

SFB 649 Discussion Paper 2010-004

# Bayesian Inference in a Stochastic Volatility Nelson-Siegel Model

Nikolaus Hautsch\*  
Fuyu Yang\*



\* Humboldt-Universität zu Berlin, Germany

This research was supported by the Deutsche  
Forschungsgemeinschaft through the SFB 649 "Economic Risk".

<http://sfb649.wiwi.hu-berlin.de>  
ISSN 1860-5664

SFB 649, Humboldt-Universität zu Berlin  
Spandauer Straße 1, D-10178 Berlin



SFB 649 ECONOMIC RISK BERLIN

# Bayesian Inference in a Stochastic Volatility Nelson-Siegel Model

Nikolaus Hautsch\*

Humboldt-Universität zu Berlin, CASE, CFS  
Quantitative Products Laboratory

Fuyu Yang†

Institute for Statistics and Econometrics  
Humboldt-Universität zu Berlin

January, 2010

## Abstract

In this paper, we develop and apply Bayesian inference for an extended Nelson-Siegel (1987) term structure model capturing interest rate risk. The so-called Stochastic Volatility Nelson-Siegel (SVNS) model allows for stochastic volatility in the underlying yield factors. We propose a Markov chain Monte Carlo (MCMC) algorithm to efficiently estimate the SVNS model using simulation-based inference. Applying the SVNS model to monthly U.S. zero-coupon yields, we find significant evidence for time-varying volatility in the yield factors. This is mostly true for the level and slope volatility revealing also the highest persistence. It turns out that the inclusion of stochastic volatility improves the model's goodness-of-fit and clearly reduces the forecasting uncertainty particularly in low-volatility periods. The proposed approach is shown to work efficiently and is easily adapted to alternative specifications of dynamic factor models revealing (multivariate) stochastic volatility.

*Keywords:* term structure of interest rates, stochastic volatility, dynamic factor

---

\*Institute for Statistics and Econometrics and Center for Applied Statistics and Economics (CASE), Humboldt-Universität zu Berlin as well as Quantitative Products Laboratory (QPL), Berlin, and Center for Financial Studies (CFS), Frankfurt. Email: nikolaus.hautsch@wiwi.hu-berlin.de. Address: Spandauer Str. 1, D-10099 Berlin, Germany.

†Institute for Statistics and Econometrics, Humboldt-Universität zu Berlin. Email: fuyu.yang@wiwi.hu-berlin.de. Address: Spandauer Str. 1, D-10099 Berlin, Germany.

model, Markov chain Monte Carlo

*JEL classification:* C5, C11, C32

## Acknowledgements

This paper has been presented on the 3rd International Conference on Computational and Financial Econometrics (CFE), 2009, Cyprus. The authors gratefully acknowledge financial support from the EU Commission through MRTN-CT-2006-034270 COMISEF as well as the Deutsche Forschungsgemeinschaft through the SFB 649 “Economic Risk”.

## 1 Introduction

Modelling the term structure of interest rates is of importance in many areas in financial economics and macroeconomics. In finance, information revealed by the yield curve is important for the pricing of bonds and interest rate derivatives, for portfolio management and asset allocation. In macroeconomics, the yield curve carries important information for the state of the economy and business cycles. While traditional approaches as, e.g., Vasicek (1977), Cox et al., (1985) or Hull and White (1990) focus on equilibrium or no-arbitrage relationships, interest rate dynamics are typically captured in terms of factor models as, for instance, proposed by Nelson and Siegel (1987). Diebold and Li (2006) re-formulate the Nelson-Siegel model in terms of a state-space representation which allows for a two-step estimation of the factors and dynamics in the factor loadings. Koopman et al. (2010) and Hautsch and Ou (2008) extend the Nelson-Siegel model to allow for time-varying volatility. While Koopman et al. (2010) allow for a common volatility component in all yield processes, Hautsch and Ou (2008) propose capturing stochastic volatility in the underlying yield factors associated with level volatility, slope volatility and curvature volatility. While it is shown that the so-called Stochastic Volatility Nelson Siegel model – henceforth SVNS model – is a powerful approach to parsimoniously capture dynamics in yields and corresponding volatilities, statistical inference for such a model is not straightforward since both yield factors and volatility factors are unobservable.

In this paper, we show how to efficiently estimate the SVNS model using Bayesian techniques and propose a specific Markov chain Monte Carlo (MCMC) approach. We illustrate the importance of accounting for stochastic volatility in the Nelson-Siegel model and show how to extract the unobservable volatility components from the data. Though the estimation procedure is specifically designed for the SVNS model, it is

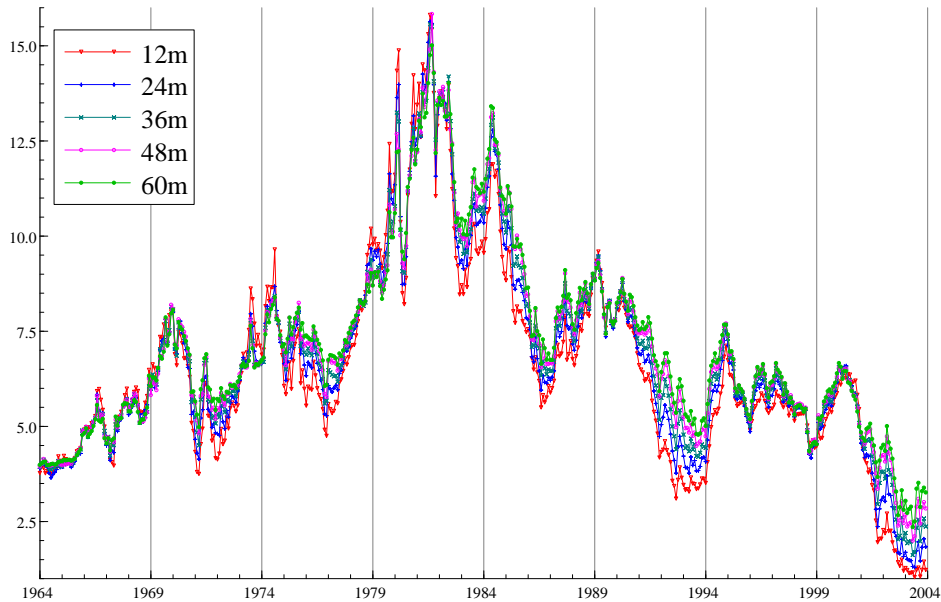
easily adapted to alternative factor specifications. In this sense, the proposed algorithm provides a general framework for the estimation of dynamic factor models revealing (multivariate) stochastic volatility. In an empirical application to U.S. bond yields, we illustrate that the proposed procedure works well and allows to efficiently extract unobservable time-varying volatility components.

The exponential components factor model proposed by Nelson and Siegel (1987) is a workhorse for the estimation and prediction of yield curves and is extensively used in financial practice and central banks. Its main power stems from the fact that it is easy to implement and is sufficiently flexible to capture a wide range of possible shapes of the yield curve. Though it is neither an equilibrium nor a no-arbitrage model, many banks use this framework to construct zero-coupon yield curves. Various extensions of the Nelson-Siegel model have been proposed, see, e.g., Björk and Christensen (1999), Rudebusch and Wu (2008), Diebold et al. (2005) and Diebold et al. (2006). Diebold and Li (2006) propose a dynamic version of the Nelson-Siegel model by allowing the underlying factor loadings to vary over time following a vector autoregressive (VAR) structure. As shown by Diebold and Li (2006) and Diebold et al. (2006), the model is able to capture interest rate dynamics and to successfully predict future yield curves. These contributions opened up a new way to model interest rate dynamics using factor models and complement the class of non-arbitrage affine models (Vasicek, 1977, Cox et al. 1985, Duffie and Kan, 1996 or Dai and Singleton, 2000, among others).<sup>1</sup>

Figure 1 gives an illustration of time series plots of yields with different maturities stemming from the data underlying this study. We observe that yields with different maturities are closely related and tend to move together. Also, we find evidence for time-varying volatility in the interest rate series. Particularly in the 1980s, yields for all maturities are very volatile. However, capturing time-varying volatility in yield curves is challenging due to their high dimensionality. Koopman et al. (2010) extend the dynamic Nelson-Siegel model by allowing for a common volatility component jointly affecting the yield processes for all maturities. A common volatility component can be associated with the volatility of an underlying bond market portfolio in the spirit of Engle et al. (1990). However, such a specification is not flexible enough to capture specific maturity-dependent volatilities. As a more flexible but still parsimonious alternative, Hautsch and Ou (2008) propose modelling stochastic volatility in the yield curve factors directly. Then, the time-varying volatilities in individual yields are captured by yield factor volatilities. These volatilities are naturally interpreted as the volatilities of underlying bond portfolios associated with short-term, medium-term and long-term

---

<sup>1</sup>See Piazzesi (2003) for a survey on affine models in continuous time.



**Figure 1:** Time series plots of U.S. zero coupon yields, Jan 1964 to Dec 2003. Maturities: 12, 24, 36, 48, 60 months, respectively.

maturities.

However, extracting the latent factors as well as their time-varying volatility components is challenging. By proposing an efficient MCMC algorithm to conduct Bayesian inference in the SVNS model, we contribute to the literature on the estimation of (extended) multivariate dynamic factor models. To extract the factors, we use the Kalman filter algorithm imbedded in an MCMC procedure. The unobservable time-varying volatilities are extracted using an approximating re-weighting approach proposed by Kim et al. (1998) and Chib et al. (2002). Using the suggested MCMC algorithm, all latent yield factors, stochastic volatilities and parameters can be sampled at once in a few blocks. Using MCMC diagnostics, we show that the proposed procedure is computationally quite efficient and is clearly superior to an element-by-element sampling of the underlying parameters and latent factors as, e.g., used in Hautsch and Ou (2008). The latter procedure requires an enormous amount of Monte Carlo drawings making the model intractable if the sample size is high. In this sense, the procedure proposed in the present paper is a contribution to make inference in this class of models feasible and tractable.

Applying the model to monthly zero-coupon yields covering a period from 1964 to 2003, we show that the SVNS model is able to successfully capture the dynamics in time-varying yields and volatilities thereof. We find strong evidence for time-varying volatility in interest rates. All yield volatilities reveal a quite high persistence. To

explain the volatilities in the underlying yield curve, the level and slope volatility are particularly important. To evaluate the model's goodness-of-fit, we compute posterior predictive p-values. Results show that the explicit inclusion of stochastic volatilities in level, slope and curvature factors improves the fit compared to the basic specification. More importantly, it turns out that the incorporation of stochastic volatilities clearly reduces parameter uncertainty. Analyzing model forecasts for three selected periods we show that the SVNS specification yields a significantly smaller variance of the forecast density. In this sense, the proposed approach produces more exact interest rate forecasts.

The remainder of the paper is organized as follows. Section 2 presents the Stochastic Volatility Nelson-Siegel model. Section 3 illustrates the underlying MCMC procedure to estimate the model. Section 4 presents the empirical results while Section 5 concludes.

## 2 The Stochastic Volatility Nelson-Siegel (SVNS) Model

Let  $t$  denote calendar time and  $p_t(\tau_i)$  denote the price of a  $\tau_i$ -period discount bond at month  $t$  with  $\tau_i, i = 1, \dots, N$ , representing the maturity. Moreover,  $y_t(\tau_i)$  is the continuously compounded zero-coupon nominal yield to maturity with  $p_t(\tau_i) = \exp[-\tau_i y_t(\tau_i)]$ . Then, the instantaneous nominal forward rate curve is given by  $f_t(\tau_i) = -p_t'(\tau_i)/p_t(\tau_i)$  or correspondingly  $y_t(\tau_i) = \tau_i^{-1} \int_0^{\tau_i} f_t(u) du$ . Nelson and Siegel (1987) propose modelling the forward rate curve as

$$f_t(\tau_i) = f_{1,t} + f_{2,t}e^{-\lambda\tau_i} + f_{3,t}\lambda e^{-\lambda\tau_i},$$

where  $e^{-\lambda\tau_i}$  and  $\lambda e^{-\lambda\tau_i}$  denote Laguerre polynomials whose shapes are determined by  $\lambda$ . Small (large) values of  $\lambda$  produce slow (fast) decays and a better fit of the curve at long (short) maturities. Correspondingly,  $f_{1,t}$ ,  $f_{2,t}$  and  $f_{3,t}$  denote time-varying factors capturing the dynamic behaviour of the forward curve.

The corresponding yield curve is given by

$$y_t(\tau_i) = f_{1,t} + f_{2,t} \left[ \frac{1 - e^{-\lambda\tau_i}}{\lambda\tau_i} \right] + f_{3,t} \left[ \frac{1 - e^{-\lambda\tau_i}}{\lambda\tau_i} - e^{-\lambda\tau_i} \right]. \quad (1)$$

Diebold and Li (2006) suggest interpreting the parameters  $f_{1,t}$ ,  $f_{2,t}$  and  $f_{3,t}$  as three latent dynamic factors with loadings 1,  $(1 - e^{-\lambda\tau_i})/\lambda\tau_i$ , and  $\{(1 - e^{-\lambda\tau_i})/\lambda\tau_i\} - e^{-\lambda\tau_i}$ , respectively. Accordingly,  $f_{1,t}$  is associated with a long-term factor whose loading is constant for all maturities. Since shocks in  $f_{1,t}$  affect all yields simultaneously,  $f_{1,t}$  is commonly referred to as a level factor. Conversely, the loading of  $f_{2,t}$  starts at one for

$\tau_i = 0$  and decays monotonically to zero. Consequently, shocks in  $f_{2,t}$  predominantly affect only short-term yields and thus induce variations in yield spreads. Therefore,  $f_{2,t}$  is referred to as a slope factor. In Rudebusch and Wu (2008), the level factor is related to inflation expectations, whereas movements in the slope factor are linked to cyclical variations in inflation and output gaps. Finally,  $f_{3,t}$  is interpreted as a medium-term factor since its loading is zero for  $\tau_i = 0$ , increases for mid-term maturities but decays to zero in the limit. Correspondingly, shocks in  $f_{3,t}$  dominantly affect the yield curve's curvature. Diebold et al. (2006) report negligible responses of macroeconomic variables to shocks in the curvature factor. Conversely, Mönch (2006) argues that a flattening of the yield curve associated with changes in the curvature factor can be linked to a slow-down of the economy.

The Nelson-Siegel model can be seen as a parsimonious parametric factor model, which is flexible enough to capture a wide range of different shapes. Moreover, the model implies desirable limiting behaviours of forward and yield curves. For  $\tau_i \rightarrow 0$ , the short rate is  $\lim_{\tau_i \rightarrow 0} y_t(\tau_i) = f_{1,t} + f_{2,t}$ . Correspondingly, we have  $\lim_{\tau_i \rightarrow \infty} y_t(\tau_i) = f_{1,t}$ .

Diebold and Li (2006) propose re-formulating the Nelson-Siegel model – henceforth Diebold-Li Nelson-Siegel (DLNS) model – in terms of a state-space system

$$\begin{pmatrix} y_t(\tau_1) \\ y_t(\tau_2) \\ \vdots \\ y_t(\tau_N) \end{pmatrix} = \begin{pmatrix} 1 & \frac{1-e^{-\lambda\tau_1}}{\lambda\tau_1} & \frac{1-e^{-\lambda\tau_1}}{\lambda\tau_1} - e^{-\lambda\tau_1} \\ 1 & \frac{1-e^{-\lambda\tau_2}}{\lambda\tau_2} & \frac{1-e^{-\lambda\tau_2}}{\lambda\tau_2} - e^{-\lambda\tau_2} \\ \vdots & \vdots & \vdots \\ 1 & \frac{1-e^{-\lambda\tau_N}}{\lambda\tau_N} & \frac{1-e^{-\lambda\tau_N}}{\lambda\tau_N} - e^{-\lambda\tau_N} \end{pmatrix} \begin{pmatrix} f_{1,t} \\ f_{2,t} \\ f_{3,t} \end{pmatrix} + \begin{pmatrix} \varepsilon_t(\tau_1) \\ \varepsilon_t(\tau_2) \\ \vdots \\ \varepsilon_t(\tau_N) \end{pmatrix}, \quad (2)$$

where the time-varying coefficients  $f_{1,t}$ ,  $f_{2,t}$ , and  $f_{3,t}$  are interpreted as latent factors following a vector autoregressive process (VAR) given by

$$\begin{pmatrix} f_{1,t} - \mu_{1,f} \\ f_{2,t} - \mu_{2,f} \\ f_{3,t} - \mu_{3,f} \end{pmatrix} = \begin{pmatrix} \phi_{11,f} & \phi_{12,f} & \phi_{13,f} \\ \phi_{21,f} & \phi_{22,f} & \phi_{23,f} \\ \phi_{31,f} & \phi_{32,f} & \phi_{33,f} \end{pmatrix} \begin{pmatrix} f_{1,t-1} - \mu_{1,f} \\ f_{2,t-1} - \mu_{2,f} \\ f_{3,t-1} - \mu_{3,f} \end{pmatrix} + \begin{pmatrix} \eta_{1,t} \\ \eta_{2,t} \\ \eta_{3,t} \end{pmatrix}. \quad (3)$$

By defining  $y_t = \{y_t(\tau_1), y_t(\tau_2), \dots, y_t(\tau_N)\}'$ ,  $\phi_f = [\phi_{ij,f}]$  as a  $(3 \times 3)$  parameter matrix, and  $\Lambda$  as the matrix of factor loadings and denoting

$$\begin{aligned} f_t &= \{f_{1,t}, f_{2,t}, f_{3,t}\}', \\ \mu_f &= \{\mu_{1,f}, \mu_{2,f}, \mu_{3,f}\}', \end{aligned}$$

$$\begin{aligned}\varepsilon_t &= \{\varepsilon_t(\tau_1), \varepsilon_t(\tau_2), \dots, \varepsilon_t(\tau_N)\}' , \\ \eta_t &= \{\eta_{1,t}, \eta_{2,t}, \eta_{3,t}\}' ,\end{aligned}$$

(2) and (3) can be written as

$$f_t - \mu_f = \phi_f (f_{t-1} - \mu_f) + \eta_t, \quad (4)$$

$$y_t = \Lambda f_t + \varepsilon_t. \quad (5)$$

To limit the computational burden and to keep the model parsimonious, we assume  $\phi_f = \text{diag}\{\phi_{11,f}, \phi_{22,f}, \phi_{33,f}\}$  to be a diagonal matrix.

For the joint distribution of  $\varepsilon_t$  and  $\eta_t$ , we assume

$$\begin{pmatrix} \eta_t \\ \varepsilon_t \end{pmatrix} \sim N \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \Sigma_\eta \Sigma'_\eta & 0 \\ 0 & \Sigma_\varepsilon \Sigma'_\varepsilon \end{bmatrix} \right), \quad (6)$$

with  $\Sigma_\eta$  and  $\Sigma_\varepsilon$  denoting the corresponding Cholesky factors of the covariance matrices of  $\eta_t$  and  $\varepsilon_t$ , respectively. The covariance matrix  $\Sigma_\varepsilon \Sigma'_\varepsilon$  is assumed to be diagonal indicating that the measurement errors are contemporaneously uncorrelated.

To allow for conditional heteroscedasticity in the yield processes, Koopman et al. (2010) propose capturing yield curve volatility by allowing for a common variance component jointly affecting all individual yields. This factor can be interpreted as the volatility of an underlying bond market portfolio in the spirit of Engle et al. (1990) and Engle and Ng (1993). However, such a specification does not allow for different volatilities in individual yield processes. As a more flexible alternative, Hautsch and Ou (2008) propose allowing for stochastic volatility in the yield factors  $f_t$  directly. Consequently,  $\Sigma_\eta \Sigma'_\eta$  is assumed to vary over time, with  $\Sigma_{\eta,t} \Sigma'_{\eta,t}$  to be specified as

$$\begin{aligned} \text{diag}(\ln \Sigma_{\eta,t} \Sigma'_{\eta,t}) &= \begin{pmatrix} h_{1,t} \\ h_{2,t} \\ h_{3,t} \end{pmatrix} \\ \begin{pmatrix} h_{1,t+1} - \mu_{1,h} \\ h_{2,t+1} - \mu_{2,h} \\ h_{3,t+1} - \mu_{3,h} \end{pmatrix} &= \begin{pmatrix} \phi_{11,h} & 0 & 0 \\ 0 & \phi_{22,h} & 0 \\ 0 & 0 & \phi_{33,h} \end{pmatrix} \begin{pmatrix} h_{1,t} - \mu_{1,h} \\ h_{2,t} - \mu_{2,h} \\ h_{3,t} - \mu_{3,h} \end{pmatrix} + \begin{pmatrix} \sigma_1 \epsilon_{1,t} \\ \sigma_2 \epsilon_{2,t} \\ \sigma_3 \epsilon_{3,t} \end{pmatrix}, \quad (7) \end{aligned}$$

where  $\epsilon_{j,t} \stackrel{i.i.d.}{\sim} N(0, 1)$ . Correspondingly, with  $\eta_t = \{\eta_{1,t}, \eta_{2,t}, \eta_{3,t}\}'$ , the model can be re-written as



$$\begin{pmatrix} \eta_{1,t} \\ \eta_{2,t} \\ \eta_{3,t} \end{pmatrix} = \begin{pmatrix} e^{h_{1,t}/2} & 0 & 0 \\ 0 & e^{h_{2,t}/2} & 0 \\ 0 & 0 & e^{h_{3,t}/2} \end{pmatrix} \begin{pmatrix} \zeta_{1,t} \\ \zeta_{2,t} \\ \zeta_{3,t} \end{pmatrix}, \quad (8)$$

where  $\zeta_{j,t} \stackrel{i.i.d.}{\sim} N(0, 1)$  with  $j = 1, 2, 3$ .

Hence,  $h_{j,t}$  with  $j = 1, 2, 3$ , are (unobservable) stochastic volatility processes and can be summarized in the  $3 \times 1$  vector  $h_t = \{h_{1,t}, h_{2,t}, h_{3,t}\}'$ . Correspondingly, by defining

$$\begin{aligned} e^{h_t/2} &= \text{diag} \{e^{h_{1,t}/2}, e^{h_{2,t}/2}, e^{h_{3,t}/2}\}, \\ \sigma &= \text{diag} \{\sigma_1, \sigma_2, \sigma_3\}, \\ \mu_h &= \{\mu_{1,h}, \mu_{2,h}, \mu_{3,h}\}', \\ \phi_h &= \text{diag} \{\phi_{11,h}, \phi_{22,h}, \phi_{33,h}\}, \\ \zeta_t &= \{\zeta_{1,t}, \zeta_{2,t}, \zeta_{3,t}\}', \\ \epsilon_t &= \{\epsilon_{1,t}, \epsilon_{2,t}, \epsilon_{3,t}\}', \end{aligned}$$

eq. (8) can be re-written as

$$\eta_t = e^{h_t/2} \zeta_t, \quad (9)$$

$$h_{t+1} - \mu_h = \phi_h (h_t - \mu_h) + \sigma \epsilon_t. \quad (10)$$

The components  $h_{1,t}$ ,  $h_{2,t}$  and  $h_{3,t}$  can be interpreted as factor volatilities associated with time-varying uncertainty in the yield curves' level, slope and curvature. Since  $h_{1,t}$  is a common variance component it is in the spirit of the joint volatility factor in Koopman et al. (2010) and can be seen as a model implied proxy of bond market volatility as used in Engle et al. (1990). The factor  $h_{2,t}$  captures time-variations in yield spreads and thus can be interpreted as yield curve slope volatility. Similarly,  $h_{3,t}$  is related to the volatility of a bond portfolio dominated by mid-term maturities and thus captures uncertainties associated with the curvature of the yield curve.

To reduce the computational burden we rule out cross-dependencies between the individual volatility components. This is in the spirit of a low-dimensional factor structure capturing high-dimensional dynamics in volatility. Unconditional moments of this specification are derived in Hautsch and Ou (2008).

### 3 Bayesian Inference

#### 3.1 Estimation Algorithm

The key idea is to estimate the highly correlated latent variables in terms of several blocks using the Kalman filter together with a simulation smoother (de Jong and Shephard, 1995) within an MCMC algorithm. With models in a state-space form, the simulation smoother allows to simulate the underlying states while avoiding degenerate sampling problems. For more details, see de Jong and Shephard (1995). Define

$$\begin{aligned} H_{j,t} &= (h_{j,1}, \dots, h_{j,t})', \\ h_j &= (h_{j,1}, \dots, h_{j,T})', \\ h_{j,-t} &= (h_{j,1}, \dots, h_{j,t-1}, h_{j,t+1}, \dots, h_{j,T})', \\ F_{j,t}^* &= (f_{j,1}^*, \dots, f_{j,t}^*), \\ f_j^* &= (f_{j,1}^*, \dots, f_{j,T}^*). \end{aligned}$$

Then, following Kim et al. (1998), using eq. (9), we can rewrite eq. (3) as

$$f_{j,t}^* = h_{j,t} + z_{j,t}, \quad j = 1, 2, 3, \quad (11)$$

where  $f_{j,t}^* = \ln \left\{ [f_{j,t} - \mu_{j,f} - \phi_{j,f} (f_{j,t-1} - \mu_{j,f})]^2 + c \right\}$ , and  $z_{j,t} = \ln \left( \zeta_{j,t}^2 \right)$ . The offset parameter  $c$  is typically set to a small value, e.g., 0.001, see also Fuller (1996). Then, the state-space system can be re-written as

$$\begin{aligned} f_{j,t}^* &= h_{j,t} + z_{j,t} \\ h_{j,t+1} - \mu_{j,h} &= \phi_{j,h} (h_{j,t} - \mu_{j,h}) + \sigma_j \epsilon_{j,t} \end{aligned}$$

allowing the use of the Kalman filter and a simulation smoother. As shown below, this allows to sample the entire block  $h_j$  at once.

To make eq. (11) linear, Kim et al. (1998) propose approximating the distribution of  $z_{j,t}$  by a mixture of normal densities,

$$z_{j,t} \mid s_t \sim f_N(m_{s_t}, \nu_{s_t}^2), \quad (12)$$

where  $s_t \in \{1, 2, \dots, 7\}$  is an index indicator with

$$\Pr(s_t = i) = q_i, \quad i \leq 7, \quad t \leq T,$$

**Table 1:** Parameters of a seven-component Gaussian mixture to approximate the distribution of  $z_t$ , Kim et al. (1998).

$s_t$	$q_i$	$m_{s_t}$	$\nu_{s_t}^2$
1	0.00730	-11.40039	5.79596
2	0.10556	-5.24321	2.61369
3	0.00002	-9.83726	5.17950
4	0.04395	1.50746	0.16735
5	0.34001	-0.65098	0.64009
6	0.24566	0.52478	0.34023
7	0.25750	-2.35859	1.26261

and the parameters  $\{q_i, m_{s_t}, \nu_{s_t}^2\}$  given in Table 1.

Due to this approximation,  $f_j^* \mid s, \mu_{j,h}, \phi_{j,h}, \sigma_j$  with  $s = (s_1, \dots, s_T)'$  becomes Gaussian, which induces substantial computational efficiency gains in the MCMC algorithm shown below. The minor approximation error can be removed by re-weighting the posterior samples afterwards.

In the following, we develop an MCMC algorithm based on eight parameter blocks. By denoting the error precision matrix as  $H_\varepsilon = (\Sigma_\varepsilon \Sigma_\varepsilon')^{-1}$  and using squared brackets  $[\cdot]$  to indicate blocks of parameters, we collect all parameters of interest in  $\theta$ , where  $\theta = \{\lambda, H_\varepsilon, [\mu_{j,f}, \phi_{j,f}], [\mu_{j,h}, \phi_{j,h}, \sigma_j]\}$  with  $j = 1, 2, 3$ .

The priors are chosen as following: (i)  $\lambda$  is assumed to follow a Uniform distribution with  $\lambda \sim U(\lambda \mid \underline{a}_\lambda, \underline{b}_\lambda)$ . (ii)  $\Sigma_\varepsilon \Sigma_\varepsilon'$  follows an Inverted Wishart distribution  $\Sigma_\varepsilon \Sigma_\varepsilon' \sim IW(\Sigma_\varepsilon \Sigma_\varepsilon' \mid \underline{A}_\varepsilon^{-1}, \underline{\nu}_\varepsilon)$ . Then, by construction, the precision matrix  $H_\varepsilon$  follows a Wishart distribution  $H_\varepsilon \sim W(H_\varepsilon \mid \underline{A}_\varepsilon, \underline{\nu}_\varepsilon)$ . (iii) By restricting the yield factors to be stationary, we elicit truncated normal priors with support  $(-1, 1)$  for  $\phi_{j,f} \sim TN(\phi_{j,f} \mid \underline{\mu}_{j,f_\phi}, \underline{V}_{j,f_\phi}) \cdot 1(|\phi_{j,f}| < 1)$ , where  $1(\cdot)$  is the indicator function. The specific choices of the resulting set of hyper-parameters  $\{\underline{a}_\lambda, \underline{b}_\lambda, \underline{A}_\varepsilon, \underline{\nu}_\varepsilon, \underline{\mu}_{j,f}, \underline{V}_{j,f}\}$  are given in the Appendix.

Denote  $y = \{y_1, \dots, y_T\}'$ . According to the model specification given by eq. (4), (5) and (7), the likelihood function can be written as

$$p(y, f, h \mid \theta) = \prod_{t=1}^T \prod_{j=1}^3 p(y_t \mid \theta, f_{j,t}, h_{j,t}) p(f_{j,t} \mid F_{j,t-1}, H_{j,t-1}, \theta) p(h_{j,t} \mid H_{j,t-1}, \theta). \quad (13)$$

With the priors elicited above and employing Bayes' theorem, the joint posterior is proportional to the product of the likelihood and priors,

$$p(\theta, f, h \mid y) \propto p(y, f, h \mid \theta) p(\theta).$$

The underlying MCMC algorithm is summarized as follows:

1. Initialize  $\theta = \{\lambda, H_\varepsilon, [\mu_{j,f}, \phi_{j,f}], [\mu_{j,h}, \phi_{j,h}, \sigma_j]\}$  and the latent variables  $f_j$  and  $h_j$  with  $f_{j,0} = \mu_{j,f}$  and  $h_{j,0} = \mu_{j,h}$  for  $j = 1, 2, 3$ .
2. Run a Gibbs sampler for steps (a) - (d) using  $S$  replications, where the initial  $S_0$  draws are discarded:
  - (a) Sample  $H_\varepsilon$  from  $H_\varepsilon | y, \lambda, [\mu_{j,f}, \phi_{j,f}], [\mu_{j,h}, \phi_{j,h}, \sigma_j], f_j, h_j, \quad j = 1, 2, 3$ .
  - (b) Sample  $\lambda$  from  $\lambda | y, H_\varepsilon, [\mu_{j,f}, \phi_{j,f}], [\mu_{j,h}, \phi_{j,h}, \sigma_j], f_j, h_j$  using a Griddy-Gibbs sampling method as illustrated in Appendix A.
  - (c) Sample  $[\mu_{j,f}, \phi_{j,f}], f_j$  from  $[\mu_{j,f}, \phi_{j,f}], f_j | y, \lambda, H_\varepsilon, h_j$  using a simulation smoother as illustrated in Appendix A.
  - (d) Run (i)-(iv) 3 times to estimate  $h_j | f_j$ , with  $j = 1, 2, 3$ , respectively:
    - (i) Compute  $f_j^*$  and run the loop (ii)-(iv) 2000 times. Discard the results from the initial 500 loops.
    - (ii) Sample  $s | f_j^*, h_j$ .
    - (iii) Sample  $h_j | f_j^*, s, [\mu_{j,h}, \phi_{j,h}, \sigma_j]$  using a simulation smoother.
    - (iv) Sample  $[\mu_{j,h}, \phi_{j,h}, \sigma_j] | f_j^*, h_j$  using the Metropolis-Hastings algorithm proposed by Chib et al. (2002) choosing a multivariate  $t$ -density as the proposal density. See Appendix A for more details.

To estimate the DLNS model, step (d) is straightforwardly simplified to allow for a constant covariance matrix  $\Sigma_\eta \Sigma'_\eta$ . In step 2 (d) (iv), as an alternative to a multivariate  $t$ -density, an adaptive mixture of Student- $t$  can be used as a candidate distribution, see Ardia et al. (2009).

This MCMC algorithm is obviously more efficient than an element-by-element drawing algorithm. Due to the blocking of highly correlated variables, the Gibbs sampler can quickly move to regions of high posterior probability. Conversely, if we draw latent variables one by one, an enormous number of draws is needed as the latent variables and thus the Gibbs draws are highly correlated. These high correlations induce a slow convergence of the chain (Chib and Greenberg, 1996). Consequently, the number of iterations required in an element-by-element algorithm is significantly higher. For instance, in Hautsch and Ou (2008), 2,500,000 iterations are required with a burn-in period of 500,000 iterations. Using the MCMC algorithm introduced above, we can reduce the number of iterations to 20,000. This makes the model significantly more flexible and applicable, particularly, if the sample size becomes large.

### 3.2 Evaluating the Model Fit using Posterior Predictive p-values

Comparing two models  $M_i$  and  $M_j$  in a Bayesian framework can be performed by calculating the posterior model probabilities  $p(M_i | y)$  and  $p(M_j | y)$  and computing the posterior odds ratio  $PO_{ij} = p(M_i | y) / p(M_j | y)$ . Alternatively, one could compute the Bayes Factors  $BF_{ij} = p(y | M_i) / p(y | M_j)$  based on the marginal likelihoods  $p(y | M_i)$  and  $p(y | M_j)$ . Computing the latter is not straightforward in the given context. However, in this paper, we are particularly interested in an evaluation of the model's goodness-of-fit yielding information to which extent the specification is able to reproduce the characteristics of the data.

In an MCMC setting, this is conveniently evaluated using so-called posterior predictive p-values (ppp-values). Suppose we have  $S$  hypothetical sample series  $y_t^{pre}$ ,  $t = 1, \dots, T$ , which are generated by a specific model. Then, the ppp-value gives the (tail) probability to observe more extreme values than in the actually observed data. Hence, if the model fits the data well, the observed data,  $y_t^{obs}$ , and the simulated data,  $y_t^{pre}$ , should not be too different and thus the ppp-value should be high. See also Meng (1994) and Gelman and Meng (1996) for more details. Denote  $\theta$  as the vector of parameters, then  $p(g(y_t^{pre}) | y_t^{obs})$  can be calculated as (see Koop, 2003)

$$\begin{aligned} p(g(y_t^{pre}) | y_t^{obs}) &= \int p(g(y_t^{pre}) | y_t^{obs}, \theta) p(\theta | y_t^{obs}) d\theta \\ &= \int p(g(y_t^{pre}) | \theta) p(\theta | y_t^{obs}) d\theta. \end{aligned} \quad (14)$$

The second equality holds since conditional on  $\theta$ ,  $y_t^{pre}$  is independent of  $y_t^{obs}$  and thus  $p(g(y_t^{pre}) | y_t^{obs}, \theta) = p(g(y_t^{pre}) | \theta)$ . Suppose we have  $S$  draws of  $\theta$  from the posterior density  $p(\theta | y_t^{obs})$ , where each draw of  $\theta$  is denoted as  $\theta^{(i)}$ , the ppp-value can be calculated as the proportion of cases in which the simulated  $g(y_t^{pre})$  exceeds the realized value  $g(y_t^{obs})$

$$\begin{aligned} ppp &= \Pr \left[ g(y_t^{pre}) \geq g(y_t^{obs}) \right] \\ &= \frac{1}{S} \sum_{i=1}^S \mathbb{1} \left[ g(y_t^{pre}) \geq g(y_t^{obs}) \right], \end{aligned} \quad (15)$$

where  $\mathbb{1}(\cdot)$  denotes the indicator function. The function  $g(\cdot)$  is typically chosen as a sample statistic. Given the normality assumption for  $\varepsilon_t$ , we choose  $g(\cdot)$  as  $E(Kurt | y)$  and  $E(Skew | y)$  with  $Kurt$  and  $Skew$  denoting the sample kurtosis and skewness, respectively. Using the MCMC algorithm proposed in Section 3.1, we simulate  $S$  draws

of  $y_t^{pre(i)}$  and thus  $g\left(y_t^{pre(i)}\right) | \theta^{(i)}$ . In particular,  $y_t^{pre(i)} | \theta$  is simulated by

$$h_{j,t} | h_{j,t-1}, \mu_{j,h}, \phi_{j,h}, \sigma_j^2 \sim f_N \left\{ \mu_{j,h} + \phi_{j,h} (h_{j,t} - \mu_{j,h}), \sigma_j^2 \right\}, j = 1, 2, 3, \quad (16)$$

$$f_t | f_{t-1}, h_t, \theta \sim f_N \left\{ f_t + \phi_f (f_{t-1} - \mu_f), \Sigma_{\eta,t} \Sigma'_{\eta,t} \right\}, \quad (17)$$

$$y_t | f_t, h_{j,t}, \theta \sim f_N \left\{ \Lambda f_t, \Sigma_\varepsilon \Sigma'_\varepsilon \right\}, \quad (18)$$

yielding simulated data  $y_t^{pre(i)}, f_j^{(i)}, h_j^{(i)}$ . Then, the sample skewness and kurtosis in draw  $(i)$  is computed from  $\varepsilon_t^{obs(i)} = y_t^{obs(i)} - \Lambda^{(i)} f_t^{(i)}$  and  $\varepsilon_t^{pre(i)} = y_t^{pre(i)} - \Lambda^{(i)} f_t^{(i)}$ . Taking the average of the  $S$  draws,  $\{Skew^{obs(1)}, \dots, Skew^{obs(S)}\}$ , produces an estimate of  $E(Skew | y_t^{obs})$ . Then, the ppp-value is computed according to (15).

As noted in Koop (2003), the ppp-value can be used as a measurement of model fit as well as for model comparisons. A small ppp-value indicates that the model is unlikely to generate data sets with more extreme properties than the data observed.

## 4 Empirical Results

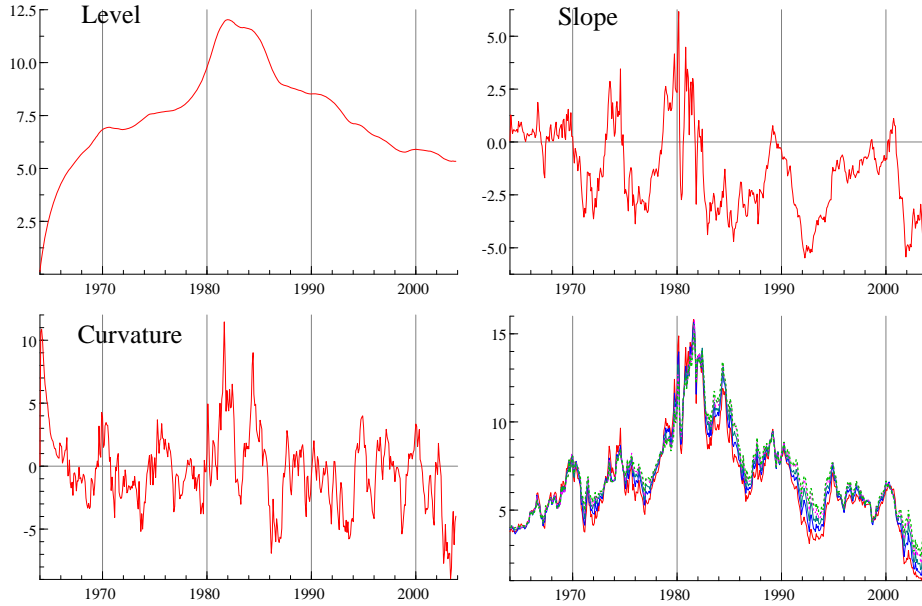
### 4.1 Estimation and MCMC diagnostics

We use monthly unsmoothed Fama-Bliss zero coupon yields from January 1964 to December 2003 with maturities from one up to five years, i.e.  $\tau_i = \{12, 24, 36, 48, 60\}$  in months. The data is available from the Centre for Research in Security Prices (CRSP) and is constructed using the method of Fama and Bliss (1987) based on end-of-month data of U.S. taxable, non-callable bonds for annual maturities up to 5 years. This data is used in Cochrane and Piazzesi (2005) and also in Hautsch and Ou (2008) and thus makes our results comparable to previous studies. In particular, since Hautsch and Ou (2008) use a Gibbs sampling algorithm based on an element-by-element drawing, a comparison of results provides insights to which extent the use of a computationally more efficient MCMC algorithm is also reflected in parameter estimates and uncertainties.

To be able to directly compare estimates of the DLNS and SVNS model, we also estimate the DLNS model using the suggested MCMC procedure. The proposed Gibbs sampler is run using 20,000 iterations with the first 5,000 iterations discarded. The algorithm is programmed in Ox building on the stochastic volatility package SvPack 2.0 (Koopman et al., 2002).<sup>2</sup> The estimation results including the estimated posterior means, posterior standard deviations, and 95% posterior density intervals (HPIV) are

---

<sup>2</sup>The algorithm is run on a machine with Intel(R) Core(TM)2Duo E8500 processor, 3.16GHz, 3.25GB RAM, and WinXP 32bit OS. For the basic model 25,000 iterations take 12 minutes, 11.01 seconds.

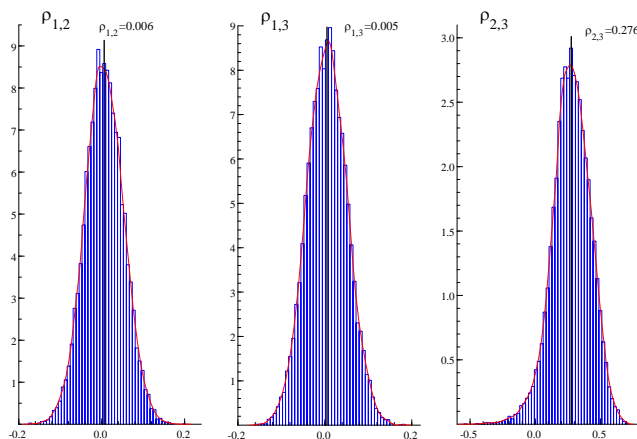


**Figure 2:** Latent Yield Factors Implied by the DLNS Model. Top left: level factor  $f_L$ ; Top right: slope factor  $f_S$ ; Bottom left: curvature factor  $f_C$ ; Bottom right: underlying yields Jan, 1964 - Dec, 2003.

given by Table 4 in Appendix B. We also report Numerical Standard Errors (NSE) as a measure of approximation errors in MCMC estimates, the Relative Numerical Efficiency (RNE) to evaluate the efficiency of the algorithm, and Convergence Diagnostics (CD) as proposed by Geweke (1992). Taking the non-i.i.d. nature of MCMC draws into account, we calculate modified NSEs and RNEs ( $NSE_q$  and  $RNE_q$ ) accounting for correlations up to lags of  $q\%$  of the size of the retained MCMC sample. According to the reported diagnostics, we do not find any indications for a lack of convergence in the Markov chain. Moreover, the RNE values indicate that the MCMC algorithm is quite effective.

Figure 2 plots the estimated latent yield factors. It turns out that all factor dynamics are highly persistent confirming also the results by Koopman et al. (2010). This is particularly true for the level factor which is virtually estimated as an integrated component. As shown below, this result is obviously induced by the fact that time-varying volatilities in levels are ignored driving the autoregressive parameter of the level factor toward one. Nevertheless, it is shown that the long-term factor  $f_{1,t}$  nicely captures the overall interest rate level, while  $f_{2,t}$  picks up variations in yield spreads. Also from the plots it is evident that time-varying variances are obviously only captured by the slope and curvature factor but not by the level factor. Using the posterior draws of the covariance  $\Sigma_\eta \Sigma'_\eta$ , we compute the error correlations  $\rho_{ik} = Corr(\eta_{i,t}, \eta_{k,t})$ ,  $i \neq k$ ,  $i \leq 3$ ,  $k \leq 3$ . Figure 3 shows the posterior distributions of  $(\rho_{12}, \rho_{13}, \rho_{23})$  with averages of

0.006, 0.005 and 0.276, respectively. Hence, apart from a distinct correlation between  $f_{2,t}$  and  $f_{3,t}$ , the mutual dependencies between the factors are comparably low widely confirming the appropriateness of modelling  $\Sigma_\eta \Sigma'_\eta$  as a diagonal matrix.

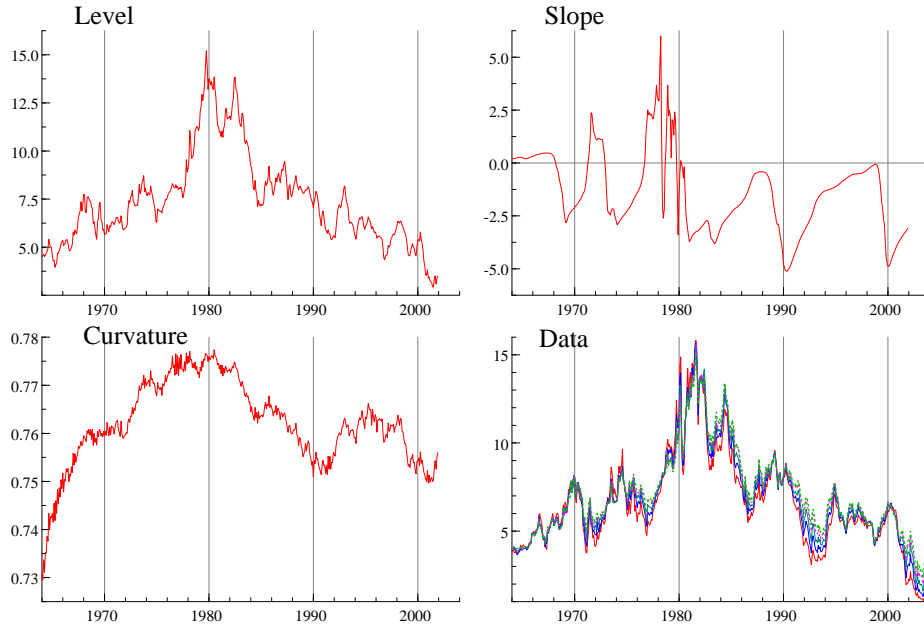


**Figure 3:** Posterior Distributions of Error Term Correlations based on the DLNS model.

The SVNS model is estimated using 18,000 runs of the Gibbs sampler with 8,000 draws used for the burn-in period. Table 2 gives the estimates of the parameters as well as the corresponding diagnostics. It turns out that – in comparison to the DLNS model – the persistence in yield factors is slightly reduced and obviously partly captured by the (high) autocorrelations in factor volatilities. Nevertheless, the persistence in the yield factor is still very high supporting the findings by Koopman et al. (2010) and Hautsch and Ou (2008). Conversely, the persistence in the curvature factor is clearly reduced. As for the DLNS model, the reported MCMC diagnostics indicate a proper convergence of the underlying Markov chain.

We find significant evidence for strong autocorrelations in factor volatilities with highest persistence in the level and slope volatility as represented by (mean) estimates of  $\phi_{h,L} = 0.89$  and  $\phi_{h,S} = 0.96$ . Hautsch and Ou (2008) report a higher persistence which might be due to a slower convergence of the Markov chain in an element-by-element drawing approach. The plots of the estimated yield factors shown in Figure 4 reveal quite similar patterns as those shown in Figure 2. However, it is evident that the SVNS model reveals also substantial volatility in the level factor which is not true in the DLNS model. These results illustrate that the inclusion of underlying stochastic volatility captures also a non-trivial part of parameter uncertainty. This will





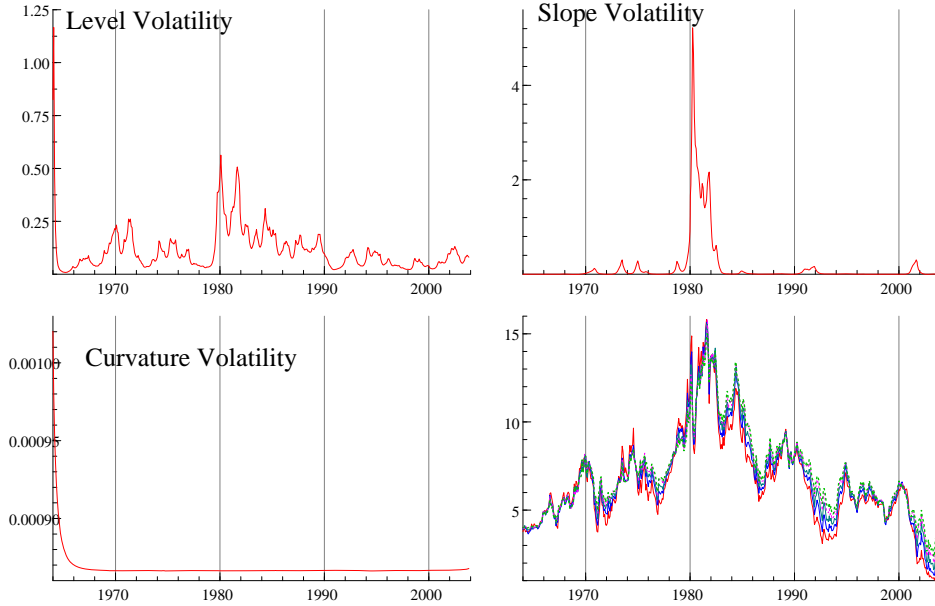
**Figure 4:** Latent Yield Factors Implied by the SVNS Model. Top left: level factor  $f_L$ ; Top right: slope factor  $f_S$ ; Bottom left: curvature factor  $f_C$ ; Bottom right: underlying yields Jan, 1964 - Dec, 2003.

be confirmed by the analysis of forecasting uncertainties induced by both models shown in Section 4.2.

Figure 5 depicts the corresponding volatilities  $e^{h_j}$  associated with the level, slope and curvature, respectively. The highest variations in volatility are shown for the level and slope factor. The level volatility seems to be positively related to the underlying (interest rate) level. Hence, in periods of high (low) interest rates we also observe higher (lower) volatility. Two main spikes are observable in 1980 and 1984 where overall interest volatility has been obviously high. In the same period, also the slope volatility has been extremely high. In contrast, during the remaining period, slope volatilities have been virtually zero reflecting comparably low fluctuations in yield spreads. Similarly, curvature volatility is widely constant through the sample period. While the time series patterns of level and slope volatilities are similar to those reported by Hautsch and Ou (2008) using the same data, the pattern of the curvature volatility is different. While in Hautsch and Ou slope and curvature volatility seem to be positively related, the present estimates seem to disentangle the individual components more clearly. Since both the underlying data and model are the same, these differences in results must be induced by the underlying MCMC sampling algorithm which is clearly more efficient in the present study. This finding indicates that the sampling of latent volatilities in a latent factor model is a challenging task and definitely requires computationally efficient algorithms as proposed in this paper.

**Table 2:** Estimation results for the SVNS model and MCMC diagnostics including means, standard deviations and 95% posterior density intervals. *cd*: Geweke convergence diagnostic values, *nse*: numerical standard errors, *rne*: relative numerical efficiency calculated based on lags covering  $q\%$  of the retained sample, with  $q = [4, 8, 15]$ .

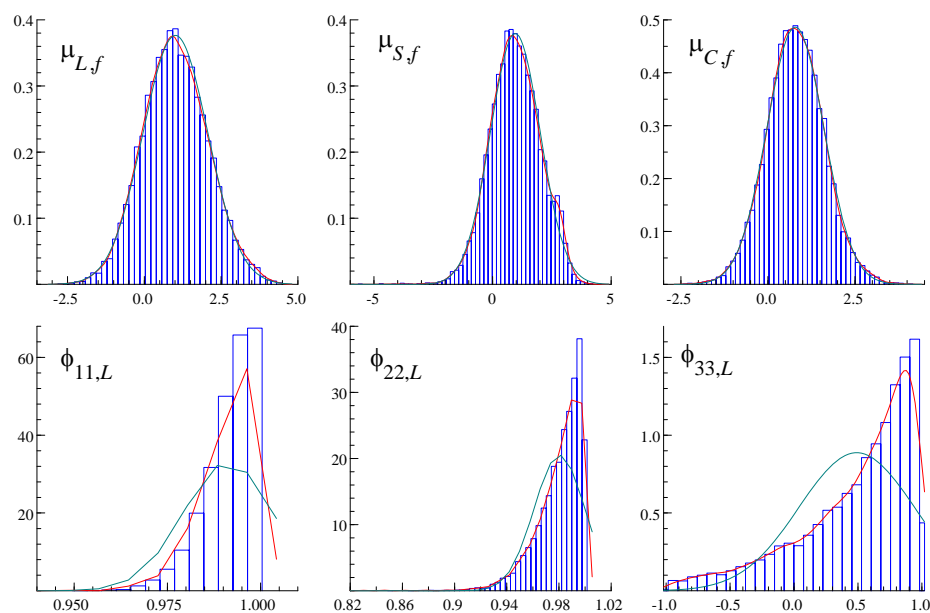
<i>Variable</i>	$\mu_{f,L}$	$\mu_{f,S}$	$\mu_{f,C}$	$\phi_{f,L}$	$\phi_{f,S}$	$\phi_{f,C}$	$\mu_{h,L}$	$\mu_{h,S}$	$\mu_{h,C}$	$\phi_{h,L}$	$\phi_{h,S}$	$\phi_{h,C}$
<i>mean</i>	1.0129	0.9596	0.7895	0.9913	0.9803	0.4900	0.2925	0.1026	0.0295	0.8983	0.9678	0.7236
<i>std</i>	1.0595	1.0515	0.8204	0.0120	0.0194	0.4498	0.0136	0.0170	0.0004	0.0118	0.0099	0.0197
<i>HPIV<sub>lo</sub></i>	-0.6911	-0.7387	-0.5559	0.9781	0.9482	-0.4617	0.2743	0.0866	0.0291	0.8852	0.9641	0.7001
<i>HPIV<sub>up</sub></i>	2.8055	2.7606	2.1339	0.9994	0.9987	0.9689	0.3155	0.1368	0.0300	0.9097	0.9712	0.7545
<i>cd</i>	0.6067	0.9823	-1.2593	0.5208	-0.0749	-0.0156	1.6743	-0.2232	0.0502	0.7655	0.7489	0.1170
<i>nse</i>	0.0106	0.0105	0.0082	0.0001	0.0002	0.0045	0.0001	0.0002	0.0000	0.0001	0.0001	0.0002
<i>nse<sub>.04</sub></i>	0.0100	0.0090	0.0156	0.0001	0.0002	0.0058	0.0002	0.0002	0.0000	0.0001	0.0001	0.0003
<i>nse<sub>.08</sub></i>	0.0088	0.0078	0.0143	0.0001	0.0002	0.0055	0.0002	0.0002	0.0000	0.0001	0.0001	0.0003
<i>nse<sub>.15</sub></i>	0.0081	0.0063	0.0102	0.0001	0.0001	0.0048	0.0002	0.0002	0.0000	0.0001	0.0001	0.0002
<i>rne<sub>.04</sub></i>	0.8583	1.1638	0.2959	1.1630	1.0442	0.5864	0.6956	0.5225	0.7630	0.8663	1.1117	0.5422
<i>rne<sub>.08</sub></i>	1.0318	1.1777	0.3016	1.2972	1.0528	0.5242	0.7619	0.5481	0.7355	0.8506	1.1457	0.5608
<i>rne<sub>.15</sub></i>	1.7384	1.2998	0.3096	1.4081	1.1060	0.5084	0.6369	0.8704	0.7289	0.7668	1.1713	0.7677
	$\sigma_{1,\varepsilon}^2$	$\sigma_{2,\varepsilon}^2$	$\sigma_{3,\varepsilon}^2$	$\sigma_{4,\varepsilon}^2$	$\sigma_{5,\varepsilon}^2$	$\sigma_{1,h}^2$	$\sigma_{2,h}^2$	$\sigma_{3,h}^2$	$\lambda$			
<i>mean</i>	0.0399	0.0217	0.0352	0.0336	0.1708	0.2049	0.3166	0.0065	0.0674			
<i>std</i>	0.0061	0.0042	0.0128	0.0122	0.0541	0.0248	0.0265	0.0012	0.0073			
<i>HPIV<sub>lo</sub></i>	0.0323	0.0171	0.0223	0.0233	0.1196	0.1640	0.2720	0.0056	0.0554			
<i>HPIV<sub>up</sub></i>	0.0512	0.0291	0.0611	0.0594	0.2886	0.2435	0.3554	0.0089	0.0788			
<i>cd</i>	-0.4623	-0.7554	-1.3177	-1.3757	-1.3802	-1.3615	0.3477	0.0153	0.3391			
<i>nse</i>	0.0001	0.0000	0.0001	0.0001	0.0005	0.0002	0.0003	0.0000	0.0001			
<i>nse<sub>.04</sub></i>	0.0001	0.0000	0.0001	0.0001	0.0006	0.0005	0.0005	0.0000	0.0001			
<i>nse<sub>.08</sub></i>	0.0001	0.0000	0.0001	0.0001	0.0007	0.0005	0.0004	0.0000	0.0001			
<i>nse<sub>.15</sub></i>	0.0001	0.0000	0.0001	0.0001	0.0008	0.0006	0.0004	0.0000	0.0000			
<i>rne<sub>.04</sub></i>	0.3741	1.4531	1.2850	1.1980	1.2128	0.4290	0.3241	0.4286	1.1188			
<i>rne<sub>.08</sub></i>	0.3779	1.2829	1.2675	1.3299	1.3379	0.4216	0.3469	0.3878	1.3171			
<i>rne<sub>.15</sub></i>	0.3666	1.1391	1.1942	1.3615	1.7101	0.2863	0.6445	0.4111	1.5010			



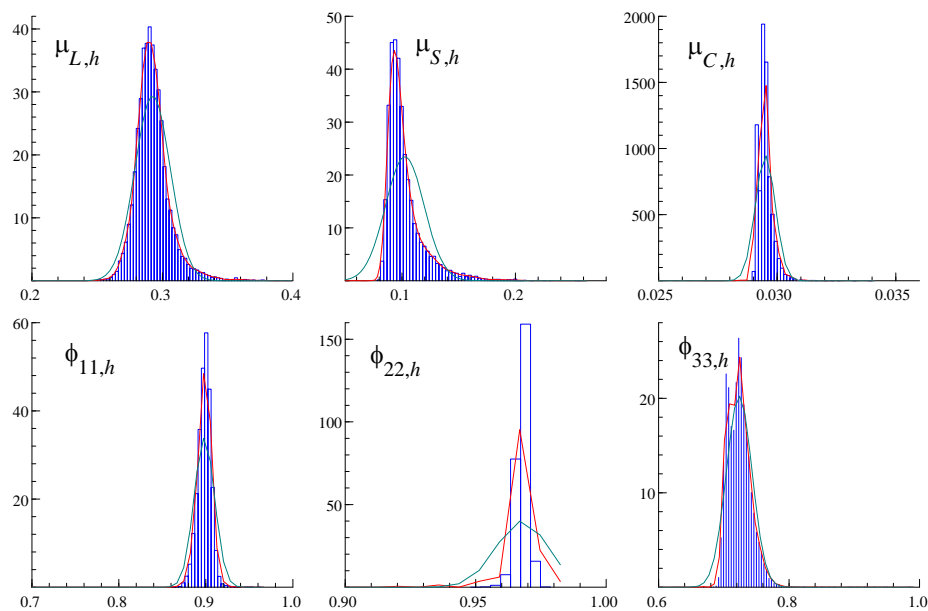
**Figure 5:** Latent Volatilities Implied by the SVNS Model. Top left: level volatility  $e^{h_L}$ ; Top right: slope volatility  $e^{h_S}$ ; Bottom left: curvature volatility  $e^{h_C}$ ; Bottom right: underlying yields Jan, 1964-Dec, 2003.

Figures 6 to 8 show the posterior distributions of the individual parameters. In most cases, the distributions strongly deviate from corresponding normal approximations which would be used asymptotically. The posterior distribution of  $\lambda$  is simulated using the Griddy-Gibbs sampler. Interestingly,  $\lambda$  is quite uniformly distributed covering a relatively small range. This result confirms the argument in Diebold and Li (2006) that factor loadings  $\Lambda$  are not very sensitive to different values of  $\lambda$ . Therefore, it is reasonable to fix  $\lambda$  to values around 0.07 as in Diebold and Li (2006), Diebold et al. (2006) or Yu and Zivot (2008). Indeed, the posterior mean of the Griddy-Gibbs sample is 0.0674, which implies the curvature loading  $\frac{1-e^{-\lambda\tau_i}}{\lambda\tau_i} - e^{-\lambda\tau_i}$  to be maximized for a maturity of  $\tau_i = 26.6$  months. Plots of the underlying MCMC draws as well as of corresponding correlograms for all parameters are given in Appendix C. It is shown that in all cases the correlations between the MCMC draws are small indicating a high efficiency of the MCMC algorithm.

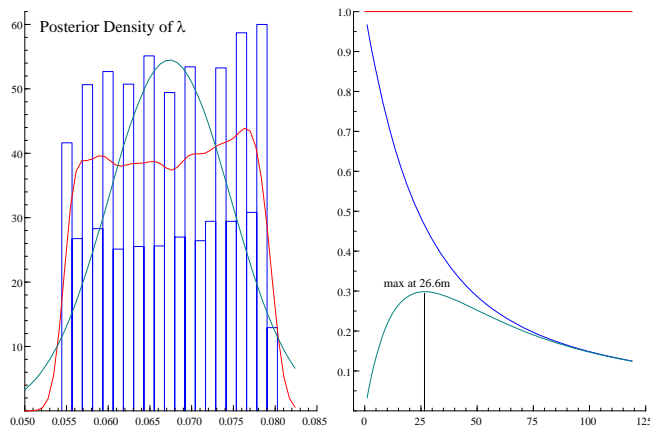
Table 3 reports the posterior predictive p-values for both the DLNS model and the SVNS model. Both models fit the data well. Neither  $E(\text{Skew} | y_t^{obs})$  nor  $E(\text{Kurt} | y_t^{obs})$  fall into the tails of the posterior predictive distribution  $E(\text{Kurt} | y_t^{pre})$  and  $E(\text{Skew} | y_t^{pre})$ . Hence, both models are able to replicate the main characteristics of the observed data. Nevertheless, the SVNS model yields ultimately larger ppp-values. For instance, the DLNS model yields  $E(\text{Skew} | y_t^{obs}) = 0.085$  and a ppp-value of 0.435 for maturities of 60 months. This indicates that 43.5% of the artificial data set generated from the



**Figure 6:** Simulated Posterior Densities based on the SVNS model: from left to right in the top panels:  $\mu_{L,f}$ ,  $\mu_{S,f}$ ,  $\mu_{C,f}$ ; from left to right in the bottom panels:  $\phi_{L,f}$ ,  $\phi_{S,f}$ ,  $\phi_{C,f}$ .



**Figure 7:** Simulated Posterior Densities based on the SVNS model: from left to right in the top panels:  $\mu_{L,h}$ ,  $\mu_{S,h}$ ,  $\mu_{C,h}$ ; from left to right in the bottom panels:  $\phi_{L,h}$ ,  $\phi_{S,h}$ ,  $\phi_{C,h}$ .



**Figure 8:** Left: Simulated Posterior Densities of  $\lambda$  based on the SVNS model. Right: calculated factor loadings using  $\lambda = 0.0674$  with the curvature loading maximized at  $\tau_i = 26.6$  months.

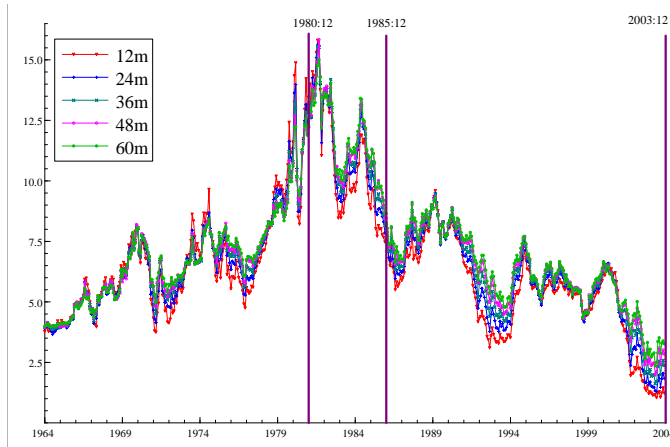
**Table 3:** Posterior Predictive p-values

$m$	<i>DLNS</i>				<i>SVNS</i>			
	<i>skewness</i>		<i>kurtosis</i>		<i>skewness</i>		<i>kurtosis</i>	
	$E(S   y_t^o)$	$p$	$E(K   y_t^o)$	$p$	$E(S   y_t^o)$	$p$	$E(K   y_t^o)$	$p$
12	0.015	0.894	2.696	0.930	0.082	0.993	3.265	0.933
24	0.022	0.839	2.689	0.935	0.104	0.993	3.109	0.931
36	0.037	0.740	2.690	0.931	0.099	0.993	3.019	0.930
48	0.059	0.585	2.695	0.924	0.092	0.993	2.953	0.926
60	0.085	0.435	2.696	0.929	0.090	0.993	2.884	0.928

DLNS model exhibit a greater degree of skewness than the actual data. Conversely, the corresponding p-values for the SVNS model are virtually one. Moreover, it turns out that the values of  $E(Kurt | y_t^{obs})$  implied by the SVNS model are larger implying that the latter specification tends to generate data with a greater degree of kurtosis. This is obviously driven by the stochastic volatility in the yield factors.

## 4.2 Evaluating the Forecasting Uncertainty

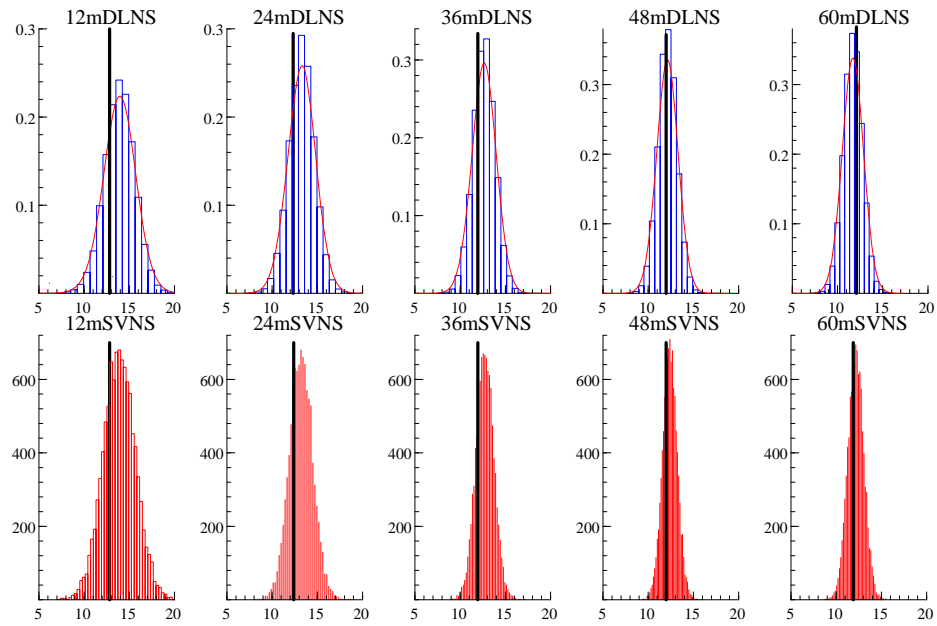
Systematic evaluations of the models' out-of-sample forecasting power is beyond the scope of this paper. Nevertheless, we aim at illustrating the importance of accounting for time-varying volatilities when forecasting densities are considered. This is performed by selecting three illustrative dates for which we produce one-step-ahead out-of-sample



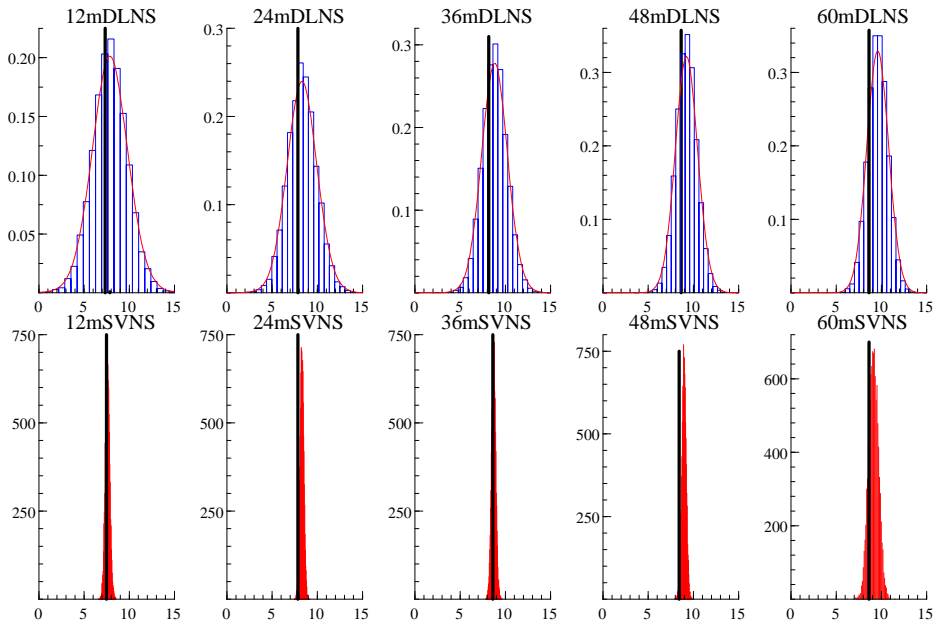
**Figure 9:** Forecast periods: December 1980, December 1985, December 2003.

forecasts. The underlying forecasting procedure considering parameter uncertainty and error uncertainty is illustrated in the Appendix. As representative dates we select December 1980, December 1985 and December 2003. Figure 9 illustrates that these dates are associated with different volatility regimes. Particularly in December 1980, volatility (and interest rates themselves) have been high. For all three periods, we compute one-step-ahead forecasts given the information set available at the previous month. The forecasts are produced based on parameter estimates employing the corresponding history of yields prevailing *before* each date. Sampling from the corresponding simulated posteriors (as shown in the Appendix) provides the full forecast density taking into account parameter uncertainty and stochastic uncertainty.

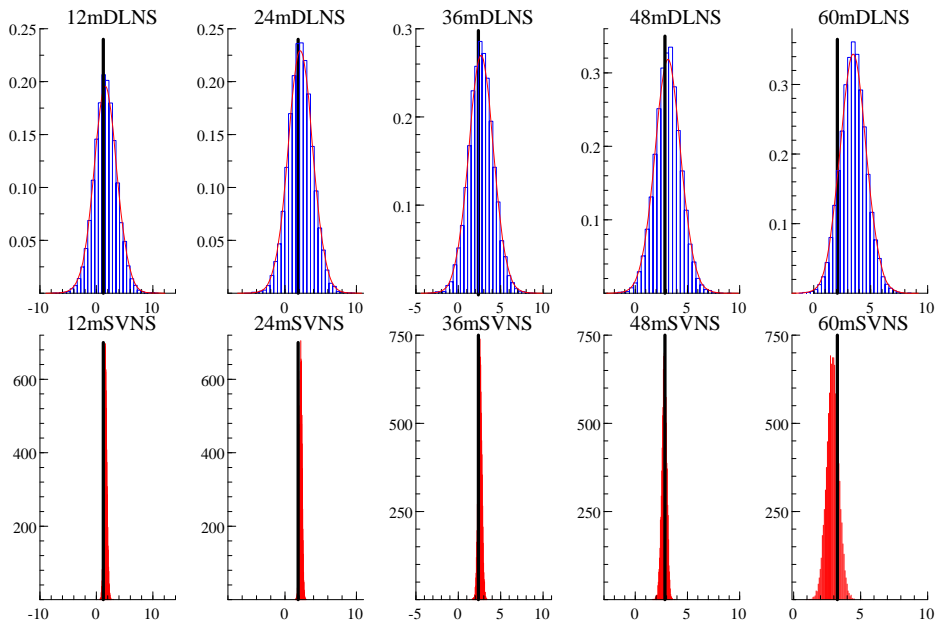
Figures 10, 11 and 12 show the corresponding density forecasts generated by the DLNS model and the SVNS model for the three dates and all maturities. Not surprisingly, the mean forecasts seem to be similar in both specifications. However, the forecast uncertainty is quite different. In all cases, the forecast uncertainty is lower in the SVNS model. This is due to the fact that in the SVNS approach, interest rate volatility – and thus an important component of parameter uncertainty and stochastic uncertainty – is explicitly taken into account. In the DLNS model, this component is missing which induces a higher stochastic uncertainty and consequently a higher forecasting uncertainty. This is most striking for December 1985 and December 2003. Both dates are associated with low volatility regimes. In these situations, the DLNS model induces a significantly higher forecasting uncertainty stemming from the fact that the ignored



**Figure 10:** Density forecasts for maturities of 12m, 24m, 36m, 48m, 60m, respectively, for December 1980. Realized values indicated by vertical lines.



**Figure 11:** Density forecasts for maturities of 12m, 24m, 36m, 48m, 60m, respectively, for December 1985. Realized values indicated by vertical lines.



**Figure 12:** Density forecasts for maturities of 12m, 24m, 36m, 48m, 60m, respectively, for December 2003. Realized values indicated by vertical lines.

stochastic and parameter uncertainty induced by high-volatility-regimes ‘spread out’ to low-volatility-regimes. The density plots show that this effect is quite distinct supporting the importance of explicitly accounting for underlying stochastic volatility. Only in December 1980, the forecasting densities are quite similar. This is due to the fact that in such a high-volatility-period both models face higher stochastic and parameter uncertainties.

## 5 Conclusions

We propose MCMC-based Bayesian inference for a so-called Stochastic Volatility Nelson-Siegel (SVNS) model. The SVNS model has been introduced by Hautsch and Ou (2008) and extends a dynamic version of the original Nelson and Siegel (1987) model as proposed by Diebold and Li (2006) to allow for stochastic volatility in the underlying yield factors. This framework allows to capture yield curve volatility in a flexible but still parsimonious way.

The model can be re-presented in a state space form with the factors in the transition equation revealing stochastic volatility and the volatility components themselves following latent dynamic processes. Our results show that the filtering of latent volatilities of latent (dynamic) variables is not straightforward and requires a high sampling precision. Therefore, we propose estimating the model using a MCMC algorithm building on the Kalman filter together with a simulation smoother. This allows to sample



the unknown parameters as well as factors block-wise making the algorithm computationally quite efficient.

Our estimation results provide evidence for distinct time-varying stochastic volatilities in yield factors. In particular, the level and slope volatilities contribute most to the volatility of the yield curve. We show that the incorporation of stochastic volatility captures an important part of stochastic uncertainty and parameter uncertainty. Analyzing the model's goodness-of-fit using posterior predictive p-values we show that the inclusion of stochastic volatility improves the model's explanatory power. Moreover, based on several illustrative examples we find that the SVNS model produces a significantly smaller forecasting uncertainty compared to a specification where variances are assumed to be constant. Actually, it turns out that ignored stochastic volatility substantially increases parameter uncertainty resulting in forecasting densities with a significantly higher variance.

Though the model is specifically designed for the modelling of yield curves it can be seen as a member of a more general class of multivariate dynamic latent factor models revealing (multivariate) stochastic volatility. Such approaches might be attractive not only for term structure dynamics but in all applications where high-dimensional heteroscedastic systems have to be modelled and forecasted. Since the proposed MCMC approach is easily adapted to alternative specifications it provides a convenient framework for statistical inference for this class of models.

## References

- [1] Ardia, D., Hoogerheide, L.F. and van Dijk, H.K. (2009). Adaptive Mixture of Student-t Distributions as a Flexible Candidate Distribution for Efficient Simulation: The R Package AdMit. *Journal of Statistical Software*, 29 (1), 1-32.
- [2] Björk, T. and Christensen, B.J. (1999). Interest Rate Dynamics and Consistent Forward Rate Curves. *Mathematical Finance*, 323-48.
- [3] Chib, S. and Greenberg, E. (1995). Understanding the Metropolis-Hastings Algorithm. *American Statistician*, 49, 327-35.
- [4] Chib, S. and Greenberg, E. (1996). Markov Chain Monte Carlo Simulation Methods in Econometrics. *Econometric Theory*, 12, 409-31.
- [5] Chib, S., Nardari, F. and Shephard, N. (2002). Markov chain Monte Carlo methods for stochastic volatility models. *Journal of Econometrics*, 108(2), 281-316.
- [6] Cochrane, J.H. and Piazzesi, M. (2005). Bond Risk Premia. *American Economic Review*, 95(1), 138-60.
- [7] Cox, J.C., Ingersoll, J.E., Jr. and Ross, S.A. (1985). A Theory of the Term Structure of Interest Rates. *Econometrica*, 53(2), 385-407.
- [8] Dai, Q. and Singleton, K.J. (2000). Specification Analysis of Affine Term Structure Models. *Journal of Finance*, 55, 1943-78.
- [9] de Jong, P. and Shephard, N. (1995). The Simulation Smoother for Time Series Models. *Biometrika*, 82(2), 339-50.
- [10] Diebold, F.X., Piazzesi, M. and Rudebusch, G.D. (2005). Modeling Bond Yields in Finance and Macroeconomics. *American Economic Review*, 95(2), 415-20.
- [11] Diebold, F.X. and Li, C. (2006). Forecasting the term structure of government bond yields. *Journal of Econometrics*, 130(2), 337-64.
- [12] Diebold, F.X., Rudebusch, G.D. and Aruoba, B.S. (2006). The macroeconomy and the yield curve: a dynamic latent factor approach. *Journal of Econometrics*, 131, 309-38.
- [13] Duffee, G. and Kan, R. (1996). A Yield-Factor Model of Interest Rates. *Mathematical Finance*, 6, 379-406.

- [14] Engle, R.F. and Ng, V.K. (1993). Time-Varying Volatility and the Dynamic Behavior of the Term Structure. *Journal of Money, Credit and Banking*, 25(3, Part 1), 336-49.
- [15] Engle, R.F., Ng, V.K. and Rothschild, M. (1990). Asset pricing with a factor-arch covariance structure : Empirical estimates for treasury bills. *Journal of Econometrics*, 45(1-2), 213-37.
- [16] Fama, E.F. and Bliss, R.R. (1987). The Information in Long-Maturity Forward Rates. *The American Economic Review*, 77(4), 680-92.
- [17] Fuller, W.A. (1996). *Introduction to statistical time series*. 2nd ed. New York: John Wiley.
- [18] Gelman, A. and Meng, X. (1996). Model Checking and Model Improvement. In Gilks, W.R. and Spiegelhalter, D.J. (eds.), *Markov chain Monte Carlo in practice*, Chapman & Hall/CRC.
- [19] Geweke, J. (1992). Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments *Bayesian Statistics* 4, 169-93.
- [20] Hautsch, N. and Ou, Y. (2008). Yield Curve Factors, Term Structure Volatility, and Bond Risk Premia. Discussion Paper DP2008-053, Collaborative Research Center 649, Humboldt-Universität zu Berlin, Germany.
- [21] Hull, J. and White, A. (1990). Pricing Interest-Rate-Derivative Securities. *The Review of Financial Studies*, 3(4), 573-92.
- [22] Kim, S., Shephard, N. and Chib, S. (1998). Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models. *The Review of Economic Studies*, 65(3), 361-93.
- [23] Koop, G. (2003). *Bayesian econometrics*. Chichester: Wiley.
- [24] Koopman, S.J., Shephard, N. and Doornik, J.A. (1999). Statistical algorithms for models in state space form using SsfPack 2.2. *Econometrics Journal*, 2(1), 113-66.
- [25] Koopman, S.J., Shephard, N. and Doornik, J.A. (2002). SVPack 2.1: Ox software for volatility models.
- [26] Koopman, S.J., Mallee, M.I.P. and van der Wel, M. (2010). Analyzing the Term Structure of Interest Rates using the Dynamic Nelson-Siegel Model with Time-Varying Parameters. *Journal of Business and Economic Statistics*.

- [27] Meng, X. (1994). Posterior Predictive p-Values. *The Annals of Statistics*, 22(3), 1142-60.
- [28] Mönch, E. (2006). Term Structure Surprises: The Predictive Content of Curvature, Level, and Slope. Humboldt University Berlin Working Paper.
- [29] Nelson, C.R. and Siegel, A.F. (1987). Parsimonious Modeling of Yield Curves. *Journal of Business*, University of Chicago Press, 60(4), 473-89.
- [30] Piazzesi, M. (2003). Affine term structure models. In Ait-Sahalia, Y. and Hansen, L. (eds.), *Handbook of Financial Econometrics*, 2009 ed. North-Holland.
- [31] Rudebusch, G.D. and Wu, T. (2008). A Macro-Finance Model of the Term Structure, Monetary Policy and the Economy. *The Economic Journal*, 118(530), 906-26.
- [32] Vasicek, O. (1977). An equilibrium characterization of the term structure. *Journal of Financial Economics*, 5(2), 177-88.
- [33] Yu, W. and Zivot, E.W. (2008). Forecasting the Term Structures of Treasury and Corporate Yields: Dynamic Nelson-Siegel Models Evaluation. SSRN Working Paper.

## A Derivations of Conditional Posterior Distributions

In the following, we derive the corresponding conditional posterior distributions and show how to sample from them.

1. Sample  $H_\varepsilon$  from a Wishart conditional posterior,  $H_\varepsilon \sim W(H_\varepsilon | \bar{A}_\varepsilon, \bar{\nu}_\varepsilon)$ , where

$$\bar{\nu}_\varepsilon = T + \underline{\nu}_\varepsilon,$$

$$\bar{A}_\varepsilon^{-1} = \underline{A}_\varepsilon^{-1} + \sum_{t=1}^T (y_t - \Lambda f_t)(y_t - \Lambda f_t)',$$

and the hyper-parameters to be selected as  $\underline{\nu}_\varepsilon = 0.001$  and  $\underline{A}_\varepsilon^{-1} = 0_{N \times N}$  associated with non-informative priors. Then  $\Sigma_\varepsilon \Sigma_\varepsilon'$  can be calculated from the inverse of  $H_\varepsilon$ .

2. To sample  $\lambda$ , we use a Griddy-Gibbs sampler posterior based on 30 grid points. Hence, with  $y$  collecting the data and  $f$  and  $h$  collecting the time series of all yield factors and factor volatilities, respectively, we have

$$\begin{aligned} p(\lambda | y, f, h) &\propto \prod_{t=1}^T p(y_t | \theta, f_t, h_t) p(\lambda) \\ &\propto \prod_{t=1}^T p(y_t | \theta, f_t, h_t, H_\varepsilon) \frac{1}{\underline{b}_\lambda - \underline{a}_\lambda} I_{(\underline{a}_\lambda, \underline{b}_\lambda)}(\lambda) \\ &\propto \frac{1}{\underline{b}_\lambda - \underline{a}_\lambda} \frac{|H_\varepsilon|^{\frac{T}{2}}}{(2\pi)^{\frac{T}{2}}} \exp \left\{ -\frac{1}{2} \sum_{t=1}^T (y_t - \Lambda f_t)' H_\varepsilon (y_t - \Lambda f_t) \right\} I_{(\underline{a}_\lambda, \underline{b}_\lambda)}(\lambda), \end{aligned}$$

where  $I_{(\underline{a}_\lambda, \underline{b}_\lambda)}(\lambda)$  denotes an indicator function, and  $\lambda \in (\underline{a}_\lambda, \underline{b}_\lambda)$ . The hyper-parameters  $\underline{b}_\lambda$  and  $\underline{a}_\lambda$  are chosen in line with previous literature. For instance, in Diebold and Li (2006),  $\lambda = 0.0609$  maximizes the curvature factor loading at a 30-month maturity. Other studies find similar results showing that  $\lambda$  is most likely in a range between 0.06 and 0.07. Accordingly, we choose  $\underline{b}_\lambda = 0.08$  and  $\underline{a}_\lambda = 0.055$ .

3. Sample  $[\mu_f, \phi_f]$  and  $f$ , using the sampled  $\lambda$ .
  - (a) Fitting eq. (4) and (5) into a state-space form, we can sample  $f$  using the Kalman filter with a simulation smoother. With  $f_t = (f_{1,t}, f_{2,t}, f_{3,t})'$  and

$y_t = \{y_t(\tau_1), y_t(\tau_2), \dots, y_t(\tau_N)\}'$ , the time-varying state-space form is

$$\begin{pmatrix} f_{t+1} \\ y_t \end{pmatrix} = \delta_t + \Phi_t f_t + u_t \quad u_t \stackrel{i.i.d.}{\sim} N(0, \Omega_t), \quad t = 1, \dots, T.$$

Initial conditions are

$$\begin{aligned} \delta_t &= \begin{pmatrix} d_t \\ c_t \end{pmatrix} = \begin{pmatrix} \mu_f - \mu_f \phi_f \\ 0_{N \times 1} \end{pmatrix} : (3 + N) \times 1, \\ \Phi_t &= \begin{pmatrix} T_t \\ Z_t \end{pmatrix} = \begin{pmatrix} \text{diag}(\phi_{11,f}, \phi_{22,f}, \phi_{33,f}) \\ \Lambda \end{pmatrix} : (3 + N) \times 3, \\ \Omega_t &= \begin{pmatrix} \Sigma_{\eta,t} \Sigma'_{\eta,t} & 0 \\ 0 & \Sigma_{\varepsilon} \Sigma'_{\varepsilon} \end{pmatrix} : (3 + N) \times (3 + N). \end{aligned}$$

As shown in in Kim et al. (1998), we obtain

$$\begin{aligned} \nu_t &= y_t - c_t - Z_t f_t : (N \times 1), \\ F_t &= Z_t P_t Z_t' + \Sigma_{\varepsilon} \Sigma'_{\varepsilon} : (N \times N), \\ K_t &= (T_t P_t Z_t') F_t^{-1} : (3 \times N), \\ f_{t+1} &= d_t + T_t f_t + K_t \nu_t : (3 \times 1) \\ P_{t+1} &= T_t P_t T_t' + \Sigma_{\eta,t} \Sigma'_{\eta,t} - K_t F_t K_t' : (3 \times 3), \end{aligned}$$

where  $a'_0 = (\mu_f)'$  using the previous draws from the Gibbs sampler and  $P = I_3 \times 10^6$  with  $I_p$  denoting a  $p$ -dimensional identity matrix. Then, the Kalman filter returns the vector  $\{\nu_t, K_t, F_t^{-1}\}$  and the simulation smoother (de Jong and Shephard, 1995) can be used to get draws  $\{f_1, f_2, \dots, f_T\} | y, \theta$ , where  $\theta$  collects all parameters in the SVNS model. Setting  $f_T = 0$  and  $N_T = 0_{3 \times 3}$ , and define

$$D_t = F_t^{-1} + K_t' N_t K_t : (N \times N),$$

$$n_t = F_t^{-1} \nu_t - K_t' f_t,$$

we run for  $t = T, \dots, 1$ ,

$$C_t = \Sigma_\varepsilon \Sigma'_\varepsilon - (\Sigma_\varepsilon \Sigma'_\varepsilon) D_t (\Sigma_\varepsilon \Sigma'_\varepsilon),$$

$$k_t \sim N(0, C_t),$$

$$f_{t-1} = Z'_t F_t^{-1} \nu_t + (T_t - K_t Z_t)' f_t - V'_t C_t^{-1} k_t,$$

$$V_t = (\Sigma_\varepsilon \Sigma'_\varepsilon) (D_t Z_t - K'_t N_t T_t),$$

$$N_{t-1} = Z'_t F_t^{-1} Z_t + (T_t - K_t Z_t)' N_t (T_t - K_t Z_t) + V'_t C_t^{-1} V_t.$$

Then,  $y_t - (\Sigma_\varepsilon \Sigma'_\varepsilon) n_t - k_t$  is a draw from  $Z_t f_t \mid y, \theta, Z_{t+1} f_{t+1}, \dots, Z_T f_T$ . The smoothed factors  $\{f_1, f_2, \dots, f_T\} \mid y, \theta$  can be obtained using the simulation smoother described above. Note that sampling  $\{f_1, f_2, \dots, f_T\}, \theta \mid y$  directly is not possible because of degeneracies (see Koopman et al., 1999).

(b) Sampling  $[\mu_f, \phi_f]$ . The likelihood is given by

$$\begin{aligned} p(y, f, h \mid \theta) &= \prod_{t=1}^T \prod_{j=1}^3 f_N(y_t \mid \Lambda f_t, \Sigma_\varepsilon \Sigma'_\varepsilon) \\ &\quad \times f_N(f_{j,t} \mid \mu_{j,f} + \phi_{j,f}(f_{j,t-1} - \mu_{j,f}), \exp(h_{j,t})) \\ &\quad \times f_N(h_{j,t} \mid \mu_{j,h} + \phi_{j,h}(h_{j,t-1} - \mu_{j,h}), \sigma_j^2). \end{aligned}$$

With a prior  $\underline{\mu}_{j,f} = 0$  and  $\underline{V}_{j,f} = 10^4$ , the posterior of  $\mu_{j,f}$  is derived as  $\mu_{j,f} \sim N(\underline{\mu}_{j,f} \mid \bar{\mu}_{j,f}, \bar{V}_{j,f})$ , where

$$\bar{V}_{j,f} = \left[ \sum_{t=1}^T \frac{(1 - \phi_{j,f})^2}{\exp(h_{j,t})} + \frac{1}{\underline{V}_{j,f}} \right]^{-1}$$

and

$$\bar{\mu}_{j,f} = \bar{V}_{j,f} \left[ \sum_{t=1}^T \frac{(1 - \phi_{j,f})(f_{j,t} - \phi_{j,f} f_{j,t-1})}{\exp(h_{j,t})} + \frac{\underline{\mu}_{j,f}}{\underline{V}_{j,f}} \right].$$

Similarly, with a truncated normal prior for  $\phi_{j,f}$ , the posterior conditional can be derived as  $\phi_{j,f} \sim TN(\phi_{j,f} \mid \bar{\mu}_{j,f_\phi}, \bar{V}_{j,f_\phi}) \mathbf{1}(|\phi_{j,f}| < 1)$  where

$$\bar{V}_{j,f_\phi} = \left[ \sum_{t=1}^T \frac{(f_{j,t-1} - \mu_{j,f})^2}{\exp(h_{j,t})} + \frac{1}{\underline{V}_{j,f_\phi}} \right]^{-1}$$

and

$$\bar{\mu}_{j,f_\phi} = \bar{V}_{j,f_\phi} \left[ \sum_{t=1}^T \frac{(f_{j,t} - \mu_{j,f})(f_{j,t-1} - \mu_{j,f})}{\exp(h_{j,t})} + \frac{\mu_{j,f_\phi}}{\bar{V}_{j,f_\phi}} \right].$$

4. Using the draws of  $f$ , we can calculate  $f^*$  from

$$f_{j,t} - \mu_{j,f} = \phi_{j,f} (f_{j,t-1} - \mu_{j,f}) + \exp\left(\frac{h_{j,t}}{2}\right) \zeta_{j,t}, \quad j = 1, 2, 3,$$

and  $f_{j,t}^* = \ln \left\{ [f_{j,t} - \mu_{j,f} - \phi_{j,f} (f_{j,t-1} - \mu_{j,f})]^2 + c \right\}$ , where  $c = 0.001$ . Then, we estimate  $[\phi_{j,h}, \sigma_j, \mu_{j,h}]$  and  $h_j$  following Kim et al. (1998):

(a) Sample  $s$  from  $s|f_j^*, h_j$ , where  $s = \{s_t\}$  for  $t = 1, \dots, T$ , using Table 1 and the following probability mass function

$$\Pr(s_t = i | f_{j,t}^*, h_{j,t}) \propto q_i f_N(f_{j,t}^* | \mu_{s_t} + h_{j,t}, \nu_{s_t}^2).$$

(b) Sampling  $h_j|f_j^*, s, [\mu_{h,j}, \phi_{h,j}, \sigma_j]$  can be achieved in one block using the Kalman filter and simulation smoother. The procedures are similar to those in Step 3(a). We treat the sampled  $f_j, j = 1, 2, 3$ , as three univariate time series, running step 4 three times to estimate and extract the stochastic volatilities from the series  $f_j$ . Since  $f_j^*$  and  $h_j$  can be fitted in a state-space form, it is straight forward to apply Kalman filter and simulation smoother. Extracting the log-volatilities terms  $h_j$  is with the same manner as we draw the smoothed factor  $\{f_1, f_2, \dots, f_T\} | y, \theta$  in Step 3(a). For matrix calculus as stated in Step 3(a), the representations of  $\delta_t, \Phi_t$  and time varying  $\Omega_t$  are given as the following:

$$\begin{pmatrix} h_{j,t+1} \\ f_{j,t}^* \end{pmatrix} = \delta_t + \Phi_t h_{j,t} + u_t \quad u_t \stackrel{i.i.d.}{\sim} N(0, \Omega_t), \quad t = 1, \dots, T.$$

Under stationary initial conditions we have

$$P = \begin{pmatrix} \sigma_j^2 / (1 - \phi_{j,h}^2) \\ \mu_{j,h} \end{pmatrix},$$

$$\delta_t = \begin{pmatrix} d_t \\ c_t \end{pmatrix} = \begin{pmatrix} \mu_{j,h} - \mu_{j,h} \phi_{j,h} \\ 0 \end{pmatrix},$$



$$\Phi_t = \begin{pmatrix} T_t \\ Z_t \end{pmatrix} = \begin{pmatrix} \phi_{j,h} \\ 1 \end{pmatrix},$$

$$u_t = \begin{pmatrix} \sigma_j \epsilon_{j,t} \\ z_{j,t} \end{pmatrix},$$

where  $z_{j,t}|s_t \sim N(m_{s_t}, \nu_{s_t}^2)$ .

$$\Omega_t = \begin{pmatrix} \sigma_j^2 & 0 \\ 0 & \nu_{s_t}^2 \end{pmatrix}$$

- (c) Sampling  $[\mu_{j,h}, \phi_{j,h}, \sigma_j]$  is achieved using the Metropolis-Hasting algorithm (Chib and Greenberg, 1995). Following Chib et al. (2002), we use a multivariate- $t$  density with  $\nu$  degrees of freedom as the proposal candidate density. We have chosen  $\nu = 5$  to cover a fat tailed distribution. For details in step (b) and (c), see Chib et al. (2002) for details.

## B Forecasting

Once all the parameters are estimated,  $h$ -step ahead point forecasts can be obtained using the posterior means/ modes. In the Bayesian context, density forecasts are easily achievable directly using the MCMC draws. Hence,

$$\hat{f}_{t+h} = \hat{\mu}_f + \hat{\phi}_f \left( \hat{f}_t - \hat{\mu}_f \right), \quad (19)$$

$$\hat{y}_{t+h} = \hat{\Lambda} \hat{f}_{t+h}, \quad (20)$$

where  $\hat{\Lambda}$  is calculated using the draws of  $\lambda$ , denoted as  $\hat{\lambda}$ . The density forecast incorporating parameter uncertainty can be conducted as follows: Denote each retained sample from the MCMC procedure as  $draw^{(s)}$ , for  $s = 1, \dots, S$ , we can produce  $S$   $h$ -step ahead forecast samples for  $\hat{f}_{t+h}$  with each forecast sample denoted by  $\hat{f}_{t+h}^{(s)}$ . Hence,

$$\hat{f}_{t+h}^{(s)} = \hat{\mu}_f^{(s)} + \hat{\phi}_f^{(s)} \left( \hat{f}_t^{(s)} - \hat{\mu}_f^{(s)} \right),$$

and

$$\hat{y}_{t+h}^{(s)} = \hat{\Lambda}^{(s)} \hat{f}_{t+h}^{(s)}.$$

For  $S \rightarrow \infty$ , we obtain an approximation of the forecast density of  $y_{t+h}$ . In the resulting forecast density parameter uncertainty is naturally taken into account.

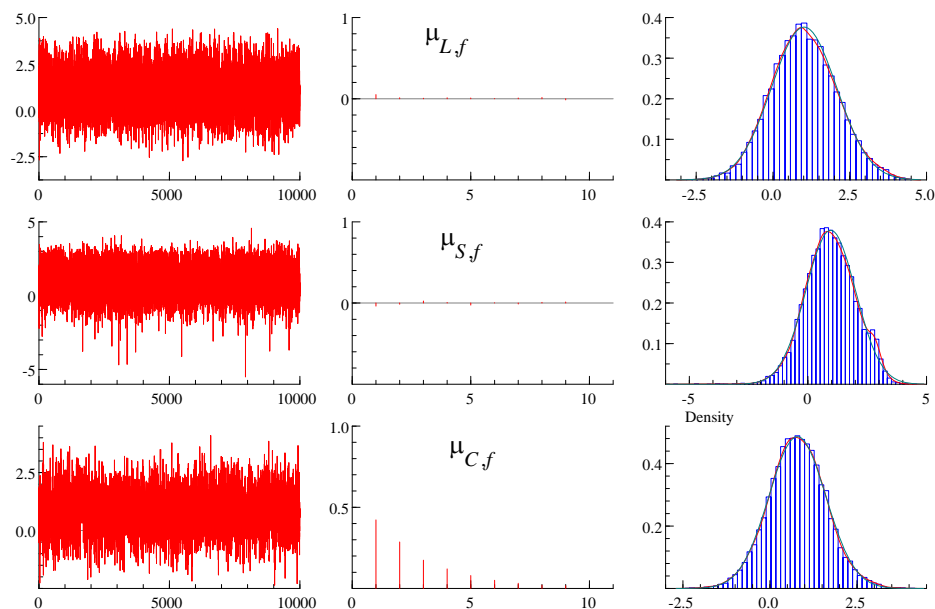
To account also for stochastic uncertainty we proceed as follows: Suppose we have a random draw of  $\Sigma_\eta^{(s)}$  and  $\Sigma_\varepsilon^{(s)}$  from the posterior conditionals, we generate a random sample  $\widehat{\eta}_{t+h}^{(s)} = \Sigma_\eta^{(s)} u_{\eta,t+h}$  and  $\widehat{\varepsilon}_{t+h}^{(s)} = \Sigma_\varepsilon^{(s)} u_{\varepsilon,t+h}$ , where  $u_{\varepsilon,t+h}$  and  $u_{\eta,t+h}$  are i.i. standard normally distributed. Then, forecasts incorporating stochastic uncertainties are

$$\widehat{f}_{t+h}^{(s)} = \widehat{\mu}_f^{(s)} + \widehat{\phi}_f^{(s)} \left( \widehat{f}_t^{(s)} - \widehat{\mu}_f^{(s)} \right) + \widehat{\eta}_{t+h}^{(s)},$$

$$\widehat{y}_{t+h}^{(s)} = \widehat{\Lambda}^{(s)} \left( \widehat{f}_{t+h}^{(s)} + \widehat{\eta}_{t+h}^{(s)} \right) + \widehat{\varepsilon}_{t+h}^{(s)},$$

where the stochastic uncertainty is induced by the errors in both the measurement equation,  $\varepsilon_{t+h}$ , and the transition equation,  $\eta_{t+h}$ . In this paper, we consider a one-step-ahead out-of-sample forecast, i.e.  $h = 1$ , taking both parameter and stochastic uncertainty into account.

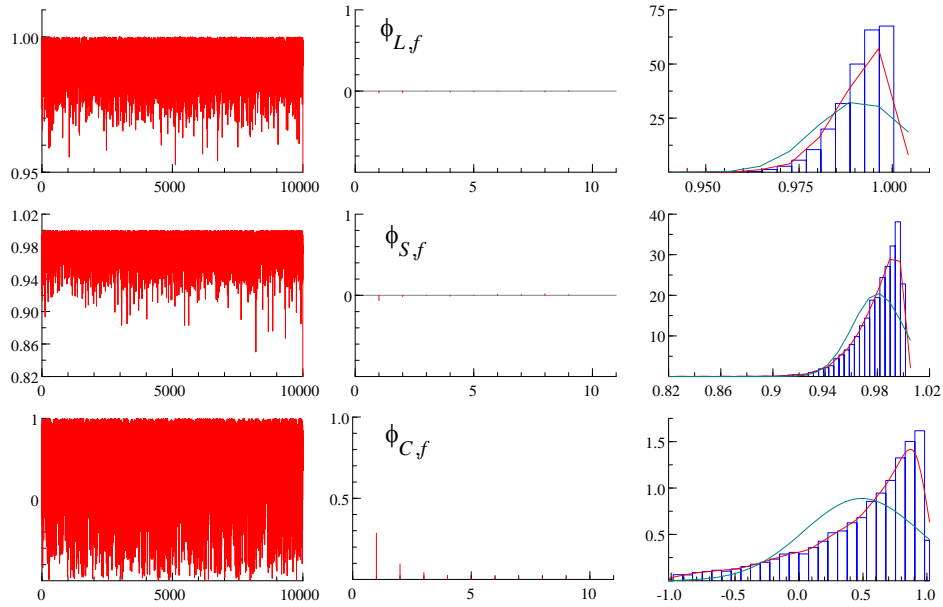
## C Tables and Figures



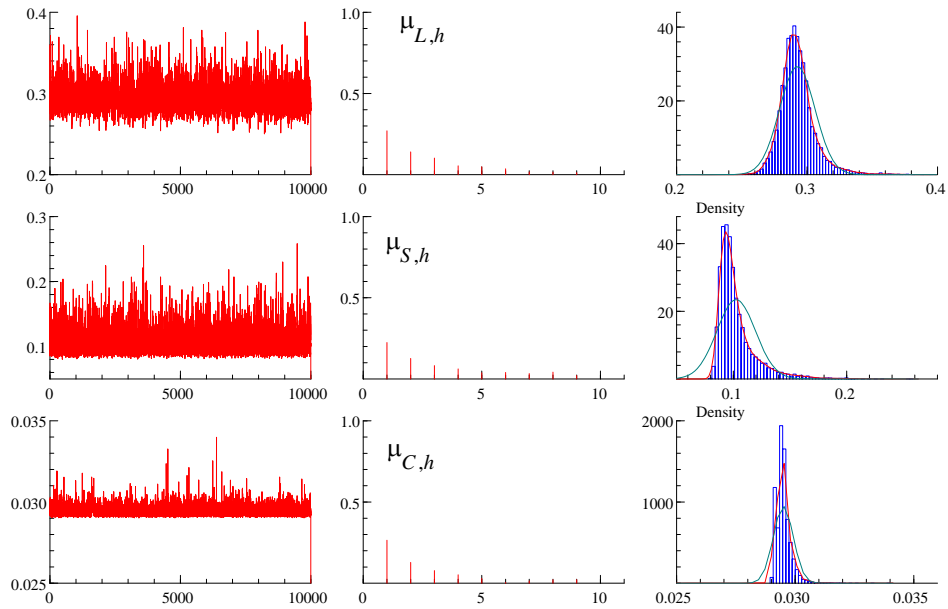
**Figure 13:** MCMC draws, Correlograms, and Posterior Densities for  $\mu_{L,f}$ ,  $\mu_{S,f}$ ,  $\mu_{C,f}$  based on the SVNS model.

**Table 4:** Estimation results for the DLNS model and MCMC diagnostics including means, standard deviations and 95% posterior density intervals. *cd*: Geweke convergence diagnostic values, *nse*: numerical standard errors, *rne*: relative numerical efficiency calculated based on lags covering  $q\%$  of the retained sample, with  $q = [4, 8, 15]$ .

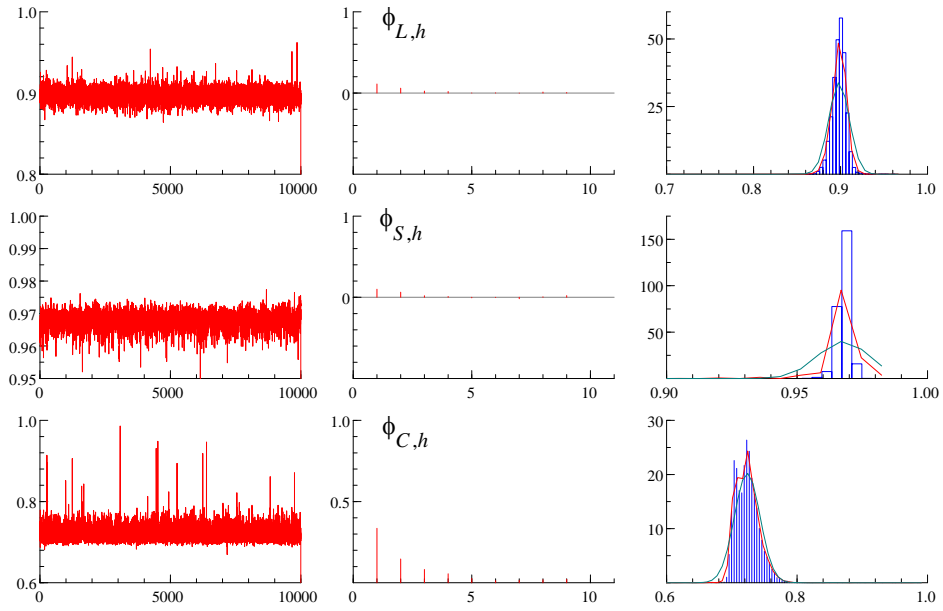
	$\mu_{f,L}$	$\mu_{f,S}$	$\mu_{f,C}$	$\phi_{f,L}$	$\phi_{f,S}$	$\phi_{f,C}$	$\lambda$	
<i>mean</i>	0.9215	0.5602	0.7113	0.9996	0.9647	0.9607	0.0673	
<i>std</i>	1.0100	0.9919	0.9615	0.0071	0.0304	0.0319	0.0075	
<i>HPIV<sub>lo</sub></i>	-0.7360	-1.0469	-0.8615	0.9990	0.9080	0.9006	0.0554	
<i>HPIV<sub>up</sub></i>	2.5784	2.2138	2.3040	1.0000	0.9975	0.9969	0.0788	
<i>cd</i>	-0.0917	0.4947	-1.6099	1.0064	-0.1886	0.5634	-0.6325	
<i>nse</i>	0.0071	0.0070	0.0068	0.0001	0.0002	0.0002	0.0001	
<i>nse<sub>.04</sub></i>	0.0075	0.0059	0.0064	0.0000	0.0002	0.0002	0.0000	
<i>nse<sub>.08</sub></i>	0.0077	0.0055	0.0065	0.0000	0.0001	0.0002	0.0000	
<i>nse<sub>.15</sub></i>	0.0074	0.0056	0.0063	0.0000	0.0001	0.0002	0.0000	
<i>rne<sub>.04</sub></i>	0.8986	1.4322	1.1204	1.0157	1.7524	1.3548	2.2903	
<i>rne<sub>.08</sub></i>	0.8603	1.6149	1.1021	1.0118	2.1974	1.6534	2.4579	
<i>rne<sub>.15</sub></i>	0.9281	1.5971	1.1683	1.0106	2.8375	1.5616	1.9414	
	$\sigma_{1,\varepsilon}^2$	$\sigma_{2,\varepsilon}^2$	$\sigma_{3,\varepsilon}^2$	$\sigma_{4,\varepsilon}^2$	$\sigma_{5,\varepsilon}^2$	$\sigma_{1,\eta}^2$	$\sigma_{2,\eta}^2$	$\sigma_{3,\eta}^2$
<i>mean</i>	0.0806	0.1346	0.1886	0.1863	0.4372	0.0723	2.6367	3.5332
<i>std</i>	0.0263	0.0110	0.0330	0.0172	0.0796	0.0047	0.5873	0.5013
<i>HPIV<sub>lo</sub></i>	0.0530	0.1198	0.1462	0.1656	0.3473	0.0647	1.9515	3.7767
<i>HPIV<sub>up</sub></i>	0.1382	0.1552	0.2534	0.2219	0.5948	0.0801	2.9000	4.4955
<i>cd</i>	1.0729	1.1482	0.9230	1.2209	1.1315	-0.7834	1.1983	-1.2735
<i>nse</i>	0.0002	0.0001	0.0002	0.0001	0.0006	0.0000	0.0042	0.0035
<i>nse<sub>.04</sub></i>	0.0002	0.0001	0.0001	0.0001	0.0006	0.0000	0.0036	0.0032
<i>nse<sub>.08</sub></i>	0.0002	0.0001	0.0001	0.0001	0.0007	0.0000	0.0036	0.0031
<i>nse<sub>.15</sub></i>	0.0002	0.0001	0.0001	0.0002	0.0007	0.0000	0.0034	0.0028
<i>rne<sub>.04</sub></i>	1.2068	1.5358	4.6890	0.8652	0.9136	0.6370	1.3262	1.2427
<i>rne<sub>.08</sub></i>	1.0198	1.5421	4.1658	0.6803	0.7065	0.5778	1.3575	1.3390
<i>rne<sub>.15</sub></i>	0.8990	1.5775	4.3326	0.6210	0.5892	0.5297	1.5189	1.6420



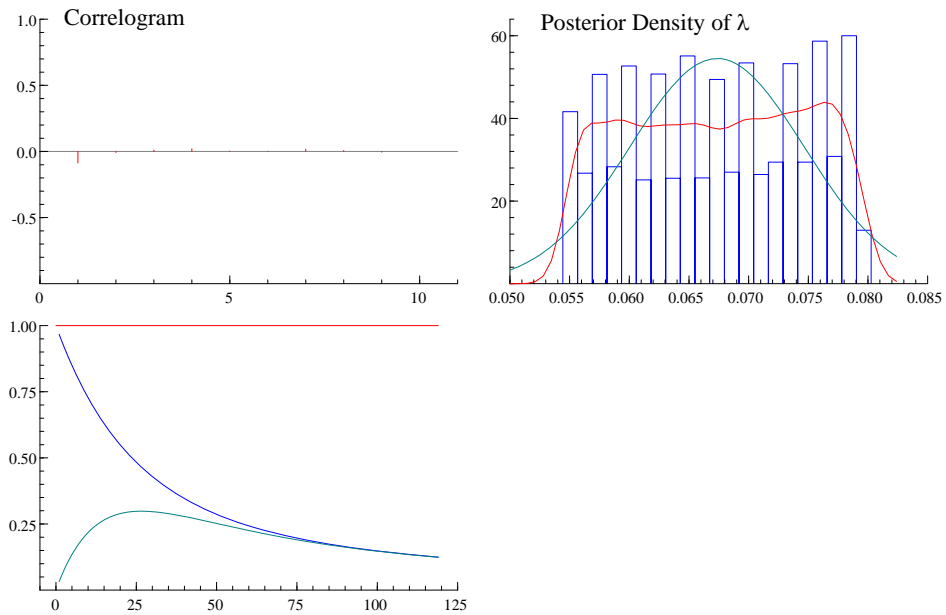
**Figure 14:** MCMC draws, Correlograms, and Posterior Densities for  $\phi_{L,f}$ ,  $\phi_{S,f}$ ,  $\phi_{C,f}$  based on the SVNS model.



**Figure 15:** MCMC draws, Correlograms, and Posterior Densities for  $\mu_{L,h}$ ,  $\mu_{S,h}$ ,  $\mu_{C,h}$  based on the SVNS model.



**Figure 16:** MCMC draws, Correlograms, and Posterior Densities for  $\phi_{L,h}$ ,  $\phi_{S,h}$ ,  $\phi_{C,h}$  based on the SVNS model.



**Figure 17:** Correlograms, and Posterior Densities for  $\lambda$  based on the SVNS model.

## SFB 649 Discussion Paper Series 2010

For a complete list of Discussion Papers published by the SFB 649, please visit <http://sfb649.wiwi.hu-berlin.de>.

- 001 "Volatility Investing with Variance Swaps" by Wolfgang Karl Härdle and Elena Silyakova, January 2010.
- 002 "Partial Linear Quantile Regression and Bootstrap Confidence Bands" by Wolfgang Karl Härdle, Ya'acov Ritov and Song Song, January 2010.
- 003 "Uniform confidence bands for pricing kernels" by Wolfgang Karl Härdle, Yarema Okhrin and Weining Wang, January 2010.
- 004 "Bayesian Inference in a Stochastic Volatility Nelson-Siegel Model" by Nikolaus Hautsch and Fuyu Yang, January 2010.