Forecasting the term structure of interest rates using Integrated Nested Laplace Approximations

Márcio Poletti Laurini[a,*], Luiz Koodi Hotta[b]

[a] *IBMEC Business School -Av. Presidente Wilson 118, 20030-020, Rio de Janeiro, RJ. Brazil. Tel (55) (21) 503-4146. fax: (55) (21) 4503-4168.*
[b] *IMECC-Unicamp*

Abstract

This article discuss the use of Bayesian methods for inference and forecasting in dynamic term structure models through Integrated Nested Laplace Approximations (INLA). This method of analytical approximations allows for accurate inferences for latent factors, parameters and forecasts in dynamic models with reduced computational cost. In the estimation of dynamic term structure models it also avoids some simplifications in the inference procedures, as the estimation in two stages.

The results obtained in the estimation of the dynamic Nelson-Siegel model indicate that this methodology performs more accurate out-of-sample forecasts compared to the methods of two-stage estimation by OLS and also Bayesian estimation methods using MCMC. These analytical approaches also allow calculating efficiently measures of model selection such as generalized cross validation and marginal likelihood, that may be computationally prohibitive in MCMC estimations.
Keywords: Term Structure, Latent Factors, Bayesian Forecasting, Laplace Approximations.

Resumo - Discutimos neste artigo o uso de inferência Bayesiana em modelos dinâmicos de taxas de juros através de Integrated Nested Laplace Approximations (INLA). Este método de aproximação analítica permite obter inferências acuradas para fatores latentes e parâmetros em modelos dinâmicos com um custo computacional reduzido,e na estimação de modelos dinâmicos de estrutura a termo de taxas de juros permite evitar o uso de algumas simplificações nos priocedimentos de inferência como a estimação em dois estágios.

Os resultados obtidos na estimação do modelo dinâmico de Nelson-Siegel indicam que esta metodologia permite obter previsões mais precisas para horizontes de previsão fora da amostra em comparação aos métodos de estimação em dois estágios por OLS e também a métodos de estimação Bayesiana utilizando MCMC. Estas aproximações analíticas também permitem calcular de forma eficiente e precisa medidas de comparação de modelos como validação cruzada generalizada e verossimilhança marginal, cujo custo computacional pode ser proibitivo em modelos baseados em Markov Chain Monte Carlo. Como uma aplicação do uso destas medidas comparativas mostramos que assumir o parâmetro de decaimento constante no modelo de Nelson-Siegel é uma restrição válida.
Palavras Chave: Estrutura a Termo, Fatores Latentes, Previsão Bayesiana, Aproximações de Laplace.

JEL Codes: C11, G12, G17.

Área de Submissão - Finanças.

---

*Corresponding author
Email addresses: marcio.laurini@ibmecrj.br (Márcio Poletti Laurini), hotta@ime.unicamp.br (Luiz Koodi Hotta)

1. Introduction

Procedures for inference and forecasting in dynamic term structure models usually are computationally complex because of the presence of latent factors and the high dimensionality of the parameter vector. The estimation methodology used in the procedure may significantly affect the performance of the estimated model, especially in the out-of-sample forecasts.

As an example of this problem, Yu & Zivot (2010) evaluated the predictive performance of the dynamic formulation of the Nelson-Siegel model proposed by Diebold & Li (2006) in forecasting yields of U.S. Treasury and corporate bonds, and they concluded that different ways of estimating the same model can affect dramatically the predictive results for the various maturities and classes of assets studied. The estimation in a single step of this model using a state space formulation using the Kalman Filter has better results for the prediction of high yield bonds over short horizons of time, while the simplest methodology using OLS in two steps has a superior performance for bonds with investment-grade and shorter time horizons, confirming that the estimation methodology used is not only important for the procedures of inference and hypothesis testing but can significantly influence the construction of forecasts.

The aim of this article is to compare the predictive performance of a Bayesian estimation methodology based on Integrated Nested Laplace Approximations (INLA) for the estimation of latent factors, parameters and forecasts in the dynamic version of the Nelson-Siegel model proposed in Diebold & Li (2006), in relationship to the two steps procedure of Diebold & Li (2006) and to the estimation methods using Bayesian Markov Chain Monte Carlo (MCMC), e.g. Migon & Abanto-Valle (2007), Laurini & Hotta (2010) and Hautsch & Yang (2010). The INLA methodology, proposed in Rue *et al.* (2009), gives accurate analytical approximations for the posterior distribution of latent factors and parameters in generalized Gaussian models without the need for procedures such as numerical simulation methods based on MCMC.

Forecasts constructed using Bayesian methods have several advantages over frequentists methods. As stated in Geweke & Amisano (2010), Bayesian estimates incorporate uncertainty of parameters in a consistent manner, avoiding all approximations used in Diebold & Li (2006) for the estimations of parameters, latent factors and out-of-sample forecasts[1].

In particular, these computational advantages in the INLA methodology allow us to test the hypothesis of a fixed decay parameter assumed in Diebold & Li (2006) to linearize the model by comparing with estimates assuming that this parameter is time varying. Another advantage of this methodology is that it gives accurate credibility intervals for the observed sample, in contrast to the method of Diebold & Li (2006) which do not control for the estimation in two steps, and also overcomes the limitations generated by the use of Monte Carlo simulation methods in Bayesian analysis using MCMC, such as problems of convergence of Markov chains and limited rate of convergence of Monte Carlo approximations, whose accuracy is limited by the $\sqrt{n}$ rate of convergence of Monte Carlo procedures. Besides, the higher accuracy obtained, this approach is computationally efficient and allows to efficiently calculate measures of comparison of models such as the marginal likelihood and cross validated log score of the model, overcoming some difficulties in the existing computational methods of Bayesian estimation based on Monte Carlo simulations.

The results indicate that the INLA methodology allows forecast gains over the methodologies based on the classical two-step estimation by ordinary least squares and the Bayesian estimation using MCMC, especially for longer maturities of the yield curve, with a computational time much lower than the MCMC methods and almost equivalent to OLS in two steps. The results also confirm that for the sample used in this study, the same database analyzed in Diebold & Li (2006), assuming a constant decay parameter in the formulation of the Nelson-Siegel model does not decrease dramatically the out-of-sample predictive performance.

This paper is organized as follows - section 2 briefly summarize the INLA methodology; section 3 shows the dynamic formulation of the model used in the Nelson-Siegel estimation; section 4 shows

[1] See also Lahiri & Martin (2010), Bauwens *et al.* (1999) and Geweke (2010) for more properties of Bayesian forecasts.

the results of the estimations and forecasts and section 5 examines the effects of using a fixed decay parameter versus a time varying decay. In this section it is proposed two methods to select the decaying parameter value. The final comments are in section 6.

2. Integrated Nested Laplace Approximations

The INLA methodology was developed by Rue *et al.* (2009) to allow accurate Bayesian inference in models known as Gaussian Markov Random Fields (GMRF), through analytical approximations for the hyperparameters θ and the latent factors ξ, as an alternative to methods based on numerical simulations as MCMC. A latent GMRF model is a hierarchical model with the first stage specifying a distribution for the observed variable y, usually assumed to be conditionally independent given the latent factors ξ and additional parameters θ, and the second stage specifying the evolution of the latent factors ξ:

$$\pi(y|\xi,\theta) = \prod_{j} \pi(y_j|\xi_j,\theta),\ j \in J \tag{1}$$

$$\xi_i = \text{Offset}_i + \sum_{k=0}^{\eta_f - 1} \omega_{ki} f_k(c_{ki}) + z_i^T \beta + \epsilon_i,\ i = 0, ..., \xi_f - 1, \tag{2}$$

where J is a subset of the latent factors, $\pi(y|\xi,\theta)$ is the likelihood function of observed variables, ϵ is a vector of unstructured random effects, Offset is a possible fixed and known a priori component to be included in linear prediction and ω_k are known weights for each observed point. The function $f_k(c_{ki})$ is the effect of generic covariates with value c_{ki} for observation i, and z_i is the effect of covariates with linear effects, with corresponding parameter β.

The formulation described by Equation 1 can represent various statistical objects such as generalized additive models, generalized additive mixed models for longitudinal data, geoadditive models and especially time series models and models formulated in state space, as dynamic linear and stochastic volatility models, as described in Ruiz-Cardenas *et al.* (2010). The INLA methodology allows us to approximate the posterior distributions of the latent factors Gaussian, denoted by:

$$\pi(\xi_i|Y) = \int \pi(\xi_i|\theta, Y)\pi(\Theta, Y) d\theta \tag{3}$$

and the marginal posterior distribution of hyperparameters, written as:

$$\pi(\theta_j|Y) = \int \pi(\theta|Y) d\theta_{-j} \tag{4}$$

using a sequence of analytical approximations based on Laplace methods for the full conditional distributions $\pi(\theta|Y)$ and $\pi(\xi_i|Y)$ and numerical integration for the hyperparameters θ, where θ_{-j} means the vector θ with the omitted j-th element. The INLA methodology of Rue *et al.* (2009) uses a sequence based on three steps to obtain these posterior distributions. The first step is to obtain an approximation to the full posterior distribution $\int \pi(\theta|Y)$, which is obtained from an approximation to the full conditional distribution of the latent factors $\pi(\xi|Y,\theta)$ using a multivariate Gaussian distribution $\tilde{\pi}_G(\xi|Y,\theta)$, and evaluated in the mode of this distribution. From this Gaussian distribution the posterior distribution of θ is obtained by Laplace approximation:

$$\tilde{\pi}(\Theta|Y) \propto \frac{\pi(\xi,\theta,Y)}{\tilde{\pi}_G(\xi|\theta,Y)} \Big|_{\xi=\xi^*(\Theta)}, \tag{5}$$

with $\xi^*(\theta)$ denoting that this approximation is evaluated in the mode of the conditional distribution of ξ given θ. Since there is no analytical formula for $\xi^*(\theta)$, this mode is obtained through a Newton-Raphson algorithm. The next step is to get the Laplace approximation to the full conditional distributions $\pi(\xi_i|\theta, Y)$ for particular values of θ, which will be used as points of evaluation in the numerical integration to obtain the posterior distribution of the latent factors in the equation 3, which is obtained by the Laplace approximation:

$$\tilde{\pi}_{LA}(\xi_i|\theta, Y) \propto \frac{\pi(\xi, \Theta, Y)}{\tilde{\pi}_G(\xi_{-i}|\xi_i, \Theta, Y)} |_{\xi_{-i}=\xi^*_{-i}(\xi_i,\Theta)}, \tag{6}$$

where ξ_{-i} means a vector of latent factors ξ with the omitted i-th element, $\tilde{\pi}_G(\xi_{-i}|\xi_i, \theta, Y)$ is the Gaussian approximation for $\pi(\xi_{-i}|\xi_i, \theta, Y)$ keeping ξ_i fixed and $\xi^*_{-i}(\xi_i, \Theta)$ is again the mode for $\pi(\xi_{-i}|\xi_i, \Theta, Y)$.

As the Laplace approximation given by equation 6 is computationally intensive because it requires a Gaussian approximation for each possible value of the hyperparameters and latent factors, two simplifications are proposed in Rue *et al.* (2009). The first approach is to use only the $\tilde{\pi}_G(\xi_{-i}|\xi_i, \Theta, Y)$, that gets results with reasonable accuracy and reduced computational cost. However, the accuracy of this approach can be affected by errors in the location or by a possible lack of asymmetry in the Gaussian approximation. The second form is the simplified Laplace approximation $\tilde{\pi}_{SLA}(\xi_i|\Theta, Y)$, obtained as a series expansion of $\tilde{\pi}_{LA}(\xi_i|\Theta, Y)$ around $\xi_i = \mu_i(\theta)$, as discussed in Rue *et al.* (2009).

The final step of the algorithm allows to obtain the marginal posterior distributions of ξ_i and θ_i by combining the two approximations in the previous steps and integrating the irrelevant factors. The approximation for the marginal distributions of the latent factors is given by:

$$\pi(\xi_i|Y) = \int \pi(\xi_i|\theta, Y)\pi(\Theta, Y)d\theta \approx \sum \pi(\xi_i|\theta_k, Y)\tilde{\pi}(\theta_k|Y)\triangle_k$$

evaluated by numerical integration in a sequence of values θ_k, with weights given by $\triangle_k$, usually chosen to be equal across the grid of values as the evaluation points of θ_k are spaced at regular intervals. The numerical approximation to the marginal posterior distribution of hyperparameters is obtained similarly:

$$\pi(\theta_j|Y) = \int \pi(\theta|Y)d\theta_{-i} \approx \int \tilde{\pi}(\theta_k|Y)d\theta_{-i}.$$

The accuracy of these numerical integrations is dependent on an adequate choice of evaluation points θ_k. The choice proposed in Rue *et al.* (2009) is obtained by spectral decomposition $S^{-1} = Q\Delta Q^\top$ of the Hessian matrix S, with Q denoting the eigenvectors and Δ the eigenvalues of the decomposition, in the mode of $\tilde{\pi}(\theta_j|Y)$, and setting a standardized value of a variable z given by:

$$z = Q^\top \Lambda^{-1/2}(\theta - \theta^*)$$

and so the corresponding points of θ around the mode θ^* are obtained as:

$$\theta(z) = \theta^* + Q\Delta^{1/2}z.$$

Starting from z=0 each additional point of z is calculated in the positive and negative directions around steps of size η_z satisfying:

$$\log \tilde{\pi}(\theta(0)|Y) - \log \tilde{\pi}(\theta(z)|Y) < \eta_\pi,$$

with the choice of η_π in accordance with the desired accuracy.

The uses of the INLA methodology for estimating dynamic (time series) models are based on state space formulation of Gaussian models, based on the following representation:

$$\begin{matrix} y_t \\ X_t \end{matrix} = \begin{matrix} F_t^{'} X_t \\ G_t^{'} X_{t-1} \end{matrix} + \begin{matrix} \nu_t \\ \omega_t, \end{matrix} \tag{7}$$

where X_t denotes the latent factors, $\nu_t \sim N(0, V_t)$, $\omega_t \sim N(0, W_t)$, with V_t and W_t denote the precision matrices for these errors.

This formulation allows accurate Bayesian estimation of any dynamic model that can be placed in this state space representation, such as ARMA and dynamic linear models. The article Ruiz-Cardenas *et al.* (2010) contains an extensive analysis of the use of INLA methodology in the estimation of time series models. In particular, any dynamic model of the term structure of interest rates, which admits a linear representation in state space can be accurately estimated by formulating INLA, as the dynamic representation of the Nelson-Siegel model used in Diebold & Li (2006).

3. Dynamic Nelson-Siegel Model

Diebold & Li (2006) proposed a dynamic version of the Nelson-Siegel model (Nelson & Siegel (1987)), interpreting the parameters for each yield curve as dynamic factors. This model can be formulated through an observation equation for the observed yield curve given by:

$$y_t(m) = \beta_{1t} + \beta_{2t}\frac{1-e^{-m\lambda}}{m\lambda} + \beta_{3t}\left[\frac{1-e^{-m\lambda}}{m\lambda} - e^{-m\lambda}\right] + \epsilon_t(m) \tag{8}$$

and a system of state equation determining the evolution of the latent factors as a first order autoregressive process:

$$\begin{bmatrix} \beta_{1t} \\ \beta_{2t} \\ \beta_{3t} \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} + \Phi \begin{bmatrix} \beta_{1t-1} \\ \beta_{2t-1} \\ \beta_{3t-1} \end{bmatrix} + \epsilon_{\beta t}, \tag{9}$$

where $y_t(m)$ denote the yield of the yield curve at time t as a function of maturity m, β_{1t}, β_{2t} and β_{3t} the latent factors with interpretations of level, slope and curvature, respectively, λ the parameter that controls the decay of the yield curve and $\epsilon_t(m)$ and $\epsilon_{\beta t}$ processes of innovations of the observation (signal) and state equations, respectively.

This representation becomes a Gaussian linear state space by assuming that the shocks $\epsilon_t(m)$ and $\epsilon_{\beta t}$ are independent Gaussian processes and the decay parameter λ is constant and known. Fixing the parameter λ as a priori known value allows a simple estimation in two steps, as proposed in Diebold & Li (2006). In this methodology keeping fixed the parameter λ is possible to estimate the parameters β_{1t}, β_{2t} and β_{3t} through a linear regression for each day of the observed yield curve in a first stage, and modeling the dynamics of these latent factors in a second stage through a vector autoregressive model for the series estimated in the first stage. Note that although computationally simple, this two-step estimation is statistically inefficient and the confidence intervals for the parameters, latent factors and forecasts are not calculated correctly, since they do not take into account that these latent factors are actually estimated in the first stage.

An alternative way is to assume that this decay parameter is constant but unobserved, and obtain inferences about the parameters and latent factors using maximum likelihood estimation by the decomposition of prediction error using the Kalman filter, as used for example in Diebold *et al.* (2006) and Yu & Zivot (2010). Yu & Zivot (2010) showed that the Kalman filter estimation compared with the method of Diebold & Li (2006) produced better forecast for longer maturities but worst for shorter maturities of the yield curve. A modified version of the Kalman filter is also used in Koopman *et al.* (2010) in the estimation of an extension of the dynamic Nelson-Siegel model which allows a factor stochastic volatility for the latent factor, through the extended Kalman filter.

Bayesian inference methods for the dynamic Nelson-Siegel models are presented by Migon & Abanto-Valle (2007) who estimated this model assuming that the decay parameter is constant but not observed. Laurini & Hotta (2010) estimated the model by a methodology based on MCMC where it is possible to consider the decay parameter as not observed and time varying. Hautsch & Yang (2010) generalized the Nelson-Siegel dynamic model including a stochastic volatility and also estimated the model by methods based on MCMC.

Although these two articles present relevant generalizations to the dynamic Nelson-Siegel model, the introduction of time varying decay parameters and volatilities do not necessarily represent improvements in the out-of-sample forecasting performance, and higher computational costs involved in the estimation by MCMC makes it almost prohibitive to use these models to construct recursive out-of-sample forecasts due to the non-sequential nature of this procedure.

The objective of the present study is to overcome these limitations with a Bayesian estimation methodology using INLA when estimating the same dynamic Nelson-Siegel formulation proposed in Diebold & Li (2006). We will show that the estimated model presents better properties in terms of inference and prediction. We initially assume that the decay parameter is constant and known a priori, but we also allow for time varying decay parameters in the next section. When we assume that the decay parameter is constant and known, there is a direct representation of this model as a Gaussian Markov Random Field, as formulated by equation 1, assuming that the weight vectors ω_k now represent the weights of the Nelson-Siegel for each maturity m, and are given by:

$$\begin{aligned} \omega_1^m &= 1 \\ \omega_2^m &= \tfrac{1-e^{-\lambda m}}{\lambda m} \\ \omega_3^m &= \tfrac{1-e^{-\lambda m}}{\lambda m} - e^{-\lambda m}, \end{aligned} \tag{10}$$

with the same dynamics of the latent factors given by equation 9.

To check the robustness of the estimation on this assumption, we conducted an alternative analysis suggesting an alternative methodology for the determination of the decay parameter, allowing it to be time varying and estimated by optimizing the marginal likelihood and the cross-validated log score. The results of the estimation assuming the parameter λ is kept fixed and time variant are showed in the following sections.

4. Results

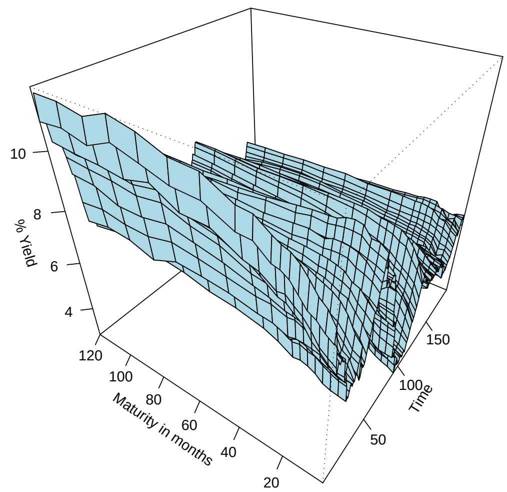
4.1. Full sample analysis

To show the potential uses of the INLA methodology in the estimation of dynamic models of the term structure of interest rates, we use this methodology to estimate the dynamic extension of the Nelson-Siegel model proposed by Diebold & Li (2006), using the same database studied in this article. The database is constructed using price quotations (average between bid-ask prices) of U.S. Treasuries 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 and 120 months using the procedure from Fama & Bliss (1987), obtained from the CRSP files and available in the website http://www.ssc.upenn.edu/~fdiebold/papers/paper49/FBFITTED.txt. Figure 1 shows the yields used in the study. Descriptive statistics and additional descriptions of this database can be found in the original study of Diebold & Li (2006).

Initially we assumed that the parameter λ in the Nelson-Siegel curve (Eq. 8) is known and constant over time. We followed the same choice of the Diebold & Li (2006) assuming a value of λ that maximizes the load factor in the medium term to maturity of 30 months, this value being equal to 0.0609. With this assumption, we estimated the model in a single step using the INLA methodology described previously. Also following the recommendation of Diebold & Li (2006) paper we assumed that each latent factor follows a first order autoregressive process, as a simplification of a more general vector autoregressive process, where each latent factor has an evolution equation given by $\beta_{it} = \phi_i \beta_{it-1} + \epsilon_{it}$. As demonstrated by Diebold & Li (2006), this simplification reduces the number of parameters to be estimated and represents gains in the out-of-sample predictive power.

To complete the Bayesian formulation, we assumed a multivariate Gaussian likelihood function, and assumed that the latent factors are orthogonal, and thus the variance matrix of the latent factors is diagonal. Although this structure can be generalized to a full matrix, the estimation with a diagonal variance matrix is computationally simpler and did not represent a significant change in model fit and forecasts. The structure of the priors for autoregressive process is given by a structure of log-Gamma prior for precision with hyperameters 1.0 and 0.001, respectively, for the shape and inverse location and a log-Gaussian structure for the autoregressive parameter with mean 0.95 and variance 0.01. According to the formulation of Ruiz-Cardenas *et al.* (2010)

Figure 1: Yield Curve - January 1985 - December 2000



stationarity of the autoregressive process is defined by the transformation $\theta = log[(1+\phi)/(1-\phi)]$ in the representation of the autoregressive process. An important result is that the posterior distributions were essentially invariant to choice of priors, and the results of latent factors and forecasts do not change significantly for alternative values for these priors.

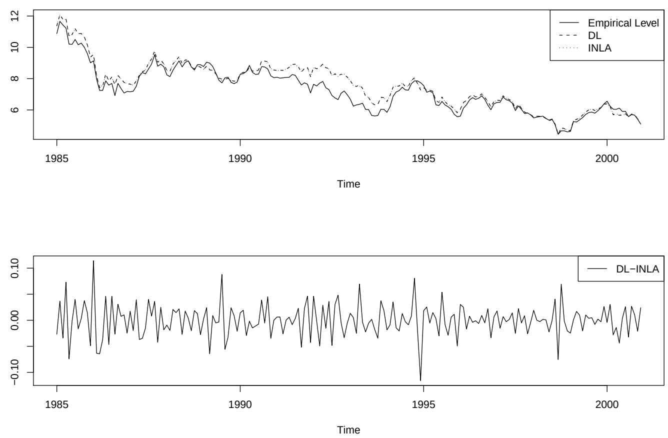
The results presented are based on the use of estimation based on the full Gaussian approximation for the Laplace approximation (Eq. 6). We also performed the same analysis using the approximation $\tilde{\pi}_G(\xi_{-i}|\xi_i, \Theta, Y)$, and the simplified Laplace approximation, but the results are basically equivalent, and for this specific estimation the computational cost is also quite similar, and therefore we chose to work with the theoretically recommended approximation. In all the analysis we report the value of the latent factors as the means for the posterior distributions, as well as the credibility intervals of 95% obtained by the empirical quantiles of .025 and .975 of the posterior distributions.

Other robustness analysis were performed to verify if the results are altered using the posterior median, and alternative specifications for the likelihood using Student-t and Laplace distributions. The results, both in terms of the fit of models, estimated latent factors and predictive performance were not altered significantly by the use of these likelihoods or the use of median, and so we used the Gaussian likelihood and the posterior mean in all the analysis.

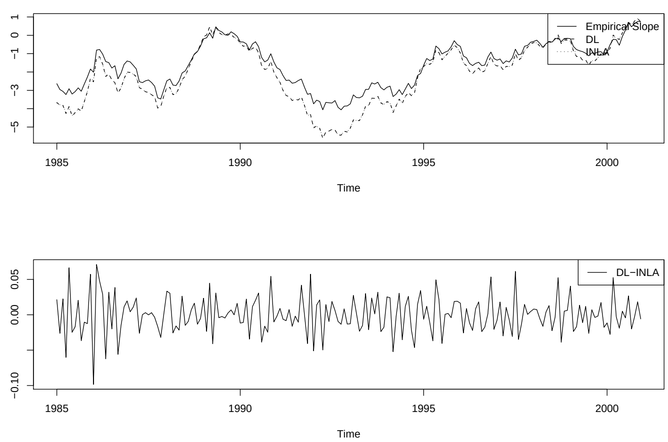
Figure 2 shows the estimates of the factors β_{1t}, β_{2t} and β_{3t} using the INLA methodology and the two steps methodology from Diebold & Li (2006). The figure also presents the and empirical estimates of the factors as defined in Diebold & Li (2006). The estimations use all the available sample. The figure also presents in the second panel the difference between the estimates by the INLA and the two steps methodologies. The latent factors estimated by the both methodologies are quite similar, with the magnitude of the differences between the estimates much smaller than the factor estimated values.

Figure 3 shows the autocorrelation function for the estimated latent factors and the residuals of each autoregressive process. The results are very similar to results obtained in the article of Diebold & Li (2006). The residuals of fit for the entire observed yield curve is placed in Figure 4, also showing that the residuals had decreased magnitude and showing that this method also

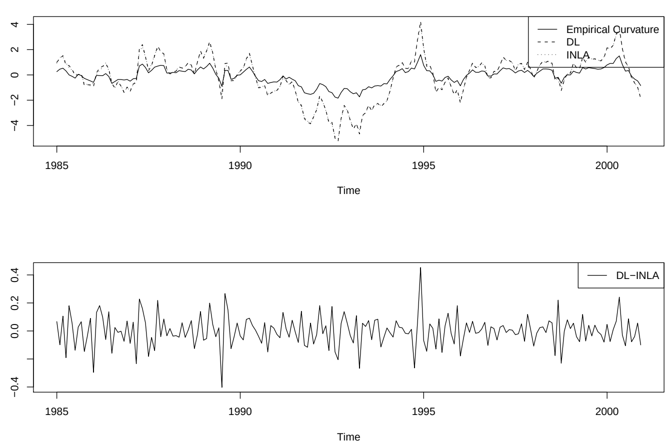
Figure 2: Estimated Factors Using INLA, the Two Steps OLS Methodologies and the Empirical Estimates. First Panel: Estimated values. Second Panel: Difference Between the Estimates.



(a) β_{1t} - Level

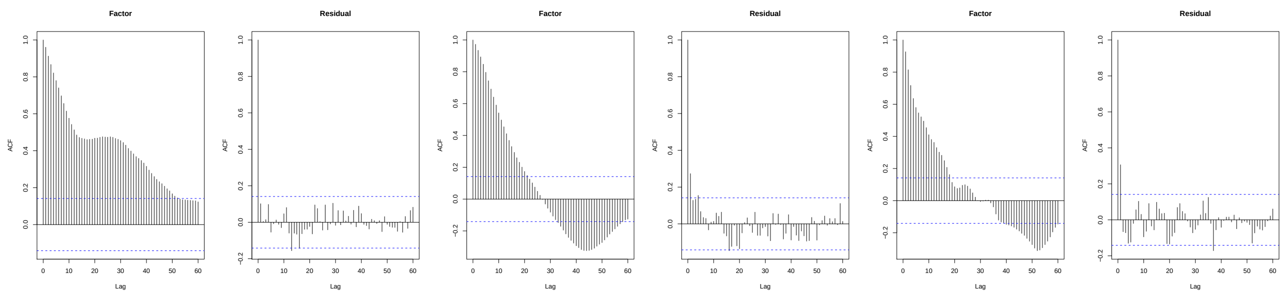


(b) β_{2t} - Slope



(c) β_{3t} - Curvature

Figure 3: Estimated Autocorrelogram Function of the Latent Factors and Residuals Autoregressive Processes



(a) β_{1t} (b) β_{2t} (c) β_{3t}

obtained a good in-sample fit.

The posterior distribution obtained for the precision of Gaussian process, autoregressive and precision parameters of latent factors are shown in Figures 5 and 6, and summarized in Table 1. The results indicate a high persistence for the first two latent factors and a lower persistence for the third factor, a result consistent with the literature of estimation of dynamic models of interest rates.

	mean	sd	0.025quant	0.5quant	0.975quant
Precision for the Gaussian observations	196.138	5.340	185.879	196.065	206.809
Precision for β_{1t}	11.434	1.290	9.114	11.362	14.165
ϕ for β_{1t}	0.999	0.003	0.992	0.999	1.000
Precision for β_{2t}	11.155	1.207	8.976	11.090	13.702
ϕ for β_{2t}	0.990	0.012	0.958	0.994	0.999
Precision for β_{3t}	2.188	0.266	1.713	2.172	2.754
ϕ for β_{3t}	0.905	0.030	0.835	0.910	0.952

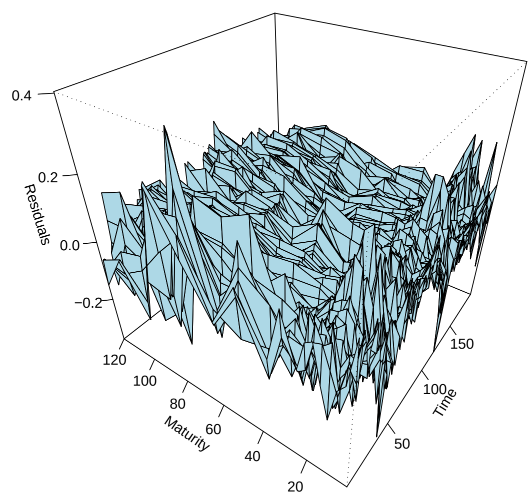
Table 1: Summary of the Hyperparameters Posterior Distributions, INLA Estimation Using Full Sample.

An obvious advantage of INLA methodology is that we do not have to worry about the convergence problem of the numerical approximations, a problem that usually occurs in the estimation of models with latent variables via MCMC methods. The INLA methodology also avoids the fairly common problems of slow convergence of chains and the existence of multimodal posterior distributions in the estimation of dynamic models with latent factors using MCMC. To show this point, we estimated the same dynamic Nelson-Siegel models by MCMC methods, using a simplification of the methodology used in the studies of Laurini & Hotta (2010) and Hautsch & Yang (2010), with the same likelihood and priors structures, and adopted the same transformation for stationarity used in the estimation of autoregressive parameters.

Figure 7 shows a comparison of the posterior distribution of the autoregressive parameters obtained by the INLA approximation and MCMC, for the latter using a long chain with 300,000 iterations and thinning every 10 iterations, which in theory should enable the convergence of these chains and minimize the problem of high autocorrelation. The results (Figure 7) indicate that even in this simple model with a linear structure in the observation equation by using a fixed decay parameter λ there are convergence problems in chains.

There are multiple modes for the autoregressive parameters of the first two latent factors, and contrary to the expectations in this estimate the persistence of the second latent factor is greater than the first factor, indicating possible identification problems in this estimation. We can identify some of the problems of convergence being generated by the transformation $\theta = log[(1+\phi)/(1-\phi)]$ in the estimation of the autoregressive parameter, which generates large instability in simulated Markov chains, but not affecting the methodology based on Integrated

Figure 4: Residuals of the Model Estimated by INLA Methodology Using Full Sample



Nested Laplace Approximations. Another obvious problem is that the estimation by MCMC, because of the problems of convergence, has out-of-sample predictive performance worse than the INLA and OLS methodologies, as will be shown in the next subsection. Another obvious problem is that the computational time spent on MCMC methodology is well above the OLS and INLA methodologies and thus the results of the MCMC methodology in terms of predictive performance are inferior to other methods by several metrics.

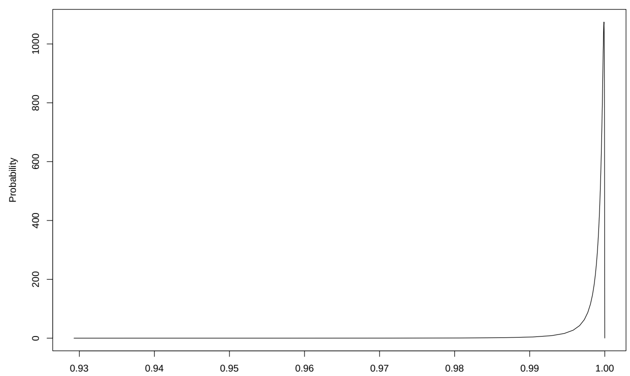
Another important point emphasized by Rue *et al.* (2009) is that when accurate results are required in MCMC applications, such as forecasting of the yield curve for hedging and trading operations, it is necessary to increase significantly the number of simulations, since in MCMC methods the rate of convergence is $\sqrt(n)$, which makes this procedure computationally very intensive and thus limit its practical application when precise estimates are required with a reduced computational cost.

4.2. Out-of-Sample Predictive Performance

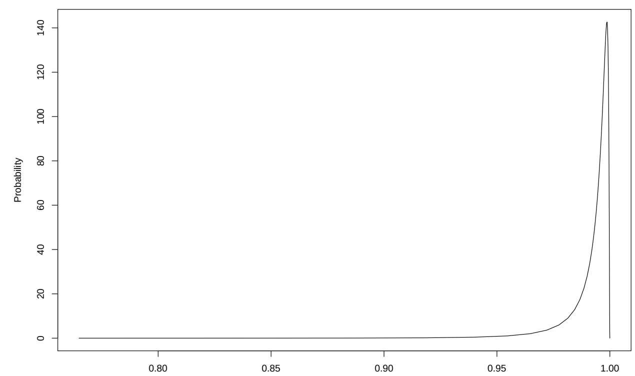
Although we have obtained in-sample evidence favoring the INLA methodology, a key point is whether the greater complexity of this estimator results in better in terms of out-of-sample predictive performance. The best performance of the estimator in two steps proposed in Diebold & Li (2006) over several other methodologies was the major result of the article, showing that this simple estimator could perform better in relation to various other widely used methodologies for forecasting yield curves. Thus, we compare the out-of-sample forecasting performance of the model estimated by the Diebold & Li (2006) two steps and INLA methodologies. Note that the model is the same.

To construct the out-of-sample forecasts, we followed the same procedure of Diebold & Li (2006) using a recursive estimation, and performing out-of-sample 1, 6 and 12 months ahead forecasting for maturities of 3 months and 1, 3, 5 and 10 years. The sample used is from 1985:1 to the date in which the forecast is made, starting from 1994:1 to 2000:12. Following the same structure as the Diebold & Li (2006) paper, we constructed the mean and standard deviation of forecast errors, and the Root Mean Squared Errors and autocorrelations of the forecast errors. In the estimation by INLA methodology λ was fixed as 0.0609, the same value used in the Diebold

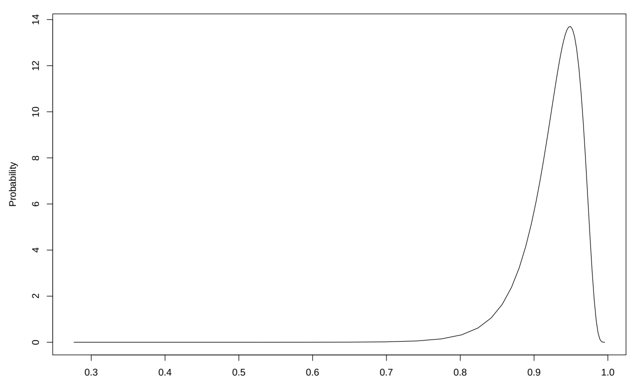
Figure 5: Posterior Distributions of the Autoregressive Parameters. INLA Estimation of the Full Sample



(a) ϕ_1

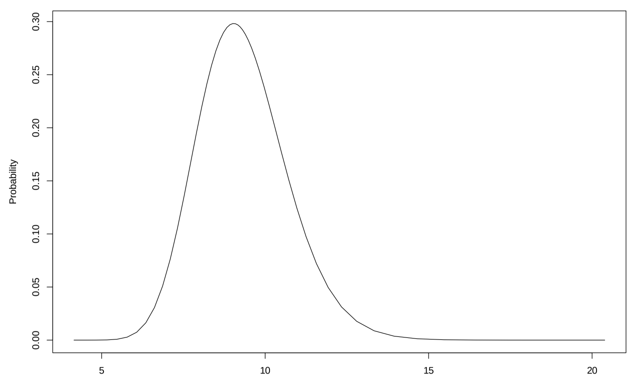


(b) ϕ_2

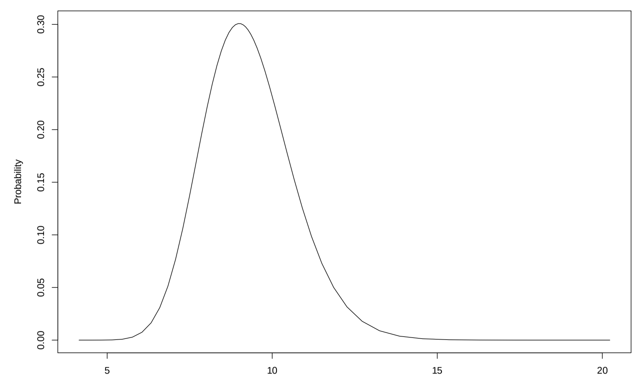


(c) ϕ_3

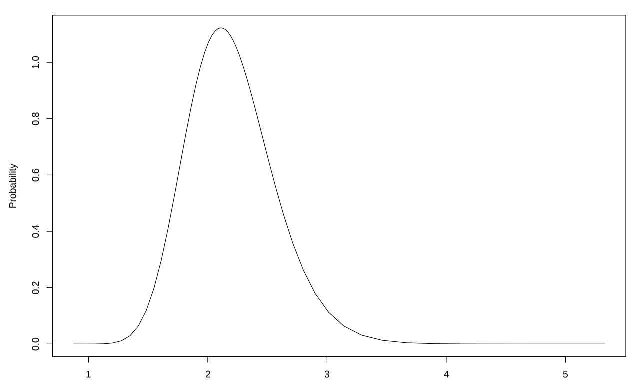
Figure 6: Posterior Distributions of the Precision Parameters. INLA Estimation of the Full Sample



(a) Precision β_{1t}

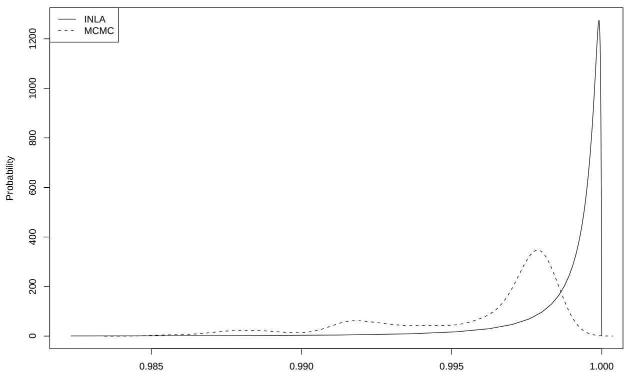


(b) Precision β_{2t}

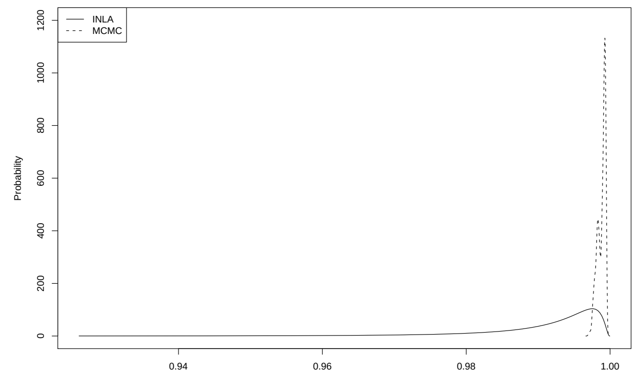


(c) Precision β_{3t}

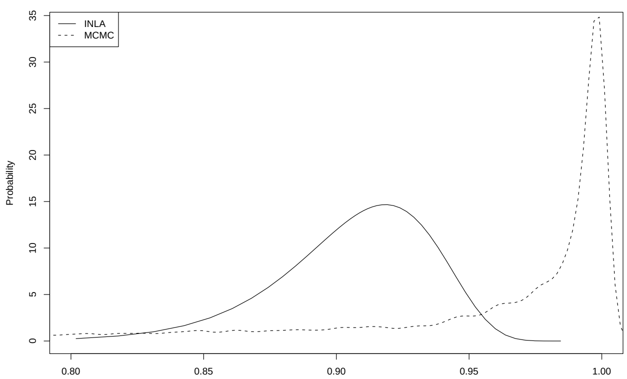
Figure 7: Posterior Distributions of the Autoregressive Parameters. INLA and MCMC Estimation of the Full Sample



(a) ϕ_1



(b) ϕ_2



(c) ϕ_3

& Li (2006) paper. We also considered the predictive performance of this same model using the estimation by Markov Chain Monte Carlo (MCMC). We first estimated the model with chains with $30,000$ iterations, thinning of 10 iterations after a burn-in of $10,000$ iterations. This number of iterations was chosen to obtain the convergence criterion of Gelman-Rubin and give a reasonable accuracy to the factors and parameters estimated. Later, in order to verify a possible improvement with models estimated using longer chain we made new estimates now using a chain of $300,000$ iterations after the burn-in process. Tables 2, 3 and 4 show the forecast results, respectively for 1, 6 and 12 months ahead prediction, for maturities of 3 months and 3, 5 and 10 years. The original results of Diebold & Li (2006) can be found in the original article and are not replicated to save space. In each table the symbol * denotes that the estimated model has a better performance, in terms of RMSE, compared to the results of the methodology of Diebold & Li (2006); When none of the three is marked it means that the model estimated by OLS was the best of the four methods in terms of RMSE.

The prediction exercise showed that the models estimated by the INLA methodology achieved the best performance for all maturities for the horizon of one month ahead, with the exception of the maturity of one year, for maturities of 3, 5 and 10 years for the forecasting horizon of six months ahead, and for the maturities of 5 and 10 years for the 12-month ahead forecasting horizon.

The analysis of the results also show a much lower predictive performance of the models estimated by MCMC, even when compared with the models estimated by OLS in two steps. With $30,000$ iterations there is no single maturity or forecast horizon in which there is an advantage of MCMC methodology. The increase to $300,000$ iterations improved just a little bit, and the MCMC method produced smaller RMSE than the OLS method only for the maturities of 1 and 3 years for 1-month ahead forecasting. Even for $300,000$ iterations the MCMC was worst than INLA for all maturities and forecasting horizon. These results indicate that given the enormous computational cost involved in the methodology of estimation by MCMC it is not justified to use this methodology for out-of-sample forecasting for this model. They also support the use of the methodology INLA, which allows accurate inferences and forecasts with a reduced computational cost.

	Mean	Std.Dev	RMSE	ACF1	ACF12
3 months	-0.004	0.164	0.163*	0.202	0.035
	-0.308	0.235	0.387	0.489	-0.024
	-0.301	0.236	0.381	0.512	-0.065
1 year	0.062	0.231	0.238	0.404	-0.216
	0.019	0.235	0.234	0.393	-0.226
	0.028	0.231	0.231*	0.376	-0.223
3 years	-0.015	0.273	0.272*	0.331	-0.123
	0.071	0.315	0.321	0.542	-0.227
	0.083	0.308	0.318	0.535	-0.221
5 years	-0.046	0.276	0.278*	0.322	-0.116
	-0.040	0.296	0.297	0.459	-0.188
	-0.025	0.291	0.290*	0.446	-0.181
10 years	-0.011	0.252	0.251*	0.262	-0.091
	-0.175	0.264	0.315	0.317	-0.100
	-0.157	0.267	0.309	0.341	-0.100

Table 2: Summary of the Out of Sample 1-Month Ahead Forecasting. First Line-INLA; Second Line: MCMC with 30,000 iterations; Third Line: MCMC with 300,000 iterations. * denotes better results than OLS; no marks means that the OLS is the best result.

	Mean	Std.Dev	RMSE	ACF6	ACF18
3 months	0.147	0.539	0.555	0.292	-0.204
	-0.318	0.507	0.596	0.154	-0.128
	-0.272	0.517	0.581	0.105	-0.071
1 year	0.181	0.678	0.698	0.053	-0.135
	-0.035	0.716	0.713	0.046	-0.170
	0.014	0.723	0.718	0.003	-0.137
3 years	0.042	0.752	0.748*	-0.091	-0.174
	-0.060	0.795	0.792	-0.031	-0.203
	0.002	0.803	0.798	-0.071	-0.178
5 years	-0.032	0.758	0.753*	-0.066	-0.209
	-0.222	0.792	0.818	-0.032	-0.233
	-0.149	0.804	0.813	-0.071	-0.206
10 years	-0.053	0.674	0.671*	-0.100	-0.212
	-0.418	0.721	0.829	-0.077	-0.243
	-0.332	0.747	0.813	-0.109	-0.221

Table 3: Summary of the Out of Sample 6-Months Ahead Forecasting. First Line-INLA; Second Line: MCMC with 30,000 iterations; Third Line: MCMC with 300,000 iterations.* denotes better results than OLS; no marks means that the OLS is the best result.

5. Robustness to Fixed λ

The INLA methodology is based on the representation of a linear additive model in the state space form, and in the application to the estimation problem of the dynamic Nelson-Siegel model it is necessary to assume that the decay parameter is fixed and known. This simplification is also adopted in Diebold & Li (2006) to allow the estimation by least squares in the first stage of the procedure.

However, it is possible to make adjustments in the methodology to deal with unknown decay parameter and also time varying. This allow us to verify whether the adoption of this assumption is valid and does not represent significant losses in terms of fit and forecast for the yield curves. We propose to use a method analogous to the methodologies used in the determination of the smoothing parameters in the smoothing spline methods (see, for instance Wang (2004)), defining measures of fit and models selection in function of this decay parameter λ.

The proposed methodology select the decay parameter λ for each sample as an optimization problem by finding the numerical value of λ which optimize some interesting measure of model fit. In this paper, this procedure is performed for two measures - the first is the marginal likelihood and the second the cross-validated log-score of the model.

In this regard, the analytical approximations given by (Rue *et al.* (2009)) INLA methodology are extremely convenient, since it allows to calculate measures of prediction and model selection in a quick and accurate way, e.g. the estimation of marginal likelihood, generalized cross validated log-score and probability integral transform (PIT), which are usually computationally prohibitive in methodologies such as MCMC, as discussed in Held *et al.* (2010). In the INLA methodology, we can calculate measures of generalized cross validation without having to re-estimate the model for each observation left out in cross validation procedure, and also to calculate accurately the marginal likelihood. We note, for instance, the difficulties encountered in calculating the marginal likelihood in complex models estimated by MCMC, such as numerical instability existing in marginal likelihood approximations using the harmonic mean of the Monte Carlo simulations, e.g. Newton & Raftery (1994).

In the INLA methodology the marginal likelihood of the model is easily calculated using the expression:

$$\tilde{\pi}(Y) = \int \frac{\pi(\theta)\pi(x|\theta)\pi(y|x,\theta)}{\tilde{\pi_G}(x|\theta, y)} d\theta. \tag{11}$$

	Mean	Std.Dev	RMSE	ACF12	ACF24
3 months	0.203	0.838	0.856	-0.228	-0.061
	-0.353	0.856	0.921	-0.231	-0.071
	-0.282	0.854	0.894	-0.225	-0.043
1 year	0.198	0.871	0.887	-0.274	-0.032
	-0.137	0.912	0.916	-0.304	-0.029
	-0.059	0.922	0.918	-0.310	-0.008
3 years	0.017	0.927	0.921	-0.357	-0.034
	-0.234	0.935	0.958	-0.341	-0.007
	-0.129	0.959	0.961	-0.365	0.005
5 years	-0.089	0.956	0.953*	-0.372	-0.054
	-0.435	0.953	1.042	-0.352	-0.027
	-0.308	0.991	1.032	-0.375	-0.023
10 years	-0.157	0.878	0.886*	-0.428	-0.079
	-0.679	0.907	1.128	-0.385	-0.054
	-0.526	0.962	1.090	-0.411	-0.055

Table 4: Summary of the Out of Sample 12-Months Ahead Forecasting. First Line-INLA; Second Line: MCMC with 30,000 iterations; Third Line: MCMC with 300,000 iterations. * denotes better results than OLS; no marks means that the OLS is the best result.

As a measure of cross-validation leaving out one observation, we can calculate the conditional predictive ordinates (CPO) of the observation y_i, conditional on all sample, except this observation left out as:

$$CPO_i = \pi(y_i^{obs}|y_{-1}), \tag{12}$$

where CPO of observation i is obtained as:

$$CPO_i = \int \pi(y_i^{obs}|y_{-1},\theta)\pi(\theta|y_{-i})d\theta \tag{13}$$

and the first term in this integral is defined as:

$$\pi(y_i^{obs}|y_{-1}) = 1/\int \frac{\pi(x_i|y,\theta)}{\pi(y_i^{obs}|x_i,\theta)} \tag{14}$$

and calculated by numerical integration, all the denominator of the equation Eq. 14 is the likelihood contribution of observation i and known. The leave-one-out cross-validation likelihood (cross-validated log-score) is finally defined as:

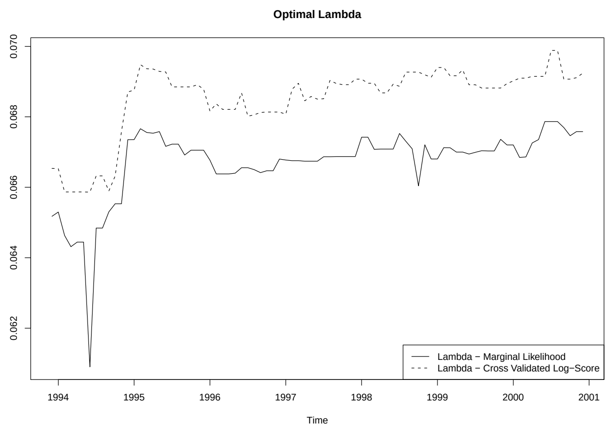
$$LSCV = \sum_{i=1}^{n} -log(CPO_i). \tag{15}$$

Thus we can select the parameter λ in order to optimize the values of the marginal likelihood and the cross-validated log-score model. This is done by estimating an optimal λ using a numerical optimization algorithm for each sample. Note that the computational efficiency of the INLA approximation is essential in this aspect, since each iteration of the numerical algorithm involves repeated estimations, which would be computationally prohibitive for estimation via MCMC, in particular for the calculation of measures of cross-validation.

With this methodology, we selected the optimal values of λ for each day of the sample used in the recursive out-of-sample forecasting procedure discussed previously. Figure 8 shows the optimal value of lambda for each of the criteria throughout the sample.

It may be noted that although the values found are higher than the value of $\lambda = 0.0609$ used in Diebold & Li (2006), values are numerically close to this value, indicating a robustness of the

Figure 8: Optimal λ, selected by Marginal Likelihood and Cross-Validated Log-Score



rule used in Diebold & Li (2006) to find the decay parameter, maximizing the load factor for the medium-term maturity of 30 months.

To verify whether it is relevant to use an optimal value of λ and how its value affects the out-of-sample forecasting, we performed the same estimation procedure with the recursive sample as in the previous section, but using the λ optimal choice for each method, and making forecasts for the horizons 1, 6 and 12 months and the same maturities as before, and computed the adjustment measures as in the previous section.

	Mean	Std.Dev	RMSE	ACF1	ACF12
3 months	0.001	0.164	0.163*	0.201	0.037
	0.002	0.164	0.163*	0.200	0.039
1 year	0.058	0.231	0.236+	0.395	-0.213
	0.057	0.230	0.236+	0.396	-0.215
3 years	-0.019	0.273	0.272*	0.328	-0.122
	-0.019	0.273	0.272*	0.327	-0.122
5 years	-0.047	0.276	0.278*	0.320	-0.114
	-0.047	0.275	0.278*	0.319	-0.114
10 years	-0.014	0.254	0.253*	0.265	-0.087
	-0.015	0.255	0.254*	0.268	-0.088

Table 5: INLA - Out of Sample 1-Month Ahead forecasting results. λ selected by Marginal Likelihood (first line) and LSCV (second line). *denotes better results than OLS; +denotes better results than INLA with fixed λ

Tables 5, 6 and 7 show the results of the predictive performance obtained by using the parameter λ selected by maximizing the marginal likelihood and by LSCV. In this table the symbol * beside the value of RMSE indicates a better performance than the estimation by OLS using a λ equal to 0.0609 as fixed, and now the + symbol to denote a superior result compared to the value found by the INLA methodology with the same parameter λ fixed. The results obtained indicate that this selection procedure using the marginal likelihood optimization in general achieves better results than the method of two steps of Diebold & Li (2006), and can improve performance of the method for five of the INLA maturities chosen. However, these maturities are those where the methodology INLA had worser performance than the OLS method, and thus would be dominated by this methodology, not justifying the use of a time varying λ parameter.

In summary, the use of these procedures to choose the parameter λ for the analyzed sample shows that it is possible to improve the performance of the INLA estimator with lambda fixed for

	Mean	Std.Dev	RMSE	ACF6	ACF18
3 months	0.147	0.537	0.554+	0.295	-0.205
	0.147	0.537	0.553+	0.295	-0.204
1 year	0.166	0.680	0.695+	0.039	-0.129
	0.162	0.679	0.694+	0.037	-0.125
3 years	0.023	0.759	0.754	-0.102	-0.172
	0.019	0.759	0.755	-0.104	-0.169
5 years	-0.047	0.764	0.761*	-0.075	-0.208
	-0.050	0.765	0.761*	-0.076	-0.205
10 years	-0.064	0.683	0.681*	-0.107	-0.211
	-0.067	0.685	0.684*	-0.108	-0.210

Table 6: INLA - Out of Sample 6-Month Ahead forecasting results. λ selected by Marginal Likelihood (first line) and LSCV (second line). *denotes better results than OLS; +denotes better results than INLA with fixed λ

	Mean	Std.Dev	RMSE	ACF12	ACF24
3 months	0.200	0.835	0.853+	-0.226	-0.063
	0.199	0.833	0.851+	-0.225	-0.063
1 year	0.180	0.873	0.886+	-0.280	-0.034
	0.174	0.873	0.885+	-0.281	-0.033
3 years	-0.007	0.938	0.932	-0.368	-0.034
	-0.014	0.941	0.934	-0.369	-0.033
5 years	-0.109	0.967	0.967*	-0.382	-0.054
	-0.115	0.969	0.970*	-0.383	-0.053
10 years	-0.169	0.891	0.901*	-0.436	-0.077
	-0.173	0.893	0.903*	-0.438	-0.074

*denotes better results than OLS

+denotes better results than INLA with fixed λ

Table 7: INLA - Out of Sample 12-Month Ahead forecasting results. λ selected by Marginal Likelihood (first line) and LSCV (second line)

the entire sample, but it does not remedy the inferior performance of the estimator for the cases where it was originally worse than the OLS methodology. However, it is not possible to generalize this result to other yield curves. For example, the study of Laurini & Hotta (2010) indicates that for curves with more complicated shapes and dynamics is essential to consider the possibility of a time varying λ. Thus, this method to select the optimum decay parameter could be important in term structure of interest rates with more complex structures.

6. Conclusions

This paper presented the use of Bayesian inference methods through INLA methodology to estimate the dynamic Nelson-Siegel model of interest rates. This methodology allows obtaining accurate Bayesian inference without the use of numerical simulation methods such as MCMC with a reduced computational cost, and especially allows the construction of accurate analytical out-of-sample predictive distributions that fully and coherently incorporate uncertainty parameter, as discussed in Geweke & Amisano (2010).

The results indicated that this methodology is able to obtain a predictive performance superior to the widely used method of Diebold & Li (2006) for most maturities studied, indicating that this methodology can achieve important practical results.

Besides the gain in predictive power, this approximation avoids some of the simplifications used in the method of Diebold & Li (2006), as the two-step estimation, which does not to construct correct confidence (credibility) intervals for the parameters, controlling for the uncertainty in parameters. Another important point is that this methodology also offers optimal selection procedures for the decay parameter λ, which are not computationally feasible for methods based on MCMC.

Although this method of inference has been studied only for a particular model used in forecasting yield curves and only a specific database, it is important to note that this methodology can be easily adapted to the estimation of any model with a linear structure, such as the Gaussian Affine models (e.g. Duffie & Kan (1996)), allowing the use of an estimation methodology that allows accurate analytical inference, without suffering the problems of convergence that may affect the estimation results using Bayesian MCMC methods, and also without suffering the problems of multiple local optima that are usually found in the estimation of affine models with latent variables via the Kalman filter, as discussed in Kim & Orphanides (2005). With these potential advantages, we believe that the INLA methodology can become an important tool for modeling and forecasting time series, with special attention to the latent factor models of the term structure of interest rates.

Bauwens, L., Lubrano, M., & Richard, J-F. 1999. *Bayesian Inference in Dynamic Econometric Models*. Cambridge University Press.

Diebold, Francis X., Rudebusch, Glenn D., & Aruoba, S. Boragan. 2006. The macroeconomy and the yield curve: a dynamic latent factor approach. *Journal of Econometrics*, **131**, 309–338.

Diebold, F.X., & Li, C. 2006. Forecasting the Term Structure of Government Bond Yields. *Journal Of Econometrics*, **130**, 337–364.

Duffie, D., & Kan, R. 1996. A yield-factor model of interest rates. *Mathematical Finance*, 6, 379–406.

Fama, Eugene, & Bliss, Robert R. 1987. The information in long-maturity forward rates. *American Economic Review*, **77**(Sept.), 680–691.

Geweke, J. 2010. *Complete and Incomplete Econometric Models*. Princeton University Press.

Geweke, J., & Amisano, G. 2010. Comparing and evaluating Bayesian predictive distributions of asset returns. *International Journal of Forecasting*, **26**(2), 216 – 230. Special Issue: Bayesian Forecasting in Economics.

Hautsch, Nikolaus, & Yang, Fuyu. 2010. Bayesian inference in a Stochastic Volatility Nelson-Siegel model. *Computational Statistics & Data Analysis*, **In Press, Corrected Proof**, –.

Held, L., Schrödle, B., & Rue, H. 2010. Posterior and Cross-validatory Predictive Checks: A Comparison of MCMC and INLA. In: Kneib, T; Tutz, G. Statistical Modelling and Regression Structures - Festschrift in Honour of Ludwig Fahrmeir.

Kim, D. H., & Orphanides, A. 2005. *Term Structure Estimation with Survey Data on Interest Rate Forecasts*. Finance and Economics Discussion Series, 2005-08, Board of Directors of Federal Reserve System.

Koopman, S. J, Mallee, M. I. P, & van der Wel, M. 2010. Analyzing the Term Structure of Interest Rates using the Dynamic Nelson-Siegel Model with Time-Varying Parameters. *Journal of Business and Economic Statistics*, **28**(329-343).

Lahiri, K., & Martin, G. (eds). 2010. Special Issue: Bayesian Forecasting in Economics. *International Journal of Forecasting*, **26**(2), 211–444.

Laurini, M. P., & Hotta, L. K. 2010. Bayesian extensions to Diebold-Li term structure model. *International Review of Financial Analysis*, **19**(5), 342–350.

Migon, H., & Abanto-Valle, C. 2007. A Bayesian Term Structure Modelling. *Pages 200–203 of:* Fernandes, C., Schimidli, H., & Kolev, N. (eds), *Proceedings of the Third Brazilian Conference on Statistical Modelling in Insurance and Finance*. IME-USP.

Nelson, C. R., & Siegel, A. F. 1987. Parsimonous Modelling of Yield Curves. *Journal of Business*, **60**(4), 473–489.

Newton, M.A., & Raftery, A.E. 1994. Approximate Bayesian inference by the weighted likelihood bootstrap (with Discussion. *Journal of the Royal Statistical Society, series B*, 3–48.

Rue, H., Martino, S., & Chopin, N. 2009. Approximated Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations (with discussion). *Journal of the Royal Statistical Society series B*, **71**, 319–392.

Ruiz-Cardenas, R., Krainski, E. T., & Rue, H. 2010. *Fitting dynamic models using integrated nested Laplace approximations – INLA*. Statistics Technical report no 12, Norges Teknisk-Naturvitenskapelige Universitet.

Wang, W. 2004. *Handbook of Computational Statistics*. Springer. Chap. Model Selection.

Yu, Wei-Choun, & Zivot, Eric. 2010. Forecasting the term structures of Treasury and corporate yields using dynamic Nelson-Siegel models. *International Journal of Forecasting*, **In Press, Corrected Proof**, –.