Journal of Empirical Finance 19 (2012) 147–161

Contents lists available at SciVerse ScienceDirect

Journal of Empirical Finance

journal homepage: www.elsevier.com/locate/jempfin

Some nonstandard stochastic volatility models and their estimation using structured hidden Markov models

Roland Langrock [a,*], Iain L. MacDonald [b], Walter Zucchini [a]

[a] Institut für Statistik und Ökonometrie, Georg-August-Universität, Platz der Göttinger Sieben 5, 37073 Göttingen, Germany
[b] Actuarial Science, University of Cape Town, 7701 Rondebosch, South Africa

ARTICLE INFO

Article history:
Received 5 August 2010
Received in revised form 16 September 2011
Accepted 22 September 2011
Available online 29 September 2011

JEL classification:
C13
C51
C53

Keywords:
State-space models
Mixture models
Financial time series
Forecasting
Pseudo-residuals
Backtesting

ABSTRACT

We introduce a number of nonstandard stochastic volatility (SV) models and examine their performance when applied to the series of daily returns on several stocks listed on the New York Stock Exchange. The nonstandard models under investigation extend both the observation process and the volatility-generating process of basic SV models. In particular, we consider dependent as well as independent mixtures of autoregressive components as the log-volatility process, and include in the observation equation a lower bound on the volatility. We also consider an experimental SV model that is based on conditionally gamma-distributed volatilities. Our estimation method is based on the fact that an SV model can be approximated arbitrarily accurately by a hidden Markov model (HMM), whose likelihood is easy to compute and to maximize. The method is close, but not identical, to those of Fridman and Harris (1998), Bartolucci and De Luca (2001, 2003) and Clements et al. (2006), and makes explicit the useful link between HMMs and the methods of those authors. Likelihood-based estimation of the parameters of SV models is usually regarded as challenging because the likelihood is a high-dimensional multiple integral. The HMM approximation is easy to implement and particularly convenient for fitting experimental extensions and variants of SV models such as those we introduce here. In addition, and in contrast to the case of SV models themselves, simple formulae are available for the forecast distributions of HMMs, for computing appropriately defined residuals, and for decoding, i.e. estimating the volatility of the process.

1. Introduction

The standard discrete-time stochastic volatility model, without leverage, for returns y_t on an asset can be written in several different forms, e.g.

$$y_t = \varepsilon_t \beta exp(g_t/2), \qquad g_{t+1} = \phi g_t + \sigma \eta_t, \qquad (t = 1, \ldots, n) \tag{1}$$

where $|\phi|<1$ and $\{\varepsilon_t\}$ and $\{\eta_t\}$ are independent sequences of independent standard normal random variables; see e.g. Shephard (1996). Following Chib et al. (2002) we use SV_0 to denote model (1). A common extension of the basic model assumes for ε_t a t distribution with ν degrees of freedom, and $\nu>0$ is then treated as an additional parameter. Again following Chib et al. (2002), we label this extension *SVt*.

Over the past two decades, stochastic volatility models such as SV_0 and *SVt* have attracted much attention in the finance literature as competitors to, *inter alia*, GARCH models (cf. Broto and Ruiz, 2004). SV models mimic several of the "stylized facts"

* Corresponding author. Centre for Research into Ecological and Environmental Modelling, The Observatory, Buchanan Gardens, University of St Andrews, St Andrews, Fife KY16 PLZ, UK. Tel.: +441334461829.
E-mail address: roland@mcs.st-and.ac.uk (R. Langrock).

0927-5398/$ – see front matter
doi:10.1016/j.jempfin.2011.09.003

attributed to asset returns: kurtosis of returns in excess of 3, zero autocorrelation of returns, and dependence of returns as revealed by the nonzero autocorrelations of squared returns. For a discussion of these stylized facts, see Taylor (2005, Chapter 4). There is evidence strongly suggesting that SV models are superior in fit to GARCH models (Danielsson, 1994).

On the other hand, SV models are not as easy to fit as GARCH models. This is because the likelihood is given by a high-order multiple integral that cannot be evaluated directly; see e.g. Jacquier et al. (1994). In the past two decades much ingenuity has been applied in the derivation of estimation methods for SV models; for a comprehensive overview of the existing methodology we recommend Broto and Ruiz (2004). Some of the most important methods are the generalized method of moments (GMM, see Melino and Turnbull, 1990), quasi-maximum-likelihood (QML, see Harvey et al., 1994), Markov chain Monte Carlo (MCMC, see Jacquier et al., 1994) and Monte Carlo likelihood (MCL, see Sandmann and Koopman, 1998). According to Shephard (2005, p.13), the methods can be categorized into those that are relatively simple but inefficient (like GMM and QML), and those that attempt to evaluate the likelihood, which are efficient but computer-intensive and rather difficult to implement (like MCMC and MCL).

In this paper we apply an alternative method, namely approximation via hidden Markov models (HMMs). The key idea, the use of iterated numerical integration, was introduced by Kitagawa (1987). In the context of stochastic volatility models it was applied by Fridman and Harris (1998), by Bartolucci and De Luca (2001, 2003), and by Clements et al. (2006), although none of these four papers explicitly makes the link between SV models and HMMs. (See also Section 13.3 of Zucchini and MacDonald (2009), where this approach is applied to an SV model with leverage.) The method involves an approximation to the SV likelihood that can be made arbitrarily accurate, and that is competitive in terms of computational effort. Advantages of the HMM formulation of the SV model 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. (For accounts of HMMs, see Ephraim and Merhav, 2002; Cappé et al., 2005, and Zucchini and MacDonald, 2009). In addition, and this is an important message of this paper, the HMM formulation makes it particularly simple to create a variety of extensions of the standard SV model that are easy to implement. We propose a number of new nonstandard SV models, in particular models with log-volatility processes $\{g_t\}$ which differ from that in Eq. (1) and involve an additional parameter representing a lower bound on the volatility. In the application that we consider, these nonstandard models appear to have some advantages. This paper therefore has two principal aims: first, a detailed investigation of the link between HMMs and numerical integration methods for SV models, and second, the theoretical and empirical analysis of some modifications or extensions of the basic SV models.

A different approach to the use of HMMs as models for share returns is that of Rossi and Gallo (2006). Unlike our work, theirs does not seek to approximate (standard or non-standard) SV models, which have a continuous-valued latent process. Their log-volatility process is a Markov chain on a finite number (e.g. seven) of equally-spaced states, and allows changes in volatility to occur one step at a time only; their transition probability matrices are tridiagonal. This is a somewhat restrictive assumption. Their model does, however, allow the transition probabilities to depend on both the sign and the magnitude of the most recent return, and does allow the returns, conditional on the volatility process, to follow an autoregressive process with innovations having a t distribution.

In Section 2 we describe how HMM techniques can be applied in order to maximize the approximate likelihood of (standard and nonstandard) SV models numerically. Section 3 introduces four nonstandard SV models. Some of these are generalizations of the basic models with additional parameters in the log-volatility process (or volatility process), which is assumed to belong to the class of conditional linear AR(1) models described by Grunwald et al. (2000). In Section 4, each of the six SV models considered is fitted to ten series of daily returns, and the relative merits of the models are assessed in terms of Akaike's information criterion (AIC) and their out-of-sample performance, especially the accuracy of their forecast distributions.

2. Model-fitting strategy

Discrete-time stochastic volatility models (without leverage) are characterized by two processes: a continuous-valued Markov state process, $\{g_t\}$, and an observation process, $\{y_t\}$, whose realizations are assumed to be conditionally independent, given the states. HMMs have the same two-process structure, except that the Markov process $\{g_t\}$ is discrete-valued instead of continuous-valued. By appropriately discretizing the state-space of an SV model into m states, the model can be approximated by an HMM. The point of using such an approximation is that, whereas the likelihood of the SV model involves a multiple integral and is difficult to compute, that of an HMM is easy to compute and to maximize.

2.1. Likelihood evaluation by iterated numerical integration

The justification (for the model SV_0, defined by Eq. (1)) of the procedure we shall use is as follows. Let the likely range of g_t-values be split into m equally-sized intervals $B_i=(b_{i-1}, b_i)$, $i=1,\ldots,m$. We denote by b_i^* a representative point in B_i, e.g. the midpoint. The likelihood of the model (1) can now be approximated as follows:

$$\begin{aligned}\mathcal{L} &= \int\ldots\int f(\mathbf{y},\mathbf{g})\mathrm{d}\mathbf{g}\\ &= \int\ldots\int f(g_1)f(y_1|g_1)\prod_{t=2}^{n} f(g_t|g_{t-1})f(y_t|g_t)\mathrm{d}g_n\ldots\mathrm{d}g_1\end{aligned} \tag{2}$$

$$\approx \sum_{i_1=1}^{m} \dots \sum_{i_n=1}^{m} \mathbb{P}\left(g_1 \in B_{i_1}\right) f(y_1 | g_1 = b_{i_1}^*) \prod_{t=2}^{n} \mathbb{P}\left(g_t \in B_{i_t} | g_{t-1} = b_{i_{t-1}}^*\right) f(y_t | g_t = b_{i_t}^*), \tag{3}$$

where f is used as a general symbol for a density.

In more detail, the innermost integral in the multiple integral (2) has been approximated as follows:

$$\begin{aligned} \int_{-\infty}^{\infty} f(g_n | g_{n-1}) f(y_n | g_n) \mathrm{d}g_n &\approx \sum_{i_n=1}^{m} \int_{B_{i_n}} f(g_n | g_{n-1}) f(y_n | g_n) \mathrm{d}g_n \\ &\approx \sum_{i_n=1}^{m} f(y_n | g_n = b_{i_n}^*) \int_{B_{i_n}} f(g_n | g_{n-1}) \mathrm{d}g_n \\ &= \sum_{i_n=1}^{m} \mathbb{P}\left(g_n \in B_{i_n} | g_{n-1}\right) f(y_n | g_n = b_{i_n}^*), \end{aligned}$$

which obviously depends on g_{n-1}. Note that we have here used an approximation of the form

$$\int_a^b f_1(x) f_2(x) \mathrm{d}x \approx f_1(c) \int_a^b f_2(x) \mathrm{d}x,$$

where c is a representative point in (a,b); throughout the interval (a,b), the function f_1 (only) has been replaced by its value at c. This is by no means the only way in which the integral $\int_{B_{i_n}} f(g_n | g_{n-1}) f(y_n | g_n) \mathrm{d}g_n$ could be approximated; we indicate another possibility below. The next integral we require is

$$\int_{B_{i_{n-1}}} f(g_{n-1} | g_{n-2}) f(y_{n-1} | g_{n-1}) \mathbb{P}\left(g_n \in B_{i_n} | g_{n-1}\right) \mathrm{d}g_{n-1},$$

which we approximate by

$$\sum_{i_{n-1}=1}^{m} \mathbb{P}\left(g_{n-1} \in B_{i_{n-1}} | g_{n-2}\right) f\left(y_{n-1} | g_{n-1} = b_{i_{n-1}}^*\right) \mathbb{P}\left(g_n \in B_{i_n} | g_{n-1} = b_{i_{n-1}}^*\right).$$

The result of the first two integrations is therefore

$$\sum_{i_{n-1}=1}^{m} \sum_{i_n=1}^{m} \mathbb{P}\left(g_{n-1} \in B_{i_{n-1}} | g_{n-2}\right) f(y_{n-1} | g_{n-1} = b_{i_{n-1}}^*) \mathbb{P}\left(g_n \in B_{i_n} | g_{n-1} = b_{i_{n-1}}^*\right) f(y_n | g_n = b_{i_n}^*).$$

After $n-2$ further integrations we get the multiple sum expression (3) as the approximation for the likelihood.

The other possibility referred to above is

$$\int_a^b f_1(x) f_2(x) \mathrm{d}x \approx (b-a) f_1(c) f_2(c); \tag{4}$$

this is an approximation of the type used by Bartolucci and De Luca (2001, 2003). Here both f_1 and f_2 have been replaced by their values at c, and the interval width therefore appears as a factor in the approximation. A third possibility is to use Gauss–Legendre quadrature to evaluate integrals such as $\int_{-\infty}^{\infty} f(g_n | g_{n-1}) f(y_n | g_n) \mathrm{d}g_n$. In this case the quadrature points are not equally spaced and the weights attached to the function evaluations are not equal. Gauss–Legendre quadrature is the approach followed by Fridman and Harris (1998). One thing that neither Fridman and Harris (1998) nor Bartolucci and De Luca (2001, 2003) mentioned explicitly is the link between the approximated SV likelihood (given here by (3)) and HMMs. In Section (2.2) we describe that link.

But note also the discretized non-linear filtering (DNF) procedure of Clements et al. (2006). Their Eq. (11), together with the prediction and update steps (9) and (10), implies a multiple-sum expression for the approximate likelihood of an SV model which is of exactly the same form as our expression (3). The only difference is that they use an approximation of the type (4); see their Eq. (6), in which the interval width appears as a factor.

2.2. Formulation in terms of hidden Markov models

We define $\{h_t\}$ to be the (discrete-time, homogeneous) Markov chain with transition probabilities $\gamma_{ij} = \mathbb{P}(h_t = j | h_{t-1} = i)$ defined by

$$\gamma_{ij} := \Phi\left(\frac{b_j - \phi b_i^*}{\sigma}\right) - \Phi\left(\frac{b_{j-1} - \phi b_i^*}{\sigma}\right), \tag{5}$$

where b_i^* is the representative point of the interval $B_i = (b_{i-1}, b_i)$ and Φ denotes the cumulative distribution function of the standard normal distribution. This is exactly the probability

$$\mathbb{P}\left(g_t \in B_j | g_{t-1} = b_i^*\right)$$

needed in the expression (3), for values of t from 2 to n. If for all i from 1 to m we define δ_i to be $\mathbb{P}(g_1 \in B_i)$, then the multiple sum (3) equals the matrix product

$$\boldsymbol{\delta}\mathbf{P}(y_1)\boldsymbol{\Gamma}\mathbf{P}(y_2)\boldsymbol{\Gamma}\cdots\boldsymbol{\Gamma}\mathbf{P}(y_{n-1})\boldsymbol{\Gamma}\mathbf{P}(y_n)\mathbf{1}', \tag{6}$$

where $\mathbf{1}'$ is a column vector of ones and $\mathbf{P}(y_t)$ is the diagonal matrix with ith diagonal entry the normal density with mean 0 and variance $\beta^2\exp(b_i^*)$, evaluated at y_t. That is, the sum (3) is precisely the likelihood of an HMM, based on the Markov chain $\{h_t\}$ with initial distribution the row vector $\boldsymbol{\delta}$. We can take the initial distribution $\boldsymbol{\delta}$ to be either the stationary distribution implied by the t.p.m. $\boldsymbol{\Gamma}$, or the stationary distribution of the AR(1) process $\{g_t\}$, discretized into the m intervals B_i.

It is then a routine matter to evaluate the matrix product (6) and to maximize it with respect to the parameters ϕ, σ and β. A technical detail that needs attention is that each of the three parameters is constrained: $-1<\phi<1$, $\sigma>0$ and $\beta>0$. Either one can apply *constrained* numerical maximization or, as was done in the applications in Section 4, one can reparametrize the model in terms of unconstrained "working parameters", namely $\log((1+\phi)/(1-\phi))$, $\log \sigma$ and $\log \beta$ in the case of the SV_0 model, and then maximize the likelihood with respect to those parameters. In this case, approximate confidence intervals for the "natural parameters" (ϕ, σ and β) can be obtained by first estimating confidence intervals for the working parameters from the inverse of the estimated information matrix, and then applying the corresponding inverse transformations to the interval boundaries for the working parameters. Alternatively the parametric bootstrap method can be applied.

In practice one has to decide what value of m, the number of states, will be adequate, and what range of g_t-values to allow for. The minimum and maximum values (g_{min} and g_{max}) for g_t have to be chosen sufficiently large to cover the essential domain of the state process, but not too large, in order to maintain sufficient fineness of the grid. Fridman and Harris (1998) suggest using $-g_{min}=g_{max}=3\sigma_g$, where σ_g denotes the standard deviation of the log-volatility process. Our experience suggests that this range is often too narrow, which confirms the findings of Bartolucci and De Luca (2001), who suggest using $-g_{min}=g_{max}=5\sigma_g$. In any case it is useful, when one has fitted a model, to examine the stationary distribution of g_t and compare its essential range to the range allowed for in the fitting process. For the SV_0 model this stationary distribution is given by $\mathrm{N}(0,\sigma^2/(1-\phi^2))$. Even if the latent process is not a Gaussian AR(1), this check can still be carried out by using the stationary distribution implied by the t.p.m. of the approximating HMM.

The choice of m has a strong influence on the accuracy of the approximation. The accuracy improves as m increases, but the size of the matrices in Eq. (6) also increases, which slows down the evaluation of the likelihood. Note that, although m needs to be large enough to provide a good approximation, the number of model parameters does not depend on m; the entries of the $m\times m$ matrix $\boldsymbol{\Gamma}$ depend only on ϕ and σ, and those of the matrix $\mathbf{P}$ depend only on β and the observations.

2.3. A simulation experiment

We describe here a simulation experiment of an SV_0 model fitted (i) by the MCL method, which is implemented in `Ox` in `ssfpack` (Koopman et al., 1999), and (ii) by the HMM method, implemented in `R`, for several values of m. Our `R` code to compute the (approximate) log-likelihood of the model is given in Appendix B.

The model was fitted to a simulated series of $n=10{,}000$ observations. The parameters were set at $\phi=0.98$, $\sigma=0.2$ and $\beta=0.05$; the starting values $\phi_0=0.9$, $\sigma_0=0.3$ and $\beta_0=0.2$ were used for both methods. Table 1 gives an indication of the influence of m on accuracy and computing time.

The results in Table 1, as well as those obtained for many observed series of returns, and for generated series that have similar properties, lead us to conclude that the parameter estimates obtained by the HMM method stabilize for m-values somewhere between 50 and 100. Secondly, for values of $m\leq 100$ the HMM method is comparable with the MCL method in terms of computing time. We note that the likelihood of the approximating HMM is larger than that obtained using MCL.

Table 1
SV_0 model: estimated parameters and computing times for MCL and HMM method ($-g_{min}=g_{max}=4$, true parameters: $\phi=0.98$, $\sigma=0.2$, $\beta=0.05$); 95% confidence intervals in parentheses.

	$\log \mathcal{L}$	time (s)	$\hat{\phi}$	$\hat{\sigma}$	$\hat{\beta}$
MCL	15,089.5	96	0.983 (0.975;0.989)	0.198 (0.178;0.219)	0.050 (0.044;0.057)
HMM					
$m=30$	15,095.6	20	0.983 (0.977;0.987)	0.188 (0.169;0.208)	0.050 (0.045;0.056)
$m=50$	15,095.6	31	0.983 (0.977;0.987)	0.198 (0.180;0.217)	0.050 (0.045;0.056)
$m=100$	15,095.6	79	0.983 (0.977;0.987)	0.202 (0.184;0.221)	0.050 (0.045;0.056)
$m=200$	15,095.6	256	0.983 (0.977;0.987)	0.202 (0.184;0.221)	0.050 (0.045;0.056)

An important motivation for applying the HMM formulation is that all kinds of extensions of the standard SV model are easy to implement by simply modifying a few lines of code for the computation of $\mathbf{\Gamma}$ and $\mathbf{P}$ in expression (6). This convenient feature of the HMM formulation is exploited in Section 4 in order to fit the nonstandard SV models that we introduce in Section 3.

Fig. 1 displays the first 500 simulated volatilities as well as their estimates using the two methods. For the MCL method the estimates were obtained using the `Ox`-package `ssfpack`; those for the HMM method were computed using the Viterbi algorithm (see Chapter 5 of Zucchini and MacDonald, 2009). The two estimates are almost identical.

2.4. Forecasts and model checking

The use of approximating HMMs provides a convenient tool for obtaining forecast distributions for SV models. For example, it is straightforward to find, for the approximating HMM, the cumulative distribution function of the one-step-ahead forecast distribution on day $t-1$, i.e. the conditional distribution of the return on day t, given all previous observations. This is given by

$$F(y_t|y_{t-1},y_{t-2},\dots,y_1)=\sum_{i=1}^{m}\zeta_i F(y_t|h_t=i), \tag{7}$$

where ζ_i is the ith entry of the vector $\boldsymbol{\alpha}_{t-1}\mathbf{\Gamma}/(\boldsymbol{\alpha}_{t-1}1')$, obtained from the “forward probabilities”

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

with $\boldsymbol{\delta}$, $\mathbf{P}(y_k)$ and $\mathbf{\Gamma}$ defined as above. The corresponding expression for longer forecast horizons is similar (see Chapter 5 of Zucchini and MacDonald, 2009). A closed-form expression for obtaining state predictions, i.e. volatility predictions, is also available.

The forecast distribution given by Eq. (7) can be used in order to perform model checking via residuals. The one-step-ahead forecast “pseudo-residual” (or quantile residual) is given by

$$r_t=\Phi^{-1}(F(y_t|y_{t-1},y_{t-2},\dots,y_1)). \tag{8}$$

In the context of SV models such residuals were first used by Kim et al. (1998). It follows immediately from a result of Rosenblatt (1952) that, if the fitted model is correct, the pseudo-residuals are distributed standard normal. (See also Zucchini and MacDonald, 2009, Chapter 6.) Thus forecast pseudo-residuals can be used to monitor time series; extreme values can be identified, and the continued suitability of the model can be checked by using, for example, qq-plots or formal tests for normality. An example of this process is given in Section 4.2.

3. Some nonstandard SV models

3.1. Shifting the volatility process

The models SV_0 and SVt can be generalized by introducing a lower bound on the volatility of the observed process. For instance, the observation equation in the model (1) can be replaced by

$$y_t=\varepsilon_t(\beta exp(g_t/2)+\xi). \tag{9}$$

The additional parameter $\xi(\geq 0)$ does appear to be worthwhile (cf. Section 4), and is plausible on the grounds that some baseline volatility is always present. In all the models that will be presented in Sections 3.2–3.5 we incorporate this additional parameter. Of course, the model with $\xi=0$ is in all cases nested in the model with $\xi\geq 0$.

In all models covered in this section, ε_t is assumed to follow a t distribution with ν degrees of freedom. We also fitted the models with Gaussian ε_t but the details are not reported here.

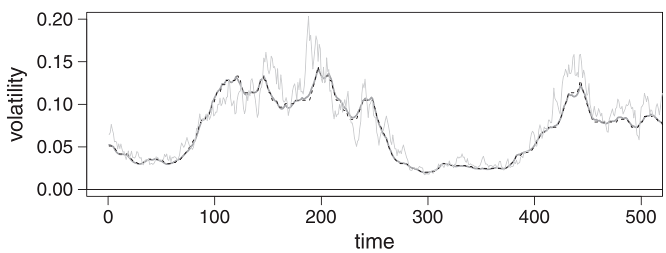

Fig. 1. Volatility and volatility estimates associated with the first 500 simulated observations (light grey line: true volatility, $\beta\exp(g_t/2)$; dashed line: volatility as estimated by the HMM method and Viterbi; thick dark grey line: volatility as estimated by the MCL method).

3.2. SVMt: using a mixture of AR(1) processes for the log-volatility

The estimates of the parameter ϕ in the SV_0 and SVt models are generally only a little smaller than one. (See, e.g., Tables A.6 and A.7). Consequently the sample paths of the Gaussian AR(1) model used to describe the log-volatility process are relatively smooth, in the sense that σ^2, the conditional variance of g_t given g_{t-1}, is much smaller than the marginal variance of g_t, namely $\sigma^2/(1-\phi^2)$. On the other hand, plots (e.g. those in Fig. 2) indicate that the level of volatility can be maintained for an extended period and then change abruptly. One possible way to model this type of behaviour is to generalize the *SVt* model by using not one but a mixture of two Gaussian AR(1) processes for the log-volatilities. We begin by defining a model based on an independent mixture and then, in the next section, consider the case of a Markov-dependent mixture.

Let y_t be given by Eq. (9), but now assume that, given g_t, g_{t+1} is distributed either $N(\phi_1 g_t, \sigma_1^2)$ (with probability α) or $N(\phi_2 g_t, \sigma_2^2)$ (with probability $1-\alpha$). Equivalently,

$$g_{t+1} = \begin{cases} \phi_1 g_t + \sigma_1 \eta_t & \text{with probability } \alpha \\ \phi_2 g_t + \sigma_2 \eta_t & \text{with probability } 1-\alpha, \end{cases} \tag{10}$$

with the innovations η_t being independent standard normal. This model, labelled *SVMt*, allows for abrupt changes in the log-volatility process and thus offers additional flexibility. The *SVt* model is nested in *SVMt*; consider the case $\alpha=1$ and $\xi=0$. One could also consider using a mixture with more than two AR(1) components, but that generalization is not pursued here.

Wong and Li (2000) give the following necessary and sufficient condition for second-order stationarity of $\{g_t\}$:

$$\alpha\phi_1^2 + (1-\alpha)\phi_2^2 < 1.$$

Note that it is possible for one of the AR(1) processes to be "explosive" (e.g. $\phi_2=1.4$) without destroying the second-order stationarity of the mixed process. The stationary mean of $\{g_t\}$ is 0 and the stationary variance of $\{g_t\}$ is

$$\sigma_g^2 = \frac{\alpha\sigma_1^2 + (1-\alpha)\sigma_2^2}{1-(\alpha\phi_1^2 + (1-\alpha)\phi_2^2)}.$$

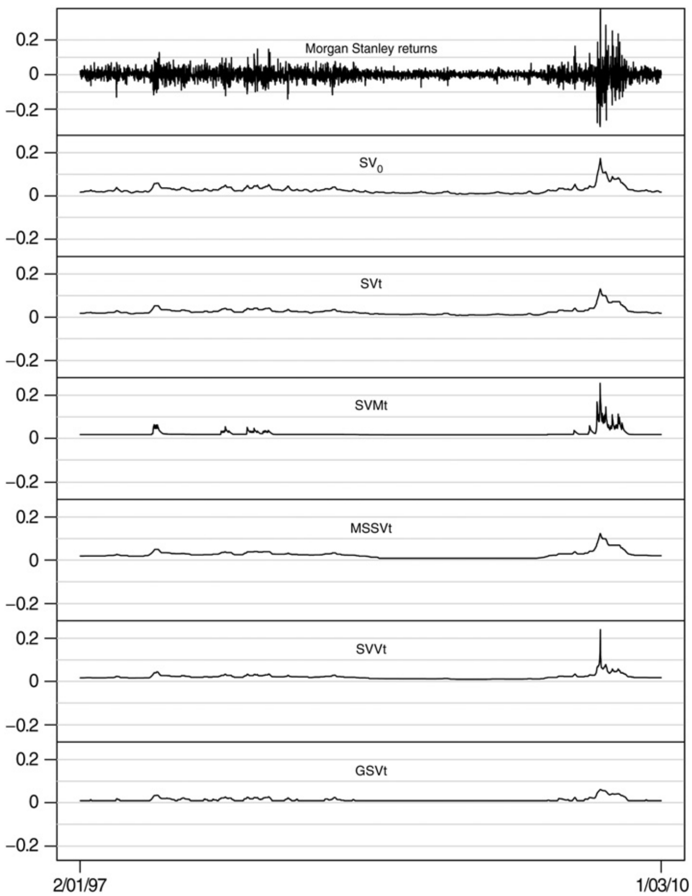

Fig. 2. Estimated volatilities for Morgan Stanley; one of the returns (0.63 on 13.08.08) fell outside the range of the top graph.

As there is no closed-form expression for the marginal distribution of g_t, the exact stationary variance and kurtosis of the observed process, $\{y_t\}$, are not available via that route. Second-order Taylor approximations for these moments are given by

$$\mathrm{var}(y_t) \approx \frac{\nu}{\nu-2}\left((\beta+\xi)^2 + c_1\sigma_g^2\right) \tag{11}$$

and

$$\mathrm{kurtosis}(y_t) \approx 3\frac{\nu-2}{\nu-4}\frac{(\beta+\xi)^4 + c_2\sigma_g^2}{\left((\beta+\xi)^2 + c_1\sigma_g^2\right)^2}, \tag{12}$$

where $c_1 = 0.5\beta^2 + 0.25\beta\xi$ and $c_2 = 3\beta^2\xi^2 + 2\beta^4 + 2.25\beta^3\xi + 0.5\beta\xi^3$.

3.3. MSSVt: Markov-switching innovations in the log-volatility process

Although the model *SVMt* allows for abrupt changes in the (two-state) process generating the log-volatilities, it assumes that these changes occur independently. This assumption may be unrealistic; some persistence in the process seems plausible, and should at least be allowed for. We therefore relax the independence assumption by replacing the independent mixture of AR components in the log-volatility by a dependent mixture, in particular a Markov-dependent mixture. Such a generalization allows the sojourn times in each of the two states underlying the log-volatilities to be (stochastically) longer than those implied by the *SVMt* model.

Let y_t again be given by Eq. (9), and assume that

$$g_{t+1} = \phi g_t + \sigma_{\alpha_t}\eta_t,$$

with $\{\alpha_t\}$ being a two-state Markov chain described by the t.p.m. $\mathbf{\Gamma}^{(\alpha)} = (\gamma_{ij}^{(\alpha)})$. The model allows for two different variances in the innovations — σ_1^2 and σ_2^2 — that are selected by the Markov chain $\{\alpha_t\}$. This model, which we label *MSSVt*, is similar to that of So et al. (1998). However, So et al. assume the innovation variance σ^2 to be constant and instead model a reparametrized β nonhomogeneously via $\{\alpha_t\}$. Clearly *SVt* is the special case of *MSSVt* with $\gamma_{11}^{(\alpha)} = \gamma_{21}^{(\alpha)} = 1$ and $\xi = 0$.

Stationarity of $\{g_t\}$ and $\{y_t\}$ is possible if and only if $|\phi| < 1$. The stationary variance of $\{g_t\}$ is

$$\sigma_g^2 = \frac{\delta_1^{(\alpha)}\sigma_1^2 + \delta_2^{(\alpha)}\sigma_2^2}{1-\phi^2},$$

where $\boldsymbol{\delta}^{(\alpha)}$ is the stationary distribution of the Markov chain $\{\alpha_t\}$; $\delta_i^{(\alpha)}$ is therefore the expected proportion of time that $\{\alpha_t\}$ spends in state i. The expressions (11) and (12) give the approximate stationary variance and kurtosis of $\{y_t\}$.

To fit the model we consider the process

$$z_t := \begin{pmatrix} g_t \\ \alpha_t \end{pmatrix},$$

which is a Markov process on $\mathbb{R}\times\{0, 1\}$. The component $\{g_t\}$ is discretized into m states, as described in Section 2, and $\{\alpha_t\}$ takes on one of two values, so the number of states of $\{z_t\}$, after discretization, is $2m$. Writing the t.p.m. of $\{z_t\}$ in terms of the model parameters, it is then straightforward to maximize the likelihood, which is given by Eq. (6).

In principle the model can be further generalized in a number of ways that we will not discuss in detail. One can allow the parameter ϕ also to depend on the state of $\{\alpha_t\}$. However, stationarity conditions then become more involved, and parameter estimation for this model proved to be unstable in practice. One can allow the parameters β, ν and ξ to depend on the current state of $\{\alpha_t\}$. The extension to more than two states for $\{\alpha_t\}$ is also easy to implement. Of course an increase in the number of states leads to an increase in the size of the t.p.m., and hence in the computational burden.

3.4. SVVt: nonhomogeneous innovations in the log-volatility process

The models *SVMt* and *MSSVt* offer substantially more flexibility in the log-volatility process than does *SVt*, but at the cost of three additional parameters. The model presented in this section, *SVVt*, involves only one additional parameter in the log-volatility process; it offers a compromise between the parsimony of *SVt* and the flexibility of *SVMt* and *MSSVt*.

Let y_t be given by Eq. (9), but now assume that

$$g_{t+1} = \phi g_t + \sigma_t\eta_t, \tag{13}$$

where $\sigma_t = \sqrt{\omega + \gamma exp(g_t)}$ with $\omega, \gamma > 0$ and $\eta_t \overset{iid}{\sim} N(0, 1)$. The motivation for this model is as follows. In the standard SV model (1) the innovations η_t can be interpreted as shocks to the intensity of the news flow (see Franses and van Dijk, 2000). Model (13) allows for possible influence of g_t, the (log-)volatility at time t, on the magnitude of such shocks at time $t+1$. High volatility at time t indicates that the markets are turbulent which, in turn, could impact on the flow of news at time $t+1$. The parameter

σ_t measures the uncertainty about future volatility, and this uncertainty can be expected to increase if the markets are nervous. Clearly the simpler model *SVt* is nested in *SVVt* (the case with $\gamma=0$).

The nonlinear influence of g_t on the variance of the innovations makes it difficult to derive necessary and sufficient conditions for second-order stationarity of the *SVVt* model. A Taylor expansion provides two approximate necessary conditions for second-order stationarity of $\{g_t\}$: $|\phi|<1$ and $\gamma<2(1-\phi^2)$. Simulation experiments suggest that these conditions provide useful approximations but that the stated range for γ is slightly conservative. This is theoretically unsatisfactory, but fortunately it is straightforward to check for stationarity of the *discretized SVVt* model, i.e. the one obtained after discretization of $\{g_t\}$. The log-volatility process is then a Markov chain with finite state space, and so stationarity holds if the initial distribution of the Markov chain, $\boldsymbol{\delta}$, is such that $\boldsymbol{\delta\Gamma}=\boldsymbol{\delta}$.

Using a Taylor expansion we can obtain the approximate stationary variance of $\{g_t\}$:

$$\sigma_g^2 \approx \frac{\omega+\gamma}{1-0.5\gamma-\phi^2}.$$

The stationary variance and kurtosis of $\{y_t\}$ can be approximated using expressions (11) and (12).

3.5. GSVt: a model with conditionally gamma-distributed volatility

Volatilities are, by definition, positive. In order to satisfy this restriction, the state processes considered so far have all modelled the *log-volatilities* (using Gaussian distributions in several different ways). However, with the HMM method it is equally easy to use a state process consisting of positive-valued random variables, i.e. to model the volatilities themselves rather than the log-volatilities. This does not lead to additional difficulties in the fitting of models. We illustrate this point with a model based on conditionally gamma-distributed volatilities.

Let y_t be defined by

$$y_t = \varepsilon_t \beta \sqrt{g_t + \xi}, \tag{14}$$

with ε_t again denoting a t distribution, and, conditional on g_t, let g_{t+1} have a gamma distribution with shape parameter $\phi g_t + \lambda$ and scale parameter 1:

$$g_{t+1} \sim \Gamma(\phi g_t + \lambda, 1). \tag{15}$$

The parameters β, ϕ, λ and ξ are all taken to be positive. We refer to this model as *GSVt*. If $\{g_t\}$ is stationary, its stationary mean is

$$\mu_g = \frac{\lambda}{1-\phi},$$

and the corresponding stationary variance is

$$\sigma_g^2 = \frac{\lambda}{(1-\phi)(1-\phi^2)}.$$

Provided $\{g_t\}$ is stationary, one obtains

$$\mathrm{var}(y_t) = \beta^2 \left(\mu_g + \xi\right) \frac{\nu}{\nu-2}$$

and

$$\mathrm{kurtosis}(y_t) = 3\frac{\nu-2}{\nu-4}\left(1 + \frac{\mu_g}{\left(\mu_g+\xi\right)^2(1-\phi^2)}\right).$$

A sufficient condition for stationarity is that $\phi \in [0,1)$; see Proposition 3 of Grunwald et al. (2000).

4. Application

4.1. Model comparisons based on ten series of returns

The above methodology was applied in order to model the daily returns for ten stocks on the New York Stock Exchange, namely Sony Corporation, Time Warner, Toyota Motor Corporation, The Travelers Companies, British Petroleum plc, Royal Dutch Shell plc, Bank of America Corporation, Citigroup Inc., Deutsche Bank AG and Morgan Stanley. The adjusted closing prices, p_t, for the period 02.01.1997–01.03.2010, were downloaded from “finance.yahoo.com”, and the daily returns were computed as

Table 2
Summary statistics for the daily returns of ten stocks on the New York Stock Exchange for the period 02.01.1997–01.03.2010.

	n	min.	max.	Std. dev.	Kurtosis
Sony	3304	−0.155	0.169	0.024	7.8
Time Warner	3310	−0.188	0.165	0.031	7.6
Toyota	3304	−0.181	0.133	0.020	8.7
Trav. Comp.	3304	−0.200	0.228	0.022	14.2
BP	3310	−0.122	0.147	0.018	9.7
Roy. D. Sh.	3303	−0.121	0.161	0.019	9.4
Bank of Am.	3310	−0.342	0.302	0.033	26.8
Citigroup	3310	−0.495	0.457	0.036	36.2
Deu. Bank	3293	−0.210	0.222	0.028	13.1
Morgan St.	3310	−0.299	0.626	0.036	41.2

$y_t = \log(p_t/p_{t-1})$, $t = 1,\ldots,n$. Summary statistics of the resulting ten series are given in Table 2. Not surprisingly, in view of the recent financial crisis, the sample standard deviations and kurtoses are high for stocks in the financial sector.

The six models covered in Sections 1 and 3 were fitted to each of the ten series. The maximum likelihood estimates are given in Tables A.6–A.11 in Appendix A. Several things are noteworthy regarding the parameter estimates (including some estimates that are not given in the tables).

- It is striking that, for all series, one of the AR(1) components of the *SVMt* model is nonstationary, i.e. has $\phi > 1$, although the mixture (10), and hence also the model for the observations, is stationary.
- With a single exception, the estimates of the parameter ξ, which constitutes a lower bound on the volatility, are all well above zero. (Fitting *SVMt* to the Citigroup series yielded $\hat{\xi} \approx 0$.) This suggests that the inclusion of such a lower bound is worthwhile.
- All models were fitted with both Gaussian and t distributions for ε_t. The latter consistently led to a substantially higher likelihood. The estimates of the parameter ν, the number of degrees of freedom, range from 7 to 23 across all series and models. This generalization of Gaussian SV models appears to be particularly fruitful.
- In the *MSSVt* model, the diagonal entries of $\hat{\boldsymbol{\Gamma}}^{(\alpha)}$, the estimated t.p.m. of the Markov chain $\{\alpha_t\}$, are usually close to one. (An exception is the estimate 0.643 in the case of the Sony series.) This indicates that the two states, which reflect high and low uncertainty about future volatility, are usually strongly persistent.

When we compare the models in terms of their AIC values, given in Table 3, the main results from the model-fitting exercise are as follows:

- For each of the ten series, SV_0 is inferior to the models *SVt*, *SVMt*, *MSSVt* and *SVVt*.
- In every case either *SVMt* or *MSSVt* performed best.
- For stocks that were relatively mildly affected by the financial crisis (Sony, Time Warner, Toyota), there is little difference in the performance of the models *SVt*, *SVMt*, *MSSVt* and *SVVt*.
- For stocks that were more strongly affected by the crisis, the nonstandard models *SVMt*, *MSSVt* and *SVVt* outperformed their simpler competitors.
- The *GSVt* model mainly yielded poor fits, and worse than SV_0 in seven cases.

In each of the six models, the return y_t is of the form ε_t times a factor which is essentially the volatility (see e.g. Eq. (9)). Fig. 2 displays the returns for Morgan Stanley (top panel) and the associated estimates of the factors for each of the six models. Notice that the factor for *SVt* is smaller than that for SV_0; that is because the t-distribution has fatter tails. The graph corresponding to the

Table 3
For the model SV_0 the AIC is given. The remaining entries in the table are AIC deviations from the AIC of the SV_0 model for the corresponding series. For example, in the case of Sony, the AIC for *SVt* is given by $-16179 - 33 = -16212$. Entries displayed in bold font indicate the model with the lowest AIC.

	SV_0	*SVt*	*SVMt*	*MSSVt*	*SVVt*	*GSVt*
# parameters	3	4	8	8	6	6
Sony	−16,179	−33	−31	**−35**	−33	−17
Time Warner	−15,472	−29	**−41**	−30	−33	−3
Toyota	−17,321	−13	−19	**−22**	−15	6
Trav. Comp.	−17,324	−35	−50	**−52**	−46	−27
BP	−18,043	−9	−30	**−33**	−28	8
Roy. D. Sh.	−17,721	−7	**−32**	−27	−27	18
Bank of Am.	−17,080	−29	−47	**−50**	−46	90
Citigroup	−16,249	−29	**−53**	−31	−43	111
Deuts. Bank	−15,955	−46	−49	**−55**	−51	1
Morgan St.	−14,955	−19	**−37**	−33	−35	58

SVMt model is less smooth than that of the dependent mixture model *MSSVt*. Notice also that, in both of these cases, the factor is constant at its minimum value ξ for substantial periods. The volatility of the nonstandard model *SVVt* is similar to that of *MSSVt* and *SVt* in periods of low volatility, but seems to respond more sharply to a volatility shock. In contrast, the model *GSVt* seems less responsive to shocks than any of the other models.

4.2. Out-of-sample analysis of forecast pseudo-residuals

We now perform an out-of-sample analysis of the forecast performance of the six models covered in Sections 1 and 3. The observation period is 02.01.97–01.03.10, as before, but for each return series the data are now divided into a calibration and a validation sample:

- Calibration sample (in-sample period): 02.01.97–08.08.07,
- Validation sample (out-of-sample period): 09.08.07–01.03.10.

The dividing date (09.08.07) has been identified as the beginning of the current financial crisis (see e.g. Swiss National Bank, 2008). That date was chosen here in order to assess how well the different SV models would have performed during the crisis, a period of unusually high volatility.

As a first step, each of the six models was fitted to the calibration sample of each series. This was done 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 of the 644 observations in the validation sample, the (one-step-ahead forecast) pseudo-residual was computed according to Eq. (8). As described in Section 2.4, non-normality of these residuals is an indication of mis-specification of the corresponding model.

The *p*-values for the Jarque–Bera tests applied to the pseudo-residuals are listed in Table 4. As an example, the qq-plots for the Morgan Stanley series are given in Fig. 3. We first concentrate on the Morgan Stanley series. Except for *MSSVt* and *SVVt* the qq-plots indicate a lack of fit in the tails; the models were unable to capture the extreme returns that were observed during the financial crisis. The fit of the SV_0 and *GSVt* models appears to be especially poor, an impression that is confirmed in both cases by the *p*-values of the Jarque–Bera test. The performance of the *SVt* model is somewhat better but still unsatisfactory; at the 10% level of significance, normality of the residuals for this model is also rejected by the Jarque–Bera test. In contrast, the *p*-values for the nonstandard SV models *SVMt*, *MSSVt* and *SVVt* are well above the conventional significance levels. In particular the results for the six-parameter model *SVVt* are surprisingly good, considering the turbulent behaviour of the Morgan Stanley return during the crisis.

In the case of the basic model SV_0, the Jarque–Bera test rejects normality for three other series at the 5% level, and for one more at the 10% level. Both the models *SVt* and *SVMt* lead to one rejection at the 10% level, while there is no rejection for the models *MSSVt* and *SVVt*. With respect to the residuals the *GSVt* model seems least suitable, with five rejections at the 5% level.

4.3. Backtesting

The plots and tests discussed in the previous section are useful for assessing the overall fit of a model but, for the purposes of assessing market risk, 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 one-day 1% – VaR is computed from the 0.01–quantile of the one-day-ahead forecast distribution. 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 α%–VaR, the number of exceptions, X, in n days is binomially distributed with parameters n and α. This distributional result makes it possible to implement *backtesting*, a procedure applied by major central banks and regulatory authorities in terms of the Basel Accords (Basel Committee on Banking Supervision, 2006). Specifically (Annex 10a, pp. 310–319), the adequacy of the time series model is assessed by the number of exceptions, X. Three zones are defined:

- *green zone:* X falls below the 95th percentile of its distribution, in which case the model is regarded as accurate;
- *red zone:* X falls above its 99.99th percentile, in which case the model is regarded as inaccurate;

Table 4
p-values of Jarque–Bera tests applied to one-step-ahead ahead forecast pseudo-residuals.

	SV_0	*SVt*	*SVMt*	*MSSVt*	*SVVt*	*GSVt*
Sony	0.235	0.323	0.232	0.253	0.262	0.395
Time Warner	0.761	0.242	0.171	0.165	0.167	0.110
Toyota	0.607	0.513	0.317	0.428	0.338	0.778
Trav. Comp.	0.105	0.849	0.759	0.779	0.792	0.422
BP	0.008**	0.093*	0.218	0.276	0.244	0.001**
Roy. D. Sh.	0.071*	0.273	0.064*	0.118	0.131	0.005**
Bank of Am.	0.033**	0.590	0.991	0.943	0.719	0.000**
Citigroup	0.196	0.893	0.620	0.881	0.864	0.000**
Deuts. Bank	0.011**	0.659	0.631	0.726	0.667	0.152
Morgan St.	0.000**	0.052*	0.380	0.346	0.807	0.000**

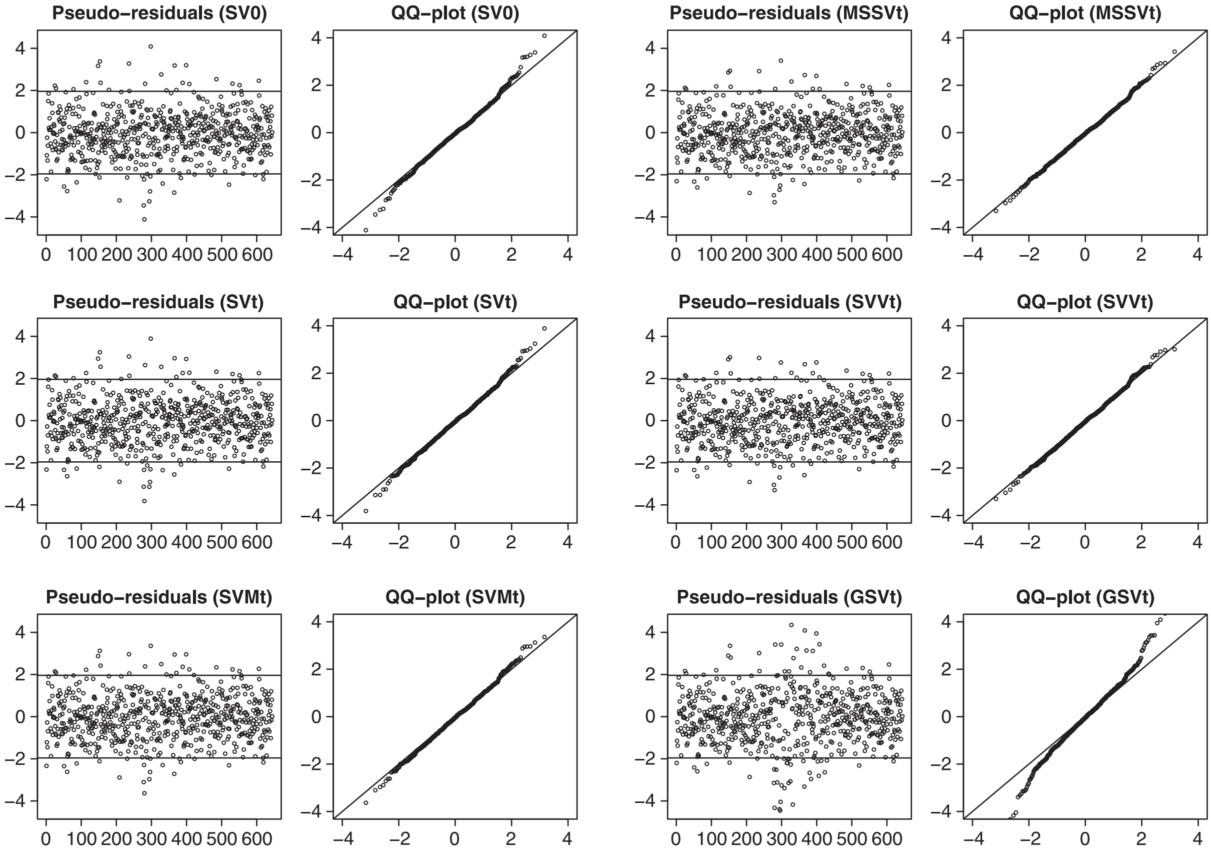

Fig. 3. Forecast pseudo-residuals for SV_0, *SVt*, *SVMt*, *MSSVt*, *SVVt* and *GSVt* for the Morgan Stanley series.

- *yellow zone: X* falls between the above percentiles, in which case "the supervisor should encourage a bank to present additional information about its model before taking action […]" (p. 315).

(For a recent account of backtesting see Wong, 2010.) Table 5 lists the number of exceptions in $n = 644$ days with $\alpha = 0.01$ for each series investigated in Section 4.1, and for each model fitted to the calibration samples as in Section 4.2. Backtesting was applied "out-of-sample", using the conditional forecast distribution computed according to Eq. (7). In each case it is indicated whether the observed counts fell into the green, red or yellow zone.

Certain aspects of the backtests in Table 5 are noteworthy. All models are in the green zone for Sony and Time Warner. For all other series and models the number of exceptions exceeded 6.4, the expected value under the hypothesis that the model is correct. In retrospect this underestimation of the market risk is not surprising in view of the enormous increase in volatility in the out-of-sample period, which had been preceded by a prolonged period of low volatility. The highest number of exceptions occurred in the financial sector. Of the models considered, the *SVVt* model led to the "least poor" results, with two outcomes

Table 5
Backtesting: The number of exceptions in $n = 644$ out-of-sample daily returns with $\alpha = 0.01$. Counts less than 11 fall in the green zone; those above 17 (marked **) fall in the red zone; the remainder (marked *) fall in the yellow zone.

	SV_0	*SVt*	*SVMt*	*MSSVt*	*SVVt*	*GSVt*
Sony	7	4	4	5	4	6
Time Warner	5	4	4	3	3	3
Toyota	10	11*	10	9	9	11*
Trav. Comp.	12*	8	8	7	8	14*
BP	13*	13*	9	9	9	18**
Roy. D. Sh.	14*	10	10	9	9	21**
Bank of Am.	19**	13*	11*	14*	10	46**
Citigroup	14*	13*	15*	10	11*	39**
Deuts. Bank	16*	12*	10	13*	11*	31**
Morgan St.	11*	10	9	8	8	21**

(Citibank and Deutsche Bank) on the lower boundary of the yellow zone, and all others in the green zone. The models *MSSVt* and *SVMt* also led to two outcomes in the yellow zone and eight in the green zone but, on the whole, the numbers of exceptions were a little higher than for *SVVt*. The next best, in terms of backtesting, was the model *SVt*, which led to five outcomes in the yellow zone and five in the green zone. The model SV_0 led to even more exceptions, although all but one of the outcomes were in the green or yellow zone. Model *GSVt*, with six outcomes in the red zone, is clearly unsuitable. These results indicate that, for the purposes of assessing market risk, it is possible to improve on SV_0 and *SVt*.

5. Concluding remarks

The evidence presented here supports the claim that nonstandard SV models can outperform the standard SV models SV_0 and *SVt* in terms of the AIC, goodness of fit (as assessed by the behaviour of residuals) and also the type of backtesting that is applied by central banks and regulatory authorities in order to assess the accuracy of models in terms of the Basel Accords. It is of course by no means guaranteed that the results obtained in this particular application will carry over to other series of returns and other observation periods, in particular periods that do not include events as extreme as the financial crisis that began in 2007. Nevertheless, we believe that series of daily returns are sufficiently long to make it worthwhile to “invest” in the few additional parameters required for such extensions, in the hope of improving the fit and the forecasting performance. Only the *GSVt* model proved to be unsuitable for most of the return series considered. We included it nevertheless in order to demonstrate that even such an unusual type of SV model can be fitted without difficulty by using the HMM approximation.

It also seems reasonable to conclude that hidden Markov approximations provide a convenient and flexible device for implementing a diverse variety of nonstandard stochastic volatility models. The programming effort involved in fitting the HMMs is very modest. The function used to compute the log-likelihood for the HMM approximation to the SV_0 model consists of 16 lines of code: see Appendix B. The nonstandard models require only straightforward changes to that code.

Acknowledgements

The authors would like to thank the editor and an anonymous referee for helpful comments on an earlier version of this manuscript.

Appendix A. Parameter estimates

Table A.6
Parameter estimates for the SV_0 model ($N=100$, $-g_{min}=g_{max}=5$).

	$\hat{\phi}$	$\hat{\sigma}$	$100\hat{\beta}$
Sony	0.960	0.238	1.947
Time Warner	0.995	0.125	2.288
Toyota	0.976	0.171	1.676
Trav. Comp.	0.969	0.239	1.625
BP	0.986	0.113	1.519
Roy. D. Sh.	0.987	0.118	1.571
Bank of Am.	0.993	0.167	1.658
Citigroup	0.991	0.179	1.919
Deu. Bank	0.988	0.150	2.034
Morgan St.	0.990	0.149	2.364

Table A.7
Parameter estimates for the *SVt* model ($N=100$, $-g_{min}=g_{max}=5$).

	$\hat{\phi}$	$\hat{\sigma}$	$100\hat{\beta}$	$\hat{\nu}$
Sony	0.983	0.141	1.759	8.5
Time Warner	0.997	0.085	2.138	11.3
Toyota	0.984	0.129	1.569	12.9
Trav. Comp.	0.987	0.140	1.456	8.0
BP	0.990	0.092	1.436	16.7
Roy. D. Sh.	0.990	0.101	1.494	19.5
Bank of Am.	0.996	0.119	1.588	11.0
Citigroup	0.995	0.122	1.822	10.0
Deu. Bank	0.994	0.101	1.817	8.5
Morgan St.	0.993	0.116	2.187	11.9

Table A.8
Parameter estimates for the SVMt model ($N=100$, $-g_{min}=g_{max}=8$).

	$\hat{\phi}$	$\hat{\sigma}$	$\hat{\alpha}$	$100\hat{\beta}$	$\hat{\nu}$	$100\hat{\xi}$
Sony	$\begin{pmatrix}0.738\\1.016\end{pmatrix}$	$\begin{pmatrix}0.004\\0.217\end{pmatrix}$	0.111	1.048	8.6	0.703
Time Warner	$\begin{pmatrix}0.974\\1.178\end{pmatrix}$	$\begin{pmatrix}0.031\\0.355\end{pmatrix}$	0.923	1.161	12.2	0.465
Toyota	$\begin{pmatrix}0.979\\1.336\end{pmatrix}$	$\begin{pmatrix}0.101\\1.187\end{pmatrix}$	0.974	0.849	13.7	0.650
Trav. Comp.	$\begin{pmatrix}0.958\\1.086\end{pmatrix}$	$\begin{pmatrix}0.021\\0.453\end{pmatrix}$	0.746	0.549	7.5	0.758
BP	$\begin{pmatrix}0.972\\1.063\end{pmatrix}$	$\begin{pmatrix}0.026\\0.441\end{pmatrix}$	0.756	0.206	16.5	0.969
Roy. D. Sh.	$\begin{pmatrix}0.975\\1.073\end{pmatrix}$	$\begin{pmatrix}0.025\\0.503\end{pmatrix}$	0.798	0.203	16.6	0.996
Bank of Am.	$\begin{pmatrix}0.970\\1.036\end{pmatrix}$	$\begin{pmatrix}0.246\\0.003\end{pmatrix}$	0.579	1.286	10.6	0.576
Citigroup	$\begin{pmatrix}0.945\\1.166\end{pmatrix}$	$\begin{pmatrix}0.035\\0.138\end{pmatrix}$	0.781	1.343	11.0	0.000
Deu. Bank	$\begin{pmatrix}0.987\\1.087\end{pmatrix}$	$\begin{pmatrix}0.174\\0.002\end{pmatrix}$	0.916	0.933	8.5	0.754
Morgan St.	$\begin{pmatrix}0.952\\1.185\end{pmatrix}$	$\begin{pmatrix}0.037\\0.220\end{pmatrix}$	0.825	1.265	13.0	0.485

Table A.9
Parameter estimates for the *MSSVt* model ($N=200$, $-g_{min}=g_{max}=6$).

	$\hat{\phi}$	$\hat{\sigma}$	$\hat{\Gamma}^{(\alpha)}$	$100\hat{\beta}$	$\hat{\nu}$	100ξ
Sony	0.986	$\begin{pmatrix}0.149\\2.253\end{pmatrix}$	$\begin{pmatrix}0.994 & 0.006\\0.357 & 0.643\end{pmatrix}$	0.962	9.1	0.708
Time Warner	0.998	$\begin{pmatrix}0.085\\0.292\end{pmatrix}$	$\begin{pmatrix}0.986 & 0.014\\0.089 & 0.911\end{pmatrix}$	1.078	11.3	0.654
Toyota	0.986	$\begin{pmatrix}0.092\\0.448\end{pmatrix}$	$\begin{pmatrix}0.989 & 0.011\\0.029 & 0.971\end{pmatrix}$	0.843	14.6	0.692
Trav. Comp.	0.988	$\begin{pmatrix}0.016\\0.423\end{pmatrix}$	$\begin{pmatrix}0.991 & 0.009\\0.013 & 0.987\end{pmatrix}$	0.487	7.4	0.775
BP	0.987	$\begin{pmatrix}0.025\\0.419\end{pmatrix}$	$\begin{pmatrix}0.989 & 0.011\\0.037 & 0.963\end{pmatrix}$	0.448	16.9	0.831
Roy. D. Sh.	0.984	$\begin{pmatrix}0.022\\0.434\end{pmatrix}$	$\begin{pmatrix}0.983 & 0.017\\0.055 & 0.945\end{pmatrix}$	0.512	17.9	0.803
Bank of Am.	0.992	$\begin{pmatrix}0.011\\0.239\end{pmatrix}$	$\begin{pmatrix}0.997 & 0.003\\0.004 & 0.996\end{pmatrix}$	1.365	11.4	0.350
Citigroup	0.987	$\begin{pmatrix}0.025\\0.491\end{pmatrix}$	$\begin{pmatrix}0.968 & 0.032\\0.162 & 0.838\end{pmatrix}$	1.049	11.2	0.359
Deu. Bank	0.993	$\begin{pmatrix}0.050\\0.367\end{pmatrix}$	$\begin{pmatrix}0.981 & 0.019\\0.049 & 0.951\end{pmatrix}$	0.777	8.8	0.765
Morgan St.	0.994	$\begin{pmatrix}0.110\\0.352\end{pmatrix}$	$\begin{pmatrix}0.998 & 0.002\\0.004 & 0.996\end{pmatrix}$	1.096	13.2	0.842

Table A.10
Parameter estimates for the SVVt model ($N=100$, $-g_{min}=g_{max}=6$).

	$\hat{\phi}$	$\hat{\omega}$	$100\hat{\gamma}$	$100\hat{\beta}$	$\hat{\nu}$	$100\hat{\xi}$
Sony	0.985	0.026	0.382	1.281	8.5	0.449
Time Warner	0.998	0.010	0.427	2.281	11.3	0.619
Toyota	0.985	0.043	0.042	0.898	13.2	0.006
Trav. Comp.	0.994	0.062	0.064	0.478	7.4	0.797
BP	0.995	0.030	0.151	0.487	16.5	0.863
Roy. D. Sh.	0.996	0.038	0.119	0.487	17.0	0.916
Bank of Am.	0.998	0.030	0.088	0.852	10.6	0.595
Citigroup	0.998	0.007	0.450	1.710	10.0	0.167
Deu. Bank	0.996	0.037	0.017	0.563	8.5	0.839
Morgan St.	0.997	0.008	0.826	2.132	12.9	0.441

Table A.11
Parameter estimates for the *GSVt* model ($N=100$, $g_{min}=0$, $g_{max}=200$).

	$\hat{\phi}$	$\hat{\lambda}$	$100\hat{\beta}$	$\hat{\nu}$	$\hat{\xi}$
Sony	0.984	0.547	0.326	8.4	5.435
Time Warner	0.999	0.146	0.333	11.8	4.503
Toyota	0.984	0.502	0.296	12.2	6.475
Trav. Comp.	0.991	0.149	0.356	8.0	4.869
BP	0.985	0.247	0.318	18.8	8.307
Roy. D. Sh.	0.987	0.187	0.355	22.8	7.129
Bank of Am.	0.999	0.013	0.532	12.9	0.619
Citigroup	0.998	0.108	0.516	9.2	0.524
Deu. Bank	0.993	0.194	0.441	9.5	3.859
Morgan St.	0.994	0.188	0.505	9.9	2.994

Appendix B. R code for the log-likelihood of the HMM approximation to the SV_0 model.

```
loglikelihood.SV0<-function(y,phi,sigma,beta,m,gmax){          #01
  K<- m+1                                                       #02
  b<- seq(-gmax,gmax,length=K)                                  #03
  bs<- (b[-1]+b[-K])*0.5                                        #04
  sey<- beta*exp(bs/2)                                          #05
  Gamma<- matrix(0,m,m)                                         #06
  for (i in 1:m) Gamma[i,]<-diff(pnorm(b,phi*bs[i],sigma))      #07
  Gamma<- Gamma/apply(Gamma,1,sum)                              #08
  foo<- solve(t(diag(m)-Gamma+1),rep(1,m))                      #09
  llk<- 0                                                       #10
  for (t in 1:length(y)){                                       #11
    foo<- foo%*%Gamma*dnorm(y[t],0,sey)                         #12
    sumfoo<- sum(foo)                                           #13
    llk<- llk+log(sumfoo)                                       #14
    foo<- foo/sumfoo}                                           #15
  return(llk)}                                                  #16
```

Notes:
`#03` Compute the interval endpoints.
`#04` Compute the interval midpoints.
`#05` Compute the std deviation of y_t for each interval midpoint.
`#07` Compute the t.p.m. $\mathbf{\Gamma}$ using expression (5).
`#08` Scale the rows of $\mathbf{\Gamma}$ so that they each sum to 1.
`#09` Compute $\boldsymbol{\delta}(=\boldsymbol{\delta}\mathbf{\Gamma})$, the stationary distribution of the Markov chain $\{h_t\}$.
`#11-#15` Loop to compute the log-likelihood, i.e. the log of expression (6).

References

Bartolucci, F., De Luca, G., 2001. Maximum likelihood estimation of a latent variable time-series model. Appl. Stoch. Model. Bus. Ind. 17, 5–17.
Bartolucci, F., De Luca, G., 2003. Likelihood-based inference for asymmetric stochastic volatility models. Comput. Stat. Data Anal. 42, 445–449.
Basel Committee on Banking Supervision, 2006. International Convergence of Capital Measurement and Capital Standards: A Revised Framework. Bank for International Settlements, Basel.
Broto, C., Ruiz, E., 2004. Estimation methods for stochastic volatility models: a survey. J. Econ. Surv. 18, 613–649.
Cappé, O., Moulines, E., Rydén, T., 2005. Inference in Hidden Markov Models. Springer, New York.
Chib, S., Nardari, F., Shephard, N., 2002. Markov chain Monte Carlo methods for stochastic volatility models. J. Econometrics 108, 281–316.
Clements, A.E., Hurn, S., White, S.I., 2006. Mixture distribution-based forecasting using stochastic volatility models. Appl. Stoch. Model. Bus. Ind. 22, 547–557.
Danielsson, J., 1994. Stochastic volatility in asset prices: estimation with simulated maximum likelihood. J. Econometrics 64, 375–400.
Ephraim, Y., Merhav, N., 2002. Hidden Markov processes. IEEE Trans. Inf. Theory 48, 1518–1569.
Franses, P.H., van Dijk, D., 2000. Non-linear Time Series Models in Empirical Finance. Cambridge University Press, Cambridge.
Fridman, M., Harris, L., 1998. A maximum likelihood approach for non-Gaussian stochastic volatility models. J. Bus. Econ. Stat. 16, 284–291.
Grunwald, G.K., Hyndman, R.J., Tedesco, L., Tweedie, R.L., 2000. Non-Gaussian conditional linear AR(1) models. Aust. N. Z. J. Stat. 42, 479–495.
Harvey, A.C., Ruiz, E., Shephard, N., 1994. Multivariate stochastic volatility models. Rev. Econ. Stud. 61, 247–264.
Jacquier, E., Polson, N.G., Rossi, P.E., 1994. Bayesian analysis of stochastic volatility models (with discussion). J. Bus. Econ. Stat. 12, 371–417.
Kim, S., Shephard, N., Chib, S., 1998. Stochastic volatility: likelihood inference and comparison with ARCH models. Rev. Econ. Stud. 65, 361–393.
Kitagawa, G., 1987. Non-Gaussian state-space modeling of nonstationary time series (with discussion). J. Am. Stat. Assoc. 82, 1032–1063.
Koopman, S.J., Shephard, N., Doornik, J., 1999. Statistical algorithms for models in state space using SsfPack 2.2. Econ. J. 2, 113–166.
Melino, A., Turnbull, S.M., 1990. Pricing foreign currency options with stochastic volatility. J. Econometrics 45, 239–265.
Rosenblatt, M., 1952. Remarks on a multivariate transformation. Ann. Math. Stat. 23, 470–472.
Rossi, A., Gallo, G.M., 2006. Volatility estimation via hidden Markov models. J. Empir. Finance 13, 203–230.
Sandmann, G., Koopman, S.J., 1998. Estimation of stochastic volatility models via Monte Carlo maximum likelihood. J. Econometrics 87, 271–301.
Shephard, N., 1996. Statistical aspects of ARCH and stochastic volatility. In: Cox, D.R., Hinkley, D.V., Barndorff-Nielsen, O.E. (Eds.), Time Series Models: In econometrics, finance and other fields. Chapman & Hall, London, pp. 1–67.

Shephard, N. (Ed.), 2005. Stochastic Volatility: Selected Readings. Oxford University Press, Oxford.
So, M.K.P., Lam, K., Li, W.K., 1998. A stochastic volatility model with Markov switching. J. Bus. Econ. Stat. 16, 244–253.
Swiss National Bank, 2008. Financial Stability Report.
Taylor, S.J., 2005. Asset Price Dynamics, Volatility, and Prediction. Princeton University Press, Princeton.
Wong, W.K., 2010. Backtesting value-at-risk based on tail losses. J. Empir. Finance 17, 526–538.
Wong, C.S., Li, W.K., 2000. On a mixture autoregressive model. J. R. Stat. Soc. B 62, 95–115.
Zucchini, W., MacDonald, I.L., 2009. Hidden Markov Models for Time Series: An Introduction Using R. Chapman & Hall/CRC Press, London and Boca Raton.