A note on calculating expected shortfall for discrete time stochastic volatility models

Michael Grabchak and Eliana Christou[*]

*Correspondence:
echris15@uncc.edu
Department of Mathematics
and Statistics, University
of North Carolina
at Charlotte, 9201 University
City Blvd, Charlotte, NC
28223, USA

Abstract

In this paper we consider the problem of estimating expected shortfall (ES) for discrete time stochastic volatility (SV) models. Specifically, we develop Monte Carlo methods to evaluate ES for a variety of commonly used SV models. This includes both models where the innovations are independent of the volatility and where there is dependence. This dependence aims to capture the well-known leverage effect. The performance of our Monte Carlo methods is analyzed through simulations and empirical analyses of four major US indices.

Keywords: Expected shortfall, Stochastic volatility, Value-at-risk

Introduction

Empirical studies consistently show that financial returns do not have a constant volatility and instead exhibit volatility clustering. This clustering is often modeled using GARCH or one of its variants. Perhaps the most prominent alternative to GARCH is the class of discrete time stochastic volatility (SV) models. These models are very flexible and capture additional stylized features of financial returns including skewness, excess kurtosis, and leverage effects (Cont and Tankov 2004). SV models are often credited to Taylor (1986) although they have a long prehistory. See Shephard (2005) or Taylor (1994) for a thorough review.

A lot of research effort has focused on option pricing for SV models, much less attention has been paid to the problem of calculating risk. However, as we will see, this problem is not trivial even in the case when all of the parameters are explicitly known. In this paper, we introduce a Monte Carlo method for calculating expected shortfall (ES) for several important classes of SV models. ES is one of the best known and most commonly used measures of financial risk. It is, arguably, second in popularity only to Value-at-Risk (VaR). However, unlike VaR, ES is a coherent risk measure (Artzner et al. 1999) and it has been chosen to replace VaR as the measure determining a bank's capital requirements in the Basel III regulatory framework (Basel Committee on Banking Supervision 2013). For more information on VaR, ES, and related risk measures, see e.g. McNeil et al. (2015) and the references therein. A well-known survey on the estimation of ES is given in Nadarajah et al. (2014). Among a long list of methodologies, that paper discusses the

estimation of ES under a GARCH model. However, we have not seen a discussion of ES estimation under an SV model in the literature.

The difficulty in calculating ES for SV models lies in the fact that one needs to work with the product of two random variables and, even in the case where both terms in the product have simple distributions, the distribution of the product may be quite complicated. This is in contrast with GARCH models, where the problem of evaluating ES, essentially, reduces to that of calculating ES for the distribution of the innovations.

The rest of this paper is organized as follows. In Sect. 2 we formally define the SV model and give a simple Monte Carlo method for evaluating ES in this case. In Sect. 3 we give a more sophisticated Monte Carlo method in the commonly used case where the innovations for the returns and for the volatility are independent. In Sect. 4 we give a similar method for an important case with dependence, which aims to model leverage. In Sect. 5 we illustrate our methodology on four major US indices. Some conclusions are given in Sect. 6.

Stochastic volatility

Discrete time stochastic volatility models commonly assume that the financial (log) return, at time t, is given by

$$r_t = e^{h_t/2}\epsilon_t, \tag{1}$$

where

$$h_t = \mu + \phi(h_{t-1} - \mu) + \sigma\eta_t \tag{2}$$

is the log variance. Here $\sigma > 0, |\phi| < 1$, and $\mu \in \mathbb{R}$ are parameters, and $\{\epsilon_t\}$ and $\{\eta_t\}$ are sequences of independent and identically distributed (iid) random variables representing the innovations for r_t and h_t, respectively. We do not, in general, assume that for a given t, ϵ_t and η_t are independent of each other. Note that the log variance is modeled by an AR(1) process. The assumption that $|\phi| < 1$ ensures that this process is weakly stationary, see Ruppert and Matteson (2015). Under general conditions on the distributions of the innovations, this model can be seen as a discretization of a continuous time SV model where the log variance is modeled by a process of Ornstein–Uhlenbeck type, see Taylor (1994) or Barndorff-Nielsen and Shephard (2001). We are interested in evaluating the ES for this model.

We begin by establishing some notation. Let $\mathcal{F}_{t-1}$ denote the information set available at time $t - 1$. For simplicity, we sometimes write P_{t-1} to denote the conditional probability $\mathrm{P}_{t-1}(\cdot) = \mathrm{P}(\cdot|\mathcal{F}_{t-1})$ and E_{t-1} to denote the conditional expectation $\mathrm{E}_{t-1}(\cdot) = \mathrm{E}(\cdot|\mathcal{F}_{t-1})$. For $\tau \in (0, 1)$, the τth VaR at time t, denoted by $\mathrm{VaR}_\tau(t)$, is the smallest number for which $\mathrm{P}_{t-1}\{r_t < -\mathrm{VaR}_\tau(t)\} \le \tau$. Note that $-\mathrm{VaR}_\tau(t)$ is the τth conditional (given $\mathcal{F}_{t-1}$) quantile of r_t. For this reason, we sometimes write $Q_\tau(r_t|\mathcal{F}_{t-1})$ for $-\mathrm{VaR}_\tau(t)$. The τth ES at time t, denoted by $\mathrm{ES}_\tau(t)$, is defined by

$$\mathrm{ES}_\tau(t) = \frac{1}{\tau}\int_0^\tau \mathrm{VaR}_s(t)\mathrm{d}s,$$

when the integral exists, and is undefined otherwise. The parameter τ is typically chosen to be a small number such as 0.01, 0.025, or 0.05. Throughout, we assume

1. that the distribution of r_t is continuous, and
2. that it satisfies

$$E_{t-1}(|r_t|) < \infty. \tag{3}$$

The second assumption ensures that $ES_\tau(t)$ is well defined, while the first allows us to use the more explicit formula

$$ES_\tau(t) = E_{t-1}[-r_t | -r_t > VaR_\tau(t)] = -\frac{1}{\tau}E_{t-1}[r_t 1\{r_t < -VaR_\tau(t)\}].$$

Here and throughout, we write $1\{\cdot\}$ to denote the indicator function.

Using the fact that the innovations are independent over time, together with basic properties of quantiles and expectations, we can write

$$ES_\tau(t) = -e^{\{\mu(1-\phi)+\phi h_{t-1}\}/2}M(\tau,\sigma), \tag{4}$$

where

$$M(\tau,\sigma) = \frac{1}{\tau}E\left[e^{\sigma Y/2}Z 1\{e^{\sigma Y/2}Z < a\}\right], \tag{5}$$

$a = Q_\tau\left(e^{\sigma Y/2}Z\right)$ is the τth (unconditional) quantile of the random variable $e^{\sigma Y/2}Z$, and the joint distribution of (Y, Z) is the same as the joint distribution of (η_t, ϵ_t). The difficulty in evaluating M is that we must work with the distribution of $X = e^{\sigma Y/2}Z$, which can be complicated even when the distributions of Y and Z are fairly simple. Little is known about the distribution of X even in the case where Y and Z are both standard normal random variables, see Yang (2008) and the references therein. For this reason, we develop Monte Carlo methods to approximate $M(\tau,\sigma)$.

We begin by approximating $a = Q_\tau(e^{\sigma Y/2}Z)$. Toward this end, fix some large integer N_1 and simulate an iid sequence of bivariate random variables $\{(Y_i, Z_i)\}_{i=1}^{N_1}$ from the joint distribution of (η_t, ϵ_t). Next, for $i = 1, 2, \ldots, N_1$, set $X_i = e^{\sigma Y_i/2}Z_i$. Now sort these from smallest to largest to get $X_{(1)} \le X_{(2)} \le \cdots \le X_{(N_1)}$. Finally, approximate $a = Q_\tau(e^{\sigma Y/2}Z)$ by

$$\widehat{a} = X_{(\lfloor \tau N_1 \rfloor)}, \tag{6}$$

where $\lfloor \cdot \rfloor$ is the floor function. One can also use a smooth approximation using kernel estimators, see e.g. Sheather and Marron (1990). However, we did not find much of an improvement when using these. Next, fix another large integer N_2 and simulate a new iid sequence $\{(Y_i, Z_i)\}_{i=1}^{N_2}$ from the joint distribution of (η_t, ϵ_t) and approximate $M(\tau,\sigma)$ by

$$\widehat{M}_1(\tau,\sigma) = \frac{1}{N_2\tau}\sum_{i=1}^{N_2} e^{\sigma Y_i/2}Z_i 1\{e^{\sigma Y_i/2}Z_i \le \widehat{a}\}. \tag{7}$$

We note that, in principle one can use the same dataset to evaluate a and $M(\tau,\sigma)$ although for smaller sample sizes this may create bias. Either way, the difficulty with this

approach is that approximately $(1 - \tau)100\%$ of the simulated values will not satisfy the condition in the indicator function in (7) and will thus be thrown out. As such, very few values will actually be used in the sum. For this reason, we may need N_2 to be an extremely large number to get a reasonable approximation. One could try to implement an importance sampling or related modification, but the fact that we are working with the product of two random variables, makes it difficult to use such an approach. Instead, we use the specific structure of this problem to implement an approach that works better in several important situations.

Independent case

It is commonly assumed that the sequences $\{\epsilon_t\}$ and $\{\eta_t\}$ are mutually independent. For simplicity and to ensure that the distribution of the returns is continuous, we assume that the distributions of ϵ_t and η_t are both continuous, having probability density functions (pdfs) f_ϵ and f_η, respectively. In order to guarantee that (3) holds, we must assume that

$$E(|\epsilon_t|) < \infty \text{ and } E(e^{0.5\sigma \eta_t}) < \infty. \tag{8}$$

By a conditioning argument, we have

$$M(\tau, \sigma) = \frac{1}{\tau} E\left[e^{\sigma Y/2} Z 1\{e^{\sigma Y/2} Z < a\} \right] = \frac{1}{\tau} E\left[e^{\sigma Y/2} E\left[Z 1\{e^{\sigma Y/2} Z < a\} | Y \right] \right]$$
$$= \frac{1}{\tau} E\left[e^{\sigma Y/2} H(Y, a, \sigma) \right],$$

where

$$H(y, a, \sigma) = E\left[Z 1\{e^{\sigma y/2} Z < a\} \right] = \int_{-\infty}^{ae^{-\sigma y/2}} x f_\epsilon(x) dx.$$

This can be used to develop a Monte Carlo method for approximating $M(\tau, \sigma)$ as follows. Fix some large integer N_1 and simulate two mutually independent sequences of iid random variables $\{Y_i\}_{i=1}^{N_1}$ and $\{Z_i\}_{i=1}^{N_1}$, where $Y_i \sim f_\eta$ and $Z_i \sim f_\epsilon$. Use these to approximate $a = Q_\tau(e^{\sigma Y/2} Z)$ by $\widehat{a}$ as in (6). Now choose another large integer N_2 and simulate $Y_1, \ldots, Y_{N_2}$ iid from f_η. We can then approximate $M(\tau, \sigma)$ by

$$\widehat{M}_2(\tau, \sigma) = \frac{1}{N_2 \tau} \sum_{i=1}^{N_2} e^{\sigma Y_i/2} H(Y_i, \widehat{a}, \sigma).$$

We now give explicit formulas for H in several important situations. Throughout we assume that $a \leq 0$, which holds for all reasonable choices of τ. Perhaps the most common assumptions are that f_ϵ is the pdf of a standard normal distribution or a t-distribution. In the standard normal case we have

$$H(y, a, \sigma) = \frac{-1}{\sqrt{2\pi}} e^{-\frac{1}{2}a^2 e^{-\sigma y}} \tag{9}$$

and in the case of a t-distribution with $\nu > 1$ degrees of freedom we have

$$H(y, a, \sigma) = \frac{-\sqrt{\nu}\Gamma\{(\nu+1)/2\}}{\sqrt{\pi}\Gamma(\nu/2)(\nu-1)}(1 + a^2 e^{-\sigma y}/\nu)^{-(\nu-1)/2}. \tag{10}$$

In the above, we need $\nu > 1$ as otherwise (8) will not hold. In practice, it is often assumed that the distributions of returns are skewed. To capture this, skewed modifications of normal and t-distributions are often used. While there are a number of ways to introduce such modifications, we follow the approach of Fernandez and Steel (1998). In general, this approach can be described as follows. If f_1 is the pdf of a distribution that is unimodal and symmetric around zero, then for $\gamma > 0$

$$f_\gamma(x) = \frac{2}{\gamma + \frac{1}{\gamma}}[f_1(x/\gamma)1\{0 \le x < \infty\} + f_1(x/\gamma)1\{-\infty < x < 0\}]$$

is a skewed modification of f_1. The parameter γ determines the skew of the distribution. When $\gamma = 1$ the distribution is symmetric, when $\gamma < 1$ it has a negative skewness, and when $\gamma > 1$ it has a positive skewness. Using change of variables, it is straightforward to show that, if H_1 corresponds to f_1, then
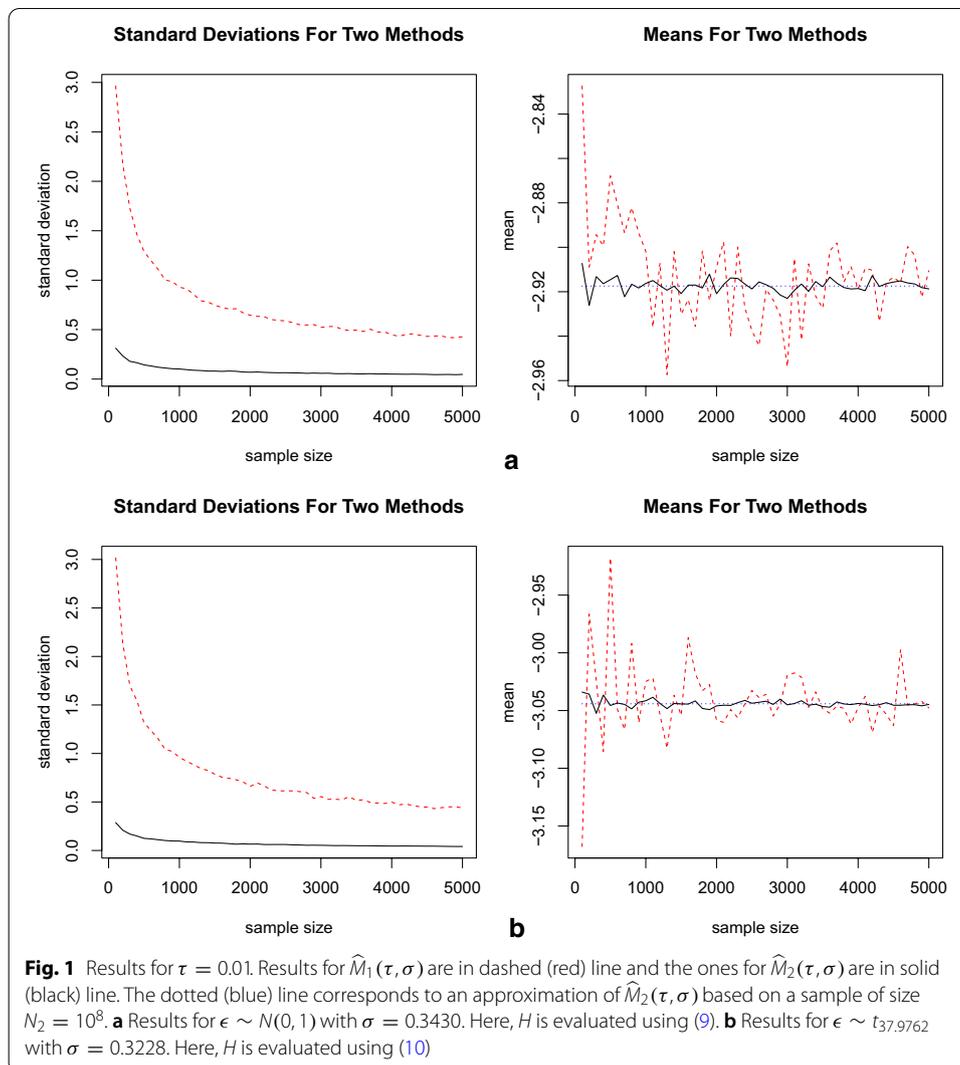
$$H_\gamma(y, a, \sigma) = \frac{2}{\gamma^3 + \gamma}H_1(y, \gamma a, \sigma)$$

corresponds to f_γ. We can easily apply this to get explicit formulas for H in the cases of skewed modifications of normal and t-distributions.

We now give a small simulation study to compare the performance of $\widehat{M}_1(\tau, \sigma)$ and $\widehat{M}_2(\tau, \sigma)$. For these simulations we assume that η_t has a standard normal distribution, while for ϵ_t we consider two distributions: standard normal and student-*t*. The values of the parameters, σ and (in the case of the student-*t* distribution) ν, were calibrated according to the daily returns from January 2014 to December 2019 of the S&P 500 Index. This was done using the stochvol package for the statistical software R, see Kastner (2016). We also performed similar simulations where the parameters were calibrated to the daily returns over the same period from the Russell 2000 Index, the Dow Jones Industrial Average, and the NASDAQ Composite Index. However, the results were similar and are not presented in the interest of space. Since our goal is to compare the two methods for evaluating $M(\tau, \sigma)$, we do not want issues with calculating a to interfere with the comparison. For this reason we choose $N_1 = 3 * 10^7$ to be a large value and use the same value of $\widehat{a}$ for all simulations with the same distribution. For N_2 we consider a range of values from 100 to 5000 in increments of 100. For each value of N_2, we estimate $\widehat{M}_1(\tau, \sigma)$ and $\widehat{M}_2(\tau, \sigma)$ 1000 times and report the standard deviations and the means over these trials. For $\tau = 0.01$, the results are presented in Fig. 1. From the plots, we can see that the second method has significantly less variance and that the mean gets close to the true value much quicker. We also repeated the procedure for $\tau = 0.025$ and 0.05, but the results were similar and are thus omitted. We note that we cannot allow η_t to have a *t*-distribution, as this would violate assumption (8).

Model with leverage

In the literature of financial returns, the leverage effect is the empirically observed phenomenon that volatility tends to be negatively correlated with returns, see e.g. Cont and Tankov (2004). In the important case where η_t and ϵ_t are jointly Gaussian, leverage is

Fig. 1 Results for $\tau = 0.01$. Results for $\widehat{M}_1(\tau, \sigma)$ are in dashed (red) line and the ones for $\widehat{M}_2(\tau, \sigma)$ are in solid (black) line. The dotted (blue) line corresponds to an approximation of $\widehat{M}_2(\tau, \sigma)$ based on a sample of size $N_2 = 10^8$. **a** Results for $\epsilon \sim N(0, 1)$ with $\sigma = 0.3430$. Here, H is evaluated using (9). **b** Results for $\epsilon \sim t_{37.9762}$ with $\sigma = 0.3228$. Here, H is evaluated using (10)

often modeled by assuming that the joint distribution of the random vector (η_t, ϵ_t) follows a bivariate normal distribution $N(0, \Sigma)$ where the covariance matrix is

$$\Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}, \tag{11}$$

for some $\rho \in (-1, 1)$, see Omori et al. (2007). When $\rho < 0$, the volatility is negatively correlated with the return, which captures the leverage effect. From properties of multivariate normal distributions, it follows that, in this case,

$$\epsilon_t \overset{d}{=} \rho W_1 + \sqrt{1 - \rho^2} W_2$$
$$\eta_t \overset{d}{=} W_1,$$

where W_1, W_2 are iid $N(0, 1)$ random variables and $\overset{d}{=}$ denotes equality in distribution. There does not seem to be a standard way to model leverage in the non-Gaussian case. However, one approach is suggested, in a continuous time setting, by Eq. (8) in

Barndorff-Nielsen and Shephard (2001). The idea is to add a constant times the innovation of the log volatility to the model for r_t. A variant of this idea, which is consistent with how the Gaussian case is treated, is to consider the model

$$r_t = e^{h_t/2}\left(\sqrt{1-\rho^2}\delta_t + \rho\eta_t\right),$$

where h_t is as in (2) and $\{\delta_t\}$ and $\{\eta_t\}$ are mutually independent sequences of iid random variables. The new parameter $\rho \in (-1, 1)$ determines the dependence between the return and the volatility. When $\rho = 0$, the model reduces to the independent case. Note that this is equivalent to taking

$$\epsilon_t = \sqrt{1-\rho^2}\delta_t + \rho\eta_t$$

in (1). For simplicity we assume that the distribution of δ_1 has pdf f_δ.

We now give an approach for evaluating $M(\tau, \sigma)$. In this case, we can write $Z \overset{d}{=} \rho W_1 + \sqrt{1-\rho^2}W_2$ and $Y \overset{d}{=} W_1$, where W_1, W_2 are independent random variables with $W_1 \sim f_\eta$ and $W_2 \sim f_\delta$. It follows that

$$
\begin{aligned}
&\mathrm{E}\left[e^{\sigma Y/2}Z1\{e^{\sigma Y/2}Z < a\}\right] \\
&= \mathrm{E}\left[e^{\sigma W_1/2}(\rho W_1 + \sqrt{1-\rho^2}W_2)1\{e^{\sigma W_1/2}(\rho W_1 + \sqrt{1-\rho^2}W_2) < a\}\right] \\
&= \rho\mathrm{E}\left[e^{\sigma W_1/2}W_11\{e^{\sigma W_1/2}(\rho W_1 + \sqrt{1-\rho^2}W_2) < a\}\right] \\
&\quad + \sqrt{1-\rho^2}\mathrm{E}\left[e^{\sigma W_1/2}W_21\{e^{\sigma W_1/2}(\rho W_1 + \sqrt{1-\rho^2}W_2) < a\}\right] \\
&= \rho\mathrm{E}\left[e^{\sigma W_1/2}W_1\mathrm{E}\left[1\{e^{\sigma W_1/2}(\rho W_1 + \sqrt{1-\rho^2}W_2) < a\}|W_1\right]\right] \\
&\quad + \sqrt{1-\rho^2}\mathrm{E}\left[e^{\sigma W_1/2}\mathrm{E}\left[W_21\{e^{\sigma W_1/2}(\rho W_1 + \sqrt{1-\rho^2}W_2) < a\}|W_1\right]\right] \\
&= \rho\mathrm{E}\left[e^{\sigma W_1/2}W_1H_1(W_1, a, \sigma, \rho)\right] \\
&\quad + \sqrt{1-\rho^2}\mathrm{E}\left[e^{\sigma W_1/2}H_2(W_1, a, \sigma, \rho)\right],
\end{aligned}
$$

where

$$H_1(y, a, \sigma, \rho) = F_\delta\left(\frac{e^{-\sigma y/2}a - \rho y}{\sqrt{1-\rho^2}}\right)$$

and

$$H_2(y, a, \sigma, \rho) = G_\delta\left(\frac{e^{-\sigma y/2}a - \rho y}{\sqrt{1-\rho^2}}\right).$$

Here

$$F_\delta(b) = \int_{-\infty}^{b} f_\delta(x)\mathrm{d}x, \quad b \in \mathbb{R}$$

is the cumulative distribution function (cdf) of the distribution of δ_t and

$$G_\delta(b) = \int_{-\infty}^{b} x f_\delta(x) \mathrm{d}x, \ \ b \in \mathbb{R}.$$

In the case where the distribution of δ_t is standard normal and $b \le 0$, $G_\delta(b) = -\varphi(b)$ where $\varphi(x) = e^{-x^2/2}/\sqrt{2\pi}$ is the pdf of the standard normal distribution.
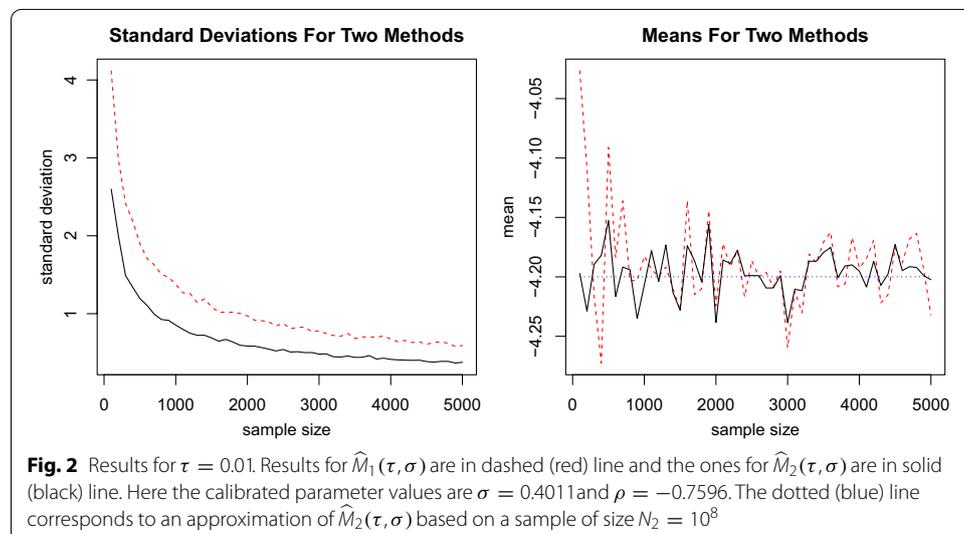
The above suggests the following Monte Carlo method. First, we approximate a by $\widehat{a}$ as in (6). Next, choose a large integer N_2 and simulate $W_1, \ldots, W_{N_2}$ iid from f_η. We can then approximate $M(\tau, \sigma)$ by

$$\widehat{M}_2(\tau, \sigma) = \frac{1}{N_2 \tau} \sum_{i=1}^{N_2} e^{\sigma W_i/2} \Big\{ \rho W_i H_1(W_i, \widehat{a}, \sigma, \rho) + \sqrt{1 - \rho^2} H_2(W_i, \widehat{a}, \sigma, \rho) \Big\}.$$

We again perform a small simulation study to compare the performance of $\widehat{M}_1(\tau, \sigma)$ and $\widehat{M}_2(\tau, \sigma)$. For these simulations we assume that δ_t and η_t are independent standard normal random variables, or equivalently that $(\eta_t, \epsilon_t) \sim N(0, \Sigma)$, where Σ is given by (11). The values of σ and ρ are calibrated according to the daily returns from January 2014 to December 2019 of the S&P 500 Index. This was again done using the stochvol package for R. For simulations we again took $N_1 = 3 * 10^7$ and N_2 from 100 to 5000 in increments of 100 and repeated each simulation over 1000 iterations. The results for $\tau = 0.01$ are given in Fig. 2. We again see that the second method has less variance and that the mean gets close to the true value quicker. However, the differences are not as strong in this case. As before, we also considered the case where the parameters were calibrated to the other three indices and when $\tau = 0.025$ and 0.05. Those results were similar and are not presented here in the interest of space.

Data analysis

In this section we perform a data analysis, where we estimate ES using SV models for four major US indices: the S&P 500 Index, the Russell 2000 Index, the Dow Jones Industrial Average, and the NASDAQ Composite Index. In all cases we use daily returns from

Fig. 2 Results for $\tau = 0.01$. Results for $\widehat{M}_1(\tau, \sigma)$ are in dashed (red) line and the ones for $\widehat{M}_2(\tau, \sigma)$ are in solid (black) line. Here the calibrated parameter values are $\sigma = 0.4011$ and $\rho = -0.7596$. The dotted (blue) line corresponds to an approximation of $\widehat{M}_2(\tau, \sigma)$ based on a sample of size $N_2 = 10^8$
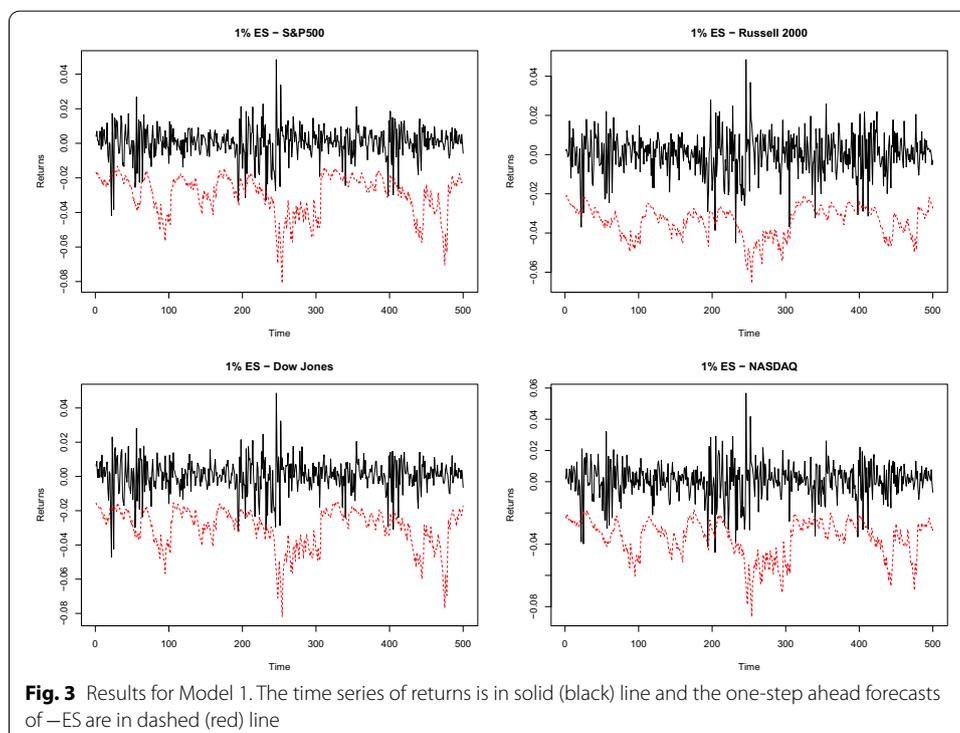
January 2014 to December 2019. For the analysis, we used the first observations as historical data and the last 500 observations for one-step ahead ES forecasts. The SV models that we consider are:

1. (SV-$\mathcal{N}$) ϵ_t and η_t are iid $N(0, 1)$;
2. (SV-t_2) ϵ_t and η_t are independent with $\epsilon_t \sim t_2$ and $\eta_t \sim N(0, 1)$;
3. (SV-t cal) ϵ_t and η_t are independent with $\epsilon_t \sim t_\nu$, where ν is calibrated to the data, and $\eta_t \sim N(0, 1)$;
4. (SV-lev) $(\eta_t, \epsilon_t) \sim N(0, \Sigma)$, where Σ is given by (11).

Models 1–3 assume that ϵ_t and η_t are independent, while Model 4 takes leverage into account.

The data analysis is performed as follows. We fix an index, an SV model, and a value of τ. For each of the last 500 observations, we use the stochvol package to calibrate the parameters of the SV model based on all of the observations before this one. The package gives multiple estimates for each parameter, and we take the mean of these as our estimate. We then use these parameter values to estimate ES_τ using (4), where we evaluate M using $\widehat{M}_2$ for the appropriate SV model. In all cases we take $N_1 = N_2 = 5000$. Note that the parameter values are recalibrated for each observation. In the interest of space we only report the results for $\tau = 0.01$, although we also repeated the procedure for $\tau = 0.025$ and 0.05. Figures 3, 4, 5 and 6 present the results for the four SV models, respectively. In each plot, we give the time series of the 500 data points, with the estimated $-ES_\tau$ overlaid. These values of $-ES_\tau$ are generally below the data, suggesting they do a good job of capturing risk. In Fig. 4, the values of $-ES_\tau$ are too far below the
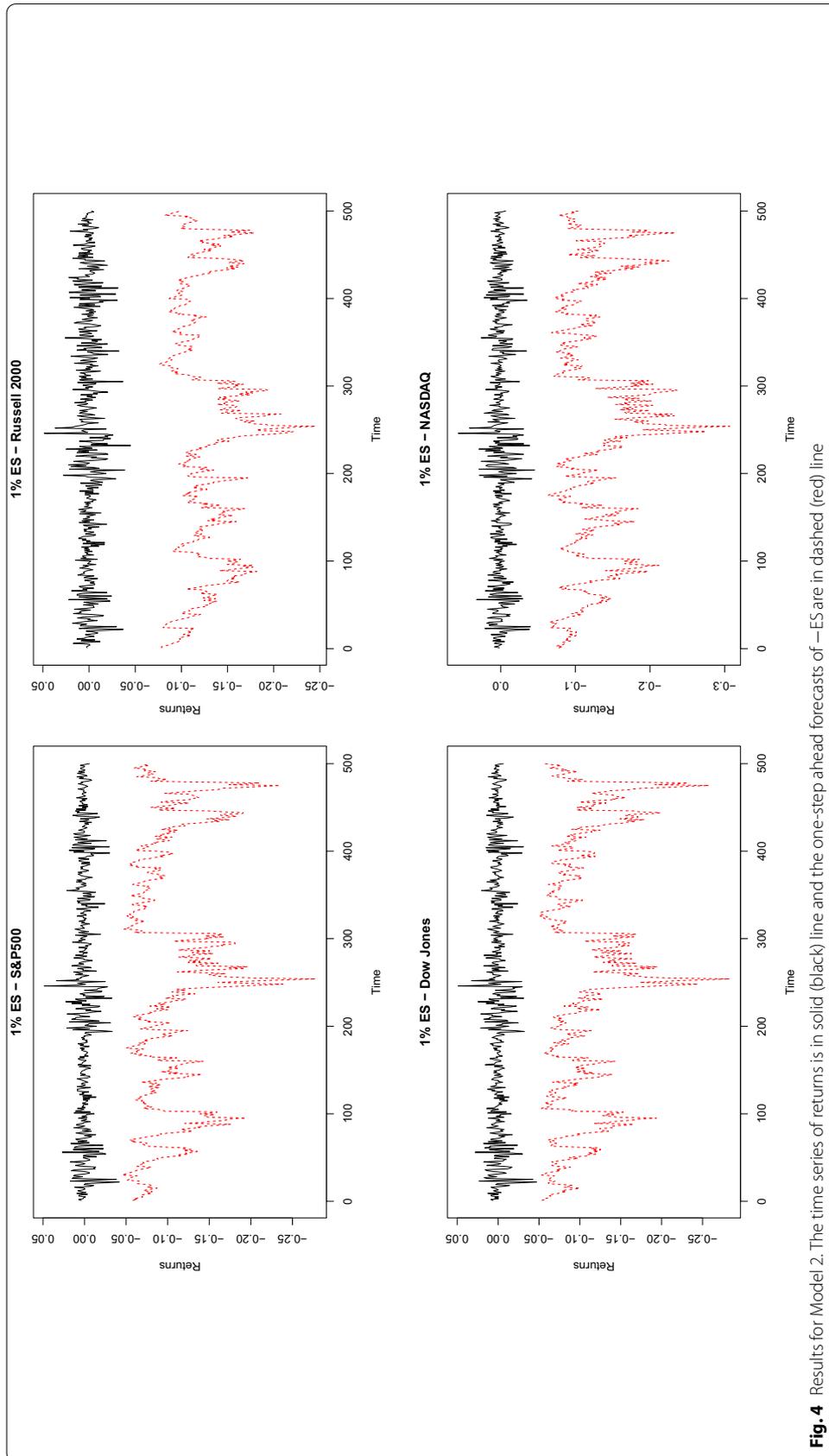

Fig. 3 Results for Model 1. The time series of returns is in solid (black) line and the one-step ahead forecasts of $-ES$ are in dashed (red) line

Fig. 4 Results for Model 2. The time series of returns is in solid (black) line and the one-step ahead forecasts of −ES are in dashed (red) line

Fig. 5 Results for Model 3. The time series of returns is in solid (black) line and the one-step ahead forecasts of −ES are in dashed (red) line

Fig. 6 Results for Model 4. The time series of returns is in solid (black) line and the one-step ahead forecasts of −ES are in dashed (red) line
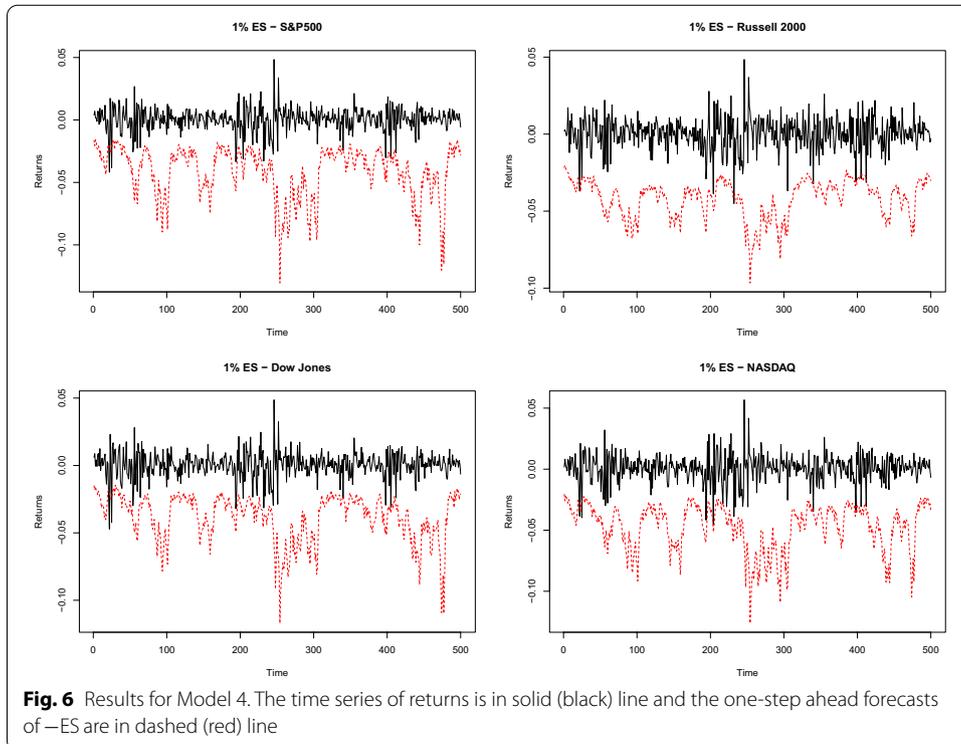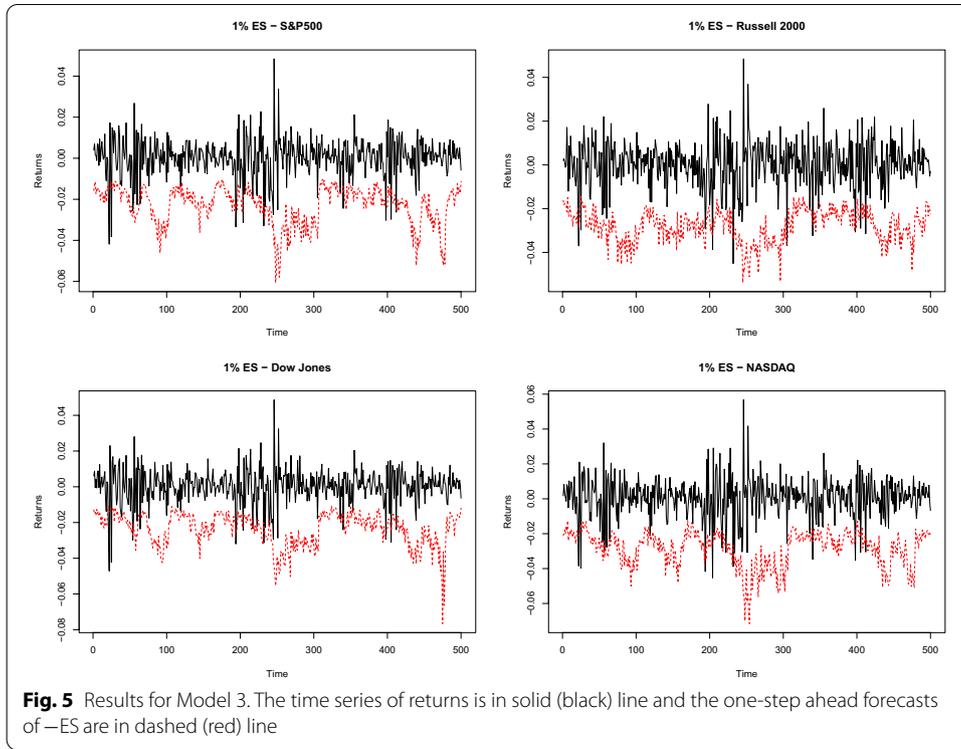
values of the time series, suggesting that the tails of this model are too heavy and that we should use a larger value for the degrees of freedom. This is done in Fig. 5, where the degrees of freedom are calibrated to the data. These calibrated values are all in the range from 20 to 40 degrees of freedom.

We also compared the performance of our approach with three well-known benchmark methods:

(1) (Hist) the historical method,
(2) (GARCH) the GARCH(1,1) method with normal innovations, and
(3) (DFGARCH) the distribution free GARCH(1,1) method.

For details see Nadarajah et al. (2014) or Christou and Grabchak (2021). We note that (2) is a special case of the QGARCH(1,1) method, where we take $h = 1$, $\mu = 0$, and estimate the standard deviation using a GARCH(1,1) model, and that (3) is the filtered historical method based on a GARCH(1,1) filter. We compare the performance of the proposed SV methods with these benchmark methods using backtesting.

There are many backtests for ES available in the literature, see Lazar and Zhang (2019) or Deng and Qiu (2021) for an overview. Many popular approaches, such as those in Du and Escanciano (2017), make parametric or semiparametric assumptions. Since we are comparing models that make a variety of different assumptions, we choose two backtests that make no such assumptions. Before giving these, we define some notation. Fix a model and assume that for each time period $t = 1, 2, \ldots, T$ we use this model to estimate $\mathrm{VaR}_\tau(t)$ by $\widehat{\mathrm{VaR}}_\tau(t)$ and $\mathrm{ES}_\tau(t)$ by $\widehat{\mathrm{ES}}_\tau(t)$. Our first backtest is from Acerbi and Székely (2014) and is based on

$$Z = \frac{1}{T\tau} \sum_{t=1}^{T} \frac{r_t I\{r_t > \widehat{\mathrm{VaR}}_t\}}{\widehat{\mathrm{ES}}_\tau(t)} + 1.$$

If we evaluate Z for several models, then the one where Z has the smallest absolute value is considered to be the best. An issue with this method is that it is sensitive to the estimate of VaR. Since different methods estimate VaR differently, this can make the results not fully comparable. The second backtest does not require estimating VaR. It is from Embrechts et al. (2005) and is based on

$$V = \frac{\sum_{t=1}^{T} D_t I\{D_t < \widehat{Q}(\tau)\}}{\sum_{t=1}^{T} I\{D_t < \widehat{Q}(\tau)\}},$$

where $D_t = r_t - \{-\widehat{\mathrm{ES}}_\tau(t)\}$ and $\widehat{Q}(\tau)$ is the empirical τth quantile of $\{D_t\}_{t=1}^{T}$. As in the previous case, the method where V has the smallest absolute value is considered to be the best. The results of these two backtests are given in Tables 1 and 2.

We begin by considering the results of the first backtest. The best method is clearly DFGARCH, which has a significantly better performance than the other methods. At first glance SV-t_2 appears to be one of the better models. However, this is likely an artifact of the way that statistic Z works. It is unbounded in the negative direction, but bounded by 1 in the positive direction, which implies that it penalizes underestimation of ES more than it penalizes overestimation. We have already seen that SV-t_2 drastically

Table 1 Backtesting results based on Z

Method	1%	2.5%	5%	1%	2.5%	5%
		SP500			*Russell*	
SV-$\mathcal{N}$	− 3.662	− 2.092	− 1.140	− 1.895	− 1.252	− 0.687
SV-t_2	0.725	0.672	0.586	1.000	0.917	0.777
SV-t cal	− 3.190	− 1.583	− 0.793	− 1.842	− 0.828	− 0.446
SV-lev	− 1.514	− 0.854	− 0.560	− 1.408	− 0.512	− 0.432
GARCH	− 2.565	− 1.039	− 0.585	− 1.936	− 0.761	− 0.365
DFGARCH	− 0.298	0.004	0.117	− 0.366	− 0.154	0.087
Hist	− 1.310	− 0.813	− 0.545	− 0.831	− 0.205	− 0.434
		Dow Jones			*NASDAQ*	
SV-$\mathcal{N}$	− 3.692	− 2.403	− 1.229	− 3.768	− 2.078	− 1.252
SV-t_2	0.719	0.707	0.585	1.000	0.701	0.652
SV-t cal	− 3.070	− 1.891	− 1.056	− 2.756	− 1.607	− 0.889
SV-lev	− 2.268	− 1.162	− 0.608	− 1.906	− 0.990	− 0.712
GARCH	− 3.121	− 1.425	− 0.608	− 2.662	− 1.082	− 0.673
DFGARCH	− 0.367	− 0.148	0.020	0.004	− 0.039	0.071
Hist	− 1.731	− 1.009	− 0.597	− 1.188	− 0.683	− 0.568

The method with the smallest absolute value is considered to be better

Table 2 Backtesting results based on V

Method	1%	2.5%	5%	1%	2.5%	5%
		SP500			*Russell*	
SV - $\mathcal{N}$	− 0.013	− 0.011	− 0.008	− 0.006	− 0.008	− 0.006
SV - t_2	0.032	0.024	0.016	0.068	0.044	0.029
SV - t cal	− 0.014	− 0.011	− 0.007	− 0.012	− 0.008	− 0.005
SV - lev	− 0.004	− 0.002	− 0.002	− 0.004	− 0.002	− 0.002
GARCH	− 0.010	− 0.007	− 0.005	− 0.010	− 0.007	− 0.004
DFGARCH	0.002	1.437×10^{-4}	0.612×10^{-4}	− 0.001	− 0.002	− 0.001
Hist	− 0.005	− 0.004	− 0.004	− 0.004	− 0.003	− 0.002
		Dow Jones			*NASDAQ*	
SV - $\mathcal{N}$	− 0.014	− 0.012	− 0.008	− 0.013	− 0.012	− 0.009
SV - t_2	0.032	0.025	0.016	0.045	0.032	0.021
SV - t cal	− 0.016	− 0.012	− 0.008	− 0.016	− 0.012	− 0.008
SV - lev	− 0.009	− 0.004	− 0.004	− 0.007	− 0.004	− 0.004
GARCH	− 0.011	− 0.007	− 0.005	− 0.012	− 0.008	− 0.006
DFGARCH	− 0.001	− 0.001	− 0.001	0.001	$− 1.521 \times 10^{-4}$	3.447×10^{-4}
Hist	− 0.007	− 0.006	− 0.005	− 0.006	− 0.004	− 0.005

overestimates ES, thus we cannot take its performance on this backtest as evidence that it works well. Aside from this, the best performing SV model is clearly SV-lev. It beats

GARCH in all cases except for 5% with Russell and NASDAQ, where it performs slightly worse. Further, it is generally comparable to, although slightly worse than, Hist. We now turn to the second backtest. This method does not seem to have an asymmetry in penalizing underestimation and overestimation of ES. We can now clearly see that SV-t_2 is the worst method. DFGARCH is again the best method. Here it is clear that SV-lev is second best, while Hist is a close third.

Conclusions

In this paper we considered the problem of estimating ES for SV models. To the best of our knowledge, this is the first paper to deal with this topic. We introduced two Monte Carlo methods, which are easy to implement in many common situations and can be used in both the case where the volatility is independent of the innovation and where there is dependence. This dependence aims to capture the leverage effect. Our simulations suggest that the second method has a lower variance and converges faster. As such it is the method that we suggest using. The other method is primarily introduced as it is more straightforward and can thus serve as a benchmark. We evaluated several variants of our method on real-world data and compared the results with three benchmark methods. We saw, in particular, that the SV model with leverage performed very well in backtests, although it was not the best. We note that we only considered a few simple distributional assumptions in the data analysis and that our methodology works with many other distributions. We leave the question of which distributions are the best to use in conjunction with an SV model for future work. We now discuss a simple extension of our work.

Thus far we have only considered SV models, where the volatility follows an AR(1) process as defined in (2). However, nothing in our approach depends on this structure and we can replace (2) by
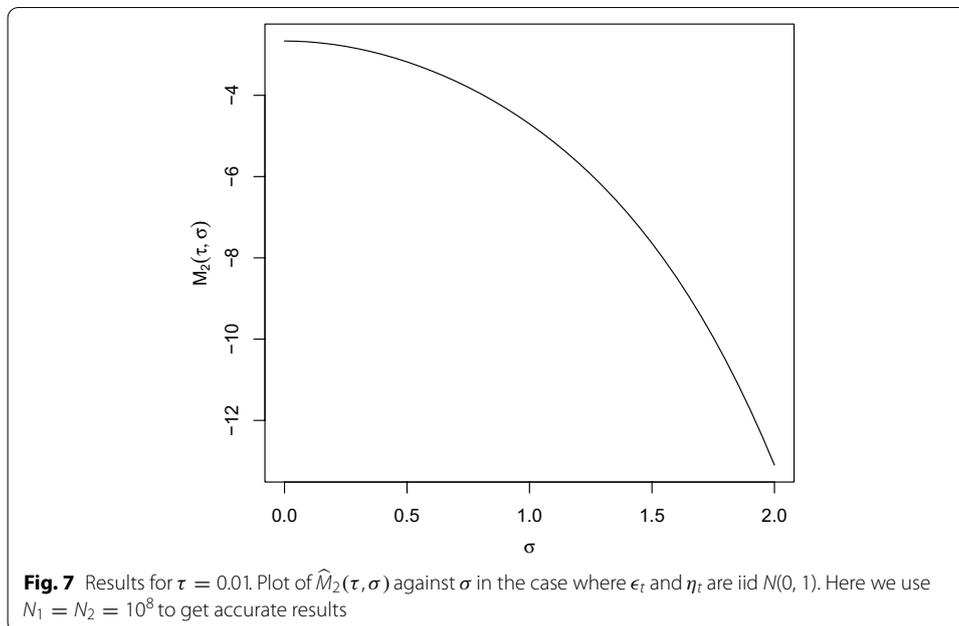
$$h_t = g_{t-1} + \sigma \eta_t$$

where g_{t-1} is any random variable that is measurable with respect to $\mathcal{F}_{t-1}$ and is independent of (η_t, ϵ_t). In this case we have

$$\text{ES}_\tau(t) = -e^{g_{t-1}/2} M(\tau, \sigma),$$

where $M(\tau, \sigma)$ is as in (5). Thus, our approach easily extends to more complicated time series structures.

We conclude by noting that in both this more general SV model and the one that we considered earlier, the main thing that needs to be evaluated is M, which only depends on τ, σ, and the joint distribution of (η_t, ϵ_t). In particular, it does not depend on g_{t-1}, which is only needed for a simple multiplicative term. Thus, if all parameters of the joint distribution of ϵ_t and η_t are known ahead of time, one can precompute the values of $M(\tau, \sigma)$ for the appropriate choice of τ on a grid of σ values. These can then be interpolated to get values very quickly. In the case where $\tau = 0.01$ and ϵ_t and η_t are independent standard normal random variables, a plot with this information is given in Fig. 7. The information in this plot, along with estimates of g_{t-1}, is all that is needed to evaluate ES in this case.

Fig. 7 Results for $\tau = 0.01$. Plot of $\widehat{M}_2(\tau, \sigma)$ against σ in the case where ϵ_t and η_t are iid $N(0, 1)$. Here we use $N_1 = N_2 = 10^8$ to get accurate results

Abbreviations
SV: Stochastic volatility; ES: Expected shortfall; VaR: Value-at-risk; iid: Independent and identically distributed.

References
Acerbi C, Székely B (2014) Back-testing expected shortfall. Risk 27:76–81
Artzner P, Delbaen F, Eber JM, Health D (1999) Coherent measures of risk. Math Finance 9:203–228
Barndorff-Nielsen OE, Shephard N (2001) Non-Gaussian Ornstein–Uhlenbeck-based models and some of their uses in financial economics. J R Stat Soc B 63:167–241
Basel Committee on Banking Supervision (2013) Consultative document, fundamental review of the trading book: a revised market risk framework. Basel, Switzerland. http://www.bis.org/publ/bcbs265.pdf
Christou E, Grabchak M (2021) Estimation of expected shortfall using quantile regression: a comparison study. Submitted
Cont R, Tankov P (2004) Financial modelling with jump processes. Chapman & Hall, Boca Raton
Deng K, Qiu J (2021) Backtesting expected shortfall and beyond. Quant Finance. https://doi.org/10.1080/14697688.2020.1834120
Du Z, Escanciano JC (2017) Backtesting expected shortfall: accounting for tail risk. Manage Sci 63:940–958
Embrechts P, Kaufmann R, Patie P (2005) Strategic long-term financial risks: single risk factors. Comput Optim Appl 32:61–90
Fernandez C, Steel MF (1998) On Bayesian modeling of fat tails and skewness. J Am Stat Assoc 93:359–371

Kastner G (2016) Dealing with stochastic volatility in time series using the R package stochvol. J Stat Softw. https://doi.org/10.18637/jss.v069.i05

Lazar E, Zhang N (2019) Model risk of expected shortfall. J Bank Finance 105:74–93

McNeil AJ, Frey R, Embrechts P (2015) Quantitative risk management: concepts, techniques and tools. Princeton University Press, Princeton

Nadarajah S, Zhang B, Chan S (2014) Estimation methods for expected shortfall. Quant Finance 14:271–291

Omori Y, Chib S, Shephard N, Nakajima J (2007) Stochastic volatility with leverage: fast and efficient likelihood inference. J Econom 140:425–449

Ruppert D, Matteson DS (2015) Statistics and data analysis for financial engineering with R examples, 2nd edn. Springer, New York

Sheather SJ, Marron JS (1990) Kernel quantile estimators. J Am Stat Assoc 85:410–416

Shephard N (2005) Stochastic volatility: selected readings. Oxford University Press, Oxford

Taylor SJ (1986) Modelling financial time series. Wiley, Chichester

Taylor SJ (1994) Modeling stochastic volatility a review and comparative study. Math Finance 4:183–204

Yang M (2008) Normal log-normal mixture, leptokurtosis and skewness. Appl Econ Lett 15:737–742

Publisher's Note