Long Memory and Term Structure of Interest Rates

Fernanda C. Valente* Márcio P. Laurini*

Abstract

This article presents a modified version of Dynamic Nelson-Siegel term structure model with long memory properties. The evolution of the latent factors is given by fractional Gaussian noise (fGn) processes, which are approximated with a weighted sum of independent first-order autoregressive components, that can be represented as a Gaussian Markov Random Field, allowing the use of computationally efficient Bayesian methods. The purpose is to assess if a long memory structure on the Dynamic Nelson-Siegel model helps to obtain more accurate forecasts of the term structure of interest rates. We compare our model with alternative specifications for the factors, including autoregressive and autoregressive fractionally integrated models. The results indicate that long memory is indeed helpful for longer forecasting horizons. However, the fGn components do not seem to improve short horizon forecasts.

JEL: G12, C22, C11.

Keywords: Term Structure, Long Memory, Bayesian Forecasting, Laplace Approximations. **Área de submissão: Área 8 - Microeconomia, Métodos Quantitativos e Finanças**

Resumo

Este artigo propõe uma versão modificada do modelo Nelson-Siegel Dinâmico com propriedades de memória longa. A evolução dos fatores latentes é dada por processos de ruido Gaussiano fracionários, aproximados por uma soma ponderada de processos autorregressivos de primeira ordem, que podem ser representados como um Campo Aleatório Gaussiano Markoviano, permitindo o uso de métodos Bayesianos computacionalmente eficientes. O objetivo é avaliar se a inclusão de memória longa no modelo Nelson-Siegel Dinâmico contribui para obter previsões mais acuradas para a estrutura a termo da taxa de juros. O modelo proposto é comparado com especificações alternativas para os fatores, como processos autorregressivos e processos autorregressivos com ordem de integração fracionária. Os resultados indicam que o componente de memória longa é, de fato, útil

*Department of Economics - FEA-RP - Universidade de São Paulo. - Av. dos Bandeirantes 3900, 14040-905, Ribeirão Preto, SP, Brazil. Tel.: +55-16-33290867 - emails - fernanda.valente@usp.br and laurini@fearp.usp.br. The authors also acknowledge the financial support by CAPES, CNPq (303738/2015-4) and FAPESP (2018/04654-9).

para previsões de longo prazo. Entretanto, componentes de persistência longa não parecem contribuir para previsões de curto prazo.

Palavras-chave: Estrutura a termo, Memória Longa, Previsões Bayesianas, Aproximações de Laplace.

1 Introduction

The term structure of interest rates represents the relationship between the interest rates and time to maturity, therefore, is a high-dimensional object, which often is not directly observed. As discussed by Pooter, Ravazzolo and Dijk (2010), long yields are risk-adjusted averages of expected future short rates, consequently, yields of different maturities are related and move together, in the cross-section as well as over time. At the same time, shocks in the economy tends to cause different effects on long and short maturities. Furthermore, monetary policy authorities are actively targeting the short end of the yield curve to achieve their macroeconomic goals. Thus, it is easy to see that many forces are at work at moving interest rates. Understanding the dynamic evolution of the yield curve and the construction of forecasts is important for many tasks, such as bond portfolio management, derivatives pricing and risk management.

Hence, it is not surprising the meaningful number of researches dedicated to model and forecast the term structure of interest rates. Nevertheless, it is not an easy task and, it is not difficult to find in financial literature simple random walk forecasts more accurate than theoretically sophisticated models (see Duffee (2002)). As discussed by Carriero, Kapetanios and Marcellino (2012), the existing methods for producing forecasts of the term structure of the interest rates can be roughly categorized in three groups: models based on forward rate regressions (Fama and Bliss (1987) and Cochrane and Piazzesi (2005)), models based on the No-arbitrage paradigm (Duffee (2002)) and a class of models that uses the Nelson and Siegel (1987) exponential components framework (Diebold and Li (2006) and Christensen, Diebold and Rudebusch (2011)).

The Nelson-Siegel class is due to Nelson and Siegel (1987), who proposed fit the term structure of interest rates adopting a flexible, smooth parametric function. The model proposed by Nelson and Siegel (1987) has been widely explored in the literature because it has some desirable characteristics, according to Diebold and Rudebusch (2013). Parsimony and flexibility are appealing features because it promotes smoothness and allows the yield curve to assume a variety of shapes at different times. Moreover, Diebold and Rudebusch (2013) comment that from a mathematical point of view, the form proposed by Nelson and Siegel (1987) is far from arbitrary. In fact, the forward rate curve is a constant plus a Laguerre function.

A step further, Diebold and Li (2006) suggest the Dynamic Nelson-Siegel (DNS) model. They used the Nelson-Siegel form with time-varying parameters, introducing dynamics in the model. According to Diebold and Rudebusch (2013), the DNS model distills the yield curve into three dynamic, latent factors, providing interpretations corresponding to level, slope, and curvature of the interest rates. The evolution of the latent factors is modeled as either an autoregressive (AR) model for each individual latent factor or as a vector autoregressive (VAR) model for the three latent factors, simultaneously. Diebold and Li (2006) show that AR forecasts performs better than many alternatives, including the VAR specification.

Due to good results of Diebold and Li (2006), many papers have been developed in different dynamic approaches of Nelson-Siegel and deliver good results in terms of forecasting. For example, Koopman, Mallee and Wel (2010) propose to introduce time-varying loading and variance parameters, Pooter (2007) adds a second slope factor and Diebold, Li and Yue (2008) extend the DNS model to a global context, modeling to a large set of country yield curves in a framework that allows for both global and country-specific factors.

However, as pointed out by Zivot and Wang (2007), many macroeconomic and financial time series like nominal and real interest rates, real exchange rates, interest rates differentials and volatility measures are very persistent. The AR specification, proposed by Diebold and Li (2006), albeit a good forecast performance compared with alternative models, often cannot capture the high degree of persistence in time series. Evidences of the presence of long memory in interest rates can be found in Backus and Zin (1993), Couchman, Gounder and Su (2006), Tsay (2000) and Goliński and Zaffaroni (2016).

One way to model persistence is by long memory dependency structures. Long memory process can be characterized in terms of the autocorrelation function that decays slowly at a hyperbolic rate. In fact, long memory implies that the autocorrelations are not absolutely summable (Palma (2007)). Fractional Brownian motion (fBm) are particularly useful to model long range dependency structures. The fBm is a self-similar continuous-time Gaussian process, with stationary increments. The increment process of fBm is defined as fractional Gaussian noise (fGn), which is stationary and a parsimoniously parameterized model, with dependency structure characterized by the Hurst exponent H, that gives long memory when $1/2 < H < 1$ (Sørbye, Myrvoll-Nilsen and Rue (2017)).

Recognizing that the notion of long memory permits to obtain a degree of persistence, we purpose to modify the structure of DNS latent factors to embody persistence into the model, using an approximation for a fGn process, proposed by Sørbye, Myrvoll-Nilsen and Rue (2017). The key point of the approximate fGn model is the possibility to represent it as a Gaussian Markov random field (GMRF). This representation allows to obtain Bayesian estimates of the latent factors and the parameter H through the Integrated Nested Laplace Approximations (INLA) method, proposed by Rue, Martino and Chopin (2009). According to Laurini and Hotta (2014), the INLA method provides accurate analytical approximations for the posterior distribution of the latent factors parameters in generalized Gaussian models without the need for procedures such as numerical simulation methods based on Markov Chain Monte Carlo (MCMC).

In this paper we assess the extent to which a persistence component on the Dynamic Nelson-Siegel model provides additional forecasting gains in relation to DNS model with alternative specification for the factors, including first order autoregressive model and autoregressive fractionally integrated moving average model. The data set considered consists 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, 30, 36, 48, 60, 72, 84, 96, 108, 120 months, and Interbank Deposits (*Depósitos Interbancários - DI*) Futures in Brazil, from January 2003 until August 2017, for maturities from 1 to 72 months. We adopt a rolling estimation to obtain out-of-sample forecasts for 1-month, 6-months and 12-months ahead. Moreover, we evaluate forecasting accuracy by root mean squared error (RMSE). The results suggest that, considering a short forecast horizon, the persistence component do not seem to improve forecasts accuracy. However, it is possible to see that when longer forecasting horizon is considered, a persistence component appears to be helpful.

This article is organized as follow. Section 2 contains a description of the DNS model with long memory. Section 3 briefly summarizes the INLA method. Section presents the results of the estimation and forecasts. Section 5 concludes.

2 Dynamic Nelson-Siegel model

Following Diebold and Li (2006), the DNS model can be represented by a state-space interpretation. The model can be formulated through a measurement equation for the observed yield curve, written as

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

and the transition equations, which represents the latent factors dynamics:

$$\begin{pmatrix} \beta_{1t} \\ \beta_{2t} \\ \beta_{3t} \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix} + \phi \begin{pmatrix} \beta_{1t-1} \\ \beta_{2t-1} \\ \beta_{3t-1} \end{pmatrix} + \epsilon_{\beta_t} \tag{2}$$

where $y_t(\tau)$ represents the yield vector, the parameters β_{it}, $i = 1, 2, 3$, are the latent factors with interpretations of level, slope and curvature, respectively, and λ is the decay parameter. Finally, $\epsilon_t(\tau)$ and ϵ_{β_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 (2006). 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 monthly yield curve 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. In order 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, Rudebusch and Aruoba (2006).

2.1 Long Memory Dynamics in DNS framework

2.1.1 Autoregressive fractionally integrated moving average (ARFIMA) model

The AR model often used to determine the evolution of the latent factors cannot capture the highly persistence of interest rates. Alternatively, the evolution of the latent factors can be modeled as a stationary process with long memory. According to Zivot and Wang (2007), a stationary process y_t has long memory if its autocorrelation function satisfies

$$\gamma(h) \sim C_1 h^{H-1} \tag{3}$$

as $k \to \infty$ and C_1 is a positive constant. Therefore, the autocorrelation function of a long memory process decays slowly at a hyperbolic rate, i.e., so slowly that are not summable:

$$\sum_{k=-\infty}^{\infty} \gamma(h) = \infty. \tag{4}$$

An approach is to model the DNS latent factors as an autoregressive fractionally integrated moving average (ARFIMA) model. According to Palma (2007), an ARFIMA $\{y_t\}$ process can be defined by

$$\phi(B)y_t = \theta(B)(1-B)^{-d}\epsilon_t, \tag{5}$$

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

$$(1-B)^{-d} = \sum_{j=0}^{\infty} \eta_j B^j = \eta(B) \tag{6}$$

where

$$\eta_j = \frac{\Gamma(j+d)}{\Gamma(j+1)\Gamma(d)}, \tag{7}$$

for $d < 1/2$, $d \neq 0, -1, -2, \ldots$, and $\{\epsilon_t\}$ is a white noise process with finite variance.

It is important to note that long memory processes are not Markovian, thus, all the state space representations are infinite dimensional and, therefore the estimation is not straightforward.

2.1.2 Fractional Gaussian noise

An alternative way of introducing long persistence is through the use of fractional Gaussian noise. Although conceptually different, ARFIMA and fGn are related processes, especially when the autoregressive and the moving average order of ARFIMA are zero. The use of a fGn is interesting in the context of state space modeling since it is possible to obtain a Markovian representation for this process, which allows one-step estimation of the DNS model with long memory. Following Sørbye, Myrvoll-Nilsen and Rue (2017) is possible to approximate fGn model by an aggregated model of a few AR(1) components and take advantages of this approach.

The evolution of the DNS latent factors with fGn dynamics is given by:

$$\beta_{it} = fGn(H, \tau), \quad i = 1, 2, 3, \tag{8}$$

where fGn is a fractional Gaussian noise defined by its autocorrelation function

$$\gamma(h) = \frac{1}{2}(|h+1|^{2H} - 2|h|^{2H} + |h-1|^{2H}), \quad h = 0, \ldots, n-1, \tag{9}$$

where $H \in (0,1)$ is referred to as the Hurst exponent (or self-similarity parameter) and τ denotes the 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.

Sørbye, Myrvoll-Nilsen and Rue (2017) propose to approximate fGn model with a weighted sum of independent first-order autoregressive process, which can be represented as a Gaussian Markov Random Field. In their approach, they propose to fit the weights and the coefficients of the approximation to mimic the autocorrelation function (9) of a fGn.

Consider m independent AR(1) process

$$z_{j,t} = \phi z_{zj,t-1} + \nu_{j,t}, \quad j = 1, \ldots, m, \quad t = 1, \ldots, n, \tag{10}$$

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 $\sigma^2_{\nu,j} = 1 - \phi_j^2$. Define the cross-sectional aggregation of the m processes as

$$\bar{x}_m = \sigma \sum_{j=1}^{m} \sqrt{w_j} z^{(j)}, \tag{11}$$

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 (11) is given by

$$\gamma_{\bar{x}_m}(h) = \sum_{j=1}^{m} w_j \phi_j^{|h|}, \quad h = 0, 1, \ldots, n-1. \tag{12}$$

The idea proposed by Sørbye, Myrvoll-Nilsen and Rue (2017) is to fit the weights $\boldsymbol{w} = \{w_j\}_{j=1}^m$ and the autocorrelation coefficients $\boldsymbol{\phi} = \{\phi_j\}_{j=1}^m$ in (11) to match the autocorrelation function of fGn process as in (9). In order to obtain values of $(\boldsymbol{w}, \boldsymbol{\phi})$, the weighted squared error is minimized

$$(\boldsymbol{w}, \boldsymbol{\phi})_H = \underset{(\boldsymbol{w},\boldsymbol{\phi})}{\operatorname{argmin}} \sum_{h=1}^{h_{\max}} \frac{1}{h} (\gamma_{\bar{x}_m}(h) - \gamma_x(h))^2, \tag{13}$$

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, proposed by Rue, Martino and Chopin (2009), provides accurate and efficient approximations on Bayesian hierarchical models that can be represented as a latent Gaussian model (LGN), focuses on models known as Gaussian Markov random field (GMRF).

A latent GMRF model is a hierarchical model with the first stage defining a distributional assumption for the observed variable y, usually assumed to be conditionally independent given the latent factors x and some additional parameter θ

$$\pi(y|x, \theta) = \prod_j \pi(y_j|x_j, \theta), \quad j \in J \tag{14}$$

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

$$x_i = \text{Offset}_i + \sum_{k=0}^{\eta_f - 1} \omega_{ki} f_k(c_{ki}) + z_i^T \beta + \epsilon_i, \quad i = 0, \ldots, \eta_x - 1 \tag{15}$$

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; 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

$$\pi(x_i|Y) = \int \pi(x_i|\theta, Y)\pi(\Theta, Y) d\theta \tag{16}$$

and the marginal posterior distribution of hyperparameters, given by

$$\pi(\theta_j|Y) = \int \pi(\theta|Y) d\theta_{\text{-}j} \tag{17}$$

where $\theta_{\text{-}j}$ denotes the vector θ without its jth element. In practice, the INLA method can be implemented in three steps. The first step is to obtain an approximation to the full posterior

[1] See Sørbye, Myrvoll-Nilsen and Rue (2017) for more details on the implementation.

distribution $\pi(\theta|y)$ by Laplace approximation. The second step is to find an approximation to the full conditional distributions $\pi(x_i|\theta, y)$ for particular values of θ. The last step is to obtain the marginal posterior distributions in (16) and (17) by combining the two approximations in the previous steps and integrating out the irrelevant factors[2].

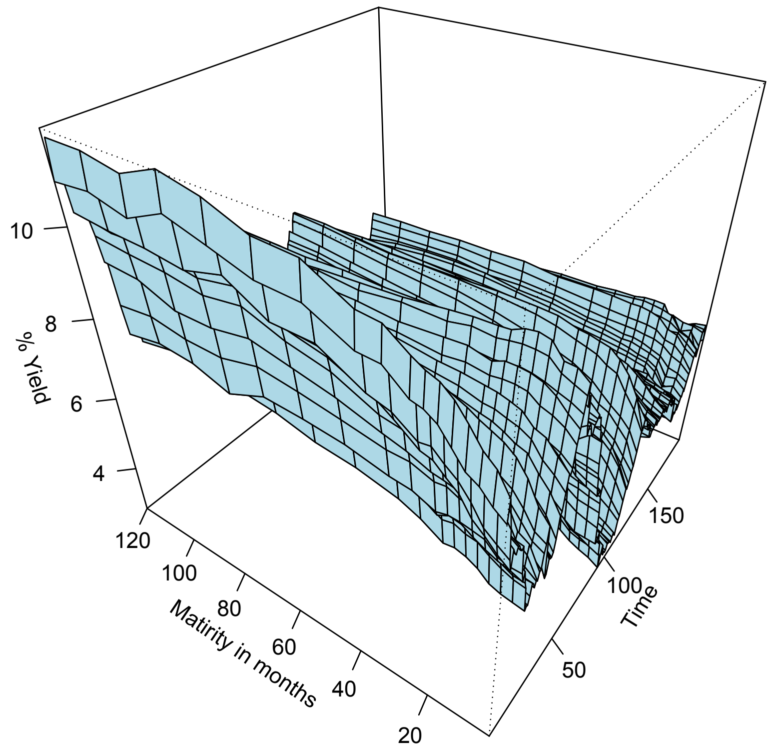
In this work we use an estimation structure generalizing the method of Laurini and Hotta (2014), which propose the use of INLA to estimate and forecast the term structure of interest rates. In the same way as this work, we use an additive structure where the decay λ parameter is kept fixed and determined before the estimation process. The use of a fixed λ parameter is necessary due to the need for an additive structure in the measurement equation. To directly measure the impact of the long memory structure on predictions for the term structure of Treasury Bonds we first use the same λ value used in Diebold and Li (2006), and after the generalized cross validation procedure proposed in Laurini and Hotta (2014) to obtain the optimal λ parameter. The analysis for the DI Futures contracts is based on the optimal choice λ by generalized cross validation.

4 Results

4.1 Fama-Bliss US Treasury zero-coupon data - Full Sample Analysis

In order to compare the DNS with long memory and the AR(1) specification proposed by Diebold and Li (2006), 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.

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



[2] See Rue, Martino and Chopin (2009) for more details.

Assuming that the parameter λ is constant and known, we follow Diebold and Li (2006), assuming the value of λ being equal to 0.0609. Under this assumption, we estimate a DNS model with long memory, as described in section 2, using the INLA method. Following Sørbye, Myrvoll-Nilsen and Rue (2017), we adopt three independent AR(1) process to approximate the fGn model. According to equation (8), the fGn process has two parameters, τ and H. The marginal precision, τ, is represented as

$$\tau = \exp(\theta_1) \tag{18}$$

and the Hurst parameter H is represented as

$$H = \frac{1}{2} + \frac{1}{2}\left(\frac{\exp(\theta_2)}{1+\exp(\theta_2)}\right) \tag{19}$$

and the prior is defined on the hyperparameters θ_1 and θ_2. Specifically

$$\theta_1 \sim pcprior(3, 0.01)$$
$$\theta_2 \sim pcpriorforfGn(0.9, 0.1).$$

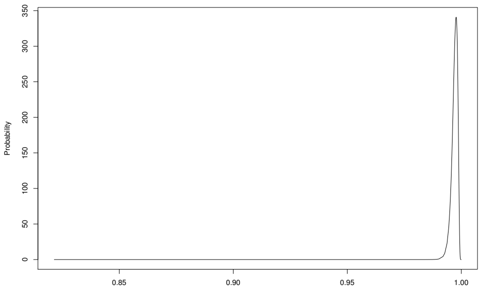
where pc prior refers to the family of penalized complexity priors for the fGn process introduced in Sørbye and Rue (2018).

Table 1 contains results about the posterior distributions obtained for the precision of the Gaussian observations, the Hurst exponent H and the precision for the latent factors. The mean of the parameters H for β_{1t}, β_{2t} and β_{3t} lie inside 1/2 and 1, consistent with persistent patterns often found in financial time series. In addition, we created 95% credibility intervals, based on the empirical quantiles of 0.025 and 0.975 of the posterior distributions, also reported in Table 1. Finally, Figures 2 and 3 shows the posterior distributions of Hurst exponent and precision parameters of latent factors, respectively.

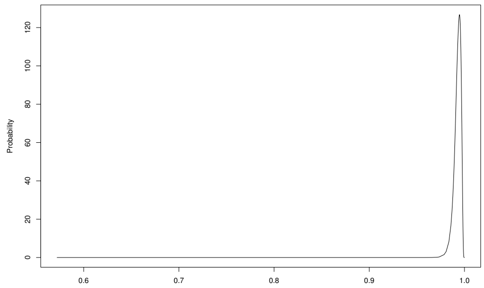
Table 1 – Summary Hyperparameters

	Mean	SD	0.025 quant.	0.5 quant.	0.975 quant.	Mode
Precision τ for Gaussian obs.	197.788	5.369	187.501	197.697	208.589	197.494
Precision τ_1 for β_{1t}	0.084	0.037	0.030	0.078	0.175	0.065
H for β_{1t}	0.997	0.001	0.994	0.997	0.999	0.998
Precision τ_2 for β_{2t}	0.153	0.075	0.049	0.140	0.336	0.112
H for β_{2t}	0.993	0.004	0.984	0.993	0.998	0.995
Precision τ_3 for β_{3t}	0.155	0.072	0.053	0.142	0.331	0.117
H for β_{3t}	0.979	0.010	0.955	0.981	0.994	0.985

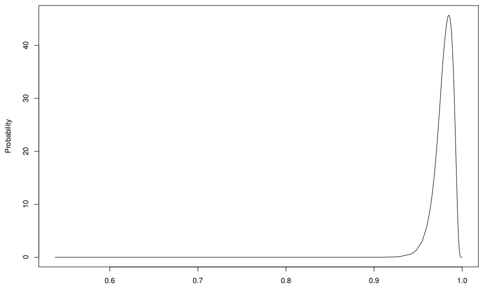
Figure 2 – Posterior distribution of Hurst exponents. (a) H(β_{1t}); (b) H(β_{2t}); (c) H(β_{3t})



(a) H(β_{1t})

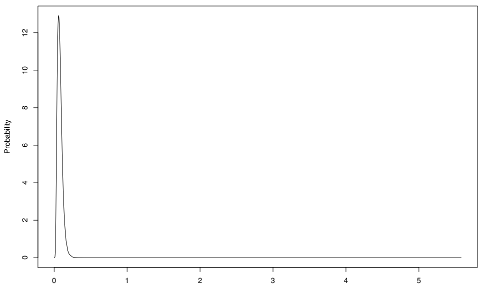


(b) H(β_{2t})

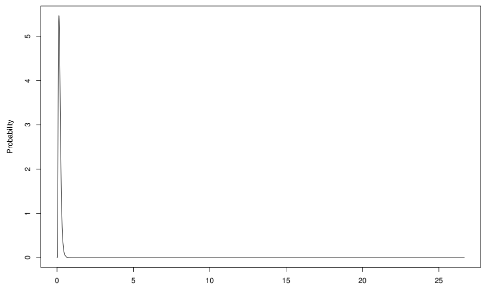


(c) H(β_{3t})

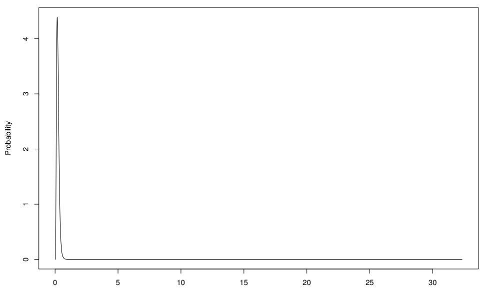
Figure 3 – Posterior distribution of precision parameters. (a) τ_1 precision β_{1t}; (b) τ_2 precision β_{2t}; (c) τ_3 precision β_{3t}



(a) τ_1 precision for β_{1t}



(b) τ_2 precision for β_{2t}



(c) τ_3 precision for β_{3t}

4.2 Out-of-sample Forecasts

We compare out-of-sample forecasting performance of the DNS model considering three different specifications for the latent factors: AR(1), ARFIMA(0,d,0) and fGn. The models will be denominated DNS-AR(1), DNS-ARFIMA and DNS-fGn henceforward. We assess out-of-sample 1-month, 6-months and 12-months ahead forecasting for maturities of 3 months and 1, 3, 5 and 10 years.

The DNS-fGn model is estimated Bayesian using INLA, following the methodology proposed in Laurini and Hotta (2014) and its modification to the fGn process proposed in our article. The estimation of the ARFIMA model is based on a modification of the two-step estimation method proposed in Diebold and Li (2006), where in the case we used an ARFIMA process in the second estimation step, where the dynamics of the time series of Betas estimated in the first step (cross-section in each day) is modeled by an ARFIMA process estimated by the maximum-likelihood method based on the Whittle estimator.

Tables 2, 3 and 4 report the RMSE for each of the forecasting models, and each of the forecasting horizons. Moreover, following the same structure as Diebold and Li (2006), Tables 2, 3 and 4 shows the mean, standard deviation and autocorrelations of forecast errors.

Table 2 – Out of Sample 1-Month Ahead Forecasting Results

	Mean	Std.Dev	RMSE	ACF1	ACF12
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 year	0.049	0.291	0.293	0.700	-0.214
3 years	-0.086	0.333	0.342	0.648	-0.213
5 years	-0.156	0.343	0.375	0.661	-0.246
10 years	-0.169	0.310	0.351	0.623	-0.213
Dynamic Nelson-Siegel with ARFIMA(0,d,0) Factor Dynamics					
3 months	0.005	0.224	0.222	0.714	-0.059
1 year	0.039	0.320	0.321	0.752	-0.202
3 years	-0.108	0.362	0.376	0.708	-0.212
5 years	-0.182	0.371	0.411	0.720	-0.248
10 years	-0.197	0.335	0.387	0.692	-0.218

Table 3 – Out of Sample 6-Month Ahead Forecasting Results

	Mean	Std.Dev	RMSE	ACF6	ACF18
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
1 year	0.043	0.666	0.663	0.225	-0.189
3 years	-0.242	0.740	0.774	0.091	-0.215
5 years	-0.409	0.746	0.846	0.106	-0.260
10 years	-0.533	0.666	0.850	0.117	-0.252
Dynamic Nelson-Siegel with ARFIMA(0,d,0) Factor Dynamics					
3 months	0.068	0.553	0.554	0.457	-0.242
1 year	0.026	0.676	0.673	0.240	-0.177
3 years	-0.268	0.745	0.787	0.099	-0.207
5 years	-0.436	0.747	0.861	0.111	-0.255
10 years	-0.560	0.665	0.866	0.125	-0.250

Table 4 – Out of Sample 12-Month Ahead Forecasting Results

	Mean	Std.Dev	RMSE	ACF12	ACF24
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	0.660	0.656	-0.145	-0.130
1 year	-0.064	0.698	0.696	-0.230	-0.088
3 years	-0.439	0.753	0.867	-0.302	-0.058
5 years	-0.660	0.767	1.008	-0.321	-0.077
10 years	-0.851	0.692	1.094	-0.332	-0.121
Dynamic Nelson-Siegel with ARFIMA(0,d,0) Factor Dynamics					
3 months	0.029	0.675	0.671	-0.121	-0.133
1 year	-0.085	0.707	0.707	-0.210	-0.091
3 years	-0.470	0.756	0.886	-0.288	-0.059
5 years	-0.693	0.766	1.029	-0.311	-0.079
10 years	-0.885	0.689	1.118	-0.320	-0.124

According to Table 2, our empirical evidence suggests that when a short forecasting horizon is considered, the DNS-AR factor dynamics performs more accurate forecasts compared to the ARFIMA and fGn specifications, for all maturities. In addition, DNS-fGn has better performance in relation to DNS-ARFIMA, for all maturities.

When longer forecasting horizons are considered (6-months and 12-months ahead), the DNS-fGn are competitive and, in some dimensions, superior to DNS-AR and DNS-ARFIMA. For 6-months ahead, the DNS-fGn shows the lowest RMSE for shorter maturities, whereas, DNS-AR forecasts are more accurate for maturities higher than 3 years. Results are shown in Table 3. However, considering 12-months ahead forecasting, the DNS-fGn performance is clearly superior than DNS-AR and DNS-ARFIMA, for all maturities. In addition, DNS-ARFIMA forecasts are also more accurate than those from DNS-AR, for all maturities. Table 4 reports the results.

In summary, our results suggest that when a short horizon is considered, a persistence component on DNS model do not provide forecasting gains, i.e., the DNS-AR forecasts are more accurate than those from DNS-fGn and DNS-ARFIMA. However, when longer forecasting horizons are considered, both, DNS-fGn and DNS-ARFIMA allows accurate forecasts.

4.3 Optimal selection of λ parameter

As discussed in Laurini and Hotta (2014), the use of the INLA approach is based on a linear additive structure in the state space form, and similarly to the methodology of Diebold and Li (2006), we must assume that the parameter λ is known to perform the estimation of the dynamic Nelson-Siegel model using this methodology. To circumvent this limitation, Laurini and Hotta (2014) propose the use of a generalized cross-validation mechanism to determine the optimal value of the λ parameter.

We repeat here the fundamentals of this procedure. The idea is to choose the parameter λ as an optimization problem by finding the value of λ which maximizes some measure of model fit. In this work we use as a measure the cross-validated log-score of the model. We can calculate measures of generalized cross validation using INLA without having to re-estimate the model for each observation left out in the cross validation procedure. The conditional predictive ordinates (CPO) relative to the observation y_i, conditional on the full sample, except for this observation, is given by:

$$CPO_i = \pi(y_i^{obs}|y_{-1}), \tag{20}$$

and the CPO of observation i is obtained as:

$$CPO_i = \int \pi(y_i^{obs}|y_{-1}, \theta)\pi(\theta|y_{-i})d\theta \tag{21}$$

with the first term in this integral represented by:

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

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

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

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 - $\lambda = 0.1482749$

	Mean	Std.Dev	RMSE	ACF12	ACF24
Out of Sample 1-Month Ahead Forecasting Results					
3 months	0.016	0.211	0.211	0.653	-0.072
1 year	0.054	0.290	0.294	0.685	-0.192
3 years	-0.112	0.324	0.341	0.645	-0.215
5 years	-0.164	0.340	0.375	0.662	-0.238
10 years	-0.150	0.327	0.358	0.649	-0.174
Out of Sample 6-Month Ahead Forecasting Results					
3 months	0.064	0.531	0.531	0.439	-0.260
1 year	0.062	0.676	0.675	0.215	-0.178
3 years	-0.291	0.725	0.777	0.097	-0.227
5 years	-0.429	0.738	0.849	0.111	-0.266
10 years	-0.492	0.687	0.841	0.111	-0.232
Out of Sample 12-Month Ahead Forecasting Results					
3 months	0.015	0.646	0.641	-0.139	-0.137
1 year	-0.036	0.707	0.703	-0.237	-0.082
3 years	-0.500	0.750	0.897	-0.297	-0.067
5 years	-0.686	0.767	1.025	-0.317	-0.082
10 years	-0.796	0.702	1.058	-0.339	-0.106

The use of this optimal selection procedure leads to forecasting gains, using the RMSE metric, for some forecasting maturities and horizons. For example, we obtain improvements in this criterion for 9 of the 15 combinations of forecast-maturity horizons analyzed by Dynamic Nelson-Siegel with fGn model. In the other comparisons the value of λ proposed by Diebold and Li (2006) seems to be robust in terms of predictive performance.

4.4 Brazilian DI Futures Term Structure

We performed a second application of the proposed model, using the term structure of Interbank Deposits (DI) Futures contracts in Brazil. This is the most important market of interest curves in Brazil, and is the main reference of term structure. This application is particularly interesting since the interest rate market in Brazil is subject to several regime changes and possible structural breaks, and in this case the use of a long memory structure can be a parsimonious way of constructing forecasts in this context.

We used in this analysis monthly data for the period January 2003 through August 2017 (176 observations) constructed from the daily data of DI operations at B3, using the end-of-period value as a month observation. The data were interpolated using cubic spline for a set of 17 maturities ranging from 1 to 72 months. Figure 4 shows the evolution of the interest curve used in this study, and the table 6 shows some descriptive statistics of this data set.

Figure 4 – DI Futures - January 2003 - August 2017

Table 6 – Descriptive Statistics - DI Futures - January 2003 - August 2017

	Mean	Std.Dev	Max	Min
1 M	12.998	4.143	26.819	6.973
3 M	12.976	4.100	27.577	7.035
6 M	12.979	4.020	28.246	7.066
9 M	13.012	3.966	28.664	7.084
12 M	13.078	3.939	29.328	7.148
15 M	13.157	3.915	30.264	7.224
18 M	13.235	3.896	31.180	7.397
21 M	13.306	3.893	32.168	7.593
24 M	13.376	3.905	33.236	7.761
27 M	13.436	3.931	34.184	7.908
30 M	13.486	3.967	35.095	7.998
33 M	13.533	4.007	35.931	8.097
36 M	13.578	4.047	36.733	8.202
40 M	13.640	4.118	37.993	8.333
48 M	13.725	4.245	39.650	8.448
60 M	13.717	4.088	40.341	8.683
72 M	13.906	4.531	40.725	8.812

We performed a prediction experiment similar to that performed with the data from Fama-Bliss Treasury Bonds data, making forecasts 1, 6 and 12 months-ahead. We used as the initial sample of estimation the period from 2003-01 to 2014-09 (140 months), and from that date increased one month at time, constructing the forecasts, and similarly followed this procedure until the end of the sample. For space issues we show only the predictive performance measures for using Dynamic Nelson-Siegel with AR(1) factor dynamics estimated by OLS and the Bayesian model using fGn as described in the methodology section. We used a λ value of 0.07317 for the two estimation methodologies, which is determined by the cross-validation criterion described in the 4.3 section.

Table 7 – Out of Sample 1-Month Ahead Forecasting Results - DI Futures

	Mean	Std.Dev	RMSE	ACF1	ACF12
Dynamic Nelson-Siegel with AR(1) Factor Dynamics					
3 months	0.133	0.425	0.439	0.768	-0.043
9 months	0.095	0.898	0.890	0.828	0.020
27 months	-0.050	0.817	0.807	0.646	-0.021
33 months	-0.042	0.796	0.785	0.614	-0.024
72 months	-0.100	0.850	0.844	0.607	-0.027
Dynamic Nelson-Siegel with fGn Factor Dynamics					
3 months	0.097	0.678	0.676	0.875	-0.043
9 months	-0.025	1.076	1.061	0.874	0.020
27 months	-0.205	0.959	0.967	0.770	-0.051
33 months	-0.190	0.936	0.942	0.755	-0.070
72 months	-0.196	0.961	0.967	0.751	-0.118

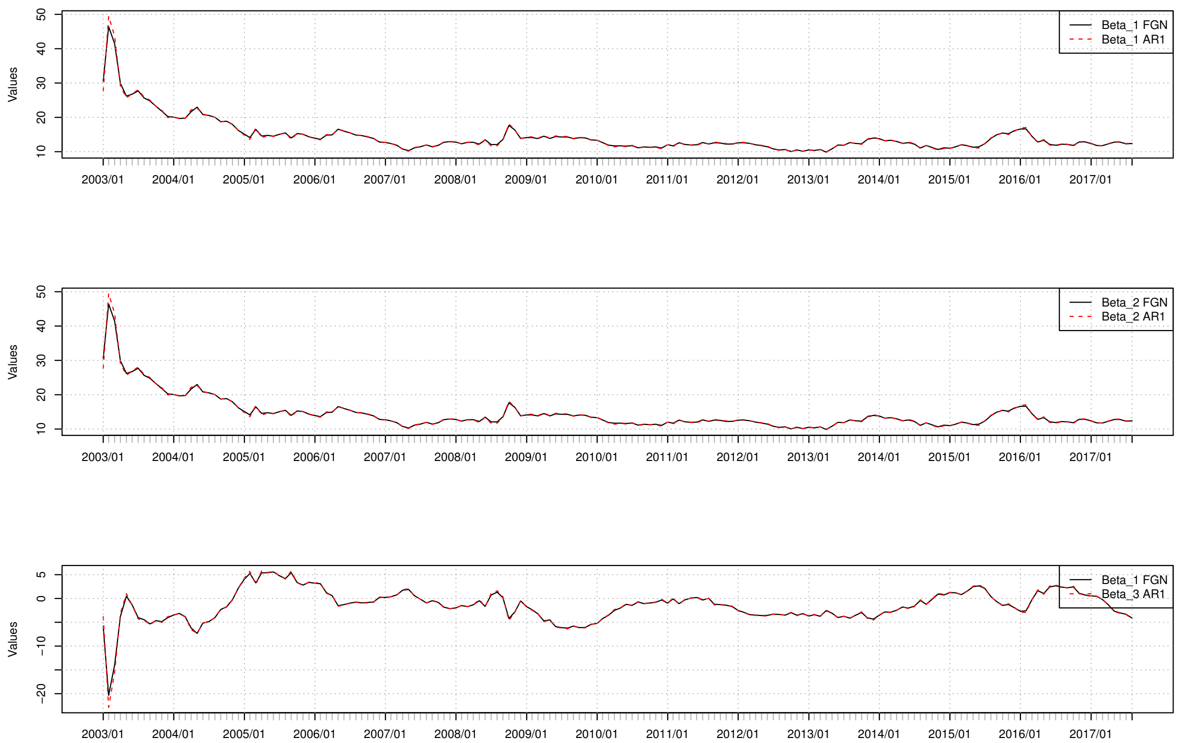
Table 8 – Out of Sample 6-Month Ahead Forecasting Results - DI Futures

	Mean	Std.Dev	RMSE	ACF6	ACF18
Dynamic Nelson-Siegel with AR(1) Factor Dynamics					
3 months	0.617	1.701	1.783	0.223	-0.258
9 months	0.144	2.269	2.235	0.347	-0.410
27 months	-0.199	2.295	2.265	0.263	-0.426
33 months	-0.183	2.270	2.240	0.213	-0.406
72 months	-0.126	2.169	2.136	0.097	-0.346
Dynamic Nelson-Siegel with fGn Factor Dynamics					
3 months	0.436	2.017	2.031	0.278	-0.297
9 months	-0.116	2.507	2.467	0.396	-0.426
27 months	-0.519	2.436	2.450	0.345	-0.445
33 months	-0.509	2.395	2.409	0.303	-0.426
72 months	-0.447	2.253	2.260	0.197	-0.365

Table 9 – Out of Sample 12-Month Ahead Forecasting Results - DI Futures

	Mean	Std.Dev	RMSE	ACF12	ACF24
Dynamic Nelson-Siegel with AR(1) Factor Dynamics					
3 months	1.324	2.161	2.496	-0.311	
9 months	0.464	2.972	2.946	-0.297	
27 months	-0.004	3.105	3.039	-0.316	
33 months	0.010	3.055	2.991	-0.322	
72 months	0.030	2.766	2.708	-0.348	
Dynamic Nelson-Siegel with fGn Factor Dynamics					
3 months	0.577	2.373	2.393	-0.223	
9 months	-0.340	3.026	2.981	-0.255	
27 months	-0.799	3.050	3.091	-0.296	
33 months	-0.769	2.993	3.029	-0.304	
72 months	-0.658	2.710	2.734	-0.328	

Figure 5 – Estimated Latent Factors - DNS-AR(1) and DNS-fGn - DI-Futures - January 2003 - August 2017



Tables 7-9 show the predictive performance measures for this data set. In a similar way to the results obtained with the data of Treasury Bonds, the results show that the methodology based on the AR(1) dynamics for latent factors and OLS estimation has better performance for short maturities and short horizons (1 and 6 months) while the Bayesian method using the fractional Gaussian noise structure for latent factors presents a superior performance in terms of RMSE for the 12-month forecast horizon and the longest maturities in the sample (27, 33 and 72 months).

To show a comparison between latent factors estimated using the two methods, we show in Figure 5 the estimated factors using the complete sample of DI data. We can observe that the estimated factors are very similar, and the main differences occur in the construction of the forecasts outside the sample of these factors.

Conclusions

In this paper we propose a different specification for the factor dynamics of the Dynamic Nelson-Siegel model in order to embody long range persistence into the model. The main goal is to assess if a long memory component helps to obtain more accurate forecasts of the term structure of interest rates. The evolution of the latent factors are given by a fractional Gaussian noise, which is approximated with a weighted sum of independent first-order autoregressive process, that can be represented as a Gaussian Markov Random Field, and through this approach, Integrated Nested Laplace Approximations method was used to obtain the Bayesian estimation of the parameters.

We compare our model with alternative specifications for the latent factors in the analysis of the Fama-Bliss dataset: first order autoregressive model and autoregressive fractionally integrated moving average model. For 3-months ahead, the DNS model with AR(1) factor dynamics outperform DNS with fGn and DNS with ARFIMA(0,d,0), for all maturities. However, for 6-months ahead, the fGn specification shows the lowest RMSE for shorter maturities, whereas, DNS with AR(1) forecasts are more accurate for maturities higher than 3 years. Finally, considering 12-months ahead forecasting, the DNS with fGn dynamics presents good forecasting properties, superior to the AR(1) and ARFIMA(0,d,0) specifications, for all maturities. The analysis for the DI Futures term structure show similar results, with a better forecasting performance for the DNS-AR(1) for short forecasting horizons and a more favorable performance to the DNS-fGn model in the longer forecasting horizon and maturities. Therefore, we provide evidence that long memory is indeed helpful for longer forecasting horizons.

Bibliography

BACKUS, D. K.; ZIN, S. E. Long-memory inflation uncertainty: Evidence from the term structure of interest rates. *Journal of Money, Credit and Banking*, [Wiley, Ohio State University Press], v. 25, n. 3, p. 681–700, 1993. ISSN 00222879, 15384616. Available at: <http://www.jstor.org/stable/2077735>.

CARRIERO, A.; KAPETANIOS, G.; MARCELLINO, M. Forecasting government bond yields with large Bayesian vector autoregressions. *Journal of Banking & Finance*, Elsevier, v. 36, n. 7, p. 2026–2047, 2012.

CHRISTENSEN, J. H.; DIEBOLD, F. X.; RUDEBUSCH, G. D. The affine arbitrage-free class of Nelson–Siegel term structure models. *Journal of Econometrics*, Elsevier, v. 164, n. 1, p. 4–20, 2011.

COCHRANE, J. H.; PIAZZESI, M. Bond risk premia. *American Economic Review*, v. 95, n. 1, p. 138–160, 2005.

COUCHMAN, J.; GOUNDER, R.; SU, J.-J. Long memory properties of real interest rates for 16 countries. *Applied Financial Economics Letters*, Routledge, v. 2, n. 1, p. 25–30, 2006.

DIEBOLD, F. X.; LI, C. Forecasting the term structure of government bond yields. *Journal of Econometrics*, Elsevier, v. 130, n. 2, p. 337–364, 2006.

DIEBOLD, F. X.; LI, C.; YUE, V. Z. Global yield curve dynamics and interactions: a dynamic Nelson–Siegel approach. *Journal of Econometrics*, Elsevier, v. 146, n. 2, p. 351–363, 2008.

DIEBOLD, F. X.; RUDEBUSCH, G. D. *Yield Curve Modeling and Forecasting: the Dynamic Nelson-Siegel Approach*. [S.l.]: Princeton University Press, 2013.

DIEBOLD, F. X.; RUDEBUSCH, G. D.; ARUOBA, S. B. The macroeconomy and the yield curve: a dynamic latent factor approach. *Journal of Econometrics*, Elsevier, v. 131, n. 1-2, p. 309–338, 2006.

DUFFEE, G. R. Term premia and interest rate forecasts in affine models. *The Journal of Finance*, Wiley Online Library, v. 57, n. 1, p. 405–443, 2002.

FAMA, E. F.; BLISS, R. R. The information in long-maturity forward rates. *The American Economic Review*, JSTOR, p. 680–692, 1987.

GOLIńSKI, A.; ZAFFARONI, P. Long memory affine term structure models. *Journal of Econometrics*, v. 191, n. 1, p. 33 – 56, 2016. ISSN 0304-4076.

KOOPMAN, S. J.; MALLEE, M. I.; WEL, M. Van der. Analyzing the term structure of interest rates using the dynamic Nelson–Siegel model with time-varying parameters. *Journal of Business & Economic Statistics*, Taylor & Francis, v. 28, n. 3, p. 329–343, 2010.

LAURINI, M. P.; HOTTA, L. K. Forecasting the term structure of interest rates using integrated nested Laplace approximations. *Journal of Forecasting*, Wiley Online Library, v. 33, n. 3, p. 214–230, 2014.

NELSON, C. R.; SIEGEL, A. F. Parsimonious modeling of yield curves. *Journal of business*, JSTOR, p. 473–489, 1987.

PALMA, W. *Long-memory Time Series: Theory and Methods*. [S.l.]: John Wiley & Sons, 2007.

POOTER, M. D. Examining the Nelson-Siegel class of term structure models: In-sample fit versus out-of-sample forecasting performance. *Available at SSRN 992748*, 2007.

POOTER, M. D.; RAVAZZOLO, F.; DIJK, D. J. V. Term structure forecasting using macro factors and forecast combination. *FRB International Finance Discussion Paper*, n. 993, 2010.

RUE, H.; MARTINO, S.; CHOPIN, N. Approximate bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, Wiley Online Library, v. 71, n. 2, p. 319–392, 2009.

SØRBYE, S. H.; MYRVOLL-NILSEN, E.; RUE, H. An approximate fractional Gaussian noise model with $\mathcal{O}(n)$ computational cost. *Statistics and Computing*, Springer, p. 1–13, 2017.

SØRBYE, S. H.; RUE, H. Fractional Gaussian noise: Prior specification and model comparison. *Environmetrics*, v. 29, n. 5-6, p. e2457, 2018.

TSAY, W.-J. Long memory story of the real interest rate. *Economics Letters*, v. 67, n. 3, p. 325 – 330, 2000. ISSN 0165-1765.

ZIVOT, E.; WANG, J. *Modeling Financial Time Series with S-Plus®*. [S.l.]: Springer Science & Business Media, 2007.