Research Article

Applied Stochastic Models in Business and Industry

Received 19 June 2015, Revised 30 November 2016, Accepted 10 February 2017 Published online in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/asmb.2246

Maximum likelihood estimation for stochastic volatility in mean models with heavy-tailed distributions

Carlos A. Abanto-Valle[a*†], Roland Langrock[b], Ming-Hui Chen[c] and Michel V. Cardoso[a]

In this article, we introduce a likelihood-based estimation method for the stochastic volatility in mean (SVM) model with scale mixtures of normal (SMN) distributions. Our estimation method is based on the fact that the powerful hidden Markov model (HMM) machinery can be applied in order to evaluate an arbitrarily accurate approximation of the likelihood of an SVM model with SMN distributions. Likelihood-based estimation of the parameters of stochastic volatility models, in general, and SVM models with SMN distributions, in particular, is usually regarded as challenging as the likelihood is a high-dimensional multiple integral. However, the HMM approximation, which is very easy to implement, makes numerical maximum of the likelihood feasible and leads to simple formulae for forecast distributions, for computing appropriately defined residuals, and for decoding, that is, estimating the volatility of the process. Copyright © 2017 John Wiley & Sons, Ltd.

Keywords: feedback effect; non-Gaussian and nonlinear state-space models; scale mixture of normal distributions; value-at-risk

1. Introduction

Over the last two decades, stochastic volatility models have proven to be useful for modeling time-varying variances, mainly in financial applications where policy makers or stockholders are constantly facing decision problems that usually depend on measures of volatility and risk. An attractive feature of the stochastic volatility model is its close relationship to financial economic theories [1] and its ability to capture the main empirical properties, that is, the stylized facts, often observed in daily series of financial returns [2].

Many empirical studies have shown strong evidence of heavy-tailed conditional mean errors in financial time series; see for example [3] and [4]. In the stochastic volatility literature, [5–7] and [8], amongst others, have provided consistent evidence that leptokurtic distributions, such as the Student's t, the generalized error or the scale mixtures of normal (SMN) distributions, which relax the normality assumption that is often made for the distribution of the returns, should be used in order to capture this feature.

Frequently, the volatility of daily stock returns has been modeled using stochastic volatility models, but the results have relied on an extensive pre-modeling of these series to avoid the problem of simultaneous estimation of the mean and variance. Koopman and Uspensky [9] introduced the stochastic volatility in mean (SVM) model to deal with this problem, incorporating the unobserved volatility as an explanatory variable in the mean equation of the returns under a normality assumption for the innovations. Abanto-Valle *et al*. [10] proposed to enhance the robustness of the specification of the innovation return in SVM models by introducing SMN distributions, referring to this generalization as the SVM–SMN class of models. This rich class contains as proper elements the SVM with normal (SVM-N), Student's t (SVM-T), and slash (SVM-S) distributions. Abanto-Valle *et al*. [10] proposed a Markov Chain Monte Carlo (MCMC) procedure for

[a] *Department of Statistics, Federal University of Rio de Janeiro, Caixa Postal 68530, CEP: 21945-970 Rio de Janeiro, Brazil*
[b] *Department of Business Administration and Economics, Bielefeld University, Postfach 10 01 31, 33501 Bielefeld, Germany*
[c] *Department of Statistics, University of Connecticut, U-4120, Storrs, CT 06269, U.S.A.*
**Correspondence to: Carlos A. Abanto-Valle, Department of Statistics, Federal University of Rio de Janeiro, Caixa Postal 68530, CEP: 21945-970 Rio de Janeiro, Brazil.*
[†] *E-mail: cabantovalle@im.ufrj.br*

Bayesian estimation of SVM–SMN models. However, the resulting MCMC algorithm has some undesirable features. In particular, the procedure is quite involved, requiring a large amount of computer-intensive simulations. In addition, the computational cost increases rapidly with the sample size.

In this paper, we apply an alternative frequentist estimation method, numerically maximizing an (virtually exact) approximation of the likelihood function, which is efficiently evaluated using recursive techniques routinely applied for hidden Markov models (HMMs). The key idea, the use of iterated numerical integration, was introduced by [11]. In the context of stochastic volatility models, it was applied by [12–15], although none of these papers explicitly makes the important link between stochastic volatility models and HMMs. The approximation to the SVM likelihood obtained through numerical integration can be made arbitrarily accurate while maintaining computational tractability, because of the powerful HMM forward algorithm becoming applicable. Further advantages of the (approximate) HMM representation of SVM–SMN models are that simple explicit formulae exist for the residuals and the forecast distributions and that estimates of the latent log-volatility process can be obtained by using the Viterbi algorithm.

Related studies [16–19] discuss other HMM-based approaches to modelling financial time series. In references [16–18] and unlike within our approach, the (log-)volatility process is modelled as a Markov chain on a (small) finite number of states, in the study of Rossi and Gallo [17] with a very specific structure assumed for the between-state transitions. While our approach does also involve a discrete-state Markov chain in the fitting procedure, this is only used as a computational tool for approximating the *continuous* log-volatility space of the model of interest. Fuertes and Papanicolaou [19] consider continuous-time models that are subject to regime switching.

The remainder of this paper is organized as follows. Section 2 gives a brief introduction to SMN distributions. Section 3 outlines the general class of SVM–SMN models as well as the maximum likelihood-based estimation procedure using HMM methods. In Section 4, we conduct a simulation study in order to verify frequentist properties of the likelihood estimators. Section 5 is devoted to the application and model comparison among particular members of the SVM–SMN models using international market indexes. Some concluding remarks and suggestions for future developments are given in Section 6.

2. Scale mixture of normal distributions

A random variable Y belongs to the SMN family if it can be expressed as

$$Y = \mu + \kappa(\lambda)^{1/2} X, \tag{1}$$

where μ is a location parameter, $X \sim \mathcal{N}(0, \sigma^2)$, λ is a positive mixing random variable with cumulative distribution function (cdf) $H(\cdot \mid \boldsymbol{\nu})$ and probability density function (pdf) $h(\cdot|\boldsymbol{\nu})$, $\boldsymbol{\nu}$ is a scalar or parameter vector indexing the distribution of λ, and $\kappa(\cdot)$ is a positive weight function. As in [20] and [21], we restrict our attention to the case where $\kappa(\lambda) = 1/\lambda$. Given λ, we have $Y|\lambda \sim \mathcal{N}(\mu, \lambda^{-1}\sigma^2)$, and the pdf of Y is given by

$$f_{SMN}(y|\mu, \sigma^2, \nu) = \int_{-\infty}^{\infty} \phi(y|\mu, \lambda^{-1}\sigma^2) dH(\lambda|\boldsymbol{\nu}), \tag{2}$$

where $\phi(\cdot \mid \mu, \sigma^2)$ denotes the density of the univariate $\mathcal{N}(\mu, \sigma^2)$ distribution. From Equation (2), we have that the cdf of the SMN distributions is given by

$$\begin{aligned} F_{SMN}(y|\mu, \sigma^2, \nu) &= \int_{-\infty}^{y} \int_{-\infty}^{\infty} \phi(u|\mu, \lambda^{-1}\sigma^2) dH(\lambda|\boldsymbol{\nu}) du \\ &= \int_{-\infty}^{\infty} \Phi\left(\frac{\lambda^{1/2}[y-\mu]}{\sigma}\right) dH(\lambda|\boldsymbol{\nu}), \end{aligned} \tag{3}$$

where $\Phi(\cdot)$ is the cdf of the standard normal distribution. The notation $Y \sim \mathcal{SMN}(\mu, \sigma^2, \boldsymbol{\nu}; H)$ will be used when Y has pdf Equation (2) and cdf Equation (3). As mentioned earlier, the SMN family constitutes a class of thick-tailed distributions, including the normal, Student's t, and Slash distributions, which are obtained, respectively, by choosing the mixing variables as: $\lambda = 1$, $\lambda \sim \mathcal{G}(\frac{\nu}{2}, \frac{\nu}{2})$, and $\lambda \sim \mathcal{B}e(\nu, 1)$, where $\mathcal{G}(\cdot, \cdot)$ and $\mathcal{B}e(\cdot, \cdot)$ denote the gamma and beta distributions, respectively.

3. The heavy-tailed stochastic volatility in mean model

3.1. Model formulation

The SVM model with heavy tails is defined by

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 \mathrm{e}^{h_t} + e^{\frac{h_t}{2}} \epsilon_t, \tag{4a}$$

$$h_{t+1} = \mu + \phi(h_t - \mu) + \sigma_\eta \eta_t, \tag{4b}$$

where y_t and h_t are, respectively, the compounded return and the log-volatility at time t. We assume that $|\phi| < 1$; that is, the log-volatility process is stationary and the initial distribution is $h_1 \sim \mathcal{N}(\mu, \frac{\sigma_\eta^2}{1-\phi^2})$. The innovations ϵ_t and η_t are assumed to be mutually independent, $\epsilon_t \sim \mathcal{SMN}(0, 1, \nu; H)$ and $\eta_t \sim \mathcal{N}(0, 1)$, respectively. The aim of the SVM–SMN class of models is to simultaneously estimate the *ex-ante* relation between returns and volatility and the volatility feedback effect in the presence of outliers. This class of models includes SVM models with normal distribution (SVM-N) [9], the Student's *t* (SVM-T), and slash (SVM-S) distributions as special cases.

3.2. Likelihood evaluation by iterated numerical integration

To formulate the likelihood, we require the conditional pdfs of the random variables y_t, given h_t and y_{t-1} $(t = 1, \dots, T)$, and of the random variables h_t, given h_{t-1} $(t = 2, \dots, T)$. We denote these by $p(y_t \mid y_{t-1}, h_t)$ and $p(h_t \mid h_{t-1})$, respectively. For a member of the class of SMN distributions, the likelihood of the model defined by Equations (4a) and (4b) can then be derived as

$$\begin{aligned}
\mathcal{L} &= \int \dots \int p(y_1, \dots, y_T, h_1, \dots, h_T \mid y_0) dh_T \dots dh_1 \\
&= \int \dots \int p(y_1, \dots, y_T \mid y_0, h_1, \dots, h_T) p(h_1, \dots, h_T) dh_T \dots dh_1 \\
&= \int \dots \int p(h_1) p(y_1 \mid y_0, h_1) \prod_{t=2}^{T} p(y_t \mid y_{t-1}, h_t) p(h_t \mid h_{t-1}) dh_T \dots dh_1,
\end{aligned}$$

exploiting the dependence structure of the stochastic volatility models in the last step. Hence, the likelihood is a high-order multiple integral that cannot be evaluated analytically. Through numerical integration, using a simple rectangular rule based on m equidistant intervals, $B_i = (b_{i-1}, b_i)$, $i = 1, \dots, m$, with midpoints b_i^* and length b, the likelihood can be approximated as follows:

$$\begin{aligned}
\mathcal{L} \approx b^T \sum_{i_1=1}^{m} \dots \sum_{i_T=1}^{m} & p(h_1 = b_{i_1}^*) p(y_1 \mid y_0, h_1 = b_{i_1}^*) \\
\times & \prod_{t=2}^{T} p(h_t = b_{i_t}^* \mid h_{t-1} = b_{i_{t-1}}^*) p(y_t \mid y_{t-1}, h_t = b_{i_t}^*) = \mathcal{L}_{\text{approx}} .
\end{aligned} \tag{5}$$

This approximation can be made arbitrarily accurate by increasing m, provided that the interval (b_0, b_m) covers the essential range of the log-volatility process. We note that this simple midpoint quadrature is by no means the only way in which the integral can be approximated (cf. [22]).

3.3. Fast evaluation and numerical maximization of the approximate likelihood using hidden Markov model techniques

In the form given in (5), the approximate likelihood can be evaluated numerically, but the evaluation will usually be computationally intractable because it involves m^T summands. However, instead of the brute force summation in (5), an efficient recursive scheme can be used to evaluate the approximate likelihood. To see this, we note that the numerical integration essentially corresponds to a discretization of the state space, that is, the support of the log-volatility process h_t. Therefore, the approximate likelihood given in (5) can be evaluated using the tools well established for HMMs, which are the models that have exactly the same dependence structure as the stochastic volatility models, but with a finite and hence discrete state space (cf. [22, 23]). In the given scenario, the discrete states correspond to the intervals B_i, $i = 1, \dots, m$, in which the state space has been partitioned. A key property of HMM, which we exploit here, is that the likelihood can be evaluated efficiently using the so-called forward algorithm, a recursive scheme that iteratively traverses forward along the time series, updating the likelihood and the state probabilities in each step [24]. For an HMM, applying the forward algorithm results in a convenient closed-form matrix product expression for the likelihood, and this is exactly what is obtained also for the SVM–SMN class of models:

$$\mathcal{L}_{\text{approx}} = \boldsymbol{\delta}\mathbf{P}(y_1)\boldsymbol{\Gamma}\mathbf{P}(y_2)\boldsymbol{\Gamma}\mathbf{P}(y_3)\cdots\boldsymbol{\Gamma}\mathbf{P}(y_{T-1})\boldsymbol{\Gamma}\mathbf{P}(y_T)\mathbf{1}^{\top}. \tag{6}$$

Here, the $m \times m$-matrix $\boldsymbol{\Gamma} = \left(\gamma_{ij}\right)$ is the analogue to the transition probability matrix in case of an HMM, defined by $\gamma_{ij} = p(h_t = b_j^* \mid h_{t-1} = b_i^*) \cdot b$, which is an approximation of the probability of the log-volatility process changing from some value in the interval B_i to some value in the interval B_j; this conditional probability is determined by Equation (4b). The vector $\boldsymbol{\delta}$ is the analogue to the Markov chain initial distribution in case of an HMM, here defined such that δ_i is the density of the $\mathcal{N}(\mu, \frac{\sigma_\eta^2}{1-\phi^2})$-distribution—the stationary distribution of the log-volatility process—multiplied by b. Furthermore, $\mathbf{P}(y_t)$ is an $m \times m$ diagonal matrix with the ith diagonal entry $p(y_t \mid y_{t-1}, h_t = b_i^*)$, hence the analogue to the matrix comprising the state-dependent probabilities in case of an HMM; this conditional probability is determined by Equation (4a). Finally, $\mathbf{1}^{\top}$ is a column vector of ones. Using the matrix product expression given in Equation (6), the computational effort required to evaluate the approximate likelihood is linear in the number of observations, T, and quadratic in the number of intervals used in the discretization, m.

In practice, this means that the likelihood can typically be calculated in a fraction of a second, even for T in the thousands and say $m = 100$, a value that renders the approximation virtually exact (see the simulation experiments later). Furthermore, the approximation can be made arbitrarily accurate by increasing m (and potentially widening the interval $[b_0, b_m]$). For any SVM–SMN model of interest, it is hence convenient and computationally inexpensive to estimate the model parameters using a numerical maximization of the approximate (log-)likelihood. Such a numerical maximization needs to address standard technical issues such as parameter constraints and numerical overflow [22].

It should perhaps be noted here that, although we are using the HMM forward algorithm to evaluate the (approximate) likelihood, the specifications of $\boldsymbol{\delta}$, $\boldsymbol{\Gamma}$, and $\mathbf{P}(x_t)$ given earlier do not define exactly an HMM, because in general the row sums of $\boldsymbol{\Gamma}$ will only approximately equal one, and the components of the vector $\boldsymbol{\delta}$ will only approximately sum to one. If desired, this can be remedied by scaling each row of $\boldsymbol{\Gamma}$ and the vector $\boldsymbol{\delta}$ to total 1.

3.4. Forecasts and model checking

The HMM forward algorithm can also be used to obtain forecast distributions for SVM models. For example, it is straightforward to find the cdf of the one-step-ahead forecast distribution on day $t-1$, that is, the conditional distribution of the

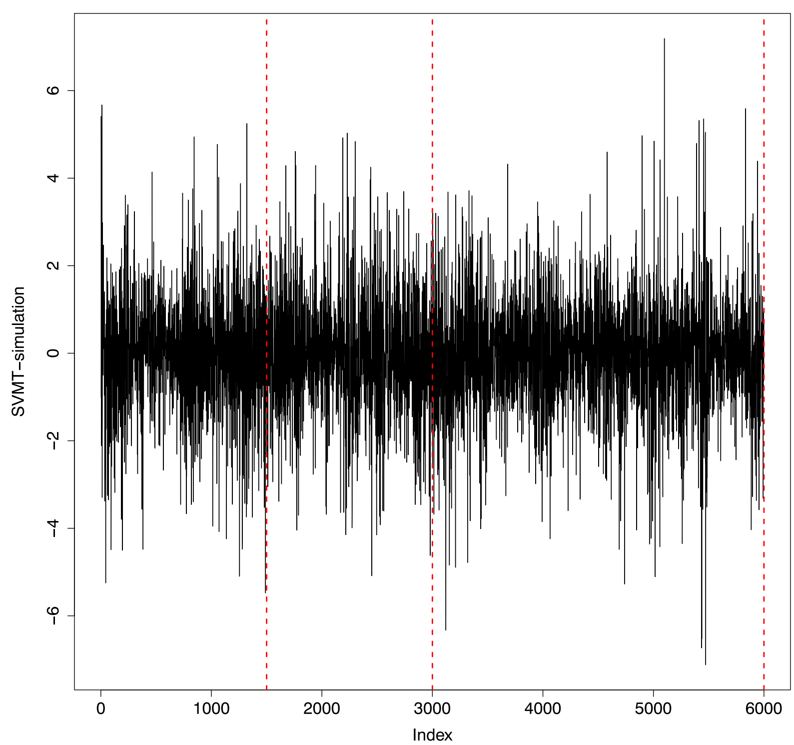

Figure 1. Simulated data set from the SVM-T with $\boldsymbol{\beta} = (0.2, 0.07, -0.18)^{\top}$, $\mu = 0.1$, $\phi = 0.98$, $\sigma_\eta = 0.1$, and $\nu = 8$ and $y_0 = 0.2$. [Colour figure can be viewed at wileyonlinelibrary.com]

Table I. SVM-T model, simulated data set: maximum likelihood estimates of the parameters and computing times in seconds for the HMM method ($b_m = -b_0 = 4$).

Size	m	$\mathcal{L}$	ϕ	σ_η	μ	β_0	β_1	β_2	ν	Time
	50	2437.26	0.9869	0.0833	0.1418	0.1918	0.0693	−0.1448	7.2929	39.45
			(0.9747, 0.9991)	(0.0504, 0.1162)	(−0.2120, 0.4957)	(0.0474, 0.3362)	(0.0203, 0.1184)	(−0.2774, −0.0124)	(4.4666, 10.1191)	
	100	2437.24	0.9880	0.0788	0.1381	0.1922	0.0687	−0.1458	7.2712	91.99
1500			(0.9738, 1.0021)	(0.0321, 0.1255)	(−0.2086, 0.4849)	(0.0471, 0.3372)	(0.0197, 0.1177)	(−0.2792, −0.0125)	(4.4592, 10.0832)	
	150	2437.24	0.9880	0.0788	0.1381	0.1922	0.0687	−0.1458	7.2712	156.68
			(0.9738, 1.0021)	(0.0321, 0.1255)	(−0.2086, 0.4849)	(0.0471, 0.3372)	(0.0197, 0.1177)	(−0.2792, −0.0125)	(4.4592, 10.0832)	
	200	2437.24	0.9880	0.0788	0.1381	0.1922	0.0687	−0.1458	7.2712	246.29
			(0.9738, 1.0021)	(0.0321, 0.1255)	(−0.2086, 0.4849)	(0.0471, 0.3372)	(0.0197, 0.1177)	(−0.2792, −0.0125)	(4.4592, 10.0832)	
	50	4901.58	0.9772	0.1191	0.1774	0.1875	0.0705	−0.1561	8.9200	96.22
			(0.9623, 0.9921)	(0.0798, 0.1585)	(−0.0301, 0.3850)	(0.0904, 0.2847)	(0.0349, 0.1063)	(−0.2415, −0.0707)	(5.8739, 11.9661)	
	100	4901.53	0.9741	0.1202	0.1206	0.1924	0.0699	−0.1632	8.1263	203.26
3000			(0.9587, 0.9895)	(0.0801, 0.1601)	(−0.0663, 0.3075)	(0.0931, 0.2917)	(0.0343, 0.1056)	(−0.2523, −0.0740)	(5.6410, 10.6116)	
	150	4901.53	0.9741	0.1202	0.1206	0.1924	0.0699	−0.1632	8.1263	335.65
			(0.9587, 0.9895)	(0.0801, 0.1601)	(−0.0663, 0.3075)	(0.0931, 0.2917)	(0.0343, 0.1056)	(−0.2523, −0.0740)	(5.6410, 10.6116)	
	200	4966.28	0.9741	0.1202	0.1206	0.1924	0.0699	−0.1632	8.1263	501.72
			(0.9587, 0.9895)	(0.0801, 0.1601)	(−0.0663, 0.3075)	(0.0931, 0.2917)	(0.0343, 0.1056)	(−0.2523, −0.0740)	(5.6410, 10.6116)	
	50	9609.13	0.9783	0.1071	0.0537	0.2048	0.0652	−0.1982	7.7256	150.45
			(0.9688, 0.9880)	(0.0823, 0.1320)	(−0.0858, 0.1934)	(0.1320, 0.2777)	(0.0403, 0.0901)	(−0.2689, −0.1276)	(6.1679, 9.2833)	
	100	9609.12	0.9783	0.1073	0.0541	0.2043	0.0652	−0.1978	7.7325	318.13
6000			(0.9684, 0.9883)	(0.0813, 0.1333)	(−0.0854, 0.1936)	(0.1316, 0.2771)	(0.0403, 0.0901)	(−0.2685, −0.1273)	(6.1711, 9.2939)	
	150	9609.12	0.9783	0.1073	0.0541	0.2043	0.0652	−0.1978	7.7325	557.27
			(0.9684, 0.9883)	(0.0813, 0.1333)	(−0.0854, 0.1936)	(0.1316, 0.2771)	(0.0403, 0.0901)	(−0.2685, −0.1273)	(6.1711, 9.2939)	
	200	9609.12	0.9783	0.1073	0.0541	0.2043	0.0652	−0.1978	7.7325	839.15
			(0.9684, 0.9883)	(0.0813, 0.1333)	(−0.0854, 0.1936)	(0.1316, 0.2771)	(0.0403, 0.0901)	(−0.2685, −0.1273)	(6.1711, 9.2939)	

True values of the parameters: $\boldsymbol{\beta} = (0.2, 0.07, -0.18)^{\top}$, $\mu = 0.1$, $\phi = 0.98$, $\sigma_\eta = 0.1$, and $\nu = 8$.

Table II. SVM-T model: simulation study results based on 300 replicates using the HMM method ($b_{max} = -b_{min} = 4$ and $T = 2500$).

Parameter	True value	Mean	MRB	MARB	MSE
		$m = 50$			
ϕ	0.98	0.9854	0.0045	0.0101	0.0002
σ_η	0.10	0.1002	0.0026	0.1879	0.0007
μ	0.10	0.2077	1.0774	1.3727	0.0307
β_1	0.20	0.1942	−0.0289	0.2424	0.0039
β_2	0.07	0.0716	0.0234	0.2320	0.0004
β_3	−0.18	−0.1711	−0.0493	0.2391	0.0032
ν	8.00	8.7098	0.0696	0.2669	1.8567
		$m = 100$			
ϕ	0.98	0.9859	0.0044	0.0099	0.0002
σ_η	0.10	0.0953	−0.0046	0.2210	0.0009
μ	0.10	0.1852	0.8524	1.3024	0.0285
β_1	0.20	0.1962	−0.0187	0.2515	0.0046
β_2	0.07	0.0713	0.0181	0.2271	0.0004
β_3	−0.18	−0.1736	−0.0351	0.2592	0.0034
ν	8.00	8.6057	0.0675	0.2612	1.8471
		$m = 150$			
ϕ	0.98	0.9857	0.0044	0.0100	0.0002
σ_η	0.10	0.0952	−0.0047	0.2221	0.0009
μ	0.10	0.1868	0.8644	1.3143	0.0289
β_1	0.20	0.1967	−0.0164	0.2451	0.0043
β_2	0.07	0.0713	0.0164	0.2455	0.0004
β_3	−0.18	−0.1736	−0.0321	0.2544	0.0037
ν	8.00	8.6056	0.0671	0.2617	1.8422
		$m = 200$			
ϕ	0.98	0.9857	0.0044	0.0100	0.0002
σ_η	0.10	0.0952	−0.0480	0.2227	0.0009
μ	0.10	0.1868	0.8644	1.3143	0.0289
β_1	0.20	0.1967	−0.0146	0.2479	0.0044
β_2	0.07	0.0711	0.0163	0.2253	0.0004
β_3	−0.18	−0.1742	−0.0304	0.2572	0.0039
ν	8.00	8.6056	0.0671	0.2619	1.8422

return on day t, given all previous observations. This is given by

$$F(y_t \mid y_{t-1}, y_{t-2}, \dots, y_0) \approx \sum_{i=1}^{m} \zeta_i F(y_t \mid y_{t-1}, h_t = b_i^*), \tag{7}$$

where ζ_i is the ith entry of the vector $\boldsymbol{\alpha}_{t-1}/(\boldsymbol{\alpha}_{t-1}\mathbf{1}^\top)$, obtained from the forward probabilities,

$$\boldsymbol{\alpha}_{t-1} = \boldsymbol{\delta}\mathbf{P}(y_1)\boldsymbol{\Gamma}\mathbf{P}(y_2)\boldsymbol{\Gamma}\mathbf{P}(y_3)\cdots\boldsymbol{\Gamma}\mathbf{P}(y_{t-1}),$$

with $\boldsymbol{\delta}$, $\mathbf{P}(y_k)$, and $\boldsymbol{\Gamma}$ defined as above. The corresponding expression for longer forecast horizons is similar (see Chapter 5 of [24] for details). The approximation in Equation (7) usually becomes virtually exact for values of m about 100. A closed-form expression for obtaining state predictions, that is, volatility predictions, is also available. Furthermore, the forecast distribution given in Equation (7) can be used in order to perform model checking via residuals [25]. The one-step-ahead forecast pseudo-residual (or quantile residual) is given by

$$r_t = \Phi^{-1}(F(y_t \mid y_{t-1}, y_{t-2}, \dots, y_0)), \tag{8}$$

for $t = 1, \dots, T$. For a correctly specified model, the r_t follows a standard normal distribution [25–29]. Thus, forecast pseudo-residuals can be used to identify extreme values, and the suitability of the model can be checked by using, for example, qq-plots or formal tests for normality.

Table III. SVM-T model: simulation study results based on 300 replicates using the HMM method ($b_{max} = -b_{min} = 4$ and $T = 5000$).

Parameter	True value	Mean	MRB	MARB	MSE
		$m = 50$			
ϕ	0.98	0.9861	0.0061	0.0087	0.0001
σ_η	0.10	0.0947	−0.0534	0.1167	0.0002
μ	0.10	0.2067	1.0774	1.4030	0.0331
β_1	0.20	0.1987	−0.0066	0.1866	0.0021
β_2	0.07	0.0714	0.0204	0.1654	0.0002
β_3	−0.18	−0.1764	−0.0197	0.1925	0.0022
ν	8.00	8.4403	0.0550	0.1124	1.3137
		$m = 100$			
ϕ	0.98	0.9865	0.0067	0.0091	0.0001
σ_η	0.10	0.0908	−0.0921	0.1542	0.0003
μ	0.10	0.2062	1.0624	1.3551	0.0320
β_1	0.20	0.2012	0.0062	0.1869	0.0023
β_2	0.07	0.0712	0.0184	0.1638	0.0002
β_3	−0.18	−0.1783	−0.0090	0.1909	0.0021
ν	8.00	8.4334	0.0541	0.1129	1.5173
		$m = 150$			
ϕ	0.98	0.9865	0.0066	0.0091	0.0001
σ_η	0.10	0.0909	−0.0901	0.1523	0.0003
μ	0.10	0.2033	1.0335	1.3411	0.0317
β_1	0.20	0.2013	0.0067	0.1858	0.0023
β_2	0.07	0.0713	0.0191	0.1633	0.0002
β_3	−0.18	−0.1784	−0.0087	0.1900	0.0021
ν	8.00	8.4138	0.0535	0.1117	1.6200
		$m = 200$			
ϕ	0.98	0.9865	0.0066	0.0091	0.0001
σ_η	0.10	0.0909	−0.0901	0.1524	0.0003
μ	0.10	0.2033	1.0335	1.3411	0.0317
β_1	0.20	0.2013	0.0067	0.1858	0.0023
β_2	0.07	0.0713	0.0191	0.1633	0.0002
β_3	−0.18	−0.1784	−0.0087	0.1900	0.0021
ν	8.00	8.4138	0.0535	0.1117	1.5090

3.5. Decoding

Again building on standard HMM machinery, estimates of the underlying log-volatility values can easily be obtained using the Viterbi algorithm, which is an efficient dynamic programming algorithm for computing the most likely Markov chain state sequence to have given rise to observations stemming from an HMM (see [22, 24] for details).

4. Simulation study

In order to assess the performance of the methodology described in the previous section, we conducted some simulation experiments. All the calculations were performed using stand-alone code developed by the authors using the Rcpp interface inside R. First, we simulated a data set comprising $T = 6000$ observations from the SVM-T model, specifying $\boldsymbol{\beta} = (0.2, 0.07, -0.18)^\top$, $\mu = 0.1$, $\phi = 0.98$, $\sigma_\eta = 0.1$, $\nu = 8$, and $y_0 = 0.2$, which correspond to typical values found in daily series of returns. Figure 1 shows the resulting artificial data set.

In order to investigate the influence of the choice of m on the accuracy of the likelihood approximation and of the sample size T on the computing time, we additionally fitted the SVM-T model using $m = 50, 100, 150, 200$ (i.e., different levels of accuracy), $b_m = -b_0 = 4$, to subsamples of length $T = 1500, 3000$ of the original simulated series. Table I reports the results. We observe that the parameter estimates obtained by the approximate maximum likelihood method become stable for values of m around 100, for all the sample sizes considered here. Although the results are not reported here, we also investigated the influence of the choice of b_0 and b_m, finding that the estimator performance was not affected much unless

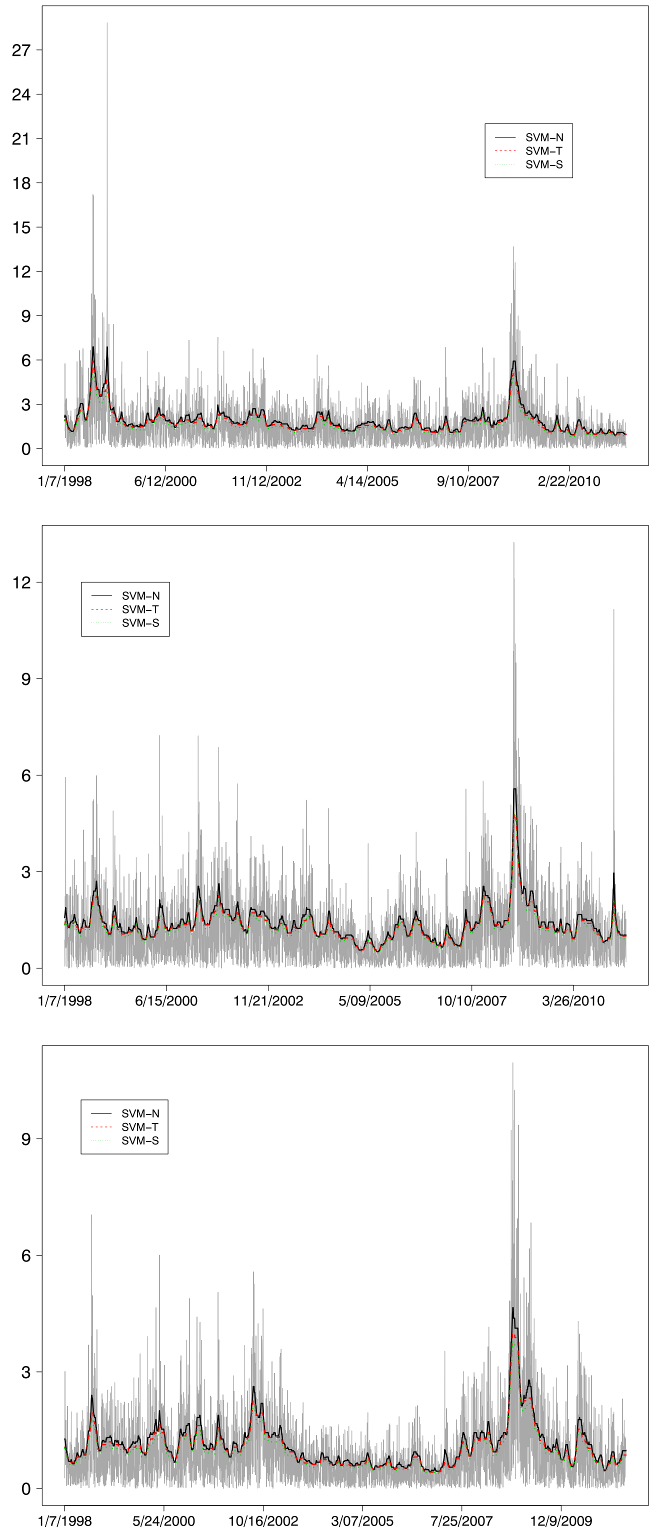

Figure 2. Decoded $e^{\frac{h_t}{2}}$ using the viterbi algorithm: top (IBVSP), center (NIKKEI), and bottom (SP500). The solid line (SVM-N), dotted red line (SVM-T), and dotted green line (SVM-S). The grey line indicates the absolute returns. [Colour figure can be viewed at wileyonlinelibrary.com]

these were chosen either much too small (thus not covering the essential support of the log-volatility process, which in the given setting would be the case for example for $b_m = -b_0 = 2$) or much too large (thus leading to a partition of the support into unnecessarily wide intervals in the numerical integration and, hence, a poor approximation of the likelihood, for example with $m = 50$ and $b_m = -b_0 = 15$). In practice, it can easily be checked *post hoc* if the chosen range, specified by b_0 and b_m, is adequate, by investigating the stationary distribution of the fitted log-volatility process. As regards the computing time to get the maximum likelihood estimators of the parameters, we observe that, for example, for the SVM-T with $m = 200$ and 6000 observations, our approach using the HMM machinery takes about 14 min, instead of almost 4 h to realize 50,000 iterations in order to achieve convergence of an MCMC procedure. With the HMM approach, the computational effort increases linearly in T.

Next, we conducted a second simulation experiment with the objective to study properties of the maximum likelihood estimators of the SVM model's parameters. We generated 300 datasets from the SVM-T model, specifying $\boldsymbol{\beta} = (0.20, 0.07, -0.18)^{\top}$, $\phi = 0.98$, $\sigma_\eta = 0.1$, $\mu = 0.10$, and $\nu = 10$. For each generated data set, we fitted the SVM-T model using $m = 50, 100, 150, 200$ and $b_m = -b_0 = 4$ for $T = 2500$ and $T = 5000$, respectively. Tables II and III report the sample mean, the mean relative bias (MRB), the mean relative absolute deviation, and the mean squared error (MSE) of the parameter estimates.

For both sample sizes, that is, $T = 2500$ and $T = 5000$, the highest MRB was found for the estimator of μ, while none of the other estimators exhibited a notable bias. The bias found for $\hat{\mu}$ does not have a strong effect of the resulting model and its performance in forecasting, because it simply means a minor shift of the volatility process. With the exception of $\hat{\mu}$, the MSEs are smaller for the larger sample size, as would be expected. The results obtained when using $m = 50$ are again similar to those using higher values of m and hence finer approximations of the likelihood.

Overall, it can be concluded that the use of the HMM machinery to numerically maximize the approximate likelihood function of SVM models leads to good estimator performance yet involves only a modest computational effort. At this point, it may be worthwhile to reiterate that, given that the likelihood approximation considered is virtually exact for $m = 100$, we are effectively conducting maximum likelihood estimation.

Additional simulation experiments, for the SVM-N and SVM-S models, are provided in the Supporting Information.

5. Empirical application

5.1. The data

In this section, we analyze the indexes from the São Paulo Stock, Mercantile and Futures Exchange, Tokyo Stock Exchange, and the New York Stock Exchange. The considered indexes are the IBOVESPA (IBVSP), Nikkei 225 (NIKKEI), and the S&P 500 (SP500), respectively. The period of analysis is from January 5, 1998, until June 30, 2011. All the datasets were downloaded from http://finance.yahoo.com. Stock returns are computed as $y_t = 100 \times (\log P_t - \log P_{t-1})$, where P_t is the (adjusted) closing price on day t. Table IV reports a summary of descriptive statistics for the series returns. The IBVSP returns show positive skewness, and the NIKKEI and SP500 returns negative skewness (NIKKEI and SP500). All the series show kurtosis greater than three, confirming a well-known stylized fact for return series, namely, the departure from normality. We analyze the data with the aim of providing robust inference by using the SMN class of distributions. In our analysis, we compare the SVM-N, SVM-T, and SVM-S models for each one of the series described earlier.

In order to obtain the maximum likelihood estimates (MLEs) of the parameters in the SVM models, we apply the HMM machinery as introduced in Section 3. To ensure a good approximation of the estimates, we use $b_m = -b_0 = 4$ and $m = 200$. As before, all the calculations were performed using the stand-alone code developed by the authors using the Rcpp interface inside R package. Table V shows the results for the SVM-N, SVM-T, and SVM-S models for each one of the return indexes series.

For all the three series of returns and all the three models considered, we find that the MLEs of ϕ were very close to 0.98, indicating a high persistence of the log-volatility process. The persistence in the log-volatility process underlying the SVM-N model was smaller than that found under the other two models. The MLEs of σ_η under the SVM-T and SVM-S were smaller than the one under the SVM-N, indicating that the volatility processes of the SVM-T and SVM-S were less variable than those in the SVM-N case.

Table IV. Summary statistics of the return indexes.

Return	Size	Mean	SD	Minimum	Maximum	Skewness	Kurtosis
IBVSP	3338	0.0531	2.1964	−17.2082	28.8325	0.5748	16.7109
NIKKEI	3310	−0.0127	1.6010	−12.1110	13.2346	−0.3206	9.0383
SP500	3393	0.0089	1.3385	−9.4695	10.9572	−0.1494	10.2474

Table V. Results obtained when fitting the SVM-N, SVM-T and SVM-S models to the three series of index returns considered (using $m = 200$ and $b_{max} = -b_{min} = -4$).

	Length	llk	ϕ	σ_η	μ	β_0	β_1	β_2	ν	Time
					SVM-N					
IBVSP	3338	−6792.32	0.9756	0.1590	1.1358	0.1971	0.0030	−0.0309	—	158.21
			(0.9655, 0.9857)	(0.1283, 0.1898)	(0.911, 1.3605)	(0.1075, 0.2867)	(−0.032, 0.038)	(−0.0568, −0.0051)	—	
Nikkei	3310	−5812.40	0.9773	0.1586	0.5763	0.1343	−0.0228	−0.0592	—	162.07
			(0.9672, 0.9874)	(0.1297, 0.1874)	(0.3354, 0.8172)	(0.0672, 0.2013)	(−0.058, 0.0123)	(−0.0947, −0.0237)	—	
SP500	3393	−5062.53	0.9806	0.1660	−0.0636	0.1146	−0.0834	−0.0726	—	147.20
			(0.9728, 0.9884)	(0.1353, 0.1967)	(−0.3595, 0.2324)	(0.0736, 0.1556)	(−0.1179, −0.0488)	(−0.1086, −0.0367)	—	
					SVM-T					
IBVSP	3338	−6793.81	0.9880	0.1218	0.8718	0.2096	0.10	−0.0338	13.1162	366.29
			(0.9796, 0.9964)	(0.0923, 0.1513)	(0.4926, 1.2509)	(0.118, 0.3012)	(−0.0333, 0.0353)	(−0.0645, −0.0032)	(8.0314, 18.201)	
NIKKEI	3310	−5815.45	0.9826	0.1288	0.3725	0.1094	−0.0323	−0.0622	12.6989	366.51
			(0.9738, 0.9913)	(0.0997, 0.1580)	(0.1057, 0.6393)	(0.0395, 0.1793)	(−0.0666, 0.0020)	(−0.1058, −0.0186)	(7.3276, 18.0701)	
SP500	3393	−5056.31	0.9878	0.1267	−0.1257	0.1196	−0.0516	−0.0799	12.0287	377.98
			(0.982, 0.9936)	(0.0997, 0.1537)	(−0.4774, 0.226)	(0.0781, 0.1612)	(−0.0848, −0.0183)	(−0.1237, −0.0361)	(7.0504, 17.007)	
					SVM-S					
IBVSP	3338	−6788.26	0.9843	0.1368	0.9315	0.2061	0.0101	−0.0455	4.4962	1070.63
			(0.9754, 0.9933)	(0.1057, 0.1678)	(0.6112, 1.2518)	(0.1153, 0.2969)	(−0.0246, 0.0449)	(−0.0802, −0.0108)	(2.3816, 6.6107)	
NIKKEI	3310	−5811.25	0.9808	0.1430	0.3284	0.1344	−0.0273	−0.0765	4.3333	922.22
			(0.9710, 0.9906)	(0.1111, 0.1749)	(0.0376, 0.6193)	(0.066, 0.2028)	(−0.0624, 0.0077)	(−0.1247, −0.0282)	(2.0783, 6.5884)	
SP500	3393	−5056.46	0.9881	0.1331	−0.3523	0.1019	−0.0605	−0.0741	3.4195	1080.35
			(0.9819, 0.9942)	(0.1057, 0.1606)	(−0.7472, 0.0426)	(0.0609, 0.1429)	(−0.0943, −0.0267)	(−0.1249, −0.0234)	(2.2619, 4.5772)	

Table VI. Model comparison via AIC.

		AIC	
Return	SVM-N	SVM-T	SVM-S
IBVSP	13596.64	13601.62	**13590.52**
NIKKEI	11636.80	**11636.40**	11636.50
SP500	10137.06	**10126.62**	10126.92

For each series considered, the minimum AIC is highlighted in bold.

Table VII. P-values of Jarque–Bera tests applied to one-step-ahead ahead forecast pseudo-residuals.

	SVM-N	SVM-S	SVM-T
IBVSP	0.97	0.51	0.49
NIKKEI	0.11	0.12	0.07
SP500	0.0002	0.0006	0.004

For the mean process, we find that the MLEs of β_0 were positive and statistically significant, because the 95% confidence intervals do not contain zero (for all models and series considered). For the IBVSP and the NIKKEI return series, there is an indication that β_1 might not be relevant. In the SP500 case, β_1 was significant. The β_2 parameter, which measures both the *ex ante* relationship between returns and volatility and the volatility feedback effect, was estimated to be negative and was deemed statistically significant, for all the models and the indexes considered. These results empirically confirm the previous results reported in the literature and indicate that when investors expect higher persistent levels of volatility in the future, they require compensation for this in the form of higher expected returns.

The magnitude of the tail fatness is measured by the shape parameter ν in the SVM-T and SVM-S models. The MLEs of ν were 13.11, 12.69, and 12.03, respectively, for the IBVSP, NIKKEI, and SP500 under the SVM-T. Figure 2 shows the decoded volatilities $exp(h_t/2)$ for the three series considered. We find smoother trajectories under the models SVM-S and SVM-T than for the SVM-N model. Extreme returns, such as during the subprime crisis, make the differences clear.

When we compare the models in terms of their relative in-sample fit, using the values of the Akaike information criterion (AIC) obtained for the different models and series considered (Table VI), the main results are as follows. For IBVSP, the SVM-S model clearly outperforms its competitors, while for NIKKEI and SP500, the SVM-T model has a slightly better AIC score than the SVM-S model. Thus, from this model selection exercise, no clear pattern emerges as to which of the three model formulations is most suited to modelling indexes, although there is an indication that the SVM-N model is less suitable than its more flexible competitors.

We additionally performed an out-of-sample analysis of the forecast performance for the models covered in Table VI. For the observation period, January 5, 1988, until September 29, 2014, each return series was divided into a calibration and a validation sample:

- Calibration sample: from January 5, 1998, until June 30, 2011.
- Validation sample: from July 1, 2011, until September 29, 2014.

As a first step, the SVM-N, SVM-S, and SVM-T models were fitted to the calibration sample of each series. This was carried out by using the HMM method with $m = 200$, a value that is large enough to ensure that any anomalies that may occur could not be attributed to inaccuracies in the approximation of the likelihood. Then, for each one of the observations in the validation sample, the (one-step-ahead forecast) pseudo-residual was computed according to Equation (8). As described in Section 3.4, non-normality of these residuals is an indication of mis-specification of the corresponding model.

Jarque–Bera tests were applied to the pseudo-residuals obtained for the different models and series considered; the corresponding P-values are listed in Table VII. The qq-plots for the three return series under the SVM-N, SVM-S, and SVM-T models are shown in Figures 3, 4, and 5, respectively. For the IBVSP return index, the qq-plot indicates a lack of fit in the left tail (SVM-N) and in the right tail (SVM-T and SVM-S). The Jarque–Bera (JB) test accepts the hypothesis of normality of the residuals under the three models at the 5% and 10% level. For the NIKKEI return index, the qq-plot reveals a poor fit in left tail (SVM-N) and a similar lack of fit in the right tail. The JB test accepts the normality assumption of the pseudo-residuals at the 10% level for the SVM-N and SVM-S models and rejects the SVM-T. At the 5% level, the JB test accepts normality of the three models. Finally, considering the SP500 returns, the qq-plot reveals a poor fit in the left tail for the three models, which is most pronounced under the SVM-N model. The JB test confirms these findings and leads to a rejection of the normality assumption of the pseudo-residuals. The indicated mis-specification could be caused by the presence of correlation between the perturbation terms defined by Equations (4a) and (4b).

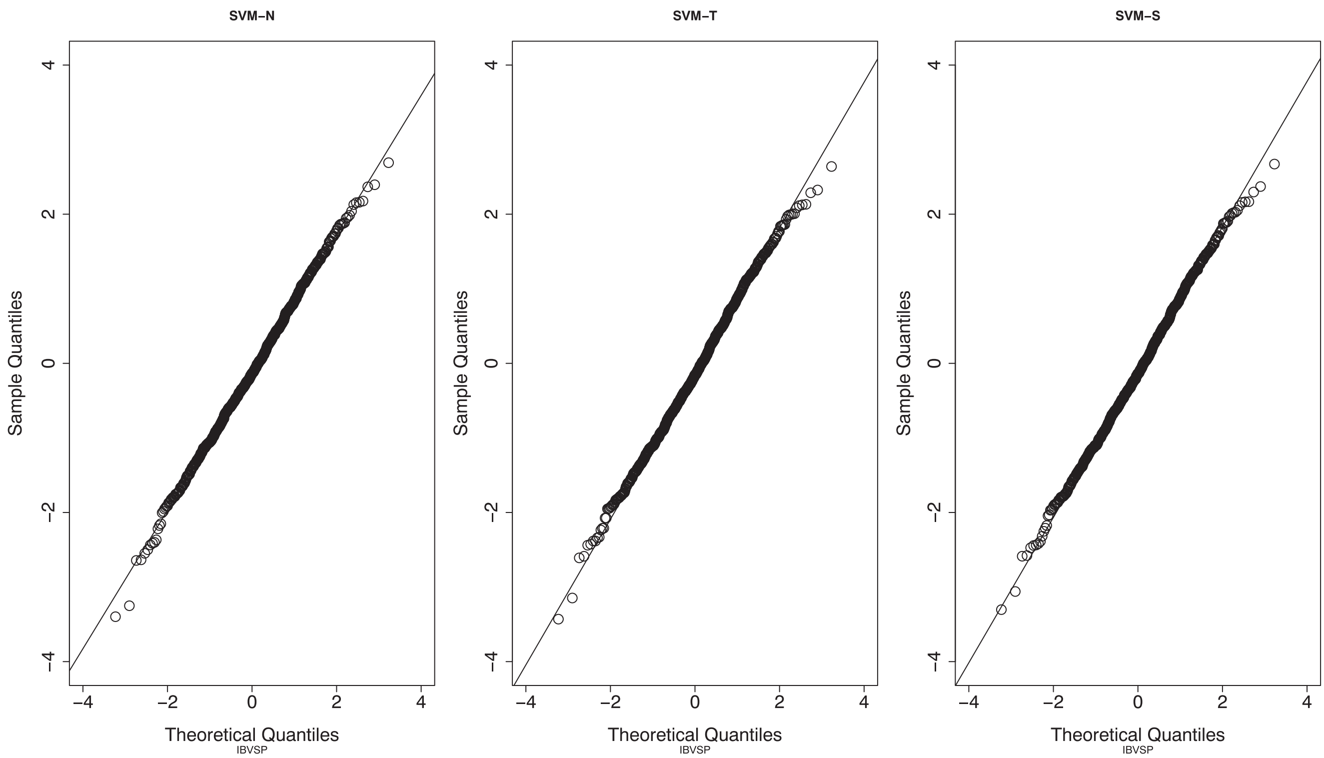

Figure 3. qq-plot of the forecast pseudo-residuals for SVM-N (left), SVM-T (middle), and SVM-S (right) for the IBVSP returns.

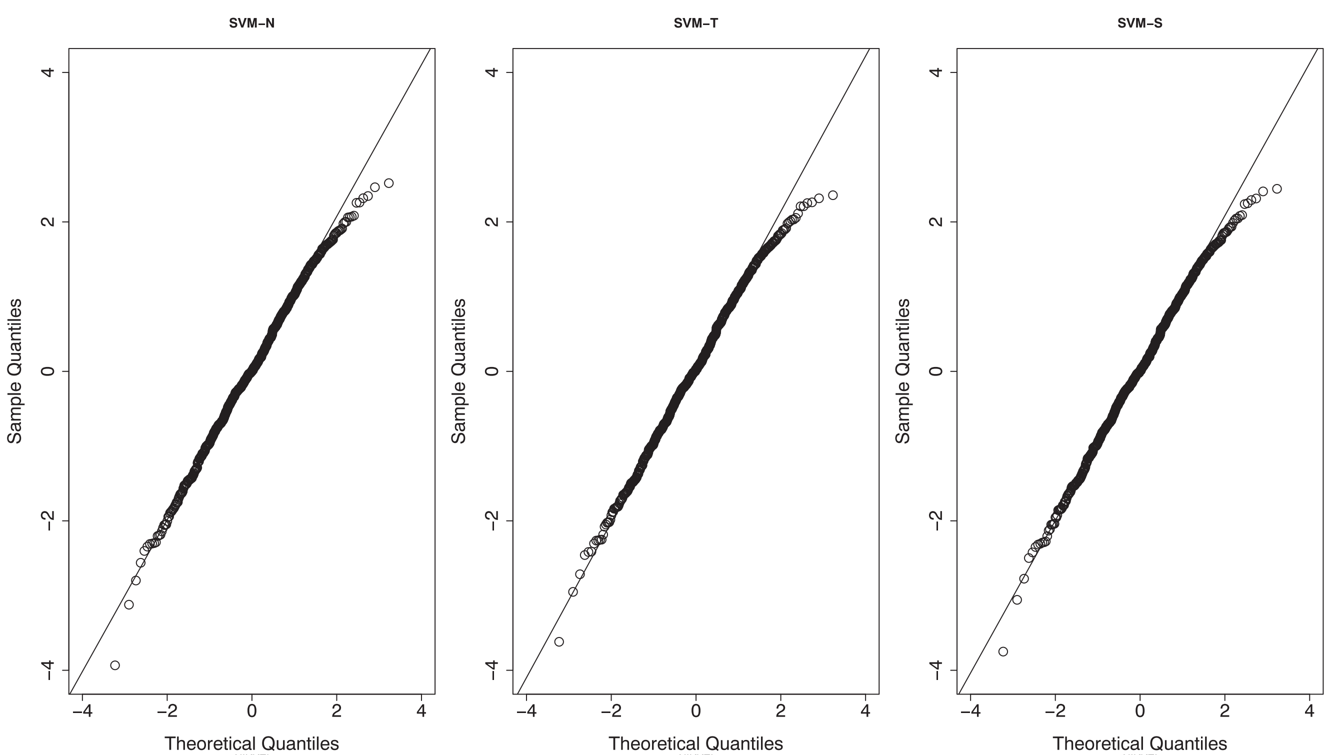

Figure 4. qq-plot of the forecast pseudo-residuals for SVM-N (left), SVM-T (middle), and SVM-S (right) for the NIKKEI returns.

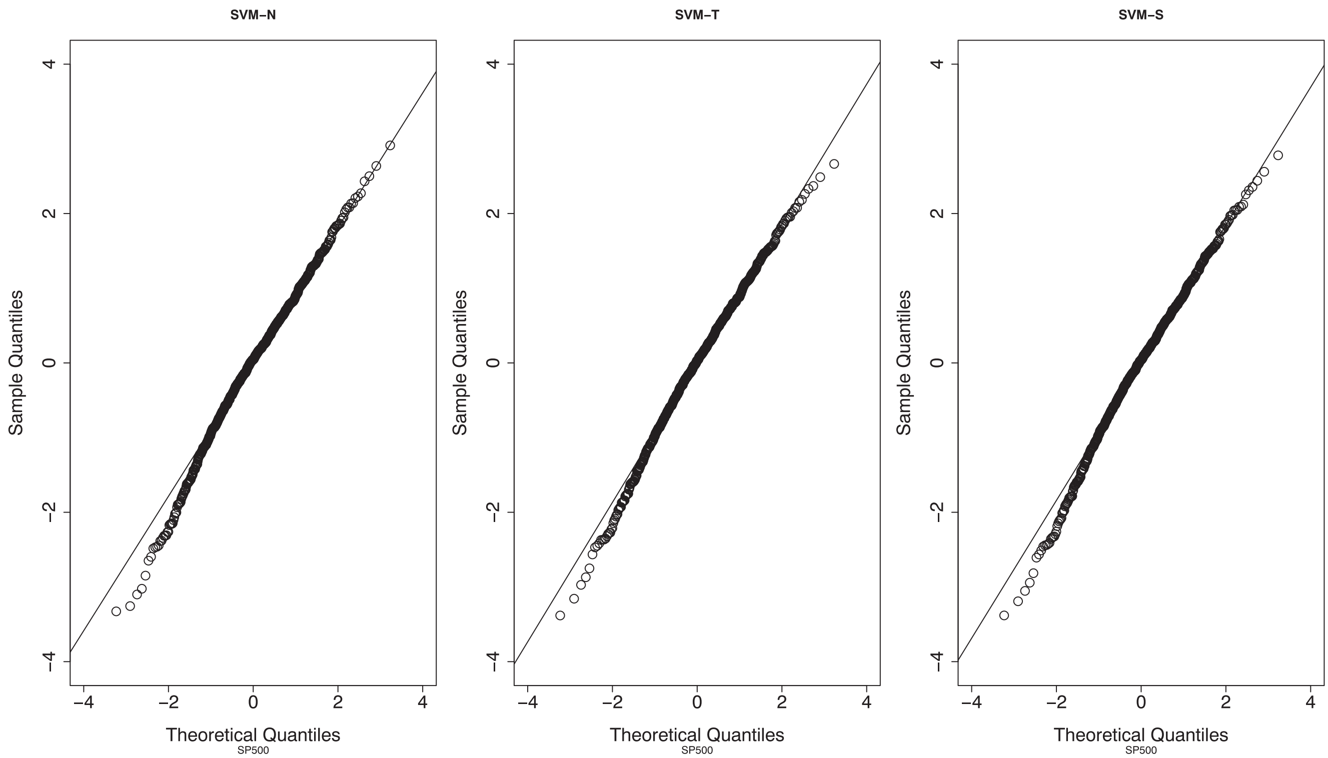

Figure 5. qq-plot of the forecast pseudo-residuals for SVM-N (left), SVM-T (middle), and SVM-S (right) for the SP500 returns.

Table VIII. Violation rate (VR) as a percentage in n one-day-ahead forecast, P-values of the unconditional coverage test at the 1% and 5% levels.

			0.01		0.05	
Return	Model	n	VR (%)	P-value	VR (%)	P-value
	SVM-N	807	0.011	0.747	0.050	0.956
IBVSP	SVM-S	807	0.010	0.975	0.052	0.791
	SVM-T	807	0.011	0.747	0.052	0.791
	SVM-N	808	0.007	0.441	0.046	0.578
NASDAQ	SVM-S	808	0.007	0.441	0.045	0.470
	SVM-T	808	0.007	0.441	0.045	0.470
	SVM-N	816	0.016	0.117	0.056	0.413
SP500	SVM-S	816	0.020	0.015	0.058	0.330
	SVM-T	816	0.016	0.117	0.056	0.413

The plots and tests discussed earlier are useful for assessing the relative and absolute fit of a model, but for the purpose of assessing the risk associated with a share or index, it is the extreme left tail of the forecast distribution that is of particular interest. It determines the value-at-risk (VaR), defined as the maximum possible loss of a portfolio (over a specified period) at a given confidence level. For example, the 1-day 1% VaR is the 0.01-quantile of the one-day-ahead forecast distribution (Table VIII). Whenever the return falls below that quantile, an *exception* is said to have occurred. If the model used for forecasting is correct, then, using a $100\alpha\%$VaR, the number of exceptions, X, in n days follows a binomial(n, α) distribution. This distributional result makes it possible to implement backtesting, where the adequacy of the time series model is assessed through a comparison of the observed number of exceptions and the corresponding theoretical distribution. A standard approach to test the accuracy of VaR forecasts is to assess the violation rate, which is estimated as $\hat{\alpha} = X/n$. In order to examine the accuracy of VaR forecasts, we adopt the unconditional coverage test introduced in [30]. This is a likelihood ratio test with χ^2_1-distributed test statistic

$$LRuc = 2\{\log[\hat{\alpha}^x(1-\hat{\alpha})^{n-x}] - \log[\alpha^x(1-\alpha)^{n-x}]\}. \tag{9}$$

The null hypothesis is that the achieved violation rate is equal to the predetermined nominal probability α. See [30] for more details. According to the unconditional coverage test, we accept the null hypothesis that the achieved violation rate is equal to 5% for all the returns under all the models considered here. We reject that the achieved violation is 1% only for SP500 under the SV-S model.

6. Discussion

In this article, we presented an implementation of an easy-to-implement maximum likelihood-based estimation approach for the SVM model. This model allows us to investigate the dynamic relationship between returns and their time-varying volatility. The commonly made Gaussian assumption of the mean innovation was replaced by univariate thick-tailed processes, known as SMN distributions. While we focused on practical and computational aspects of fitting these models to real data, there may of course also be interest in deriving theoretical properties of the estimators, which could be accomplished using general maximum likelihood theory for state-space models (see, e.g., [31]).

For all the indexes and the models considered in our real data application, the β_2 estimate, which measures both the *ex ante* relationship between returns and volatility and the volatility feedback effect, was found to be negative. The results are in line with those of [32], who found a similar relationship between unexpected volatility dynamics and returns and confirm the hypothesis that investors require higher expected returns when unanticipated increases in future volatility are highly persistent. This is consistent with our findings of higher values of ϕ combined with larger negative values for the in-mean parameter.

Our SVM–SMN models showed considerable flexibility to accommodate outliers; however, their robustness aspects could be seriously affected by the presence of skewness and heavy-tailedness simultaneously. To remedy this problem, the scale mixtures of skew-normal distributions can be used, or alternatively, the conditional distribution of the returns could be modeled nonparametrically [33].

Finally, another important stylized fact often attributed to financial time series, the so-called leverage effect, is not explicitly incorporated in the class of models presented in this paper. Models with leverage effect involve a (negative) correlation between the innovations in the returns and subsequent innovations in the log-volatility process. While relatively easy to accomplish when both innovations are Gaussian—in which case the joint distribution of the innovations can simply be taken to be a bivariate normal—it is not quite as straightforward to formulate corresponding models where the distribution of the innovations in the returns is from the general class of SMN distributions. We believe that the way forward to constructing corresponding models is via the use of copulas, which can be used to couple arbitrary marginal densities, in particular, SMN distributions and normal distributions, respectively. This strategy was first proposed by [34] but has since not been pursued further and, in particular, not within classes of models as flexible as the one discussed in the present paper. The copula-based extension of SVM models, which is beyond the scope of the present paper, is currently under investigation.

Acknowledgements

We would to thank the Editor, an Associate Editor, and the two anonymous referees for their very helpful comments and suggestions, which have led to a much improved version of the paper. The research of Carlos A. Abanto-Valle was partially supported by the CNPq-Brazil and Fundo de Amparo à Pesquisa do Estado de Rio de Janeiro (FAPERJ). The research of Michel V. Cardoso was supported by the Comissão de Aperfeiçoamento de Pessoal de Nível Superior de Pessoal (CAPES). The research of Ming-Hui Chen was partially supported by US NIH grants #GM70335 and #P01CA142538.

References

1. Melino A, Turnbull SM. Pricing foreign options with stochastic volatility. *Journal of Econometrics* 1990; **45**:239–265.
2. Carnero MA, Peña D, Ruiz E. Persistence and kurtosis in GARCH and stochastic volatility models. *Journal of Financial Econometrics* 2004; **2**:319–342.
3. Mandelbrot B. The variation of certain speculative prices. *Journal of Business* 1963; **36**:314–419.
4. Fama E. Portfolio analysis in a stable Paretian market. *Management Science* 1965; **11**:404–419.
5. Liesenfeld R, Jung CR. Stochastic volatility models: conditional normality versus heavy-tailed distributions. *Journal of Applied Econometrics* 2000; **15**(2):137–160.
6. Chib S, Nardari F, Shephard N. Markov chain Monte Carlo methods for stochastic volatility models. *Journal of Econometrics* 2002; **108**(2): 281–316.
7. Jacquier E, Polson NG, Rossi PE. Bayesian analysis of stochastic volatility models with fat-tails and correlated errors. *Journal of Econometrics* 2004; **122**(1):185–212.
8. Abanto-Valle CA, Bandyopadhyay D, Lachos V, Enriquez I. Robust Bayesian analysis of heavy-tailed stochastic volatility models using scale mixtures of normal distributions. *Computational Statistics and Data Analysis* 2010; **54**:2883–2898.

9. Koopman SJ, Uspensky EH. The stochastic volatility in mean model: empirical evidence from international tock markets. *Journal of Applied Econometrics* 2002; **17**:667–689.
10. Abanto-Valle CA, Migon HS, Lachos V. Stochastic volatility in mean models with heavy-tailed distributions. *Brazilian Journal of Probability and Statistics* 2012; **26**:402–422.
11. Kitagawa G. Non-Gaussian state-space modeling of nonstationary time series (with discussion). *Journal of the American Statistical Association* 1987; **82**:1032–1041.
12. Fridman M, Harris L. A maximum likelihood approach for non-Gaussian stochastic volatility models. *Journal of Business and Economic Statistics* 1998; **16**:284–291.
13. Bartolucci F, De Luca G. Maximum likelihood estimation of a latent variable time-series model. *Applied Stochastic Models in Business and Industry* 2001; **17**:5–17.
14. Bartolucci F, De Luca G. Likelihood-based inference for asymmetric stochastic volatility models. *Computational Statictics and Data Analysis* 2003; **42**:445–449.
15. Clements AE, Hurn S, White SI. Mixture distribution-based forecasting using stochastic volatility model. *Applied Stochastic Models in Business and Industry* 2006; **22**:547–557.
16. Rydén T, Terärsvirta T, ÅBsbrink S. Stylized facts of daily return series and the hidden Markov model. *Journal of Applied Econometrics* 1998; **13**:217–244.
17. Rossi A, Gallo GM. Volatility estimation via hidden Markov models. *Journal of Empirical Finance* 2006; **13**:203–230.
18. Bulla J, Bulla I. Stylized facts of financial time series and hidden semi-Markov models. *Computational Statistics and Data Analysis* 2006; **51**:2192–2209.
19. Fuertes C, Papanicolaou A. Implied filtering densities on volatility's hidden state. *Applied Mathematical Finance* 2014; **21**:483–522.
20. Lange KL, Sinsheimer JS. Normal/independent distributions and their applications in robust regression. *Journal of Computational and Graphical Statistics* 1993; **2**:175–198.
21. Choy S, Wan WY, Chan C. Bayesian student-t stochastic volatility models via scale mixtures. *Advances in Econometrics* 2008; **23**:595–618.
22. Langrock R, MacDonald IL, Zucchini W. Some nonstandard stochastic volatility models and their estimation using structured hidden Markov models. *Journal of Empirical Finance* 2012; **19**:147–161.
23. Langrock R. Some applications of nonlinear and non-Gaussian state-space modelling by means of hidden Markov models. *Journal of Applied Statistics* 2011; **38**:2955–2970.
24. Zucchini W, MacDonald IL, Langrock R. *Hidden Markov Models for Time Series: An Introduction Using R* 2nd ed. Chapman & Hall: Boca Raton, FL, 2016.
25. Kim S, Shephard N, Chib S. Stochastic volatility: likelihood inference and comparison with ARCH models. *Review of Economic Studies* 1998; **65**:361–393.
26. Gerlach R, Carter C, Kohn R. Diagnostics for time series analysis. *Journal of Time Series Analysis* 1999; **20**:309–330.
27. Liesenfeld R, Richard J-F. Univariate and multivariate stochastic volatility models: estimation and diagnostics. *Journal of Empirical Finance* 2003; **10**:505–531.
28. Rosenblatt M. Remarks on a multivariate transformation. *Annals of Mathematical Statistics* 1952; **23**(23):470–472.
29. Smith JQ. Diagnostic checks of non-standard time series models. *Journal of Forecasting* 1985; **4**:283–291.
30. Kupiec PH. Techniques for verifying the accuracy of risk measurement models. *The Journal of Derivatives* 1995; **3**:73–84.
31. Douc R, Moulines E, Olsson J, van Handel R. Consistency of the maximum likelihood estimator for general hidden Markov models. *The Annals of Statistics* 2011; **39**:474–513.
32. French KR, Schert WG, Stambugh RF. Expected stock return and volatility. *Journal of Financial Economics* 1987; **19**:3–29.
33. Langrock R, Michelot T, Sohn A, Kneib T. Semiparametric stochastic volatility modelling using penalized splines. *Computational Statistics* 2015; **30**(2):517–537.
34. Smith DR. A Stochastic Volatility Model with Fat Tails, Skewness and Leverage Effects. Available at: https://doi.org/10.2139/ssrn.1078625 [Accessed on 23 August 2013], 2007.

Supporting information

Additional supporting information may be found online in the supporting information tab for this article.