A Bayesian Term tructure Modeling using heavy-tailed distributions

C. A. Abanto-Valle†[1], V. H. Lachos‡ and Pulak Ghosh§

†Department of Statistics, Federal University of Rio de Janeiro, Brazil

‡Department of Statistics, Campinas State University, Brazil

§Department of QM & IS, Indian institute of Management, Bangalore

Abstract

In this paper we introduce the three-factor model of Diebold and Li (2006) using the class of symmetric scale mixtures of normal distributions is introduced to analyze the term structure of interest rates. Specific distributions examined include the multivariate normal, Student-t, slash and variance gamma distributions. In the face of non-normality, these distributions provide an appealing robust alternative to the use of the normal distribution. Using a Bayesian paradigm, an efficient Markov chain Monte Carlo algorithm is developed for parameter estimation. Moreover, the mixing parameters obtained as a by-product of the scale mixture representation can be used to identify outliers. Bayesian model selection criteria as well as out-of- sample forecasting results reveal that the Diebold and Li (DL) models based on Student-t distribution provide significant improvement in model fit as well as prediction to the US yields data over the usual normal model.

keywords: Term structure, interest rates, scale mixture of normal distributions, Markov chain Monte Carlo, state space models.

1 Introduction

The term structure of interest rates of the bonds describes how interest rates evolve over time. The interest rate term structures of both government and corporate bonds have important applications in economics and finance, including pricing fixed income derivatives, in risk management and in designing macroeconomic policies. Obtaining accurate estimates of both Treasury and corporate term structures is thus essential. The recent literature has shown the major advances in the theoretical models for predicting the term structure of the interest rates. However, modeling and predicting the term structure of interest rates is not straightforward (de Pooter et al., 2006). Approaches to term structure models include no-arbitrage and equilibrium models. Equilibrium models usually start with modeling the short rate process, and output the term structure for all times t. No-arbitrage models start with an initial (today's) term structure (or a forward rate curve) and they provide an evolution of this process in future. During the last decades significant progress has been made in modeling the term structure, mainly in the no-arbitrage factor models. The literature on no-arbitrage factor dates back to the seminal papers of Vasicek (1977) and Cox et al. (1985), characterized by Duffie and Kan (1996) and classified by Dai and Singleton (2000). Affine models identify a small number of latent factors that can be extracted from the panel of yields for different maturities and impose cross-equation restrictions that rule out arbitrage opportunities. Affine models, provided they are properly specified, have been shown to accurately fit

[1]Correspondence to: Carlos A. Abanto-Valle, Department of Statistics, Federal University of Rio de Janeiro, Caixa Postal 68530, CEP: 21945-970, RJ, Brazil. E-mail: cabantovalle@im.ufrj.br

the term structure, see for example Dai and Singleton (2000).

Nelson and Siegel (1987) developed a model for term structure moving away from the no-arbitrage interest models and Diebold and Li (2006) and Moench (2006) have shown this kind of model can be beneficial to reflect the state of the economy. The Nelson-Siegel models have the potential of providing satisfactory answers to both the questions of accurate estimation of the current term structure of interest and to forecast the future term structure. Nelson and Siegel (1987) proposed to fit the term structure using a flexible, smooth parametric function. They demonstrated that their proposed model is capable of capturing many of the typically observed shapes that the yield curve assumes over time. Since then various extensions have been proposed that incorporate additional flexibility with a popular extension being the Svensson (1994) model. Due to these successes it is not surprising that the Nelson-Siegel model is increasingly being used in other applications as well. For example, Diebold et al. (2006b) use the model to study the interactions between the macroeconomy and the yield curve (see also Diebold et al., 2005) whereas Diebold et al. (2006a) apply it to identify systematic risk sources and to construct a generalized duration measure. Academic literature and practitioner oriented publications show that the parametric yield curve model suggested by Nelson and Siegel (1987) fits data well. Due to its intuitive appeal and implementational easiness the Nelson-Siegel model has grown to be very popular among practitioners, in particular in central banks. Recently, Diebold and Li (2006) have shown that the three-factor Nelson-Siegel model can also be used to construct accurate term structure forecasts. By using a straightforward two-step estimation procedure they demonstrate that the model performs well, relative to competing models, especially for longer forecast horizons. Moench (2008) partially confirms these results and Fabozzi et al. (2005) show that the Nelson-Siegel model produces forecasts that are not only statistically accurate but also economically meaningful as these can be used to generate substantial investment returns.

Much of this research, however, focuses, solely focuses on the original three-factor Nelson-Siegel model, where a basic assumption is the use of the normal distribution as the basis for parameter inference. Unfortunately, such normality assumptions are too restrictive and suffer from the lack of robustness in the presence of outliers, and thus can have an important effect on the inferences. Extensions of this model to derive robust estimates have not yet been investigated. It is well known that bond prices often exhibit heavy tails with possible outliers (Schwartz, 1998; Jarrow et al., 2004). Schwartz (1998) uses a robust measure and finds that almost 10% of the US Treasury securities in the Fixed Income database are outliers. According to Schwartz (1998), there could be outliers in bond prices in the term structure data due to a variety of reasons: out-of-date bond ratings; large spreads in illiquid bonds; nonsynchronous trades or data entry errors; use of the wrong interest rate function. However, the use of a piecewise constant curve in Schwartz (1998) for the interest rate gives large fitted bond pricing errors. There has been few methods to derive a robust term-structure model based on splines , e.g., McCulloch (1971, 1975), Shea (1984), Chambers et al. (1984), Jarrow et al. (2004) and Li and Yu (2006). But these spline based methods are not robust to the distributional assumption and thus may not accurately estimate the term structure model in the presence of skewness or outliers.

In this article we develop a state space framework for the term-structure model introduced by Nelson and Siegel (1987) and re-interpreted by Diebold and Li (2006) (henceforth DL) as a modern three-factor model of level, slope and curvature . We further address the issue of robustness by extending the DL model with a flexible class of scale mixtures of normal (SMN) distributions (Andrews and Mallows, 1974; Lange and Sinsheimer, 1993; Chow and Chan, 2008). The use of the SMN class of distributions is motivated by the following considerations: (1) the DL model can have substantial non-normality, like outliers which may not be captured by the routine use of normal distribution; (2) the class of SMN distributions is a rich class as it contains as proper elements the normal, the Student-t, the slash and variance gamma distribution. All these distributions have heavier tails than the normal one, and thus can be used for robust inference in these type of models. We refer to this generalization of the SMN class for DL models as DL–SMN distributions; (3) this generalization is easy to implement and has the potential to make the inference robust to departures from a normal distribution. Thus, in this paper the use of SMN distribution extend the existing methods and develop a more flexible model for the DL model. Also, in this setup, since the likelihood function depends on high-dimensional integrals, we develop a novel Bayesian simulation-based method for estimating the parameters of the model without evaluating the likelihood function. The method is based on the Gibbs sampler a Markov Chain Monte Carlo (MCMC) algorithm.

The remainder of the paper proceeds as follows. Section 2 gives a brief description of SMN distributions. Section 3 outlines the general class of the DL-SMN models as well as the Bayesian estimation procedure using MCMC methods. Section 4 shows an application on the US yield data. Finally section 5 concludes.

2 Scale mixtures of normal distribution

A random vector has a $\mathbf{Y}$ has a scale mixtures of normal (SMN) distribution (Andrews and Mallows, 1974), if it can be expressed as follows

$$\mathbf{Y} = \boldsymbol{\mu} + \kappa^{1/2}(U)\mathbf{Z}, \tag{1}$$

where $\boldsymbol{\mu}$ is a location vector, $\mathbf{Z}$ is a normal random vector with mean vector $\mathbf{0}$, variance–covariance matrix $\boldsymbol{\Sigma}$, $\kappa(.)$ is a weight function and U is a mixing positive random variable with cumulative distribution function (cdf) $P(u \mid \boldsymbol{\nu})$ and probability density function (pdf) $p(u \mid \boldsymbol{\nu})$, independent of $\mathbf{Z}$, where $\boldsymbol{\nu}$ is a scalar or parameter vector indexing the distribution of U. Given U, $\mathbf{Y}$ follows a multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and variance–covariance matrix $\kappa(u)\boldsymbol{\Sigma}$, i.e., $\mathbf{Y}|U = u \sim \mathcal{N}_p(\boldsymbol{\mu}, \kappa(u)\boldsymbol{\Sigma})$. Then, the pdf of $\mathbf{Y}$ is given by

$$f(\mathbf{y}) \quad = \quad \int_0^\infty \mathcal{N}_p(\mathbf{y}|\boldsymbol{\mu}, \kappa(u)\boldsymbol{\Sigma}) dP(u \mid \boldsymbol{\nu}), \tag{2}$$

where $\mathcal{N}_p(.|\boldsymbol{\mu}, \kappa(u)\boldsymbol{\Sigma})$ stands for the pdf of the p–variate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. One particular case of this distribution is the normal distribution, for which H is degenerate, with $\kappa(u) = 1$, $u > 0$. In the sequel, we denote $\mathrm{SMN}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}; P)$ as the SMN distribution with the pdf (2). We notice that when $k(u) = u^{-1}$ in (1), the distribution of $\mathbf{Y}$ reduces to the NI distribution family discussed, for instance, in Lange and Sinsheimer (1993).

The symmetrical class of SMN distribution include distributions such as the Student–t, the slash, the contaminated–normal and the variance gamma distributions; all these distributions have heavier tails than the normal ones and can be used for robust inference. In the next, we present some special cases of SMN distributions. Other members of SMN distributions can be found in Andrews and Mallows (1974). However, for many of the cases, the scale distribution $P(u \mid \boldsymbol{\nu})$ does not have a computationally attractive form and thus it will not be dealt with in this work.

2.1 Multivariate Student–t distribution

The multivariate Student–t distribution with ν degrees of freedom, $\mathcal{T}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \nu)$, can be derived from the mixture model (2), by taking $\kappa(u) = 1/u$, with $U \sim \mathcal{G}(\nu/2, \nu/2)$, $u > 0$, $\nu > 0$, where $\mathcal{G}(.,.)$. denotes the gamma distribution. The pdf of $\mathbf{Y}$ takes the following form:

$$f(\mathbf{y}) = \frac{\Gamma(\frac{p+\nu}{2})}{\Gamma(\frac{\nu}{2})\pi^{p/2}} \nu^{-p/2} |\Sigma|^{-1/2} \left(1 + \frac{d}{\nu}\right)^{-(p+\nu)/2},$$

for $\mathbf{y} \in \mathbb{R}^p$. When $\nu \uparrow \infty$, we get the normal distribution as the limiting case. That is, $Y \sim \mathcal{T}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$ is equivalent to the following hierarchical form:

$$Y \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu, u \sim \mathcal{N}\left(\boldsymbol{\mu}, u^{-1}\boldsymbol{\Sigma}\right), \qquad u \mid \nu \sim \mathcal{G}(\nu/2, \nu/2). \tag{3}$$

2.2 Multivariate slash distribution

Another SMN distribution, termed multivariate slash distribution and denoted by $\mathcal{S}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \nu)$, arise when $\kappa(u) = 1/u$ and the distribution of $U \sim \mathcal{B}e(\nu, 1)$, $0 < u < 1$ and $\nu > 0$, where $\mathcal{B}e(.,.)$ denotes the beta distribution. Its pdf is given by

$$f(\mathbf{y}) = \nu \int_0^1 u^{\nu-1} \mathcal{N}_p(\mathbf{y}|\boldsymbol{\mu}, u^{-1}\boldsymbol{\Sigma}) du, \quad \mathbf{y} \in \mathbb{R}^p.$$

The slash distribution reduces to the normal distribution when $\nu \uparrow \infty$. Thus, the slash distribution is equivalent to the following hierarchical form:

$$Y \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu, u \sim \mathcal{N}\left(\mu, u^{-1}\boldsymbol{\Sigma}\right), \qquad u \mid \nu \sim \mathcal{B}e(\nu, 1). \tag{4}$$

The slash distribution has been mainly used in simulation studies because it represents some extreme situations depending on the value of ν; see for example Andrews et al. (1972), Gross (1973), Morgenthaler and Tukey (1991) and Wang and Genton (2006).

2.3 Multivariate Variance Gamma distribution

The variance gamma distribution, $Y \sim \mathcal{VG}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$, $\nu > 0$ was first proposed by Madan and Seneta (1990) to model share market returns. The VG distribution is controlled by the shape parameter $\nu > 0$, presents heavier tails than those of the normal distribution and has a similar SMN density representation to the student-t distribution. It can be shown that the VG density can be expressed as

$$f(\mathbf{y}) = \int_0^\infty \mathcal{N}_p(\mathbf{y}|\boldsymbol{\mu}, u^{-1}\boldsymbol{\Sigma}) \mathcal{IG}(u|\frac{\nu}{2}, \frac{\nu}{2}) d\lambda. \tag{5}$$

Thus, the VG distribution is equivalent to the following hierarchical form:

$$\mathbf{Y}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, u \sim \mathcal{N}\left(\boldsymbol{\mu}, u^{-1}\boldsymbol{\Sigma}\right), \qquad u|\nu \sim \mathcal{IG}(\frac{\nu}{2}, \frac{\nu}{2}), \tag{6}$$

where $\mathcal{IG}(.,.)$ denotes the inverse gamma distribution.

3 Heavy-tailed version of the DL model

Diebold and Li (2006) show that a generalized Nelson-Siegel model accurately approximates yield curve dynamics and produces good forecasts. The resulting specification is given by

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

where $y_t(\tau)$ is the continuously-compounded zero-coupon nominal yield at maturity τ, $\epsilon_t(\tau)$ is a disturbance with zero mean and standard deviation $\sigma_t(\tau)$. In this setup β_{1t}, β_{2t}, β_{3t} can be considered as latent factors and interpreted as level, slope and curvature. The parameter λ governs the exponential decay toward zero of the factor loadings on β_{2t} and β_{3t}. This generalized Nelson-Siegel model is very flexible and capable of accommodating several stylized facts on the term structure and its dynamics.

In this article, we extend the framework of Diebold and Li (2006) in order to capture heavy-tailed features in the marginal distribution of the disturbances by using the SMN class of distributions. The resulting model is written in a general vectorial fashion using a state space model framework, that is

$$\mathbf{y}_t \quad = \quad \boldsymbol{\Lambda}\boldsymbol{\beta}_t + U_t^{-1/2}\boldsymbol{\epsilon}_t \tag{7}$$

$$\boldsymbol{\beta}_t - \boldsymbol{\mu} \quad = \quad \mathbf{A}(\boldsymbol{\beta}_{t-1} - \boldsymbol{\mu}) + \boldsymbol{\eta}_t \tag{8}$$

$$U_t \quad \sim \quad p(u \mid \boldsymbol{\nu}) \tag{9}$$

where $\mathbf{y}_t = [y_t(\tau_1), y_t(\tau_2), \ldots, y_t(\tau_r)]'$ is the vector of term structure's observed at time t at maturities $\tau_1, \ldots, \tau_r$, $\boldsymbol{\beta}_t = (\beta_{1t}, \beta_{2t}, \beta_{3t})'$ is the latent factor, U_t is the mixing variable, $\boldsymbol{\Lambda}$ is the matrix of loadings which has the following structure

$$\boldsymbol{\Lambda} \quad = \quad \begin{pmatrix} 1 & \frac{1-\mathrm{e}^{-\lambda\tau_1}}{\lambda\tau_1} & \frac{1-\mathrm{e}^{-\lambda\tau_1}}{\lambda\tau_1} - \mathrm{e}^{-\lambda\tau_1} \\ 1 & \frac{1-\mathrm{e}^{-\lambda\tau_2}}{\lambda\tau_2} & \frac{1-\mathrm{e}^{-\lambda\tau_2}}{\lambda\tau_2} - \mathrm{e}^{-\lambda\tau_2} \\ \vdots & \vdots & \vdots \\ 1 & \frac{1-\mathrm{e}^{-\lambda\tau_r}}{\lambda\tau_r} & \frac{1-\mathrm{e}^{-\lambda\tau_r}}{\lambda\tau_r} - \mathrm{e}^{-\lambda\tau_r} \end{pmatrix}, \tag{10}$$

$\boldsymbol{\mu}$ is the unconditional mean of $\boldsymbol{\beta}_t$, $\mathbf{A}$ is the transition matrix of the latent factors, $\boldsymbol{\epsilon}_t = [\epsilon_t(\tau_1), \epsilon_t(\tau_2), \ldots, \epsilon_t(\tau_r)]'$. We assume that $\boldsymbol{\epsilon}_t \sim \mathcal{N}_r(0, \mathrm{diag}(\sigma_1^2, \ldots, \sigma_r^2))$ and $\boldsymbol{\eta}_t \sim \mathcal{N}_3(0, \mathbf{W})$.

The class of model given by equations (7)-(9) is named the DL-SMN models. As depicted in Section 2, this class of models includes the DL with normal (DL-N), student-t (DL-t), slash (DL-S) and variance gamma distributions (DL-VG) as special cases. All these distributions have heavier tails than the normal density and thus provide an appealing robust alternative to the usual Gaussian process in DL framework. Under a Bayesian paradigm, we use MCMC methods to conduct the posterior analysis in the next subsection. Conditionally to U_t, some derivations are common to all members of the DL-SMN family (see Appendix for details).

3.1 Parameter estimation via MCMC

A Bayesian approach to parameter estimation of the DL-SMN class of models relies on simulation based methods. We propose an MCMC algorithm to make Bayesian analysis feasible.

Let $\boldsymbol{\theta} = (\lambda, \boldsymbol{\mu}', \text{vec}(\mathbf{A})', \text{vech}(\mathbf{W})', \{\sigma_j^2\}_{j=1}^r, \nu)'$ be the vector of parameters of the DL-SMN class of models, $\mathbf{U}_{1:T} = (U_1, \ldots, U_T)'$ the mixing variables, $\boldsymbol{\beta}_{0:T} = (\boldsymbol{\beta}_0', \boldsymbol{\beta}_1', \ldots, \boldsymbol{\beta}_T')'$ the latent factors and $\mathbf{y}_{1:T} = (\mathbf{y}_1', \ldots, \mathbf{y}_T')'$ all the information available up time T. The Bayesian approach for estimating the parameters in the DL-SMN class of models uses the data augmentation principle, which considers $\boldsymbol{\beta}_{0:T}$ and $\mathbf{U}_{1:T}$ as a latent variables. By Bayes' theorem, the joint posterior density of parameters and latent variables can be written as

$$\begin{aligned} p(\boldsymbol{\theta}, \mathbf{U}_{1:T}, \boldsymbol{\beta}_{0:T} \mid \mathbf{y}_{1:T}) &\propto \prod_{t=1}^{T} p(\mathbf{y}_t \mid \boldsymbol{\theta}, \boldsymbol{\beta}_t, U_t) \\ &\times \prod_{t=1}^{T} p(\boldsymbol{\beta}_t \mid \boldsymbol{\theta}, \boldsymbol{\beta}_{t-1}) p(\boldsymbol{\beta}_0) \\ &\times \prod_{t=1}^{T} p(U_t \mid \boldsymbol{\theta}) p(\boldsymbol{\theta}) \end{aligned} \tag{11}$$

where

$$\begin{aligned} \mathbf{y}_t \mid \boldsymbol{\theta}, \boldsymbol{\beta}_t, U_t &\sim \mathcal{N}_r(\boldsymbol{\Lambda}\boldsymbol{\beta}_t, U_t^{-1}\text{diag}(\sigma_1^2, \ldots, \sigma_r^2)) \\ \boldsymbol{\beta}_t \mid \boldsymbol{\theta}, \boldsymbol{\beta}_{t-1} &\sim \mathcal{N}_3((\mathbf{I}_3 - \mathbf{A})\boldsymbol{\mu} + \mathbf{A}\boldsymbol{\beta}_{t-1}, \mathbf{W}) \\ \beta_0 \mid \boldsymbol{\theta} &\sim \mathcal{N}_3(\boldsymbol{\mu}, \Xi) \\ U_t &\sim p(U_t \mid \boldsymbol{\nu}), \end{aligned}$$

where Ξ is the unconditional variance of $\boldsymbol{\beta}_t$ and $\text{vec}(\Xi) = [\mathbf{I}_9 - (\mathbf{A} \otimes \mathbf{A})]^{-1}\text{vec}(\mathbf{W})$. $\mathcal{N}_k(.,.)$, $\mathbf{I}_k$ and vec(.) indicate the k-variate normal distribution, the $k \times k$ identity matrix of and the vector column operator respectively. We assume the prior distribution as

$$p(\boldsymbol{\theta}) = p(\lambda)p(\boldsymbol{\mu})p(\mathbf{A}, \mathbf{W}) \prod_{j=1}^{r} p(\sigma_j^2) p(\boldsymbol{\nu}).$$

The components of the prior distribution are specified as

$$\begin{aligned} \mathbf{A} \mid W &\sim \mathcal{N}_{3\times 3}(\bar{\mathbf{A}}, \mathbf{C}, \mathbf{W})\mathbb{I}_{\Omega}(\mathbf{A}), \\ \mathbf{W} &\sim \mathcal{IW}_3(\upsilon, \mathbf{S}), \\ \boldsymbol{\mu} &\sim \mathcal{N}_3(\boldsymbol{\mu}_0, S_{\boldsymbol{\mu}}) \\ \sigma_j^2 &\sim \mathcal{IG}(\frac{a_j}{2}, \frac{b_j}{2}), \qquad j = 1, \ldots, r \\ \lambda &\sim \mathcal{G}(\alpha, \delta), \\ \boldsymbol{\nu} &\sim p(\boldsymbol{\nu}), \end{aligned} \tag{12}$$

where $\mathcal{N}_{k\times s}(.,.,)$ and $\mathcal{IW}_k(.,.)$ represents the $k \times s$ matrix variate normal and the k inverse Wishart distributions, respectively. $\mathbb{I}_{\Omega}(.)$ is an indicator function and $\Omega = \{\mathbf{A} \in R^{3\times 3}$ such that max| eigenvalue of $\mathbf{A}| < 1\}$ is the stationarity condition of the VAR specification (8). $p(\boldsymbol{\nu})$ is specified for each member of the SMN distributions.

As the posterior distribution in (11) is intractable analytically, we sample the parameters as well the mixing variables and latent factors from their full conditional distributions using the Gibbs sampler. The sampling scheme is described by the following algorithm:

Algorithm 3.1

1. Set $i = 0$ and initialize $\boldsymbol{\theta}^{(i)}$, $\mathbf{U}_{1:T}^{(i)}$ and $\boldsymbol{\beta}_{0:T}^{(i)}$

2. Drawn $\boldsymbol{\theta}^{(i+1)} \sim p(\boldsymbol{\theta} \mid \mathbf{U}_{1:T}^{(i)}, \boldsymbol{\beta}_{0:T}^{(i)}, \mathbf{y}_{1:T})$

3. Drawn $\mathbf{U}_{1:T}^{(i+1)} \sim p(\mathbf{U}_{1:T} \mid \boldsymbol{\theta}^{(i+1)}, \boldsymbol{\beta}_{0:T}^{(i)}, \mathbf{y}_{1:T})$

4. Drawn $\boldsymbol{\beta}_{0:T}^{(i+1)} \sim p(\boldsymbol{\beta}_{0:T} \mid \boldsymbol{\theta}^{(i+1)}, \mathbf{U}_{1:T}^{(i+1)}, \mathbf{y}_{1:T})$

5. $i = i + 1$ and return to 2.

Cycling through 2 to 4 is a complete sweep of this sampler. The MCMC sampler will require us to perform many thousands of sweeps to generate samples from the posterior distribution $p(\boldsymbol{\theta}, \mathbf{U}_{1:T}, \boldsymbol{\beta}_{0:T} \mid \mathbf{y}_{1:T})$. Details on the full conditionals of $\boldsymbol{\theta}$, the mixing variables $\mathbf{U}_{1:T}$ and the latent factors $\boldsymbol{\beta}_{0:T}$ are given in the Appendix.

3.2 Bayesian model selection

To assess the goodness-of-fit for the proposed class of models and to compare among several candidate models, we adopt Spiegelhalter et al. (2002) deviance information criterion (DIC) for model selection within the MCMC framework. It serves a measure that reasonably balances between model fit and model complexity. The DIC is defined as:

$$\mathrm{DIC} \quad = \quad -2E_{\boldsymbol{\theta}|\mathbf{y}_{1:T}}[\log L(\mathbf{y}_{1:T} \mid \boldsymbol{\theta})] + p_D. \tag{13}$$

The second term in (13) measures the complexity of the model by the effective number of parameters, p_D, defined as the difference between the posterior mean of the deviance and the deviance evaluated at the posterior mean of the parameters:

$$p_D \quad = \quad 2[\log L(\mathbf{y}_{1:T} \mid \bar{\boldsymbol{\theta}}) - E_{\boldsymbol{\theta}|\mathbf{y}_{1:T}}[\log L(\mathbf{y}_{1:T} \mid \boldsymbol{\theta})]]. \tag{14}$$

The best model is chosen as the one that has the minimum DIC. To calculate the DIC, in the context of DL-SMN models, the conditional likelihood $L(\mathbf{y}_{1:T} \mid \lambda, \boldsymbol{\mu}, \mathbf{A}, \mathbf{W}, \{\sigma_j^2\}_{j=1}^r, \boldsymbol{\nu},\ \mathbf{U}_{1:T}, \boldsymbol{\beta}_{0:T})$ is used in equation (13), where $\boldsymbol{\theta}$ encompasses $(\lambda, \boldsymbol{\mu}', \mathrm{vec}(\mathbf{A})', \mathrm{vech}(\mathbf{W})', \{\sigma_j^2\}_{j=1}^r, \boldsymbol{\nu}')'$, $\mathbf{U}_{1:T}$ and $\boldsymbol{\beta}_{0:T}$.

3.3 Out-of-sample forecasting

We have that $K-$step ahead prediction density can be calculated using the composition method through the following recursive procedure:

$$\begin{aligned}
p(\mathbf{y}_{T+K} \mid \mathbf{y}_{1:T}) \quad &= \quad \int \Big[p(y_{T+K} \mid U_{T+K}, \boldsymbol{\beta}_{T+K}, \boldsymbol{\theta}) p(\boldsymbol{\beta}_{T+K} \mid \boldsymbol{\theta}, \mathbf{y}_{1:T}) p(U_{T+K} \mid \boldsymbol{\theta}) p(\boldsymbol{\theta} \mid \mathbf{y}_{1:T}) \Big] d\boldsymbol{\beta}_{T+K} dU_{T+K} d\boldsymbol{\theta}, \\
p(\boldsymbol{\beta}_{T+K} \mid \boldsymbol{\theta}, \mathbf{y}_{1:T}) \quad &= \quad \int \Big[p(\boldsymbol{\beta}_{T+K} \mid \boldsymbol{\theta}, \boldsymbol{\beta}_{T+K-1}) p(\boldsymbol{\beta}_{T+K-1} \mid \boldsymbol{\theta}, \mathbf{y}_{1:T}) \Big] d\boldsymbol{\beta}_{T+K-1}.
\end{aligned}$$

Evaluation of these integrals is straightforward, by using Monte Carlo approximation. To initialize a recursion, we use N draws $\{\boldsymbol{\beta}_T^{(i)}, U_T^{(i)}, \boldsymbol{\theta}^{(i)}\}_{i=1}^N$ from the MCMC sample. Then given these N draws, sample N draws

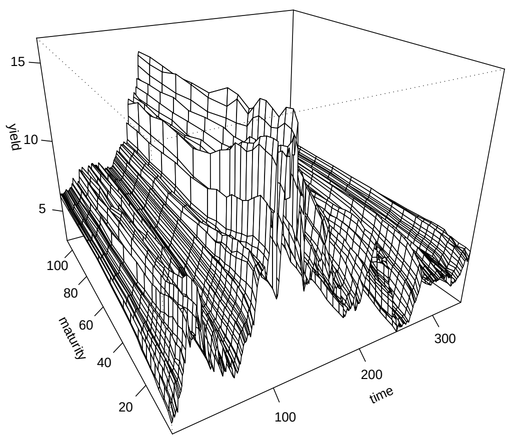

Figure 1: U.S. yield curves. The sample consists of monthly yield data from January 1972 to July 2000 at maturities 3, 6, 9, 12, 15, 18, 21, 24, 30, 36, 48, 60, 72, 84, 96, 108, and 120 months.

$\boldsymbol{\beta}^{(i)}_{T+k}$ from $p(\boldsymbol{\beta}_{T+k} \mid \boldsymbol{\theta}^{(i)}, \boldsymbol{\beta}^{(i)}_{T+k-1})$ and $U^{(i)}_{T+k}$ from $p(U_{T+k} \mid \boldsymbol{\theta}^{(i)})$ for $i = 1, \ldots, N$ and $k = 1, \ldots, K$, by using equations (8) and (9), respectively. Finally, sample N draws $\{\mathbf{y}^{(i)}_{T+k}\}_{i=1}^{N}$ from $p(\mathbf{y}_{T+k} \mid U_{T+k}, \boldsymbol{\beta}^{(i)}_{T+k}, \boldsymbol{\theta}^{(i)})$. With draws from $\boldsymbol{\beta}_{T+k}$ and $\mathbf{y}_{T+k}$ some summary statistics can be calculated.

4 Empirical application

In this section we present an empirical application of the method described in section 3. Our basic data set consists of a set of zero-coupon equivalent US yields with maturities: 3, 6, 9, 12, 15, 24, 36, 60, 84 and 120 months. The yields are derived from bid/ask average quotes, from January 1972 through December 2000[2], using the unsmoothed Fama and Bliss (1987) approach.[3]

We divide our data set in an initial period which covers the sample from 1972:1-200:07 (343 observations). The rest is used for out-of-sample forecast purposes. In Table 1, we present descriptive statistics for the yields at representative maturities in the estimation period. For this maturities we show the mean, standard deviation, minimum, maximum, kurtosis and some autocorrelation coefficients. In addition we show these statistics for proxies of the level, slope and curvature coefficients. These proxies are in accordance with the construction of Diebold and Li (2006). All yields are highly persistent, with average first-order autocorrelation greater than 0.97. The kurtosis in all the maturities is greater than 3, specially in earlier maturities. This fact could give us some empirical evidence of the presence of heavy-tailed distributions. In Figure 1, we provide a three-dimensional plot of our yield curve data. The large amount of temporal variation in the level is visually apparent, because the sample period covers a wide range of interest rate regimes, from very low levels to the high rate regime. It

[2]We thank F. X. Diebold for making the data set available on his website: http://www.ssc.upenn.edu/~fdiebold/YieldCurve.html

[3]A more detailed description can be found in Diebold et al. (2006b)

is clear that the typical yield curve is upward sloping, that the long rates are less volatile and more persistent than short rates, that the level (120-month yield) is highly persistent but varies only moderately relative to its mean, that the slope is less persistent than any individual yield but quite highly variable relative to its mean, and the curvature is the least persistent of all factors and the most highly variable relative to its mean.

Table 1: Descriptive statistics, US yields curve data set.

Maturity	Mean	Std. Dev.	Minimum	Maximum	kurtosis	$\hat{\rho}(1)$	$\hat{\rho}(12)$
3	6.861	2.717	2.732	16.020	4.595	0.970	0.698
12	7.321	2.619	3.107	15.822	3.852	0.971	0.727
60	7.967	2.283	4.347	15.005	3.371	0.981	0.778
120 (level)	8.181	2.158	3.402	14.925	3.592	0.983	0.770
Slope	1.321	1.454	-3.505	4.060	3.180	0.929	0.416
Curvature	-0.033	0.825	-2.427	3.501	4.779	0.824	0.398

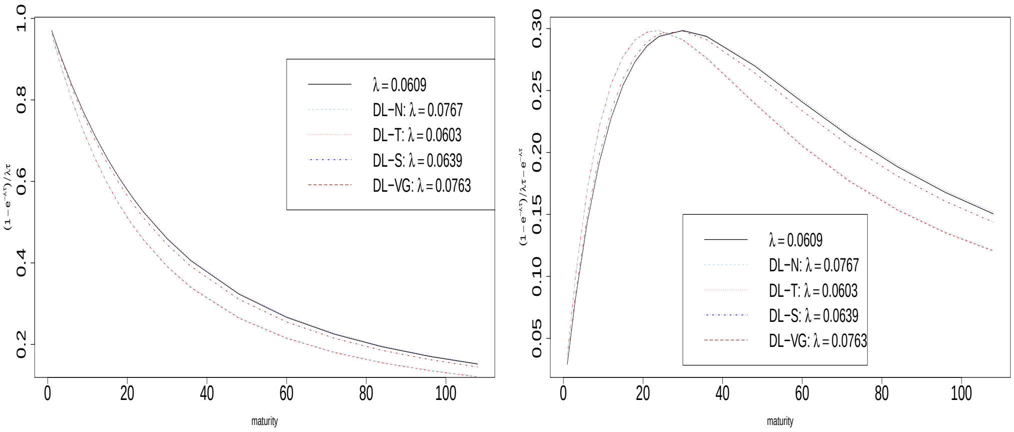

Figure 2: US yield data set. Left : Factor loading Slope $(\frac{1-e^{-\lambda\tau}}{\lambda\tau})$ vs maturity. Bottom: Factor loading Curvature $(\frac{1-e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau})$ vs maturity

We fitted the DL-N, DL-T, DL-S and DL-VG models. We set the prior distributions as in equation (12), with $\boldsymbol{\mu}_0 = (7.43, -1.37, 0.54)'$ and $S_\mu = 10\mathbf{I}_3$, $a_j = 5$, $b_j = 0.2$, $\bar{\mathbf{A}} = 0.9\mathbf{I}_3$, $\mathbf{C} = 100\mathbf{I}_3$, $\upsilon = 5$, $\mathbf{S} = 10\mathbf{I}_3$. We assume that $\nu \sim \mathcal{G}(12, 0.8)$ (DL-t model), $\nu \sim \mathcal{G}(0.2, 0.05)$ (DL-S model) and $\nu \sim \mathcal{G}(12, 0.8)$ (DL-VG model) and $\log(\lambda) \sim \mathcal{N}(-2.8, 10)$.

We start the MCMC simulation and choose the initial values of the chain as follows: initial values for parameters and the mixing variables $\mathbf{U}_{1:T}$ are randomly generated from the priors distributions and the latent factors $\boldsymbol{\beta}_{0:T}$ are initialized using the OLS estimates. Next, we run a chain by 35000 iterations. The first 5000 are discarded as a burn-in period and then, the next were recorded. All the calculations were performed running

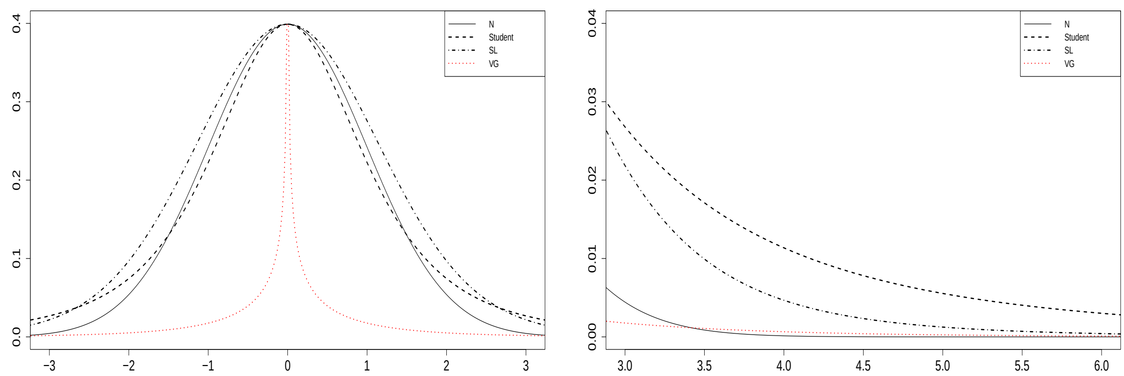

Figure 3: Density curves of the univariate normal, student-t, slash and variance gamma using the estimated tail-fatness parameter from the respective DL model.

stand alone code developed by the authors using an open source C++ library for statistical computation, the Scythe statistical library (Pemstein et al., 2007), which is available for free download at http://scythe.wustl.edu.

Tables 2 and 3 present the posterior mean and the 95% posterior credibility intervals for parameters in the VAR representation in equation (8), for the DL-N, DL-t, DL-S and DL-VG models. The high persistence from the proxies for the level, slope and curvature that we report in Tables 1 are confirmed by the high diagonal elements of the VAR coefficient matrix **A**. The heavy-tailed versions of the DL model show are slightly more persistent than the normal one. From Table 3, we found that the curvature and slope factors and more volatile than the level factor. All the results are coherent with previous result of Diebold et al. (2006b).

Table 4 shows the posterior mean and the 95% posterior credibility interval for lambda for the four models. The DL-N has a posterior mean of 0.0767 and the credibility interval (0.0728,0.0813), which is quite similar to the results of Diebold et al. (2006b). This result shows that λ is high significant. The posterior mean and credibility intervals for the heavy-tailed distributions are slightly different. The posterior means for the DL-T, DL-S and DL-VG models are 0.0603, 0.0639 and 0.0763. This result confirm that interest rates can be informative about λ and that small changes in the loadings can have a significant effect on the estimates. This affirmation is confirmed by Figure 2, where the left panel shows the factor loading of the slope and right panel the factor loading of curvature evaluated at each one of the posterior means obtained for the DL-N, DL-t, DL-S and DL-VG models. There are significative differences between the graph of such functions.

The magnitude of the tail fatness is measured by the shape parameter ν in the DL-t, DL-S and DL-VG models. The posterior mean of ν in the DL-t model is 2.6439. In the DL-S model, the posterior mean of ν is 1.7253, and in the DL-VG model the posterior mean of ν is 0.6451. These results seem to indicate that the measurement errors are better explained by heavy-tailed distributions.

Table 2: US data yield data set. Summary results for the matrix $\mathbf{A}$ and the vector $\boldsymbol{\mu}$ for the DL class of models. We report the posterior mean and in parenthesis the 95% posterior credibility interval obtained from the MCMC output.

		DL-N		
	Level$_{t-1}$ (β_{t-1})	Slope$_{t-1}$ ($\beta_{1,t-1}$)	Curvature$_{t-1}$ ($\beta_{3,t-1}$)	Constant ($\boldsymbol{\mu}$)
Level$_t$ ($\beta_{1,t}$)	0.9914	0.0270	-0.0160	7.9517
	(0.9737,1.0075)	(0.0070,0.0468)	(-0.0400,0.0081)	(4.9137,10.5588)
Slope$_t$ ($\beta_{2,t}$)	-0.0246	0.9412	0.0350	-1.4816
	(-0.0571,0.0058)	(0.9059,0.9765)	(-0.0089,0.0779)	(-2.9371,-0.0112)
Curvature$_t$ ($\beta_{3,t}$)	0.0244	0.0257	0.8453	-0.3747
	(-0.0226,0.0691)	(-0.0251,0.0778)	(0.7833,0.9064)	(-1.4214,0.5891)
		DL-t		
	Level$_{t-1}$ (β_{t-1})	Slope$_{t-1}$ ($\beta_{1,t-1}$)	Curvature$_{t-1}$ ($\beta_{3,t-1}$)	Constant ($\boldsymbol{\mu}$)
Level$_t$ ($\beta_{1,t}$)	0.9916	0.0219	-0.0171	7.9610
	(0.9748,1.0072)	(0.0036,0.0408)	(-0.0419,0.0081)	(4.7760,10.6325)
Slope$_t$ ($\beta_{2,t}$)	-0.0175	0.9564	0.0004	-1.4880
	(-0.0485,0.0118)	(0.9241,0.9864)	(-0.0432,0.0448)	(-3.1289,0.1922)
Curvature$_t$ ($\beta_{3,t}$)	0.0177	0.0002	0.8718	-0.1236
	(-0.0209,0.0582)	(-0.0427,0.0438)	(0.8134,0.9275)	(-1.1067,0.7872)
		DL-S		
	Level$_{t-1}$ (β_{t-1})	Slope$_{t-1}$ ($\beta_{1,t-1}$)	Curvature$_{t-1}$ ($\beta_{3,t-1}$)	Constant ($\boldsymbol{\mu}$)
Level$_t$ ($\beta_{1,t}$)	0.9919	0.0234	-0.0178	7.9604
	(0.9752,1.0079)	(0.0046,0.0426)	(-0.0422,0.0064)	(4.9194,10.6148)
Slope$_t$ ($\beta_{2,t}$)	-0.0185	0.9545	0.0062	-1.4943
	(-0.0496,0.0108)	(0.9216, 0.9858)	(-0.0375,0.0511)	(-3.057,0.1390)
Curvature$_t$ ($\beta_{3,t}$)	0.0185	0.0047	0.8678	-0.1729
	(-0.0215,0.0599)	(-0.0397,0.0491)	(0.8101,0.9246)	(-1.1540,0.7531)
		DL-VG		
	Level$_{t-1}$ (β_{t-1})	Slope$_{t-1}$ ($\beta_{1,t-1}$)	Curvature$_{t-1}$ ($\beta_{3,t-1}$)	Constant ($\boldsymbol{\mu}$)
Level$_t$ ($\beta_{1,t}$)	0.9921	0.0273	-0.0181	7.9394
	(0.9747,1.0079)	(0.0078,0.0478)	(-0.0419,0.0057)	(4.8925,10.5584)
Slope$_t$ ($\beta_{2,t}$)	-0.0246	0.9421	0.0344	1.4727
	(-0.0571,0.0055)	(0.9064, 0.9761)	(-0.0086,0.0779)	(-2.8974, 0.0218)
Curvature$_t$ ($\beta_{3,t}$)	0.0225	0.0244	0.8467	-0.3788
	(-0.0231,0.0689)	(-0.0254,0.0760)	(0.7828,0.9086)	(-1.4607,0.6166)

To illustrate the tail behavior, we plot the normal ($\mathcal{N}(0,1)$) density, student's-t ($\mathcal{T}(0,1,\nu)$) density with ν degrees of freedom, the slash ($\mathcal{S}(0,1,\nu)$) density with shape parameter ν and the variance gamma ($\mathcal{VG}(0,1,\nu)$) density with shape parameter ν. We set ν as the posterior mean of the respective DL model (see Table 4). Figure 3 depicts the four density curves (the student-t, slash and variance gamma have been rescaled to be comparable, see Wang and Genton, 2006). The density of the variance gamma emphasizes on the sharpness around the mean rather than the tails fatness, so that the Student-t and slash distributions have fatter tails than the standard normal and variance gamma distributions.

To assess the goodness of the estimated models, we calculate the deviance information criterion, DIC. From Table 5, according to the DIC criterion, the DL-t model best fits the US yields data set. The DL-VG shows the worst fit.

Table 3: US data yield data set. Summary results for the covariance matrix $\mathbf{W}$ of the latent factors for the DL class of models. We report the posterior mean and in parenthesis the 95% posterior credibility interval obtained from the MCMC output.

		DL-N	
	Level$_t$ (β_t)	Slope$_t$ ($\beta_{1,t}$)	Curvature$_{t-1}$ ($\beta_{3,t}$)
Level$_t$ ($\beta_{1,t}$)	0.1342	-0.0150	0.0216
	(0.1140,0.1577)	(-0.0432,0.0118)	(-0.0227,0.0636)
Slope$_t$ ($\beta_{2,t}$)		0.4171	0.0158
		(0.3563,0.4879)	(-0.0559,0.0878)
Curvature$_t$ ($\beta_{3,t}$)			0.8353
			(0.6817,1.0090)
		DL-t	
	Level$_t$ (β_t)	Slope$_t$ ($\beta_{1,t}$)	Curvature$_{t-1}$ ($\beta_{3,t}$)
Level$_t$ ($\beta_{1,t}$)	0.1259	-0.0186	0.0205
	(0.1062,0.1484)	(-0.0453,.0074)	(-0.0170,0.0571)
Slope$_t$ ($\beta_{2,t}$)		0.3832	0.0513
		(0.3246,0.4519)	(-0.0135,0.1182)
Curvature$_t$ ($\beta_{3,t}$)			0.6328
			(0.5138,0.7739)
		DL-S	
	Level$_t$ (β_t)	Slope$_t$ ($\beta_{1,t}$)	Curvature$_{t-1}$ ($\beta_{3,t}$)
Level$_t$ ($\beta_{1,t}$)	0.1271	-0.0187	0.0247
	(0.1077,0.1499)	(-0.0458,0.0072)	(-0.0150,0.0621)
Slope$_t$ ($\beta_{2,t}$)		0.3873	0.0501
		(0.3276,0.4569)	(-0.0164,0.1170)
Curvature$_t$ ($\beta_{3,t}$)			0.6614
			(0.5352,0.8101)
		DL-VG	
	Level$_t$ (β_t)	Slope$_t$ ($\beta_{1,t}$)	Curvature$_{t-1}$ ($\beta_{3,t}$)
Level$_t$ ($\beta_{1,t}$)	0.1319	-0.0149	0.0264
	(0.1115,0.1553)	(-0.0419,0.0115)	(-0.0163,0.0680)
Slope$_t$ ($\beta_{2,t}$)		0.4166	0.0169
		(0.3555,0.4865)	(-0.0543,0.0885)
Curvature$_t$ ($\beta_{3,t}$)			0.8210
			(0.6664,0.9981)

In Figure 4 the posterior smoothed means of the latent factors obtained from the DL-N, DL-t, DL-S and DL-VG are compared with their data-base proxies. Each of the factors agrees with their data-based proxy: the level factor is close to the 120-month yield, the slope is close to spread of the 3-month over the 120 month yields and the curvature is close to the twice times the 24-month minus the 3 and 120 month yield. There are some differences between the DL-SMN in relation to DL-N. Finally, Figure 5 shows the posterior mean of $\mathbf{\Lambda}\boldsymbol{\beta}_t$ for March 1989, July 1989, May 1997 and August 1999, for each one of models fitted. The results are quite similar for all models.

Figures 6 and 7 show the posterior mean and 95% credibility intervals for the k-step forecast of 3-month and 12-month yields respectively for the next $k = 1, \ldots, 5$ months, that is for 2000:9 to 2000:12. All the true values are near to the posterior means for each one of the DL class of models fitted. Which give us that the

Table 4: US data yield data set. Summary results for the factor loading parameter λ and the heavy-tail parameter ν. We report the posterior mean and 95% posterior credibility interval obtained from the MCMC output.

λ			
DL-N	DL-t	DL-S	DL-VG
0.0767	0.0603	0.0639	0.0763
(0.0728,0.0813)	(0.0565,0.0635)	(0.0604,0.0674)	(0.0721,0.0808)
ν			
DL-N	DL-t	DL-S	DL-VG
	2.6439	1.7253	0.6451
	(2.1707,3.1865)	(1.5047,1.9684)	(0.5669,0.7289)

Table 5: DIC: deviance information criterion

	DIC	p_D
Model	Value	Ranking
DL-N	-9964.9	3
DL-t	-11771.2	1
DL-S	-11720.6	2
DL-VG	-9211.3	4

models give good out-of-sample forecasts.

5 Conclusions

We have used and applied the term-structure model introduced by Nelson and Siegel (1987) and reinterpreted by Diebold and Li (2006) as a modern three-factor model of level, slope and curvature. Using a state space model framework, the Gaussian assumption of the measurement innovation was replaced by thick-tailed processes, known as scale mixtures of normal distributions. We study three specific sub-classes, viz. the Student-t, the slash and the variance gamma distributions and compare parameter estimates and model fit with the default normal model. We consider that $\lambda_t = \lambda$ is time-invariant as in Diebold and Li (2006) but we estimate it from the model. Under a Bayesian perspective, we constructed an algorithm based on Markov Chain Monte Carlo (MCMC) simulation methods to estimate all the parameters and latent quantities in our proposed DL-SMN class of models. As a by product of the MCMC algorithm, we were able to produce an estimate of the latent information process which can be used in financial modeling. The use of mixing variable, $\mathbf{U}_{1:T}$ for normal scale mixture distributions not only simplifies the full conditional distributions required for the Gibbs sampling algorithm, but also provides a mean for outlier diagnostics.

Some contributions have been introduced in this article, but some other extensions are still possible. We can include a dynamic for λ_t and to model the error term in the observational equation as a skewed distributions and include macro-economical variables as a latent factors. We are currently exploring these extensions.

Appendix: The Full conditionals

In this appendix, we describe the full conditional distributions for the parameters, the mixing variables $\mathbf{U}_{1:T}$ and the latent factors $\boldsymbol{\beta}_{0:T}$ of the DL-SMN class of models.

Full conditional of σ_j^2

Under the specified inverse-gamma prior, the full conditional of σ_j^2 is given by

$$p(\sigma_j^2 \mid \mathbf{U}_{1:T}, \mathbf{y}_{1:T}, \boldsymbol{\beta}_{0:T}) \quad \propto \quad \left(\frac{1}{\sigma_j^2}\right)^{\frac{a_j+T}{2}+1} \mathrm{e}^{\{-\frac{1}{2\sigma_j^2}(\sum_{t=1}^{T} U_t[(y_t(\tau_j)-\Lambda_j\beta_t)^2]+b_j)\}}$$

that is,

$$\sigma_j^2 \mid \mathbf{U}_{1:T}, \mathbf{y}_{1:T}, \boldsymbol{\beta}_{0:T} \quad \sim \quad \mathcal{IG}(\frac{a_j^*}{2}, \frac{b_j^*}{2}) \tag{15}$$

where $a_j^* = a_j + T$, $b_j^* = b_j + \sum_{t=1}^{T} U_t(y_t(\tau_j) - \Lambda_j\beta_t)^2$ and $\Lambda_j = [1, \frac{1-e^{-\lambda\tau_j}}{\lambda\tau_j}, \frac{1-e^{-\lambda\tau_j}}{\lambda\tau_j} - e^{-\lambda\tau_j}]$

Full conditional of $\boldsymbol{\mu}$

Assuming that $\boldsymbol{\mu} \sim \mathcal{N}_3(\boldsymbol{\mu}_0, S_{\boldsymbol{\mu}})$ is the prior distribution. Then, the full conditional of $\boldsymbol{\mu}$ is given by

$$\begin{aligned} p(\boldsymbol{\mu} \mid \boldsymbol{\beta}_{0:T}, \mathbf{W}, \mathbf{A}) \quad &\propto \quad \mathrm{e}^{-\frac{1}{2}(\boldsymbol{\mu}'[\mathbf{S}_{\boldsymbol{\mu}}^{-1}+T(I-\mathbf{A})'\mathbf{W}^{-1}(I-\mathbf{A})+\Xi^{-1}]\boldsymbol{\mu}} \\ &\times \quad \mathrm{e}^{-\frac{1}{2}[-2\boldsymbol{\mu}_0'\mathbf{S}_{\boldsymbol{\mu}}^{-1}-2\sum_{t=1}^{T}(\boldsymbol{\beta}_t-\mathbf{A}\boldsymbol{\beta}_{t-1})'\mathbf{W}^{-1}(I-\mathbf{A})-2\boldsymbol{\beta}_0'\Xi^{-1}]\boldsymbol{\mu}}, \end{aligned}$$

which gives

$$\boldsymbol{\mu} \mid \boldsymbol{\beta}_{0:T}, \mathbf{W}, \mathbf{A} \quad \sim \quad \mathcal{N}(\mathbf{b}_{\boldsymbol{\mu}}, \Sigma_{\boldsymbol{\mu}}), \tag{16}$$

where

$$\boldsymbol{\Sigma}_{\boldsymbol{\mu}} = [\mathbf{S}_{\boldsymbol{\mu}}^{-1} + T(\mathrm{I}_3 - \mathbf{A})'\mathbf{W}^{-1}(\mathbf{I}_3 - \mathbf{A}) + \Xi^{-1}]^{-1}$$

and

$$\mathbf{b}_{\boldsymbol{\mu}} = \boldsymbol{\Sigma}_{\boldsymbol{\mu}}[\mathbf{S}_{\boldsymbol{\mu}}^{-1}\boldsymbol{\mu}_0 + \sum_{t=1}^{T}(\mathbf{I}_3 - \mathbf{A})'\mathbf{W}^{-1}(\boldsymbol{\beta}_t - \mathbf{A}\boldsymbol{\beta}_{t-1}) + \Xi^{-1}\boldsymbol{\beta}_0].$$

Full conditionals of A and W

Sampling the $\mathbf{A}$ and $\mathbf{W}$ one at a time is straightforward. We use the method of Chibb and Greenberg (1994), which is based on the Metropolis-Hastings algorithm. Under the Matrix Normal-Inverse Wishart prior, the full conditional distribution of $\mathbf{A}$ is proportional to

$$p(\mathbf{A} \mid \boldsymbol{\mu}, \mathbf{W}, \boldsymbol{\beta}_{0:T}) \quad \propto \quad \mathrm{e}^{-\frac{1}{2}\mathrm{tr}[(\mathbf{S}_1+\mathbf{S}_2)\mathbf{W}^{-1}]} g(\mathbf{A}, \mathbf{W})\mathbb{I}_\Omega(\mathbf{A}), \tag{17}$$

where

$$\begin{aligned} \mathbf{S}_1 \quad &= \quad (\mathbf{A} - \bar{\mathbf{A}})\mathbf{C}^{-1}(\mathbf{A} - \bar{\mathbf{A}})', \\ \mathbf{S}_2 \quad &= \quad \sum_{t=1}^{T}[\boldsymbol{\beta}_t - \boldsymbol{\mu} - \mathbf{A}(\boldsymbol{\beta}_{t-1} - \boldsymbol{\mu})][\boldsymbol{\beta}_t - \boldsymbol{\mu} - \mathbf{A}(\boldsymbol{\beta}_{t-1} - \boldsymbol{\mu})]' \end{aligned}$$

and

$$g(\mathbf{A},\mathbf{W}) \quad = \quad |\, \Xi \,|^{-1/2}\, \mathrm{e}^{-\frac{1}{2}(\boldsymbol{\beta}_0-\boldsymbol{\mu})'\Xi^{-1}(\boldsymbol{\beta}_0-\boldsymbol{\mu})}.$$

As $p(\mathbf{A} \mid \boldsymbol{\mu}, \mathbf{W}, \boldsymbol{\beta}_{0:T})$ in (17) does not have closed form, we sample from using the Metropolis-Hastings algorithm with proposal density given by

$$\mathrm{vec}(\mathbf{A}') \mid \boldsymbol{\mu}, \mathbf{W}, \boldsymbol{\beta}_{0:T} \quad \sim \quad \mathcal{N}(\mathrm{vec}(\tilde{\mathbf{A}}'), \mathbf{W} \otimes \tilde{C})\mathbb{I}_{\Omega}(\mathbf{A}), \tag{18}$$

where

$$\tilde{\mathbf{A}} = [\bar{\mathbf{A}}\mathbf{C}^{-1} + \sum_{t=1}^{T}(\boldsymbol{\beta}_t - \boldsymbol{\mu})(\boldsymbol{\beta}_{t-1} - \boldsymbol{\mu})']\tilde{\mathbf{C}}$$

and

$$\tilde{\mathbf{C}} = [\mathbf{C}^{-1} + \sum_{t=1}^{T}(\boldsymbol{\beta}_{t-1} - \boldsymbol{\mu})(\boldsymbol{\beta}_{t-1} - \boldsymbol{\mu})']^{-1}.$$

Given the current value $\mathbf{A}^{(i-1)}$ at the $(i-1)-th$ iteration, sample a proposed value $\mathbf{A}^{(*)}$ from (18), accept this proposal value as $\mathbf{A}^{(i)}$ with probability $\frac{g(\mathbf{A}^{(*)},\mathbf{W})}{g(\mathbf{A}^{(i-1)},\mathbf{W})}$. If the proposal value is rejected, set $\mathbf{A}^{(i)} = \mathbf{A}^{(i-1)}$.

For $\mathbf{W}$, we sample from the full conditional distribution in a similar fashion. Under the specified prior, the full conditional distribution of $\mathbf{W}$ is proportional to

$$p(\mathbf{W} \mid \boldsymbol{\mu}, \boldsymbol{\beta}_{0:T}, \mathbf{A}) \quad \propto \quad |\, \mathbf{W} \,|^{-\frac{T+\nu+7}{2}}\, \mathrm{e}^{-\frac{1}{2}\mathrm{tr}[(\mathbf{S}_1+\mathbf{S}_2+\mathbf{S})\mathbf{W}^{-1}]} g(\mathbf{A},\mathbf{W}),$$

which does not have closed form. The Metropolis-Hastings algorithm is used to sample from with proposal density given by $\mathcal{IW}_3(T + v + 3, \mathbf{S}_1 + \mathbf{S}_2 + \mathbf{S})$. Given the current value $\mathbf{W}^{(i-1)}$ at the $(i-1)-th$ iteration and the proposed value $\mathbf{W}^{(*)}$, we accept $\mathbf{W}^{(*)}$ with probability $\frac{g(\mathbf{A},\mathbf{W}^{(*)})}{g(\mathbf{A},\mathbf{W}^{(i-1)})}$. If the proposal value is rejected, set $\mathbf{W}^{(i)} = \mathbf{W}^{(i-1)}$.

Full conditional of λ

Under the specified gamma prior, $\mathcal{G}(\alpha, \delta)$, the full conditional of λ is given by

$$p(\lambda \mid \mathbf{U}_{1:T}, \sigma_1^2, \ldots, \sigma_r^2, \mathbf{y}_{1:T}) \quad \propto \quad \lambda^{a-1} e^{-b\lambda} \mathrm{e}^{-\frac{1}{2}\sum_{t=1}^{T}\sum_{j=1}^{r}\frac{U_t[y_t(\tau_j)-\Lambda_j\boldsymbol{\beta}_t]^2}{\sigma_j^2}}.$$

This distribution does not have closed form, then we use the Metropolis-Hasting algorithm to sample it, using the proposal based on Geweke and Tanizaki (2001). Let $q(.)$ the log of the full conditional of λ. Then, the proposal density is $\mathcal{N}(\mu_\lambda, \sigma_\lambda^2)$, where $\mu_\lambda = x - \frac{q'(x)}{q''(x)}$ and $\sigma_\lambda^2 = [-q''(x)]^{-1}$, x is the value of the previous iteration, $q'(.)$ and $q''(.)$ are the first and second derivatives respectively.

Full conditional of U_t and ν

- **SV-t case**

As $U_t \sim \mathcal{G}(\frac{\nu}{2}, \frac{\nu}{2})$, the full conditional of U_t is given by

$$p(U_t \mid \mathbf{y}_t, \boldsymbol{\beta}_t, \nu) \quad \propto U_t^{\frac{\nu+r}{2}-1} \mathrm{e}^{-\frac{U_t}{2}[\nu+\sum_{j=1}^{r}\frac{(y_t(\tau_j)-\Lambda_j\beta_t)^2}{\sigma_j^2}]}, \tag{19}$$

which is $\mathcal{G}(\frac{\nu+r}{2}, \frac{1}{2}[\nu + \sum_{j=1}^{r} \frac{(y_t(\tau_j)-\Lambda_j\beta_t)^2}{\sigma_j^2}])$.

We assume the prior distribution of ν as $\mathcal{G}(a_\nu, b_\nu)\mathbb{I}_{2<\nu\leq 40}$. Then, the full conditional of ν is

$$p(\nu \mid \mathbf{U}_{1:T}) \quad \propto \quad \frac{\left[\frac{\nu}{2}\right]^{\frac{T\nu}{2}} \nu^{a_\nu - 1} \mathrm{e}^{-\frac{\nu}{2}[\sum_{t=1}^{T}(U_t - \log U_t)+2b_\nu]}}{[\Gamma(\frac{\nu}{2})]^T} \mathbb{I}_{2<\nu\leq 40}. \tag{20}$$

We sample ν by the Metropolis-Hastings acceptance-rejection algorithm (Tierney, 1994; Chib, 1995). Let ν^* denote the mode (or approximate mode) of $p(\nu \mid \mathbf{U}_{1:T})$, and let $\ell(\nu) = \log p(\nu \mid \mathbf{U}_{1:T})$. As $\ell(\nu)$ is concave, we use the proposal density $\mathcal{N}_{(2,40)}(\mu_\nu, \sigma_\nu^2)$, where $\mu_\nu = \nu^* - \ell'(\nu^*)/\ell''(\nu^*)$ and $\sigma_\nu^2 = -1/\ell''(\nu^*)$. $\ell'(\nu^*)$ and $\ell''(\nu^*)$ are the first and second derivatives of $\ell(\nu)$ evaluated at $\nu = \nu^*$. To proof the concavity of $\ell(\nu)$, we use the result of Abramowitz and Stegun (1970), in which the $\log\Gamma(\nu)$ could be approximated as

$$\log\Gamma(\nu) \quad = \quad \frac{\log(2\pi)}{2} + \frac{2\nu - 1}{2}\log(\nu) - \nu + \frac{\theta}{12\nu}, \tag{21}$$

for $0 < \theta < 1$. Taking the second derivative of $\ell(\nu)$ from (20) and using (21), we have that

$$\ell''(\nu) \quad = \quad -\frac{T\theta}{3\nu^3} - \frac{(T + 2a_\nu - 2)\nu}{2\nu^2} < 0.$$

- **SV-S case**

Using the fact that $U_t \sim \mathcal{B}(\nu, 1)$, we have that the full conditional of λ_t is given by

$$p(U_t \mid \mathbf{y}_t, \boldsymbol{\beta}_t, \nu) \quad \propto U_t^{\nu + \frac{r}{2} - 1} \mathrm{e}^{-\frac{U_t}{2}\sum_{j=1}^{r}\frac{(y_t(\tau_j)-\Lambda_j\beta_t)^2}{\sigma_j^2}} \mathbb{I}_{0<U_t<1}, \tag{22}$$

that is $U_t \sim \mathcal{G}_{(0<U_t<1)}(\nu + \frac{r}{2}, \frac{1}{2}\sum_{j=1}^{r}\frac{(y_t(\tau_j)-\Lambda_j\beta_t)^2}{\sigma_j^2})$, i.e., the right truncated gamma distribution.

Assuming that a prior distribution of $\nu \sim \mathcal{G}(a_\nu, b_\nu)$, the full conditional distribution of ν is given by

$$p(\nu \mid \mathbf{U}_{1:T}) \quad \propto \nu^{T + a_\nu - 1} \mathrm{e}^{-\nu(b_\nu - \sum_{t=1}^{T}\log U_t)} \mathbb{I}_{\nu>1}, \tag{23}$$

we have, $\nu \sim \mathcal{G}_{(\nu>1)}(T + a_\nu, b_\nu - \sum_{t=1}^{T}\log U_t)$, i.e. the left truncated gamma distribution. We simulate from the right and left truncated gamma distributions using the algorithm proposed by Philippe (1997).

- **SV-VG case**

As $U_t \sim \mathcal{IG}(\frac{\nu}{2}, \frac{\nu}{2})$, the full conditional of U_t is given by

$$p(U_t \mid \mathbf{y}_t, \boldsymbol{\beta}_t, \nu) \quad \propto U_t^{-\frac{\nu}{2} + \frac{r}{2} - 1} \mathrm{e}^{-\frac{1}{2}(U_t\sum_{j=1}^{r}\frac{(y_t(\tau_j)-\Lambda_j\beta_t)^2}{\sigma_j^2} + \frac{\nu}{U_t})}, \tag{24}$$

which is $\mathcal{GIG}(-\frac{\nu}{2} + \frac{r}{2}, \sum_{j=1}^{r}\frac{(y_t(\tau_j)-\Lambda_j\beta_t)^2}{\sigma_j^2}, \nu)$, the generalized inverse gaussian distribution.

We set the prior distribution of ν as $\mathcal{G}(a_\nu, b_\nu)\mathbb{I}_{0<\nu\leq 40}$. Then, the full conditional of ν is

$$p(\nu \mid \mathbf{U}_{1:T}) \quad \propto \quad \frac{\left[\frac{\nu}{2}\right]^{\frac{T\nu}{2}} \nu^{a_\nu - 1} \mathrm{e}^{-\frac{\nu}{2}[\sum_{t=1}^{T}(\frac{1}{U_t} + \log U_t)+2b_\nu]}}{[\Gamma(\frac{\nu}{2})]^T} \mathbb{I}_{0<\nu\leq 40} \tag{25}$$

which is log-concave. Thus, we sample ν by the Metropolis-Hastings acceptance-rejection algorithm as in the case of the SV-t model with proposal density $\mathcal{N}_{(0,40)}(\mu_\nu, \sigma_\nu^2)$.

Full condition of $\boldsymbol{\beta}_{0:T}$

In order to facilitate the exposition we supprese the dependence in λ, $\mathbf{A}$, $\sigma_1^2, \ldots, \sigma_r^2$ and $\mathbf{W}^{-1}$. An efficient way to sample $\boldsymbol{\beta}_{0:T}$ is sampling them in one go. The key to doing so is the following decomposition of the posterior density

$$p(\boldsymbol{\beta}_{0:T} \mid .) \quad = \quad p(\boldsymbol{\beta}_T \mid \mathbf{y}_{0:T}) \prod_{t=0}^{T-1} p(\boldsymbol{\beta}_t \mid \boldsymbol{\beta}_{t+1}, \mathbf{y}_{0:t}) \tag{26}$$

Since $p(\boldsymbol{\beta}_T \mid \mathbf{y}_{0:T})$ and $p(\boldsymbol{\beta}_t \mid \boldsymbol{\beta}_{t+1}, \mathbf{y}_{0:t})$ are normally distributed, it is straightforward to obtain a sample using the decomposition in (26) since the quantities needed for the conditional densities are supplied by the Kalman filter. Then the simulation scheme runs as follows

- Sample $\boldsymbol{\beta}_T^{(i)} \sim \mathcal{N}_p(\mathbf{m}_T, \mathbf{C}_T)$
- For $t = T-1, \ldots, 0$ sample $\boldsymbol{\beta}_t^{(i)}$ de $p(\boldsymbol{\beta}_t \mid \boldsymbol{\beta}_{t+1}^{(i)}, \mathbf{y}_{0:t})$.

Then, $\boldsymbol{\beta}_{0:T}^{(i)}$ is a sample from $p(\boldsymbol{\beta}_{0:T} \mid \mathbf{y}_{0:T})$. Repeat the simulation scheme to obtain an i.i.d. sample from $p(\boldsymbol{\beta}_{0:T} \mid \mathbf{y}_{1:T})$. Alternatively the algorithm of de Jong and Shepard (1995) could be used to simulate the latent factors $\boldsymbol{\beta}_{0:T}^{(i)}$.

References

Abramowitz M, Stegun N. 1970. *Handbook of Mathematical Functions.* New York: Dover Publications, Inc.

Andrews DF, Bickel PJ, Hampel FR, Huber PJ, Rogers WH, Tukey J. 1972. *Robust Estimates of Location: Survey and Advances.* Princeton, NJ: Princeton University Press.

Andrews DF, Mallows SL. 1974. Scale mixtures of normal distributions. *Journal of the Royal Statistical Society, Series B* **36**: 99–102.

Chambers DR, Carleton WT, Waldman DW. 1984. A new approach to estimation of the term structure of interest rates. *Journal of Financial and Quantitative Analysis* **19**: 233–252.

Chib S. 1995. Marginal likelihood from the Gibbs output. *Journal of the American Statistical Association* **90**: 1313–1321.

Chibb S, Greenberg E. 1994. Bayes inference in regression models with ARMA(p,q) errors. *Journal of Econometrics* **64**: 183–206.

Chow STB, Chan JSK. 2008. Scale mixtures distributions in statistical modelling. *Australian & New Zeland Journal of Statistics* **50**: 135–146.

Cox JC, Ingersoll JE, Ross SA. 1985. A theory of the term structure of interest rates. *Econometrica* **53**: 385–407.

Dai Q, Singleton KJ. 2000. Specification analysis of affine term structure models. *Journal of Finance* **55**: 1943–1978.

de Jong P, Shepard N. 1995. The simulation smoother for time series models. *Biometrika* **82**: 339–350.

de Pooter MD, Segers R, van Dijk HK. 2006. On the practice of bayesian inference in basic economic time series models using gibbs sampling. Discussion paper 2006-076/4, Tinbergen Institute, Erasmus University Rotterdam.

Diebold FX, Ji L, Li C. 2006a. A three-factor yield curve model: non-affine structure, systematic risk sources, and generalized duration. In Klein L (ed.) *Macroeconomics, Finance and Econometrics: Essays in Memory of Albert Ando*. Cheltenham, U.K.: Edward Elgar.

Diebold FX, Li C. 2006. Forecasting the term structure of goverments bond yields. *Journal of Econometrics* **130**: 337–364.

Diebold FX, Piazzesi M, Rudebusch GD. 2005. Modeling bond yields in finance and macroeconomics. *American Economic Review* **95**: 415–420.

Diebold FX, Rudebusch GD, Aruoba B. 2006b. The macroeconomy and the yield curve: a dynamic latent factor approach. *Journal of Econometrics* **131**: 309–338.

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

Fabozzi FJ, Martellini L, Priaulet P. 2005. Predictability in the shape of the term structure of interest rates. *Journal of Fixed Income* **15**: 40–53.

Fama EF, Bliss RR. 1987. The information in long-maturity forward rates. *American Economic Review* **77**: 680–692.

Geweke J, Tanizaki H. 2001. Bayesian estimation of state space models using Metropolis-Hastings algorithm within Gibbs sampling. *Computacional Statistics & Data Analysis* **37**: 151–170.

Gross AM. 1973. A Monte Carlo swindle for estimators of location. *Journal of the Royal Statistical Society, Series C. Applied Statistics* **22**: 347–353.

Jarrow R, Ruppert D, Yu Y. 2004. Estimating the term structure of corporate debt with a semiparametric penalized spline model. *Journal of the American Statistical Association* **99**: 57–66.

Lange KL, Sinsheimer JS. 1993. Normal/independent distributions and their applications in robust regression. *J. Comput. Graph. Stat* **2**: 175–198.

Li M, Yu Y. 2006. A robust approach to the interest rates term structure estimation. *Journal of Data Science* **4**: 169–188.

Madan D, Seneta E. 1990. The variance gamma (v.g) model for share market return. *Journal Business* **63**: 511–524.

McCulloch JH. 1971. Measuring the term structure of interest rates. *Journal Business* **44**: 19–31.

McCulloch JH. 1975. The tax-adjusted yield curve. *Journal of Finance* **30**: 811–830.

Moench E. 2006. Term structure surprises: The predictive content of curvature, level, and slope. Working paper, Federal Reserve Bank of New York.

Moench E. 2008. Forecasting the yield curve in a data-rich environment: A no-arbitrage factor-augmented VAR approach. *Journal of Econometrics* **146**: 26–43.

Morgenthaler S, Tukey J. 1991. *Configural Polysampling: A Route to Practical Robustness.* New York: Wiley.

Nelson CR, Siegel AF. 1987. Parsimonious modeling of yield curve. *Journal of Business* **60**: 473–489.

Pemstein D, Quinn KV, Martin AD. 2007. The Scythe statistical library: An open source C++ library for statistical computation. *Journal of Statistical Software* **V**: 1–29.

Philippe A. 1997. Simulation of right and left truncated gamma distributions by mixtures. *Statistics and Computing* **7**: 173–181.

Schwartz T. 1998. Estimating the term structure of corporate debt. *Review of Derivatives Research* **2**: 193–230.

Shea G. 1984. Pitfalls in smoothing interest rate term structure data: Equilibrium models and spline approximations. *Journal of Financial and Quantitative Studiess* **19**: 253–269.

Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A. 2002. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, Series B* **64**: 621–622.

Svensson LEO. 1994. Estimating and interpreting forward interest rates: Sweden 1992-94. Working paper, 114, International Monetary Fund, Washington DC.

Tierney L. 1994. Markov chains for exploring posterior distributions (with discussion). *Annal of Statistics* **21**: 1701–1762.

Vasicek O. 1977. An equilibrium characterization of term structure. *Journal of Financial Economics* **5**: 177–188.

Wang J, Genton MG. 2006. The multivariate skew-slash distribution. *Journal of Statistical Planning and Inference* **136**: 209–220.

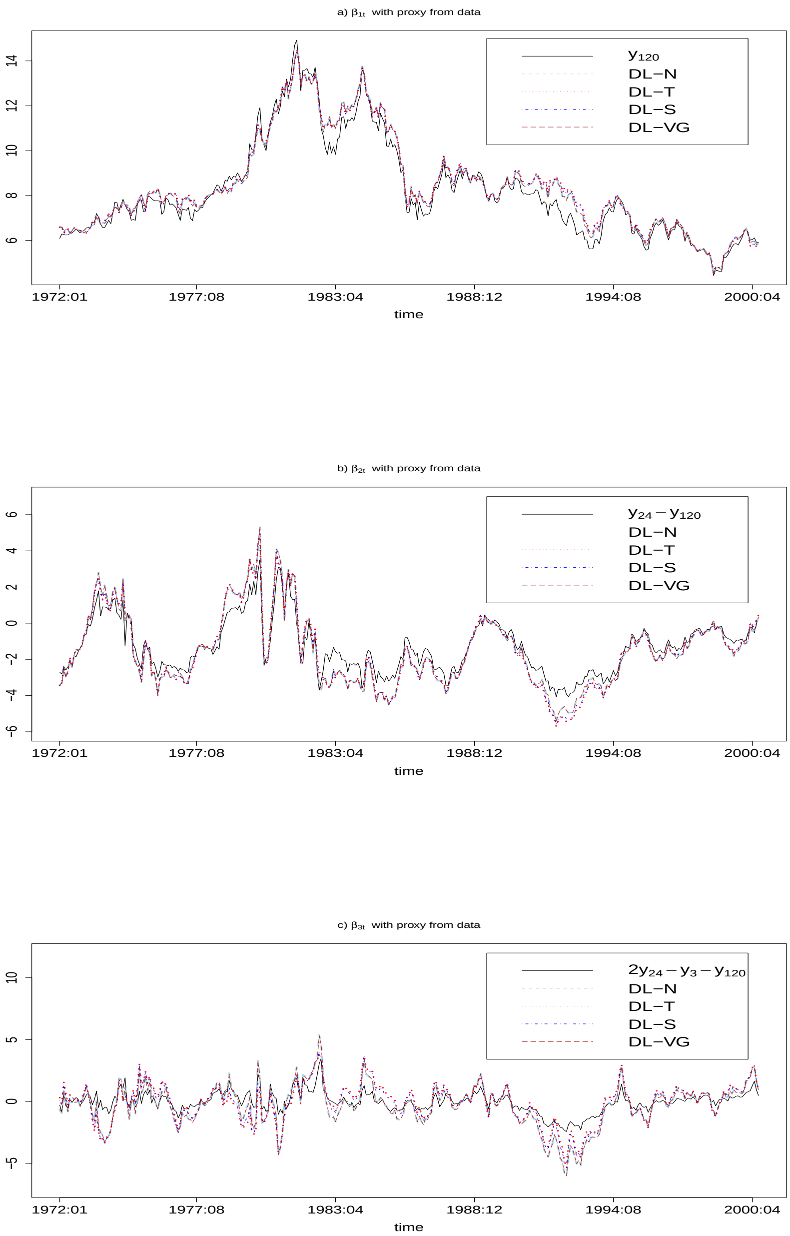

Figure 4: US yield data set. Posterior smoothed mean: a) Level ($\beta_{1,t}$), b) Slope ($\beta_{2,t}$) and c) Curvature ($\beta_{3,t}$)

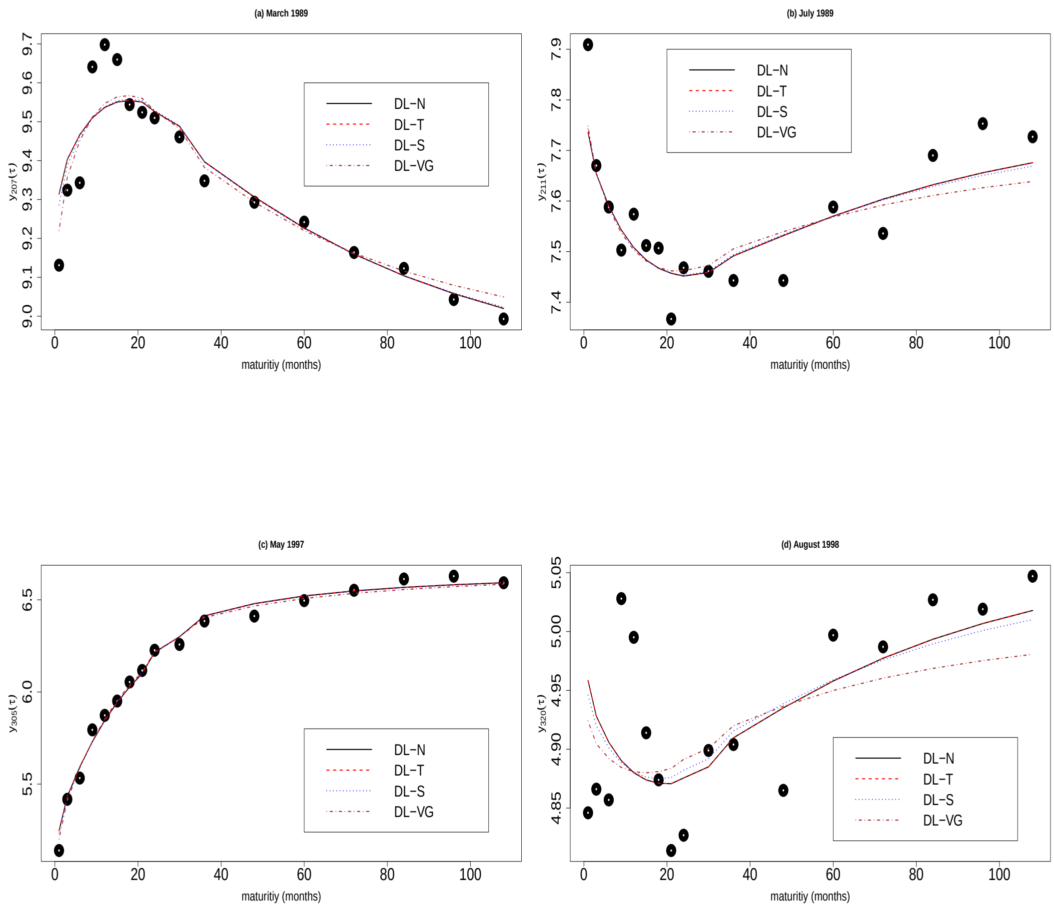

Figure 5: US yield data set. The Figure represents the posterior mean of $\mathbf{\Lambda}\boldsymbol{\beta}$ for each model. The points represent the actual yield curve.

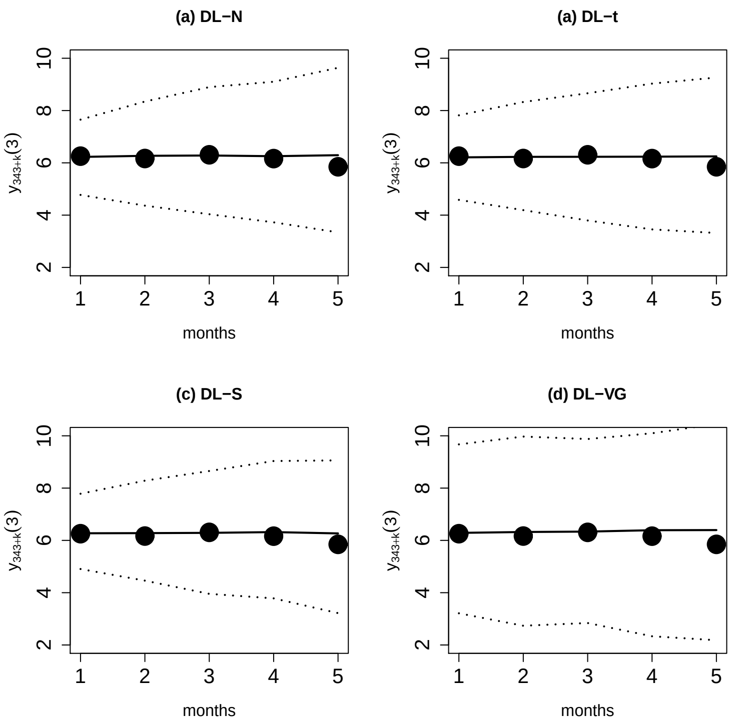

Figure 6: US yield data set. One step forecast

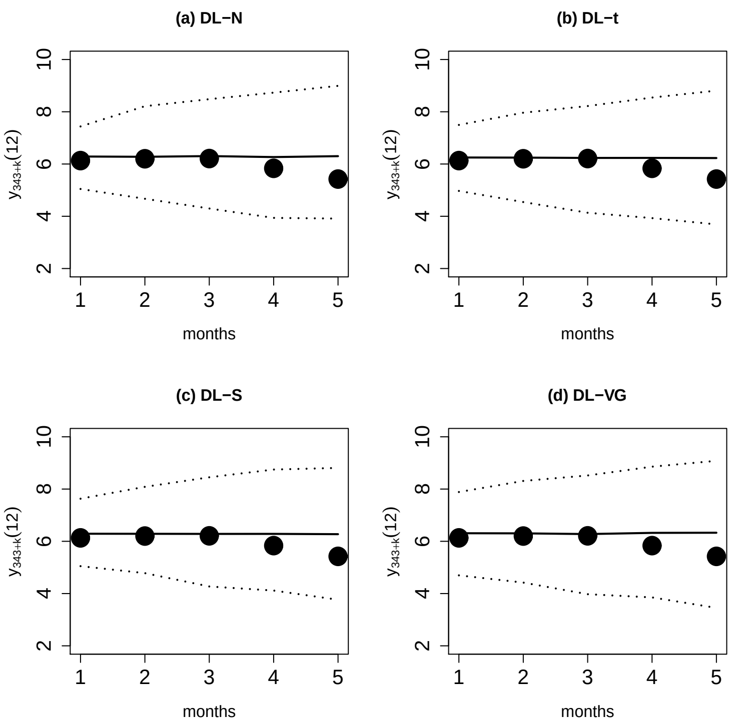

Figure 7: US yield data set. One step forecast