DEM Working Papers

N. 2017/03

**Likelihood-based Risk Estimation for
Variance-Gamma Models**

Marco Bee, Maria Michela Dickson, Flavio Santi

UNIVERSITÀ DEGLI STUDI DI TRENTO
Dipartimento di Economia e Management

Università degli Studi di Trento

Department of Economics and Management, University of Trento, Italy.

Editors

Luciano ANDREOZZI luciano.andreozzi@unitn.it

Roberto GABRIELE roberto.gabriele@unitn.it

Technical officer

Marco TECILLA marco.tecilla@unitn.it

Guidelines for authors

Papers may be written in Italian or in English. Faculty members of the Department must submit to one of the editors in pdf format. Management papers should be submitted to R. Gabriele. Economics Papers should be submitted to L. Andreozzi. External members should indicate an internal faculty member that acts as a referee of the paper.

Typesetting rules:

1. papers must contain a first page with title, authors with emails and affiliations, abstract, keywords and codes. Page numbering starts from the first page;
2. a template is available upon request from the managing editors.

Likelihood-based Risk Estimation for Variance-Gamma Models

Marco Bee

Department of Economics and Management, University of Trento

Maria Michela Dickson

Department of Economics and Management, University of Trento

Flavio Santi

Department of Economics and Management, University of Trento

Abstract

Although the variance-gamma distribution is a flexible model for log-returns of financial assets, so far it has found rather limited applications in finance and risk management. One of the reasons is that maximum likelihood estimation of its parameters is not straightforward. We develop an EM-type algorithm that bypasses the evaluation of the full likelihood, which may be difficult because the density is not in closed form and is unbounded for small values of the shape parameter. Moreover, we study the relative efficiency of our approach with respect to the maximum likelihood estimation procedures implemented in the `VarianceGamma` and `ghyp` R packages. Extensive simulation experiments and real-data analyses suggest that the multicycle ECM algorithm and the routines in the `ghyp` R package give the best results in terms of root-mean-squared-error, for both parameter and Value-at-Risk estimation.

Keywords: Multicycle EM algorithm; maximum likelihood; numerical optimization; risk estimation.

1 Introduction

The variance-gamma (VG) distribution is the marginal distribution of the VG process, first introduced by Madan and Seneta (1990). Whereas geometric Brownian motion implies that log-returns are normally distributed, the log-returns computed from the VG process are, in general, asymmetric and leptokurtic, because the VG distribution has two additional parameters that account for skewness and kurtosis. According to the well-known stylized facts about financial returns (e.g. Cont, 2001), for modeling purposes this is a very desirable feature.

Given that the VG distribution is the marginal distribution of a stochastic process that generalizes geometric Brownian motion, the model has been used for option pricing (see, e.g., Madan and Milne, 1991; Madan et al, 1998; Carr and Madan, 1999; Seneta, 2004; Musiela and Rutkowski, 2005, p. 312); in particular, Madan et al (1998) obtain the prices of European options in closed form. However, as the risk-neutral parameters needed for option pricing are not explicit functions of the statistical parameters, the impact of the estimation error on option prices cannot be evaluated. The risk-neutral parameters are usually calibrated from option market prices, so that estimation issues are not so important for option pricing.

On the other hand, the capabilities of the marginal distribution to model skewness and fat tails make it well-suited also for risk management (McNeil et al, 2005, Section 3.2.5; Hu and Kercheval, 2007). In this case, measuring the effect of the sampling distribution of the parameters estimators on the estimated risk measures is quite important.

The VG distribution can also be obtained from a different perspective, noting that it is a special case of the family of multivariate Generalised Hyperbolic Distributions (GHD) developed by Barndorff-Nielsen (1978); see also Barndorff-Nielsen and Blæsild (1981). The GHD family is obtained via a mean-variance mixture construction, where the mixing distribution is a Generalized Inverse Gaussian (GIG) distribution, and has proved to be a very flexible multivariate model for financial data (Eberlein and Keller, 1995; McNeil et al, 2005, Section 3.2.3).

Maximum likelihood estimation (MLE) of the VG distribution is a challenging problem. The basic reason is twofold: first, the probability density function (pdf) cannot be written in closed form, because it includes the modified Bessel function of the second kind; second, it is unbounded for values of the shape parameter smaller than $d/2$, where d is the dimension of the

distribution. As a result, to compute the MLEs it is necessary to use numerical optimization algorithms and it is advisable to address the possible singularity of the likelihood.

Over the years, a couple of papers (Seneta, 2004; Cervellera and Tucci, 2016) have tried without success to replicate the MLEs obtained by Madan et al (1998, Section 5). Cervellera and Tucci (2016)’s investigation suggests that standard optimization algorithms, such as the Nelder-Mead simplex algorithm implemented in R’s `optim` (R Core Team, 2016) or MatlabTM’s `fminsearch` functions, are very sensitive to the starting values and produce highly unstable parameter estimates.

The literature dealing with MLE of the GHD generally takes a different approach. Instead of performing brute force maximization of the likelihood function, it tries to exploit the stochastic structure of the data-generating process, namely the mean-variance mixture construction. Noting that MLE would be easier if we knew the latent GIG mixing variables, MLE can be carried out by means of an EM-type algorithm (Dempster et al, 1977; McLachlan and Krishnan, 2008). An approach based on an extension of the EM algorithm known as MCECM (MultiCycle, Expectation, Conditional Maximization; Meng and Rubin, 1993; Meng, 1994; McLachlan and Krishnan, 2008, Section 5.3) has been developed by McNeil et al (2005, Section 3.2.4) for the general multivariate GHD and by Nitithumbundit and Chan (2015) for the multivariate VG distribution.

In all cases the implementation of the algorithms is not straightforward, particularly when the shape parameter is smaller than $1/2$ and the estimation procedures have to deal with the unboundedness of the pdf at the location parameter.

The contribution of this paper is threefold. First, we develop three EM-types algorithms, based on Nitithumbundit and Chan (2015)’s corresponding multivariate versions, for MLE of the univariate VG distribution, and compare them with the standard optimization routines implemented in the R `VarianceGamma` (Scott and Yang Dong, 2015) and `ghyp` (Luethi and Breymann, 2016) packages. We study the computational efficiency of the three EM-type algorithms as well as the impact of the input parameters. Second, extensive simulation experiments suggest that the MCECM algorithm performs similarly or better than the `fit.VGuv` R function of the `ghyp` R package, whereas the `vgFit` function of the `VarianceGamma` package is the worst alternative in almost all circumstances. Third, we use the distribution

as a model for financial returns and study the impact of the sampling variability of the parameters estimators obtained with MCECM, **VarianceGamma** and **ghyp** on Value-at-Risk (VaR) estimation.

The rest of the paper is organized as follows: Section 2 formally defines the VG distribution and the parametrization used; Section 3 gives a detailed derivation of the EM-type algorithms as well as of the approaches implemented in the **VarianceGamma** and **ghyp** R packages; Section 4 illustrates the outcomes of simulation experiments aimed at performing a comparison of the three EM-type algorithms, analyzing the relevance of the inputs of MCECM, and comparing MCECM to **VarianceGamma** and **ghyp**; Section 5 estimates the VaR of two financial time series by means of the VG and two other classical approaches. Finally, Section 6 concludes.

2 The Variance-Gamma Distribution

We first introduce the VG distribution as the marginal density of the VG process. Consider the following stochastic process, possibly describing the time evolution of the price of a risky asset P_t :

$$P_t = P_0 \exp\{\mu t + \gamma T_t + \sigma W(T_t)\},$$

where $\mu \in \mathbb{R}$, $\gamma \in \mathbb{R}$ and $\sigma^2 \in \mathbb{R}^+$. $\{T_t\}$ is a positive increasing stochastic process, usually called (market) activity time, whose differences $\tau_t = T_t - T_{t-1}$ ($t > 1$) are stationary, independent of the standard Brownian motion $\{W(t)\}$ and follow a gamma distribution with rate and shape parameters both equal to $\nu \in \mathbb{R}^+$. The corresponding log-returns are given by

$$X_t = \log P_t - \log P_{t-1} = \mu + \gamma(T_t - T_{t-1}) + \sigma(W(T_t) - W(T_{t-1})).$$

The common assumption $E(T_t) = 1$ does not imply any loss of generality and allows one to derive closed form expressions for the moments (Seneta, 2004, pp. 177-78) as well as the marginal pdf, given by:

$$f(x; \mu, \sigma^2, \gamma, \nu) = \frac{2 \exp\{\gamma(x - \mu)/\sigma^2\}}{\sigma \sqrt{2\pi} (1/\nu)^\nu \Gamma(\nu)} \left(\frac{|x - \mu|}{\sqrt{2\gamma^2 \nu + \sigma^2}} \right)^{\nu-1/2} K_{\nu-1/2} \left(\frac{|x - \mu| \sqrt{2\sigma^2 \nu + \gamma^2}}{\sigma^2} \right), \quad (1)$$

where $x \in \mathbb{R} \setminus \{\mu\}$, $K_\nu(\cdot)$ is the modified Bessel function of the second kind with index ν , μ is the location parameter, σ^2 is the scale parameter, γ is the skewness parameter and ν is the shape parameter. When $x = \mu$, two cases need to be distinguished: when $\nu \leq 1/2$, the density is singular at μ , i.e. $\lim_{x \rightarrow \mu} f(x; \mu, \sigma^2, \gamma, \nu) = +\infty$; when $\nu > 1/2$, the density reduces to

$$f(x; \mu, \sigma^2, \gamma, \nu) = \frac{\Gamma(\nu - 1/2)}{\sigma \sqrt{2\pi} (1/\nu)^\nu \Gamma(\nu)} \left(\frac{2\sigma^2}{\sqrt{2\sigma^2\nu + \gamma^2}} \right)^{\nu-1/2}. \quad (2)$$

The parametrization leading to (1) is used in McNeil et al (2005) and in the **ghyp** package; with respect to Seneta (2004) and the **VarianceGamma** package, the only difference is that ν in (1) replaces $1/\nu$ in Formula (15) of Seneta (2004).

Although γ is not equal to the skewness and ν is not equal to the kurtosis, the VG distribution is positively (negatively) skewed when γ is positive (negative), and the absolute value of the skewness is a function of the absolute value of γ ; similarly, the kurtosis is negatively related to ν , being always larger than 3 and converging to 3 when $\nu \rightarrow \infty$ (see Seneta, 2004, for the expressions of the skewness and of the kurtosis). When $\gamma = 0$ and $\nu \rightarrow +\infty$, the VG distribution converges to the normal $N(\mu, \sigma^2)$ distribution. Figure 1 plots the VG pdf for various values of ν along with the standard normal pdf. The singularity at μ and the convergence to the normal are clear from the graph.

As said above, the VG pdf (1) is also a special case of the GHD, which, in turn, can be seen as a mean-variance mixture. Let $\lambda \sim \text{GIG}(\nu, \chi, \psi)$ and

$$X = \mu + \gamma\lambda + \sqrt{\lambda}\sigma Z,$$

where $Z \sim N(0, 1)$, so that $X|\lambda \sim N(\mu + \gamma\lambda, \lambda\sigma^2)$. The marginal pdf of X , obtained by evaluating the integral $\int f_{X|\lambda}(x|\lambda) \cdot f_{\text{GIG}}(\lambda) d\lambda$, is the GHD pdf (McNeil et al, 2005, Formula (3.30)):

$$f(x; \mu, \sigma^2, \gamma, \nu, \chi, \psi) = \frac{(\sqrt{\chi\psi})^{-\nu} (\psi)^\nu (\psi + \gamma^2/\sigma^2)^{1/2-\nu}}{\sqrt{2\pi}\sigma K_\nu(\sqrt{\chi\psi})} \cdot \frac{K_{\nu-1/2} \sqrt{(\chi + (x - \mu)^2/\sigma^2)(\psi + \gamma^2/\sigma^2)} \exp\{(x - \mu)\gamma/\sigma^2\}}{(\sqrt{(\chi + (x - \mu)^2/\sigma^2)(\psi + \gamma^2/\sigma^2)})^{1/2-\nu}}. \quad (3)$$

The VG pdf is obtained when $\nu > 0$ and $\chi = 0$, because it can be shown that, under these conditions, (3) reduces to (1). For the general version (3) of the

Figure 1: The VG density for various values of ν , with $\mu = 0$, $\sigma^2 = 1$, $\gamma = -0.1$. For comparison, the standard normal is reported as well.

GHD we use the so-called “chi/psi” representation, whereas the VG pdf (1) is written in the “alpha.bar” parametrization, which is more convenient in terms of interpretation of the parameters (Prause, 1999, Appendix C). Using the latter notation, both χ and ψ are redundant because $\chi = 0$ and $\psi = 2\nu$, so that (1) is a four-parameter density. We do not attempt to detail here the various parameterizations of the GHD: McNeil et al (2005, p. 79) give a list of the parameterizations as well as of their relationships, and the `coef` function of the `ghyp` package allows one to switch among parameterizations. It is also worth noting that the mean-variance mixture representation is simpler for (1) than for (3) because a GIG r.v. $\lambda \sim \text{GIG}(\nu, 0, \psi)$ is a gamma with shape and rate parameters respectively equal to ν and $\psi/2$. But, as we have just recalled, $\psi = 2\nu$ in “alpha.bar”, so that the mixing r.v. is $\lambda \sim \Gamma(\nu, \nu)$.

3 Maximum Likelihood Estimation

The literature about MLE of the univariate VG distribution is mostly based on standard optimization algorithms. Nevertheless, the outcomes obtained with different methods are often different. This may have four, not necessarily exclusive, causes: the different performances of the algorithms employed, the starting values, the computation of the likelihood function, particularly with respect to the singularity at μ , and the different parametrization used.

3.1 Existing approaches to MLE

Direct numerical maximization of the log-likelihood using standard algorithms is the approach taken by the `VarianceGamma` and (in the univariate case) `ghyp` R packages. The `vgFit` function of the former allows the user to employ the Nelder-Mead, the BFGS method or a Newton-type algorithm, while the `fit.VGuv` of the latter is based on the Nelder-Mead algorithm. Both packages use the Nelder-Mead algorithm as implemented in the `optim` R function. BFGS is also implemented in `optim`, whereas the Newton-type algorithm is from the `nlm` function. It is also worth noting that the `ghyp` package uses the MCECM algorithm for estimating the multivariate, but not the univariate, VG distribution.

When $x \rightarrow \mu$, the pdf and the likelihood are computed in different ways in the two packages.

- If $|x - \mu|/\sigma < \epsilon$ and $\nu \leq 1/2$, **ghyp** sets $x = \epsilon$ and computes the density via (1); if $|x - \mu|/\sigma < \epsilon$ and $\nu > 1/2$, **ghyp** uses spline interpolation to compute (2); in both cases, $\epsilon \approx 2.22 \cdot 10^{-16}$ is the machine epsilon.
- If $x = \mu$ and $\nu \leq 1/2$, **VarianceGamma** returns $f(x; \mu, \sigma^2, \gamma, \nu) = \text{Inf}$; if $x = \mu$ and $\nu > 1/2$ **varianceGamma** computes the density via (2).

Starting values for the parameters in the **VarianceGamma** package are either based on a fitted skew-Laplace distribution or on the method of moments, whereas the **ghyp** package uses $\mu^{(0)} = \text{median}(\mathbf{y})$, $\sigma^{(0)} = \text{MAD}(\mathbf{y})$, $\gamma^{(0)} = 0$, where MAD is the Mean Absolute Deviation, and $\nu^{(0)} = 1$. For the EM-type algorithms we use the same starting values of the **ghyp** package.

3.2 Three EM-type algorithms for the VG distribution

The EM algorithm is an iterative procedure for MLE with missing data. It is well known that often MLE in non-missing data cases can be restated so as to become a missing-data problem, and that in many such situations the algorithm is computationally more efficient than standard numerical optimization procedures. This happens, for example, when estimating the parameters of finite mixtures (i.e., mixtures where the mixing variable is discrete; see, e.g., McLachlan and Peel, 2000). More recently, it has been shown that the algorithm is quite effective also in mixtures with continuous mixing distribution (Protassov, 2004; McNeil et al, 2005).

The basic version of the EM algorithm alternates an Expectation (E) and a Maximization (M) step. The former computes the expected value of the complete log-likelihood function given the current value of the parameters and the observed data. The latter maximizes the conditional expectation computed in the E-step. Over the years, various extensions have been developed to handle more complicated cases where the E and/or the M step are not in closed form.

We next explain in detail three EM-type algorithms for the VG distribution, following the derivation developed by Nitithumbundit and Chan (2015) in a multivariate setup. Given a random sample $\mathbf{x} = (x_1, \dots, x_n)'$ from (1), the observed log-likelihood is given by

$$l_{VG}(\mu, \sigma^2, \gamma, \nu; \mathbf{x}) = \sum_{i=1}^n \log(f(x_i; \mu, \sigma^2, \gamma, \nu)), \quad (4)$$

where $f(x; \mu, \sigma^2, \gamma, \nu)$ is the pdf (1).

When dealing with the VG distribution, it is convenient to treat $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_n)'$ as missing data, because, if they were known, the log-likelihood would just be the sum of conditional normal and gamma log-densities:

$$\begin{aligned}
l_c(\mu, \sigma^2, \gamma, \nu; \mathbf{x}, \boldsymbol{\lambda}) &= \sum_{i=1}^n \log(\phi_{\mu, \sigma^2, \gamma}(x_i | \lambda_i) f_{\Gamma}(\lambda_i; \nu, \nu)) = \\
&= \sum_{i=1}^n \{\log(\phi_{\mu, \sigma^2, \gamma}(x_i | \lambda_i)) + \log(f_{\Gamma}(\lambda_i; \nu, \nu))\} = \\
&= -\frac{n}{2} \log \sigma^2 - \frac{1}{2} \sum_{i=1}^n \frac{1}{\lambda_i} \frac{(x_i - \mu - \lambda_i \gamma)^2}{\sigma^2} + \\
&\quad + n\nu \log \nu - n \log \Gamma(\nu) + (\nu - 1) \sum_{i=1}^n \log \lambda_i - \nu \sum_{i=1}^n \lambda_i = \\
&= l_N(\mu, \sigma^2, \gamma | \mathbf{x}, \boldsymbol{\lambda}) + l_{\Gamma}(\nu | \boldsymbol{\lambda}), \tag{5}
\end{aligned}$$

where l_N and l_{Γ} are respectively the conditional normal and gamma log-likelihood. In terms of conditional distributions, the two summands depend on different parameters, so that they can be maximized separately. If $\boldsymbol{\lambda}$ were known, it would be possible to carry out a conditional maximization (CM) of $l_N(\mu, \sigma^2, \gamma | \mathbf{x}, \boldsymbol{\lambda})$ to estimate the normal parameters μ , σ^2 and γ and another CM of $l_{\Gamma}(\nu | \boldsymbol{\lambda})$ to estimate ν . With these premises, it is possible to write down a step-by-step description of the algorithm.

E-step. The E-step requires to take the conditional expectation of the complete log-likelihood function (5). This amounts to computing $E(\lambda_i)$, $E(1/\lambda_i)$ and $E(\log \lambda_i)$, so that we need to know the conditional distribution of $\lambda_i | x_i, \mu, \sigma^2, \gamma, \nu$. Embrechts (1983) has shown that

$$f(\lambda_i | x_i, \mu, \sigma^2, \gamma, \nu) \propto \lambda_i^{\nu-3/2} \exp \left\{ -\frac{(x_i - \mu)^2}{2\lambda_i \sigma^2} - \lambda_i \left(\frac{\gamma^2}{2\sigma^2} + \nu \right) \right\},$$

which is the pdf of a Generalized Inverse Gamma distribution. Accordingly,

the required expectations are given by

$$\hat{\lambda}_i = E(\lambda_i | \mathbf{x}, \mu, \sigma^2, \gamma, \nu) = \frac{\delta_i K_{\nu+1/2}(\sqrt{2\nu + \gamma^2/\sigma^2} \delta_i)}{\sqrt{2\nu + \gamma^2/\sigma^2} K_{\nu-1/2}(\sqrt{2\nu + \gamma^2/\sigma^2} \delta_i)}, \quad (6)$$

$$\widehat{1/\lambda_i} = E\left(\frac{1}{\lambda_i} | \mathbf{x}, \mu, \sigma^2, \gamma, \nu\right) = \frac{\sqrt{2\nu + \gamma^2/\sigma^2} K_{\nu-3/2}(\sqrt{2\nu + \gamma^2/\sigma^2} \delta_i)}{\delta_i K_{\nu-1/2}(\sqrt{2\nu + \gamma^2/\sigma^2} \delta_i)}, \quad (7)$$

$$\widehat{\log \lambda_i} = E(\log \lambda_i | \mathbf{x}, \mu, \sigma^2, \gamma, \nu) = \log\left(\frac{\delta_i}{\sqrt{2\nu + \gamma^2/\sigma^2}}\right) + \frac{\hat{K}_{\nu-1/2}^{(1,0)}(\sqrt{2\nu + \gamma^2/\sigma^2} \delta_i)}{K_{\nu-1/2}(\sqrt{2\nu + \gamma^2/\sigma^2} \delta_i)}, \quad (8)$$

where $\hat{K}_{\nu}^{(1,0)} = (K_{\nu+h}(z) - K_{\nu-h}(z))/2$ is an approximation of $K_{\nu-1/2}^{(1,0)} := \partial K_{\nu}(z)/\partial \alpha|_{\alpha=\nu}$ and $\delta_i = (x_i - \mu)^2/\sigma^2$.

CM-step 1. The first CM-step estimates the parameters μ , σ^2 and γ using the expectations (6) and (7) by maximizing the conditional normal log-likelihood $l_N(\mu, \sigma^2, \gamma | \mathbf{x}, \boldsymbol{\lambda})$. By equating to zero the likelihood equations one gets:

$$\hat{\mu} = \frac{S_{x,\lambda} S_{\lambda} - n S_x}{S_{1,\lambda} S_{\lambda} - n^2}, \quad (9)$$

$$\hat{\gamma} = \frac{S_x - n \hat{\mu}}{S_{\lambda}}, \quad (10)$$

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n \frac{1}{\lambda_i} (x_i - \hat{\mu})^2 - \frac{1}{n} \hat{\gamma}^2 S_{\lambda}, \quad (11)$$

where $S_x = \sum_{i=1}^n x_i$, $S_{x,\lambda} = \sum_{i=1}^n x_i/\lambda_i$, $S_{\lambda} = \sum_{i=1}^n \lambda_i$ and $S_{1,\lambda} = \sum_{i=1}^n \lambda_i^{-1}$ are the complete-data sufficient statistics.

CM-step 2a. The last parameter to be estimated is ν . As in the first CM-step, the maximization is carried out conditionally on λ . The maximum of $l_{\Gamma}(\nu | \lambda)$ cannot be found explicitly, so that this step must be performed numerically via the Newton-Raphson algorithm. In this case one needs the first two partial derivatives of the log-likelihood function as well as the estimate of $E(\log \lambda_i)$, i.e. (8), in the E-step. The partial derivatives are given

by Nitithumbundit and Chan (2015):

$$\begin{aligned}\frac{\partial l}{\partial \nu} &= n + n \log \nu - n\psi^{(0)}(\nu) + S_{\log \lambda} - S_{\lambda}, \\ \frac{\partial^2 l}{\partial \nu^2} &= \frac{n}{\nu} - n\psi^{(1)}(\nu),\end{aligned}$$

where $S_{\log \lambda} = \sum_{i=1}^n \log(\lambda_i)$, $\psi^{(0)}(x) = d \log \Gamma(x)/dx$ is the digamma function and $\psi^{(1)}(x) = d\psi^{(0)}(x)/dx$.

CM-step 2b. Alternatively, it is possible to estimate ν by maximizing $l_{VG}(\nu|\mu^{(t+1)}, \sigma^{2(t+1)}, \gamma^{(t+1)}, \mathbf{x})$ with respect to ν using some standard numerical optimization routine, such as the `optimize` R function.

A pseudo-code for the t -th iteration of the algorithm is as follows.

- E step 1: compute $\hat{\lambda}_i^{(t+1/2)}$ and $\widehat{1/\lambda_i}^{(t+1/2)}$ using (6) and (7) and the current estimates $\hat{\mu}^{(t)}$, $\hat{\sigma}^{2(t)}$, $\hat{\gamma}^{(t)}$ and $\hat{\nu}^{(t)}$; compute also the sufficient statistics $S_{x,\lambda}$, S_{λ} and $S_{1,\lambda}$.
- CM step 1: compute $\hat{\mu}^{(t+1)}$, $\hat{\gamma}^{(t+1)}$ and $\hat{\sigma}^{2(t+1)}$ using (9), (10) and (11).
- E step 2: compute $\hat{\lambda}_i^{(t+1)}$ and $\widehat{\log \lambda_i}^{(t+1)}$ using the updated estimates $\hat{\mu}^{(t+1)}$, $\hat{\sigma}^{2(t+1)}$ and $\hat{\gamma}^{(t+1)}$ as well as $\hat{\nu}^{(t)}$. Compute also the sufficient statistics $S_{\lambda}^{(t+1)}$ and $S_{\log \lambda}^{(t+1)}$.
- CM step 2a: compute $\hat{\nu}^{(t+1)}$ using the sufficient statistics $S_{\lambda}^{(t+1)}$ and $S_{\log \lambda}^{(t+1)}$ and the Newton-Raphson method.
- CM step 2b: alternatively, maximize $l_{VG}(\nu|\mu^{(t+1)}, \sigma^{2(t+1)}, \gamma^{(t+1)}, \mathbf{x})$.

Given the rather complicated form of the VG log-likelihood and the possible singularity at μ , the algorithm is sometimes numerically unstable for certain values of the parameters. The first difficulty may arise in the E step. Nitithumbundit and Chan (2015) have obtained limit expressions for $E(\lambda_i|\mu, \sigma^2, \gamma, \nu, \mathbf{x})$, $E(\lambda_i^{-1}|\mu, \sigma^2, \gamma, \nu, \mathbf{x})$ and $E(\log \lambda_i|\mu, \sigma^2, \gamma, \nu, \mathbf{x})$. As the last two expectations do not exist, they propose to replace δ_i by $\delta_i^* := \Delta/\sqrt{2\nu + \gamma^2/\sigma^2}$ when δ_i is smaller than some predefined small value. It will be shown in Section 4 that the actual value of $\Delta \in \mathbb{R}^+$ has a non-negligible impact on the properties of the estimators.

The second difficulty has to do with the estimation of σ^2 , because the quantity $\widehat{1/\lambda_i}(x_i - \hat{\mu})^2$ diverges if $\widehat{1/\lambda_i} \rightarrow \infty$. To rule out this possibility, we insert an additional E-step to estimate $1/\lambda_i$ and λ_i after updating μ and γ (Nitithumbundit and Chan, 2015, Section 3.3.1).

The algorithm based on CM step 2a is a MultiCycle, Expectation, Conditional Maximization algorithm (MCECM - see Meng and Rubin, 1993 and Meng, 1994), whereas the version that uses CM step 2b is called Expectation Conditional Maximization Either (ECME - see Liu and Rubin, 1995). A likely advantage of the former is that we do not need to adjust the VG pdf, as its evaluation is not needed in the maximization process; on the other hand the E step of ECME does not require the computation of $E(\log \lambda_i | \mathbf{x}, \mu, \sigma^2, \gamma, \nu)$. However, it is not clear a priori which of the two advantages is more important; more generally, the computational efficiency of the two algorithms has to be assessed via simulation.

Both MCECM and ECME possess the stable monotone convergence property, i.e. the likelihood is increased at each iteration. In Nitithumbundit and Chan (2015)'s implementation, the algorithm stops when the difference between the log-likelihood values at iterations $t + 1$ and t is smaller than 10^{-8} . Given the monotone convergence property, however, it is equivalent to stop the algorithm when $\max_{i=1, \dots, 4} (\theta_i^{(t+1)} - \theta_i^{(t)})$ is smaller than a predefined threshold η , where $\boldsymbol{\theta} = (\mu, \sigma^2, \gamma, \nu)'$. In this paper we use the latter criterion, mainly because this way of proceeding has the desirable consequence that MCECM does not require the evaluation of the full log-likelihood function.

In addition to MCECM and ECME, Nitithumbundit and Chan (2015) also propose an hybrid algorithm (HECM) combining MCECM and ECME: more precisely, HECM starts with MCECM and then carries out ECME using the estimates of the first step as starting values.

4 Simulation

The Monte Carlo analysis performed in this section has various goals. First, we compare the performance of the MCECM, ECME and HECM algorithms. As will be seen in Section 4.1 below, MCECM turns out to be preferable; hence, we restrict the attention to MCECM in the rest of the paper. Second, we aim at finding the impact of the parameter Δ on the efficiency of MCECM, trying to identify its optimal value. Third, we carry out a detailed comparison

of MCECM and of the procedures implemented in the `VarianceGamma` and `ghyp` R packages. This will be done for various sample sizes and parameter values. Finally, we analyze how the sampling variability of the estimators extends to VaR measures.

All the experiments consist of simulating n observations from the VG distribution, estimating the parameters and computing the bias, the variance and the root-mean-squared-error (RMSE). The number of replications is equal to $B = 100$ throughout the paper. The algorithm stops when $\max_{i=1,\dots,4}(\theta_i^{(t+1)} - \theta_i^{(t)}) < 10^{-8}$ or when the maximum number of iterations, which is set to 2000, is reached. As for the true numerical values of the parameters, we choose values already employed in the literature and close to those typically found in financial applications, i.e. $\mu = 0.000952372$, $\sigma^2 \in \{0.17, 0.14, 0.04, 0.03, 6.44 \cdot 10^{-5}\}$, $\gamma \in \{-0.000843568, -0.2, -0.3\}$ and $\nu \in \{0.5, 1.991437, 7.509199\}$

4.1 EM-type algorithms: a comparison

In this section we compare the three EM-type algorithms for various values of Δ . Tables A.1 and A.2 in the supplementary material display the outcomes respectively for the case $\mu = 0.00095$, $\sigma^2 = 0.04294$, $\gamma = -0.00084$, $\nu = 1.99144$ and $\mu = 0.00095$, $\sigma^2 = 0.04294$, $\gamma = -0.3$, $\nu = 1.99144$ with $n = 500$.

Tables A.1 and A.2 show that, in terms of statistical accuracy, the results of the three algorithms are not significantly different, yet MCECM seems to be slightly preferable for all parameters and all values of Δ . Moreover, even though in terms of number of iterations it is not the best performer, the computational cost of a single iteration is by far smaller, so that the total time taken by MCECM is smaller as well. Similar outcomes are obtained for different parameter configurations and are therefore not shown here. We have tried to repeat the simulations with smaller sample size ($n \leq 200$). In this case, ECME and HECM encounter some numerical difficulties: the two reasons are that the evaluation of the likelihood sometimes becomes impossible (and the algorithm breaks down), or the algorithm aborts without reaching convergence within the maximum number of iterations. On the other hand, MCECM works well even for sample sizes as small as 50.

The main message conveyed by these experiments is that MCECM is overall preferable; accordingly, we use MCECM in the rest of the paper.

4.2 Impact of Δ

In the univariate case, Nitithumbundit and Chan (2015, Table 4) give the results of a simulation for $\nu = 0.3$, $\sigma^2 = 1$ and $\gamma = 0$, which suggests that very little difference is observed for Δ in the range $[10^{-300}, 10^{-3}]$. However, their experiment is focused on the symmetric VG model and on a single set of parameter values. Moreover, they only base the comparison on the bias of the estimators, without any detail about the variability.

Since the features of the distribution change markedly with the numerical values of the parameters, we carry out a more comprehensive experiment, considering the models corresponding to all combinations of the parameter values listed in Section 4. Tables A.3 to A.47 show the RMSEs of the estimators for the 45 parameter combinations.

In Section A.1, which refers to the case $\nu = 0.5$, overall $\Delta = 0.01$ provides the best results among the 8 values of Δ considered, typically because of a much smaller bias. In the remaining tables (Sections A.2 and A.3) it is difficult to identify a pattern depending on Δ , but the case $\Delta = 0.01$ is always one of the best performers, so that we use $\Delta = 0.01$ throughout the paper. Note that this conclusion is different from Nitithumbundit and Chan (2015): this may be due to the difference between the univariate and the multivariate estimation problem as well as to the different numerical values of the parameters.

4.3 MCECM vs standard approaches

Starting from the fact that the MCECM algorithm with $\Delta = 0.01$ is the best EM-type solution, the analysis related to the precision of the estimators is based on a comparison among the MCECM algorithm with $\Delta = 0.01$ and the standard approaches implemented in the `VarianceGamma` and `ghyp` R packages.

The first three lines of Table 1 show the average RMSE-ranking of the estimators over all the 45 models considered. The MCECM algorithm has the best performance for all parameters; in all cases `ghyp` is second and `VarianceGamma` is third. The rest of the table gives the “marginal” rankings for the models with one parameter fixed: for example, the case $\nu = 0.5$ is the ranking of all the models with $\nu = 0.5$. In this way we try to assess whether any method works particularly well (or badly) when some parameter takes a specific value. If we consider only the estimators of σ^2 , γ and ν , which are

		GHYP	VG	MCECM
All models	μ	1.87	2.53	1.60
	σ^2	2.02	2.76	1.22
	γ	2.13	2.6	1.27
	ν	2.20	2.36	1.44
$\nu = 0.5$	μ	1.20	2.87	1.93
	σ^2	2	3	1
	γ	2	3	1
	ν	1.87	3	1.13
$\nu = 1.99144$	μ	1.80	3	1.20
	σ^2	1.73	3	1.27
	γ	1.73	3	1.27
	ν	1.80	3	1.20
$\nu = 7.50920$	μ	2.60	1.73	1.67
	σ^2	2.33	2.27	1.40
	γ	2.6	1.8	1.6
	ν	2.93	1.07	2.00
$\sigma^2 = 6.44 \cdot 10^{-5}$	μ	1.44	3	1.56
	σ^2	1.44	2.89	1.67
	γ	1.56	3	1.44
	ν	1.78	2.33	1.89
$\sigma^2 = 0.03$	μ	1.89	2.44	1.67
	σ^2	2.11	2.89	1.00
	γ	2.22	2.56	1.22
	ν	2.33	2.44	1.22
$\sigma^2 = 0.04$	μ	2.11	2.44	1.44
	σ^2	2.00	2.89	1.11
	γ	2.22	2.56	1.22
	ν	2.33	2.33	1.33
$\sigma^2 = 0.14$	μ	2	2.44	1.56
	σ^2	2.22	2.67	1.11
	γ	2.33	2.56	1.11
	ν	2.22	2.33	1.44
$\sigma^2 = 0.17$	μ	1.89	2.33	1.78
	σ^2	2.33	2.44	1.22
	γ	2.22	2.33	1.44
	ν	2.33	2.33	1.33
$\gamma = -0.00084$	μ	1.87	2.40	1.73
	σ^2	2.07	2.87	1.07
	γ	2.13	2.53	1.33
	ν	2.13	2.33	1.53
$\gamma = -0.2$	μ	1.87	2.53	1.60
	σ^2	2.00	2.73	1.27
	γ	2.13	2.60	1.27
	ν	2.20	2.33	1.47
$\gamma = -0.3$	μ	1.87	2.67	1.47
	σ^2	2.00	2.67	1.33
	γ	2.07	2.67	1.27
	ν	2.27	2.40	1.33

Table 1: Average RMSE-ranking of the estimators of the parameters of the VG distribution, computed using 100 replications and a sample size $n = 500$.

much more important than μ , there are two cases in which MCECM is not best: when $\sigma^2 = 6.44 \cdot 10^{-5}$ the estimators of σ^2 and ν are second, and when $\nu = 7.50920$ only the estimator of ν is second.

When $\nu = 7.50920$ the VG estimators perform considerably better than in the models with smaller values of ν (the only instance where one of the VG estimators is best is the case of the estimator of ν in the models with $\nu = 7.50920$). However, the precision in estimating ν when ν is large is somewhat less important in practice, as the impact of ν on the shape of the pdf is less and less important when ν gets large.

When $\sigma^2 = 6.44 \cdot 10^{-5}$ the `ghyp` estimators of σ^2 and ν are preferable. Considering that the “small- σ^2 ” setup is challenging and important with financial data, we will study it more in detail in Section 4.5.

In the simulation experiments we have tried different starting values. Whereas the MCECM results do not change, both `ghyp` and `VarianceGamma` give different outcomes if initialized differently; in addition, the `vgFit` function of the `VarianceGamma` package, if initialized using the estimators obtained by means of the method of moments, often breaks down immediately because the log-likelihood cannot be evaluated at the initial parameters. The insensitivity to the starting values (under mild conditions: McLachlan and Krishnan, 2008, p. 28) is a well-known feature of the EM algorithm, and is another advantage of MCECM in the present setup.

4.4 Impact on risk measures estimation

The aim of this section is the assessment of the impact of the variability of the estimators on the VaR. For each of the models considered above, we compute via simulation the true 95% VaR (using the true parameter values) and the VaRs based on the B estimated parameter vectors. Finally we compute the bias, variance and RMSE of the B VaRs.

The average RMSE-ranking of the VaR estimates is reported in the first line of Table 2. Similarly to the outcomes shown in Table 1, MCECM performs best. The rest of the table gives the marginal rankings for the models with one parameter fixed, and MCECM is best also in all the marginal models.

	GHYP	VG	MCECM
All models	2.00	2.53	1.47
$\nu = 0.5$	2	3	1
$\nu = 1.99144$	1.6	3	1.4
$\nu = 7.50920$	2.67	1.73	1.60
$\sigma^2 = 6.44 \cdot 10^{-5}$	1.56	3	1.44
$\sigma^2 = 0.03$	2.22	2.44	1.33
$\sigma^2 = 0.04$	2.22	2.56	1.22
$\sigma^2 = 0.14$	2.22	2.56	1.22
$\sigma^2 = 0.17$	2.22	2.33	1.44
$\gamma = -0.00084$	2.07	2.53	1.4
$\gamma = -0.2$	2	2.6	1.4
$\gamma = -0.3$	2.2	2.6	1.2

Table 2: RMSE-rankings of the estimated 95% VaRs, computed using 100 replications and a sample size $n = 500$. The total number of VG distributions is equal to 45, i.e. the number of combinations of $\mu = 0.000952372$, $\sigma^2 \in \{0.17, 0.14, 0.04, 0.03, 6.44 \cdot 10^{-5}\}$, $\gamma \in \{-0.000843568, -.2, -.3\}$, $\nu \in \{0.5, 1.991437, 7.50920\}$.

		γ	σ^2	ν
Bias	GHYP	$1.25 \cdot 10^{-6}$	$2.03 \cdot 10^{-5}$	0.02
	MCECM	$-7.54 \cdot 10^{-6}$	$8.93 \cdot 10^{-5}$	0.04
	Rel. perf.	-0.17	0.23	0.49
RMSE	GHYP	$1.08 \cdot 10^{-10}$	$2.86 \cdot 10^{-7}$	$4.32 \cdot 10^{-3}$
	MCECM	$9.47 \cdot 10^{-11}$	$1.62 \cdot 10^{-7}$	$4.92 \cdot 10^{-3}$
	Rel. perf.	1.14	1.77	0.88

Table 3: Bias, RMSE and relative performance of the estimators computed using 100 replications and a sample size $n = 500$. The parameters are equal to $\mu = 0.000952372$, $\sigma^2 = 6.44 \cdot 10^{-5}$, $\gamma = -0.000843568$, $\nu = 0.5$. Relative performance is given by $\text{bias}^{GHYP}/\text{bias}^{MCECM}$ or $\text{RMSE}^{GHYP}/\text{RMSE}^{MCECM}$.

		γ	σ^2	ν
Bias	GHYP	$5.57 \cdot 10^{-3}$	$-4.39 \cdot 10^{-2}$	$-3.75 \cdot 10^{-2}$
	MCECM	$-6.43 \cdot 10^{-5}$	$-2.49 \cdot 10^{-2}$	$5.92 \cdot 10^{-3}$
	Rel. perf.	-86.49	1.76	-6.33
RMSE	GHYP	$1.05 \cdot 10^{-4}$	$9.08 \cdot 10^{-3}$	$5.56 \cdot 10^{-3}$
	MCECM	$4.14 \cdot 10^{-9}$	$8.08 \cdot 10^{-4}$	$5.60 \cdot 10^{-4}$
	Rel. perf.	25 386.22	11.24	9.93

Table 4: Bias, RMSE and relative performance of the estimators computed using 100 replications and a sample size $n = 500$. The parameters are equal to $\mu = 0.000952372$, $\sigma^2 = 6.44 \cdot 10^{-5}$, $\gamma = -0.2$, $\nu = 0.5$. Relative performance is given by $\text{bias}^{GHYP}/\text{bias}^{MCECM}$ or $\text{RMSE}^{GHYP}/\text{RMSE}^{MCECM}$.

4.5 A closer look at two important cases

The models with $\sigma^2 = 6.44 \cdot 10^{-5}$ are among the most challenging for the algorithms. Hence, we look more closely at the results of two “small- σ^2 ” parameter configurations, i.e. $\sigma^2 = 6.44 \cdot 10^{-5}$, $\gamma \in \{-0.000843568, -0.2\}$, $\nu = 0.5$. These two cases are especially challenging because the density is very skewed and concentrated near the singularity. Moreover, the “small- σ^2 ” framework is the setup where the advantage of MCECM is smallest. Tables 3 and 4 show the bias and the RMSE of the estimators of γ , σ^2 and ν for the two models.

In the former case ($\mu = 0.000952372$, $\sigma^2 = 6.44 \cdot 10^{-5}$, $\gamma = -0.000843568$,

$\nu = 0.5$, Table 3) the two approaches perform similarly: **ghyp** is slightly preferable in terms of bias, and MCECM in terms of RMSE. On the other hand, in the latter, MCECM is much better than **ghyp** both in terms of bias and of RMSE, particularly as concerns the estimation of γ . The histograms in Figure 2 confirm that the VaRs corresponding to MCECM are much better.

5 Real data analysis

5.1 Standard & Poor's index

We first consider the computation of VaR for the Standard & Poor's 500 index using daily log-returns from December 1, 2015 to December 1, 2016. As the returns are typically dependent, we first apply a GARCH filter (McNeil and Frey, 1998; Hu and Kercheval, 2007). Formally, let $r_t^* = \log(P_t/P_{t-1})$, $r_t = r_t^* - \bar{r}$ and

$$r_t = \sigma_t Z_t, \quad \sigma_t^2 = \alpha_0 + \alpha_1 r_{t-1}^2 + \beta_1 \sigma_{t-1}^2,$$

i.e. a GARCH(1,1) model. We estimate σ_t^2 using quasi maximum likelihood as implemented in the **garchFit** function of the **fGarch** R package (Wuertz, 2016) and compute the filtered return series $\hat{Z}_t \stackrel{\text{def}}{=} r_t / \hat{\sigma}_t$, where $\hat{\sigma}_t$ is the estimated GARCH(1,1) volatility. The correlograms in Figure A.1 confirm that r_t^2 is autocorrelated but $\hat{Z}_t^2$ is not, so that the series $\hat{Z}_t$ can be taken as approximately iid and can be used for estimating the VG parameters. The $(1 - \alpha)\%$ dynamic VaR is finally given by

$$\text{VaR}_{1-\alpha} = \hat{\sigma}_t \hat{F}^{-1}(\alpha), \quad (12)$$

where $\hat{F}^{-1}$ is the quantile function corresponding to the estimated parameters of the chosen distribution. For comparison purposes, we also compute the normal and Student t VaR; in these cases the quantile function $\hat{F}^{-1}$ in (12) is respectively given by the standard normal and Student t (with number of degrees of freedom estimated via MLE) quantile function.

The sample skewness and kurtosis are $(sk, \kappa)' = (-0.351, 4.552)'$, and the estimated parameters obtained with the three algorithms are reported in Table 5. It can be seen that the **VarianceGamma** estimators are markedly different from **ghyp** and MCECM.

Figure 2: Simulated distributions of the 95% VaR computed using **ghyp** (top) and MCECM (bottom). The left column refers to the case $\sigma^2 = 6.44 \cdot 10^{-5}$, $\gamma = -0.000843568$, $\nu = 0.5$, the right column to the model $\sigma^2 = 6.44 \cdot 10^{-5}$, $\gamma = -0.2$, $\nu = 0.5$. The vertical line is the true VaR.

Parameter	VG	GHYP	MCECM
μ	-0.257	0.048	0.048
σ^2	6.338	0.987	0.987
γ	1.742	-0.056	-0.056
ν	0.348	1.688	1.687

Table 5: Estimated parameters for the Standard & Poor's example.

Figure 3: Standard & Poor's returns (Dec. 1, 2015 - Dec. 1, 2016) and 95% estimated VaRs.

level		VG	GHYP	MCECM	t	Norm
$\alpha = 10\%$	Quant.	-1.016	-1.199	-1.203	-1.467	-1.282
	Freq.	0.150	0.099	0.099	0.063	0.083
$\alpha = 5\%$	Quant.	-1.688	-1.655	-1.666	-1.996	-1.645
	Freq.	0.040	0.040	0.040	0.024	0.043
$\alpha = 1\%$	Quant.	-3.502	-2.680	-2.687	-3.306	-2.326
	Freq.	0.008	0.008	0.008	0.008	0.012

Table 6: Estimated quantiles (Quant.) and relative frequency of violations (Freq.) for the Standard & Poor's data.

Parameter	VG	GHYP	MCECM
μ	0.186	0.120	0.120
σ^2	1.047	1.003	1.004
γ	-0.211	-0.144	-0.144
ν	0.805	0.983	0.981

Table 7: Estimated parameters for the Banca Monte dei Paschi example.

Figure 3 shows the returns and the VaRs, and Table 6 gives the relative frequency of VaR violations. MCECM and **ghyp** have the best performance for $\alpha = 10\%$, whereas all the approaches produce similar results for $\alpha = 5\%$, with the exception of the Student t , which overestimates the VaR. When $\alpha = 1\%$, the algorithms do not differ much in terms of frequency of violations, but the VaR computed with **VarianceGamma** is much larger in absolute value.

5.2 Monte dei Paschi di Siena

The same analysis has been carried out using daily closing prices of the Banca Monte dei Paschi di Siena stock from December 1, 2015 to December 1, 2016. This was a quite turbulent period for Banca Monte dei Paschi, whose return series is indeed more skewed and leptokurtic than the S&P's index: the sample skewness and kurtosis are equal to $(sk, \kappa)' = (0.537, 6.540)'$. Similarly to the preceding example, Figure A.2 suggests that $\hat{Z}_t$ can be treated as an approximately iid series. Table 7 displays the estimated parameters, whereas the outcomes of the VaR analysis are shown in Figure 4 and Table 8. When $\alpha = 10\%$, all the methods except the Student t distribution give the same outcomes, whereas **ghyp** and MCECM are best for $\alpha = 5\%$ and $\alpha = 1\%$.

6 Conclusion

In this paper we have studied MLE of the variance-gamma distribution. The distribution is flexible enough to accommodate skewness and leptokurtosis but MLE is difficult because the pdf is not in closed form and is unbounded for small values of the shape parameter. Two R packages (**VarianceGamma** and **ghyp**) compute the MLEs of the VG distribution, and both of them use standard optimization routines.

Figure 4: Estimated parameters for the Standard & Poor's example.

level		VG	GHYP	MCECM	t	Norm
$\alpha = 10\%$	Quant.	-1.245	-1.209	-1.227	-1.614	-1.282
	Perc.	0.0916	0.0916	0.0916	0.0573	0.0916
$\alpha = 5\%$	Quant.	-1.859	-1.763	-1.751	-2.302	-1.645
	Perc.	0.0420	0.0496	0.0496	0.0229	0.0573
$\alpha = 1\%$	Quant.	-3.309	-2.970	-3.029	-4.349	-2.326
	Perc.	0.0076	0.0114	0.0114	0	0.0229

Table 8: Estimated quantiles (Quant.) and relative frequency of violations (Freq.) for the Monte dei Paschi data.

As the VG is a normal mean-variance mixture, it is natural to use an EM-type algorithm for estimation. Accordingly, we develop three extensions of the EM algorithm, based on the multivariate approach proposed by Nitithumbundit and Chan (2015). Extensive simulation experiments suggest that MCECM is the most computationally efficient of the three algorithms and allow us to find an appropriate value of the input parameter Δ ; this is quite important because they are found to have a strong impact on the statistical properties of the estimators. Moreover, we find that MCECM is preferable to both `VarianceGamma` and `ghyp`, but the latter is more precise than the former, and its performance is similar to MCECM. A valuable outcome of the simulation is that MCECM works particularly well when the scale and shape parameters are small. Furthermore, MCECM, unlike `VarianceGamma` and `ghyp`, is insensitive to the starting values of the parameters.

Finally, we use stock market observations to compute the Value-at-Risk. Besides confirming that MCECM and `ghyp` are the most precise algorithms, the applications suggest that the fit of the VG distribution to financial returns is very good, also in comparison to other models, and the VaR is computed with high precision at all confidence levels.

The results obtained in this paper suggest that algorithms that take explicitly into account the stochastic nature of the maximization problem related to MLE may be significantly more efficient than general-purpose optimization techniques. A possible explanation is that the MCECM algorithm splits the maximization problem into two simpler problems, corresponding to a conditional normal and a gamma likelihood, and, as often happens in numerical optimization, maximizing separately two lower-dimensional functions is easier than maximizing a single higher-dimensional function.

Acknowledgments. An earlier version of this paper was presented at the 10th International Conference on Computational and Financial Econometrics (CFE 2016), Seville, December 9-11, 2016. We thank the participants for helpful comments and suggestions. The R codes used for estimating the VG distribution using the the MCECM, ECME and HECM algorithms and for computing the VaR are available at <http://marcobee.weebly.com/software.html>.

Appendix

Supplementary material related to this article can be found online.

References

- Barndorff-Nielsen OE (1978) Hyperbolic distributions and distributions on hyperbolae. *Scandinavian Journal of Statistics* 5:151–157
- Barndorff-Nielsen OE, Blæsild P (1981) Hyperbolic distributions and ramifications: contributions to theory and application. In: Taillie C, G P, Baldessari B (eds) *Statistical Distributions in Scientific Work*, vol 4, Reidel, Dordrecht:, pp 19–44
- Carr PP, Madan DB (1999) Option valuation using the fast Fourier transform. *Journal of Computational Finance* 1:61–73
- Cervellera GP, Tucci MP (2016) A note on the estimation of a gamma-variance process: Learning from a failure. *Computational Economics* DOI 10.1007/s10614-016-9566-3
- Cont R (2001) Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance* 1:223–236
- Dempster N, Laird N, Rubin D (1977) Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society B* 39:1–38
- Eberlein E, Keller U (1995) Hyperbolic distributions in finance. *Bernoulli* 1:281–299
- Embrechts P (1983) A property of the generalized inverse Gaussian distribution with some applications. *Journal of Applied Probability* 20:537–544
- Hu W, Kercheval A (2007) Risk management with generalized hyperbolic distributions. In: *Proceedings of the Fourth IASTED International Conference on Financial Engineering and Applications*, ACTA Press, Anaheim, CA, USA, FEA '07, pp 19–24, URL <http://dl.acm.org/citation.cfm?id=1650320.1650324>

- Liu C, Rubin DB (1995) ML estimation of the t distribution using EM and its extensions, ECM and ECME. *Statistica Sinica* 5:19–39
- Luethi D, Breymann W (2016) ghyp: A package on the generalized hyperbolic distribution and its special cases. URL <https://cran.r-project.org/web/packages/ghyp/index.html>, R package version 1.5.7
- Madan DB, Milne F (1991) Option pricing with VG martingale components. *Mathematical Finance* 1:39–55
- Madan DB, Seneta E (1990) The variance gamma (vg) model for share market returns. *Journal of Business* 63:511–524
- Madan DB, Carr P, Chang E (1998) The variance gamma process and option pricing. *European Finance Review* 2:135–156
- McLachlan G, Krishnan T (2008) *The EM Algorithm and Extensions*, 2nd edn. Wiley
- McLachlan G, Peel D (2000) *Finite Mixture Models*. Wiley
- McNeil A, Frey R, Embrechts P (2005) *Quantitative Risk Management: Concepts, Techniques, Tools*. Princeton University Press
- McNeil AJ, Frey R (1998) Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. *Journal of Empirical Finance* 7:271–300
- Meng XL (1994) On the rate of convergence of the ECM algorithm. *Annals of Statistics* 22:326–339
- Meng XL, Rubin D (1993) Maximum likelihood estimation via the ECM algorithm: a general framework. *Biometrika* 80:267–278
- Musiela M, Rutkowski M (2005) *Martingale Methods in Financial Modelling*, 2nd edn. Springer
- Nitithumbundit T, Chan JSK (2015) An ECM algorithm for Skewed Multivariate Variance Gamma Distribution in Normal Mean-Variance Representation. ArXiv e-prints 1504.01239

- Prause K (1999) The generalized hyperbolic model: Estimation, financial derivatives, and risk measures. PhD thesis, Albert-Ludwigs-Universität Freiburg
- Protasov R (2004) EM-based maximum likelihood parameter estimation for multivariate generalized hyperbolic distributions with fixed λ . *Statistics and Computing* 14:67–77
- R Core Team (2016) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, URL <http://www.R-project.org/>
- Scott D, Yang Dong C (2015) VarianceGamma: The Variance Gamma distribution. URL <https://cran.r-project.org/web/packages/VarianceGamma/index.html>, R package version 0.3-1
- Seneta E (2004) Fitting the variance-gamma model to financial data. *Journal of Applied Probability* 41:177–187
- Wuertz D (2016) fGarch: GARCH modelling package. R package version 3010.82.1