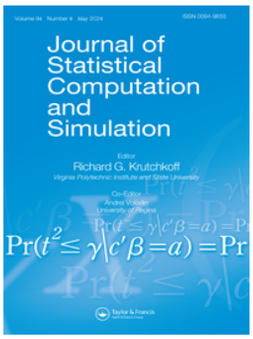


Taylor & Francis
Taylor & Francis Group

Journal of Statistical Computation and Simulation

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/gscs20

Bayesian inference for long memory term structure models

Fernanda Valente & Márcio Laurini

To cite this article: Fernanda Valente & Márcio Laurini (2024) Bayesian inference for long memory term structure models, Journal of Statistical Computation and Simulation, 94:8, 1735-1759, DOI: 10.1080/00949655.2023.2299938

To link to this article: https://doi.org/10.1080/00949655.2023.2299938

JOURNAL OF STATISTICAL COMPUTATION AND SIMULATION
2024, VOL. 94, NO. 8, 1735–1759
https://doi.org/10.1080/00949655.2023.2299938

Bayesian inference for long memory term structure models

Fernanda Valente and Márcio Laurini

Department of Economics, FEARP – University of São Paulo, São Paulo, Brazil

ABSTRACT

In this study, we propose a novel adaptation of the Dynamic Nelson–Siegel term structure model, incorporating long memory properties to enhance its forecasting accuracy. Our approach involves modelling the evolution of latent factors using fractional Gaussian noise processes, approximated by a weighted sum of independent first-order autoregressive components. The resulting formulation allows for a Gaussian Markov Random Field representation, facilitating the application of computationally efficient Bayesian techniques through Integrated Nested Laplace Approximations. Extensive simulation and empirical analysis demonstrate that integrating long memory significantly improves the model's forecasting performance, particularly for longer time horizons. By shedding light on the potential benefits of incorporating long memory concepts into traditional term structure models, our research highlights its utility in capturing intricate temporal dependencies and enhancing prediction precision.

ARTICLE HISTORY
Received 22 July 2023
Accepted 21 December 2023

KEYWORDS
Term structure; long memory; Bayesian forecasting; Laplace approximations

1. Introduction

The term structure of interest rates represents the relationship between the interest rates and time to maturity in fixed income financial instruments and is a fundamental object in the asset pricing process. Methods for modelling and forecasting the term structure of interest rates are a central theme in theoretical and empirical research in finance, and one of the main modelling instruments used in practical aspects in financial markets [1,2].

Modeling interest curves is a complex process with several challenges. One of the main difficulties is that the term structure is a high-dimensional object that is not directly observable, and in some cases, it can even be an infinite-dimensional object [3,p. 93–103]. As noted by De Pooter et al. [4], long-term yields are risk-adjusted averages of expected future short-term rates, which means that yields of different maturities are interconnected and move together over time. However, shocks in the economy can have different effects on long and short maturities, and monetary policy authorities actively target the short end of the yield curve to achieve their macroeconomic objectives. These factors contribute to the complexity of interest rate movements.

CONTACT Márcio Laurini laurini@fearp.usp.br Department of Economics, University of São Paulo, School of Economics, Business Administration and Accounting at Ribeirão Preto (FEA-RP/USP), Av. dos Bandeirantes 3900, 14040-905, Ribeirão Preto, SP, Brazil

Supplemental data for this article can be accessed online at http://dx.doi.org/10.1080/00949655.2023.2299938.

Understanding the dynamic evolution of the yield curve and developing accurate forecasts is essential for various financial tasks, such as bond portfolio management, derivatives pricing, and risk management.

Hence, it is not surprising the meaningful number of articles dedicated to model and forecast the term structure of interest rates. As discussed by Carriero et al. [5], the existing methods for producing forecasts of the term structure of the interest rates can be practically divided in three groups: models based on forward rate regressions (e.g. Fama and Bliss [6] and Cochrane and Piazzesi [7]), models based on the no-arbitrage paradigm (e.g. Duffee [8]) and a class of models that uses the Nelson–Siegel Nelson and Siegel [9] exponential components framework (e.g. Diebold and Li [10] and Christensen et al. [11]).

The Nelson–Siegel class, introduced by Nelson and Siegel [9], is a widely used approach for fitting the term structure of interest rates. The method involves adopting a flexible, smooth parametric function to model the yield curve. This model has become popular in the finance literature due to its desirable characteristics, including parsimony and flexibility, which promote smoothness and enable the yield curve to assume various shapes at different times [2]. Furthermore, it is noteworthy that the forward rate curve can be expressed as a constant plus a Laguerre function, which has important mathematical implications for the modelling of interest rates.

Taking a step further, Diebold and Li [10] proposed the Dynamic Nelson–Siegel (DNS) model as an extension of the original Nelson–Siegel approach. The DNS model incorporates time-varying parameters and introduces dynamics to the model. According to Diebold and Rudebusch [2], the DNS model decomposes the yield curve into three dynamic, latent factors, corresponding to the level, slope, and curvature of the interest rates. The evolution of these latent factors can be modelled as either an autoregressive (AR) process for each individual latent factor or as a vector autoregressive (VAR) model for all three factors simultaneously.

However, as demonstrated by Diebold and Li [10], AR forecasts tend to outperform many alternatives, including the VAR specification. The success of the DNS model has led to the development of several other dynamic approaches to Nelson–Siegel that have demonstrated promising results in terms of forecasting accuracy. For example, Koopman et al. [12] proposed the introduction of time-varying loading and variance parameters, De Pooter [13] added a second slope factor, and Diebold et al. [14] extended the DNS model to a global context, modelling a set of country yield curves for both global and country-specific factors.

However, it should be noted that many macroeconomic and financial time series, such as nominal and real interest rates, real exchange rates, interest rate differentials, and volatility measures, exhibit high persistence, as highlighted by Zivot and Wang [15,p. 271–272]. The study of long memory properties of time series in finance and macroeconomics is widespread, with several researchers such as Ding et al. [16], Breidt et al. [17], Hassler and Wolters [18], and Baillie et al. [19] investigating this phenomenon. The existence of long memory in the context of term structure was first recognized by Backus and Zin [20], while Tsay [21] found evidence of long memory properties in monthly and quarterly U.S. post-war data. Moreover, Couchman et al. [22] analysed the long memory properties of interest rates for 16 countries and found that most of the examined countries have long memory parameters between zero and one. More recently, Goliński and Zaffaroni [23] developed an affine term structure model that incorporates long memory.

The AR specification proposed by Diebold and Li [10] performs well compared to alternative models, but it often fails to capture the high degree of persistence present in yields' time series. Long memory dependency structures are one way to model persistence. These processes exhibit autocorrelation functions that decay slowly at a hyperbolic rate, meaning that the autocorrelations are not absolutely summable [24,p. 40–43]. To model long-range dependencies, fractional Brownian motion (fBm) is particularly useful. fBm is a self-similar, continuous-time Gaussian process with stationary increments. Its increment process is characterized by the Hurst exponent H, that gives long memory when $1/2 < H < 1$ [25]. Another parameterization for long memory processes is through models of the ARFIMA class. This class uses fractional differentiation processes to produce long memory processes [15,24,26].

There exists a close relationship between fGn and ARFIMA models. Specifically, an ARFIMA (0,d,0) model, where the relationship between the Hurst H coefficient used in the fGn representation is given by $d = H-0.5$, displays the same hyperbolic decay pattern as an fGn process, as pointed out in previous studies [25,27]. While both models can be used to represent long memory processes, the fGn process has some advantages over the ARFIMA representation. For example, the fGn process can be obtained through a fractional derivative of a continuous Brownian motion, making it a discretization of a continuous time process, and is more analytically treatable [28], whereas an ARFIMA process only exists in discrete time [27]. Additionally, many asymptotic relationships of fGn processes apply to finite samples, as discussed in Taqqu et al. [29], making the fGn process a powerful tool for analysing long memory properties in finance and economics.

As the long memory property of interest rates is well-documented, we aim to enhance the DNS latent factor model to account for this persistence. To achieve this, we introduce a modification that incorporates the long memory property using an approximation of a fractional Gaussian noise process, as suggested by Sørbye et al. [25]. Our proposed model, DNS-fGn, uses fractional Gaussian noise processes to represent the latent factors, with different persistence parameters denoted by H. To estimate the model's parameters, latent factors, and predictions, we adopt a Bayesian inference approach.

The main advantage of the approximate fGn model is that it can be represented as a Gaussian Markov random field (GMRF), which enables us to estimate the latent factors and the persistence parameter H using the Integrated Nested Laplace Approximations (INLA) method developed by Rue et al. [30]. According to Laurini and Hotta [31], the INLA method provides accurate analytical approximations for the posterior distribution of the latent factors parameters in generalized Gaussian models, without the need for numerical simulation methods such as Markov Chain Monte Carlo (MCMC). This work can be viewed as a generalization of the Bayesian estimation methodology using Laplace Approximations to the Nelson–Siegel dynamic model proposed by Laurini and Hotta [31] in the context of long memory processes.

In this study, we evaluate the performance of our proposed DNS-fGn model in both simulated and real data scenarios. First, we conduct a Monte Carlo experiment to compare the INLA-based parameter estimates of the DNS-fGn model with those obtained through MCMC simulation. Our results demonstrate that the INLA-based estimation approach achieves similar performance to MCMC methods but with much higher computational efficiency.

Next, we analyse the properties of the DNS-fGn model using the Fama-Bliss US Treasury zero-coupon yields dataset from January 1985 until December 2000, which is also used in previous literature. We compare our proposed model with alternative factor specifications, including first-order autoregressive and autoregressive fractionally integrated moving average models. We adopt a rolling estimation approach to obtain out-of-sample forecasts for horizons of 1-month, 6-months, and 12-months, and evaluate the predictive accuracy of these models. Our findings suggest that the DNS-fGn model outperforms the alternative models, particularly for longer forecasting horizons of 12 months ahead.

To further understand this result and the possible relationship between parameter changes and long memory on the interest curve, we conduct a Monte Carlo study using a model with a Markov regime switching as the data generating process. Our experiment demonstrates that the presence of a regime change process induces better predictive performance of the DNS-fGn model in comparison to the model based on first-order autoregressive processes.

Overall, our study provides evidence that incorporating long memory processes using fGn models can improve the forecasting performance of the DNS model in the term structure of interest rates.

This article is organized as follow. Section 2 contains a description of the DNS and DNS-fGn models. Section 3 briefly summarizes the INLA method. Section 4 presents the empirical analysis. The Monte Carlo experiment with the Markov switching model is presented in Section 5, and Section 6 concludes. The Supplementary Material presents the results of the Monte Carlo experiment comparing the INLA and MCMC parameters estimates of the DNS-fGn model, and the implementation of the basic DNS-fGn model using r-inla package.

2. Dynamic Nelson–Siegel model

The Nelson–Siegel dynamic model is a powerful tool used in finance to understand the term structure of interest rates. It is a dynamic factor structure where the observed yield vector is represented as a polynomial function of three common factors. The factor loadings are determined by Laguerre polynomials, which have excellent mathematical approximation properties within the domain of nominal interest rates, i.e. the interval $[0, \infty)$. The Laguerre polynomials also allow for the imposition of no-arbitrage conditions on this class of models, as discussed in Christensen et al. [11] and Diebold and Rudebusch [2]. Through the construction of an affine structure in the Duffie-Kan sense [32], this model provides a powerful framework for understanding the term structure of interest rates in a way that is both theoretically sound and empirically relevant.

This structure is interesting for modelling interest rates due to the exponential format implied by this curve, which has properties useful in modelling discount curves, as discussed in Diebold and Li [10] and Diebold and Rudebusch [2]. Following Diebold et al. [33], the DNS model can be represented by a state-space representation. The model can be formulated through a measurement equation for the observed yield curve, written as

$$y_t(\tau) = \beta_{1t} + \beta_{2t}\left(\frac{1-\mathrm{e}^{-\tau\lambda}}{\tau\lambda}\right) + \beta_{3t}\left(\frac{1-\mathrm{e}^{\tau\lambda}}{\tau\lambda} - \mathrm{e}^{-\tau\lambda}\right) + \epsilon_t(\tau) \tag{1}$$

and the transition equations, using a parameterization in deviations from the mean, for the latent factor dynamics:

$$\begin{pmatrix} \beta_{1t} - \mu_1 \\ \beta_{2t} - \mu_2 \\ \beta_{3t} - \mu_3 \end{pmatrix} = \begin{pmatrix} \phi_1 & 0 & 0 \\ 0 & \phi_2 & 0 \\ 0 & 0 & \phi_3 \end{pmatrix} \begin{pmatrix} \beta_{1t-1} - \mu_1 \\ \beta_{2t-1} - \mu_2 \\ \beta_{3t-1} - \mu_3 \end{pmatrix} + \begin{pmatrix} \epsilon_{\beta_{1t}} \\ \epsilon_{\beta_{1t}} \\ \epsilon_{\beta_{3t}} \end{pmatrix} \tag{2}$$

where $y_t(\tau)$ represents the yield vector in time t, β_{it}, $i = 1, 2, 3$, are the latent factors with interpretations of level, slope and curvature, with means μ_1, μ_2 and $\mu 3$, respectively, and λ is the decay parameter. Finally, $\epsilon_t(\tau)$ and $\epsilon_{\beta_i t}$ are innovations process of the measurement and transition equations, respectively.

Under the assumption that the shocks $\epsilon_t(\tau)$ and ϵ_{β_t} are independent Gaussian process and the decay parameter λ is constant and known, the state space representation becomes Gaussian and linear, allowing a simple estimation in two-steps, as proposed by Diebold and Li [10]. In this method, in the first stage, with λ held fixed, it is possible to estimate the level, slope and curvature parameters through a linear regression for each yield curve in period t and, in the second stage, model the dynamic of the latent factors through a first-order autoregressive model to the time series of the factors obtained in the first stage.

Alternatively, one can assume that the decay parameter λ is constant but unknown. To obtain inferences about the parameters and the latent factors, maximum likelihood estimation via Kalman filter can be adopted, as used, for example, in Diebold et al. [33]. The use of Bayesian methods to estimate DNS models was discussed in Laurini and Hotta [34] and Hautsch and Yang [35], using Markov Chain Monte Carlo methods; the use of Hamiltonian Monte Carlo for the DNS model was proposed in Batista and Laurini [36], and the use of Integrated Nested Laplace Approximations (INLA) in Laurini and Hotta [31]. We extend the INLA estimation for the case of the DNS model with dynamics given by fractional Gaussian noise in the following sections.

2.1. Long memory dynamics in DNS framework

2.1.1. Autoregressive fractionally integrated moving average (ARFIMA) model

The AR model used to determine the evolution of the latent factors may not effectively capture the high persistence of interest rates. Instead, a more suitable approach is to model the evolution of the latent factors as a stationary process with long memory. According to Beran [37,p. 41–59], a weakly stationary process presents long memory if its autocovariance function $\gamma(k)$ for distant lags k obeys

$$\gamma(k) \sim C_1 k^{2(H-1)-1}, \quad \text{as } k \to \infty,$$

for $C_1 > 0$, with the Hurst parameter $1/2 < H < 1$. The same property can be defined in the frequency domain if the spectral density $f(\omega)$ for frequencies $\omega \in [-\pi, \pi]$ near to zero satisfy

$$f(\omega) \sim C_2 |\omega|^{-2H+1}, \quad \text{as } \omega \to 0,$$

for $C_2 > 0$ and $1/2 < H < 1$.

One approach to modelling the DNS latent factors is through an autoregressive fractionally integrated moving average (ARFIMA) model. As described by Palma [24,p. 43–49], an

ARFIMA $\{y_t\}$ process can be defined by

$$\phi(L)y_t = \theta(L)(1-L)^{-d}\epsilon_t, \tag{3}$$

where L is a lag operator and $\phi(L)$ and $\theta(L)$ are the autoregressive and moving average operators, respectively. Plus, $\phi(L)$ and $\theta(L)$ have no common roots, $(1-L^{-d}$ is a fractional differencing operator defined by the binomial expansion

$$(1-L)^{-d} = \sum_{j=0}^{\infty} \eta_j L^j = \eta(L) \tag{4}$$

where

$$\eta_j = \frac{\Gamma(j+d)}{\Gamma(j+1)\Gamma(d)}, \tag{5}$$

for $d < 1/2$, $d \neq 0, -1, -2, \ldots$, and $\{\epsilon_t\}$ is a white noise process with finite variance and $\Gamma(.)$ is the Gamma function [38,p. 255–258].

To estimate a DNS model with ARFIMA dynamics, the second step of the original method proposed by Diebold and Li [10] needs to be modified. It's important to note that long memory processes are not Markovian, making all state space representations infinite dimensional, which can complicate estimation. The use of ARFIMA specifications to incorporate long memory structures in interest curve models has its origins in Backus and Zin [20]. Comte and Renault [39] also proposed using long memory models in continuous time. Another formulation of a continuous-time long memory model for modelling one-factor interest rates models using a fractional Brownian motion structure can be found in Laurini and Hotta [40]. More recent applications of long memory models in interest rate modelling can be found in Osterrieder [41] and Abbritti et al. [42]. Additionally, Goliński and Zaffaroni [23] developed an affine term structure model based on ARFIMA formulations.

2.1.2. Fractional Gaussian noise

Another approach to introduce long persistence is by using fractional Gaussian noise (fGn). While ARFIMA and fGn are conceptually different, they are related processes, particularly when the autoregressive and moving average orders of ARFIMA are zero. Using fGn in state space modelling is particularly interesting because it's possible to obtain a Markovian representation of this process, enabling one-step estimation of the DNS model with long memory. According to Sørbye et al. [25], it's possible to approximate an fGn model by an aggregated model of a few AR(1) components and take advantage of this approach.

The fractional Gaussian noise is defined by its autocorrelation function

$$\gamma(h) = \frac{1}{2}(|h+1|^{2H} - 2|h|^{2H} + |h-1|^{2H}), \quad h = 0, \ldots, n-1, \tag{6}$$

where $H \in (0,1)$ is the Hurst exponent (or self-similarity parameter) and τ denotes the marginal precision parameter. Note that the fGn reduces to uncorrelated white noise when $H = 0.5$. If $H > 0.5$ the process has positive correlation, and, similarly, if $H < 0.5$, the

autocorrelation is negative. In this work we allow each latent factor to have a distinct H parameter.

Another representation of the fGn process can be made by the definition of the fGn as the difference of a fractional Brownian Motion (e.g. Shi et al. [43]). Using the definition of a fractional Brownian motion [44]:

$$B_{H_t} = \frac{1}{\Gamma(H+0.5)} \int_{-\infty}^{t} (t-s)^{H-0.5} - (-s)^{H-0.5} \, \mathrm{d}B(s) + \int_{0}^{t} (t-s)^{H-0.5} \, \mathrm{d}B(s) \quad (7)$$

B_{H_t} is the fractional Brownian motion process with Hurst parameter H, $dB(s)$ is the standard Brownian motion, and $\Gamma()$ is the gamma function. Using y_t as the trajectory of the process, the fGn can be defined as:

$$y_t = \sigma \left(B_{H_t} - B_{H_{1t-1}} \right) \quad (8)$$

with σ being the standard deviation of the process.

Our formulation for a dynamic Nelson–Siegel model with fGn dynamics for the latent factors is given by a substitution of AR dynamics for the fGn processes:

$$y_t(\tau) = \beta_{1t} + \beta_{2t} \left(\frac{1 - \mathrm{e}^{-\tau\lambda}}{\tau\lambda} \right) + \beta_{3t} \left(\frac{1 - \mathrm{e}^{\tau\lambda}}{\tau\lambda} - \mathrm{e}^{-\tau\lambda} \right) + \epsilon_t(\tau) \quad (9)$$

$$\begin{pmatrix} \beta_{1t} - \mu_1 \\ \beta_{2t} - \mu_2 \\ \beta_{3t} - \mu_3 \end{pmatrix} = \begin{pmatrix} fGn(H_1, \tau_1) - \mu_1 \\ fGn(H_2, \tau_2) - \mu_2 \\ fGn(H_2, \tau_3) - \mu_3 \end{pmatrix} \quad (10)$$

Using the definition of the fGn as the diference of the fractional Brownian motion we also can define the latent factors as

$$\begin{pmatrix} \beta_{1t} - \mu_1 \\ \beta_{2t} - \mu_2 \\ \beta_{3t} - \mu_3 \end{pmatrix} = \begin{pmatrix} \sigma_1 \left(B_{H_{1t}} - B_{H_{1t-1}} \right) - \mu_1 \\ \sigma_2 \left(B_{H_{2t}} - B_{H_{2t-1}} \right) - \mu_2 \\ \sigma_3 \left(B_{H_{3t}} - B_{H_{3t-1}} \right) - \mu_3 \end{pmatrix} \quad (11)$$

where $\sigma_i = 1/\tau_i^2$ using τ_i as the precision parameter of each latent factor.

The derivation of no-arbitrage conditions for the dynamic Nelson–Siegel model with long memory, as employed in our study, remains a challenge. No-arbitrage conditions for the DNS model with AR dynamics are obtained in Christensen et al. [11], but the assumptions used in this work are not valid for latent factors given by fractional Gaussian noise processes. Ohashi [45] established no-arbitrage and consistency results for the Heath-Jarrow-Morton term structure models under the assumption of dynamics driven by Fractional Brownian Motion. However, Ohashi [45] methodology differs significantly from that employed by Christensen et al. [11]. Consequently, the direct adaptation of the techniques used to construct the no-arbitrage framework in our formulation of the long-memory DNS model remains an unresolved challenge.

To develop computationally efficient estimation methods for this representation, we suggest combining the approximate representation of fGn processes, as presented in Sørbye et al. [25], with Bayesian estimation using Integrated Nested Laplace Approximations (INLA) [30]. This approach generalizes the use of INLA for the estimation of DNS proposed by Laurini and Hotta [31]. We will now describe these components in more detail.

2.2. Gaussian Markov random field approximation of fGn process

Long-memory processes such as the fGn process pose a computational challenge due to their complex dependency structure, which does not follow a Markovian structure, and the fact that process innovations are not independent. Unlike an autoregressive process where the conditional distribution of the process depends only on the most recent past, fGn processes do not have a Markovian representation.

As discussed in Rue and Held [46], a Markovian structure is associated with sparse and block-diagonal covariance/precision matrices that allow computationally efficient representations of dependent stochastic processes. Under certain conditions, the numerical evaluation of Markovian processes is associated with Gaussian state space representations. However, outside of this context, the evaluation of the conditional distribution of dynamic processes requires the use of numerically intensive approximations and truncations, as discussed again in Rue and Held [46].

The class of Gaussian Markov Random Field (GMRF) processes defines the class of stochastic processes that admit a Markovian representation with a structure of innovations (random effects) given by Gaussian distributions. If a process can be represented as a GMRF, there are efficient methods for evaluating its conditional distribution. For example, the models that admit a linear Gaussian state space representation are an example of GMRF, where the likelihood function evaluation can be performed using the Kalman filter.

The main characteristic of a process that can be represented as a GMRF is that, conditional on the nearest neighbour, the process is conditionally independent of all other process observations. This simplifies the computational representation and numerical evaluation of the process.

To define Gaussian Markov Random Fields, we use a representation of undirected graphs to operationalize the definition of the neighbourhood. An undirected graph is a tuple $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$ where $\mathcal{V}$ are the nodes and $\mathcal{E}$ the set of vertices $\{i, j\}$ of the graph, with i, $j \in \mathcal{V}$. The set of neighbours of a node i is given by all elements in $\mathcal{G}$ having a vertex for node i.

Defining $\mathbf{x} = (x_1, x_2, \ldots, x_k)^\top$ as a Gaussian vector with $E(\mathbf{x}) = \mu$ and covariance matrix Σ, we can define a graph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$, with $\mathcal{V} = \{1, 2, \ldots, n\}$ and $\mathcal{E}$ such that there is no vertex between nodes i and j if and only $x_i \perp x_j \mid x_{-ij}$, with x_{-ij} a notation for the elements of x_{ij} except the elements i, j. In this case the vector x defines a GMRF with respect to $\mathcal{G}$. Let $\mathbf{x}$ be a random vector with $E(\mathbf{x}) = \boldsymbol{\mu}$ and precision matrix $\boldsymbol{Q} > 0$, where $\boldsymbol{Q}$ the inverse of the covariance matrix $\boldsymbol{\Sigma}$. So each $i \neq j \iff \boldsymbol{Q}_{ij} = 0$. This result defines the conditionally independent elements of the vector of interest. With that we can give the precise definition of a GMRF.

A random vector $\boldsymbol{x} = (x_1, x_2, \ldots, x_k)^\top \in \mathbb{R}^n$ is called a Gaussian Markov Random Field with respect to the graph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$ with mean $E(\boldsymbol{x}) = \boldsymbol{\mu}$ e precision matrix $\boldsymbol{Q}$ if and only if its density is given by:

$$\pi(\boldsymbol{x}) = (2\pi)^{-n/2} |\boldsymbol{Q}|^{1/2} \exp\left(-\frac{1}{2} (\boldsymbol{x} - \boldsymbol{\mu})^\top \boldsymbol{Q} (\boldsymbol{x} - \boldsymbol{u}) \right.$$

In their work, Sørbye et al. [25] suggested an approximation method for fGn processes by utilizing a weighted sum of independent first-order autoregressive processes, which can be represented as a Gaussian Markov Random Field. This approach aims to estimate the

weights and coefficients of the approximation in a way that can mimic the autocorrelation function (6) of an fGn. Specifically, the authors proposed to use an optimization algorithm to find the optimal values for the weights and coefficients, such that the autocorrelation function of the approximation matches the target fGn autocorrelation function. This way, the fGn process can be approximated by the GMRF with a sparse and block-diagonal precision matrix, which facilitates the computational representation and numerical evaluation of the process.

Consider m independent AR(1) process

$$z_{j,t} = \phi_j z_{j,t-1} + \nu_{j,t}, \quad j = 1, \ldots, m, \ t = 1, \ldots, n, \tag{12}$$

where $0 < \phi_j < 1$ denotes the first-order autoregressive parameter of the jth process. In addition, let $\nu_{j,t}$ be independent zero-mean Gaussian shocks with variance $\sigma^2_{\nu,j} = 1 - \phi^2_j$. Define the cross-sectional aggregation of the m processes as

$$\bar{x}_m = \sigma \sum_{j=1}^{m} \sqrt{w_j} z^{(j)}, \tag{13}$$

where $z^{(j)} = (z_{j,1}, \ldots, z_{j,n})^T$ and the weights w_j sum to one. This implies that $var(\bar{x}_m) = \sigma^2$. The autocorrelation function of (13) is given by

$$\gamma_{\bar{x}_m}(h) = \sum_{j=1}^{m} w_j \phi_j^{|h|}, \quad h = 0, 1, \ldots, n-1. \tag{14}$$

The idea proposed by Sørbye et al. [25] is to fit the weights $\boldsymbol{w} = \{w_j\}_{j=1}^m$ and the autocorrelation coefficients $\boldsymbol{\phi} = \{\phi_j\}_{j=1}^m$ in (13) to match the autocorrelation function of fGn process as in (6). To obtain values of $(\boldsymbol{w}, \boldsymbol{\phi})$, the weighted squared error is minimized

$$(\boldsymbol{w}, \boldsymbol{\phi})_H = \underset{(\boldsymbol{w},\boldsymbol{\phi})}{\operatorname{argmin}} \sum_{h=1}^{h_{\max}} \frac{1}{h} (\gamma_{\bar{x}_m}(h) - \gamma_x(h))^2, \tag{15}$$

where $h_{\max}$ represents an arbitrary upper limit to the number of lags included.[1] The key point of this approximation of the fGn process is the possibility to represent the DNS model with long memory as a latent GMRF, which allow us to estimate it using the INLA method.

3. Integrated Nested Laplace approximations

The INLA method [30] provides accurate and efficient approximations on Bayesian hierarchical models that can be represented as a latent Gaussian model, focuses on Gaussian Markov random field, which is defined as an N-dimensional Gaussian vector with mean μ and precision matrix Q with Markov properties. The GMRF has been widely applied for Bayesian hierarchical models, which is characterized by stages, where the first one typically defines a distributional assumption for the observed variable y, usually assumed to be conditionally independent given the latent factors x and some additional parameter θ, in the

form

$$\pi(y \mid x, \theta) = \prod_j \pi(y_j \mid x_j, \theta), \quad j \in J \tag{16}$$

where y_j with $j \in J$, are observed values and J is a subset of the latent factors, and $\pi(y \mid x, \theta)$ is a likelihood function of observed variables. The second stage defines the latent parameter,

$$x_i = \text{Offset}_i + \sum_{k=0}^{\eta_f - 1} \omega_{ki} f_k(c_{ki}) + z_i^T \beta + \epsilon_i, \quad i = 0, \ldots, \eta_x - 1 \tag{17}$$

where Offset is a prior known component to be included in the linear prediction, ω_k are known weights for each observed data point, $f_k(c_{ki})$ is the effect of generic covariates with value c_{ki} for observation i, and β are the regression parameters of linear covariates z_i. Finally, the third stage of the model consists of the prior distribution for the hyperparameters θ.

The INLA approach obtains accurate approximations of the posterior distributions of the latent factors, written as

$$\pi(x_i \mid Y) = \int \pi(x_i \mid \theta, Y) \pi(\Theta, Y) \, \mathrm{d}\theta \tag{18}$$

and the marginal posterior distribution of hyperparameters, given by

$$\pi(\theta_j \mid Y) = \int \pi(\theta \mid Y) \, \mathrm{d}\theta_{\_j} \tag{19}$$

where $\theta_{\_j}$ denotes the vector θ without its jth element.

In summary, the INLA method can be implemented in three steps. First, we obtain an approximation of the full posterior distribution, $\pi(\theta \mid y)$, by Laplace approximation. Second, we obtain an approximation to the full conditional distributions, $\pi(x_i \mid \theta, y)$, for particular values of θ. In the last step, we obtain the marginal posterior distributions in (18) and (19) by combining the two approximations in the previous steps and integrating out the irrelevant factors. More specifically, we integrate over all the latent variables, x_i, that are not required to calculate the marginal posterior distributions of interest. For more details, please refer to Rue et al. [30].

In the Appendix we provide a Monte Carlo analysis comparing the INLA estimation of the DNS-fGn process with a MCMC estimation. Based on the Monte Carlo experiment conducted, it can be inferred that the INLA technique is a viable option for estimating a dynamic Nelson–Siegel model with long memory structure that involves fractional Gaussian Noise for latent factors. The results obtained from this approach indicate that it performs similarly to MCMC in terms of estimation propertis, with a huge efficiency gain in terms of computational performance.

In this work, we adopt an estimation structure that builds on the method proposed by Laurini and Hotta [31] for estimating and forecasting the term structure of interest rates using INLA. Similar to Laurini and Hotta [31], we use an additive structure in which the decay parameter λ is kept fixed and determined prior to the estimation process. This is

necessary due to the need for an additive structure in the measurement equation. However, it is possible to perform an optimal choice procedure for this parameter in a two-step procedure. Specifically, we can optimize some loss function in terms of this parameter in an estimation sample, using a generalized cross-validation procedure, and then use the selected parameter for the forecast sample. This procedure is described in detail in Laurini and Hotta [31] and is also used in this article, as presented in Section 4.1.3.

4. Empirical results

4.1. Fama-Bliss US Treasury zero-coupon data

4.1.1. Full sample analysis

To compare the DNS with fGn dynamics and the AR(1) specification proposed by Diebold and Li [10], we use the same database of time series of monthly unsmoothed Fama-Bliss US Treasury zero-coupon yields from January 1985 until December 2000, for maturities of 3, 6, 9, 12, 15, 18, 21, 24, 30, 36, 48, 60, 72, 84, 96, 108, 120 months. Figure 1 shows a three-dimensional plot of the yields used in this paper.

We assume that the parameter λ is constant and known, following Diebold and Li [10], and set it equal to 0.0609. Under this assumption, we estimate a DNS-fGn using the INLA method, as described in Section 2. To approximate the fGn model, we follow Sørbye et al. [25] and employ three independent AR(1) processes. We use a Penalized Complexity (PC) prior with parameters (3, 0.01) for the precision parameters of the Gaussian observations, and a PC prior with parameters (0.9, 0.1) for H. Table 1 displays the results for the posterior distributions of the precision of the Gaussian observations, the Hurst exponent H. and the precision for the latent factors. The mean values of the parameters H for β_{1t}, β_{2t}

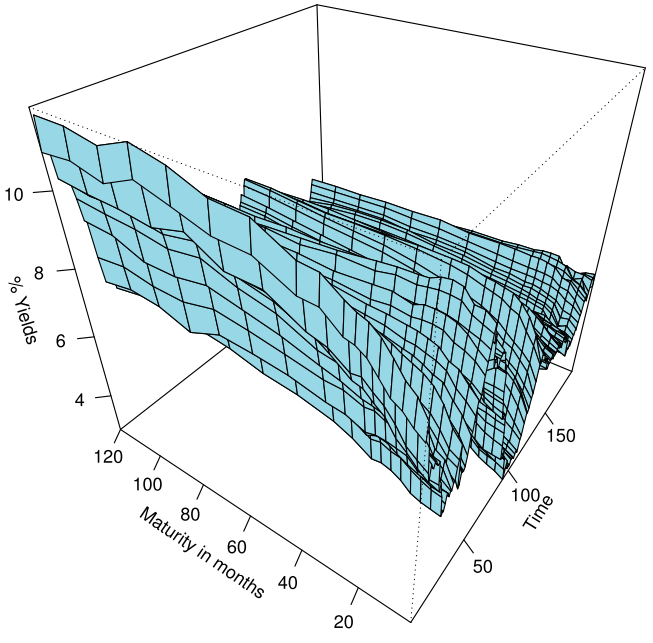


Figure 1. Unsmoothed Fama-Bliss US Treasury zero-coupon yields. January 1985–December 2000.

Table 1. Summary hyperparameters.

	Mean	SD	0.025quant.	0.5 quant.	0.975quant.
Precision τ for Gaussian obs.	197.523	4.783	188.335	197.452	207.145
μ_1	7.579	0.005	7.569	7.579	7.589
Precision τ_1 for β_{1t}	0.263	0.026	0.215	0.261	0.321
H for β_{1t}	0.991	0.001	0.989	0.991	0.992
μ_2	−2.098	0.005	−2.108	−2.098	−2.088
Precision τ_2 for β_{2t}	0.161	0.044	0.084	0.155	0.276
H for β_{2t}	0.992	0.002	0.987	0.992	0.996
μ_3	−0.162	0.019	−0.201	−0.162	−0.124
Precision τ_3 for β_{3t}	0.185	0.017	0.153	0.183	0.224
H for β_{3t}	0.975	0.003	0.969	0.975	0.980

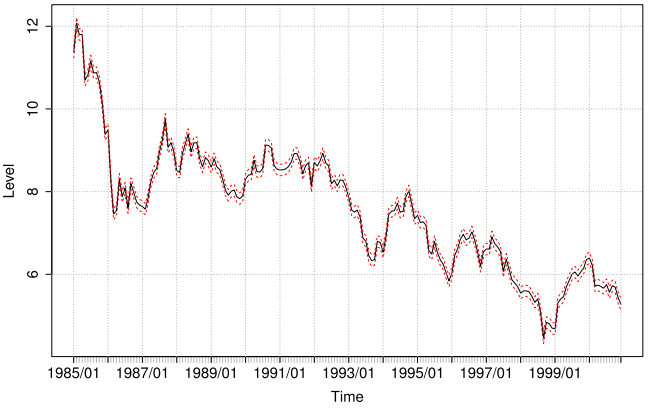
and β_{3t} are close to one, which is consistent with the persistent patterns typically found in financial time series. Moreover, this reproduces the pattern of high persistence observed in yield curves, as discussed in Goliński and Zaffaroni [23]. We also computed 95% credibility intervals based on the empirical quantiles of 0.025 and 0.975 of the posterior distributions, which are reported in Table 1.

Figure 2 shows the posterior mean and the 95% credibility interval for the estimated latent factors. The results are similar to the results obtained with the two-step OLS estimation used in Diebold and Li [10]. The main question is to verify if the use of the DNS-fGn structure leads to gains in terms of out-of-sample forecasting, which is discussed in the next section.

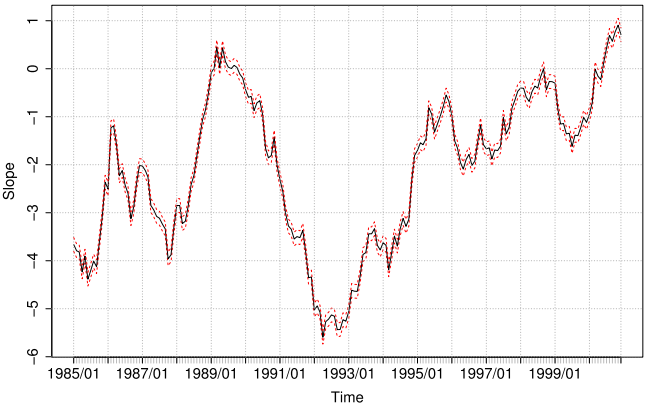
We also compare our proposed long-memory formulation using fGn with the Bayesian estimation of the Nelson–Siegel dynamic model with AR(1) dynamics for the latent factors using INLA. We perform this comparison using the marginal likelihood and log-Bayes factor. The estimated marginal log likelihood for the fGn dynamic model for the latent sample is −2736.763, while the model with AR(1) dynamics has an estimated marginal likelihood of −2857.737. Thus, the log-Bayes factor between the two models is 120.974, which according to Kass and Raftery [47], provides strong evidence in favour of the fGn dynamic model for latent factors

4.1.2. Out-of-sample forecasts

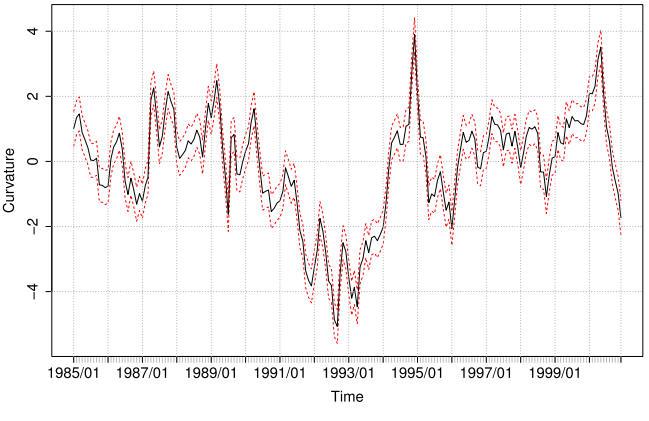
We compare the out-of-sample forecasting performance of the DNS model using three different specifications for the latent factors: AR(1), ARFIMA (0,d,0), and fGn. We refer to these models as DNS-AR(1), DNS-ARFIMA, and DNS-fGn, respectively. We also tested alternative specifications using AR(1)+AR(12) and AR(12) dynamics for the latent factors, but the results in general are dominated by the other dynamic structures for the latent factors, and for space reasons are not displayed in the text. We assess the forecasting performance for 1-month, 6-months, and 12-months ahead for maturities of 3 months and 1, 3, 5, and 10 years. The construction of the forecasts follows the same specification used in Diebold and Li [10]. We perform estimation and forecasting recursively, using an estimation sample that starts in January of 1985 until the forecast period, which starts in January of 1994 and goes until December 2000. We use the same value of λ used in Diebold and Li [10] for the DNS and DNS-fGn specifications. The DNS-fGn model is estimated Bayesian using INLA, following the methodology proposed in Sections 2



(a) Level - β_1



(b) Slope - β_2



(c) Curvature - β_3

Figure 2. Posterior mean and 95% Credibility Intervals – Latent Factors – Fama-Bliss US Treasury zero-coupon data. (a) Level – β_1. (b) Slope – β_2 and (c) Curvature – β_3.

Table 2. Out of sample 1-month ahead forecasting results.

	Mean	Std.Dev	RMSE	ACF1	ACF12	DM	*p*value
Dynamic Nelson–Siegel with AR(1) Factor Dynamics							
3 months	−0.036	0.169	0.172	0.275	0.013	.	.
1 year	0.033	0.235	0.236	0.449	−0.227	.	.
3 years	−0.047	0.274	0.276	0.352	−0.132	.	.
5 years	−0.081	0.278	0.288	0.352	−0.128	.	.
10 years	−0.053	0.253	0.257	0.269	−0.115	.	.
Dynamic Nelson–Siegel with fGn Factor Dynamics							
3 months	0.005	0.198	0.197	0.618	−0.037	−1.87	0.06
1 year	0.049	0.291	0.293	0.702	−0.214	−2.78	0.01
3 years	−0.086	0.333	0.342	0.649	−0.213	−3.40	0.00
5 years	−0.156	0.343	0.375	0.662	−0.246	−4.60	0.00
10 years	−0.169	0.310	0.351	0.624	−0.213	−4.74	0.00
Dynamic Nelson–Siegel with ARFIMA(0,d,0) Factor Dynamics							
3 months	0.005	0.224	0.222	0.714	−0.059	−2.885	0.004
1 year	0.039	0.320	0.321	0.752	−0.202	−3.223	0.001
3 years	−0.108	0.362	0.376	0.708	−0.212	−4.056	1.13e−04
5 years	−0.182	0.371	0.411	0.720	−0.248	−5.096	2.18e−06
10 years	−0.197	0.335	0.387	0.692	−0.218	−5.250	1.17e−06

Table 3. Out of sample 6-month ahead forecasting results.

	Mean	Std.Dev	RMSE	ACF6	ACF18	DM	*p*value
Dynamic Nelson–Siegel with AR(1) Factor Dynamics							
3 months	−0.056	0.543	0.542	0.378	−0.242	.	.
1 year	−0.031	0.668	0.664	0.153	−0.190	.	.
3 years	−0.220	0.737	0.765	0.004	−0.219	.	.
5 years	−0.336	0.745	0.813	0.037	−0.247	.	.
10 years	−0.410	0.650	0.765	0.009	−0.253	.	.
Dynamic Nelson–Siegel with fGn Factor Dynamics							
3 months	0.076	0.539	0.541	0.440	−0.258	0.01	0.99
1 year	0.043	0.666	0.663	0.226	−0.189	0.01	0.99
3 years	−0.242	0.740	0.774	0.091	−0.216	−0.13	0.89
5 years	−0.409	0.745	0.846	0.106	−0.260	−0.49	0.63
10 years	−0.533	0.665	0.849	0.117	−0.252	−1.21	0.23
Dynamic Nelson–Siegel with ARFIMA(0,d,0) Factor Dynamics							
3 months	0.068	0.553	0.554	0.457	−0.242	−0.138	0.890
1 year	0.026	0.676	0.673	0.240	−0.177	−0.098	0.921
3 years	−0.268	0.745	0.787	0.099	−0.207	−0.289	0.772
5 years	−0.436	0.747	0.861	0.111	−0.255	−0.633	0.528
10 years	−0.560	0.665	0.866	0.125	−0.250	−1.307	0.195

and 3. The estimation of the ARFIMA model is based on a modified two-step estimation method proposed in Diebold and Li [10]. In the second estimation step, we use an ARFIMA (0,d,0) process to model the dynamics of the time series of Betas estimated in the first step (cross-section in each day). We estimate the ARFIMA (0,d,0) process using the maximum-likelihood method based on the Whittle estimator (e.g. Palma [24,p. 78–81]).

Tables 2, 3 and 4 present a comprehensive analysis of the forecasting performance of the DNS-AR(1), DNS-ARFIMA, and DNS-fGn models. Specifically, these tables report the mean error, standard deviation, RMSE, autocorrelations of forecasting errors of order 1 and 12, for each of the models and for different forecasting horizons of 1-month, 6-months and 12-months, for maturities of 3 months and 1, 3, 5 and 10 years.

Table 4. Out of sample 12-month ahead forecasting results.

	Mean	Std.Dev	RMSE	ACF12	ACF24	DM	*p*value
Dynamic Nelson–Siegel with AR(1) Factor Dynamics							
3 months	−0.248	0.750	0.785	−0.065	−0.084	.	.
1 year	−0.322	0.803	0.860	−0.099	−0.070	.	.
3 years	−0.654	0.851	1.069	−0.182	−0.076	.	.
5 years	−0.856	0.867	1.214	−0.187	−0.093	.	.
10 years	−1.030	0.761	1.277	−0.228	−0.138	.	.
Dynamic Nelson–Siegel with fGn Factor Dynamics							
3 months	0.040	0.659	0.656	−0.145	−0.130	0.92	0.36
1 year	−0.064	0.698	0.696	−0.229	−0.088	1.76	0.08
3 years	−0.439	0.753	0.867	−0.301	−0.058	2.05	0.04
5 years	−0.660	0.767	1.008	−0.320	−0.077	2.00	0.05
10 years	−0.851	0.692	1.094	−0.332	−0.121	1.98	0.05
Dynamic Nelson–Siegel with ARFIMA(0,d,0) Factor Dynamics							
3 months	0.029	0.675	0.671	−0.121	−0.133	0.762	0.448
1 year	−0.085	0.707	0.707	−0.210	−0.091	1.489	0.140
3 years	−0.470	0.756	0.886	−0.288	−0.059	1.716	0.090
5 years	−0.693	0.766	1.029	−0.311	−0.079	1.692	0.094
10 years	−0.885	0.689	1.118	−0.320	−0.124	1.646	0.104

To compare the predictive performance of these models with the Dynamic Nelson–Siegel with AR(1) Factor Dynamics, we employ the Diebold-Mariano (DM) statistic, which measures the statistical significance of the difference in forecast accuracy between two models. We follow the same structure adopted by Diebold and Li [10] and report the DM statistic and the associated *p*value for each forecasting horizon.

The DM test is based on the squared difference between the forecast errors of the two models, and the alternative hypothesis assumes a difference between the predictive performance of the two models. The result of the test is interpreted based on the sign of the DM statistic. A negative sign in the DM statistic indicates that the compared model has a greater mean squared forecast error than the reference model, while a positive sign suggests a smaller squared error than the reference model.

Based on the results in Table 2, our analysis reveals that the DNS-AR(1) factor dynamics specification performs better for short (1-month ahead) and medium (6-months ahead) forecast horizons than the ARFIMA and fGn specifications, across most of the maturities. The differences in predictive performance are statistically significant based on the results of the DM tests. Furthermore, DNS-fGn outperforms DNS-ARFIMA for all maturities. For the 6-month ahead forecast horizon, the DM tests do not reveal any statistically significant differences in predictive performance with respect to the specification with AR(1) dynamics, although there are some slight variations in the model with the best performance.

When examining longer forecasting horizons (12-months ahead), our results indicate that the DNS-fGn specification outperforms both DNS-AR and DNS-ARFIMA specifications. Specifically, DNS-fGn demonstrates superior forecasting performance for all maturities, while DNS-ARFIMA forecasts are also more accurate than those of DNS-AR for all maturities. The detailed results for these comparisons are reported in Table 4.

Overall, our results suggest that the performance of the DNS model depends on the forecast horizon considered. For short and medium horizons, a persistence component on

the DNS model does not provide forecasting gains, as the DNS-AR specification outperforms both the DNS-fGn and DNS-ARFIMA specifications. However, for longer horizons, both the DNS-fGn and DNS-ARFIMA specifications allow for relatively more accurate forecasts.

4.1.3. Optimal selection of λ parameter

As pointed out in Laurini and Hotta [31], the use of the INLA approach for estimating the dynamic Nelson–Siegel model relies on a linear additive structure in the state space form, which requires the value of a tuning/decay parameter λ to be known. However, assuming a fixed value for λ may not be optimal in practice. To address this issue, Laurini and Hotta [31] propose a generalized cross-validation (GCV) approach to choose an optimal value of λ. We will compare this approach with two other alternative ways of estimating the lambda parameter. The first involves estimating the model using Markov Chain Monte Carlo, using the Hurst coefficient expansion proposed in Chan and Palma [26] and Chan and Petris [48]. The second way is to combine the INLA estimation with a Metropolis-Hastings step to sample the distribution of the lambda parameter, using the methodology proposed in Gómez-Rubio and Rue [49].

The GCV approach involves finding the value of λ that maximizes a measure of model fit, which in our case is the cross-validated log-score. To calculate the GCV score, we can use the conditional predictive ordinates (CPO) calculated by INLA for each observation. The CPO for observation y_i given the rest of the data is defined as follows:

$$\text{CPO}_i = \pi(y_i^{obs} \mid y_{-1}), \tag{20}$$

and the CPO of observation i is obtained as:

$$\text{CPO}_i = \int \pi(y_i^{obs} \mid y_{-1}, \theta)\pi(\theta \mid y_{-i})\, \mathrm{d}\theta \tag{21}$$

with the first term in this integral represented by:

$$\pi(y_i^{obs} \mid y_{-1}) = 1/\int \frac{\pi(x_i \mid y, \theta)}{\pi(y_i^{obs} \mid x_i, \theta)} \tag{22}$$

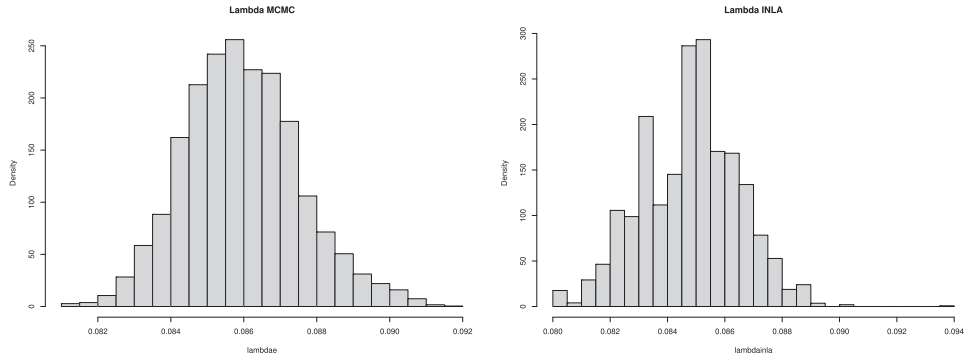
and calculated by numerical integration. The leave-one-out cross-validation likelihood (cross-validated log-score) is defined as:

$$LSCV = \sum_{i=1}^{n} -log(\text{CPO}_i). \tag{23}$$

With this approximation we can estimate the optimal λ parameter optimizing the cross-validated log-score model using a numerical minimization algorithm. In this problem we use a box constrained BFGS optimization method. We apply this method to select the optimal λ parameter for the first estimation sample for the Bayesian estimation with the fGn process in the forecasting procedure, and maintain this fixed value for the other forecast periods. The estimated optimal value was 0.1482749, and the out-of-sample forecast results with this value are shown in Table 5.

Table 5. Forecasting results with optimal selection of λ for the Dynamic Nelson–Siegel with fGn factor dynamics.

	Mean	Std.Dev	RMSE	ACF12	ACF24	DM	*p*value
Out of Sample 1-Month Ahead Forecasting Results							
INLA-GCV – $\lambda = 0.1482749$							
3 months	0.015	0.212	0.211	0.653	−0.071	−2.13	0.04
1 year	0.055	0.291	0.294	0.685	−0.192	−0.15	0.88
3 years	−0.113	0.324	0.341	0.644	−0.215	0.04	0.97
5 years	−0.164	0.340	0.375	0.662	−0.238	−0.12	0.91
10 years	−0.150	0.327	0.358	0.651	−0.173	−1.16	0.25
INLA-MH – $\lambda = 0.08479648$							
3 months	0.017	0.205	0.205	0.638	−0.065	−2.70	0.01
1 year	0.046	0.288	0.289	0.689	−0.204	2.66	0.01
3 years	−0.091	0.330	0.340	0.646	−0.216	0.81	0.42
5 years	−0.156	0.342	0.374	0.662	−0.243	1.03	0.31
10 years	−0.166	0.312	0.352	0.624	−0.197	−0.53	0.59
MCMC – $\lambda = 0.08595922$							
3 months	0.017	0.206	0.205	0.639	−0.066	−2.72	0.01
1 year	0.045	0.287	0.289	0.689	−0.2042	2.53	0.01
3 years	−0.092	0.330	0.341	0.647	−0.216	0.73	0.47
5 years	−0.157	0.342	0.374	0.662	−0.243	0.93	0.36
10 years	−0.166	0.313	0.352	0.625	−0.197	−0.59	0.55
Out of Sample 6-Month Ahead Forecasting Results							
INLA-GCV – $\lambda = 0.1482749$							
3 months	0.063	0.531	0.531	0.439	−0.260	0.69	0.49
1 year	0.062	0.677	0.675	0.214	−0.178	−0.78	0.44
3 years	−0.293	0.725	0.778	0.097	−0.228	−0.15	0.88
5 years	−0.429	0.738	0.850	0.111	−0.266	−0.46	0.65
10 years	−0.491	0.688	0.841	0.111	−0.231	0.53	0.60
INLA-MH – $\lambda = 0.08479648$							
3 months	0.081	0.538	0.541	0.440	−0.261	0.15	0.88
1 year	0.044	0.668	0.665	0.220	−0.185	−0.64	0.52
3 years	−0.253	0.736	0.774	0.092	−0.218	0.04	0.97
5 years	−0.415	0.741	0.846	0.108	−0.262	0.11	0.92
10 years	−0.525	0.672	0.849	0.113	−0.245	0.00	1.00
MCMC – $\lambda = 0.08595922$							
3 months	0.080	0.538	0.540	0.440	−0.261	0.24	0.81
1 year	0.043	0.668	0.665	0.220	−0.185	−0.64	0.52
3 years	−0.254	0.736	0.774	0.092	−0.219	0.02	0.98
5 years	−0.416	0.741	0.846	0.108	−0.263	0.05	0.96
10 years	−0.524	0.673	0.850	0.113	−0.245	−0.00	1.00
Out of Sample 12-Month Ahead Forecasting Results							
INLA-GCV – $\lambda = 0.1482749$							
3 months	0.013	0.646	0.642	−0.139	−0.138	0.66	0.51
1 year	−0.036	0.707	0.703	−0.237	−0.082	−0.48	0.63
3 years	−0.502	0.750	0.898	−0.297	−0.067	−1.87	0.07
5 years	−0.687	0.767	1.025	−0.317	−0.082	−2.33	0.02
10 years	−0.795	0.703	1.057	−0.339	−0.105	2.84	0.01
INLA-MH – $\lambda = 0.08479648$							
3 months	0.040	0.655	0.651	−0.144	−0.132	0.89	0.38
1 year	−0.061	0.700	0.698	−0.233	−0.085	−0.64	0.53
3 years	−0.453	0.753	0.874	−0.301	−0.060	−1.74	0.09
5 years	−0.670	0.766	1.014	−0.318	−0.079	−1.92	0.06
10 years	−0.838	0.696	1.086	−0.337	−0.115	2.46	0.02
MCMC – $\lambda = 0.08595922$							
3 months	0.039	0.654	0.651	−0.143	−0.132	0.85	0.40
1 year	−0.061	0.701	0.698	−0.233	−0.086	−0.71	0.48
3 years	−0.454	0.753	0.875	−0.301	−0.061	−1.80	0.08
5 years	−0.671	0.766	1.014	−0.318	−0.080	−1.93	0.06
10 years	−0.837	0.697	1.086	−0.337	−0.115	2.59	0.01



(a) MCMC (b) INLA-MH

Figure 3. λ Estimated by MCMC and INLA-Metropolis Hastings. (a) MCMC. and (b) INLA-MH.

The second way to determine the parameter λ used in the out-of-sample forecast is to use a Markov Chain Monte Carlo estimation, placing a prior distribution for this parameter. From this estimated posterior distribution we can obtain an estimate for the parameter λ to be used as a fixed parameter in the estimation using INLA. The details of the MCMC estimation procedure are presented in the Appendix of the article, and the only change is to introduce a log-normal prior for the parameter λ with a mean parameter with the value log(0.0609), that is, the log of mean is centred on the value used in Diebold and Li [10] for the parameter λ and variance.1.We use the posterior mean of the lambda parameter estimated by MCMC as a fixed λ parameter in the forecasting procedure.

The third way to estimate the posterior distribution of the λ parameter is constructed by combining the INLA approximation with a Metropolis-Hastings step, using the INLA-MH methodology proposed in Gómez-Rubio and Rue [49]. This methodology allows estimating non-linear models using INLA, in cases where the model is conditionally linear if the parameter is known. In this procedure we use an additional Metropolis-Hastings step to obtain a sample of λ, and using this value we estimate the other parameters of the model using INLA in the usual way. The central idea of the method is to combine the Metropolis-Hastings method with a Bayesian Model Averaging using the marginal likelihood of the model estimated using INLA. Due to space issues, we do not present the details of the method, but the interested reader can consult Gómez-Rubio and Rue [49] and the documentation of the function INLAMH in the INLABMA package[2]. We used a logit-Normal distribution with standard deviation 0.15 as an auxiliary distribution in the Metropolis-Hastings procedure, and for the prior of λ the same log-normal distribution used in the MCMC estimation. The posterior distributions obtained by the MCMC and INLA-MH algorithms for λ are shown in Figure 3, using chains of size 30,000 with a burn-in of 4,000 for MCMC estimation and 5,000 samples with burn-in 1,000 for estimation by INLA-MH. Due to the computational cost of estimation using the INLA-MH procedure, we used a smaller sample size, but the results indicate low correlation and convergence in this chain.

The posterior mean was estimated with the value of O.08479648 by INLA-MH and 0.08595922 by MCMC. Based on these values we performed out-of-sample forecast analyses, with the results presented in Table 5. In this table the results of the Diebold-Mariano

test compare the results with the predictions for the Dynamic Nelson–Siegel model with fGn Factor Dynamics assuming $\lambda = 0.0609$.

We can see that the results in terms of predictive gain by looking at the RMSE results are mixed. For the forecast horizon of 6 months ahead there is no predictive gain at any maturity in relation to the model using $\lambda = 0.0609$. We can observe significant forecast gains only for the 10-year maturity in the 12-months ahead forecast horizon by the three optimal selection methods for λ, and for the other maturities and selection methods there are statistically significant losses in relation to the value of λ chosen by Diebold and Li [10], indicating the robustness of this value for the sample analysed.

5. Regime switching and long memory

The empirical analysis of the U.S. Treasury Bonds dataset reveals that long memory structures in latent factors result in improved forecasting accuracy for long horizons. However, the presence of non-Markovian dynamics in the evolution of prices in financial instruments such as the U.S. Treasury market seems counter-intuitive given their high liquidity ([50]). This raises the question of why long memory processes are observed in yields.

One possible explanation for this phenomenon is that parameter changes, such as structural or regime changes, in the data-generating process could be responsible for the observed long memory. In fact, the equivalence between long memory processes and parameter changes has been extensively studied in the literature ([51–53]). As highlighted in Diebold and Inoue [54], even if the true data-generating process is a regime change model, a long memory model can still provide predictive gains over a short memory model by approximating the dependence structure induced by the regime change processes.

Therefore, our perspective is aligned with Diebold and Inoue [54], and we suggest that the observed long memory in yields may be an approximation of the dependence structure induced by the parameter changes. Nevertheless, the use of long memory models in forecasting can still be justified by their improved predictive performance compared to short memory models, regardless of the underlying data-generating process.

We hypothesize that the superior forecasting properties of the DNS-fGn model for long horizons in our analysis are due to the presence of regime changes. The existence of regime changes in the term structure of interest rates has been studied extensively in the literature, with several studies (e.g. Hevia et al. [55], Zhu and Rahman [56] and Levant and Ma [57]) suggesting their presence. To investigate this possibility further, we conducted a Monte Carlo study to analyse the impact of a Markovian regime switching structure on the predictive performance of the Nelson–Siegel dynamic model with the original specification of Diebold and Li using short-memory AR(1) processes for latent factors, and the DNS model with dynamics given by fGn processes as discussed in this article.

In our study, we used the basic framework of models with Markov switching, where a latent S_t state determines the parameter set in each period of time, given by the realization of a Markov chain. The probability of transition between regimes is given by $P(S_t = j \mid S_{t-1} = i) = P_{ij}$, where we assumed two regimes $i, j = 1, 2$, and the constraint $\sum_{j=1}^{2} P(S_t = j \mid S_{t-1} = i) = 1$ holds.

We simulated 200 interest rate curves assuming that the data-generating process followed a Dynamic Nelson–Siegel model (Equation 2), but with the latent factors modelled by a Markov Switching process with two regimes modelled by first-order autoregressive

processes. In the first experiment, we assumed that only the autoregressive parameters were subject to regime changes, while the other parameters remained fixed between regimes (mean and precision of each latent factor and the measurement error of the observation equation). In a second experiment, we also allowed the mean of yields to depend on the current regime, allowing the mean of the level factor to be regime-dependent. The latent factor dynamics for the specification with regime changes in autoregressive and mean parameters are given by:

$$\begin{pmatrix} \beta_{1t} - \mu_1^i \\ \beta_{2t} - \mu_2 \\ \beta_{3t} - \mu_3 \end{pmatrix} = \begin{pmatrix} \phi_1^i & 0 & 0 \\ 0 & \phi_2^i & 0 \\ 0 & 0 & \phi_3^i \end{pmatrix} \begin{pmatrix} \beta_{1t-1} - \mu_1^i \\ \beta_{2t-1} - \mu_2 \\ \beta_{3t-1} - \mu_3 \end{pmatrix} + \begin{pmatrix} \epsilon_{\beta_{1t}} \\ \epsilon_{\beta_{1t}} \\ \epsilon_{\beta_{3t}} \end{pmatrix} \tag{24}$$

where we assume that $i = 1$ if $S_t = 1$ and $i = 2$ if $S_t = 2$, with S_t modelled as a first-order Markov Chain. In each realization we assume that the curve has the same sample size as the Fama-Bliss data analysed in this study (378 observations), and we performed exactly the same out-of-sample prediction experiment performed in Subsection 4.1.2 for each replication of the Monte Carlo experiment. We save the forecast error measures obtained in each replication, and we report it in Tables 8 and 9 the average of the results obtained in each process simulation.

The latent factor dynamics for the specification with regime changes in autoregressive and mean parameters are given by:

Regime 1: $(\phi_1^1, \phi_2^1, \phi_3^1) = (0.98, 0.95, 0.93)$

Regime 2: $(\phi_1^2, \phi_2^2, \phi_3^2) = (0.95, 0.90, 0.8)$.

In the first experiment, the mean parameters (μ_1, μ_2, μ_3) in the two regimes are given by $(7, 0, 0)$, the precision of the latent factors τ_1, τ_2, and τ_3 are given by $(0.8264463, 51.02, 4.93)$, and the precision of the measurement error τ_e is equal to 400. We also assumed that $\lambda = 0.0609$ and used the true value in the estimations. The Markov chain with the transition probabilities is given by:

$$P = \begin{pmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{pmatrix} \tag{25}$$

The results of the forecast analysis for one-step and twelve-step ahead predictions in the first configuration are presented in Tables 6 and 7. It is noteworthy that the results are quite similar to those obtained in the empirical analysis of the Fama-Bliss database. The one-step ahead forecasts (Table 6) constructed using the two-step method of Diebold and Li [10] exhibit better overall performance. On the other hand, for the twelve-step forecasts (Table 7), the DNS-fGn model demonstrates the best performance in terms of RMSE, reproducing the same pattern observed in the empirical analysis.

To investigate the impact of regime changes in the mean level of yields on the predictive performance, we conducted a second Monte Carlo experiment. In this experiment, we allowed the mean parameter of the first latent factor to be subject to regime changes, while keeping the other parameters at the same values as in the first experiment with Markovian regime switching. Specifically, we set the value of parameter μ_1^1 to 8 in the first regime, and 6 in the second regime.

Based on the results presented in Tables 8 and 9 for one-step and twelve-step ahead predictions, respectively, we observed a similar pattern of results as in the first experiment.

Table 6. Out of sample 1-month ahead forecasting results – Monte Carlo results with a regime switching DNS – Distinct persistence and equal means between regimes.

	Mean	Std.Dev	RMSE	ACF12	ACF24
Dynamic Nelson–Siegel with AR(1) Factor Dynamics					
3 months	0.001	0.180	0.181	0.006	−0.000
1 year	0.002	0.186	0.187	0.035	−0.006
3 years	0.001	0.188	0.190	0.039	−0.013
5 years	0.000	0.166	0.167	0.034	−0.015
10 years	−0.001	0.135	0.136	0.013	−0.005
Dynamic Nelson–Siegel with fGn Factor Dynamics					
3 months	0.003	0.197	0.202	0.355	−0.017
1 year	0.004	0.203	0.208	0.349	−0.024
3 years	0.002	0.205	0.209	0.342	−0.027
5 years	0.001	0.181	0.185	0.335	−0.023
10 years	−0.000	0.148	0.153	0.350	0.002

Table 7. Out of Sample 12-Month Ahead Forecasting Results – Monte Carlo results with a regime switching DNS – Distinct persistence and equal means between regimes.

	Mean	Std.Dev	RMSE	ACF12	ACF24
Dynamic Nelson–Siegel with AR(1) Factor Dynamics					
3 months	0.006	0.435	0.492	−0.060	−0.152
1 year	0.005	0.440	0.496	−0.060	−0.143
3 years	−0.001	0.431	0.484	−0.058	−0.137
5 years	−0.005	0.385	0.434	−0.054	−0.141
10 years	−0.010	0.327	0.370	−0.044	−0.155
Dynamic Nelson–Siegel with fGn Factor Dynamics					
3 months	0.009	0.430	0.479	−0.049	−0.156
1 year	0.010	0.439	0.486	−0.058	−0.150
3 years	0.006	0.432	0.475	−0.058	−0.139
5 years	0.002	0.384	0.426	−0.041	−0.146
10 years	−0.003	0.323	0.364	−0.001	−0.163

Table 8. Out of Sample 1-Month Ahead Forecasting Results – Monte Carlo results with a regime switching DNS – Distinct persistence and equal means between regimes.

	Mean	Std.Dev	RMSE	ACF12	ACF24
Dynamic Nelson–Siegel with AR(1) Factor Dynamics					
3 months	0.000	0.183	0.184	0.021	−0.006
1 year	0.001	0.189	0.190	0.049	−0.012
3 years	−0.000	0.191	0.192	0.054	−0.017
5 years	−0.001	0.169	0.170	0.051	−0.019
10 years	−0.002	0.139	0.139	0.035	−0.010
Dynamic Nelson–Siegel with fGn Factor Dynamics					
3 months	−0.000	0.205	0.211	0.392	−0.021
1 year	.001	0.211	0.216	0.385	−0.026
3 years	−0.001	0.213	0.217	0.377	−0.027
5 years	−0.002	0.190	0.195	0.381	−0.022
10 years	−0.003	0.158	0.164	0.410	0.000

The two-step OLS estimation and first-order autoregressive dynamics yielded better performance for one-step ahead forecasts, while the DNS-fGn model dominated in all criteria for twelve-step ahead forecasts.

These findings suggest that the superior predictive performance of the fractional Gaussian noise model with long memory may be due to the possible presence of parameter

Table 9. Out of Sample 12-Month Ahead Forecasting Results – Monte Carlo results with a regime switching DNS – Distinct persistence and equal means between regimes.

	Mean	Std.Dev	RMSE	ACF12	ACF24
Dynamic Nelson–Siegel with AR(1) Factor Dynamics					
3 months	−0.005	0.473	0.535	−0.061	−0.150
1 year	−0.007	0.479	0.539	−0.056	−0.142
3 years	−0.013	0.471	0.527	−0.049	−0.136
5 years	−0.017	0.429	0.481	−0.045	−0.140
10 years	−0.022	0.374	0.423	−0.038	−0.148
Dynamic Nelson–Siegel with fGn Factor Dynamics					
3 months	−0.004	0.466	0.521	−0.040	−0.153
1 year	−0.003	0.475	0.527	−0.047	−0.145
3 years	−0.007	0.468	0.517	−0.045	−0.135
5 years	−0.010	0.424	0.473	−0.029	−0.140
10 years	−0.016	0.367	0.417	0.006	−0.154

changes in the data generating process. By using the fGn structure as a dynamic mechanism for latent factors, we obtain a better approximation for the correlation structure that arises from potential changes in the data parameters. Therefore, it can be inferred that the DNS-fGn model can be an efficient tool for long-horizon forecasts in the presence of regime changes in the data generating process.

It is worth noting that this approach is directly linked to the method proposed by Sørbye et al. [25] for representing the fGn process in a Gaussian Markov Random Field structure via the aggregation of first-order autoregressive processes.

6. Conclusions

In this paper, we propose a novel specification for the factor dynamics of the Dynamic Nelson–Siegel (DNS) model to incorporate long-range persistence. Our aim is to investigate whether this approach improves the accuracy of interest rate term structure forecasts. We model the evolution of the latent factors with fractional Gaussian noise, which we approximate using a weighted sum of independent first-order autoregressive processes. This can be represented as a Gaussian Markov Random Field, and we use Integrated Nested Laplace Approximations (INLA) to obtain Bayesian parameter estimation.

To assess the properties of our approximations, we compare parameter estimation using INLA and Markov Chain Monte Carlo (MCMC) methods in a Monte Carlo experiment, reported in the Appendix. Our results show that INLA provides comparable performance to MCMC, with the added benefit of reduced computational costs.

For the empirical analysis, we compare our proposed DNS-fGn model with alternative specifications for the latent factors using the Fama-Bliss dataset. Specifically, we compare it with an autoregressive model and an autoregressive fractionally integrated moving average model. Using a fixed and known decay parameter, our out-of-sample forecasts show that the DNS model with AR(1) factor dynamics outperforms DNS with fGn and DNS with ARFIMA(0,d,0) for 3-months and 6-months ahead forecasts for all maturities. However, for 12-months ahead forecasting, the DNS with fGn dynamics performs better than the AR(1) and ARFIMA(0,d,0) specifications for all maturities. Therefore, we provide evidence that long memory is helpful for longer forecasting horizons. Moreover, we show

that selecting the decay parameter based on optimization leads to improved forecasting performance for the DNS-fGn model.

To account for the possibility of parameter changes in the data generating process, we conduct a Monte Carlo analysis to evaluate the impact of a Markovian regime switching structure on the forecasting performance of the DNS-fGn model with the DNS model with AR(1) factor dynamics. Our results suggest that the DNS-fGn model outperforms the DNS model with AR(1) factor dynamics in the presence of regime changes.

Notes

1. See Sørbye et al. [25] for more details on the implementation.
2. https://cran.r-project.org/web/packages/INLABMA/index.html

Disclosure statement

The authors report that there are no conflicts of interest of any kind.

Funding

The authors acknowledge funding from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) [grant number 306023/2018-0], Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) [grant number 2018/04654-9], Coordenação de Aperfeiçoamento de Pessoal de Nível (Capes) [grant number Finance Code 001].

References

[1] Brigo D, Mercurio F. Interest rate models – theory and practice: with smile, inflation and credit. Berlin, Heidelberg: Springer; 2006.
[2] Diebold FX, Rudebusch GD. Yield curve modeling and forecasting: the dynamic Nelson–Siegel approach. Princeton, NJ: Princeton University Press; 2013.
[3] Filipovic D. Term structure models: a graduate course. Berlin, Heidelberg: Springer; 2009.
[4] De Pooter M, Ravazzolo F, Van Dijk DJ. Term structure forecasting using macro factors and forecast combination. FRB International Finance Discussion Paper; 2010. (Technical Report No. 993).
[5] Carriero A, Kapetanios G, Marcellino M. Forecasting government bond yields with large Bayesian vector autoregressions. J Bank Financ. 2012;36(7):2026–2047. doi: 10.1016/j.jbankfin.2012.03.008
[6] Fama EF, Bliss RR. The information in long-maturity forward rates. Am Econ Rev. 1987;77:680–692.
[7] Cochrane JH, Piazzesi M. Bond risk premia. Am Econ Rev. 2005;95(1):138–160. doi: 10.1257/0002828053828581
[8] Duffee GR. Term premia and interest rate forecasts in affine models. J Finance. 2002;57(1):405–443. doi: 10.1111/jofi.2002.57.issue-1
[9] Nelson CR, Siegel AF. Parsimonious modeling of yield curves. J Bus. 1987;60(4):473–489. doi: 10.1086/jb.1987.60.issue-4
[10] Diebold FX, Li C. Forecasting the term structure of government bond yields. J Econom. 2006;130(2):337–364. doi: 10.1016/j.jeconom.2005.03.005
[11] Christensen JH, Diebold FX, Rudebusch GD. The affine arbitrage-free class of Nelson–Siegel term structure models. J Econom. 2011;164(1):4–20. doi: 10.1016/j.jeconom.2011.02.011
[12] Koopman SJ, Mallee MI, Van der Wel M. Analyzing the term structure of interest rates using the dynamic Nelson–Siegel model with time-varying parameters. J Bus Econ Stat. 2010;28(3):329–343. doi: 10.1198/jbes.2009.07295

[13] De Pooter M. Examining the Nelson–Siegel class of term structure models. In-sample fit versus out-of-sample forecasting performance: SSRN; 2007. (Technical Report). Available from: SSRN 992748.
[14] Diebold FX, Li C, Yue VZ. Global yield curve dynamics and interactions: a dynamic Nelson–Siegel approach. J Econom. 2008;146(2):351–363. doi: 10.1016/j.jeconom.2008.08.017
[15] Zivot E, Wang J. Modeling financial time series with s-plus®. Berlin, Heidelberg: Springer Science & Business Media; 2007.
[16] Ding Z, Granger CW, Engle RF. A long memory property of stock market returns and a new model. J Empir Finance. 1993;1(1):83–106. doi: 10.1016/0927-5398(93)90006-D
[17] Breidt FJ, Crato N, de Lima P. Modeling long-memory stochastic volatility. J Econom. 1993;73:325–334.
[18] Hassler U, Wolters J. Long memory in inflation rates: international evidence. J Bus Econ Stat. 1995;13(1):37–45.
[19] Baillie RT. Long memory processes and fractional integration in econometrics. J Econom. 1996;73(1):5–59. doi: 10.1016/0304-4076(95)01732-1
[20] Backus DK, Zin SE. Long-memory inflation uncertainty: evidence from the term structure of interest rates. J Money Credit Bank. 1993;25(3):681–700. doi: 10.2307/2077735
[21] Tsay WJ. Long memory story of the real interest rate. Econ Lett. 2000;67(3):325–330. doi: 10.1016/S0165-1765(99)00272-4
[22] Couchman J, Gounder R, Su JJ. Long memory properties of real interest rates for 16 countries. Appl Financial Econ Lett. 2006;2(1):25–30. doi: 10.1080/17446540500396990
[23] Goliński A, Zaffaroni P. Long memory affine term structure models. J Econom. 2016;191(1):33–56. doi: 10.1016/j.jeconom.2015.09.006
[24] Palma W. Long-memory time series: theory and methods. Hoboken, NJ: John Wiley & Sons; 2007.
[25] Sørbye SH, Myrvoll-Nilsen E, Rue H. An approximate fractional Gaussian noise model with $\mathcal{O}(n)$ computational cost. Stat Comput. 2019;29:821–833. doi: 10.1007/s11222-018-9843-1
[26] Chan NH, Palma W. State space modelling of long memory processes. Ann Stat. 1998;26:719–740. doi: 10.1214/aos/1028144856
[27] Hosking J. Fractional differencing. Biometrika. 1981;68:165–176. doi: 10.1093/biomet/68.1.165
[28] Purczyński J, Włodarski P. On fast generation of fractional Gaussian noise. Comput Stat Data Anal. 2006;50(10):2537–2551. doi: 10.1016/j.csda.2005.04.021
[29] Taqqu M, Teverovsky V, Willinger W. Estimators for long-range dependence: an empirical study. Fractals. 1995;3:785–798. doi: 10.1142/S0218348X95000692
[30] Rue H, Martino S, Chopin N. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. J R Stat Soc Series B (Stat Methodol). 2009;71(2):319–392. doi: 10.1111/j.1467-9868.2008.00700.x
[31] Laurini MP, Hotta LK. Forecasting the term structure of interest rates using integrated nested Laplace approximations. J Forecast. 2014;33(3):214–230. doi: 10.1002/for.v33.3
[32] Duffie D, Kan R. A yield-factor model of interest rates. Math Financ. 1996;6(4):379–406. doi: 10.1111/mafi.1996.6.issue-4
[33] Diebold FX, Rudebusch GD, Aruoba SB. The macroeconomy and the yield curve: a dynamic latent factor approach. J Econom. 2006;131(1-2):309–338. doi: 10.1016/j.jeconom.2005.01.011
[34] Laurini MP, Hotta LK. Bayesian extensions to Diebold-Li term structure model. Int Rev Financial Anal. 2010;19(5):342–350. doi: 10.1016/j.irfa.2010.08.010
[35] Hautsch N, Yang F. Bayesian inference in a stochastic volatility Nelson–Siegel model. Comput Stat Data Anal. 2012;56(11):3774–3792. doi: 10.1016/j.csda.2010.07.0031st issue of the Annals of Computational and Financial Econometrics Sixth Special Issue on Computational Econometrics.
[36] Batista RL, Laurini M. Bayesian estimation of term structure models: an application of the Hamiltonian Monte Carlo method. Commun Stat Case Stud Data Anal Appl. 2016a;2(3–4):79–91.
[37] Beran J. Statistics for long-memory processes. Milton Park, Oxfordshire: Routledge; 2017.

[38] Abramowitz M, Stegun IA. Handbook of mathematical functions with formulas, graphs, and mathematical tables (ninth Dover printing, tenth GPO printing ed.). New York: Dover; 1964.
[39] Comte F, Renault E. Long memory continuous time models. J Econom. 1996;73(1):101–149. doi: 10.1016/0304-4076(95)01735-6
[40] Laurini MP, Hotta LK. Indirect inference in fractional short-term interest rate diffusions. Math Comput Simul. 2013;94:109–126. doi: 10.1016/j.matcom.2013.06.003
[41] Osterrieder D. Interest rates with long memory a: generalized affine term structure model; 2013. (Research Paper No. 2013–17).
[42] Abbritti M, Gil-Alana LA, Lovcha Y, et al. Term structure persistence. J Financial Econ. 2015;14(2):331–352.
[43] Shi S, Yu J, Zhang C. Fractional Gaussian noise. Spectral density and estimation methods; 2022. (Working Paper).
[44] Mandelbrot BB, Van Ness JW. Fractional Brownian motions, fractional noises and applications. SIAM Rev. 1968;10(4):422–437. doi: 10.1137/1010093
[45] Ohashi A. Fractional term structure models: no-arbitrage and consistency. Ann Appl Probab. 2009;19(4):1553–1580. doi: 10.1214/08-AAP586 [2023-10-13] https://www.jstor.org/stable/30243632.
[46] Rue H, Held L. Gaussian markov random fields: theory and applications. London, UK: Chapman & Hall – CRC; 2005.
[47] Kass RE, Raftery AE. Bayes factors. J Am Stat Assoc. 1995;90(430):773–795. doi: 10.1080/01621459.1995.10476572
[48] Chan NH, Petris G. Long memory stochastic volatility: a Bayesian approach. Commun Stat Theory Methods. 2000;29:1367–1378. doi: 10.1080/03610920008832549
[49] Gómez-Rubio V, Rue H. Markov chain Monte Carlo with the integrated nested Laplace approximation. Stat Comput. 2018;28:1033–1051. doi: 10.1007/s11222-017-9778-y
[50] Fleming MJ. Measuring treasury market liquidity. Econ Policy Rev. 2003;9(3):83–108.
[51] Granger C. Long memory relationships and the aggregation of dynamic models. J Econom. 1980;14(2):227–238. doi: 10.1016/0304-4076(80)90092-5
[52] Granger CWJ, Teräsvirta T. A simple nonlinear time series model with misleading linear. Econ Lett. 1999;62:161–165. doi: 10.1016/S0165-1765(98)00228-6
[53] Gourieroux C, Jasiak J. Memory and infrequent breaks. Econ Lett. 1999;70:29–41. doi: 10.1016/S0165-1765(00)00346-3
[54] Diebold FX, Inoue A. Long memory and regime switching. J Econom. 2001;105(1):131–159. doi: 10.1016/S0304-4076(01)00073-2Forecasting and empirical methods in finance and macroeconomics.
[55] Hevia C, Gonzales-Rozada M, Spagnolo F. Estimating and forecasting the yield curve using a Markov switching dynamic Nelson and Siegel model. J Appl Econ. 2015;30(6):987–1009. doi: 10.1002/jae.v30.6
[56] Zhu X, Rahman S. A regime-switching Nelson–Siegel term structure model of the macroeconomy. J Macroecon. 2015;44(C):1–17. doi: 10.1016/j.jmacro.2014.12.007
[57] Levant J, Ma J. A dynamic Nelson–Siegel yield curve model with Markov switching. Econ Model. 2017;67:73–87. doi: 10.1016/j.econmod.2016.10.003