Comput Stat (2015) 30:517–537
DOI 10.1007/s00180-014-0547-5

ORIGINAL PAPER

Semiparametric stochastic volatility modelling using penalized splines

Roland Langrock · Théo Michelot · Alexander Sohn · Thomas Kneib

Received: 17 June 2014 / Accepted: 21 November 2014 / Published online: 4 December 2014

Abstract Stochastic volatility (SV) models mimic many of the stylized facts attributed to time series of asset returns, while maintaining conceptual simplicity. The commonly made assumption of conditionally normally distributed or Student-t-distributed returns, given the volatility, has however been questioned. In this manuscript, we introduce a novel maximum penalized likelihood approach for estimating the conditional distribution in an SV model in a nonparametric way, thus avoiding any potentially critical assumptions on the shape. The considered framework exploits the strengths both of the hidden Markov model machinery and of penalized B-splines, and constitutes a powerful alternative to recently developed Bayesian approaches to semiparametric SV modelling. We demonstrate the feasibility of the approach in a simulation study before outlining its potential in applications to three series of returns on stocks and one series of stock index returns.

Keywords B-splines · Cross-validation · Forward algorithm · Hidden Markov model · Numerical integration · Penalized likelihood

Electronic supplementary material The online version of this article (doi:10.1007/s00180-014-0547-5) contains supplementary material, which is available to authorized users.

R. Langrock (✉)
University of St Andrews, St Andrews, UK
e-mail: roland.langrock@st-andrews.ac.uk

T. Michelot
INSA de Rouen, Rouen, France

A. Sohn · T. Kneib
Georg August University of Göttingen, Göttingen, Germany

1 Introduction

Stochastic volatility (SV) models are immensely popular tools for the analysis of financial time series. This subclass of state-space models (SSMs) constitutes one of the two most widely used approaches for modelling stock market volatility, the other one being ARCH/GARCH-type models. For log-returns $y_1, \ldots, y_T$ on an asset, the structure of the standard discrete-time SV model, labeled SV_0 in the following, is as follows:

$$y_t = \varepsilon_t^{(0)} \beta \exp(g_t/2), \qquad g_t = \phi g_{t-1} + \sigma \eta_t, \tag{1}$$

where $\beta, \sigma > 0$, and where $\{\varepsilon_t^{(0)}\}$ and $\{\eta_t\}$ are independent sequences of independent standard normal random variables (e.g., Shephard 1996). Stationarity is obtained for $|\phi| < 1$. The unobserved sequence $\{g_t\}$, commonly referred to as the log-volatility process, represents the time-varying "nervousness" of the market. This model captures several of the stylized facts attributed to asset returns, including positive autocorrelation of squared returns (indicating a volatility that slowly varies over time and hence volatility clustering), zero autocorrelation of the unsquared returns, and a kurtosis in excess of 3. However, the basic model tends to underestimate the probability of relatively extreme returns, such that it is often more adequate to consider a Student-t distribution with ν degrees of freedom for $\varepsilon_t^{(0)}$ (Chib et al. 2002); we label this second model SV_t.

In the existing literature, several different model formulations have been considered that extend the flexibility of the log-volatility process, $\{g_t\}$ (e.g., Gallant et al. 1997; Abraham et al. 2006; Langrock et al. 2012). Yet Durham (2006) found "no evidence that even simple single factor models are unable to capture the dynamics of the volatility process" (p. 276). Instead, Durham considers the shape of the conditional distribution in SV models—i.e., of the conditional distribution of y_t, given g_t—to be "the more critical problem" (p. 304). In addition to heavy tails, which are accounted for in the SV_t formulation, evidence of asymmetries has been found (e.g., Gallant et al. 1997; Harvey and Siddique 2000; Jondeau and Rockinger 2003; Durham 2006). To get the shape right, and in particular to accurately estimate the tails of the conditional distribution, is of high importance for example in risk management. While parametric models can be constructed that enable the inclusion of heavy tails and skewness (e.g., using a skewed Student-t distribution, as proposed for example by Nakajima and Omori 2012), nonparametric approaches have the considerable advantage that no restriction to a particular class of distributions is made a priori.

Some recent work in this direction has been conducted in the Bayesian context, where the normal distribution can be used as a building block to formulate more complex models that still utilize the benefits of the normal formulation for constructing convenient update schemes in a Markov chain Monte Carlo simulation. Abanto-Valle et al. (2010) consider scale mixtures of normals for the conditional distribution, where the variance of the normal distribution is supplemented with suitable prior specifications that yield a larger class of potential marginal distributions after integrating out the mixing distribution for the variance. Nonparametric specifications relying on an infinite-dimensional mixture of normals, generated by a Dirichlet process mixture prior, have been developed in Jensen and Maheu (2010) and Delatola and Griffin (2011).

While Jensen and Maheu (2010) directly tackle the conditional distribution of the returns, Delatola and Griffin (2011) employ a different representation where the logarithm of the squared noise in the return process is considered for the analysis. Delatola and Griffin (2013) extend the latter approach by including a leverage effect, allowing for the potential correlation of the two error terms, in the return and in the log-volatility process, respectively.

In this manuscript, we develop a novel frequentist approach for nonparametrically estimating the conditional distribution in an SV model. The proposed maximum penalized likelihood approach is similar in spirit to that proposed in Langrock et al. (2014, to appear), and exploits the strengths both of likelihood-based hidden Markov model (HMM) machinery and of penalized B-splines (i.e., P-splines). The former is employed to deal with a well-known difficulty with SV models, which is that their likelihood is given by a high-order multiple integral that is analytically intractable. It has however been shown that methods available for HMMs—which have the same dependence structure as SV models and constitute another subclass of SSMs, with finite state space—can be applied in order to perform a fast and accurate numerical integration of the SV model likelihood. More specifically, such a numerical integration corresponds to a fine discretization of the support of the log-volatility process. The associated transformation of the continuous support of $\{g_t\}$ to a finite support renders the powerful HMM forward algorithm applicable, making it feasible to evaluate an arbitrarily accurate approximation to the SV model likelihood (Fridman and Harris 1998; Bartolucci and Luca 2001, 2003; Langrock et al. 2012). We extend this likelihood-based approach to allow for a nonparametric estimation of the conditional distribution, by representing the density of this distribution as a linear combination of a large number of standardized B-spline basis functions, including a roughness penalty in the likelihood in order to arrive at an appropriate balance between goodness of fit and smoothness for the fitted density. Since we still model the log-volatility process in a parametric way, we use the label SV_{sp} to refer to the resulting *semiparametric* SV model with nonparametrically modelled conditional distribution.

The paper is structured as follows. In Sect. 2, we begin by describing the likelihood evaluation, then introducing the B-spline-based representation of the conditional distribution and discussing associated inferential issues. The performance of the suggested approach is investigated in a simulation study in Sect. 3. In Sect. 4, we apply the approach to real data related to three stocks and one stock index, comparing the predictive performance of our model to popular parametric counterparts.

2 Semiparametric SV modelling

2.1 SV model likelihood

We consider a model SV_{sp} of the form

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

with the η_t iid standard normal, but where, in contrast to the models SV_0 and SV_{t}, we do not make any assumptions on the distributional form of the random variables ε_t.

However, we do assume these variables to be iid, and to be independent of $\{\eta_t\}$. Our aim is to nonparametrically estimate the probability density function (pdf) f_ε of the variables ε_t. Compared to the SV_0 model, given in (1), we have omitted the parameter β in (2), since otherwise the semiparametric model would not be identifiable; in the SV_{sp} model, the effect of β will be absorbed within f_ε. Before we introduce our strategy for estimating f_ε in a nonparametric way (alongside the other model parameters), we will derive a tractable likelihood function for general f_ε, including the nonparametric case, but for example also those of a normal distribution and of a Student-t distribution. In the following, we will use f as a general symbol for a density function. To formulate the likelihood, we will require the conditional pdfs of the random variables y_t, given g_t $(t = 1, \ldots, T)$. We denote these conditional pdfs by $f(y_t|g_t)$, for $t = 1, \ldots, T$. These pdfs are simple transformations of the density f_ε:

$$f(y_t|g_t) = \exp(-g_t/2) f_\varepsilon(y_t \exp(-g_t/2)).$$

For any f_ε, the likelihood of the model defined by (2) can then be derived as

$$\begin{aligned}
\mathcal{L} &= \int \cdots \int f(y_1, \ldots, y_T, g_1, \ldots, g_T)\, dg_T \ldots dg_1 \\
&= \int \cdots \int f(y_1, \ldots, y_T|g_1, \ldots, g_T) f(g_1, \ldots, g_T)\, dg_T \ldots dg_1 \\
&= \int \cdots \int f(g_1) f(y_1|g_1) \prod_{t=2}^{T} f(g_t|g_{t-1}) f(y_t|g_t)\, dg_T \ldots dg_1 \\
&= \int \cdots \int f(g_1) \exp(-g_1/2) f_\varepsilon(y_1 \exp(-g_1/2)) \\
&\quad \times \prod_{t=2}^{T} f(g_t|g_{t-1}) \exp(-g_t/2) f_\varepsilon(y_t \exp(-g_t/2))\, dg_T \ldots dg_1 . \qquad (3)
\end{aligned}$$

In the second last step, we exploited the dependence structure that is characteristic of SV models, HMMs and general SSMs. Hence, the likelihood is a high-order multiple integral that cannot be evaluated directly. Via numerical integration, using a simple rectangular rule based on m equidistant intervals, $B_i = (b_{i-1}, b_i)$, $i = 1, \ldots, m$, with midpoints b_i^* and of length b, the likelihood can be approximated as follows:

$$\begin{aligned}
\mathcal{L} &\approx b^T \sum_{i_1=1}^{m} \cdots \sum_{i_T=1}^{m} f(b_{i_1}^*) \exp(-b_{i_1}^*/2) f_\varepsilon(y_1 \exp(-b_{i_1}^*/2)) \\
&\quad \times \prod_{t=2}^{T} f(b_{i_t}^*|b_{i_{t-1}}^*) \exp(-b_{i_t}^*/2) f_\varepsilon(y_t \exp(-b_{i_t}^*/2)) = \mathcal{L}_{\text{approx}}. \qquad (4)
\end{aligned}$$

This approximation can be made arbitrarily accurate by increasing m, and in fact virtually exact for m around 100, provided that the interval (b_0, b_m) covers the essential range of the log-volatility process (for more details, see Sect. 2.3.1). We note that, in

principle, other quadrature rules could be implemented (see, e.g., Fridman and Harris 1998), which could improve the efficiency of the approximation. However, the simple midpoint quadrature is very easy to implement, and at the same time is extremely accurate for values of m which will still guarantee computational tractability when HMM machinery is applied. Thus, throughout this work, we will be limiting ourselves to the consideration of this simple yet functional numerical integration procedure.

In the given form, the approximate likelihood (4) is computationally intractable already for small T, since its evaluation requires $O(Tm^T)$ operations. However, a much more efficient recursive scheme can be used to evaluate the approximate likelihood. To see this, note that the numerical integration essentially corresponds to a discretization of the state space, i.e., of the support of the log-volatility process $\{g_t\}$. Therefore, the approximate likelihood given in (4) can be evaluated using the well-developed and powerful machinery of the subclass of SSMs given by HMMs, which are SSMs with a finite state space (cf. Langrock 2011; Langrock et al. 2012). We sketch the relevant HMM methodology in the "Appendix" to this manuscript. In particular, in the "Appendix" we highlight a key property of HMMs, which is that the likelihood can be evaluated efficiently using the so-called forward algorithm (Zucchini and MacDonald 2009). Rather than separately considering all possible underlying state (and hence log-volatility) sequences, as in (4), the forward algorithm exploits the conditional independence assumptions to perform the likelihood calculation recursively, traversing along the time series and updating the likelihood and state probabilities at every step. For an HMM, applying the forward algorithm results in a matrix product expression for the likelihood, and this is exactly what we obtain also in the present context:

$$\mathcal{L}_{\text{approx}} = \boldsymbol{\delta}\mathbf{P}(y_1)\boldsymbol{\Omega}\mathbf{P}(y_2)\boldsymbol{\Omega}\mathbf{P}(y_3)\ldots\boldsymbol{\Omega}\mathbf{P}(y_{T-1})\boldsymbol{\Omega}\mathbf{P}(y_T)\mathbf{1}. \tag{5}$$

Here, the $m \times m$-matrix $\boldsymbol{\Omega} = \left(\omega_{ij}\right)$ is the analogue to the transition probability matrix in case of an HMM (see the "Appendix"), defined as $\omega_{ij} = f(b_j^*|b_i^*) \cdot b$. Furthermore, the vector $\boldsymbol{\delta}$ is the analogue to the Markov chain initial distribution in case of an HMM, here defined such that δ_i, $i = 1, \ldots, m$, is the density of the normal distribution with mean zero and standard deviation $\sigma/\sqrt{1-\phi^2}$—the stationary distribution of the autoregressive process used to model the log-volatility—evaluated at b_i^* and multiplied by b. Finally, $\mathbf{P}(y_t)$ is an $m \times m$ diagonal matrix with ith diagonal entry $\exp(-b_i^*/2)\, f_\varepsilon(y_t \exp(-b_i^*/2))$, hence the analogue to the matrix comprising the state-dependent probabilities in case of an HMM. Using the matrix product expression given in (5), the approximate likelihood can be evaluated in $O(Tm^2)$ operations. In practice, this means that the likelihood of an SV model can typically be calculated in a fraction of a second, even for T in the thousands and say $m = 100$, a value which renders the approximation virtually exact. Furthermore, $\mathcal{L}_{\text{approx}} \to \mathcal{L}$ as $b_m, m \to \infty$ and $b_0 \to -\infty$.

2.2 Nonparametric modelling

We now turn to the nonparametric modelling of the distribution of ε_t. Following Schellhase and Kauermann (2012), we suggest to estimate the pdf of ε_t by considering

finite linear combinations of a large number of basis functions:

$$\hat{f}_\varepsilon(x) = \sum_{k=-K}^{K} a_k \psi_k(x). \tag{6}$$

Here the basis functions $\psi_{-K}, \ldots, \psi_K$ are known and fixed pdfs. Clearly, $\hat{f}_\varepsilon(x)$ is a pdf if $\sum_{k=-K}^{K} a_k = 1$ and $a_j \geq 0$ for all $j = -K, \ldots, K$. To enforce these constraints, the coefficients to be estimated, $a_{-K}, \ldots, a_K$, are transformed using the multinomial logit link function

$$a_k = \frac{\exp(\beta_k)}{\sum_{j=-K}^{K} \exp(\beta_j)}, \tag{7}$$

where we set $\beta_0 = 0$ for identifiability. In principle, any set of densities $\psi_{-K}, \ldots, \psi_K$ can be used to approximate $f_\varepsilon(x)$ as in (6). We follow Schellhase and Kauermann (2012) and use B-splines, in ascending order in the basis used in (6), and standardized such that they integrate to 1. For more details on B-splines, see for example Boor (1978) and Eilers and Marx (1996). Since each B-spline basis function is associated with a separate coefficient, this model formulation in fact leads to a finite-dimensional parameter space. However, the coefficients themselves are not of interest, and we can use arbitrarily many of them in order to allow for virtually any desired shaped of the conditional distribution. We therefore refer to our estimation approach as nonparametric despite the fact that it does rely on a parametric specification with a large number of parameters. This is in line with the standard terminology used in the literature, where (penalized) spline approaches are subsumed under nonparametric approaches (see for example Ruppert et al. 2003).

The approximate likelihood of the resulting SV_{sp} model is given by (5), plugging in $\hat{f}_\varepsilon$ for f_ε in the matrices $\mathbf{P}(y_t)$, $t = 1, \ldots, T$. Following Eilers and Marx (1996), we modify the (approximate) log-likelihood by including a penalty on (q-th order) differences between coefficients associated with adjacent B-splines, yielding the penalized log-likelihood

$$l_p = \log\left(\mathcal{L}_{\text{approx}}\right) - \frac{\lambda}{2} \sum_{k=-K+q}^{K} \left(\Delta^q a_k\right)^2, \tag{8}$$

with a_k parameterized as in (7) and smoothing parameter $\lambda \geq 0$. The penalty term involves the difference operator Δ, where $\Delta a_k = a_k - a_{k-1}$ and $\Delta^q a_k = \Delta(\Delta^{q-1} a_k)$. This results in a penalization of roughness of the estimator, with λ controlling how much emphasis is put on goodness of fit and on smoothness, respectively. In particular, unpenalized estimates are obtained for $\lambda = 0$. For $\lambda \to \infty$ the penalty will dominate the likelihood, resulting in a sequence of weights a_k that follow a polynomial of order $q - 1$ in k. The difference order therefore also affects the smoothness of the estimates indirectly (and to a much smaller extent than the degree of the spline basis). We will use $q = 2$ in the remainder since this provides an approximation to the integrated squared second derivative penalty that is popular in the context of smoothing splines.

Including the penalty term in the likelihood avoids the problem of selecting an optimal number of basis elements, since the penalty effectively reduces the number of free basis parameters and yields an adaptive fit to the data, provided the smoothing parameter is chosen in a data-driven way. The number of basis elements needs to be large enough to give sufficient flexibility for reflecting the structure of the conditional distribution f_ε, but once this threshold is passed, increasing the number of basis elements further does no longer change the fit to the data much due to the impact of the penalty. For moderately smooth regression functions, Ruppert (2002) recommends to use a default of about 35 (or 40) basis functions. To capture the pdf of ε_t in an SV model, we expect such a choice to easily provide sufficient flexibility, and hence have chosen K accordingly in our analyses (see below). To select the smoothing parameter in a data-driven way, we will consider cross-validation (see Sect. 2.3.2).

In preliminary simulation experiments, we found that with an equidistant spacing of the knots our approach tended to produce estimated densities that were overly smooth around the peak of the true distribution and too wiggly in the tails. This is related to the fact that the basis coefficients systematically decay towards the tails of the estimated distribution, which would require an adaptive amount of smoothing instead of a global smoothing parameter. As a simple yet effective strategy to achieve such adaptiveness, we consider increasingly wider distances between the B-spline basis densities towards the tails instead of the common equidistant specification. Since we still rely on the unweighted difference penalty in (8), this effectively increases the penalty for the tails of the distribution.

2.3 Inference

2.3.1 Parameter estimation

The use of the forward algorithm allows for a very fast evaluation of the penalized log-likelihood given in (8). A numerical maximization of the penalized log-likelihood is therefore feasible in typical cases, even for high m and hence very close approximations to the likelihood in (3); some computing times are given in Sect. 4. Since the first part of expression (8) is susceptible to numerical overflow, it is required to compute its logarithm, which involves a minor difficulty since we are dealing with a matrix product. However, techniques to address this issue are standard: Zucchini and MacDonald (2009) describe a straightforward scaling strategy for calculating the logarithm of an HMM-type matrix product likelihood (see their Chapter 3).

In practice, one also has to select the value of m, the number of intervals used in the discretization of the log-volatility process, and the range of possible g_t-values considered in the numerical integration. In our experience, estimates usually stabilize for values of m around 50 (cf. Langrock et al. 2012; Langrock and King 2013). The minimum and maximum values for g_t have to be chosen sufficiently large to cover the essential domain of the log-volatility process, but not too large, in order to maintain sufficient fineness of the grid. More guidance on this issue is provided in Langrock et al. (2012). Another technical issue in the numerical maximization is that of local maxima: it may sometimes happen that the numerical search fails to find the MLE,

and returns a local maximum instead. The best way to address this issue seems to be to use a number of different sets of initial values in order to find and verify the global maximum. Uncertainty quantification, for both the parameters of the underlying log-volatility process and the density of ε_t, can be conducted using a parametric bootstrap.

2.3.2 Choice of the smoothing parameter

Cross-validation techniques can be used to choose the smoothing parameter. We proceed along the lines of Racine (2000) and use a k-fold h-block cross-validation approach, noting that this is only one of several cross-validation techniques available for time series (see, e.g., Arlot and Celisse 2010; Bergmeier and Benítez 2012). More specifically, we first partition any given series of log-returns into k subsets containing consecutive observations, denoted by $I_1, \ldots, I_k$ (e.g., $I_1 = \{1, 2, \ldots, 200\}$, $I_2 = \{201, 202, \ldots, 400\}$, etc.). For $i = 1, \ldots, k$, the model is then calibrated by estimating the parameters using all available data except those from I_i. We here maintain the existing time series structure, by fitting the model to the time series $y_1^*, \ldots, y_T^*$, where $y_t^* = y_t$ if the observation made at time t is in the calibration sample, and where y_t^* is treated as a missing data point if the corresponding observation is in I_i. (Dealing with missing data is straightforward using the HMM machinery—see the "Appendix".) The calibrated model is then assessed on a validation sample, namely I_i but excluding the first h and the last h observations within I_i; this procedure is to ensure near-independence of the calibration and validation samples (Racine 2000). Proper scoring rules (Gneiting and Raftery 2007) are used to assess the calibrated model for the given λ. For computational convenience, we consider the log-likelihood of the validation sample, under the model fitted in the calibration stage, as the score of interest. Alternatively, we could consider the log-likelihood of the validation sample conditional on the observations from the calibration sample—which seems sensible if the focus lies on predictive capacity, as opposed to goodness of fit—but with the given cross-validation scheme, which ensures near-independence of calibration and validation samples, it is anticipated that this would in any case not make much difference in practice. From some pre-specified set of possible smoothing parameters, e.g., $\{2^n | n = r, r+1, \ldots, s\}$, where r and s are integers, we select the λ with the highest mean score over all k cross-validation samples.

2.3.3 Model checking

We have already seen that the use of the HMM forward algorithm provides an efficient and convenient way to evaluate the SV model likelihood. Moreover, HMM techniques can also be used in order to check the goodness of fit of a given model. Following Zucchini and MacDonald (2009), we consider one-step-ahead forecast pseudo-residuals for model checking. Pseudo-residuals have the same purpose as residuals in regression analyses, but can be applied much more generally (e.g., in time series analyses). The idea behind pseudo-residuals is to compare any given observation to the distribution of that observation under the fitted model. In our context, the forecast

pseudo-residuals are given by

$$r_t = \Phi^{-1}\big(F(y_t \mid y_{t-1}, y_{t-2}, \ldots, y_1)\big).$$

Here Φ denotes the cumulative distribution function of the standard normal distribution, and $F(y_t \mid y_{t-1}, y_{t-2}, \ldots, y_1)$ is the cumulative distribution function of y_t given all observations up to time $t-1$. Using the HMM-type approximation, this can be written as

$$\begin{aligned} F(y_t \mid y_{t-1}, y_{t-2}, \ldots, y_1) &\approx \sum_{i=1}^{m} \zeta_i F(y_t \mid g_t = b_i^*) \\ &= \sum_{i=1}^{m} \zeta_i \int_{-\infty}^{y_t} \exp(-b_i^*/2) f_\varepsilon \left(x \exp(-b_i^*/2)\right) dx \,, \end{aligned} \tag{9}$$

where ζ_i is the ith entry of the vector $\tilde{\boldsymbol{\alpha}}_{t-1}\boldsymbol{\Omega}/(\tilde{\boldsymbol{\alpha}}_{t-1}\mathbf{1}')$, which is defined as

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

$t = 2, \ldots, T$, with $\boldsymbol{\delta}$, $\mathbf{P}(y_k)$ and $\boldsymbol{\Omega}$ defined as above. These $\tilde{\boldsymbol{\alpha}}_t$'s constitute the SV model analogue to the HMM forward probabilities; see the "Appendix" for more details on the latter. The representation given in (9) is only approximate due to the discretization of the log-volatility process, but as for the likelihood the accuracy also of this approximation can be made arbitrarily accurate by increasing m. 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, then the pseudo-residuals follow a standard normal distribution (see also Zucchini and MacDonald 2009). Thus, forecast pseudo-residuals can be used to identify extreme values, and the general suitability of the model can be checked by using, for example, quantile–quantile plots or formal tests for normality.

2.3.4 Decoding

Again building on existing HMM machinery, estimates of the underlying log-volatility 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. Furthermore, the formula for computing state probabilities of an HMM—Eq. (5.6) in Zucchini and MacDonald 2009—can be used to quantify the uncertainty in the log-volatility estimates.

3 Simulation experiments

To generate artificial data which adhere to many of the stylized facts discussed above, we use an SV model as in (2), with $\phi = 0.98$, $\sigma = 0.1$ and ε_t specified as $\varepsilon_t =$

$0.02(\zeta_{1,t}-1)^{\alpha_t}(\zeta_{2,t}+1)^{1-\alpha_t}+0.006$, where $\zeta_{1,t}$ and $\zeta_{2,t}$ are mutually independent iid sequences of Student-t random variables with 6 and 8 degrees of freedom, respectively, and α_t are iid Bernoulli variables, each taking on the value 1 with probability 0.35. This specification results in a skewed and leptokurtic distribution (skewness ≈ -0.22; kurtosis ≈ 3.58) with zero mean. For an illustration of the shape of this distribution, see Fig. 1.

For this model, we conducted 200 simulation runs, with $T = 4{,}000$ observations being generated in each run. In each run, the final 1,000 observations of the generated series were used only to assess the predictive capacity of various models, which were previously fitted to the first 3,000 observations. To make a fairly extensive simulation study feasible, we did not conduct a cross-validation for the smoothing parameter λ within each simulation run. Instead, we ran cross-validations only in 10 preliminary simulation runs, trying the values 256, 512, 1,024, 2,048, 4,096 and 8,192, and then fixed λ for the main 200 simulation runs at the value which was selected most often by cross-validation in the preliminary runs (namely $\lambda = 1{,}024$). This procedure resulted in a good performance (see below), but in fact the results could potentially be further improved by conducting a cross-validation within each simulation run. We set $K = 15$, resulting in 31 B-spline basis densities that were used in the estimation. To obtain a benchmark for the semiparametric model $\mathrm{SV_{sp}}$, we further fitted the basic models $\mathrm{SV_0}$ and $\mathrm{SV_t}$ to each generated series, also using the HMM-based discretization approach described in Sect. 2.1.

For the $\mathrm{SV_{sp}}$ model, the sample mean estimates of the parameters ϕ and σ were obtained as 0.978 (sample standard error: 0.007) and 0.103 (0.017), respectively. For the $\mathrm{SV_0}$/$\mathrm{SV_t}$ models, the sample mean estimates of the parameters ϕ and σ were obtained as 0.972/0.977 (sample standard errors: 0.010/0.007) and 0.117/0.102

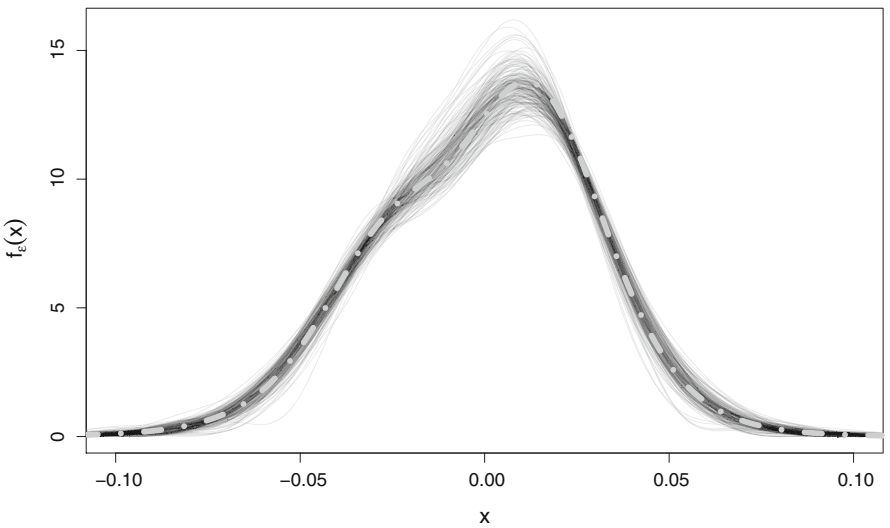

Fig. 1 True density of ε_t considered in the simulation experiments (*dashed grey line*) and its 200 estimates obtained using the nonparametric approach (*grey lines*)

(0.023/0.017), respectively. The SV_0 model underestimated the persistence parameter and overestimated the variance of the log-volatility process. The latter is due to the model's inability to capture the slight excess kurtosis of the true conditional distribution (see further comments on this issue in Sect. 4.2). In contrast, both the SV_{sp} model and the SV_t model yielded approximately unbiased estimates of the parameters related to the log-volatility process. Concerning the conditional process, y_t, Fig. 1 displays the true pdf of ε_t and the corresponding pdfs that were estimated using the nonparametric approach. From the graphic we can see that all 200 fits seem fairly reasonable.

In the given simulation scenario, we also assessed the predictive capacity of the three different modelling approaches, represented by the models SV_0, SV_t and SV_{sp}. We did this by calculating, in each simulation run and under each of the three types of models fitted to the first 3,000 observations, the log-likelihood score for the final 1,000 observations, denoted by $\text{llk}_i(SV_0)$, $\text{llk}_i(SV_t)$ and $\text{llk}_i(SV_{sp})$, with i indicating the simulation run. These scores were compared to the corresponding score obtained when using the true model, i.e., the one that was actually used to generate the artificial data; we denote this score by $\text{llk}_i(SV_{\text{true}})$. The average differences between the scores obtained for a given fitted model SV_* (either SV_0, SV_t or SV_{sp}) and the score obtained for the true model, i.e.,

$$\Delta_{SV_*,SV_{\text{true}}} = \frac{1}{200}\sum_{i=1}^{200}\big(\text{llk}_i(SV_*) - \text{llk}_i(SV_{\text{true}})\big),$$

are provided in Table 1. On average, the SV_{sp} model fitted the out-of-sample data substantially better than its parametric counterparts. Considering the individual simulation runs, the SV_{sp} model had a better predictive performance than the two parametric models in 176 out of 200 cases. These results are hardly surprising given the skewness of the distribution chosen for ε_t. Nevertheless, they do demonstrate both the practical feasibility and the potential benefits of our approach.

In order to have a benchmark for these results, we considered a second simulation scenario in which we generated data from the SV_t model, specifying $\varepsilon_t^{(0)}$ in (1) to be a Student-t distribution with 10 degrees of freedom and $\mu = 0.02$. Except of the true conditional distribution, this second simulation experiment was configured exactly as the first. The averaged differences between the scores obtained for the three different models fitted, SV_0, SV_t and SV_{sp}, and the score obtained for the true model are also provided in Table 1. In this scenario, the (correct) SV_t model had a better predictive performance than the two other models in 152 out of 200 cases, while the SV_{sp} model still showed a better predictive performance than the (incorrect) SV_0 model in 164

Table 1 Average differences between the out-of-sample predictive scores obtained for the three different models fitted, SV_0, SV_t and SV_{sp}, and the score obtained for the true model, for both simulation scenarios considered

	$\Delta_{SV_0,SV_{\text{true}}}$	$\Delta_{SV_t,SV_{\text{true}}}$	$\Delta_{SV_{sp},SV_{\text{true}}}$
Scenario with skewed and leptokurtic distribution	−9.48	−9.14	−3.52
Scenario with Student-t distribution	−5.44	−0.62	−2.30

cases. Thus, overall, the results demonstrate i) the potential of the nonparametric approach to considerably improve the predictive capacity in scenarios where the true conditional distribution deviates from the form imposed by either the normal or the Student-t distribution (e.g., if it is skewed), and ii) that it can perform almost as well as parametric modelling approaches in scenarios where those are adequate.

4 Application to real data

4.1 The data

The SV_{sp} model was fitted to series of daily log-returns for three stocks, namely Sony Corporation, Merck & Co. and Microsoft Corporation, and for the stock index S&P 500. The adjusted closing prices, p_t, for the period 03.01.2000–01.08.2013, were downloaded from "finance.yahoo.com", and the daily log-returns were computed as $y_t = \log(p_t/p_{t-1})$. To assess the out-of-sample predictive performance of various models, we divided each of the four series into two parts:

- In-sample period: 03.01.2000–31.12.2007,
- Out-of-sample period: 02.01.2008–01.08.2013.

The dividing date was chosen to lie before the outburst of the recent financial crisis, which culminated in the collapse of Lehman Brothers Holdings Inc. in September 2008. The four time series that were analyzed are displayed in Fig. 2.

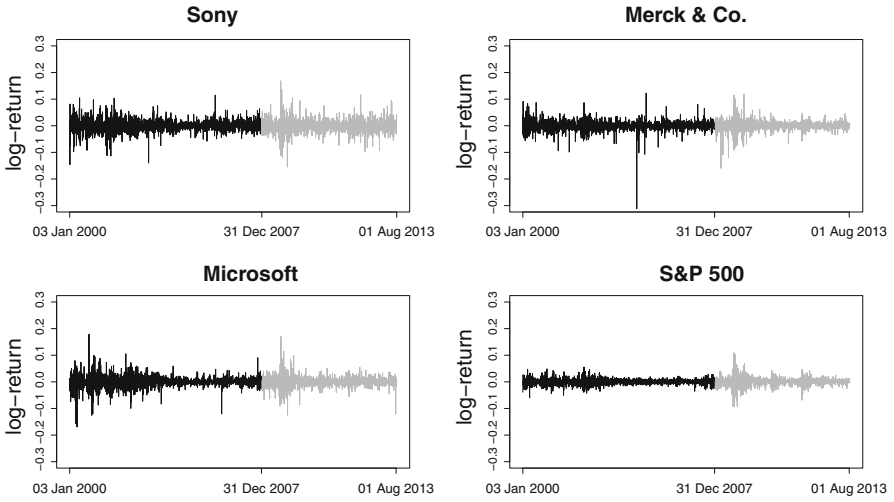

Fig. 2 Time series of log-returns on shares of three companies (Sony, Merck and Microsoft) and on the stock index S&P 500; the observations from the in-sample and out-of-sample periods are displayed in *black* and *grey*, respectively

4.2 Results

To each of the four series, the SV_{sp} model was fitted using $K = 20$ and hence 41 B-spline basis densities to represent the density of ε_t, $m = 100$ intervals in the discretization of the log-volatility process, numerically integrating over the log-volatility values from the interval $[b_0, b_{100}] = [-5, 5]$. Smoothing parameters were selected via cross-validation as described in Sect. 2.3.2, where we took $k = 10$ as the number of cross-validation partitions and $h = 50$ as the number of observations to be excluded from each validation block at both the start and the end. Fitting the SV_{sp} model took about 10 min per series on an i7 CPU, at 2.7 GHz and with 4 GB RAM.

For comparison purposes, we also considered the two basic models SV_0 and SV_t, and additionally a nonstandard parametric SV model with conditional distribution the skew Student-t discussed in Fernandez and Steel (1998); the latter model will be labeled SV_{skew-t}, and we denote the associated skewness parameter by γ.

Each model considered was fitted to the four series using the partition into in-sample period and out-of-sample period given in Sect. 4.1. Models were fitted using the data from the in-sample period only, and the data from the out-of-sample period were used to assess the predictive performance of the four different models (as detailed below).

The estimates of the parameters ϕ, σ, β (only for the parametric models), ν (only for the SV_t model and the SV_{skew-t} model) and γ (only for the SV_{skew-t} model), and the associated 95 % confidence intervals, for each of the four series considered, are given in Table 2. For each model considered, the confidence interval of each model

Table 2 Parameter estimates and 95 % bootstrap confidence intervals obtained for the four different models fitted to the four series of log-returns considered

	Sony	Merck	Microsoft	S&P 500
	SV_0			
$\hat{\phi}$	0.957 (0.931; 0.976)	0.825 (0.752; 0.878)	0.979 (0.964; 0.989)	0.991 (0.979; 0.997)
$\hat{\sigma}$	0.249 (0.196; 0.308)	0.545 (0.441; 0.660)	0.239 (0.195; 0.281)	0.114 (0.085; 0.144)
$\hat{\beta}$	0.019 (0.016; 0.021)	0.014 (0.013; 0.015)	0.015 (0.012; 0.020)	0.010 (0.007; 0.013)
	SV_t			
$\hat{\phi}$	0.992 (0.979; 0.999)	0.992 (0.976; 1.000)	0.994 (0.984; 1.000)	0.992 (0.983; 0.999)
$\hat{\sigma}$	0.092 (0.060; 0.132)	0.086 (0.050; 0.126)	0.116 (0.085; 0.151)	0.104 (0.073; 0.135)
$\hat{\beta}$	0.017 (0.013; 0.022)	0.012 (0.010; 0.015)	0.014 (0.009; 0.020)	0.009 (0.007; 0.012)
$\hat{\nu}$	6.70 (5.21; 9.81)	4.67 (3.89; 6.16)	6.31 (4.81; 8.62)	25.72 (12.82; ∞)
	SV_{skew-t}			
$\hat{\phi}$	0.993 (0.979; 0.999)	0.992 (0.976; 0.999)	0.994 (0.983; 1.000)	0.992 (0.980; 0.999)
$\hat{\sigma}$	0.092 (0.060; 0.131)	0.086 (0.052; 0.128)	0.116 (0.084; 0.151)	0.104 (0.075; 0.135)
$\hat{\beta}$	0.017 (0.013; 0.022)	0.012 (0.010; 0.015)	0.014 (0.009; 0.021)	0.009 (0.007; 0.012)
$\hat{\nu}$	6.69 (5.26; 10.08)	4.66 (3.92; 5.97)	6.33 (5.08; 8.93)	25.89 (13.44; ∞)
$\hat{\gamma}$	0.997 (0.970; 1.025)	1.014 (0.984; 1.043)	1.008 (0.977; 1.041)	1.010 (0.983; 1.037)
	SV_{sp}			
$\hat{\phi}$	0.994 (0.984; 0.999)	0.993 (0.982; 0.999)	0.994 (0.985; 0.999)	0.998 (0.993; 1.000)
$\hat{\sigma}$	0.087 (0.060; 0.119)	0.078 (0.052; 0.111)	0.122 (0.089; 0.161)	0.084 (0.062; 0.111)

parameter was taken as the empirical quantiles of 500 bootstrap replicates obtained from refitting the model to time series that were artificially generated from the fitted model (i.e., a parametric bootstrap).

The results obtained for the SV_0 model illustrate the problems of the conditional normal distribution to capture the extreme returns. Especially for the Merck stock, where on September 30, 2004, the withdrawal of the drug Rofecoxib from the market caused heavy losses, SV_0 performs badly. Indeed, the only way for the SV_0 model to cope with the associated extreme negative return of -0.31, which occurs in a period of calm market, is to assign a very high uncertainty to the log-volatility process (as expressed by a high $\hat{\sigma}$ and a small $\hat{\phi}$). This results in an undersmoothing of the volatility. By contrast, the SV_t and SV_{skew-t} models' leptokurtic conditional distribution leads to much more plausible estimates for ϕ and σ, with the results being similar to those obtained for the SV_{sp} model. In these three models, extreme returns are assigned to the tail of the return distribution, rather than to big jumps in the log-volatility process, as in the SV_0 model. The same pattern is found for the other two stock return series—although to a lesser extent. Only for the stock index S&P 500, the estimate of σ obtained using the SV_0 model is of the same magnitude as the corresponding estimates obtained when applying SV_t, SV_{skew-t} and SV_{sp}. This is not surprising since in the index the extreme returns of individual companies play a smaller role, which is also reflected by the much lighter tail of the conditional distribution in the fitted SV_t and SV_{skew-t} models.

Figure 3 displays the nonparametrically estimated densities of the conditional distribution in the SV_{sp} model and the associated pointwise 95 % confidence intervals, for the four series considered. Using 500 parametric bootstrap replicates, the point-

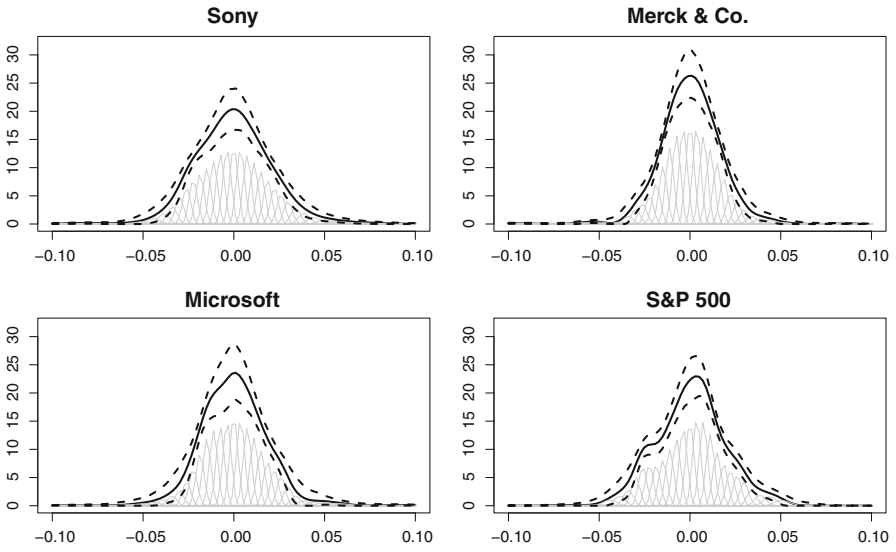

Fig. 3 Conditional densities of ε_t estimated using the nonparametric approach (*solid black lines*) and the associated bootstrap pointwise 95 % confidence intervals (*dashed black lines*), for the four series of log-returns considered, and underlying weighted B-splines that generate these densities via a linear combination (in *grey*)

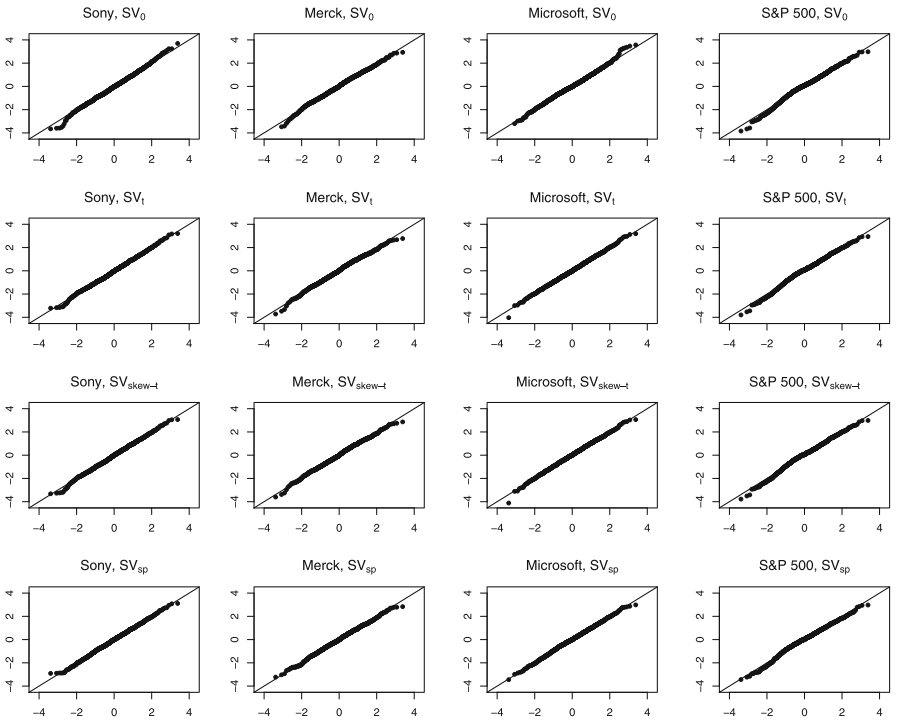

Fig. 4 Assessment of the absolute fit of the models in terms of their predictive performance: quantile–quantile plots of the out-of-sample forecast pseudo-residuals obtained from the fitted models SV_0 (*top row*), SV_t (*second row*), SV_{skew-t} (*third row*) and SV_{sp} (*bottom row*), for Sony (*first column*), Merck (*second column*), Microsoft (*third column*) and S&P 500 (*fourth column*). Sample quantiles are given on the *vertical axes*, and quantiles of the standard normal are given on the *horizontal axes*

wise confidence intervals for the densities of the state-dependent distributions were obtained as the quantiles of the bootstrap density estimates at a specific point in the support. The skewness of the nonparametrically estimated distribution f_ε is -0.73, -2.61, -0.20 and 0.54, for Sony, Merck, Microsoft and S&P 500, respectively. With the exception of S&P 500, this is in line with the stylized facts attributed to financial return series which propagate a gain/loss asymmetry as large drawdowns generally exceed large upward movements (Cont 2001), resulting in a left tail in the conditional distribution that is more extreme than the right tail (Durham 2006). However, to some extent the skewness also stems from an asymmetric density close to the center, a phenomenon that could be related to the lack of a leverage effect in our model. Indeed, Delatola and Griffin (2013) gave evidence that if data stem from an SV model with a strong leverage effect, then a nonparametric SV model not including this effect may infer a multimodal conditional distribution.

Quantile–quantile plots of the out-of-sample one-step-ahead forecast pseudo-residuals associated with the returns observed in the out-of-sample period 02.01.2008–01.08.2013, under the models fitted to the data from the in-sample period, are given in Fig. 4. The p values for the Jarque–Bera, Anderson–Darling and Kolmogorov–Smirnov tests applied to the pseudo-residuals are listed in Table 3. The results show

Table 3 Assessment of the absolute fit of the models in terms of their predictive performance: p values of Jarque–Bera, Anderson–Darling and Kolmogorov–Smirnov tests for normality applied to out-of-sample one-step-ahead ahead forecast pseudo-residuals

	Sony	Merck	Microsoft	S&P 500
	Jarque–Bera			
SV_0	0.048**	0.002**	<0.001**	<0.001**
SV_t	0.896	0.055*	0.795	<0.001**
SV_{skew-t}	0.886	0.045**	0.780	<0.001**
SV_{sp}	0.818	0.333	0.411	<0.001**
	Anderson–Darling			
SV_0	0.439	0.083*	0.015**	<0.001**
SV_t	0.450	0.008**	0.955	<0.001**
SV_{skew-t}	0.438	0.007**	0.970	<0.001**
SV_{sp}	0.694	0.233	0.206	<0.001**
	Kolmogorov–Smirnov			
SV_0	0.269	0.130	0.062*	<0.001**
SV_t	0.190	0.023**	0.523	<0.001**
SV_{skew-t}	0.192	0.015**	0.593	<0.001**
SV_{sp}	0.747	0.365	0.066*	<0.001**

* normality rejected at 10% level of significance;
** normality rejected at 5% level of significance

that the SV_0 model provides poor out-of-sample forecasts for all three stock price time series, as the corresponding pseudo-residuals show large deviations from normality. The SV_t and SV_{skew-t} models are able to adequately forecast the Sony and Microsoft stocks, but exhibit some problems in the forecasts for the Merck series. Overall, the SV_{sp} shows a slightly higher accuracy in the forecasts of the stock returns, especially with respect to the extreme negative returns. For the S&P 500 index, all four models perform badly, with each of the normality tests rejecting the null hypothesis of normally distributed pseudo-residuals for each of the considered models. While for the models SV_0, SV_t and SV_{skew-t} at least part of the reason for the poor performance is the inaccurate forecast of extreme negative returns, for the SV_{sp} model the reason for the poor performance lies solely in the inaccurate forecast of moderate losses, while more extreme losses are again captured well. We believe the reason for the bad performance in this particular case to lie in the long persistent decline of stock prices during the financial and economic crisis which ensued the collapse of Lehman Brothers.

As in the simulation study, we also calculated the log-likelihood scores for the observations from the out-of-sample period, for the four different models and each of the four series analyzed. The results of this comparative assessment of the out-of-sample predictive performance of the models are displayed in Table 4. For the stock returns, the log-likelihood scores portray a similar picture as above, with the SV_0 model performing worse than the three models SV_t, SV_{skew-t} and SV_{sp}, which again perform similarly well. For the index S&P 500, the performance of the SV_{sp} model is relatively poor, and the simple model SV_0 performs about as well as the more flexible models, since the index exhibits less extreme dynamics due to the averaging over

Table 4 Assessment of the relative fit of the models in terms of their predictive performance: log-likelihood scores for the out-of-sample period

	Sony	Merck	Microsoft	S&P 500
$\text{llk}(\text{SV}_0)$	3,265.31	3,891.75	3,778.64	4,228.95
$\text{llk}(\text{SV}_\text{t})$	3,269.58	3,913.12	3,799.58	4,230.53
$\text{llk}(\text{SV}_{\text{skew}-\text{t}})$	3,269.67	3,913.74	3,799.98	4,231.09
$\text{llk}(\text{SV}_{\text{sp}})$	3,271.38	3,914.65	3,799.08	4,226.11

For each company, the highest score is underlined

several stocks. However, the quantile–quantile plots indicate that even for S&P 500 our semiparametric model has a slightly improved forecast accuracy at the extreme end of the lower tail.

5 Discussion

The stylized facts of asset returns indicate that simple parametric distributions, such as the normal or the Student's t-distribution, may not be well-suited to describe the shape of the conditional distribution in SV models. Thus, a nonparametric modelling of the conditional distribution, which allows for heavy tails, gain/loss asymmetry and other unusual features, may bear considerable advantages. In this manuscript, we developed a powerful and flexible frequentist framework for a nonparametric estimation of the conditional distribution in a discrete-time SV model. The approach exploits the strengths of the HMM machinery, in particular allowing for model checking, forecasting and volatility estimation. Compared to alternative Bayesian estimation approaches (Jensen and Maheu 2010; Delatola and Griffin 2011, 2013), an advantage of our approach is its conceptual simplicity and the associated easy-to-implement estimation algorithm. In particular, with the HMM approach it is straightforward to modify the estimation code to allow for a variety of different model formulations, for example with log-volatility processes more complicated than an AR(1) (Langrock et al. 2012). Furthermore, model selection can easily be carried out based on information criteria, whereas in a Bayesian framework model selection can be more difficult.

The computational burden for estimating a model of the proposed type is low, namely in the order of a couple of minutes for the considered series and fixed smoothing parameter. Applying cross-validation techniques to choose a data-driven smoothing parameter is, however, computationally demanding: computing times for this part of the analysis were about 2–3 h per series we analyzed. These computing times could be substantially reduced by employing parallel computing. Although the model specifications are not directly comparable, Delatola and Griffin (2011) report computing times of up to a day for the model fitting with their Bayesian approach, for series that were slightly longer than the ones we considered and when fixing the concentration parameter for the Dirichlet process mixture (which represents an analogue to the smoothing parameter in our setting). On the other hand, uncertainty quantification is

routinely achieved in the Bayesian framework by studying the variability of the posterior samples, whereas we employed computationally expensive bootstrap techniques.

A technical issue with the presented method which calls for further research concerns the configuration of the B-spline basis densities used in the estimation. We employed an ad hoc approach to account for the fact that in the tails of the conditional distribution only few observations are available to infer the shape of the density. Our approach effectively increases the penalty for non-smoothness in the tails of the distribution. The use of equally spaced sample quantiles, as suggested by Ruppert (2002), seems a promising avenue to explore in this regard. An alternative would be to follow the literature on adaptive smoothing parameter selection, e.g., Krivobokova et al. (2008), where the smoothing parameter would be specified as another spline function on the log-volatility domain. However, this would considerably increase the complexity of the likelihood-based analysis and may therefore easily lead to numerical or identifiability problems.

While we modelled the conditional distribution in the SV model in a nonparametric way, we still assumed a parametric distribution form of the innovations in the log-volatility process, which is not necessary. Furthermore, the possible incorporation of leverage effects into the model—i.e., the explicit modelling of a (negative) correlation between returns and subsequent log-volatilities, as often done in parametric SV modelling (e.g., Harvey and Shephard 1996; Jacquier et al. 2004)—was not discussed in the present manuscript, since we felt that in this first step towards a frequentist framework for semiparametric SV modelling it is advisable to focus on the inferential machinery, rather than on exploring the various possible variations in the model structure. Corresponding extensions are to be explored in future research.

Overall, the approach shows promise as a useful novel tool for analyzing time series of daily log-returns. We have illustrated that the approach *can* lead to an improved predictive capacity compared to basic parametric SV models, and in the real data analyses we found some notable distributional shapes. In particular, our model revealed negative skewness and heavy tails in the conditional distribution of the returns we analyzed, while still identifying the behavior of the log-volatilities that is typical of SV models. In out-of-sample comparisons, the parametric model with Student-t conditional distribution performed about as well as our semiparametric model, and both performed much better than the model with Gaussian conditional distribution, at least for stock returns. It should be noted here that all validation samples that we considered involve rather extreme dynamics, since they comprise the recent financial crisis. Further research needs to be done to investigate the performance of our approach in different scenarios, including calmer markets. The present manuscript is intended mainly to introduce the frequentist estimation framework and to outline its potential. While much work remains to be done, we strongly believe that the model's flexibility with regard to describing asymmetries and extreme events, and the relative accessibility of the maximum likelihood framework combined with P-spline techniques, will render our approach and potential future extensions a useful tool in portfolio management.

6 Supplementary Material

R and C++ code 1) to generate artificial data as in the simulation study and 2) to fit, to the generated data, the SV_0, SV_t and SV_{sp} models (format: main R code in .R file and additional C++ code in .cpp file).

Acknowledgments The authors would like to thank the two anonymous referees who provided useful comments on an earlier version of this manuscript.

Appendix: HMM essentials

This appendix reviews some HMM basics. A standard m-state HMM has the same two process structure as SV models and SSMs, only that the unobserved process is a Markov chain and hence discrete-valued rather than continuous-valued. Consider an HMM with observable process $\{X_t\}_{t=1}^T$ and underlying Markov chain $\{S_t\}_{t=1}^T$. Given the current state of S_t, the variable X_t is usually assumed to be conditionally independent from previous and future observations and states. The Markov chain is typically considered to be of first order, and the probabilities of transitions between the different states are summarized in the $m \times m$ transition probability matrix $\mathbf{\Gamma} = \big(\gamma_{ij}\big)$, where $\gamma_{ij} = \Pr\big(S_{t+1} = j | S_t = i\big)$, $i, j = 1, \ldots, m$. The initial state probabilities are summarized in the vector $\boldsymbol{\pi}$, where $\pi_i = \Pr(S_1 = i)$, $i = 1, \ldots, m$. It is usually convenient and appropriate to assume $\boldsymbol{\pi}$ to be the stationary distribution. For the described HMM, with observations given by $x_1, \ldots, x_T$ and underlying states denoted by $s_1, \ldots, s_T$, the likelihood is given by

$$\begin{aligned} \mathcal{L}^{\text{HMM}} = f(x_1, \ldots, x_T) &= \sum_{s_1=1}^{m} \cdots \sum_{s_T=1}^{m} f(x_1, \ldots, x_T | s_1, \ldots, s_T) f(s_1, \ldots, s_T) \\ &= \sum_{s_1=1}^{m} \cdots \sum_{s_T=1}^{m} \pi_{s_1} \prod_{t=1}^{T} f(x_t | s_t) \prod_{t=2}^{T} \gamma_{s_{t-1}, s_t}. \end{aligned}$$

In this form the likelihood involves m^T summands, which would make a numerical maximization infeasible in most cases. However, there is a much more efficient way of calculating the likelihood $\mathcal{L}^{\text{HMM}}$, given by a recursive scheme called the *forward algorithm*. To see this, we consider the vectors of forward variables, defined as $\boldsymbol{\alpha}_t = \big(\alpha_t(1), \ldots, \alpha_t(m)\big)$, $t = 1, \ldots, T$, where $\alpha_t(j) = f(x_1, \ldots, x_t, S_t = j)$, $j = 1, \ldots, m$. We then have the recursion:

$$\boldsymbol{\alpha}_1 = \boldsymbol{\pi} \mathbf{Q}(x_1), \qquad \boldsymbol{\alpha}_{t+1} = \boldsymbol{\alpha}_t \mathbf{\Gamma} \mathbf{Q}(x_{t+1}), \tag{10}$$

where $\mathbf{Q}(x_t) = \text{diag}\big(f_1(x_t), \ldots, f_m(x_t)\big)$, with $f_i(x_t) = f(x_t | S_t = i)$. The recursion (10) can be derived in a straightforward manner using the HMM dependence structure.

The likelihood can then be written as a matrix product:

$$\mathcal{L}^{\text{HMM}} = \sum_{i=1}^{m} \alpha_T(i) = \boldsymbol{\pi}\mathbf{Q}(x_1)\boldsymbol{\Gamma}\mathbf{Q}(x_2)\ldots\boldsymbol{\Gamma}\mathbf{Q}(x_T)\mathbf{1},$$

where $\mathbf{1} \in \mathbb{R}^m$ is a column vector of ones. For a missing observation x_t, the associated matrix $\mathbf{Q}(x_t)$ is simply replaced by the $m \times m$ identity matrix. For more details on HMMs, see for example Zucchini and MacDonald (2009).

References

Abanto-Valle CA, Bandyopadhyay D, Lachos V, Enriquez I (2010) Robust Bayesian analysis of heavy-tailed stochastic volatility models using scale mixtures of normal distributions. Comput Stat Data Anal 54:2883–2898

Abraham B, Balakrishna N, Sivakumar R (2006) Gamma stochastic volatility models. J Forecast 25:153–171

Arlot S, Celisse A (2010) Model selection. Stat Surv 4:40–79

Bartolucci F, De Luca G (2001) Maximum likelihood estimation of a latent variable time-series model. Appl Stoch Models 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

Bergmeier C, Benítez JM (2012) On the use of cross-validation for time series predictor evaluation. Inf Sci 191:192–213

Chib S, Nardari F, Shephard N (2002) Markov chain Monte Carlo methods for stochastic volatility models. J Econom 108:281–316

Cont R (2001) Empirical properties of asset returns: stylized facts and statistical issues. Quant Financ 1:223–236

de Boor C (1978) A practical guide to splines. Springer, Berlin

Delatola E-I, Griffin JE (2011) Bayesian nonparametric modelling of the return distribution with stochastic volatility. Bayesian Anal 6:901–926

Delatola E-I, Griffin JE (2013) A Bayesian semiparametric model for volatility with a leverage effect. Comput Stat Data Anal 60:97–110

Durham GB (2006) Monte Carlo methods for estimating, smoothing, and filtering one- and two-factor stochastic volatility models. J Econom 133:273–305

Eilers PHC, Marx BD (1996) Flexible smoothing with B-splines and penalties. Stati Sci 11:89–121

Fridman M, Harris L (1998) A maximum likelihood approach for non-Gaussian stochastic volatility models. J Bus Econ Stat 16:284–291

Fernandez C, Steel MFJ (1998) On Bayesian modeling of fat tails and skewness. J Am Stat Assoc 93: 359–371

Gallant AR, Hsieh D, Tauchen GE (1997) Estimation of stochastic volatility models with diagnostics. J Econom 81:159–192

Gneiting T, Raftery AE (2007) Strictly proper scoring rules, prediction, and estimation. J Am Stat Assoc 102:359–378

Harvey AC, Shephard N (1996) Estimation of an asymmetric stochastic volatility model for asset returns. J Bus Econ Stat 14:429–434

Harvey CR, Siddique A (2000) Conditional skewness in asset pricing tests. J Financ 55:1263–1295

Jacquier E, Polson NG, Rossi PE (2004) Bayesian analysis of stochastic volatility models with fat-tails and correlated errors. J Econom 122:185–212

Jondeau E, Rockinger M (2003) Conditional volatility, skewness and kurtosis: existence, persistence and co-movements. J Econ Dyn Control 27:1699–1737

Jensen MJ, Maheu JM (2010) Bayesian semiparametric stochastic volatility modeling. J Econom 157: 306–316

Kim S, Shephard N, Chib S (1998) Stochastic volatility: likelihood inference and comparison with ARCH models. Rev Econ Stud 65:361–393

Krivobokova T, Crainiceanu CM, Kauermann G (2008) Fast adaptive penalized splines. J Comput Graph Stat 17:1–20
Langrock R (2011) Some applications of nonlinear and non-Gaussian state-space modelling by means of hidden Markov models. J Appl Stat 38:2955–2970
Langrock R, King R (2013) Maximum likelihood estimation of mark-recapture-recovery models in the presence of continuous covariates. Ann Appl Stat 7:1709–1732
Langrock R, Kneib T, Sohn A, DeRuiter SL (2014) Nonparametric inference in hidden Markov models using P-splines. Biometrics (to appear). arXiv:1309.0423v2
Langrock R, MacDonald IL, Zucchini W (2012) Some nonstandard stochastic volatility models and their estimation using structured hidden Markov models. J Empir Financ 19:147–161
Nakajima J, Omori Y (2012) Stochastic volatility model with leverage and asymmetrically heavy-tailed error using GH skew Student's -distribution. Comput Stat Data Anal 56:3690–3704
Racine J (2000) Consistent cross-validatory model-selection for dependent data: hv-block cross-validation. J Econom 99:39–61
Rosenblatt M (1952) Remarks on a multivariate transformation. Ann Math Stat 23:470–472
Ruppert D (2002) Selecting the number of knots for penalized splines. J Comput Graph Stat 11:735–757
Ruppert D, Wand MP, Carroll RJ (2003) Semiparametric regression. Cambridge University Press, Cambridge
Schellhase C, Kauermann G (2012) Density estimation and comparison with a penalized mixture approach. Comput Stat 27:757–777
Shephard N (1996) Statistical aspects of ARCH and stochastic volatility. In: Cox DR, Hinkley DV, Barndorff-Nielsen OE (eds) Time series models: in econometrics, finance and other fields. Chapman & Hall, London, pp 1–67
Zucchini W, MacDonald IL (2009) Hidden Markov Models for time series: an introduction using R. Chapman & Hall, London