Bayesian Inference Applied to Dynamic Nelson-Siegel Model with Stochastic Volatility

João F. Caldeira**
Márcio P. Laurini***
Marcelo S. Portugal****

Abstract
In this paper, we propose to estimate a dynamic model for the term structure of interest rates considering two alternative specifications of the Nelson and Siegel (1987) model, generalizing several models found in the literature. At first, we consider that the weights of factors vary over time with conditional heteroskedasticity using a stochastic volatility model with common factors. In the second case, we consider a model in which the latent factors individually follow autoregressive processes with stochastic volatility, including the possibility of leverage effects. The so-called volatility factors seek to capture the uncertainty over time associated with the level, slope and curvature of the yield curve. The estimation is performed through Bayesian inference using the Monte Carlo Markov Chain method. The results for the term structure of DI futures and U.S. Treasury bonds used in this study show that the volatility factors are highly persistent, and also indicate that the use of stochastic volatility leads to better in-sample fits to the observed yield curve.

Keywords: Term structure of interest rates, stochastic volatility, MCMC, factor models.

JEL Codes: C53, E43, G17.
1. Introduction

Fitting and forecasting the term structure of interest rates is of utmost importance for policymakers. In finances, the information provided by interest rate curves is important for the pricing of bonds, interest rate derivatives, portfolio management and asset allocation. In macroeconomics, the interest rate gives important information on the state of the economy and of business. Therefore, several researchers have worked on this topic, giving rise to a comprehensive literature on the fitting and forecast of the term structure. Notwithstanding the sizeable amount of studies dealing with this issue, only very recently has attention been turned to the fitting and forecast of interest rates along with the conditional heteroskedasticity of the term structure of interest rates.¹

This issue is particularly important because the assumption of constant interest rate volatility in these models often has remarkable practical implications for risk management policies, which can be too simple, neglecting the risk of a time-varying volatility structure. Interest rate hedging and arbitrage operations could be influenced by the presence of time-varying volatility as, in these operations, it is necessary to compensate for the market price of interest rate risk. Another important implication is that in the presence of conditional volatilities the confidence intervals of fits and forecasts derived from these models will be calculated incorrectly in finite samples.

To assess the effects of time-varying volatility structures, it is important to retrieve characteristics from the major models used for the modeling of the term structure of interest rates. These models can be categorized into three main classes: no-arbitrage models; equilibrium models; and statistical models.

No-arbitrage models for the term structure of interest rates are obtained by the imposition of consistency conditions between the interest rates of several maturities so as to prevent the systematic existence of arbitrage opportunities, and are based on the existence of an equivalent martingale. Examples of these models include those of Hull and White (1990), and the general framework that defines these models is described in the seminal work of Heath et al. (1992). With respect to time-varying volatilities, note that the structure defined in Heath et al. (1992) (henceforth HJM) allows for stochastic volatility structures.

However, the presence of stochastic volatility structures in the class of HJM models leads to two important problems. The first problem arises because these models are often used in the pricing of derivative tools. In this case, the parameters of the HJM model are obtained by the perfect calibration of the interest rate curve observed on a given day through the prices of market instruments. Although this method enables the arbitrage-free pricing of derivative tools, it does not allow for interest rate forecasts, given that this is a cross-sectional fit and does not allow directly obtaining the dynamics observed in the term structure of interest

¹See, for instance, Filipovic (2009) for a review on interest rate modeling.
Bayesian Inference Applied to Dynamic Nelson-Siegel Model with Stochastic Volatility

The second problem is that the imposition of general stochastic volatility structures on no-arbitrage models for the term structure of interest rates often leads to non-Markov dynamics for these interest rates, which hinders the use of econometric methods, such as the Kalman filter, for parameter estimation.

The second class of models, equilibrium models, is obtained by the imposition of equilibrium conditions between the yields of several maturities for the interest rates. These models are usually based on the modeling of the instantaneous forward rate, typically using the structure of affine models, in which the rates of other maturities can be derived by assuming that the risk premium is given by an affine function. These models were developed in Vasicek (1977), Cox et al. (1985), Duffie and Kan (1996) or in Dai and Singleton (2002). Although they are analytically tractable and allow obtaining closed formulas for the pricing of bonds and derivatives, these models have a major problem with the fit and forecasts of the interest rate curve. The literature indicates that these models are usually deficient in these aspects, and the forecasts produced by equilibrium models do not compare to those generated by the random walk model as shown, for example, in Duffie and Kan (1996). In regard to the presence of time-varying volatilities, the imposition of stochastic volatility structures on this class of models leads to problems such as the lack of affine formulations or the necessity for time non-homogeneous models, resulting in the need of calibration methods which, as with no-arbitrage models, prevents the dynamic fit of the interest rate curve and the construction of forecasts. A general discussion about these problems is available in Brigo and Mercurio (2006).

The third class of models includes statistical ones. These models are obtained as purely statistical representations of the evolution of the term structure of interest rates, and are therefore obtained without the direct imposition of no-arbitrage or equilibrium conditions. This class comprises principal components models, factor models or latent variable models, as well as interpolation models. According to Matzner-Lober and Villa (2004), most of the intuition about the dynamics of the prices of bonds and yields comes from these models, as in Litterman and Scheinkman (1991) and in Pearson and Sun (1994). Even though these models are obtained as purely statistical representations, note that for some classes of models, as occurs with the vector autoregressive representation for yields and the generalization of Nelson-Siegel model proposed in Christensen et al. (2010) and Christensen et al. (2009), the imposition of no-arbitrage conditions does not represent gains in the fit or forecasting of these models, as discussed in Joslin et al. (2011) and, consequently, if the objective consists of the fit and forecast of the interest rate curve, the imposition of no-arbitrage conditions is redundant.

Among statistical models, Nelson-Siegel model (Nelson and Siegel, 1987) and its variants are the most popular amidst fixed income managers and central banks. The attractiveness of factor models in Nelson-Siegel class is due to their parsimony and good empirical performance. Such models can capture most of the evolution of
the term structure of interest rates with the use of three factors. Many extensions of Nelson-Siegel models have been proposed (see Almeida et al. (2009), Laurini and Hotta (2010), Bjork and Christensen (1999), Rudebusch and Wu (2008), among others).

The works of Diebold and Li (2006) and Diebold et al. (2006) turned their attention to Nelson-Siegel model and reinterpreted it as a three-factor statistical model to describe the interest rate curve over time. The three factors are interpreted as level, slope, and curvature, being consistent with the interpretation made in Litterman and Scheinkman (1991), providing some economic sense. The major contribution in Diebold and Li (2006) and Diebold et al. (2006) is that the forecasts generated by this class of models are superior to those of several other models available in the term structure of interest rates literature, including time series models, such as vector autoregressive (VAR) models and dynamic error correction models. Diebold et al. (2006) extend the structure by adding non-latent macroeconomic factors, such as inflation and output gap.

Parameter estimation in Diebold and Li (2006) and Diebold et al. (2006) is based on two simplifying assumptions. The factor loadings in Nelson-Siegel model rely on a single weighting parameter (decay). To allow for the estimation of time-varying latent factors in a linear framework, factor loadings are kept constant over time for each maturity. In Diebold and Li (2006), the decay parameter is determined ad hoc, being fixed in the whole sample, and consequently, factor loadings do not vary over time. The second hypothesis, also assumed in Diebold et al. (2006), Yu and Zivot (2011), Almeida et al. (2009), Caldeira et al. (2010), is concerned with volatility, which is maintained constant for all maturities throughout the sampling period.

These assumptions are related to the greater estimation difficulty in the presence of time-varying volatilities and decay (weight) factors. By setting these assumptions aside, estimation by usual methods such as ordinary least squares can no longer be used, as well as simultaneous estimation techniques such as maximum likelihood obtained by the decomposition of the forecast error through the Kalman filter used in Caldeira et al. (2010) since, with time-varying volatilities and decay parameters, it is not possible to obtain a linear and Gaussian state-space model.

Another problem is the high dimensionality of the term structure, which becomes even more problematic with the use of multivariate stochastic volatilities. In this problem, if the volatility structure is utilized for all latent factors, the volatility process is turned into a multivariate process, making estimation problems much more complex, as multivariate stochastic volatility models are plagued by problems with non-trivial inference, as shown, for instance, in Chib et al. (2009). Some ways to overcome this issue were proposed by Bianchi et al. (2009), Koopman et al. (2010), Laurini and Hotta (2010) and Haustsch and Ou (2010). In Koopman et al. (2010), a factor volatility structure is adopted for the latent variables, using a specification based on GARCH models. However, this representation assumes
some simplifications to avoid identification and estimation problems in multivariate GARCH models. The volatility structure is based on an essentially univariate framework, as in this specification the volatility of other latent factors is obtained by the linear drift of only one volatility factor. It should also be remarked that the estimation method used is based on an approximation to the likelihood of the process, which might not have statistical optimality. Another drawback to this model is the incompatibility of GARCH models with continuous time representations, which hinders the interpretation of these models in terms of pricing of financial assets.

In Laurini and Hotta (2010) and Haustsch and Ou (2010), the stochastic volatility process is based on models known as log-normal stochastic volatility models, introduced by Taylor (1986). In Laurini and Hotta (2010) model, the log-normal stochastic volatility specification is assumed only for the measurement errors of the interest rate curve, whereas the latent factor structure is deemed to be homoskedastic. In the specification proposed by Haustsch and Ou (2010), a log-normal stochastic volatility structure is assumed for each latent factor, thus resulting in a more general stochastic volatility process. Note that, in these two approaches, the estimation is based on Bayesian methods using a Markov Chain Monte Carlo method, owing to the difficulty in estimating stochastic volatility models, in which it is not possible to obtain analytical formulas for the evaluation of the likelihood function.

In this paper, we present a generalization for the volatility structures described in Koopman et al. (2010), Laurini and Hotta (2010) and Haustsch and Ou (2010). By using data on interbank deposit (ID) futures contracts traded at the Brazilian Mercantile and Futures Exchange (BM&F) and the U.S. Treasury data from Diebold and Li (2006) and Diebold et al. (2006), we estimate the dynamic Nelson-Siegel model with time-varying factor loadings and stochastic volatility. We take into account two specifications in order to model the conditional volatility. In the first one, analogous to the one utilized by Koopman et al. (2010), the term structure risk is captured via a common stochastic volatility component. In the second alternative, our goal is to model the stochastic volatility of each latent factor in Nelson-Siegel model in a straightforward manner. Modeling volatility directly in terms of factors is a more flexible approach and allows capturing the uncertainty over the level, slope, and curvature of the interest rate. Thus, the so-called level volatility represents the volatility of the general interest rate level, while slope volatility captures the uncertainty over the spread between long- and short-term rates. In turn, curvature volatility is associated with the risk of changes in the curvature of the term structure. In this specification, it is also possible to model the interest rate curve measurement error as a process with stochastic volatility, as in Laurini and Hotta (2010). In this specification, we also discuss the possibility of leverage effects, i.e., the possible correlation between measurement equation errors and the errors of each latent factor, as outlined in Jacquier et al. (1994).
Unlike the specifications used in the formulation of Nelson-Siegel model with stochastic volatility defined in Haustsch and Ou (2010), in both specifications proposed here, the parameter that determines factor loadings, $\lambda_t$, is also regarded as a latent process that varies stochastically over time, being addressed as a fourth latent factor. This addition is important for the modeling of interest rate curves of emerging countries, characterized by frequent changes in formats and maximum maturities in interest rate curves, as discussed in Laurini and Hotta (2010). The dynamics of level, slope, and curvature latent factors, along with the parameter that determines the factor loadings is modeled jointly as an autoregressive process.

Modeling of conditional heteroskedasticity is targeted at improving the quality of fit and forecasting, including the periods of change in volatility, and at capturing the stylized fact about the presence of different volatility structures for each maturity of the term structure, and also at adding stylized facts about the interest rate curves of emerging markets, which usually have higher volatility than interest rate curves of developed countries, characterized by a larger stability over time and lower volatilities.

The paper is organized as follows. Section 2 reviews the basic structure of Nelson-Siegel models. Section 3 presents the two formulations of stochastic volatility models. Section 4 describes the Bayesian estimation method using the Markov Chain Monte Carlo approach. Section 5 provides the term structures of the interest rates used in this study, the ID futures contract and the U.S. term structure of Treasury bonds, the estimation results, and the comparison between models. Finally, Section 6 concludes.

2. Nelson-Siegel Factor Model for the Term Structure of Interest Rates

At a given time, $t$, the interest rate curve, denoted here by $y_t(\tau)$, represents the interest rate as a function of maturities $\tau$. The exponential model of the interest rate curve proposed by Nelson and Siegel (1987) and reinterpreted by Diebold and Li (2006) considers a parametric structure for the evolution of the term structure of interest rates over time, in which the coefficients can be interpreted as level, slope, and curvature factors. The corresponding interest rate curve (in the representation for the spot curve) is:

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

(1)

The shape of the interest rate curve is determined by three parameters ($\beta_{1,t}$, $\beta_{2,t}$ and $\beta_{3,t}$) and by the weights assigned to them. Parameter $\lambda_t$, regarded as fixed in Diebold and Li (2006), governs the exponential decay rate, and small (large) values of $\lambda_t$ are associated with a smooth (fast) decay, and provide longer (shorter) maturities with a better fit. The weight of the first component is 1 (constant) for all maturities, thus $\beta_{1,t}$ stands for the long-term factor and is interpreted as the
Bayesian Inference Applied to Dynamic Nelson-Siegel Model with Stochastic Volatility

interest rate curve level, which equally influences the short- and long-term rates. The weight of the second component, \( \left( \frac{1-e^{\lambda \tau t}}{\lambda \tau} \right) \), begins at 1 and converges to zero monotonically and rapidly, and \( \beta_{2,t} \) is interpreted as the interest rate curve slope, or the short-term factor, given that this factor has a strong influence on short-term interest rates. The weight of the third component, \( \left( \frac{1-e^{\lambda \tau t}}{\lambda \tau} - e^{\lambda \tau t} \right) \), is a concave function, which takes value zero for maturity zero, increases, and then converges monotonically to zero in longer maturities. Therefore, \( \beta_{3,t} \) is interpreted as curvature, or medium-term factor, as it is associated with medium-term interest rates.

At each time period, there are interest rates \( y_t(\tau) \) for several maturities \( \tau \). Therefore, the equation can be estimated at each time period, obtaining time series for parameters \( \beta_t \). Hence, if it is possible to model and predict the behavior of latent factors \( \beta_t \), interest rate curve forecasts can be obtained from these factors. Diebold and Li (2006), Almeida et al. (2009), among others, use a two-step procedure to estimate the model. In the first step, time series are obtained by least squares regressions at each \( t \), and the three time series of \( \beta_{1,t}, \beta_{2,t} \) and \( \beta_{3,t} \) are often modeled by a first-order autoregressive process.

An alternative is to represent the model in state-space form and to estimate all parameters simultaneously. The system, containing the transition and measurement equations, is given by:

\[
y_t(\tau) = \Lambda(\lambda_t)\beta_t + \epsilon_t \quad (2)
\]

where \( \Lambda(\lambda_t) \) is an \((N \times 3)\) matrix for factor loadings, which will vary over time only if the decay parameter \( \lambda \) is variable.

\[
\beta_t = (I - \Phi)\mu_t + \Phi \beta_{t-1} + \eta_t \quad (3)
\]

Measurement equation (2) defines the vector of interest rates \((T \times N)\) for \( N \) maturities, as the sum of factors multiplied by their weights, with a vector of Gaussian-distributed errors that is independent across maturities. The vector \( \beta_t \) of \((3 \times 1)\) size represents the factors, and \( \Phi \) is the coefficient matrix of VAR that models the dynamics of states (latent factors) over time. To estimate the model in linear form through the Kalman filter, for example, the factor loadings need to be kept constant over time for each maturity, yielding \( \lambda_t = \lambda \). For further references on this estimation technique, see, for instance, Caldeira et al. (2010).

3. Generalized Nelson-Siegel Models

3.1 Factor SV model

The first generalization of dynamic Nelson-Siegel models shown in this paper generalizes the original model by adding a time-varying decay parameter \( \lambda \) and a generalized stochastic volatility structure, made up of two blocks. The first block
of the SV structure is an equation that makes the measurement error volatility of the yield observation equation vary over time, whereas the second block describes the evolution of the volatility of latent factors $\beta_{1,t}$, $\beta_{2,t}$ and $\beta_{3,t}$ as a factor stochastic volatility model. This model can be represented by the following system of equations in nonlinear state-space form:

$$y(\tau) = \beta_{1,t} + \beta_{2,t} \left( \frac{1 - e^{\lambda \tau}}{\lambda \tau} \right) + \beta_{3,t} \left( \frac{1 - e^{\lambda \tau}}{\lambda \tau} - e^{\lambda \tau} \right) + \epsilon_{t}$$  \hspace{1cm} (4)

$$\ln \sigma_{\epsilon,t}^2 = \mu_{\sigma} + \phi_{\sigma} \ln \sigma_{\epsilon,t-1}^2 + \epsilon_{\sigma,t}$$  \hspace{1cm} (5)

$$\ln \lambda_{t} = \mu_{\lambda} + \phi_{\lambda} \ln \lambda_{t-1} + \epsilon_{\lambda t}$$  \hspace{1cm} (6)

$$\begin{pmatrix} \beta_{1,t} \\ \beta_{2,t} \\ \beta_{3,t} \end{pmatrix} = \begin{pmatrix} \mu_{1} \\ \mu_{2} \\ \mu_{3} \end{pmatrix} + \begin{pmatrix} \phi_{11} & \phi_{12} & \phi_{13} \\ \phi_{21} & \phi_{22} & \phi_{23} \\ \phi_{31} & \phi_{32} & \phi_{33} \end{pmatrix} \begin{pmatrix} \beta_{1,t-1} \\ \beta_{2,t-1} \\ \beta_{3,t-1} \end{pmatrix} + \begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \\ \epsilon_{3,t} \end{pmatrix}$$  \hspace{1cm} (7)

$$\ln \sigma_{1,t}^2 = \mu_{\sigma,1} + \phi_{\sigma,11} \ln \sigma_{1,t-1}^2 + \epsilon_{\sigma,1,t}$$  \hspace{1cm} (8)

$$\ln \sigma_{2,t}^2 = \alpha_{2} + \gamma_{2} \ln \sigma_{1,t}^2 + \epsilon_{\sigma,2,t}$$  \hspace{1cm} (9)

$$\ln \sigma_{3,t}^2 = \alpha_{3} + \gamma_{3} \ln \sigma_{1,t}^2 + \epsilon_{\sigma,3,t}$$  \hspace{1cm} (10)

In this system, Equation 4 is the Nelson-Siegel observation equation for the observed yields; equation 5 describes the evolution of volatility $\sigma_{\epsilon,t}$ of measurement error $\epsilon_{t}$ as a log-normal stochastic volatility model. Equation 6 shows the evolution of the log of parameter $\lambda_{t}$ as a first-order autoregressive process. The log in this equation is used to guarantee the positive sign of this parameter in the observation equation. We assume that error processes $\epsilon_{\sigma,t}$, $\epsilon_{\lambda t}$, $\epsilon_{\sigma,1,t}$, $\epsilon_{\sigma,2,t}$ and $\epsilon_{\sigma,3,t}$ are independent processes (in time and among themselves) with zero mean and constant variance. The system of equations described in 7 shows the evolution of latent factors $\beta_{1,t}$, $\beta_{2,t}$ and $\beta_{3,t}$ as a vector autoregressive process, with error components given by $\epsilon_{i,t}$, where $i$ stands for each latent factor and $\sigma_{i,t}$ is the time-varying volatility of each latent factor. In this first specification, similarly to the factor GARCH specification used in Koopman et al. (2010), the evolution of the volatility of latent factors is described by a factor model, in which we identify the only latent volatility factor with the volatility of latent factor $\sigma_{1,t}$ for $\beta_{1,t}$. Other normalizations are possible and do not affect the results. In this normalization, factors $\sigma_{2,t}$ and $\sigma_{3,t}$ are obtained as linear functions of latent factor $\sigma_{1,t}$, with intercept parameters given by $\alpha_{2}$ and $\alpha_{3}$ and slope parameters $\gamma_{2}$ and $\gamma_{3}$.
Factor parameterization reduces the dimension of the problem by excluding two additional latent factors for volatilities from the estimation, replacing them with the linear functions of the first factor. Although this parameterization is similar to the one used by Koopman et al. (2010), some important differences are noted. In the parameterization conducted by Koopman et al. (2010), volatility structure is given by a GARCH volatility structure, and thus volatility is a deterministic function of the squared error and of volatility at $t - 1$ plus one intercept. Note that in the model of Koopman et al. (2010), this volatility intercept is defined a priori by identification issues. The estimation of the model proposed by Koopman et al. (2010) is based on quasi-maximum likelihood through the decomposition of the forecast error by the extended Kalman filter, which is based on a local log-linearization of the nonlinear state-space model.

Nevertheless, these series approximations used in the extended Kalman filter could be poor representations of true nonlinear equations and of probability distributions involved in the process, causing problems of filter divergence which usually result in high mean squared errors in the fit of these models, as discussed in Wan and van der Merwe (2000) and Doucet et al. (2001). Owing to these approximations, the quasi-maximum likelihood properties derived from these approximations imply nonoptimal properties in finite samples for parameter estimators and latent factors.

In this paper, we replaced the GARCH model with log-normal stochastic volatility models. The choice of this model is justified by several factors. There is empirical evidence indicating that these models are better at capturing stylized facts about financial series and at forecasting than are other volatility models (e.g., Koopman et al. (2005)). These models also have natural continuous time formulations, which does not occur with conventional GARCH models, and thus their interpretation is straightforward in continuous time asset pricing models. Finally, these additional stochastic volatility components can be formulated as additional latent factors in state-space form and dealt with by the same Bayesian method using Markov Chain Monte Carlo, which is used for the estimation of other latent volatility factors, through a hierarchical formulation.

3.2 SV model with individual factors

While the factor model specified above and the factor model proposed by Koopman et al. (2010) are relatively parsimonious by eliminating two additional factors from the model, these formulations also introduce important restrictions on the dynamics of the volatility of latent factors. By assuming that the dynamics of each latent factor are linear transformations of a single latent volatility factor, the factor model disregards an important aspect concerning the dynamics of Nelson-Siegel models. A remarkable feature in the specification of Nelson-Siegel model is that latent factors can be considered empirical decompositions of level, slope, and curvature. This empirical decomposition implies that the behavior expected
from latent factors is poorly correlated with movements in the level, slope, and curvature. This behavior can be explained by the high correlation among level, slope, and curvature factors obtained by Nelson-Siegel model and the level, slope, and curvature factors obtained by the orthogonal decomposition of principal components in Litterman and Scheinkman (1991). Therefore, even though the decomposition of level, slope, and curvature implies orthogonality for the mean of latent factors, it also affects the movement of the second moment of latent factors; so, assuming the dynamics of a single latent factor for the evolution of volatility can be an invalid restriction.

To avert the imposition of this restriction, the second parameterization used in this paper utilizes a complete evolution structure for the volatility of latent factors, based on the log-normal SV model structure for each of the latent factors, maintaining the other structures used in the previous model (log-normal SV for measurement errors and time-varying decay parameter $\lambda$). The full structure of this model is given by:

$$y(\tau) = \beta_{1,t} + \beta_{2,t} \left(1 - \frac{e^{\lambda \Delta \tau}}{\lambda \Delta \tau}\right) + \beta_{3,t} \left(1 - \frac{e^{\lambda \Delta \tau}}{\lambda \Delta \tau} - e^{\lambda \Delta \tau}\right) + \epsilon_{\epsilon,t} \quad (11)$$

$$\ln \sigma_{\epsilon,t}^2 = \mu_\sigma + \phi_\sigma \ln \sigma_{\epsilon,t-1}^2 + \epsilon_{\sigma,t} \quad (12)$$

$$\ln \lambda_t = \mu_\lambda + \phi_\lambda \ln \lambda_{t-1} + \epsilon_\lambda \quad (13)$$

$$\begin{pmatrix}
\beta_{1,t} \\
\beta_{2,t} \\
\beta_{3,t}
\end{pmatrix} = 
\begin{pmatrix}
\mu_1 \\
\mu_2 \\
\mu_3
\end{pmatrix} + 
\begin{pmatrix}
\phi_{11} & \phi_{12} & \phi_{13} \\
\phi_{21} & \phi_{22} & \phi_{23} \\
\phi_{31} & \phi_{32} & \phi_{33}
\end{pmatrix} 
\begin{pmatrix}
\beta_{1,t-1} \\
\beta_{2,t-1} \\
\beta_{3,t-1}
\end{pmatrix} + 
\begin{pmatrix}
\epsilon_{\beta_{1,t}} \\
\epsilon_{\beta_{2,t}} \\
\epsilon_{\beta_{3,t}}
\end{pmatrix} \quad (14)$$

$$\begin{pmatrix}
\ln \sigma_{1,t}^2 \\
\ln \sigma_{2,t}^2 \\
\ln \sigma_{3,t}^2
\end{pmatrix} = 
\begin{pmatrix}
\mu_{\sigma,1} \\
\mu_{\sigma,2} \\
\mu_{\sigma,3}
\end{pmatrix} + 
\begin{pmatrix}
\phi_{\sigma,11} & \phi_{\sigma,12} & \phi_{\sigma,13} \\
\phi_{\sigma,21} & \phi_{\sigma,22} & \phi_{\sigma,23} \\
\phi_{\sigma,31} & \phi_{\sigma,32} & \phi_{\sigma,33}
\end{pmatrix} 
\begin{pmatrix}
\ln \sigma_{1,t-1}^2 \\
\ln \sigma_{2,t-1}^2 \\
\ln \sigma_{3,t-1}^2
\end{pmatrix} + 
\begin{pmatrix}
\epsilon_{\sigma_{1,t}} \\
\epsilon_{\sigma_{2,t}} \\
\epsilon_{\sigma_{3,t}}
\end{pmatrix} \quad (15)$$

$$\text{var} \begin{pmatrix}
\epsilon_{\epsilon,t} \\
\epsilon_{\sigma_{1,t}} \\
\epsilon_{\sigma_{2,t}} \\
\epsilon_{\sigma_{3,t}}
\end{pmatrix} = \Omega \quad (16)$$

In this parameterization, equations 8, 9 and 10 are substituted into the system of equations described in Eq. 15, and now instead of a single latent factor for the volatility process, each component $\beta_{1,t}$, $\beta_{2,t}$, $\beta_{3,t}$ has a specific stochastic volatility component. In this parameterization, we also allow for leverage of volatility.
Bayesian Inference Applied to Dynamic Nelson-Siegel Model with Stochastic Volatility

processes, enabling a contemporaneous correlation between the error components of stochastic volatility equations of the observation and sign equation, similarly to the stochastic volatility model with leverage defined in Jacquier et al. (1994). The covariance matrix between these factors is defined by equation 16 through the symbol \( \Omega \), and by the addition of another six correlation parameters to the model. Equations 4-9 and 11-16 represent the general models of both classes. We can obtain several submodels by adding restrictions on these general models by assuming that parameter \( \lambda \), for instance, is constant over time, that the matrix of vector autoregressive parameters for latent factors \( \beta_{1,t}, \beta_{2,t}, \beta_{3,t} \) is diagonal and by eliminating the leverage effect, rendering matrix \( \Omega \) diagonal. A specification analysis with the several estimated submodels is shown in Section 6.

4. Bayesian Estimation Based on MCMC

The systems of equations described in equations 4-9, representing the factor SV model, and in equations 11-16, representing the model with individual SVs, are systems represented in nonlinear state-space form. This nonlinearity hinders the use of the conventional estimation method using state-space models, the maximum likelihood obtained by the decomposition of forecast errors through the Kalman filter. The presence of time-varying component \( \lambda_t \) turns the observation equation into a nonlinear function of this latent factor. Note that the specification of this Nelson-Siegel observation equation also hinders the use of this method for the estimation of stochastic volatility components, as the conventional estimation by quasi-maximum likelihood through the Kalman filter is based on a linearization of the observation equation, e.g. Jungbacker and Koopman (2009), which cannot be directly applied to this class of models. A possible alternative is the use of the extended Kalman filter proposed in Anderson and Moore (1979), but it is plagued by divergence problems and poor fits in finite samples, as mentioned earlier.

Given that it is of particular interest to maintain the specification and to avoid approximations, the simultaneous estimation of the proposed models should be performed by Bayesian inference using the Markov Chain Monte Carlo approach. This method allows obtaining estimates for the a posteriori distributions of parameters and latent factors using Markov chain Monte Carlo simulations. Other reasons for the use of Bayesian methods in term structure models are presented in Chib and Ergashev (2009), who point out the advantages of these methods over dimensionality, identification and inference problems in multifactor term structure models. In Bayesian inference procedures, the aim is to estimate the posterior distribution of parameters \( \Theta \) conditional on the observed sample, denoted by \( p(\Theta|y) \). The following expression is used to find the distribution of parameters \( \Theta \) conditional on the sample:

\[
p(\Theta|y) = p(\Theta, y)/p(\Theta) = p(y|\Theta)p(\Theta)/p(y) \tag{17}
\]

Brazilian Review of Econometrics 30(1) May 2010 133
where \( p(y|\Theta) \) is the model’s likelihood, \( p(\Theta) \) is the prior distribution assumed for the parameter and \( p(y) \) is the marginal distribution of the sample. As the sample is observed and fixed, this marginal distribution only has to be known up to a constant of integration, which makes inference procedures easier. Using the Bayes lemma, we can write:

\[
p(\Theta|y) = \frac{p(\Theta, y)}{p(\Theta)} = \frac{p(y|\Theta)p(\Theta)}{c}
\]

Then the posterior distribution is proportional to the product of the likelihood and to the prior distribution:

\[
p(\Theta|y) \propto p(y|\Theta)p(\Theta)
\]

Bayesian estimation methods enable natural interpretation of information updates, as the information in the prior distribution is updated by the information in the sample represented by likelihood, allowing the estimation of the posterior distribution. Bayesian estimation methods also have some important computational advantages over classic methods, especially with regard to complex dimensionality problems. In these contexts, the use of prior information allows circumventing usual identification problems in higher-dimensional models, mainly through the use of hierarchical formulations. Identification, both in classic and Bayesian contexts, is defined as the existence of only one vector of parameters that defines the same likelihood function, e.g. Rothenberg (1971), Kadane (1974), Poirier (1998) and Aldrich (2002). The use of prior information allows estimating unidentified models by reducing the sample space in the posterior distribution of data. As discussed in Florens et al. (1990), the selection of appropriate priors allows estimating posterior distributions econometrically identified by the reduction in the sigma-algebra that generates the posterior distribution. This concept is especially useful in hierarchical models, in which the posterior distribution of each hierarchy serves as the prior of the subsequent level of the model. As state-space models are a type of hierarchical structure, the information on each level allows identifying the other levels, with an adequate choice of priors. This method allows solving usual non-identification problems in classic models by using the Kalman filter for higher dimensional models, as is the case of interest rate latent factor models (e.g. Duffee (2002)). Kim and Orphanides (2005) demonstrate that affine term structure models can be characterized by observationally equivalent (unidentified) estimations, but with different economic interpretations.

Both classes of models discussed in this paper correspond to latent hierarchical models with three levels and six latent processes. Let \( \Theta \) be the collection of parameters of the model, and let \( F_t := (\beta_{1t}, \beta_{2t}, \beta_{3t}) \) and \( V_t := (\ln \sigma^2_{1,t}, \ln \sigma^2_{2,t}, \ln \sigma^2_{3,t}) \). Then, the likelihood function of the model is given by:
Bayesian Inference Applied to Dynamic Nelson-Siegel Model with Stochastic Volatility

\[ p(\Theta | y) = \int_{F_1} \int_{F_2} \ldots \int_{F_T} p(Y | \Theta, F_1, F_2, \ldots, F_T) p(F_1, F_2, \ldots, F_T | \Theta) dF_1 dF_2 \ldots dF_T \]  

(20)

where \( p(Y | \Theta, F_1, F_2, \ldots, F_T) \) denotes the (conditional) density of observations \( Y \) given the parameters \( \Theta \) and latent factors, and \( p(F_1, F_2, \ldots, F_T | \Theta) \) denotes the conditional density of latent factors given the parameters of the model \( \Theta \). Because factors are unobservable, they need to be integrated, resulting in an integral of dimension (3T). Evidently, \( p(F_1, F_2, \ldots, F_T | \Theta) \) depends on a new set of unknown components represented by volatility factors \( V_1, \ldots, V_T \). It is written as:

\[ p(F_1, F_2, \ldots, F_T | \Theta) = \int_{V_1} \int_{V_2} \ldots \int_{V_T} p(F_1, F_2, \ldots, F_T | \Theta, V_1, V_2, \ldots, V_T) \times p(V_1, \ldots, V_T | \Theta) dV_1 dV_2 \ldots dV_T, \]  

(21)

where \( (V_1, \ldots, V_T | \Theta) \) denotes the density of stochastic volatility components.

The structure of priors is given by a univariate normal distribution with the precision matrix given by an inverse gamma distribution for the vector autoregressive parameters of latent factors; the decay parameters follow log-normal distributions in the hierarchical formulation with precision parameters given by an inverse gamma distribution, latent factors of level, slope, and curvature, hierarchical Gaussian priors, also using inverse gamma distributions as precision parameters. The intercept parameters and the persistence of volatility equations are specified by the normal-inverse gamma distribution and in the case of the factor model, in the volatility given by equation (5), the loading matrix is also given by a normal-inverse gamma distribution. The general structure for latent volatility factors is given by hierarchical log-normal priors with inverse gamma precisions. The model’s structure then assumes a multivariate normal distribution for the interest rate curve yields.

The model basically follows the same structure of priors and hyperparameters adopted in Haustsch and Ou (2010), except for decay parameters and time-varying volatilities. Note that the fit of the model basically does not change with the choices of these hyperparameters, which can be understood by the large number of observations in the term structure of interest rates over time. The values of the hyperparameters used in the estimations can be obtained from the authors, and were omitted here for space restrictions.

By defining the structure of prior distributions and of the likelihood, it is possible to retrieve the posterior distribution of parameters and latent factors. Due to the use of nonlinear structures and to the use of nonconjugate priors, it is not possible to obtain posterior distributions in an analytical fashion, and thus it is necessary to use numerical methods. In general, analytical formulas for the posterior distributions are only known in conjugate families, in which prior and
posterior distributions belong to the same family. In other cases, it is necessary to use numerical methods to calculate the posterior distribution of data and, in high dimensionality problems, it is necessary to use Monte Carlo simulations. A successful Monte Carlo method used in high dimensionality problems is the Markov Chain Monte Carlo (MCMC) (e.g. Robert and Casella (2005), Gamerman and Lopes (2006)). MCMC methods are based on the simulation of Markov chains whose stationary (invariant) distribution converges to the posterior distribution of data.

An especially important property to reduce the computational cost of these simulations is that the estimation of \( p(\Theta|y) \) can be factored by the simulation of conditional distributions of parameters, a procedure known as Componentwise Metropolis-Hastings (e.g. Ntzoufras (2009)). These conditional simulations allow separating the total parameter space into smaller blocks. The procedure is shown in the following iterations:

\[
\begin{align*}
 p(\Theta_1|\Theta_2, \Theta_3, ..., \Theta_n, y) \\
p(\Theta_2|\Theta_1, \Theta_3, ..., \Theta_n, y) \\
\vdots \\
p(\Theta_n|\Theta_2, \Theta_3, ..., \Theta_{n-1}, y)
\end{align*}
\]  

(22)

This property is based on the Clifford-Hammersley theorem (Robert and Casella, 2005), which states that under certain conditions for regularity of Markov chains these conditional distributions converge only to the distribution of \( p(\Theta|y) \). Note that this simulation procedure avoids the use of numerical maximization, which is essentially problematic in latent factor models for the term structure of interest rates, as described in Kim and Orphanides (2005).

The factorization used in this paper is defined as follows: Let \( \Omega := \{\Theta, F_1, ..., F_T, V_1, ..., V_T\} \) be a random vector whose posterior distribution \( p(\Omega|Y) \) can be written as:

\[
\begin{align*}
 p(\Omega|Y) &= p(F_1, F_2, ..., F_T, V_1, V_2, ..., V_T, \Theta|Y) \\
 &\propto p(Y|F_1, F_2, ..., F_T, V_1, V_2, ..., V_T, \Theta) \\
 &\times p(F_1, F_2, ..., F_T|V_1, V_2, ..., V_T, \Theta) \\
 &\times p(V_1, V_2, ..., V_T, \Theta) \\
 &\times p(\Theta)
\end{align*}
\]

When all conditional distributions are known, the MCMC algorithm is reduced to the so-called Gibbs sampler, in which the estimation is carried out by the simulation of all analytically known conditional distributions. When some conditional distribution is not known, it is possible to use the Metropolis-Hastings algorithm (e.g. Robert and Casella (2005), Gamerman and Lopes (2006)), which is a generalization of the acceptance-rejection method used with random variables for the
Bayesian Inference Applied to Dynamic Nelson-Siegel Model with Stochastic Volatility simulation of analytically unknown conditional distributions. Therefore, a usual MCMC estimation method uses the Gibbs algorithm to simulate known conditional distributions and the Metropolis-Hastings algorithm to simulate other conditional distributions. Algorithms that use Gibbs and Metropolis-Hastings steps are known as hybrid algorithms, and were initially proposed in Tierney (1994). A hybrid MCMC algorithm (Robert and Casella, 2005) can be seen as iterations as follows:

For $i=1,\ldots,p$ and given $(\theta^{(t+1)}_1, \ldots, \theta^{(t+1)}_{i-1}, \theta^{(t)}_i, \ldots, \theta^{(t)}_p)$,

1 - Sample

$$\tilde{\theta}_i \sim q_i(\theta|\theta^{(t+1)}_1, \ldots, \theta^{(t+1)}_{i-1}, \theta^{(t)}_i, \ldots, \theta^{(t)}_p)$$

2 - Accept

$$\theta^{(t+1)}_i, = \begin{cases} 
\tilde{\theta}_i & \text{with probability } 1 - \rho \\
\theta^{(t)}_i & \text{with probability } \rho 
\end{cases}$$

where

$$\rho = 1 \land \min \left\{ \frac{q_i(\theta|\theta^{(t+1)}_1, \ldots, \theta^{(t+1)}_{i-1}, \theta^{(t)}_i, \ldots, \theta^{(t)}_p)}{q_i(\theta^{(t)}_1, \ldots, \theta^{(t+1)}_{i-1}, \theta^{(t)}_i, \ldots, \theta^{(t)}_p)}, \frac{q_i(\theta^{(t+1)}_1, \ldots, \theta^{(t+1)}_{i-1}, \theta^{(t)}_i, \ldots, \theta^{(t)}_p)}{q_i(\theta^{(t)}_1, \ldots, \theta^{(t+1)}_{i-1}, \theta^{(t)}_i, \ldots, \theta^{(t)}_p)} \right\}$$

where $q$ is the so-called tentative or auxiliary distribution, which is used to approximate the unknown distribution to be simulated.

In the structure of priors assumed here, the Metropolis-Hastings algorithm is necessary to assess the posterior distributions of parameters and latent factors associated with the decay parameter and with stochastic volatility, which represent the nonlinear components of the model. In other cases, conditional distributions are known and the Gibbs algorithm can be used. For further details, see Haustsch and Ou (2010). In the Markov Chain Monte Carlo procedure, we started the estimation with 20,000 burn-in iterations, and used another 20,000 iterations in the construction of posterior distributions. To avoid non-convergence problems in the chains, we used each tenth iteration only, discarding any other iterations on this interval. Convergence is ascertained by the Gelman-Rubin diagnostics. To summarize the results, we used the posterior mean, and to construct the confidence intervals and the 2.5% and 97.5% percentiles of the posterior distribution obtained by the MCMC procedure.

5. Data

In this section, we describe the databases used in the estimation of the models. For the empirical assessment of interest rate curves, we used two different databases, one with daily data and closing prices of the interest rates in ID futures contracts traded at BM&F. The other database consisted of monthly data
obtained from the prices of U.S. treasury bonds, and also closing prices (at the end of the month). The database is available from website: www.ssc.upenn.edu/~fdiebold.

The ID futures contract with maturity $\tau$ is a futures contract of which the basic asset is the interest rate accumulated on a daily basis, capitalized between trading time $t$ and $\tau$. The value of the contract is given by its value at maturity, R$100,000.00 discounted according to the accumulated interest rate, traded between the seller and the buyer of the contract. When buying an ID futures contract for the price at time $t$ and keeping it up to maturity $\tau$, the gain or loss is given by:

$$100,000 \left( \prod_{i=1}^{\zeta(t,\tau)} \frac{(1 + y_i)}{(1 + DI^*)^{\frac{1}{365}}} - 1 \right)$$

where $y_i$ denotes the ID rate $(i-1)$ days after the trading day. The function $\zeta$ represents the number of days between $t$ and $\tau$.

The ID futures contract is very similar to a zero-coupon bond, except for the fact that marginal calls are paid on a daily basis. Every day the cash flow is the difference between the adjustment price (liquidation) of the current day and the adjustment price of the previous day, corrected by the ID rate of the previous day.

The ID futures contracts are traded at BM&F, which determines the number of maturities with authorized contracts. In general, there are around 20 maturities with authorized contracts every day, but not all of them have liquidity. Often, approximately 10 maturities have contracts with higher liquidity. There are contracts with monthly maturities for the months that begin each quarter (January, April, July, and October). In addition, there are contracts with maturities for the four months that follow the current month. The maturity date is the first business day of the month on which the contract is due.

The data used in this paper consist of daily observations of the interest rates of ID futures contracts, closing prices. In practice, contracts with all maturities are not observed every day. Therefore, from the observed rates, the data were converted to fixed maturities of 1, 2, 3, 4, 6, 9, 12, 15, 18, 24, 27, 29, 31 and 33 months, using cubic spline interpolations. The database contains the maturities with highest liquidity for January 2006 through February 2009, totaling 756 observations.

Table 1 displays the descriptive statistics for the Brazilian interest rate curve. Mean, standard deviation, minimum, maximum, and some autocorrelation coef-
Bayesian Inference Applied to Dynamic Nelson-Siegel Model with Stochastic Volatility

Table 1
Descriptive statistics of the Brazilian interest rate curve. Sampling period from 2006:01 to 2009:02

<table>
<thead>
<tr>
<th>Maturity</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Minimum</th>
<th>Maximum</th>
<th>( \hat{\rho} (5) )</th>
<th>( \hat{\rho} (21) )</th>
<th>( \hat{\rho} (63) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>13.126</td>
<td>1.731</td>
<td>11.05</td>
<td>17.71</td>
<td>0.995</td>
<td>0.894</td>
<td>0.672</td>
</tr>
<tr>
<td>2</td>
<td>13.082</td>
<td>1.677</td>
<td>11.04</td>
<td>17.60</td>
<td>0.996</td>
<td>0.890</td>
<td>0.662</td>
</tr>
<tr>
<td>3</td>
<td>13.050</td>
<td>1.620</td>
<td>11.02</td>
<td>17.40</td>
<td>0.997</td>
<td>0.885</td>
<td>0.650</td>
</tr>
<tr>
<td>4</td>
<td>13.027</td>
<td>1.566</td>
<td>11.02</td>
<td>17.21</td>
<td>0.997</td>
<td>0.879</td>
<td>0.635</td>
</tr>
<tr>
<td>6</td>
<td>13.028</td>
<td>1.534</td>
<td>10.90</td>
<td>16.72</td>
<td>0.996</td>
<td>0.866</td>
<td>0.597</td>
</tr>
<tr>
<td>9</td>
<td>13.058</td>
<td>1.534</td>
<td>10.78</td>
<td>16.54</td>
<td>0.993</td>
<td>0.858</td>
<td>0.575</td>
</tr>
<tr>
<td>12</td>
<td>13.121</td>
<td>1.567</td>
<td>10.68</td>
<td>16.47</td>
<td>0.991</td>
<td>0.851</td>
<td>0.560</td>
</tr>
<tr>
<td>15</td>
<td>13.193</td>
<td>1.600</td>
<td>10.60</td>
<td>16.91</td>
<td>0.987</td>
<td>0.849</td>
<td>0.553</td>
</tr>
<tr>
<td>18</td>
<td>13.248</td>
<td>1.614</td>
<td>10.46</td>
<td>17.15</td>
<td>0.985</td>
<td>0.845</td>
<td>0.550</td>
</tr>
<tr>
<td>24</td>
<td>13.287</td>
<td>1.615</td>
<td>10.37</td>
<td>17.36</td>
<td>0.982</td>
<td>0.840</td>
<td>0.543</td>
</tr>
<tr>
<td>27</td>
<td>13.311</td>
<td>1.625</td>
<td>10.28</td>
<td>17.57</td>
<td>0.978</td>
<td>0.836</td>
<td>0.536</td>
</tr>
<tr>
<td>29</td>
<td>13.321</td>
<td>1.630</td>
<td>10.21</td>
<td>17.72</td>
<td>0.975</td>
<td>0.831</td>
<td>0.527</td>
</tr>
<tr>
<td>31</td>
<td>13.325</td>
<td>1.632</td>
<td>10.15</td>
<td>17.83</td>
<td>0.972</td>
<td>0.828</td>
<td>0.521</td>
</tr>
<tr>
<td>33</td>
<td>13.326</td>
<td>1.633</td>
<td>10.13</td>
<td>17.87</td>
<td>0.966</td>
<td>0.827</td>
<td>0.517</td>
</tr>
</tbody>
</table>

Note: The last three columns show the autocorrelations for the lags of 1 week, 1 month and 3 months.

For the empirical assessment of the proposed generalizations, we also used the same database analyzed by Diebold and Li (2006), Diebold et al. (2006) and Koopman et al. (2010). The database consists of the unsmoothed forward rates described in Fama and Bliss (1987), obtained from U.S. Treasury bonds, with fixed maturities of 3, 6, 9, 12, 15, 18, 21, 24, 30, 36, 48, 60, 72, 84, 96, 108 and 120 months.

Table 2 shows the descriptive statistics for the U.S. Treasury database. Mean, standard deviation, minimum, maximum, and some autocorrelation coefficients are reported for each maturity. Some common characteristics shared by the interest rate curves can be observed, such as positive slope, longer maturities with smaller volatility and lower persistence than shorter maturities, and level as the most persistent component.

Figures 1 and 2 show the evolution of the Brazilian and U.S. interest rate over time. In both cases, there is a wide time range in the interest rate curve level, with less intense variations in slope and curvature. In the case of the U.S. interest rate curve, the decline in level over time is quite evident, which is not that evident in the Brazilian case because of the database used, which covers a shorter time period. In both cases, databases with fixed maturities are used.
Table 2
Descriptive statistics for the U.S. Treasury curve. Sampling period from 1985:01 to 2000:12

<table>
<thead>
<tr>
<th>Maturity (Months)</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Minimum</th>
<th>Maximum</th>
<th>( \hat{\rho}(1) )</th>
<th>( \hat{\rho}(12) )</th>
<th>( \hat{\rho}(30) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>5.630</td>
<td>1.488</td>
<td>2.732</td>
<td>9.131</td>
<td>0.978</td>
<td>0.969</td>
<td>-0.079</td>
</tr>
<tr>
<td>6</td>
<td>5.785</td>
<td>1.482</td>
<td>2.891</td>
<td>9.324</td>
<td>0.976</td>
<td>0.555</td>
<td>-0.042</td>
</tr>
<tr>
<td>9</td>
<td>5.907</td>
<td>1.492</td>
<td>2.984</td>
<td>9.343</td>
<td>0.973</td>
<td>0.545</td>
<td>-0.005</td>
</tr>
<tr>
<td>12</td>
<td>6.067</td>
<td>1.501</td>
<td>3.107</td>
<td>9.683</td>
<td>0.969</td>
<td>0.539</td>
<td>0.021</td>
</tr>
<tr>
<td>15</td>
<td>6.225</td>
<td>1.504</td>
<td>3.288</td>
<td>9.988</td>
<td>0.968</td>
<td>0.527</td>
<td>0.060</td>
</tr>
<tr>
<td>18</td>
<td>6.308</td>
<td>1.496</td>
<td>3.482</td>
<td>10.188</td>
<td>0.965</td>
<td>0.513</td>
<td>0.089</td>
</tr>
<tr>
<td>21</td>
<td>6.375</td>
<td>1.484</td>
<td>3.638</td>
<td>10.274</td>
<td>0.963</td>
<td>0.502</td>
<td>0.115</td>
</tr>
<tr>
<td>24</td>
<td>6.401</td>
<td>1.464</td>
<td>3.777</td>
<td>10.413</td>
<td>0.960</td>
<td>0.481</td>
<td>0.133</td>
</tr>
<tr>
<td>30</td>
<td>6.550</td>
<td>1.462</td>
<td>4.043</td>
<td>10.748</td>
<td>0.957</td>
<td>0.479</td>
<td>0.190</td>
</tr>
<tr>
<td>36</td>
<td>6.644</td>
<td>1.439</td>
<td>4.204</td>
<td>10.787</td>
<td>0.956</td>
<td>0.471</td>
<td>0.226</td>
</tr>
<tr>
<td>48</td>
<td>6.838</td>
<td>1.439</td>
<td>4.308</td>
<td>11.269</td>
<td>0.951</td>
<td>0.457</td>
<td>0.294</td>
</tr>
<tr>
<td>60</td>
<td>6.928</td>
<td>1.439</td>
<td>4.347</td>
<td>11.313</td>
<td>0.951</td>
<td>0.464</td>
<td>0.336</td>
</tr>
<tr>
<td>72</td>
<td>7.082</td>
<td>1.457</td>
<td>4.384</td>
<td>11.653</td>
<td>0.953</td>
<td>0.454</td>
<td>0.372</td>
</tr>
<tr>
<td>84</td>
<td>7.142</td>
<td>1.425</td>
<td>4.352</td>
<td>11.841</td>
<td>0.948</td>
<td>0.448</td>
<td>0.391</td>
</tr>
<tr>
<td>96</td>
<td>7.226</td>
<td>1.413</td>
<td>4.433</td>
<td>11.512</td>
<td>0.954</td>
<td>0.468</td>
<td>0.417</td>
</tr>
<tr>
<td>108</td>
<td>7.27</td>
<td>1.428</td>
<td>4.429</td>
<td>11.664</td>
<td>0.953</td>
<td>0.475</td>
<td>0.426</td>
</tr>
<tr>
<td>120</td>
<td>7.254</td>
<td>1.432</td>
<td>4.443</td>
<td>11.663</td>
<td>0.953</td>
<td>0.467</td>
<td>0.428</td>
</tr>
</tbody>
</table>

Note: The last three columns show the autocorrelations for the lags of 1, 12, and 30 months.

Figure 1
Dynamics of the term structure over time This graph illustrates the evolution of the term structure of interest rates (ID futures contract database) for the period from 2006:01 to 2009:02

Note: The sample consists of daily data for maturities of 1, 2, 3, 4, 6, 9, 12, 15, 18, 24, 27, 29, 31 and 33 months.
Bayesian Inference Applied to Dynamic Nelson-Siegel Model with Stochastic Volatility

Figure 2

Dynamics of the term structure over time This graph illustrates the evolution of the term structure of interest rates (U.S. Treasury database) for the period from 1985:01-2000:12

Note: The sample consists of monthly data for maturities of 3, 6, 9, 12, 15, 18, 21, 24, 30, 36, 48, 60, 72, 84, 96, 108 and 120 months.

6. Analysis of Results

In this section, we examine the adjustment of the generalizations proposed for the volatility structure. For the sake of comparison, we also estimate some alternative parameterizations: structure proposed by Diebold and Li (2006), two-step estimation with fixed \( \lambda \) (P1) defined \textit{a priori}; same parameterization as before, but including the decay factor \( \lambda \) as time-varying, without taking into account the conditional heteroskedasticity (P2).

We also estimate the model with the same Diebold-Li parameterization, with fixed \( \lambda \) defined \textit{a priori}, modeling the conditional heteroskedasticity with specific stochastic volatility factors (P3); same specification as before but with fixed \( \lambda \), estimated together with the other parameters (P4); model with varying \( \lambda \) and common stochastic volatility factor (P5); model with varying \( \lambda \) and stochastic volatility for each latent factor, considering two alternative specifications for the dynamics of factors, AR(1) and VAR(1) (P6 and P7); parameterization with total leverage, i.e., full correlation matrix across all stochastic volatilities (P8). Table 6 shows the number of parameters estimated in each parameterization.

6.1 ID futures contract database

At first, we estimated the model using the two-step procedure, with a fixed \( \lambda \) parameter and disregarding the conditional heteroskedasticity (parameterization
Table 3
Number of parameters estimated in each model

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Total</th>
<th>P1</th>
<th>P2</th>
<th>P3</th>
<th>P4</th>
<th>P5</th>
<th>P6</th>
<th>P7</th>
<th>P8</th>
</tr>
</thead>
<tbody>
<tr>
<td>λ</td>
<td>—</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td></td>
<td>6</td>
</tr>
<tr>
<td>σ</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SV</td>
<td>—</td>
<td>9</td>
<td>9</td>
<td>7</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td></td>
<td>9</td>
</tr>
<tr>
<td>VAR (β’s)</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>12</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Correl</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>6</td>
<td></td>
</tr>
</tbody>
</table>

Note: σ refers to the variances of β’s and of the measurement error for the models estimated by the two-step method. The correlations refer to the specification with leverage (full correlation matrix across all SVs).

Parameter λ, which determines the shape of the interest rate curve, was chosen such that it minimizes the root mean squared error (RMSE) for in-sample fit (λ = 0.1046).

Table 4 shows the RMSE for the fit of the ID futures interest rate curve, taking into account all the analyzed parameterizations. The RMSE is expressed as basis points (bps). Note that even the model estimated by the two-step procedure, using OLS, captures a substantial part of the dynamics of the interest rate curve, especially for medium-term maturities. The average RMSE for this parameterization is 0.45 bps. However, the presence of high autocorrelation in the residuals indicates that important information on mean and variance is being neglected.

In the second parameterization, the decay parameter, λ, is deemed to be time-varying and is estimated jointly with the other factors (Parameterization P2). Thus, the model has four latent factors, β₁ᵗ, β₂ᵗ, β₃ᵗ, and β₄ᵗ (λᵗ); so, it is nonlinear and is estimated by Bayesian inference (MCMC) in a single step. This specification does not include conditional heteroskedasticity. Considering that λ varies over time significantly improves the quality of fit, and the RMSE mean drops to 0.07. In the third parameterization (P3), the parameter λ is defined a priori, but conditional heteroskedasticity is modeled by a structure with specific stochastic volatility component for each latent factor. Taking into account the conditional variance of factors separately substantially improves the quality of fit, and the RMSE mean drops from 0.45 to 0.15 bps (Table 4).

In another alternative, we estimate the model in a single step and deal with conditional variance using specific stochastic volatility components, as in the previous case (P4). Nonetheless, in this case, the decay factor, λ, is estimated jointly with the other parameters and kept constant for all time periods. As highlighted earlier, in this case, the model is nonlinear and the parameters are estimated simul-

4This parameterization is often used as a benchmark for alternative specifications.
Bayesian Inference Applied to Dynamic Nelson-Siegel Model with Stochastic Volatility

### Table 4
RMSE of the Estimated Models (ID futures contract database). Period between 2006:01 and 2009:02

<table>
<thead>
<tr>
<th>Maturity Months</th>
<th>P1</th>
<th>P2</th>
<th>P3</th>
<th>P4</th>
<th>P5</th>
<th>P6</th>
<th>P7</th>
<th>P8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Short</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P1</td>
<td>0.5100</td>
<td>0.0473</td>
<td>0.1453</td>
<td>0.0420</td>
<td>0.0348</td>
<td>0.0337</td>
<td>0.0342</td>
<td>0.0346</td>
</tr>
<tr>
<td>P2</td>
<td>0.0880</td>
<td>0.0447</td>
<td>0.0990</td>
<td>0.0503</td>
<td>0.0481</td>
<td>0.0472</td>
<td>0.0462</td>
<td>0.0454</td>
</tr>
<tr>
<td>P3</td>
<td>0.0900</td>
<td>0.0664</td>
<td>0.0693</td>
<td>0.0673</td>
<td>0.0658</td>
<td>0.0665</td>
<td>0.0653</td>
<td>0.0652</td>
</tr>
<tr>
<td>P4</td>
<td>0.3800</td>
<td>0.0655</td>
<td>0.1051</td>
<td>0.0655</td>
<td>0.0657</td>
<td>0.0658</td>
<td>0.0658</td>
<td>0.0654</td>
</tr>
<tr>
<td>P5</td>
<td>0.2600</td>
<td>0.0649</td>
<td>0.2166</td>
<td>0.0621</td>
<td>0.0639</td>
<td>0.0638</td>
<td>0.0638</td>
<td>0.0638</td>
</tr>
<tr>
<td>P6</td>
<td>0.2600</td>
<td>0.0649</td>
<td>0.2166</td>
<td>0.0621</td>
<td>0.0639</td>
<td>0.0638</td>
<td>0.0638</td>
<td>0.0638</td>
</tr>
<tr>
<td>P7</td>
<td>0.2600</td>
<td>0.0649</td>
<td>0.2166</td>
<td>0.0621</td>
<td>0.0639</td>
<td>0.0638</td>
<td>0.0638</td>
<td>0.0638</td>
</tr>
<tr>
<td>P8</td>
<td>0.2600</td>
<td>0.0649</td>
<td>0.2166</td>
<td>0.0621</td>
<td>0.0639</td>
<td>0.0638</td>
<td>0.0638</td>
<td>0.0638</td>
</tr>
<tr>
<td>Medium</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P1</td>
<td>0.4100</td>
<td>0.0778</td>
<td>0.1098</td>
<td>0.0648</td>
<td>0.0657</td>
<td>0.0518</td>
<td>0.0508</td>
<td>0.0524</td>
</tr>
<tr>
<td>P2</td>
<td>0.2100</td>
<td>0.0763</td>
<td>0.1525</td>
<td>0.0894</td>
<td>0.0861</td>
<td>0.0830</td>
<td>0.0813</td>
<td>0.0820</td>
</tr>
<tr>
<td>P3</td>
<td>0.3400</td>
<td>0.0905</td>
<td>0.0715</td>
<td>0.0398</td>
<td>0.0339</td>
<td>0.0354</td>
<td>0.0344</td>
<td>0.0354</td>
</tr>
<tr>
<td>P4</td>
<td>0.4800</td>
<td>0.0589</td>
<td>0.1241</td>
<td>0.0470</td>
<td>0.0386</td>
<td>0.0324</td>
<td>0.0328</td>
<td>0.0302</td>
</tr>
<tr>
<td>P5</td>
<td>0.5400</td>
<td>0.0860</td>
<td>0.1682</td>
<td>0.0726</td>
<td>0.0658</td>
<td>0.0574</td>
<td>0.0591</td>
<td>0.0571</td>
</tr>
<tr>
<td>P6</td>
<td>0.5700</td>
<td>0.0976</td>
<td>0.1965</td>
<td>0.0945</td>
<td>0.0872</td>
<td>0.0786</td>
<td>0.0805</td>
<td>0.0794</td>
</tr>
<tr>
<td>P7</td>
<td>0.5500</td>
<td>0.0973</td>
<td>0.1962</td>
<td>0.0942</td>
<td>0.0871</td>
<td>0.0785</td>
<td>0.0803</td>
<td>0.0792</td>
</tr>
<tr>
<td>P8</td>
<td>0.5700</td>
<td>0.0976</td>
<td>0.1965</td>
<td>0.0945</td>
<td>0.0872</td>
<td>0.0786</td>
<td>0.0805</td>
<td>0.0794</td>
</tr>
<tr>
<td>Long</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P1</td>
<td>0.3200</td>
<td>0.0371</td>
<td>0.0391</td>
<td>0.0686</td>
<td>0.0633</td>
<td>0.0818</td>
<td>0.0635</td>
<td></td>
</tr>
<tr>
<td>P2</td>
<td>0.4100</td>
<td>0.0702</td>
<td>0.1486</td>
<td>0.0670</td>
<td>0.0586</td>
<td>0.0584</td>
<td>0.0584</td>
<td>0.0580</td>
</tr>
<tr>
<td>P3</td>
<td>0.4900</td>
<td>0.0708</td>
<td>0.1571</td>
<td>0.0665</td>
<td>0.0634</td>
<td>0.0604</td>
<td>0.0605</td>
<td>0.0605</td>
</tr>
<tr>
<td>P4</td>
<td>0.5600</td>
<td>0.0708</td>
<td>0.1571</td>
<td>0.0665</td>
<td>0.0634</td>
<td>0.0604</td>
<td>0.0605</td>
<td>0.0605</td>
</tr>
<tr>
<td>P5</td>
<td>0.4900</td>
<td>0.0708</td>
<td>0.1571</td>
<td>0.0665</td>
<td>0.0634</td>
<td>0.0604</td>
<td>0.0605</td>
<td>0.0605</td>
</tr>
<tr>
<td>P6</td>
<td>0.4900</td>
<td>0.0708</td>
<td>0.1571</td>
<td>0.0665</td>
<td>0.0634</td>
<td>0.0604</td>
<td>0.0605</td>
<td>0.0605</td>
</tr>
<tr>
<td>P7</td>
<td>0.4900</td>
<td>0.0708</td>
<td>0.1571</td>
<td>0.0665</td>
<td>0.0634</td>
<td>0.0604</td>
<td>0.0605</td>
<td>0.0605</td>
</tr>
<tr>
<td>P8</td>
<td>0.4900</td>
<td>0.0708</td>
<td>0.1571</td>
<td>0.0665</td>
<td>0.0634</td>
<td>0.0604</td>
<td>0.0605</td>
<td>0.0605</td>
</tr>
</tbody>
</table>

The last line contains the number of times the parameterization outperformed the other ones. Boldface indicates better performance at maturity. RMSE expressed as basis points. P1 = Two-step estimation (2S) with a fixed $\lambda$; P2 = One-step estimation (1S) with time-varying $\lambda$; P3 = Estimation (2S), with $\lambda$ defined a priori and SV for each factor; P4 = Estimation (1S), with estimated $\lambda$ and SV for each factor; P5 = Estimation (1S), time-varying $\lambda$ and factor SV; P6 and P7 = Estimation (1S), time-varying $\lambda$, SV for each factor and dynamics of factors modeled by AR(1) and VAR(1), respectively; P8 = Estimation (1S), parameterization with full leverage.

The quality of fit improves compared to the ones considered previously, and the RMSE mean is 0.067, outperforming the other ones, mainly at longer maturities (Table 4).

The first alternative specification proposed for Diebold and Li (2006) model consists in including a common stochastic volatility factor in measurement, as defined in equations (8, 9 and 10), discussed in Section 3. In this case, we also regard the decay factor $\lambda_t$ as a fourth latent factor, $\beta_4$ (Parameterization P5). The vector $\beta_t$ is then modeled by a VAR process (equation 14), as in Koopman et al. (2010). Figure 3 shows the time series for the latent factors, including the decay parameter, $\lambda$.

In addition to being useful for the pricing and management of fixed income assets, modeling the conditional volatility of the interest rate curve, in this case using a common stochastic volatility component, improves the fit of the model, and the RMSE mean drops to 0.06 bps. Note that in the case of the Brazilian interest rate curve, the parameterization has a better performance for 5 out of the 14 maturities analyzed, especially for the intermediate portion of the curve. Besides taking into account the conditional variance of the interest rate curve, regarding the decay factor $\lambda_t$ as time-varying allows observing changes in the term structure over time as, for instance, the maturity at which the curvature factor loading is maximized.

Figure 4 shows the estimated factor and measurement error volatilities. Observe that the measurement error volatility is especially high in the second half.
Figure 3
*Estimates of Latent and Decay Factors, Parameterization P5*

Note: Specification with common stochastic volatility component and time-varying $\lambda_t$. 
Figure 4
Time-varying volatility. Parameterization P5

Note: Specification with common stochastic volatility component and time-varying $\lambda_t$. 
of 2008, period which coincides with the most critical phase of the international financial crisis. The estimated stochastic volatility component demonstrates the capacity of the model to capture the conditional heteroskedasticity observed in interest rates, identifying the moments of change, such as inverted shape, at the end of 2006 and in the last quarter of 2008, periods marked by changes in monetary policy conduct.

The second generalization consists in estimating the model by using conditional heteroskedasticity separately in latent factors and in measurement errors. In this specification, both factor volatility and the decay factor $\lambda_t$ are deemed to be time-varying. Figure 5 shows the estimates of latent factors, including the decay factor. In this case, full (Parameterization P6) and diagonal (Parameterization P7) specifications were considered for the transition equation that models the factor dynamics, but it should be noted that in the latter case the dynamics of $\lambda_t$ is modeled by an AR(1). A parameterization with total leverage, with full correlation matrix across all stochastic volatilities, was also used (Parameterization P8). Once again, note that the estimated factors describe well the changes in the interest rate curve level observed in Figure 1.

Figure 5
Estimates of Latent and Decay Factors. Parameterization P7

Note: Diagonal specification for the equation of the transition matrix and AR(1) for $\lambda_t$.

The last three columns of Table 4 show the RMSE for the second change in the
Bayesian Inference Applied to Dynamic Nelson-Siegel Model with Stochastic Volatility

model, with the full and diagonal VAR matrix for the factor dynamics (P6 and P7), in addition to leverage for volatility components (P8). Estimation details are discussed in Section 4. Note that with this parameterization the model has the lowest RMSE means (0.0584, 0.0581 and 0.0580, respectively). Compared to the specification that imposes a common volatility factor, it has a better fit for 9 out of 14 maturities, especially the longest ones, outperforming other parameterizations in all vertices.

The evolution of volatility factors, $\ln(h_t^{\beta_1})$, $\ln(h_t^{\beta_2})$ and $\ln(h_t^{\beta_3})$, and the volatility of measurement errors, $\sigma_{\epsilon,t}$, are shown in Figure 6.

Figure 6
Specific Time-Varying Stochastic Volatility Factors. Parameterization P7

![](image)

Note: Specification with factor stochastic volatility and time-varying $\lambda_t$.

Confidence intervals for the estimation of parameters related to the dynamics of latent factors and stochastic volatility factors, model specified by equations (5-10) and model specified by equations (12-16), are shown in Table 5. In both cases, the latent factors are highly persistent. In addition, we found strong evidence of high autocorrelation in volatility factors, with autoregressive coefficients of 0.998, 0.997 and 0.998 for level, slope, and curvature, respectively.
### Table 5: 95% Confidence Intervals (Parameterization P5)

<table>
<thead>
<tr>
<th>Brazil</th>
<th>Par</th>
<th>$\mu$</th>
<th>$\beta_{11}$</th>
<th>$\beta_{21}$</th>
<th>$\beta_{31}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\beta_{11}$</td>
<td>0.0921</td>
<td>0.392</td>
<td>0.106</td>
<td>-0.011</td>
</tr>
<tr>
<td></td>
<td>$\beta_{21}$</td>
<td>-0.043</td>
<td>0.278</td>
<td>0.944</td>
<td>0.006</td>
</tr>
<tr>
<td></td>
<td>$\beta_{31}$</td>
<td>-0.058</td>
<td>0.372</td>
<td>-0.089</td>
<td>1.001</td>
</tr>
<tr>
<td>$\sigma$s</td>
<td>$\sigma_{11}$</td>
<td>0.062</td>
<td>—</td>
<td>0.988</td>
<td>—</td>
</tr>
<tr>
<td>SV</td>
<td>$\sigma_{21}$</td>
<td>0.710</td>
<td>0.985</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>$\sigma_{31}$</td>
<td>0.87</td>
<td>0.870</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

Note: 95% confidence intervals in brackets. Parameterization using VAR for latent factors and specific SV.

### Table 6: 95% Confidence Intervals (Parameterization P7)

<table>
<thead>
<tr>
<th>Brazil</th>
<th>Par</th>
<th>$\mu$</th>
<th>$\beta_{11}$</th>
<th>$\beta_{21}$</th>
<th>$\beta_{31}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\beta_{11}$</td>
<td>0.094</td>
<td>0.392</td>
<td>0.106</td>
<td>-0.011</td>
</tr>
<tr>
<td></td>
<td>$\beta_{21}$</td>
<td>-0.043</td>
<td>0.278</td>
<td>0.944</td>
<td>0.006</td>
</tr>
<tr>
<td></td>
<td>$\beta_{31}$</td>
<td>-0.058</td>
<td>0.372</td>
<td>-0.089</td>
<td>1.001</td>
</tr>
<tr>
<td>$\sigma$s</td>
<td>$\sigma_{11}$</td>
<td>0.062</td>
<td>—</td>
<td>0.988</td>
<td>—</td>
</tr>
<tr>
<td>SV</td>
<td>$\sigma_{21}$</td>
<td>0.710</td>
<td>0.985</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>$\sigma_{31}$</td>
<td>0.87</td>
<td>0.870</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

Note: 95% confidence intervals in brackets. Parameterization using VAR for latent factors and specific SV.

Confidence intervals for the stochastic volatility parameters of measurement errors, $\sigma_{\epsilon,t}$, and for the dynamics of the decay factor, $\lambda_t$, are shown in Table 7. 95% confidence intervals for $\lambda_t$ and $\sigma_{\epsilon,t}$ (BRA).
Table 7: 95% confidence intervals for $\lambda_t$ and $\sigma_{\epsilon,t}$ (BRA)

<table>
<thead>
<tr>
<th>Model</th>
<th>Specific SV</th>
<th>Factor SV</th>
</tr>
</thead>
<tbody>
<tr>
<td>CI</td>
<td>$\mu_\lambda$</td>
<td>$\phi_\lambda$</td>
</tr>
<tr>
<td>$C^0_{10}$ 0.25</td>
<td>-0.204</td>
<td>0.964</td>
</tr>
<tr>
<td>$C^0_{10}$ 0.50</td>
<td>0.233</td>
<td>0.982</td>
</tr>
</tbody>
</table>

Note: 95% confidence intervals for the decay factor and for the measurement error variance.

Therefore, it seems clear that regarding the parameter $\lambda_t$ as fixed and ignoring the presence of conditional heteroskedasticity, although it facilitates the estimation procedures, implies loss of the quality of fit. Results also indicate that the maturity for which curvature has maximum weight is not constant over time. Thus, keeping $\lambda$ constant throughout the sampling period could be very restrictive, as the data cover a long time period. In particular, the maturity at which the curvature factor $\beta_3t$ is maximized and the speed of decay of the slope factor $\beta_2t$ depend only on parameter $\lambda$, and are maintained constant as a consequence. However, these features of the interest rate curve can change over time, especially for the Brazilian interest rate curve, which went through several changes in the past few years owing to changes in macroeconomic fundamentals.

In the analysis in which we used ID futures contract data, we concluded that regarding the decay parameter as time-varying and including conditional volatility improve the quality of fit. The extension with specific stochastic volatility components had the best performance.

6.2 U.S. Treasury database

In this section, we estimate the same parameterizations for the U.S. Treasury database described in Section 5. Table 8 shows the RMSE for each maturity, taking into account all the estimated versions of the model.

The first parameterization of the model (P1) follows Diebold and Li (2006), with a decay parameter $\lambda$ equal to 0.0609. The model also was estimated using the time-varying decay parameter $\lambda$, regarded as a fourth latent factor, without considering conditional heteroskedasticity (P2). Unlike the ID futures contract database, this parameterization does not improve the fit of the model compared to the previous case, and the RMSE mean is then slightly higher (equal to 0.0881 bps).

The model also was estimated using the parameter $\lambda$ defined a priori and a stochastic volatility structure with specific component for each factor (P3). This parameterization outperforms the previous ones in terms of RMSE at all maturities, as shown in Table 8, highlighting the importance of conditional heteroskedasticity. In the subsequent specification, the model is estimated in a single step using the fixed $\lambda$ parameter, jointly estimated with the other factors (P4). The same structure is used for the volatility of the previous case. In this case, the fit
Table 8

<table>
<thead>
<tr>
<th>Maturity Months</th>
<th>P1</th>
<th>P2</th>
<th>P3</th>
<th>P4</th>
<th>P5</th>
<th>P6</th>
<th>P7</th>
<th>P8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Short</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.1240</td>
<td>0.1459</td>
<td>0.0847</td>
<td>0.0870</td>
<td>0.0681</td>
<td>0.0687</td>
<td>0.0716</td>
<td>0.0658</td>
</tr>
<tr>
<td>6</td>
<td>0.0706</td>
<td>0.0702</td>
<td>0.0668</td>
<td>0.0755</td>
<td>0.0659</td>
<td>0.0667</td>
<td>0.0664</td>
<td>0.0674</td>
</tr>
<tr>
<td>9</td>
<td>0.0841</td>
<td>0.0851</td>
<td>0.0811</td>
<td>0.0829</td>
<td>0.0746</td>
<td>0.0741</td>
<td>0.0743</td>
<td>0.0731</td>
</tr>
<tr>
<td>Medium</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>0.0879</td>
<td>0.0905</td>
<td>0.0866</td>
<td>0.0667</td>
<td>0.0674</td>
<td>0.0704</td>
<td>0.0719</td>
<td>0.0695</td>
</tr>
<tr>
<td>18</td>
<td>0.0728</td>
<td>0.0758</td>
<td>0.0596</td>
<td>0.0465</td>
<td>0.0543</td>
<td>0.0528</td>
<td>0.0539</td>
<td>0.0522</td>
</tr>
<tr>
<td>21</td>
<td>0.0586</td>
<td>0.0624</td>
<td>0.0386</td>
<td>0.0339</td>
<td>0.0378</td>
<td>0.0381</td>
<td>0.0389</td>
<td>0.0372</td>
</tr>
<tr>
<td>24</td>
<td>0.0702</td>
<td>0.0777</td>
<td>0.0519</td>
<td>0.0717</td>
<td>0.0538</td>
<td>0.0560</td>
<td>0.0537</td>
<td>0.0540</td>
</tr>
<tr>
<td>30</td>
<td>0.0575</td>
<td>0.0633</td>
<td>0.0411</td>
<td>0.0610</td>
<td>0.0388</td>
<td>0.0392</td>
<td>0.0382</td>
<td>0.0416</td>
</tr>
<tr>
<td>36</td>
<td>0.0678</td>
<td>0.0728</td>
<td>0.0582</td>
<td>0.0830</td>
<td>0.0561</td>
<td>0.0552</td>
<td>0.0520</td>
<td>0.0558</td>
</tr>
<tr>
<td>Long</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>48</td>
<td>0.0667</td>
<td>0.0682</td>
<td>0.0681</td>
<td>0.0770</td>
<td>0.0625</td>
<td>0.0607</td>
<td>0.0606</td>
<td>0.0638</td>
</tr>
<tr>
<td>60</td>
<td>0.0857</td>
<td>0.0836</td>
<td>0.0779</td>
<td>0.1018</td>
<td>0.0708</td>
<td>0.0696</td>
<td>0.0690</td>
<td>0.0699</td>
</tr>
<tr>
<td>72</td>
<td>0.0886</td>
<td>0.0940</td>
<td>0.0821</td>
<td>0.0817</td>
<td>0.0761</td>
<td>0.0779</td>
<td>0.0794</td>
<td>0.0778</td>
</tr>
<tr>
<td>84</td>
<td>0.0892</td>
<td>0.0855</td>
<td>0.0611</td>
<td>0.0716</td>
<td>0.0623</td>
<td>0.0602</td>
<td>0.0614</td>
<td>0.0606</td>
</tr>
<tr>
<td>96</td>
<td>0.0999</td>
<td>0.1026</td>
<td>0.0578</td>
<td>0.0586</td>
<td>0.0549</td>
<td>0.0532</td>
<td>0.0544</td>
<td>0.0553</td>
</tr>
<tr>
<td>108</td>
<td>0.117</td>
<td>0.1181</td>
<td>0.0613</td>
<td>0.0644</td>
<td>0.0558</td>
<td>0.0579</td>
<td>0.0555</td>
<td>0.0542</td>
</tr>
<tr>
<td>120</td>
<td>0.1237</td>
<td>0.1340</td>
<td>0.0807</td>
<td>0.0884</td>
<td>0.0864</td>
<td>0.0797</td>
<td>0.0782</td>
<td>0.0811</td>
</tr>
<tr>
<td>Mean</td>
<td>0.0843</td>
<td>0.0881</td>
<td>0.0647</td>
<td>0.0710</td>
<td>0.0604</td>
<td>0.0605</td>
<td>0.0601</td>
<td>0.0605</td>
</tr>
<tr>
<td>Median</td>
<td>0.0841</td>
<td>0.0856</td>
<td>0.0613</td>
<td>0.0717</td>
<td>0.0623</td>
<td>0.0602</td>
<td>0.0606</td>
<td>0.0606</td>
</tr>
</tbody>
</table>

Note: The last line shows the number of times the parameterization outperformed the other ones. Boldface indicates better performance at maturity. RMSE expressed as basis points. P1 = Two-step procedure (2S) with constant $\lambda$; P2 = One-step procedure (1S) with time-varying $\lambda$; P3 = 2S estimation with $\lambda$ defined a priori and SV for each factor; P4 = 1S estimation with estimated $\lambda$ and SV for each factor; P5 = 1S estimation with time-varying $\lambda$ and factor SV; P6 and P7 = 1S estimation with time-varying $\lambda$, SV for each factor, and factor dynamics modeled by AR(1) and VAR(1), respectively; P8 = 1S estimation, parameterization with total leverage. 1.
Bayesian Inference Applied to Dynamic Nelson-Siegel Model with Stochastic Volatility

improves marginally for some medium-term maturities, but the RMSE mean is higher than in the previous case.

The two generalizations proposed for the model were also estimated. The first one utilizes a common stochastic volatility component and the second one includes specific stochastic volatility components for each factor (P5). Thus, as with ID futures contract data, the results show that the extension of the model with specific volatility components improves the quality of fit, mainly for longer maturities, outperforming the most basic specifications. The parameterization with a common stochastic volatility component has a lower RMSE than other parameterizations in 4 out of 17 maturities, with a mean RMSE of 0.0604 bps, outperforming the parameterizations assessed previously.

Figure 7

Estimates for Latent and Decay Factors. Parameterization P5

Figure 7 shows the time series for latent factors estimated using a specific volatility component and a time-varying decay factor. Note that the downtrend of the interest rate curve level observed in Figure 2 is well captured by the estimated level component, $\beta_1$, which has a clear downtrend over time. On the other hand, parameter $\lambda$ assumes values between 0.05 and 0.10 most of the time, and thus the weight of component $\beta_3$ is maximized to maturities between 20 and 30 months.

Figure 8 shows the estimates for the measurement error and factor volatilities. The volatility components are especially high in the second half of the 1980s,
taking on moderate values in the subsequent period.

Figure 8

Time-Varying Volatility. Parameterization P5

Note: Factor and Measurement Error Volatilities.

The last three columns in Table 8 display the RMSE for all the maturities of the parameterization that includes specific stochastic volatility components for each latent factor. As with the case of ID futures contract data, complete (P7) and diagonal (P7) specifications were considered for the transition equation that models the factor dynamics, but in the latter case the dynamics of $\lambda_t$ is modeled by an AR(1). A parameterization with total leverage (P8), full correlation matrix across all stochastic volatilities, was also estimated. Specifications with specific volatility components outperform other specifications in 10 out of 17 maturities, especially in terms of the longest portion of the interest rate curve. Note that, in the case of the U.S. Treasury interest rate curve, the best fit was obtained by modeling the factor dynamics with a VAR(1), with mean RMSE of 0.0601 basis points.

Figure 9 shows the time series for the latent and decay factors for the model estimated with specific volatility components and complete VAR for the factor dynamics. Once again, note that the downward trend in the general interest rate curve level is well captured by the component $\beta_1$. Also in this case, parameter $\lambda_t$ ranges between 0.05 and 0.10 most of the time. Figure 10 depicts the behavior of
specific volatility factors and measurement error variance at each $t$.

Figure 9

Estimates of Latent Factors and Decay Parameter. Parameterization P7

Note: Diagonal specification for the transition equation matrix and AR(1) for the decay factor.

Again, the results provide evidence of how important it is to take into account conditional variance in interest rate curve models. Moreover, the results found here, using both the ID futures contract and the U.S. Treasury databases, indicate that imposing a specific volatility component structure for each latent factor of Diebold and Li (2006) model and regarding the decay factor as time-varying has significant gains in quality of fit. However, unlike Brazilian data, regarding the parameter $\lambda$ as time-varying does not have significant gains in quality of fit in the U.S. Treasury database. With respect to conditional heteroskedasticity, the quality of fit improved significantly in both cases.

Tables 9 and 10 show the confidence intervals for the factor dynamics parameters for the model with factor stochastic volatility and for the model with specific stochastic volatility factors, respectively. Observe that the imposition of a volatility component structure slightly changes the persistence of factors, increasing the autoregressive components of latent factors in both cases. Parameters for the dynamics of volatility factors, ($h_t^{\beta_1}, h_t^{\beta_2}$ and $h_t^{\beta_3}$), are also shown, all with high persistence.
Figure 10

Specific Time-Varying Stochastic Volatility Factor

Note: Stochastic volatility latent factors and measurement error volatility.
Table 9: 95% confidence intervals (Parameterization P5)

<table>
<thead>
<tr>
<th>EUA Par</th>
<th>$\beta_{1t}$</th>
<th>$\beta_{2t}$</th>
<th>$\beta_{3t}$</th>
<th>$\beta_{s}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_{1t}$</td>
<td>0.001</td>
<td>0.983</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>$\beta_{2t}$</td>
<td>0.000</td>
<td>0.992</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>$\beta_{3t}$</td>
<td>-0.01</td>
<td>917</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

Note: 95% confidence intervals in brackets. Parameterization using VAR for latent factors and specific SV.

Table 10: 95% confidence intervals (Parameterization P7)

<table>
<thead>
<tr>
<th>Esp. Par</th>
<th>$\beta_{1t}$</th>
<th>$\beta_{2t}$</th>
<th>$\beta_{3t}$</th>
<th>$\beta_{s}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_{1t}$</td>
<td>0.001</td>
<td>0.981</td>
<td>-0.005</td>
<td>0.000</td>
</tr>
<tr>
<td>$\beta_{2t}$</td>
<td>0.000</td>
<td>-0.174</td>
<td>0.934</td>
<td>0.072</td>
</tr>
<tr>
<td>$\beta_{3t}$</td>
<td>-0.001</td>
<td>0.049</td>
<td>0.015</td>
<td>0.813</td>
</tr>
</tbody>
</table>

Note: 95% confidence intervals in brackets. Parameterization using VAR for latent factors and specific SV.
Table 11 shows the confidence intervals for the stochastic volatility parameters of measurement errors, $\sigma_{\epsilon,t}$, and for the decay factor dynamics, $\lambda_t$.

Table 11: 95% confidence intervals for $\lambda_t$ and $\sigma_{\epsilon,t}$

<table>
<thead>
<tr>
<th>Model</th>
<th>Specific SV</th>
<th>Factor SV</th>
</tr>
</thead>
</table>
| CI    | $\mu_\lambda$ | $\phi_\lambda$ | $\mu_{\sigma_{\epsilon}}$ | $\phi_{\sigma_{\epsilon}}$ | $\mu_\lambda$ | $\phi_\lambda$ | $\mu_{\sigma_{\epsilon}}$ | $\phi_{\sigma_{\epsilon}}$
| CI_{0.025} | -0.807 | 0.353 | 4.094 | 0.997 | -0.447 | 0.834 | 0.205 | 0.997 |
| CI_{0.50} | -1.186 | 0.716 | 4.475 | 1.001 | -0.284 | 0.894 | 0.221 | 1.000 |
| CI_{0.975} | -0.501 | 0.823 | 4.082 | 1.004 | -0.089 | 0.965 | 0.243 | 1.070 |

Note: 95% confidence intervals for the decay factor and measurement error variance.

7. Conclusion

In this paper, we proposed two extensions for Nelson and Siegel (1987) term structure model developed by Diebold and Li (2006). Using two different databases – the first one containing data on ID futures contracts traded at BM&F and the same database used by Diebold and Li (2006), Diebold et al. (2006) and Koopman et al. (2010), containing U.S. Treasury data – we estimated Nelson-Siegel model with factor loadings and time-varying volatility. In the first specification, we followed the structure proposed by Koopman et al. (2010), using a common stochastic volatility component for measurement errors, modeled by a stochastic volatility process. In the second generalization, we extended the formulation introduced by Haustsch and Ou (2010) and proposed modeling the stochastic volatility of latent factors in Nelson and Siegel (1987) model in a straightforward fashion, regarding factor loadings as time-varying. This structure allows capturing the interest rate curve volatility in a flexible and also parsimonious way. The model is represented in state-space form, in which the factors in the transition equation have stochastic volatility and the stochastic volatility components follow a dynamic latent process. As the model is nonlinear, we proposed the use of Bayesian inference, based on the MCMC method, to estimate Nelson-Siegel model with stochastic volatility.

In the specification with a common volatility factor, as proposed by Koopman et al. (2010), this factor can be interpreted as the volatility of a benchmark bond portfolio, structured as a GARCH model as in Engle et al. (1990) and Engle and Ng (1993). This specification does not allow different volatilities in individual interest rates. In the specification proposed herein, the volatility factor associated with the interest rate curve level $h_{t}^{\beta_{1}}$ is a common variance component as in the joint volatility factor in Koopman et al. (2010) and can be seen as an implicit proxy for the volatility of a market bond as that used by Engle et al. (1990). Volatility factor $h_{t}^{\beta_{2}}$ captures the changes in spread over time between the rates obtained for long and short maturities and is interpreted as slope volatility. Similarly, $h_{t}^{\beta_{3}}$ is the volatility of a portfolio of bonds with medium-term maturities, which captures the uncertainty associated with the interest rate curve.
Koopman et al. (2010) estimated Nelson-Siegel model with two changes, compared to what Diebold and Li (2006) proposed; the first one is that the model is estimated with a time-varying decay factor, and the second one is that they consider the interest rate volatility to be time-varying using a common (GARCH) volatility component. They demonstrated that regarding parameter $\lambda$ as time-varying improves the fit, but that modeling volatility has more significant gains in terms of fit. Our results followed the same pattern when we regarded the decay parameter as time-varying, as the fit improved, mainly in the case of the ID futures contract database. Just as in Koopman et al. (2010), the most significant gains were obtained when we considered the conditional heteroskedasticity of the interest rate curve, either by using a common stochastic volatility component or mainly by imposing a structure with specific volatility components for each factor, which is not allowed in the specification proposed by Koopman et al. (2010).

The specification proposed herein also has some advantages over the inference methods used in Koopman et al. (2010). The Bayesian method allows for accurate inferences in finite samples for parameters and latent factors, whereas the method of Koopman et al. (2010) is based on a series of approximations and linearizations, which could lead to losses when fitting the data and to problems with convergence and local maxima in the likelihood function. Also note that the stochastic volatility model allows for continuous time representation, which does not occur in GARCH models.

Our estimation results provide evidence of different time-varying stochastic volatilities for interest rate curve factors, with highly persistent volatility factors. In particular, the level factor is the one that most contributes to interest rate volatility. With respect to the ID futures contract database, regarding the decay parameter as time-varying improves the quality of fit. However, specifications that include a common volatility component or which deal with latent factor volatility separately are superior to others, with lower RMSE for all maturities. For shorter maturities, the model with common volatility components has a better quality of fit. On the other hand, for longer maturities, the specification with specific volatility components outperforms the other ones, especially with respect to the results obtained for the U.S. Treasury interest rate curve.

In addition, modeling the specific stochastic volatility of each factor can give important information about risk assessment in the portfolio of public and derivative securities. Alternatives that could be addressed in forthcoming studies include additional factors for curvature and slope, as proposed by Laurini and Westin (2010), rendering the model more flexible, as well as other specifications for the volatility structure, such as multivariate or regime switch models. It should be noted that the findings obtained here are valid for in-sample fit, as an analysis of out-of-sample forecasts could yield different results, thus favoring more parsimonious specifications, but we highlight that the focus of the present paper was on checking which volatility structure is more suitable to the interest rate curve.
behavior observed and not on carrying out straightforward out-of-sample forecasts.

References


