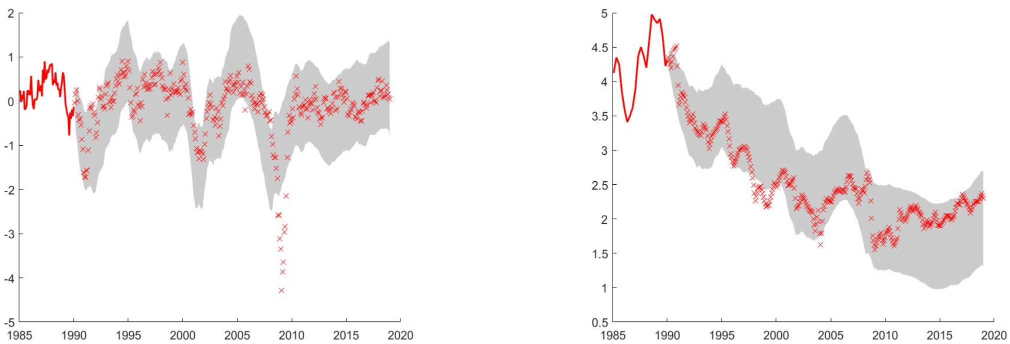
I use such exercises to test whether the term structure of interest rates possesses sufficient information regarding the dynamics of the inflation and real economic activity by conditioning on the actual paths of the non-macro factors. Second, I will also use such projections to test whether conditioning on the actual paths of the macro factors is sufficient to capture the salient dynamics of the bond yields after the year 1989.

In the first exercise, the forecasts for the real activity and inflation are computed by conditioning on the actual paths of the non-macro factors. Figure A2 shows the conditional forecasts for the short memory model. In most of the forecast period, the actual real activity and inflation data fall within the 95% confidence band. Similarly, Figure A3 clearly demonstrates that the long memory model forecasts are broadly in line with the actual data. In this exercise, the long memory model slightly outperforms the short memory model in forecasting the actual paths of the macro variables. It is worth noting that while I use an unspanned

Figure A.2: Conditional projections with short memory

This figure depicts the conditional projections for the inflation and real activity measure. These forecasts are computed conditioning on the actual paths of the three non-macro factors. The shaded area indicates the 95% confidence interval, whereas the actual data is shown with "x" signs. The forecast covers the period January, 1990 - December, 2018. The first five years are used as initial conditions.

(a) GRO **(b)** Inflation



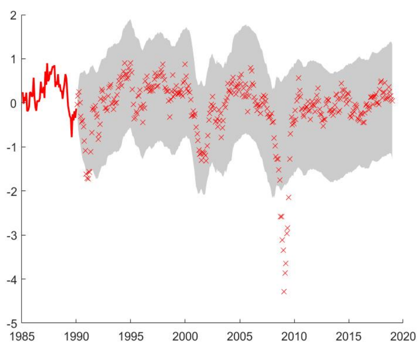
specification, the term structure of interest rates seems to possess sufficient information regarding the future paths of the macro variables over most of the forecast period.

In the second forecasting exercise, I condition on the actual paths of inflation and real economic activity and investigate the model forecasts for the term structure of the yields. The computed conditional forecasts for the 3-month and 10-year bond yields are displayed in Figure A4. Both models forecast that the 3-month yield should be lifted up after the year 2011. Thus, it seems that the factors other than the inflation and real activity index played an important role in keeping the short term interest rates close to the zero level. For the 10-year yield, both models effectively capture the secular decline in the long rate over the last two decades.

Figure A.3: Conditional projections with long memory

This figure depicts the conditional projections for the inflation and real activity measure. These forecasts are computed conditioning on the actual paths of the three non-macro factors. The shaded area indicates the 95% confidence interval, whereas the actual data is shown with "x" signs. The forecast covers the period January, 1990 - December, 2018. The first five years are used as initial conditions.

(a) GRO



(b) Inflation

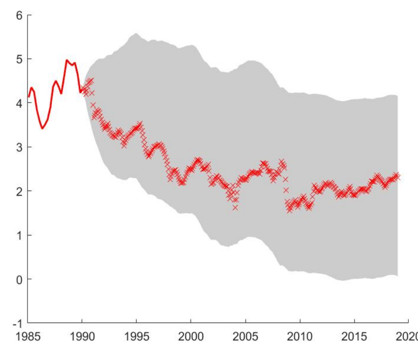
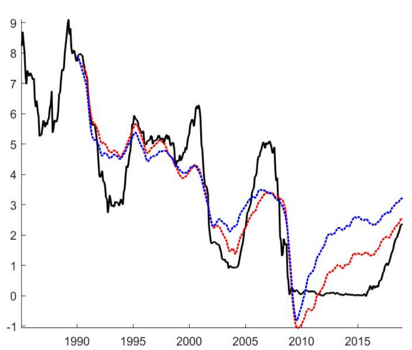


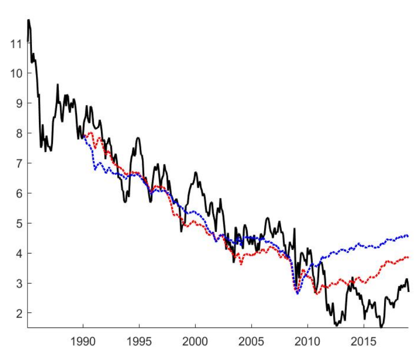
Figure A.4: Conditional projections

This figure depicts the conditional projections for the 3 month and 10 year yield. These forecasts are computed conditioning on the actual paths of the two macro factors. The black solid line shows the actual data. The forecasts of the long memory model are shown with the blue dotted line, whereas the forecasts of the short memory are given with the red dotted line. The forecast covers the period January, 1990 - December, 2018. The first five years are used as initial conditions.

(a) GRO



(b) Inflation

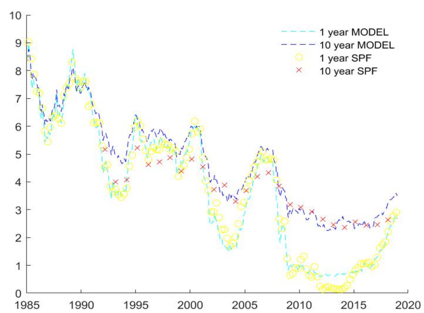


A.10 Matching survey expectations

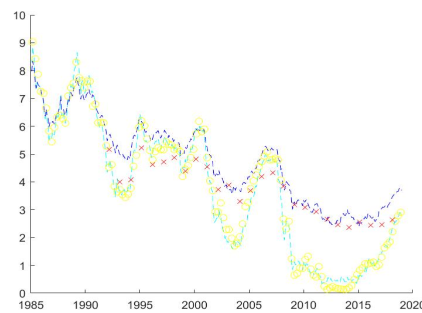
Figure A.5: Matching the short-rate survey expectations

This figure displays the ability of the models in matching the 3-month short-rate expectations from the Survey of Professional Forecasters (SPF) using the yield data augmented with the short horizon survey expectations. The 12-month-ahead short-rate expectations from the SPF is denoted by "1-year SPF" whereas the 10-year-ahead average short-rate expectations as "10-year SPF". The label "1-year MODEL" shows the 1-year-ahead short-rate forecasts from the respective model and similarly, the label "10-year MODEL". The abbreviation "SM" denotes short memory, whereas the "LM" indicates long memory.

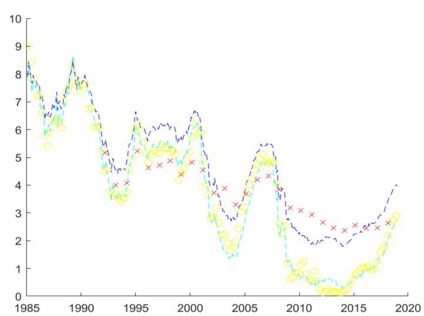
(a) Affine SM



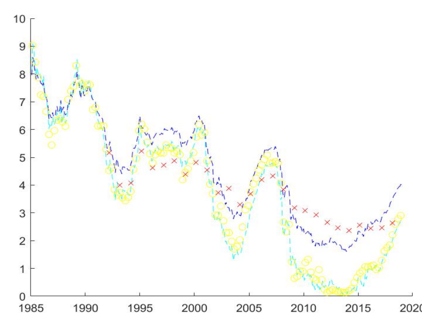
(b) SRM SM



(c) Affine LM



(d) SRM LM



Research Papers 2021

CREATES
Center for Research in Econometric Analysis of Time Series

2020-19:	Eric Hillebrand, Jakob Mikkelsen, Lars Spreng and Giovanni Urga: Exchange Rates and Macroeconomic Fundamentals: Evidence of Instabilities from Time-Varying Factor Loadings
2021-01:	Martin M. Andreasen: The New Keynesian Model and Bond Yields
2021-02:	Daniel Borup, David E. Rapach and Erik Christian Montes Schütte: Now- and Backcasting Initial Claims with High-Dimensional Daily Internet Search-Volume Data
2021-03:	Kim Christensen, Mathias Siggaard and Bezirgen Veliyev: A machine learning approach to volatility forecasting
2021-04:	Fabrizio Iacone, Morten Ørregaard Nielsen and Robert Taylor: Semiparametric Tests for the Order of Integration in the Possible Presence of Level Breaks
2021-05:	Stefano Grassi and Francesco Violante: Asset Pricing Using Block-Cholesky GARCH and Time-Varying Betas
2021-06:	Gloria González-Rivera, Carlos Vladimir Rodríguez-Caballero and Esther Ruiz Ortega: Expecting the unexpected: economic growth under stress
2021-07:	Matei Demetrescu and Robinson Kruse-Becher: Is U.S. real output growth really non-normal? Testing distributional assumptions in time-varying location-scale models
2021-08:	Luisa Corrado, Stefano Grassi and Aldo Paolillo: Modelling and Estimating Large Macroeconomic Shocks During the Pandemic
2021-09:	Leopoldo Catania, Alessandra Luati and Pierluigi Vallarino: Economic vulnerability is state dependent
2021-10:	Søren Johansen and Anders Ryghn Swensen: Adjustment coefficients and exact rational expectations in cointegrated vector autoregressive models
2021-11:	Bent Jesper Christensen, Mads Markvart Kjær and Bezirgen Veliyev: The incremental information in the yield curve about future interest rate risk
2021-12:	Mikkel Bennedsen, Asger Lunde, Neil Shephard and Almut E. D. Veraart: Inference and forecasting for continuous-time integer-valued trawl processes and their use in financial economics
2021-13:	Anthony D. Hall, Annastiina Silvennoinen and Timo Teräsvirta: Four Australian Banks and the Multivariate Time-Varying Smooth Transition Correlation GARCH model
2021-14:	Ulrich Hounyo and Kajal Lahiri: Estimating the Variance of a Combined Forecast: Bootstrap-Based Approach
2021-15:	Salman Huseynov: Long and short memory in dynamic term structure models