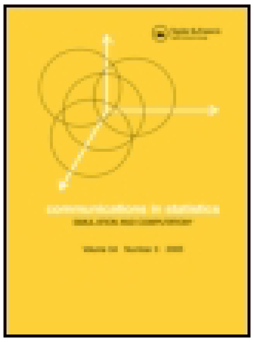

Communications in Statistics - Simulation and Computation

ISSN: 0361-0918 (Print) 1532-4141 (Online) Journal homepage: http://www.tandfonline.com/loi/lssp20

Riemann manifold Langevin methods on stochastic volatility estimation

Mauricio Zevallos, Loretta Gasco & Ricardo Ehlers

COMMUNICATIONS IN STATISTICS—SIMULATION AND COMPUTATION®
2017, VOL. 0, NO. 0, 1–15
https://doi.org/10.1080/03610918.2016.1255972

Taylor & Francis
Taylor & Francis Group

Check for updates

Riemann manifold Langevin methods on stochastic volatility estimation

Mauricio Zevallos[a], Loretta Gasco[b], and Ricardo Ehlers[c]

[a]Department of Statistics, University of Campinas (UNICAMP), Campinas - SP, Brazil; [b]Facultad de Ciencias e Ingeniería, Pontificia Universidad Católica del Perú, San Miguel Lima, Peru; [c]Department of Applied Mathematics and Statistics, University of São Paulo, São Paulo, Brazil

ABSTRACT
In this article, we perform Bayesian estimation of stochastic volatility models with heavy tail distributions using Metropolis adjusted Langevin (MALA) and Riemman manifold Langevin (MMALA) methods. We provide analytical expressions for the application of these methods, assess the performance of these methodologies in simulated data, and illustrate their use on two financial time series datasets.

ARTICLE HISTORY
Received 10 September 2016
Accepted 20 October 2016

KEYWORDS
Bayesian analysis; Langevin methods; Markov chain Monte Carlo; Metropolis–Hastings; Value at Risk

MATHEMATICS SUBJECT CLASSIFICATION
62F15; 62M10

1. Introduction

Stochastic volatility (SV) models were proposed by Taylor (1986). This model and its generalizations has been applied successfully to model the time-varying volatility present in financial time series. To estimate these models, several estimation methods have been proposed in the literature, for example, quasi-maximum likelihood methods (Harvey et al., 1994), generalized method of moments (Andersen and Sorensen, 1996), Markov chain Monte Carlo Methods (MCMC) pioneered by Jacquier et al. (1994), and Integrated Nested Laplace Approximations (Martino et al., 2010), to name a few. For an account of recent developments in the estimation of SV models, see Broto and Ruiz (2004) and Shephard and Andersen (2009) and the references therein. In particular, MCMC methods are considered one of the most efficient estimation methods. Proposals include, for example, Jacquier et al. (1994); Kim et al. (1998); Nakajima and Omori (2012).

Recently, Girolami and Calderhead (2011) proposed a methodology based on Metropolis adjusted Langevin and Hamiltonian Monte Carlo sampling methods. These methods take advantage of the relationship between Riemann geometry and statistics to overcome some of the shortcomings of existing Monte Carlo algorithms. They provide evidence that some sort of local calibration in the MCMC scheme may lead to strong improvements in large dimensional problems.

In particular, one of the examples discussed by these authors is the estimation of SV models with normal perturbations. Since these models often give rise to posterior distributions

CONTACT Ricardo Ehlers ehlers@icmc.usp.br Department of Applied Mathematics and Statistics, University of São Paulo, São Paulo 03178-200, Brazil.

with high correlations, the methods proposed can be particularly useful for estimation. More recently, Nugroho and Morimoto (2015) presented an algorithm based on Hamiltonian Monte Carlo methods for the estimation of realized stochastic volatility models.

In this article, we discuss the use of Langevin and modified Langevin methods to the estimation of SV models with normal, t-Student, and GED perturbations for the observations. One contribution of the article is to provide closed-form analytical expressions for the t-Student and GED cases and assess the statistical performance via simulation studies. We also illustrate with two real datasets.

Because the computational time is critical for estimating stochastic volatility models, we implemented a hybrid method in which a Riemann manifold MALA (MMALA) scheme is applied for the parameters and a MALA scheme is applied for the volatilities. From a methodological viewpoint, this is another contribution of the article as the hybrid method proves a useful tool for the analysis of long time series. Our preference for Langevin type sampling over Hamiltonian Monte Carlo (HMC) is mainly for computational reasons. In particular, the Riemann manifold HMC methods as proposed in Girolami and Calderhead (2011) are computationally more expensive to implement in the context of SV models. All the computations in this article were implemented using the open-source statistical software language and environment R (R Development Core Team, 2010).

The remainder of this article is organized as follows. The models are presented in Section 2 and the methodology for estimation is discussed in Section 3. To assess the estimation methodology, some Monte Carlo experiments are presented in Section 4. Section 5 illustrates with empirical data, and some final remarks are given in Section 6.

2. Models

We consider the following stochastic volatility (SV) model,

$$y_t = \beta \exp(h_t/2)\varepsilon_t, \tag{1}$$

$$h_t = \phi h_{t-1} + \eta_t, \tag{2}$$

where $\{\varepsilon_t\}$ is a sequence of independent identically distributed (IID) random variables with zero mean and unit variance, $\{\eta_t\}$ is an IID sequence of random variables such that $\eta_t \sim N(0, \sigma^2)$, and η_t and ε_t are independent for all t. In addition, we assume that $\beta > 0$ and $|\phi| < 1$.

In the SV model, conditional to the information set $\mathcal{F}_t = \{y_t, y_{t-1}, \ldots\}$, the standard deviation of y_t is given by

$$\sigma_t = \beta \exp(h_t/2).$$

In Finance, if y_t represents the tth return then σ_t is the *volatility* at time t.

The original formulation of the SV model by Taylor (1986) considers ε_t following a standard normal distribution. However, many empirical studies indicate that this model does not account for the kurtosis observed in most financial time series returns. Consequently, several other error distributions have been considered (see, e.g., Wang et al., 2013). For example, we consider ε_t following an exponential power distribution (or generalized error distribution, GED) with zero mean, unit variance (see Box and Tiao, 1973; Nelson, 1991) with density function,

$$f(\varepsilon_t) = \frac{\nu}{\lambda 2^{1+1/\nu}\Gamma(1/\nu)} \exp\left\{-\frac{1}{2}\left|\frac{\varepsilon_t}{\lambda}\right|^{\nu}\right\}, \tag{3}$$

where $\lambda^2 = 2^{-2/\nu}\Gamma(1/\nu)/\Gamma(3/\nu)$ and shape parameter $\nu > 0$. Important special cases are the Laplace (or double exponential) distribution for $\nu = 1$ and the standard normal distribution when $\nu = 2$. The kurtosis is given by $\Gamma(1/\nu)\Gamma(5/\nu)/\Gamma(3/\nu)^2 - 3$ so that when $\nu < 2$, this distribution reproduces heavy-tails. In addition, we consider ε_t following a t-Student distribution with ν degrees of freedom and density function,

$$f(\varepsilon_t) = \frac{1}{\sqrt{\pi(\nu-2)}}\frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})}\left\{1 + \frac{\varepsilon_t^2}{\nu-2}\right\}^{-(\nu+1)/2}. \tag{4}$$

When $\nu \to \infty$, this distribution approaches the standard normal distribution.

3. Estimation

Let $\boldsymbol{y} = (y_1, \ldots, y_n)$ be the observed time series and let $\boldsymbol{h} = (h_1, \ldots, h_n)$ be the latent variables. In order to estimate the SV models, we use the Metropolis adjusted Langevin (MALA) method (Roberts and Stramer, 2003) and the Riemannian manifold Metropolis adjusted Langevin (MMALA) Monte Carlo method proposed by Girolami and Calderhead (2011). For both methods, the estimation procedure, described by Girolami and Calderhead (2011), is performed in a two-step blocking approach. In the first step, the latent variables $\{h_t\}$ (the log-squared volatilities) are sampled and then, conditional on these sampled values, we sample the parameters $\boldsymbol{\theta} = (\beta, \sigma, \phi, \nu)$. At each step, a Metropolis–Hastings sampling scheme is applied using the methods described below.

3.1. Metropolis adjusted Langevin algorithm (MALA)

Let $\boldsymbol{\xi} \in \mathbb{R}^D$ be the random vector of interest with density $f(\boldsymbol{\xi})$. The Metropolis adjusted Langevin algorithm is based on a Langevin diffusion process whose stationary distribution is $f(\boldsymbol{\xi})$ and its stochastic differential equation is discretized to give the following proposal mechanism,

$$\boldsymbol{\xi} = \boldsymbol{\xi}^{[n]} + \frac{\epsilon^2}{2}\nabla_{\xi} \ln f(\boldsymbol{\xi}^{[n]}) + \epsilon\boldsymbol{z}, \tag{5}$$

where $\boldsymbol{z} \sim N(0, \boldsymbol{I})$, $\boldsymbol{I}$ is the identity matrix of order D, ϵ is the integration step size, and $\nabla_{\xi} \ln f(\boldsymbol{\xi}^{[n]})$ is the gradient of $\ln f(\cdot)$ with respect to ξ evaluated at $\boldsymbol{\xi}^{[n]}$. A Metropolis acceptance probability is then employed to ensure convergence to the invariant distribution as follows. A new value $\boldsymbol{\xi}$ is sampled from a multivariate normal distribution with mean $\mu(\boldsymbol{\xi}^{[n]}, \epsilon) = \boldsymbol{\xi}^{[n]} + \frac{\epsilon^2}{2}\nabla_{\xi} \ln f(\boldsymbol{\xi}^{[n]})$ and variance-covariance matrix $\epsilon^2\boldsymbol{I}$. This value is accepted with probability given by $\min\{1, f(\boldsymbol{\xi})q(\boldsymbol{\xi}^{[n]}|\boldsymbol{\xi})/f(\boldsymbol{\xi}^{[n]})q(\boldsymbol{\xi}|\boldsymbol{\xi}^{[n]})\}$ where the proposal density is $q(\boldsymbol{\xi}|\boldsymbol{\xi}^{[n]}) = N(\mu(\boldsymbol{\xi}^{[n]}, \epsilon), \epsilon^2\boldsymbol{I})$.

This algorithm is then employed to estimate the SV model following the two steps below.

1. *Sample the latent variables* $\boldsymbol{h}$. Assuming the parameters as constants, apply (5) with $f = f(\boldsymbol{y}, \boldsymbol{h})$ and gradient ∇ calculated with respect to $\boldsymbol{h}$.
2. *Sample parameters* $\boldsymbol{\theta}$. Given $(\boldsymbol{y}, \boldsymbol{h})$, apply (5) with $f = f(\boldsymbol{y}, \boldsymbol{h}|\boldsymbol{\theta})f(\boldsymbol{\theta})$ and gradient ∇ calculated with respect to $\boldsymbol{\theta}$.

3.2. Riemann manifold MALA (MMALA)

Girolami and Calderhead (2011) developed a modification in the Metropolis proposal mechanism in which the moves in $\mathbb{R}^D$ are according to a Riemann metric instead of the standard

Euclidian distance. This procedure is referred to as Riemann manifold MALA or MMALA. The proposal mechanism is now given by

$$\xi_i = \mu(\boldsymbol{\xi}^{[n]}, \epsilon)_i + \left\{\epsilon\sqrt{\boldsymbol{G}^{-1}}(\boldsymbol{\xi}^{[n]})\boldsymbol{z}\right\}_i, \tag{6}$$

$$\begin{aligned}\mu(\boldsymbol{\xi}^{[n]}, \epsilon)_i = \xi_i^{[n]} &+ \frac{\epsilon^2}{2}\left\{\boldsymbol{G}^{-1}(\boldsymbol{\xi}^{[n]})\nabla_\xi \ln f(\boldsymbol{\xi}^{[n]})\right\}_i \\ &- \epsilon^2\sum_{j=1}^{D}\left\{\boldsymbol{G}^{-1}(\boldsymbol{\xi}^{[n]})\frac{d\boldsymbol{G}(\boldsymbol{\xi}^{[n]})}{d\xi_j}\boldsymbol{G}^{-1}(\boldsymbol{\xi}^{[n]})\right\}_{ij} \\ &+ \frac{\epsilon^2}{2}\sum_{j=1}^{D}\left\{\boldsymbol{G}^{-1}(\boldsymbol{\xi}^{[n]})\right\}_{ij} tr\left\{\boldsymbol{G}^{-1}(\boldsymbol{\xi}^{[n]})\frac{d\boldsymbol{G}(\boldsymbol{\xi}^{[n]})}{d\xi_j}\right\},\end{aligned} \tag{7}$$

where $\boldsymbol{z} \sim N(0, \boldsymbol{I})$ and

$$\boldsymbol{G}(\boldsymbol{\xi}) = -E\left(\frac{d^2 \ln f(\boldsymbol{\xi})}{d\boldsymbol{\xi}^\top \boldsymbol{\xi}}\right).$$

Then, employing a Metropolis mechanism with proposal density given by $q(\boldsymbol{\xi}|\boldsymbol{\xi}^{[n]}) = N(\mu(\boldsymbol{\xi}^{[n]}, \epsilon), \epsilon^2\boldsymbol{G}^{-1}(\boldsymbol{\xi}^{[n]}))$ and the usual acceptance probability given by the quantity $\min\{1, f(\boldsymbol{\xi})q(\boldsymbol{\xi}^{[n]}|\boldsymbol{\xi})/f(\boldsymbol{\xi}^{[n]})q(\boldsymbol{\xi}|\boldsymbol{\xi}^{[n]})\}$ ensures convergence to the invariant distribution. We note that in this case both the mean vector and covariance matrix in the proposal distribution depend on the current state of the Markov chain.

A simplified proposal mechanism is obtained when a constant curvature is assumed. In this case, the last two terms in (7) vanish and the proposal mean becomes

$$\boldsymbol{\mu}(\boldsymbol{\xi}^{[n]}, \epsilon) = \boldsymbol{\xi}^{[n]} + \frac{\epsilon^2}{2}\boldsymbol{G}^{-1}(\boldsymbol{\xi}^{[n]})\nabla_\xi \ln f(\boldsymbol{\xi}^{[n]}).$$

In this simplified version of MMALA, the state-dependent covariance matrix in the proposal mechanism still allows adaptation to the local curvature of the target $f(\boldsymbol{\xi})$, which has been shown to increase algorithm efficiency in a number of applications (Girolami and Calderhead, 2011; Xifara et al., 2014). This is the approach adopted here. We show in the simulation study that, in particular for stochastic volatility models, we have an algorithm for estimation and prediction with a lower computational cost, which is an important feature in practice.

In our SV model, this algorithm is then applied following the two steps below.

1. *Sample the latent variables* $\boldsymbol{h}$. Assuming the parameters as constants, apply (5) with $f = f(\boldsymbol{y}, \boldsymbol{h})$ and gradient ∇ calculated with respect to $\boldsymbol{h}$.
2. *Sample parameters* $\boldsymbol{\theta}$. Given $(\boldsymbol{y}, \boldsymbol{h})$, apply (6) and (7) with $f = f(\boldsymbol{y}, \boldsymbol{h}|\boldsymbol{\theta})f(\boldsymbol{\theta})$, gradient ∇ and matrix $\boldsymbol{G}$ calculated with respect to $\boldsymbol{\theta}$.

In the Appendix, we provide details on the required expressions of partial derivatives and metric tensors for both MALA and MMALA. Also, it is worth mentioning that matrix inversion is less computationally demanding in the SV model since G has a sparse tridiagonal form.

3.3. A hybrid proposal

In real-life applications, the computational time for estimating the model is crucial. In SV models, the vector of volatilities has the same dimension as the sample size (usually thousands of observations), then we eventually have a very high-dimensional estimation problem. In this sense, it is worth to implement fast and reliable computational estimation procedures. In this article, we combine the MALA and MMALA procedures. This method, named *Hybrid* thereafter, consists of using the MMALA scheme for estimating the parameters ($\boldsymbol{\theta}$) and the MALA scheme for estimating the volatilities ($\boldsymbol{h}$).

3.4. Likelihood and priors

The log-likelihood $L_{y|\theta} = \ln[f(\boldsymbol{y}, \boldsymbol{h}|\boldsymbol{\theta})]$ is given by

$$f(\boldsymbol{y}, \boldsymbol{h}|\beta, \phi, \sigma, \nu) = f(h_1|\phi, \sigma)\prod_{t=2}^{n} f(h_t|h_{t-1}, \phi, \sigma)\prod_{t=1}^{n} f(y_t|h_t, \beta, \nu),$$

where $h_1|(\phi, \sigma) \sim N(0, \sigma^2/(1-\phi^2))$ and $h_t|(h_{t-1}, \phi, \sigma) \sim N(\phi h_{t-1}, \sigma^2)$. In addition,

$$f(y_t|h_t, \beta, \nu) = \frac{\nu}{\beta\lambda 2^{1+1/\nu}\Gamma(1/\nu)} \exp\left\{-\frac{h_t}{2} - \frac{1}{2\lambda^\nu}\left|\frac{y_t}{\beta\exp(h_t/2)}\right|^\nu\right\}$$

for GED errors and

$$f(y_t|h_t, \beta, \nu) = \frac{1}{\beta\sqrt{\pi(\nu-2)}}\frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})}\left\{1 + \frac{y_t^2}{\beta^2(\nu-2)\exp(h_t)}\right\}^{-(\nu+1)/2} \exp(-h_t/2)$$

for t-Student errors.

Following the Bayesian paradigm, we need to complete the model specification with appropriate prior distributions for the parameters. Independent prior distributions were assigned for ϕ and σ as in Liu (2001) and Girolami and Calderhead (2011), that is, $\sigma^2 \sim$ Inv-χ^2(10,0.05) and $(\phi+1)/2 \sim \text{Beta}(20, 1.5)$. In addition, we propose an exponential distribution with mean one as the prior for β. The prior for the tail parameter ν depends on the distribution adopted for the error terms. For GED errors, we propose the prior for $\nu \sim$ Inv-χ^2(10,0.05) while for Student-t errors, following Watanabe and Asai (2001), we consider the truncated exponential density,

$$f(\nu) = \lambda \exp\{-\lambda(\nu-4)\}$$

for $\nu > 4$ and zero otherwise, as the prior for ν. Differently from Watanabe and Asai (2001), we specified $\lambda = 1/3$.

Now, denoting the joint prior density of $\boldsymbol{\theta}$ by $\pi(\boldsymbol{\theta})$, the log prior is then given by

$$L_\theta = \ln\pi(\boldsymbol{\theta}) = -\beta - \frac{1}{4\sigma^2} - 11\ln(\sigma) + 19\ln\left(\frac{1+\phi}{2}\right) + \frac{1}{2}\ln\left(\frac{1-\phi}{2}\right) + \ln f(\nu),$$

where $\ln f(\nu) = -\frac{4}{\nu} - 3\ln(\nu)$ for GED errors and $\ln f(\nu) = \ln(\lambda) - \lambda(\nu-4)$ for t-Student errors.

It is worth noting that, in order to employ the algorithms described in the previous sections, we need to implement a transformation of σ, ϕ, and ν to the real line. Here, we set $\sigma = \exp(\gamma)$ and $\phi = \tanh(\alpha)$ as in Girolami and Calderhead (2011), and we propose to set $\nu = \exp(p)$ and $\nu = \exp(p) + 4$ for GED and t-Student errors, respectively. Of course this introduces Jacobian factors into the acceptance ratios given by $\frac{d\sigma}{d\gamma} = \exp(\gamma) = \sigma$, $\frac{d\phi}{d\alpha} = 1 - \tanh^2(\alpha) = 1 - \phi^2$. For GED errors, $\frac{d\nu}{dp} = \exp(p) = \nu$ and for t-Student errors $\frac{d\nu}{dp} = \nu - 4$.

4. Simulations

To assess the methodology described in the previous section, we conducted a Monte Carlo study. We generated $m =$ 1,000 replications of 1,000 observations from the SV model (1)-(2)

Table 1. Monte Carlo experiments.

		β		ϕ		σ		ν	
Errors	Method	bias	smse	bias	smse	bias	smse	bias	smse
Gaussian	MALA	− 0.001	0.038	− 0.022	0.028	0.051	0.056		
	Hybrid	0.024	0.038	− 0.011	0.015	0.000	0.014		
GED	MALA	− 0.002	0.032	− 0.042	0.051	0.090	0.099	− 0.011	0.128
	Hybrid	0.002	0.029	− 0.027	0.032	0.050	0.054	0.048	0.115
Student's *t*	MALA	− 0.003	0.031	− 0.063	0.072	0.122	0.131	0.912	2.311
	Hybrid	− 0.010	0.030	− 0.101	0.107	0.180	0.185	0.287	1.428

Note: Bias and square root of the mean squared error of posterior means. Parameters: $\beta = 0.65$, $\phi = 0.98$, $\sigma = 0.15$, and $\nu = 1.6$ (for GED) and $\nu = 7$ (for Student's *t*).

with parameters $\beta = 0.65$, $\phi = 0.98$ and two values for σ, $\sigma \in \{0.05, 0.15\}$. These parameter values were used by Liu (2001) and Girolami and Calderhead (2011) among others. We also considered three distributions for the errors: Gaussian, GED with parameter $\nu = 1.6$, and Student's t with $\nu = 7$ degrees of freedom. We then evaluated two estimation schemes: (i) MALA scheme for both the parameters and the volatilities and (ii) MMALA scheme for the parameters and MALA scheme for the volatilities (hybrid method, see Section 3.3).

The true parameter values were used as initial values for the MCMC samplers and the prior distributions are as described in Section 3.4 For each time series, we drew 20,000 MCMC samples discarding the first 10,000 samples as a burn-in.

To evaluate the performance of the estimation methods, two criteria were considered: the bias and square root of the mean square error (smse), which are defined as

$$\text{bias} = \frac{1}{m}\sum_{i=1}^{m}\hat{\theta}^{(i)} - \theta, \tag{8}$$

$$\text{smse}^2 = \frac{1}{m}\sum_{i=1}^{m}(\hat{\theta}^{(i)} - \theta)^2, \tag{9}$$

where $\hat{\theta}^{(i)}$ is the estimate of parameter θ for the ith replication, $i = 1, \ldots, m$. In this article, we take the posterior means of θ as point estimates.

The estimation results are presented in Tables 1 and 2. In Table 1 (when $\sigma = 0.15$), we obtained good results in terms of bias and smse for all the parameters in the Gaussian and GED cases with superiority of the hybrid over the MALA method except for the bias of ν in the GED case. In contrast, for the Student's t case, we obtained better results using MALA instead of the hybrid method. The results with Student's t perturbations are worse than the

Table 2. Monte Carlo experiments.

		β		ϕ		σ		ν	
Errors	Method	bias	smse	bias	smse	bias	smse	bias	smse
Gaussian	MALA	− 0.007	0.019	− 0.194	0.211	0.152	0.153		
	Hybrid	− 0.007	0.023	− 0.067	0.071	0.085	0.086		
GED	MALA	− 0.012	0.022	− 0.196	0.210	0.199	0.205	0.059	0.142
	Hybrid	− 0.013	0.025	− 0.107	0.112	0.132	0.133	0.109	0.152
Student's *t*	MALA	− 0.014	0.027	− 0.193	0.205	0.231	0.238	2.163	3.145
	Hybrid	− 0.020	0.030	− 0.199	0.205	0.256	0.260	1.419	2.156

Note: Bias and square root of the mean squared error of posterior means. Parameters: $\beta = 0.65$, $\phi = 0.98$, $\sigma = 0.05$, and $\nu = 1.6$ (for GED) and $\nu = 7$ (for Student's *t*).

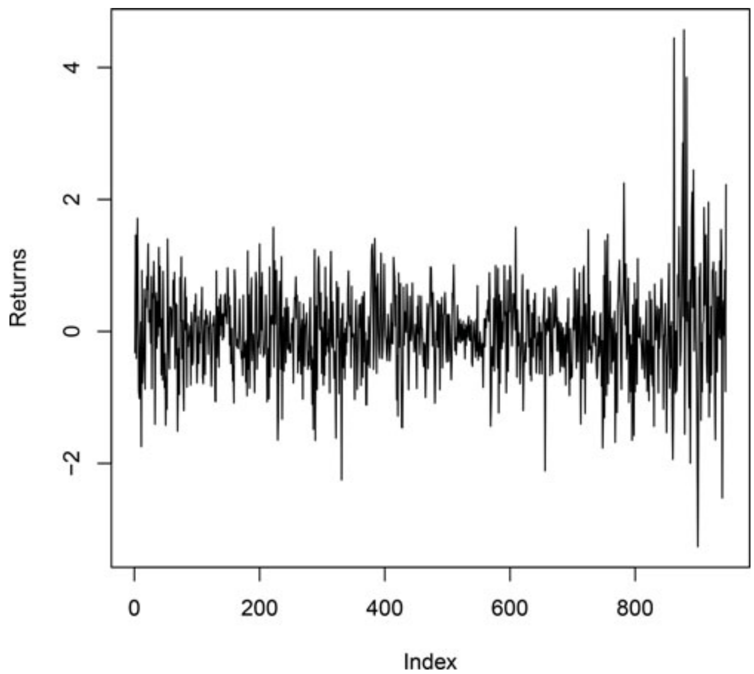

Figure 1. Pound/Dollar time series returns.

ones obtained with the other two distributions. In particular, note the large bias and smse values for σ.

When $\sigma = 0.15$ (see Table 2), overall we obtained worse results compared to Table 1.[1] In all cases, the parameter ϕ is underestimated and the parameter σ presents a large positive bias, as well as large smse for σ MALA estimates. For Gaussian and GED distributions, we observe an important bias and smse reduction for ϕ and σ estimates when using the hybrid method instead of MALA. For Student's t errors, the performance of MALA and the hybrid method are barely the same, but we obtain better results in terms of the bias of ν for hybrid estimates.

In summary, overall we obtained good results and the hybrid method tends to perform better than the MALA method.

5. Illustrations

In this section, we applied the described methodology to estimate two exchange rate time series data: the Pound/Dollar (£/USD) and the Canadian dollar /Dollar (CAN/USD). The time series under study are the daily continuously compounded returns in percentage, defined as $r_t = 100[\log(P_t) - \log(P_{t-1})]$ where P_t is the price at time t.

The £/USD time series returns cover the period from 1/10/81 to 28/6/85 and the SV model was estimated by Harvey et al. (1994) using quasi-maximum likelihood methods and by Durbin and Koopman (2001, pp. 236) using quasi-maximum likelihood and Monte Carlo importance sampling methods. In both cases the authors assumed Gaussian errors.

The CAN/USD returns are based on daily noon rates prices. The time series prices were obtained from the website http://www.bankofcanada.ca/rates/exchange/ and cover the period from January 2, 2007 to February 7, 2013.

We have 945 and 2,509 returns for the £/USD and CAN/USD time series, respectively. In Figs. 1 and 2, we show the time series returns and Table 3 consigns some descriptive statistics.

[1] These results were expected and are in line with the literature, see, for example, Jacquier et al. (1994).

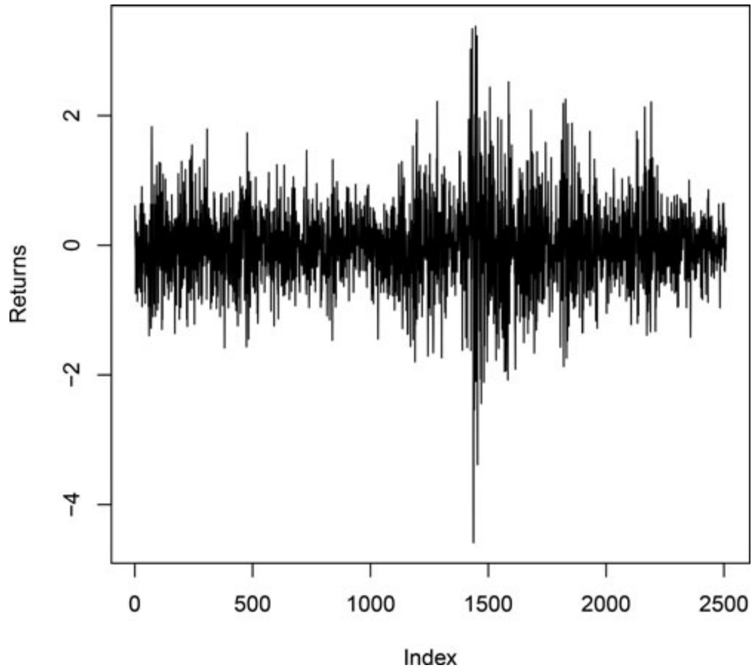

Figure 2. Canadian dollar/Dollar time series returns.

Table 3. Descriptive statistics.

Time series	n	Mean	Std. Dev.	Skewness	Kurtosis
£/USD	945	− 0.0353	0.7111	0.60	7.85
CAN/USD	2509	− 0.0168	0.6380	0.14	6.18

Note: n is the number of observations.

From this table, we observe a little skewness and high kurtosis, indicating asymmetric distributions with heavy tails. In addition, even not shown, the autocorrelation function indicates no serial correlation.

The analyses were performed on the demeaned returns. For each time series, we estimated SV models considering the following three different distributions for the errors ε_t in (1): the Gaussian, the GED distribution with parameter ν, and the Student's t distribution with ν degrees of freedom. Two estimation methods were applied: the MALA and the hybrid method.

For each time series, we drew 100,000 MCMC samples of parameters and volatilities. We discarded the first 50,000 as burn-in and skipped every 25th thus resulting in a final sample of 4,000 values from the posterior distribution.

The estimated posterior means and standard deviations for each parameter are shown in Table 4. [2] We can observe high persistence estimates (ϕ) and moderate values of the degrees of freedom ν in the Student's t distribution, thus indicating not too heavy tails. In addition, for the hybrid and Student's t errors case posterior standard deviations of β are large. For each time series and error distribution, when comparing point estimates under MALA and hybrid schemes, we note the following. First, for the £/USD, estimates almost do not change under Gaussian errors but change under GED and Student's t errors with a large change in ν for Student's t errors. Second, for the CAN/USD, estimates change slightly under Gaussian errors

[2] The maximum likelihood estimates in Harvey et al. (1994) are $\hat{\phi} = 0.9912$, $\hat{\sigma}^2 = 0.0069$, and $\hat{\gamma} = -0.0879$, then $\hat{\sigma} = 0.0831$ and $\hat{\beta} = \exp(-\hat{\gamma}/2) = 0.9570$. Durbin and Koopman (2001) report the following maximum likelihood estimates: $\hat{\phi} = 0.9731$, $\hat{\sigma} = 0.1726$, and $\hat{\beta} = 0.6338$ but do not report the Bayesian estimates.

Table 4. Estimation of stochastic volatility models.

Time series	Method	Errors	β	ϕ	σ	ν
£/USD	MALA	Gaussian	0.6156 (0.0115)	0.9824 (0.0042)	0.0903 (0.0016)	
		GED	0.3351 (0.0056)	0.9980 (0.0008)	0.0904 (0.0014)	2.0572 (0.1118)
		Student's *t*	0.6353 (0.0136)	0.9827 (0.0043)	0.0841 (0.0015)	10.5511 (1.8144)
	Hybrid	Gaussian	0.6311 (0.0146)	0.9847 (0.0052)	0.0752 (0.0017)	
		GED	0.6095 (0.0157)	0.9920 (0.0036)	0.0665 (0.0015)	1.6538 (0.1047)
		Student's *t*	0.9832 (0.2201)	0.9875 (0.0046)	0.0584 (0.0013)	4.7574 (0.4305)
CAN/USD	MALA	Gaussian	0.5524 (0.0066)	0.9873 (0.0022)	0.0812 (0.0009)	
		GED	0.5546 (0.0069)	0.9875 (0.0022)	0.0839 (0.0009)	1.7670 (0.0590)
		Student's *t*	0.5699 (0.0071)	0.9905 (0.0019)	0.0606 (0.0006)	12.6578 (2.0301)
	Hybrid	Gaussian	0.5579 (0.0079)	0.9921 (0.0023)	0.0628 (0.0009)	
		GED	0.5701 (0.0089)	0.9853 (0.0032)	0.0815 (0.0011)	1.7311 (0.0777)
		Student's *t*	0.8182 (0.1503)	0.9895 (0.0027)	0.0631 (0.0027)	5.1043 (0.9192)

Note: Posterior means and standard deviations (in parentheses).

but do not change under GED errors. For Student's t errors, we notice changes in β and ν. Third, the estimates of ν for the hybrid method are larger than the MALA estimates.

Figure 3 shows the sample autocorrelations, sample paths, and marginal posterior densities of parameters β, σ, ϕ, and ν for the CAN/USD series using the hybrid sampling scheme under GED errors. The autocorrelations vanish fairly rapidly and the sample paths show relatively good mixing in the parameter space.

In Figs. 4 and 5, the estimated volatilities $\exp(\hat{h}_t/2)$ taking the posterior medians of h_t as the point estimates $\hat{h}_t$ are depicted. As can be seen, the volatilities follow very well the observed volatility clustering of returns for both series.

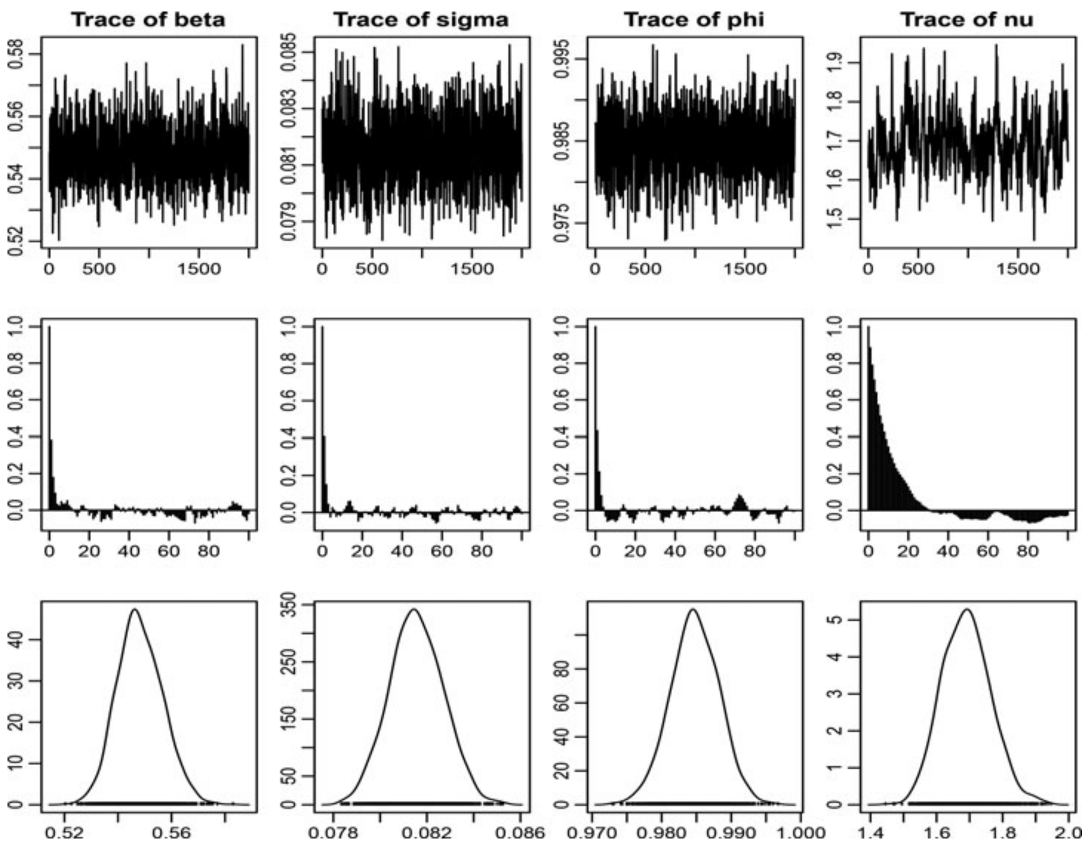

Figure 3. Sample autocorrelations, sample paths, and marginal posterior densities for the CAN/USD series using the Hybrid sampling scheme under GED errors.

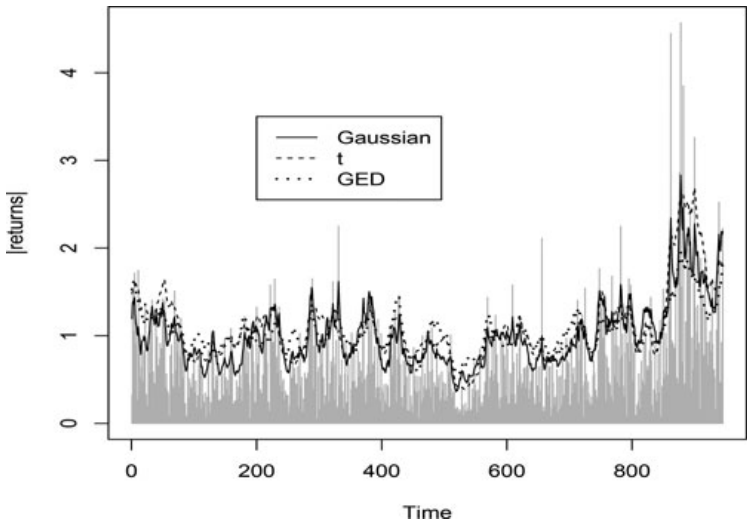

Figure 4. Absolute returns for the Pound/Dollar series and estimated volatilities using the hybrid sampling scheme under three different errors.

The performance of the proposed models and methods can also be assessed by estimating the Value at Risk (VaR) for multiple time horizons. From a Bayesian perspective, given the observed values of returns $\boldsymbol{y} = \{y_1, \ldots, y_n\}$ point estimates of the one-step ahead VaR could be obtained using a sample of values drawn from its predictive distribution, that is,

$$E(\mathrm{VaR}_{n+1|y}) \approx \frac{1}{J}\sum_{j=1}^{J} \mathrm{VaR}_{n+1}^{(j)}, \tag{10}$$

where $\mathrm{VaR}_{n+1}^{(j)}$ is the predicted one-step ahead VaR in the MCMC iteration. Because they are not available analytically, we adopt the following procedure. Given the parameter values and log-volatilities in the jth iteration, we obtain values of $\{h_{n+1}^{(j)}\}$ by drawing $\eta_{n+1}^{(j)} \sim N(0, \sigma^{2(j)})$ and setting $h_{n+1}^{(j)} = \phi^{(j)} h_n^{(j)} + \eta_{n+1}^{(j)}$. Next, we generate L replications $\{\epsilon_{n+1}^{(j,1)}, \ldots, \epsilon_{n+1}^{(j,L)}\}$ from the error distribution (with tail parameter $\nu^{(j)}$ for Student's t or GED distributions). Finally, we form a sample of returns by setting $y_{n+1}^{(j,k)} = \beta^{(j)} \exp(h_{n+1}^{(j,k)}/2)\epsilon_{n+1}^{(j,k)}$, which allow us to approximate $\mathrm{VaR}_{n+1}^{(j)}$ of confidence α by the value of the sample $(1-\alpha)$-quantile.

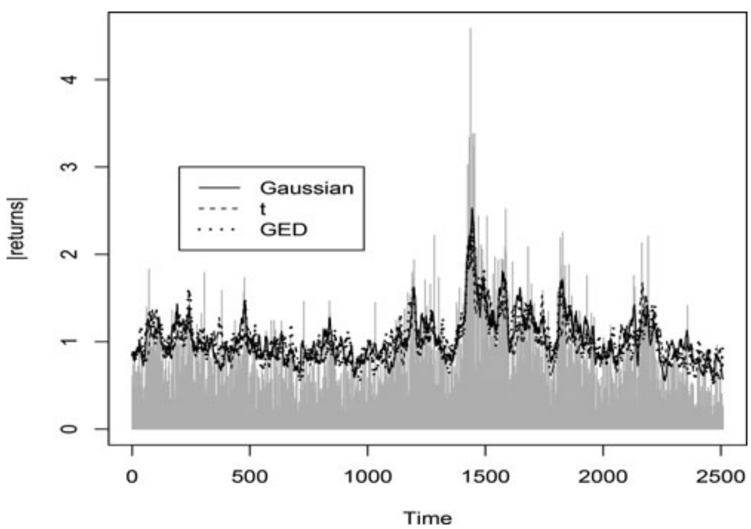

Figure 5. Absolute returns for the Canadian Dollar/Dollar series and estimated volatilities using the hybrid sampling scheme under three different errors.

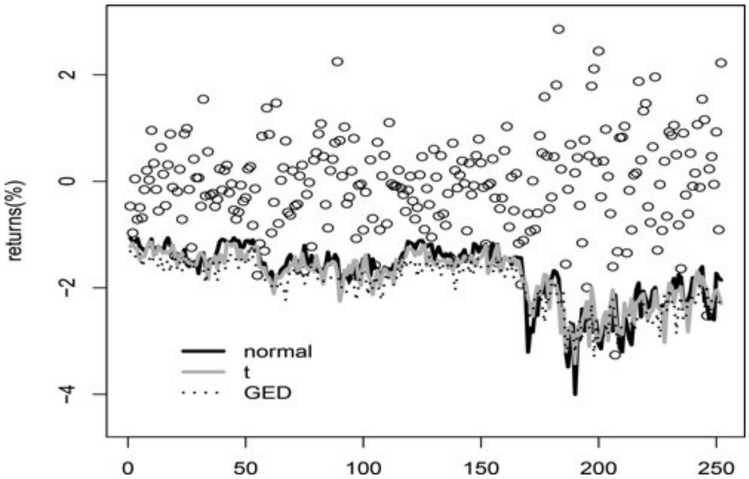

Figure 6. 99% Value at risk of Pound/Dollar exchange rates using the hybrid scheme.

For illustration, we estimated the one-day-ahead 99% VaR for the last 252 observations (which covers one stock market year approximately) of both the £/Dollar and the Canadian-Dollar/Dollar time series. Since we wanted to reproduce a real scenario, the model parameters were estimated and the VaR calculated based on observations $y_1, \ldots, y_{n-252+i}$, $i = 0, \ldots, 251$. Consequently, we estimated the model 252 times.

Figure 6 shows the last 252 returns and the VaR estimates using our hybrid MMALA algorithm for the £/Dollar series. We note that the VaR estimates follow very well the volatility in the market and react well to extreme down movements (large negative return values). In addition, we obtained 8, 7, and 5 observations outside the VaR limits for the Gaussian, Student's t, and GED distributions, respectively, more than the expected figure, 2.5 (1% of 252 observations).

As for the Canadian-Dollar/Dollar series, we note from Fig. 7 that, qualitatively the results for the Gaussian and GED errors are better and we obtained two observations outside the VaR limits in both cases, indicating a good performance. The VaR's for Student's t errors on the other hand are quite large (unnecessarily large from a financial viewpoint) including all the observations. This was indeed expected given the estimates of ν in Table 4. The estimate of β is also large compared to Gaussian and GED errors. In our empirical experience, it is usually better to work with GED distributions instead of Student's t distributions for the errors.

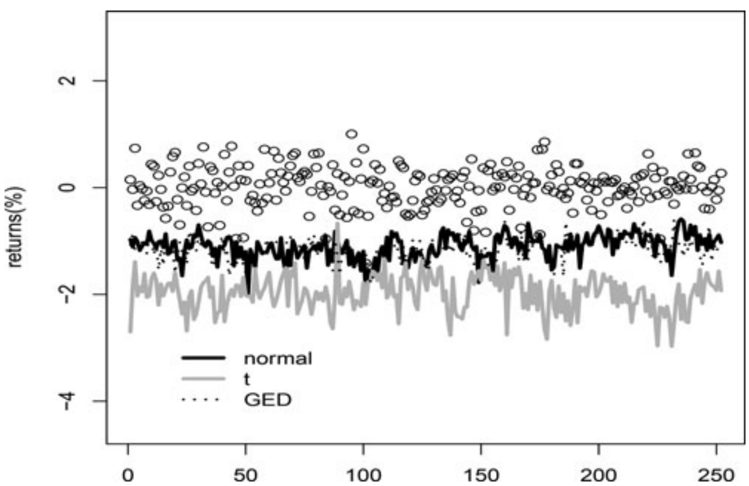

Figure 7. 99% Value at risk of Canadian-Dollar/Dollar exchange rates using the hybrid scheme.

6. Conclusions

In this article, we discuss the Bayesian estimation of the stochastic volatility model with Gaussian and two heavy-tailed distributions: GED and Student's t. Specifically, we implemented the Metropolis adjusted Langevin (MALA) and Riemann manifold MALA algorithms. Since the volatility has dimension equal to the sample size, the computational time could be high in real-life applications. Then we implemented a hybrid method: MMALA estimation for the parameters and MALA for sampling volatilities. These methods were assessed in both simulated data and real time series returns.

As in any Metropolis–Hastings like algorithm, our hybrid sampling scheme may be sensitive to the choice of the step size parameter ϵ. Tuning the sampler is simply unavoidable in practice and we recommend trying two different tuning parameters during the burn-in period and the stationary phase of the Markov chain (from which the final sample will be collected).

It is worth noting that (Riemann manifold) Hamiltonian Monte Carlo methods would provide alternative sampling schemes, however at a higher computational cost. Since we are allowing the metric tensor to vary with $\boldsymbol{\theta}$ along the iterations, the Hamiltonian function would not be separable thus requiring the use of third derivatives and a number of fixed point iterations. In consequence, we left this topic as a further research issue.

Appendix: Gradients and Matrix Tensors

In this appendix, we present the expressions of gradients and matrix tensors needed for the implementation of MALA and MMALA for GED and Student's t errors. For the Gaussian case, see Girolami and Calderhead (2011). In what follows, let $\varepsilon_t = \beta^{-1}\exp(-h_t/2)y_t$.

A.1. For GED errors

A.1.1. Sampling volatilities

The target function is proportional to

$$L_h = -\frac{(1-\phi^2)}{2\sigma^2}h_1^2 - \frac{1}{2\sigma^2}\sum_{t=2}^{n}(h_t-\phi h_{t-1})^2 - \frac{1}{2}\sum_{t=1}^{n}h_t - \frac{1}{2}\sum_{t=1}^{n}\left|\frac{\varepsilon_t}{\lambda}\right|^{\nu},$$

therefore the gradient $\nabla_h L_h = \frac{dL_h}{dh} = \boldsymbol{s} - \boldsymbol{r}$, where $\boldsymbol{s} = (s_1,\ldots,s_n)$ and $\boldsymbol{r} = (r_1,\ldots,r_n)$ assume values

$$\begin{aligned}
s_i &= -\frac{1}{2} + \frac{\nu}{4}\left|\frac{\varepsilon_i}{\lambda}\right|^{\nu}, \quad i = 1,\ldots,n\\
r_1 &= \frac{1}{\sigma^2}(h_1-\phi h_2), \quad r_n = \frac{1}{\sigma^2}(h_n-\phi h_{n-1}),\\
r_i &= \frac{1}{\sigma^2}\left[(h_i-\phi h_{i-1}) - \phi(h_{i+1}-\phi h_i)\right], \quad i = 2,\ldots,n-1.
\end{aligned}$$

On the other hand, the matrix tensor is a symmetric tridiagonal matrix with elements $\boldsymbol{G}_h(i,j) = -E(\frac{d^2L_h}{dh_i dh_j})$ for $i,j = 1,\ldots,n$,

$$\begin{aligned}
\boldsymbol{G}_h(i,i) &= \frac{\nu}{4} + \frac{1}{\sigma^2}, \quad i = 1, n\\
\boldsymbol{G}_h(i,i) &= \frac{\nu}{4} + \frac{1}{\sigma^2}(1+\phi^2), \quad i = 2,\ldots,n-1\\
\boldsymbol{G}_h(i,i+1) &= -\frac{\phi}{\sigma^2}, \quad i = 1,\ldots,n-1.
\end{aligned}$$

A.1.2. Sampling parameters

Here $L_{y|\theta} = \ln[f(\boldsymbol{y}, \boldsymbol{h}|\boldsymbol{\theta})]$, that is,

$$L_{y|\theta} = \frac{1}{2}\ln(1-\phi^2) - n\ln(\sigma) - n\ln(\beta) - \frac{(1-\phi^2)}{2\sigma^2}h_1^2 - \frac{1}{2\sigma^2}\sum_{t=2}^{n}(h_t - \phi h_{t-1})^2$$
$$- \frac{1}{2}\sum_{t=1}^{n}\left|\frac{\varepsilon_t}{\lambda}\right|^{\nu}.$$

The partial derivatives of this log-density with respect to the transformed parameters $(\delta, \gamma, \alpha, p)$ are

$$\frac{dL_{y|\theta}}{d\delta} = -n + \frac{\nu}{2}\sum_{t=1}^{n}\left|\frac{\varepsilon_t}{\lambda}\right|^{\nu},$$
$$\frac{dL_{y|\theta}}{d\gamma} = -n + \frac{1}{\sigma^2}(1-\phi^2)h_1^2 + \frac{1}{\sigma^2}\sum_{t=2}^{n}(h_t - \phi h_{t-1})^2,$$
$$\frac{dL_{y|\theta}}{d\alpha} = -\phi + \frac{\phi}{\sigma^2}(1-\phi^2)h_1^2 + \frac{(1-\phi^2)}{\sigma^2}\sum_{t=2}^{n}h_{t-1}(h_t - \phi h_{t-1})$$
$$\frac{dL_{y|\theta}}{dp} = \frac{n}{\nu}\left[\nu - \nu\left(\frac{\nu}{\lambda}\frac{d\lambda}{d\nu}\right) + \psi(1/\nu) + \ln(2)\right] - \frac{1}{2}\sum_{t=1}^{n}\left|\frac{\varepsilon_t}{\lambda}\right|^{\nu}\left\{\ln\left|\frac{\varepsilon_t}{\lambda}\right|^{\nu} - \nu\left(\frac{\nu}{\lambda}\frac{d\lambda}{d\nu}\right)\right\},$$

where

$$\nu\left(\frac{\nu}{\lambda}\frac{d\lambda}{d\nu}\right) = \ln(2) - \frac{1}{2}\psi(1/\nu) + \frac{3}{2}\psi(3/\nu).$$

In addition,

$$E\left(\frac{\partial^2 L_{y|\theta}}{\partial\delta^2}\right) = -n\nu, \quad E\left(\frac{\partial^2 L_{y|\theta}}{\partial\delta\partial\gamma}\right) = E\left(\frac{\partial^2 L_{y|\theta}}{\partial\delta\partial\alpha}\right) = 0$$
$$E\left(\frac{\partial^2 L_{y|\theta}}{\partial\delta\partial p}\right) = n\left\{1 + \psi(1+1/\nu) + \ln(2) - \nu\left(\frac{\nu}{\lambda}\frac{d\lambda}{d\nu}\right)\right\}$$
$$E\left(\frac{\partial^2 L_{y|\theta}}{\partial\gamma^2}\right) = -2n, \quad E\left(\frac{\partial^2 L_{y|\theta}}{\partial\gamma\partial\alpha}\right) = -2\phi, \quad E\left(\frac{\partial^2 L_{y|\theta}}{\partial\gamma\partial p}\right) = 0$$
$$E\left(\frac{\partial^2 L_{y|\theta}}{\partial\alpha^2}\right) = -2\phi^2 - (n-1)(1-\phi^2), \quad E\left(\frac{\partial^2 L_{y|\theta}}{\partial\alpha\partial p}\right) = 0$$
$$E\left(\frac{\partial^2 L_{y|\theta}}{\partial p^2}\right) = -n\nu\left(\frac{\nu}{\lambda}\frac{d\lambda}{d\nu}\right)^2 + \frac{n}{\nu}\left\{(1-1/\nu)\psi_1(1+1/\nu) + [\psi(1+1/\nu) + \ln(2)]^2\right\},$$

where ψ and ψ_1 are, respectively, the digamma and trigamma functions.

Now let $L_\theta = \ln\pi(\boldsymbol{\theta}) = \ln[f(\beta, \phi, \sigma, \nu)]$. Then

$$\frac{dL_\theta}{d\beta} = -1, \quad \frac{dL_\theta}{d\gamma} = \frac{1}{2\sigma^2} - 11, \quad \frac{dL_\theta}{d\alpha} = 19(1-\phi) - \frac{1}{2}(1+\phi), \quad \frac{dL_\theta}{dp} = \frac{4}{\nu} - 3$$

and the expectations of the second-order derivatives of L_θ are given by

$$E\left(\frac{\partial^2 L_\theta}{\partial\gamma^2}\right) = -\frac{1}{\sigma^2}, \quad E\left(\frac{\partial^2 L_\theta}{\partial\alpha^2}\right) = -\frac{39}{2}(1-\phi^2), \quad E\left(\frac{\partial^2 L_\theta}{\partial p^2}\right) = -\frac{4}{\nu}$$

and zero elsewhere. Finally, we use $\nabla_\theta \ln f = \frac{dL_{y|\theta}}{d\theta} + \frac{dL_\theta}{d\theta}$ and $\mathbf{G}_\theta = -E(\frac{\partial^2 L_{y|\theta}}{\partial\theta^2}) - E(\frac{\partial^2 L_\theta}{\partial\theta^2})$.

A.2. For t-Student errors

Next, we present those expressions that are different compared with the GED case.

A.2.1. Sampling volatilities

The target function is proportional to

$$L_h = -\frac{(1-\phi^2)}{2\sigma^2}h_1^2 - \frac{1}{2\sigma^2}\sum_{t=2}^{n}(h_t-\phi h_{t-1})^2 - \frac{1}{2}\sum_{t=1}^{n}h_t - \frac{(\nu+1)}{2}\sum_{t=1}^{n}\ln\left(1+\frac{\varepsilon_t^2}{\nu-2}\right).$$

$$s_i = -\frac{1}{2} + \frac{1}{2}\frac{(\nu+1)}{(\nu-2)}\frac{\varepsilon_i^2}{1+\varepsilon_i^2/(\nu-2)}, \quad i=1,\dots,n$$

$$\boldsymbol{G}_h(i,i) = \frac{\nu}{2(\nu+3)} + \frac{1}{\sigma^2}, \quad i=1,n$$

$$\boldsymbol{G}_h(i,i) = \frac{\nu}{2(\nu+3)} + \frac{1}{\sigma^2}(1+\phi^2), \quad i=2,\dots,n-1.$$

A.2.2. Sampling parameters

Here $L_{y|\theta} = \ln[f(\boldsymbol{y},\boldsymbol{h}|\boldsymbol{\theta})]$,

$$L_{y|\theta} = \frac{1}{2}\ln(1-\phi^2) - n\ln(\sigma) - n\ln(\beta) - \frac{(1-\phi^2)}{2\sigma^2}h_1^2 - \frac{1}{2\sigma^2}\sum_{t=2}^{n}(h_t-\phi h_{t-1})^2$$
$$-\frac{n}{2}\ln(\nu-2) + n\ln\Gamma\left(\frac{\nu}{2}+\frac{1}{2}\right) - n\ln\Gamma\left(\frac{\nu}{2}\right) - \frac{(\nu+1)}{2}\sum_{t=1}^{n}\ln\left(1+\frac{\varepsilon_t^2}{\nu-2}\right).$$

Let $p = \ln(\nu-4)$

$$\frac{dL_{y|\theta}}{d\beta} = -\frac{n}{\beta} + \frac{\nu+1}{\beta}\sum_{t=1}^{n}\frac{\varepsilon_t^2/(\nu-2)}{1+\varepsilon_t^2/(\nu-2)},$$

$$\frac{2}{(\nu-4)}\frac{dL_{y|\theta}}{dp} = n\left[\psi\left(\frac{\nu}{2}+\frac{1}{2}\right) - \psi\left(\frac{\nu}{2}\right) - (\nu-2)^{-1}\right] + \frac{(\nu+1)}{(\nu-2)}\sum_{t=1}^{n}\frac{\varepsilon_t^2/(\nu-2)}{1+\varepsilon_t^2/(\nu-2)}$$
$$-\sum_{t=1}^{n}\ln\left(1+\varepsilon_t^2/(\nu-2)\right)$$

$$E\left(\frac{\partial^2 L_{y|\theta}}{\partial\delta^2}\right) = -\frac{2n\nu}{\nu+3}$$

$$E\left(\frac{\partial^2 L_{y|\theta}}{\partial\delta\partial p}\right) = \frac{-6n(\nu-4)}{(\nu-2)(\nu+1)(\nu+3)}$$

$$E\left(\frac{\partial^2 L_{y|\theta}}{\partial p^2}\right) = \frac{n}{2}\frac{(\nu-4)^2}{(\nu-2)^2}\left\{\frac{(\nu-3)(\nu+4)}{(\nu+1)(\nu+3)} + \frac{(\nu-2)^2}{2}\left[\psi_1\left(\frac{\nu}{2}+\frac{1}{2}\right) - \psi_1\left(\frac{\nu}{2}\right)\right]\right\}.$$

Finally, $\frac{dL_\theta}{dp} = E\left(\frac{\partial^2 L_\theta}{\partial p^2}\right) = -\lambda(\nu-4)$.

Funding

This research was partially supported by FAPESP grant number 2013/00506-1 for the first author. Ricardo Ehlers received support from São Paulo Research Foundation (FAPESP) - Brazil, under grant number 2015/00627-9.

References

Andersen, T., Sorensen, B. (1996). GMM estimation of a stochastic volatility model: A Monte Carlo study. *Journal of Business and Economic Statistics* 13:329–352.

Box, G. E. P., Tiao, G. C. (1973). *Bayesian Inference in Statistical Analysis*. Reading, MA: Addison-Wesley Publishing.

Broto, C., Ruiz, E. (2004). Estimation methods for stochastic volatility models: A survey. *Journal of Economic Surveys* 18:613–649.

Durbin, J., Koopman, S. J. (2001). *Time Series Analysis by State Space Methods*. Oxford: Oxford University Press.

Girolami, M., Calderhead, B. (2011). Riemann manifold Langevin and Hamiltonian Monte Carlo methods. *Journal of the Royal Statistical Society, Series B* 73:123–214.

Harvey, A. C., Ruiz, E., Shephard, N. (1994). Multivariate stochastic variance models. *Reviews of Economic Studies* 61:247–264.

Jacquier, E., Polson, N. G., Rossi, P. E. (1994). Bayesian analysis of stochastic volatility models (with discussion). *Journal of Business and Economic Statistics* 12:371–418.

Kim, S., Shepard, N., Chib, S. (1998). Stochastic volatility: Likelihood inference comparison with ARCH models. *Review of Economic Studies* 65:361–393.

Liu, J. S. (2001). *Monte Carlo Strategies in Scientific Computing*. New York: Springer.

Martino, S., Aas, K., Lindqvist, O., Neef, L., Rue, H. (2010). Estimating stochastic volatility models using integrated nested Laplace approximations. *The EuropeanJournal of Finance* 17(7):487–503.

Nakajima, J., Omori, Y. (2012). Stochastic volatility model with leverage and asymmetrically heavy-tailed error using GH skew student distribution. *Computational Statistics and Data Analysis* 56(11):3690–3704.

Nelson, D. B. (1991). Conditional heteroscedasticity in asset returns: A new approach. *Econometrica* 59:347–370.

Nugroho, D. B., Morimoto, T. (2015). Estimation of realized stochastic volatility models using Hamiltonian Monte Carlo-Based methods. *Computational Statistics* 30(2):491–516.

R Development Core Team, (2010). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Roberts, G. O., Stramer, O. (2003). Langevin diffusions and Metropolis-Hastings algorithms. *Methodology and Computing in Applied Probability* 4:337–358.

Shephard, N., Andersen, T. G. (2009). Stochastic volatility: Origins and overview. In: Mikosch, T., Krei, J.-P., Davis, R. A., Gustav, T., eds. *Handbook of Financial Time Series*. New York: Springer, pp. 233–254.

Taylor, S. (1986). *Modelling Financial Time Series*. New York: Wiley.

Wang, J., Choy, S., Chan, J. (2013). Modelling stochastic volatility using generalised t distribution. *Journal of Statistical Computation and Simulation* 83(2):340–354.

Watanabe, T., Asai, M. (2001). *Stochastic Volatility Models with Heavy-Tailed Distributions: A Bayesian Analysis*. Discussion Paper 2001-E-17, Institute for Monetary and Economic Studies, Bank of Japan.

Xifara, T., Sherlock, C., Livingstone, S., Byrne, S., Girolami, M. (2014). Langevin diffusions and the Metropolis-adjusted Langevin algorithm. *Statistics and Probability Letters* 91:14–19.