Bayesian Risk Evaluation in State Space Models using Importance Sampling

Agnieszka Borowska, Lennart F. Hoogerheide, Siem Jan Koopman

Abstract We present a novel approach to Bayesian estimation of two financial risk measures, Value at Risk and Expected Shortfall, in nonlinear, non-Gaussian state space models. In particular, we consider two specifications of the stochastic volatility model: with normal and Student's t observation disturbances. The key insight behind our proposed importance sampling based approach is to accurately approximate the optimal importance density, which focuses on the augmented parameter subspace corresponding to high losses. By oversampling the extreme scenarios and punishing them by lower importance weights, we achieve a much higher precision in characterising the properties of the left tail. We report substantial gains in the accuracy of estimates in an empirical study on daily financial data.
Keywords: Bayesian inference; Value at Risk; Expected Shortfall; efficient importance sampling; mixture of Student's t distributions; nonlinear non-Gaussian state space models.

1 Introduction

The experience from the recent global financial crisis shows that precise market risk estimation is crucial for a large number of agents in the worldwide economy. This type of risk is related to changes in the investment value in response to the moves of the market risk factors, e.g. stock prices. There are two standard measures of market risk: *Value at Risk* (VaR), which is a specified quantile of the percentage return distribution, and *Expected Shortfall* (ES), defined as the expected loss given it ex-

Agnieszka Borowska
FEWEB, VU Amsterdam, and Tinbergen Institute, the Netherlands, e-mail: a.borowska@vu.nl

Lennart F. Hoogerheide
FEWEB, VU Amsterdam, and Tinbergen Institute, the Netherlands, e-mail: l.f.hoogerheide@vu.nl

Siem Jan Koopman
FEWEB, VU Amsterdam, and Tinbergen Institute, the Netherlands, e-mail: s.j.koopman@vu.nl

ceeds the VaR. However, it is a well documented fact that the *volatility* of returns is subject to a dynamic process. Two distinct strands in the literature have emerged to model the unobserved volatility: observation driven models (ODM), which are non-probabilistic models for volatility and are typically formulated as generalized autoregressive conditional heteroskedasticity (GARCH) models; and parameter driven models (PDM), which are probabilistic models for volatility, with stochastic volatility (SV) model being a common example. The latter group has sound theoretical foundations and it allows for the observed prices to be driven by two independent processes (for the level and for the volatility).

The price for this greater flexibility of SV-type models is a more difficult analysis, which is usually performed using the methods for nonlinear, non-Gaussian state space models (SSM). In particular, Bayesian inference is computationally challenging, as the unobserved volatility sequence needs to be treated as a parameter to be estimated. Standard approaches to sampling of the latent volatility rely on Markov chain Monte Carlo (MCMC) methods, cf. [5], [7]. In the context of risk estimation these methods are clearly inefficient, as they result in a vast majority of the draws (consisting of model parameters and latent states) corresponding to the non-extreme scenarios and only very few leading to extreme losses of interest. Intuitively, computationally efficient Bayesian estimation of the tail-related quantities must focus on the extreme events, which can be achieved by resorting to importance sampling (IS). [3] developed an IS-based algorithm, *Quick Evaluation of Risk using Mixture of t approximations* (QERMit), which allows for efficient estimation of VaR and ES based on an appropriately constructed importance density. Their approach, however, is restricted to the class of ODM, as it requires a closed-form formula for the likelihood.

We present a novel approach to Bayesian risk evaluation in the context of nonlinear non-Gaussian SSM, which we call *Extended QERMit* (EQERMit). The key insight of [3] is that the optimal importance candidate for VaR estimation allocates half of the probability mass to the "high-loss" scenarios and half to the "regular" ones. We approximate this optimal candidate for PDM to report a substantial gain in the accuracy of evaluated VaR and ES. Further, the basic method is augmented by including the most recent realisations of the latent state variables in the "high-loss" scenarios. We observe an even closer approximation to the optimal target density, which should lead to an additional gain in the precision of our results.

2 Bayesian Risk Evaluation

Let $y = \{y_t\}_{t=1}^{T}$ be a sample of T historical logreturns, and let $\tilde{\theta}$ denote the (augmented) vector of all the model parameters (potentially including the latent state). We are interested in the h-day-ahead forecast of the $100\alpha\%$ VaR or ES, which are determined by the profit-loss function PL, mapping the h-vector of the future logreturns $y^* = \{y_{T+1}, \ldots, y_{T+h}\}$ into a scalar. A straightforward approach to Bayesian estimation of both risk measures, called the *direct approach*, is discussed below.

Direct Approach

1. Simulate $\tilde{\theta}^{(i)}$, $i = 1,\dots,N$, from the posterior $p(\tilde{\theta}|y)$.
2. Generate $y^{*(i)} \sim p(y^*|\tilde{\theta}^{(i)}, y)$, $i = 1,\dots,N$, the corresponding paths of future logreturns, given $\tilde{\theta}^{(i)}$ and y.
3. Compute $PL(y^{*(i)})$, the corresponding profit-loss values.
4. Sort $PL(y^{*(i)})$ ascending to obtain the permutation $PL^{(j)} := PL(y^{*(j)})$, $j = 1,\dots,N$.
5. Compute the $100\alpha\%$ VaR and ES estimates as $\widehat{VaR}_{DA} = PL^{((1-\alpha)N)}$ and $\widehat{ES}_{DA} = \frac{1}{(1-\alpha)N}\sum_{j=1}^{(1-\alpha)N} PL^{(j)}$, respectively.

Clearly, such an approach is inefficient: to gain a satisfactory insight into the "high loss region" (HLR), with the $100(1-\alpha\%)$ lowest returns, one needs to "waste" many draws by sampling from the "whole" posterior posterior and predictive distributions (as essentially we are interested only in the tail). An alternative solution, the *IS approach*, focuses on the relevant part of the posterior. By oversampling from the tail and punishing the "excessive" draws by lower importance weights one can achieve a much higher precision in characterising the properties of the HLR.

IS Approach

1. Simulate $\tilde{\theta}^{(i)}$, $i = 1,\dots,N$, from the candidate $q(\tilde{\theta}|y)$.
2. Compute $w^{(i)} = \frac{p(\tilde{\theta}^{(i)}, y)}{q(\tilde{\theta}^{(i)}|y)}$, $i = 1,\dots,N$, the importance weights of draws $\tilde{\theta}^{(i)}$.
3. Generate $y^{*(i)} \sim p(y^*|\tilde{\theta}^{(i)}, y)$, $i = 1,\dots,N$, the corresponding paths of future logreturns, given $\tilde{\theta}^{(i)}$ and y.
4. Compute $PL(y^{*(i)})$, the corresponding profit-loss values.
5. Sort $PL(y^{*(i)})$ ascending to obtain the permutation $PL^{(j)} := PL(y^{*(j)})$, $j = 1,\dots,N$.
6. Set $\widehat{VaR}_{IS}$ as $PL(y^{*(k)})$ for which $\sum_{j=1}^{k} w(\tilde{\theta}^{(j)}) \leq 1-\alpha$ and $\sum_{j=1}^{k+1} w(\tilde{\theta}^{(j)}) > 1-\alpha$. Given $\widehat{VaR}_{IS}$, $\widehat{ES}_{IS} = \sum_{j=1}^{k} w(\tilde{\theta}^{(j)}) PL^{(j)} / \sum_{j=1}^{k} w(\tilde{\theta}^{(j)})$.

Optimal Candidate Density The choice of the importance density is crucial for the performance of the IS estimation. The optimal candidate distribution (OCD) ought to minimise, given the specified number of draws, the numerical standard error (NSE) of the IS estimator. [2] discusses the condition which the OCD q_{opt} satisfies, which for the case of the probability estimation of a set S results in

$$\int_{x \in S} q_{opt}(x)dx = \int_{x \notin S} q_{opt}(x)dx = \frac{1}{2}. \tag{1}$$

Condition (1) implies that half of the total mass of the OCD shall be put in the region of interest S, while the remaining half – outside that region[1]. Hence, $q_{opt} = 0.5q_1 + 0.5q_2$, where q_1 and q_2 are candidates for the kernel on S^C and S,

[1] Such a split is the consequence of the fact that we only have the kernel of the target posterior and predictive distributions and not the exact normalised densities, which makes it necessary to adequately normalise the weights via sampling from the whole domain.

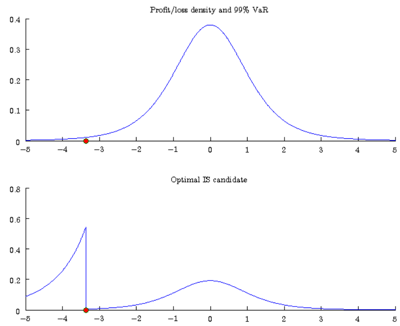

Fig. 1 Construction of the optimal 50%-50% importance density. Exemplary profit/loss function (Student's t with 5 degrees of freedom) and the implied 99% VaR (top). The optimal importance candidate density for the VaR evaluation (bottom).

respectively. [3] apply the above result in the context of VaR estimation by interpreting S as the HLR. Since the measure of S is then small, [3] take simply q_1 as the candidate on the whole space. Figure 1 illustrates the construction of the OCD for VaR estimation. In the QERMit algorithm, to determine the border of the HLR, first the *preliminary* VaR estimate is found with the direct approach by sampling from q_1; second, the HLR is approximated by q_2; third, the OCD q_{opt} is used to perform IS estimation. In QERMit q_{opt} is a functions of model parameters *and* future *observation* disturbances.

Approximation to the OCD in SSM The idea of adopting the OCD is conceptually appealing; it is also feasible in the class of ODM, originally discussed in [3], as these models allow for an explicit characterisation of the HLR. For PDM, however, such a specification is unavailable, as the whole volatility process is latent. Hence, our aim is to approximate the OCD as accurately as possible, by capturing these augmented parameter vectors which result in extreme losses.

Let $y = \{y_t\}_{t=1}^T$ be an observed time series driven by a latent state $x = \{x_t\}_{t=1}^T$. We consider the following nonlinear non-Gaussian SSM parametrised by θ

$$
\begin{aligned}
y_t &= f_t(x_t, \theta, \varepsilon_t), \qquad \varepsilon_t \overset{i.i.d.}{\sim} p(\theta), \\
x_{t+1} &= d_t + T_t x_t + \eta_t, \quad \eta_t \overset{i.i.d.}{\sim} \mathcal{N}(0, Q_t).
\end{aligned} \tag{2}
$$

We assume that the state transition is linear and Gaussian, which is not too restrictive in most of applications. The state vectors and matrices d_t, T_t and Q_t depend deterministically on t and θ; f_t is a deterministic measurement function of t, θ, x_t and of the realisation of observation disturbance ε_t. Notice that the OCD for model (2) is a joint candidate for θ and x (and future observation *and signal* disturbances), which makes it high-dimensional. In approximating the "whole" component q_1 we use the decomposition $q_1(x, \theta|y) = q_1(x|\theta, y) q_1(\theta|y)$ and follow [1] by targeting the marginal posterior of θ by a mixture of Student's t distributions (cf. [4]), and the conditional state density – by a Gaussian density $q(x|\theta, y)$ obtained with a numerically efficient algorithm of [6]. We use this approximation to obtain the preliminary VaR estimate, which we use to locate the HLR. Then, in our basic EQERMit algorithm, we approximate the target distribution in the HLR with $q_2(x,\theta,\eta^*,\varepsilon^*|y) = q_2(x|\theta,\eta^*,\varepsilon^*,y)q_2(\theta,\eta^*,\varepsilon^*|y)$, where $\eta^* = \{\eta_t\}_{t=T+1}^{T+h}$, $\varepsilon^* = \{\varepsilon_t\}_{t=T+1}^{T+h}$, where both components of q_2 are targeted with the same algorithms as for q_1. Then, we evaluate both risk measures with the IS approach, with the formula for the importance weight given by $w(\theta,x,\eta^*,\varepsilon^*) = q(y|\theta)\frac{p(y|\theta,x)}{q(y|\theta,x)}\frac{p(\theta)p(\eta^*)p(\varepsilon^*)}{q_1(\theta|y)p(\eta^*)p(\varepsilon^*)+q_2(\theta,\eta^*,\varepsilon^*|y)}$.

Including both the observation *and* the state disturbances in q_2 allows us for a more precise characterisation of the HLR for (2). Further efficiency gains can be obtained if we augment this approximation by the most recent realisations $x_T, x_{T-1},\dots,x_{T-r}$ of the latent state, as the volatility processes are typically highly persistent (in the simplest version $r=0$). Then, the weight formula must be adjusted by the transition probability from the old, "regular" states, to the most recent, "high-loss" states. We refer to this approach as *Augmented EQERMit*.

3 SV Model Example

To illustrate the performance of our EQERMit approach, we carry out an empirical study, considering SV model with normal and Student's t (SVt) observation errors.

Study Design We apply both specifications to the IBM stock daily logreturns from January 3, 2007 to December 30, 2011 (1259 observations). The model is

$$\begin{aligned} y_t &= \exp(x_t/2)\,\varepsilon_t, & \varepsilon_t &\overset{i.i.d.}{\sim} \mathcal{N}(0,1), \\ x_{t+1} &= c+\phi(x_t-c)+\sigma_\eta\eta_t, & \eta_t &\overset{i.i.d.}{\sim} \mathcal{N}(0,1). \end{aligned} \tag{3}$$

where y_t is an observed time series of logreturns. For the model parameters c, ϕ and σ^2_η, we adopt the following (proper, non-informative) prior specifications: $c \sim \mathcal{N}(0,1)$, $\frac{\phi+1}{2} \sim \text{Beta}(20,1.5)$ and $\frac{1}{\sigma^2_\eta} \sim \text{Gamma}\left(\frac{5}{2},\frac{0.05}{2}\right)$, which are standard in the literature, cf. [7]. In the case of Student's t errors, we take $\varepsilon_t \overset{i.i.d.}{\sim} t(\nu)$, where a priori $\nu-2 \sim \text{Exp}(1)$, and put $y_t = \exp(x_t/2)\left(\frac{\nu-2}{\nu}\right)^{1/2}\varepsilon_t$.

Results Table 1 shows the results for the 1-day-ahead 99% VaR and ES for the SV and SVt models based on $N = 10{,}000$ draws and 100 replications to obtain NSEs. EQERMit achieves usually at least twice higher precision than the direct approach, based on a single Student's t component. This accuracy gain is due to explicitly focusing on the HLR, i.e. approximating the OCD with two components, q_1 and q_2. Using just an approximation to the posterior, q_1, even when optimised as in [1] (Preliminary), does not improve much upon adopting a direct candidate.

Figure 2 illustrates the grounds for the superiority of our IS based methods. With the direct approach only 1% of all the draws correspond to the tail. Due to the focus on the HLR, our basic EQERMit method generates much more high-loss scenarios (around 20%). The Augmented EQERMit gets even closer to the OCD: almost 50% of the draws come from the HLR, as required by condition (1).

Table 1: Results for the 1-day-ahead 99% VaR and ES, in the SV and SVt models, based on $N = 10,000$ candidate draws and 100 replications to obtain NSEs.

	SV					SVt				
Method	AR[a]	VaR	(NSE)	ES	(NSE)	AR[a]	VaR	(NSE)	ES	(NSE)
Direct[b]	0.3963	-3.6810	(0.0902)	-4.4534	(0.1284)	0.0900	-4.0663	(0.1194)	-5.1718	(0.1683)
Preliminary[c]	0.5033	-3.6804	(0.0809)	-4.4663	(0.1159)	0.6668	-4.0746	(0.0847)	-5.2324	(0.1464)
EQERMit	–	-3.6555	(0.0487)	-4.4368	(0.0578)	–	-4.0679	(0.0457)	-5.2010	(0.0787)

[a] Acceptance rate for the independence Metropolis-Hastings algorithm, drawing from $q_1(\theta, x|y)$.
[b] Based on a single Student's t component, with mode and scale based on the simulated maximum likelihood estimates, with 5 degrees of freedom.
[c] Based on a mixture of Student's t distributions, with the mixture parameters optimised as in [1].

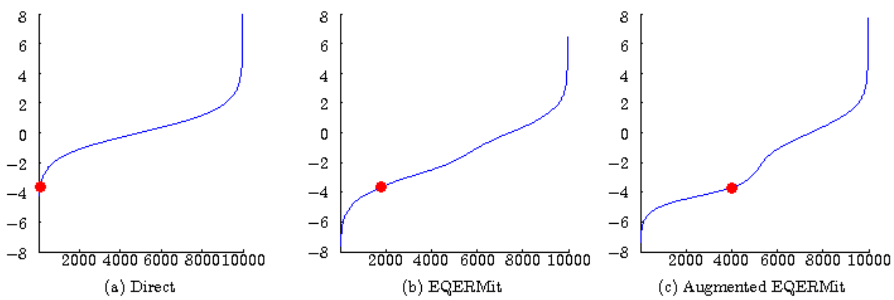

Fig. 2: Sorted future profit/losses $PL(y_{T+1}^{(i)})$, based on $N = 10,000$ candidate draws, and the $99\%\widehat{VaR}$ for the SV model. If the true (infeasible) OCD were used, $99\%\widehat{VaR}$ would be in middle of the sample.

References

1. Barra, I., Hoogerheide, L.F., Koopman, S.J., Lucas, A.: Joint Bayesian Analysis of Parameters and States in Nonlinear, Non-Gaussian State Space Models. TI Discussion Paper, Amsterdam, Tinbergen Institute **14-118/III** (2014)
2. Geweke, J.: Bayesian Inference in Econometric Models using Monte Carlo Integration. Econometrica **57**, 1317–1739 (1989)
3. Hoogerheide, L., van Dijk, H.K.: Bayesian Forecasting of Value at Risk and Expected Shortfall using Adaptive Importance Sampling. International Journal of Forecasting **26**, 231–247 (2010)
4. Hoogerheide, L.F., Opschoor, A., van Dijk, H.K.: A Class of Adaptive Importance Sampling Weighted EM Algorithms for Efficient and Robust Posterior and Predictive Simulation. Journal of Econometrics **171**, 101–120 (2012)
5. Kim, S., Shephard, N., Chib, S.: Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models. Review of Economic Studies **65**, 361–393 (1998)
6. Koopman, S.J., Lucas, A., Scharth, M.: Numerically Accelerated Importance Sampling for Nonlinear Non-Gaussian State Space Models. Journal of Business and Economic Statistics **33**, 114–127 (2015)
7. Omori, Y., Chib, S., Shephard, N., Nakajima, J.: Stochastic Volatility with Leverage: Fast and Efficient Likelihood Inference. Journal of Econometrics **140**, 425–449 (2007)