



The Variance-Gamma Distribution: A Review

Document Version

Accepted author manuscript

[Link to publication record in Manchester Research Explorer](#)

Citation for published version (APA):

Fischer, A., Gaunt, R., & Sarantsev, A. (in press). The Variance-Gamma Distribution: A Review. *Statistical Science*.

Published in:

Statistical Science

Citing this paper

Please note that where the full-text provided on Manchester Research Explorer is the Author Accepted Manuscript or Proof version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version.

General rights

Copyright and moral rights for the publications made accessible in the Research Explorer are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Takedown policy

If you believe that this document breaches copyright please refer to the University of Manchester's Takedown Procedures [<http://man.ac.uk/04Y6Bo>] or contact openresearch@manchester.ac.uk providing relevant details, so we can investigate your claim.



The Variance-Gamma Distribution: A Review

Adrian Fischer , Robert E. Gaunt and Andrey Sarantsev

Abstract. The variance-gamma (VG) distributions form a four-parameter family which includes as special and limiting cases the normal, gamma and Laplace distributions. Some of the numerous applications include financial modelling and distributional approximation on Wiener space. In this review, we provide an up-to-date account of the basic distributional theory of the VG distribution. Properties covered include probability and cumulative distribution functions, generating functions, moments and cumulants, mode and median, Stein characterisations, representations in terms of other random variables, and a list of related distributions. We also review methods for parameter estimation and some applications of the VG distribution, including the aforementioned applications to financial modelling and distributional approximation on Wiener space.

Key words and phrases: Variance-gamma distribution, distributional theory, estimation, variance-gamma process, financial modelling, approximation on Wiener space.

1. INTRODUCTION

The variance-gamma (VG) distribution with parameters $r > 0, \theta \in \mathbb{R}, \sigma > 0, \mu \in \mathbb{R}$, denoted by $\text{VG}(r, \theta, \sigma, \mu)$, has probability density function (PDF)

$$(1) \quad p(x) = \frac{1}{\sigma \sqrt{\pi} \Gamma(r/2)} e^{\theta(x-\mu)/\sigma^2} \left(\frac{|x-\mu|}{2\sqrt{\theta^2 + \sigma^2}} \right)^{\frac{r-1}{2}} \times K_{\frac{r-1}{2}} \left(\frac{\sqrt{\theta^2 + \sigma^2}}{\sigma^2} |x-\mu| \right), \quad x \in \mathbb{R}.$$

In the limit $\sigma \rightarrow 0$ the support becomes the interval (μ, ∞) if $\theta > 0$, and is $(-\infty, \mu)$ if $\theta < 0$. Here $K_\nu(x)$ is a modified Bessel function of the second kind; a definition and basic properties that are used throughout the paper are given in Appendix A. The parametrisation (1) is taken from Gaunt [36] and is similar to one given by Finlay and Seneta [31]. Alternative parametrisations are given in Bibby and Sørensen [15] and the book of Kotz, Kozubowski and Podgórski [64], in which they refer to the distribution as the generalized (asymmetric) Laplace distribution. Other names for the VG distribution include the Bessel function distribution (McKay [87]) and the McKay Type II distribution (Holm and Alouini [57]).

Adrian Fischer is Postdoctoral Research Associate, University of Oxford, Oxford, UK (e-mail: adrian.fischer@stats.ox.ac.uk).
Robert E. Gaunt is Senior Lecturer, University of Manchester, Manchester, UK (e-mail: robert.gaunt@manchester.ac.uk).
Andrey Sarantsev is Assistant Professor, University of Nevada, Reno, USA (e-mail: asarantsev@unr.edu).

The PDF (1) was first written down by Pearson, Jefferey and Elderton [100] as the exact PDF of the sample covariance for a random sample drawn from a bivariate normal population. Further distributional properties and applications were given by McKay [87] and Pearson, Stouffer and David [101]. The flexibility offered by the four parameters and the modified Bessel function of the second kind in the PDF mean that the VG distribution is often well-suited to statistical modelling; for example, Sichel [117] reported that the VG distribution provided an excellent fit when modelling the size of diamonds mined in South West Africa. The VG distribution (in the symmetric case $\theta = 0$) was introduced to the financial literature in a seminal work of Madan and Seneta [83], and has since found application in financial modelling; see, for example, Madan, Carr and Chang [79], Madan and Milne [80] and Seneta [115]. The VG distribution is a special case of the generalized hyperbolic (GH) distribution that is also popular in financial modelling; see, for example, Eberlein and Keller [25], Eberlein and Prause [25] and Rydberg [109]. The $\text{VG}(r, 0, \sigma, 0)$ PDF is (up to a constant) the Matérn covariance function, which is widely used in spatial statistics (Stein [119]). In addition to its suitability in statistical modelling, the VG distribution has a rich distributional theory with special or limiting cases including the normal, gamma and Laplace distributions, and the product of two zero mean normals and the difference of two independent gammas. In virtue of this, starting with Gaunt [36] and Eichelsbacher and Thäle [28], the VG distribution has recently found application in the probability literature as a limiting distribution, most notably in analysis on Wiener space.

To date, the most comprehensive account of the VG distribution in the literature is given in Chapter 4 of the excellent book of Kotz et al. [64], in which the VG distribution is viewed as a natural generalisation of the classical (asymmetric) Laplace distribution. As the VG distribution is a special case of the GH distribution, distributional properties can be inferred from results for the GH distribution found in, for example, Bibby and Sørensen [15] and Hammerstein [54]. The aforementioned references, however, do not contain even some of the most basic distributional properties of the VG distribution, such as formulas for absolute moments or the mode, and the interested researcher is left to search a rather large and difficult to navigate literature to track down many such results.

In this review, we fill in a gap in the literature by providing an up-to-date account of the basic distributional theory of the VG distribution. The end result is that many of the most important distributional properties of the VG distribution are now contained in a single reference. Most results are already stated in the literature or can be inferred from the fact that VG distribution is a special case of the GH distribution and appealing to known results for this distribution. For the small number of results that we could not locate in the literature, we provide straightforward and concise derivations.

This review covers formulas for the PDF (Section 2.1), the cumulative distribution function (Section 2.2), generating functions, infinite divisibility and self-decomposability (Section 2.3), representations in terms of other random variables (Section 2.4), lists of related distributions (Section 2.5), Stein characterisation (Section 2.6), moments and cumulants (Section 2.7), and mode and median (Section 2.8). The review covers some of the most basic and important properties of a probability distribution, but is not comprehensive; for example, we only briefly touch upon multivariate extensions.

In Section 3, we review methods for parameter estimation for the VG distribution. This section provides a concise overview of a literature that is rather large on account of the popularity of the VG distribution in statistical modelling, particularly in mathematical finance. In Section 4, we provide applications of the VG distribution in several research domains. In Section 4.1, we provide examples in which the VG distribution appears as an exact distribution in connection to sample covariances, Wishart matrices and Matérn covariance functions. We review the well-known connection to the variance-gamma process (also referred to as Laplace motion; see for example, Kotz et al. [64]) and financial modelling in Sections 4.2 and 4.3, respectively. We provide an overview of the recent literature concerning application of the VG distribution to time series modelling in Section 4.4. In Section 4.5, we review some of the recent activity in the probability literature in

which the VG distribution has arisen as a limiting distribution.

Notation. To simplify formulae, we shall define $\lambda_{\pm} := (\sqrt{\theta^2 + \sigma^2} \pm \theta)/\sigma^2$.

2. DISTRIBUTIONAL PROPERTIES

2.1 Density and parametrisations

We begin by recalling some other parametrisations given in the literature. The first is given in Finlay and Seneta [31]. For $x \in \mathbb{R}$,

$$(2) \quad p(x) = \sqrt{\frac{2}{\pi}} \frac{\alpha^\alpha e^{\theta_0(x-\mu)/\sigma_0^2}}{\sigma_0 \Gamma(\alpha)} \left(\frac{|x-\mu|}{\sqrt{\theta_0^2 + 2\alpha\sigma_0^2}} \right)^{\alpha-1/2} \times K_{\alpha-1/2} \left(\frac{\sqrt{\theta_0^2 + 2\alpha\sigma_0^2}}{\sigma_0} |x-\mu| \right).$$

It is related to the parametrisation in (1) by $r = 2\alpha$, $\sigma^2 = \sigma_0^2/(2\alpha)$, $\theta = \theta_0/(2\alpha)$. If a random variable has PDF (2), we write $X \sim \text{VG}_2(\alpha, \theta_0, \sigma_0, \mu)$. Setting $\alpha = 1/\nu$ gives the parametrisation of Madan et al. [79], and further setting $\theta_0 = 0$ yields the parametrisation of the (symmetric) VG distribution of Madan and Seneta [83]. Another parametrisation is given in Bibby and Sørensen [15]:

$$(3) \quad p(x) = \frac{\gamma^{2\lambda}}{\sqrt{\pi} \Gamma(\lambda)} \left(\frac{|x-\mu|}{2\alpha} \right)^{\lambda-1/2} e^{\beta(x-\mu)} \times K_{\lambda-1/2}(\alpha|x-\mu|), \quad x \in \mathbb{R},$$

where $\gamma^2 = \alpha^2 - \beta^2$, and is related to the parametrisation in (1) by $r = 2\lambda$, $\theta = \beta/\gamma^2$, $\sigma = 1/\gamma$. Finally, we note the following parametrisation used by Kotz et al. [64]:

$$p(x) = \frac{\sqrt{2} e^{\frac{\sqrt{2}}{2\sigma_0}(1/\kappa - \kappa)(x-\mu)}}{\sigma_0^{\tau+1/2} \Gamma(\tau)} \left(\frac{\sqrt{2}|x-\mu|}{\kappa + 1/\kappa} \right)^{\tau-1/2} \times K_{\tau-1/2} \left(\frac{\sqrt{2}}{2\sigma_0} \left(\frac{1}{\kappa} - \kappa \right) |x-\mu| \right), \quad x \in \mathbb{R},$$

which is related to the parametrisation in (1) by $r = 2\tau$, $\theta = \sigma_0(1/\kappa - \kappa)/2^{3/2}$, $\sigma^2 = \sigma_0^2/2$. Henceforth, we shall work mostly with the parametrisation (1), with results in the other parametrisations being readily deduced using the above change of parameters.

The presence of the modified Bessel function in the PDF (1) makes it a little difficult to parse on first inspection. The following limiting forms can help in this regard. Suppose $r > 0$, $\theta \in \mathbb{R}$, $\sigma > 0$, $\mu \in \mathbb{R}$. Using the limiting form (65) gives that, as $x \rightarrow \mu$,

$$(4) \quad p(x) \sim \begin{cases} \frac{(1 + \theta^2/\sigma^2)^{-(r-1)/2}}{2\sigma\sqrt{\pi}} \frac{\Gamma((r-1)/2)}{\Gamma(r/2)}, & r > 1, \\ -\frac{1}{\pi\sigma} \log|x-\mu|, & r = 1, \\ \frac{1}{(2\sigma)^r \sqrt{\pi}} \frac{\Gamma((1-r)/2)}{\Gamma(r/2)} |x-\mu|^{r-1}, & r < 1 \end{cases}$$

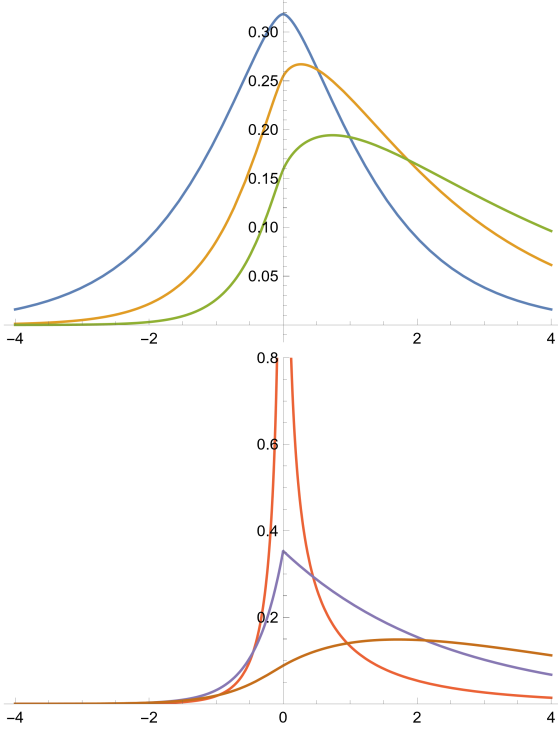


FIG 1. The VG PDF for different parameter constellations:
 ● VG(3, 0, 1, 0), ● VG(3, 0.5, 1, 0), ● VG(3, 1, 1, 0) (top image) and
 ● VG(0.5, 1, 1, 0), ● VG(2, 1, 1, 0), ● VG(4, 1, 1, 0) (bottom image).

(see Gaunt [35]). We see from (4) that the PDF has a singularity at the origin for $r \leq 1$. Moreover, for all parameter values ($r > 0$, $\theta \in \mathbb{R}$, $\sigma > 0$, $\mu \in \mathbb{R}$), the VG(r, θ, σ, μ) distribution is unimodal; further details are given in Section 2.8. The density is bounded for $r > 1$. Plots of the VG PDF (1) that show the effect of varying the shape parameter r (which agree with these assertions) and the skewness parameter θ are given in Figure 1. Also, applying the limiting form (66) to (1) gives that (see Gaunt [41])

$$(5) \quad p(x) \sim \frac{x^{r/2-1}}{2^{r/2}(\theta^2 + \sigma^2)^{r/4}\Gamma(r/2)} e^{-\lambda_-(x-\mu)}, \quad x \rightarrow \infty,$$

$$(6) \quad p(x) \sim \frac{(-x)^{r/2-1}}{2^{r/2}(\theta^2 + \sigma^2)^{r/4}\Gamma(r/2)} e^{\lambda_+(x-\mu)}, \quad x \rightarrow -\infty.$$

We observe that the tails of the VG distribution are heavier than the tails of the normal distribution. This feature is important in financial modelling and allows for a better fit to real financial data than the normal distribution; see Section 4.3 for further details.

When $r \in 2\mathbb{Z}^+ = \{2, 4, \dots\}$ is even, we can use a standard simplification of the modified Bessel function of the second kind (see (64)) to obtain a formula for the PDF in terms of elementary functions (see Kotz et al. [64], and earlier Teichroew [121] for the case $\theta = 0$): for $x \in \mathbb{R}$,

$$p(x) = \frac{|x - \mu|^{r/2-1}}{2^{r/2}(\theta^2 + \sigma^2)^{r/4}\Gamma(r/2)}$$

$$(7) \quad \times \exp\left(\frac{\theta}{\sigma^2}(x - \mu) - \frac{\sqrt{\theta^2 + \sigma^2}}{\sigma^2}|x - \mu|\right) \times \sum_{j=0}^{r/2-1} \frac{(r/2 - 1 + j)!}{(r/2 - 1 - j)!j!} \left(\frac{\sigma^2}{2\sqrt{\theta^2 + \sigma^2}|x - \mu|}\right)^j.$$

It is immediate from (1) that the class of VG distributions is closed under affine transformations (see Kotz et al. [64]). Let $X \sim \text{VG}(r, \theta, \sigma, \mu)$. Then, for $a \neq 0$ and $b \in \mathbb{R}$,

$$(8) \quad aX + b \sim \text{VG}(r, a\theta, |a|\sigma, a\mu + b).$$

To simplify expressions, we shall sometimes set $\mu = 0$, with results for the general case $\mu \in \mathbb{R}$ immediately following because $\mu + \text{VG}(r, \theta, \sigma, 0) =_d \text{VG}(r, \theta, \sigma, \mu)$ (with obvious abuse of notation). We also observe that if $X \sim \text{VG}(r, \theta, \sigma, \mu)$ then $-X \sim \text{VG}(r, -\theta, \sigma, -\mu)$, whilst if $X \sim \text{VG}(r, 0, \sigma, 0)$ then $-X =_d X$. In line with terminology of Kotz et al. [64], we say that $X \sim \text{VG}(r, 0, \sigma, 0)$ has a *symmetric variance-gamma* distribution.

2.2 Cumulative distribution function

Exact formulas for the CDF of the VG distribution, for general parameter values $r > 0$, $\theta \in \mathbb{R}$, $\sigma > 0$, $\mu \in \mathbb{R}$, as infinite series involving the modified Bessel function of the second kind and the modified Lommel function of the first kind were recently obtained by Gaunt [44]. Being rather complicated, we do not report these formulas here. Instead, we note some cases for which simple formulas can be given. For $X \sim \text{VG}(r, \theta, \sigma, \mu)$, let $F(x) = \mathbb{P}(X \leq x)$.

Suppose $\theta = 0$. Then, by the symmetry of the VG($r, 0, \sigma, \mu$) distribution about $x = \mu$, it follows that $F(x) = 1/2 + \text{sgn}(x) \int_0^{|x|} p(t) dt$, where $\text{sgn}(x)$ is the sign of x . Calculating the integral using (68), gives that, for $x \in \mathbb{R}$,

$$F(x) = \frac{1}{2} + \frac{(x - \mu)}{2\sigma} \left[K_{\frac{r-1}{2}}\left(\frac{|x - \mu|}{\sigma}\right) \mathbf{L}_{\frac{r-3}{2}}\left(\frac{|x - \mu|}{\sigma}\right) + \mathbf{L}_{\frac{r-1}{2}}\left(\frac{|x - \mu|}{\sigma}\right) K_{\frac{r-3}{2}}\left(\frac{|x - \mu|}{\sigma}\right) \right],$$

where $\mathbf{L}_\nu(x)$ is a modified Struve function of the first kind (see Olver et al. [99, Chapter 11] for a definition and basic properties). Other formulas for the CDF for the case $\theta = 0$ are given by Jankov Maširević and Pogány [59] and Nadarajah, Srivastava and Gupta [92].

Now suppose $r \in 2\mathbb{Z}^+$ and $\theta \in \mathbb{R}$. Then, making use of the formula (7) one readily obtains the following formulas (see Nadarajah et al. [92]). For $x \leq \mu$,

$$F(x) = \frac{(\theta^2 + \sigma^2)^{-r/4}}{(2\lambda_+)^{r/2}(r/2 - 1)!} \sum_{j=0}^{r/2-1} \frac{(r/2 - 1 + j)!}{(r/2 - 1 - j)!j!} \times \left(\frac{\theta + \sqrt{\theta^2 + \sigma^2}}{\sqrt{\theta^2 + \sigma^2}}\right)^j \Gamma\left(\frac{r}{2} - j, -\lambda_+(x - \mu)\right),$$

and, for $x > \mu$,

$$F(x) = 1 - \frac{(\theta^2 + \sigma^2)^{-r/4}}{(2\lambda_-)^{r/2}(r/2 - 1)!} \sum_{j=0}^{r/2-1} \frac{(r/2 - 1 + j)!}{(r/2 - 1 - j)!j!} \\ \times \left(\frac{\sqrt{\theta^2 + \sigma^2} - \theta}{\sqrt{\theta^2 + \sigma^2}} \right)^j \Gamma\left(\frac{r}{2} - j, \lambda_-(x - \mu)\right),$$

where $\Gamma(a, x) = \int_x^\infty t^{a-1} e^{-t} dt$ is the upper incomplete gamma function.

Let us now return to the general setting $r > 0$, $\theta \in \mathbb{R}$, $\sigma > 0$, $\mu \in \mathbb{R}$. Since a closed-form formula is not available for the CDF, asymptotic approximations for the tail probabilities are of interest. Let $\bar{F}(x) = 1 - F(x) = \mathbb{P}(X > x)$ for $X \sim \text{VG}(r, \theta, \sigma, \mu)$. Then, using (1) and the limiting form (70) gives that, as $x \rightarrow \infty$,

$$(9) \quad \bar{F}(x) \sim \frac{1}{2^{r/2}(\theta^2 + \sigma^2)^{r/4}\lambda_- \Gamma(r/2)} x^{r/2-1} e^{-\lambda_-(x-\mu)},$$

and, by symmetry, as $x \rightarrow -\infty$,

$$F(x) \sim \frac{1}{2^{r/2}(\theta^2 + \sigma^2)^{r/4}\lambda_+ \Gamma(r/2)} (-x)^{r/2-1} e^{\lambda_+(x-\mu)}.$$

Upper and lower bounds for $F(x)$ and $\bar{F}(x)$ can be obtained by using bounds of Gaunt [37, 39] for the integral $\int_x^\infty e^{\beta t} t^\nu K_\nu(t) dt$, $x > 0$, $-1 < \beta < 1$, $\nu > -1/2$. As an example, from inequality (2.9) of Gaunt [39] we have that $\int_x^\infty e^{\beta t} t^\nu K_\nu(\alpha t) dt \leq (\alpha - \beta)^{-1} e^{\beta x} x^\nu K_\nu(\alpha x)$, for $x > 0$, $0 \leq \beta < \alpha$, $-1/2 < \nu \leq 1/2$, with equality if and only if $\nu = 1/2$, and the inequality is reversed if $\nu > 1/2$. Applying this bound to (1) yields that, for $\theta \geq 0$, $x > \mu$,

$$(10) \quad \bar{F}(x) \leq \frac{p(x)}{\lambda_-},$$

with equality if and only if $r = 2$, and the inequality is reversed if $r > 2$. Here p is the PDF (1). Inequality (10) is tight as $x \rightarrow \infty$, which can be seen by applying the limiting form (66) to the bound in (10) and comparing to (9).

2.3 Generating functions, infinite divisibility and self-decomposability

The moment generating function of $X \sim \text{VG}(r, \theta, \sigma, \mu)$ is easily calculated using the integral formula (69) (see, for example, Bibby and Sørensen [15]),

$$(11) \quad M(t) = \mathbb{E}[e^{tX}] = e^{\mu t} (1 - 2\theta t - \sigma^2 t^2)^{-r/2},$$

which exists provided $-\lambda_+ < t < \lambda_-$ (see (5) and (6)). The characteristic function is

$$(12) \quad \varphi(t) = \mathbb{E}[e^{itX}] = e^{i\mu t} (1 - 2i\theta t + \sigma^2 t^2)^{-r/2}.$$

The cumulant generating function is defined for $-\lambda_+ < t < \lambda_-$, and given by

$$K(t) = \log \mathbb{E}[e^{tX}] = \mu t - \frac{r}{2} \log(1 - 2\theta t - \sigma^2 t^2) \\ (13) \quad = \mu t - \frac{r}{2} \log\left(1 - \frac{t}{\lambda_-}\right) - \frac{r}{2} \log\left(1 + \frac{t}{\lambda_+}\right).$$

From the characteristic function (12) formula, it is clear that if $X_1 \sim \text{VG}(r_1, \theta, \sigma, \mu_1)$ and $X_2 \sim \text{VG}(r_2, \theta, \sigma, \mu_2)$ are independent, then

$$(14) \quad X_1 + X_2 \sim \text{VG}(r_1 + r_2, \theta, \sigma, \mu_1 + \mu_2)$$

(see Bibby and Sørensen [15]). Thus, the class of VG distributions is closed under convolution, provided the random variables have common values of θ and σ .

Recall that a probability distribution F is said to be *infinitely divisible* if for every positive integer n there exist n i.i.d. random variables $X_{n,1}, \dots, X_{n,n}$ such that the sum $\sum_{i=1}^n X_{i,n}$ has the same distribution as F . Infinitely divisible distributions have the Lévy-Hinchin representation. It is clear from (14) that the $\text{VG}(r, \theta, \sigma, \mu)$ distribution is infinitely divisible. We emphasise that the fact that the VG distribution is infinitely divisible together with the formula (12) for the characteristic function are fundamental in understanding the VG Lévy process; for further details on the VG Lévy process see Section 4.3.

Moreover, the VG characteristic function has the following Lévy-Hinchin representation (see Madan et al. [79]):

$$(15) \quad \varphi(t) = \exp\left(i\mu t + \int_{-\infty}^{\infty} (e^{itx} - 1)\nu(x) dx\right),$$

where the Lévy density is given by

$$(16) \quad \nu(x) = \frac{r}{2|x|} e^{-\lambda_+|x|} \mathbf{1}_{x < 0} + \frac{r}{2x} e^{-\lambda_-x} \mathbf{1}_{x > 0}.$$

It is elementary to deduce the formula (16) for the VG Lévy density from the characteristic function (12). Indeed, we have

$$\varphi(t) = (1 - it/\lambda_-)^{-r/2} (1 + it/\lambda_+)^{-r/2}.$$

recalling that the characteristic function of the gamma $\Gamma(r, \lambda)$ distribution with PDF $p(x) = \lambda^r x^{r-1} e^{-\lambda x} / \Gamma(r)$, $x > 0$, is given by $\varphi(t) = (1 - it/\lambda)^{-r}$, we deduce (16) using the well-known Lévy-Hinchin representation of the gamma distribution (see, for example, Kyprianou [70, Section 1.2.4]).

A distribution \mathcal{Q} on the real line is *self-decomposable* (Sato [110, Definition 15.1, p. 90]) if for any constant $c \in (0, 1)$ there exists a probability space with a random variable $X \sim \mathcal{Q}$ and independent random variable Y such that $cX + Y \sim \mathcal{Q}$. Self-decomposable distributions are infinitely divisible and have a special form of Lévy-Hinchin representation. As the Lévy density $\nu(x)$, given by (16),

can be written in the form $\nu(x) = h(x)/|x|$, where $h(x) \geq 0$ is increasing for negative x and decreasing for positive x , it is immediate from the representation (15) (see Sato [110, Corollary 15.11]) that the $\text{VG}(r, \theta, \sigma, \mu)$ distribution is self-decomposable. An explicit construction of the self-decomposability of the $\text{VG}(r, \theta, \sigma, \mu)$ distribution is also available for the case that $r \in 2\mathbb{Z}^+$. Let $X \sim \text{VG}(2n, \theta, \sigma, \mu)$, where $n \in \mathbb{Z}^+$. Then, for any $c \in [0, 1]$,

$$(17) \quad X =_d cX + (1-c)\mu + \sum_{i=1}^n V_i,$$

where V_1, \dots, V_n are i.i.d. random variables that can be expressed as $V_1 =_d \sigma(\delta_1 W_1 / (\sigma \lambda_-) - \sigma \lambda_- \delta_2 W_2)$. Here, δ_1 and δ_2 are 0-1 random variables with probabilities

$$\mathbb{P}(\delta_1 = 0, \delta_2 = 0) = c^2, \quad \mathbb{P}(\delta_1 = 1, \delta_2 = 1) = 0,$$

$$\mathbb{P}(\delta_1 = 1, \delta_2 = 0) = (1-c) \left(c + \frac{1-c}{1 + \sigma^2 \lambda_-^2} \right),$$

$$\mathbb{P}(\delta_1 = 0, \delta_2 = 1) = (1-c) \left(c + \frac{(1-c)\sigma^2 \lambda_-^2}{1 + \sigma^2 \lambda_-^2} \right),$$

where W_1 and W_2 are exponential with rate parameter 1, and X, W_1, W_2 , and (δ_1, δ_2) are mutually independent. The representation (17) is given in Kotz et al. [64, Proposition 4.1.4], in which they generalised the representation of Ramachandran [104] for the asymmetric Laplace distribution by combining the convolution property (14) of the VG distribution together with the fact that the $\text{VG}(2, \theta, \sigma, \mu)$ distribution is an asymmetric Laplace distribution (see part 1 of Section 2.5.1).

2.4 Representation in terms of other random variables

1. The VG distribution has the following fundamental representation in terms of independent normal and gamma random variables; see Kotz et al. [64, Proposition 4.1.2] for a statement of the result and a short proof involving characteristic functions. To fix notation, consider the $\Gamma(r, \lambda)$ distribution with PDF $p(x) = \lambda^r x^{r-1} e^{-\lambda x} / \Gamma(r)$, $x > 0$. Suppose that $S \sim \Gamma(r/2, 1/2)$ and $T \sim N(0, 1)$ are independent random variables. Then

$$(18) \quad \mu + \theta S + \sigma \sqrt{ST} \sim \text{VG}(r, \theta, \sigma, \mu).$$

The VG distribution is therefore a univariate normal variance-mean mixture distribution (Barndorff-Nielsen, Kent and Sørensen [12]). Indeed, conditional on S , the random variable $\mu + \theta S + \sigma \sqrt{ST}$ has the $N(\mu + \theta S, \sigma^2 S)$ distribution.

2. When $r \geq 1$ is a positive integer, the VG distribution can be represented in terms of independent standard normal random variables X_1, \dots, X_r and Y_1, \dots, Y_r :

$$(19) \quad \mu + \theta \sum_{i=1}^r X_i^2 + \sigma \sum_{i=1}^r X_i Y_i \sim \text{VG}(r, \theta, \sigma, \mu).$$

(see Gaunt [36, Corollary 2.5]). To see this, first we define $Z_1 = \mu + \theta X_1^2 + \sigma X_1 Y_1$ and $Z_i = \theta X_i^2 + \sigma X_i Y_i$, $i = 2, \dots, r$. Observe that, for $i = 1, 2, \dots, r$, $X_i Y_i =_d |X_i| Y_i$ and that $X_i^2 \sim \Gamma(r/2, 1/2)$. Hence, by (18), we have that $Z_1 \sim \text{VG}_1(1, \theta, \sigma, \mu)$ and $Z_i \sim \text{VG}_1(1, \theta, \sigma, 0)$, $i = 2, \dots, r$. It therefore follows from (14) that $\sum_{i=1}^r Z_i \sim \text{VG}(r, \theta, \sigma, \mu)$.

As pointed out by the referee, another, perhaps more appealing proof of (19) from (18) is to note that, by conditioning on the X_i 's, we have $(\sum_{i=1}^r X_i^2, \sum_{i=1}^r X_i Y_i) =_d (S, \sqrt{S} Y_1)$, where $S := \sum_{i=1}^r X_i^2 \sim \Gamma(r/2, 1/2)$.

3. The VG distribution has a neat representation as a difference of independent gamma random variables (this was shown in the $\theta = 0$ case by McLeish [88]; for the general case see, for example, Press [103] or Kotz et al. [64]). Suppose $S \sim \Gamma(r/2, (\sqrt{\theta^2 + \sigma^2} + \theta)^{-1})$ and $S' \sim \Gamma(r/2, (\sqrt{\theta^2 + \sigma^2} - \theta)^{-1})$ are independent. Then

$$(20) \quad \mu + S - S' \sim \text{VG}(r, \theta, \sigma, \mu).$$

This representation is efficiently proved using a standard characteristic function argument with the $\text{VG}(r, \theta, \sigma, \mu)$ characteristic function formula (12) and the formula for the characteristic function of the $\Gamma(r, \lambda)$ distribution, $\varphi(t) = (1 - it/\lambda)^{-r}$.

Suppose now that $r \geq 1$ is an integer. Then $S =_d (\sqrt{\theta^2 + \sigma^2} + \theta)V/2$ and $S' =_d (\sqrt{\theta^2 + \sigma^2} - \theta)V'/2$, where V and V' are independent χ_r^2 random variables (chi-square distribution with r degrees of freedom). From the representation of the χ_r^2 distribution as the sum of the squares of r independent $N(0, 1)$ random variables and the representation (20), we deduce that

$$\mu + r\theta + \sum_{j=1}^{2r} \lambda_j (N_j^2 - 1) \sim \text{VG}(r, \theta, \sigma, \mu),$$

where N_1, \dots, N_{2r} are independent $N(0, 1)$ random variables, $\lambda_1 = \dots = \lambda_r = (\sqrt{\theta^2 + \sigma^2} + \theta)/2$ and $\lambda_{r+1} = \dots = \lambda_{2r} = -(\sqrt{\theta^2 + \sigma^2} - \theta)/2$. Thus, as noted by Azmoodeh, Eichelsbacher and Thäle [7, Corollary 1.2], for integer $r \geq 1$ the $\text{VG}(r, \theta, \sigma, -r\theta)$ distribution is a member of the second Wiener chaos (see Nourdin and Peccati [95, Section 2.2]).

4. If $r \in 2\mathbb{Z}^+$, the independent random variables $V \sim \chi_r^2$ and $V' \sim \chi_r^2$ can be represented in terms of independent uniform $U(0, 1)$ random variables U_1, \dots, U_r by $V =_d \sum_{j=1}^{r/2} \log U_j$ and $V' =_d \sum_{j=r/2+1}^r \log U_j$. Thus, from (20),

$$\begin{aligned} \mu + \frac{\sqrt{\theta^2 + \sigma^2} + \theta}{2} \sum_{j=1}^{r/2} \log U_j \\ - \frac{\sqrt{\theta^2 + \sigma^2} - \theta}{2} \sum_{j=r/2+1}^r \log U_j \sim \text{VG}(r, \theta, \sigma, \mu). \end{aligned}$$

This representation is convenient for simulating the VG distribution when $r \in 2\mathbb{Z}^+$. For other values of $r > 0$, the $\text{VG}(r, \theta, \sigma, \mu)$ distribution can be simulated using the representation (20) and simulating the gamma distributions of V and V' using the methods given in Chapter 9, Section 3 of Devroye [24].

2.5 Related distributions

2.5.1 Subclasses and limiting cases 1. The asymmetric Laplace distribution corresponds to the $\text{VG}(2, \theta, \sigma, \mu)$ distribution. Indeed, setting $r = 2$ in (7) yields (21)

$$p(x) = \frac{1}{2\sqrt{\theta^2 + \sigma^2}} \exp\left(\frac{\theta}{\sigma^2}(x - \mu) - \frac{\sqrt{\theta^2 + \sigma^2}}{\sigma^2}|x - \mu|\right).$$

Setting $\theta = 0$ in (21) gives the PDF of the classical Laplace distribution. A comprehensive account of the distributional theory of the Laplace and asymmetric Laplace distributions is given in Chapters 2 and 3 of Kotz et al. [64].

2. The gamma distribution is a limiting case of the VG distribution (see, for example, Kotz et al. [64]). For fixed, $r, \lambda > 0$, the sequence of random variables $X_\sigma \sim \text{VG}(2r, (2\lambda)^{-1}, \sigma, 0)$ converges in distribution to a $\Gamma(r, \lambda)$ random variable, as $\sigma \rightarrow 0$. This can be seen by letting $\sigma \rightarrow 0$ in the formula (12) for the characteristic function of VG distribution and comparing to the gamma characteristic function using Lévy's continuity theorem.

3. The normal distribution is also a limiting case of the VG distribution (see, for example, Kotz et al. [64]). For fixed, $\theta \in \mathbb{R}$ and $\sigma > 0$, the sequence of random variables $Y_r \sim \text{VG}(r, \theta/\sqrt{r}, \sigma/\sqrt{r}, -\theta\sqrt{r})$ converges to a $N(0, \sigma^2 + 2\theta^2)$ random variable, as $r \rightarrow \infty$. As in part 2, this can be deduced from a characteristic function argument. More insightfully, from the convolution property, for $r \in \mathbb{Z}^+$, we can write $Y_r =_d r^{-1/2} \sum_{i=1}^r Y_{r,i}$, where $Y_{r,i} \sim \text{VG}(1, \theta, \sigma, -\theta)$. The result now follows from the central limit theorem and the formulas $\mathbb{E}[Y_{r,i}] = 0$ and $\text{Var}(Y_{r,i}) = \sigma^2 + 2\theta^2$ (see (30) and (31)). We also note that the result is a trivial consequence of relation (19) and the central limit theorem.

4. Let (X, Y) be a bivariate normal random vector with zero mean vector, variances (σ_X^2, σ_Y^2) and correlation coefficient ρ . Denote the product of these correlated normal random variables by $Z = XY$. We will denote this distribution by $\text{PN}(\rho, \sigma_X, \sigma_Y)$. Consider also the mean $\bar{Z}_n = n^{-1}(Z_1 + \dots + Z_n)$, where Z_1, \dots, Z_n are independent copies of Z . It was noted in the thesis of Gaunt [35] that $Z \sim \text{VG}(1, \rho\sigma_X\sigma_Y, \sigma_X\sigma_Y\sqrt{1-\rho^2}, 0)$ and later by Gaunt [40] that also

$$(22) \quad \bar{Z}_n \sim \text{VG}(n, \rho\sigma_X\sigma_Y/n, \sigma_X\sigma_Y\sqrt{1-\rho^2}/n, 0).$$

To see this, suppose $\sigma_X = \sigma_Y = 1$, with the general case following by rescaling. One can verify that X and $W =$

$(Y - \rho X)/\sqrt{1-\rho^2}$ are independent $N(0, 1)$ random variables. Thus, $Z = XY = X(\sqrt{1-\rho^2}W + \rho X) = \sqrt{1-\rho^2}XW + \rho X^2$, so $Z \sim \text{VG}(1, \rho, \sqrt{1-\rho^2}, 0)$ by (19). Finally, we deduce (22) from properties (8) and (14).

With (22), Gaunt [40] proved that the PDF of \bar{Z}_n (previously given independently by Mangilli, Plaszczyński and Tristram [85] and Nadarajah and Pogány [91]) is given by

$$p(x) = \frac{n^{(n+1)/2} 2^{(1-n)/2} |x|^{(n-1)/2}}{s^{(n+1)/2} \sqrt{\pi(1-\rho^2)} \Gamma(n/2)} \exp\left(\frac{\rho n x}{s(1-\rho^2)}\right) \times K_{\frac{n-1}{2}}\left(\frac{n|x|}{s(1-\rho^2)}\right), \quad x \in \mathbb{R},$$

where $s = \sigma_X \sigma_Y$, with the PDF for Z following on setting $n = 1$. For an account of the distributional theory of Z and \bar{Z} , see Gaunt [42]. We note that the product of two normal random variables with non-zero means does not follow the VG distribution; the rather complicated exact formula for the PDF is given in Cui et al. [22].

2.5.2 Superclasses 1. The generalized hyperbolic (GH) distribution was introduced by Barndorff-Nielsen [10, 11], who studied it in the context of modelling dune movements. Like the VG distribution, GH distributions are widely used in financial modelling; see for example, Bibby and Sørensen [15], Eberlein and Keller [26] and Eberlein and Prause [27]. Properties of the GH distribution are given in Bibby and Sørensen [15] and Hammerstein [54]. For $x \in \mathbb{R}$, the PDF is given by (23)

$$p(x) = \frac{(\gamma/\delta)^\lambda}{\sqrt{2\pi} K_\lambda(\delta\gamma)} e^{\beta(x-\mu)} \frac{K_{\lambda-1/2}(\alpha\sqrt{\delta^2 + (x-\mu)^2})}{(\sqrt{\delta^2 + (x-\mu)^2}/\alpha)^{1/2-\lambda}},$$

where $\gamma = \sqrt{\alpha^2 - \beta^2}$. The parameter domain is given by

$$\delta \geq 0, \quad \gamma > 0, \quad \lambda > 0,$$

$$\delta > 0, \quad \gamma > 0, \quad \lambda = 0,$$

$$\delta > 0, \quad \gamma \geq 0, \quad \lambda < 0,$$

and in each case $\mu \in \mathbb{R}$. If $\delta = 0$ or $\gamma = 0$, the PDF (23) is defined as the limit obtained by using (65). In particular, taking $\delta \downarrow 0$ in (23) yields the density (3) of the VG distribution in the parametrisation of Bibby and Sørensen [15]. A detailed study of limiting cases of GH distributions is given by Eberlein and Hammerstein [25].

2. The VG distribution is a special case of the CGMY distribution of Carr et al. [20], which was introduced in the context of financial modelling. The distribution is defined by its Lévy-Hinchin representation $\varphi(t) = \exp\left(\int_{\mathbb{R}} (e^{itx} - 1) \nu_{\text{CGMY}}(x) dx\right)$, where the Lévy density is given by

$$\nu_{\text{CGMY}}(x) = \frac{C}{|x|^{1+Y}} e^{-G|x|} \mathbf{1}_{x < 0} + \frac{C}{x^{1+Y}} e^{-Mx} \mathbf{1}_{x > 0}.$$

The VG distribution corresponds to the case $Y = 0$. A closed-form formula is not available for the PDF of the CGMY distribution, although the characteristic function takes an elementary form (see Carr et al. [20]):

$$\varphi(t) = \exp \left(C\Gamma(-Y) \{ (M - it)^Y - M^Y + (G - it)^Y - G^Y \} \right),$$

from which one readily obtains formulas for lower order moments, as well as the variance, skewness and kurtosis (again, see Carr et al. [20]).

Being a subclass of the CGMY distributions, the VG distributions form a subclass of the tempered stable distributions that was introduced by Koponen [62]. For properties and references for this class of distributions see, for example, K uchler and Tappe [69].

3. Consider the distribution of the random variable $X = Z + L$, with independent $Z \sim N(0, \sigma^2)$ and L is asymmetric Laplace. This distribution, introduced by Reed [106], called the *normal-asymmetric Laplace*, has four parameters. Some basic distributional properties are given in Reed [106], such as infinite divisibility, exact formulas for the PDF, CDF, moment and cumulants, and estimation methods are also discussed.

More generally, if L is VG distributed (generalized asymmetric Laplace), then X is said to be *generalized normal-Laplace* (GNL) distributed, introduced in Reed [107]. This distribution has five parameters. The article of Reed [107] includes financial applications. In the thesis of Wu [127], one can find a systematic treatment and bibliography for GNL distributions.

4. A multivariate generalisation of the VG distribution in the symmetric case was first proposed by Madan and Seneta [83]. A number of basic properties of the (general non-symmetric) multivariate VG distribution are reviewed by Kozubowski, Podg orski and Rychlik [68] (they refer to it as the multivariate generalized Laplace distribution). For an account of the matrix variate VG distribution, see Kozubowski, Mazur and Podg orski [65] and references therein. Here, we collect just a few of the most fundamental properties of the multivariate VG distribution, noting that they generalise properties of the univariate VG distribution in a natural manner.

The multivariate VG distribution is often defined through its characteristic function. In line with Kozubowski et al. [68], we say that a random vector \mathbf{X} in \mathbb{R}^d follows the multivariate VG distribution if, for $\mathbf{t} \in \mathbb{R}^d$, its characteristic function is given by

$$(24) \quad \varphi(\mathbf{t}) = \mathbb{E}[e^{i\mathbf{t}^\top \mathbf{X}}] = e^{i\boldsymbol{\mu}^\top \mathbf{t}} \left(1 - i\boldsymbol{\theta}^\top \mathbf{t} + \frac{1}{2} \mathbf{t}^\top \boldsymbol{\Sigma} \mathbf{t} \right)^{-s},$$

where $s > 0$, $\boldsymbol{\theta} \in \mathbb{R}^d$, $\boldsymbol{\mu} \in \mathbb{R}^d$, and $\boldsymbol{\Sigma}$ is $d \times d$ non-negative definite symmetric matrix. Note that when $d = 1$, the characteristic function (24) reduces to the characteristic

function (12) of the univariate VG distribution (under a slightly different parametrisation). We can represent the multivariate VG distribution as a normal variance-mean mixture

$$\mathbf{X} =_d \boldsymbol{\mu} + \boldsymbol{\theta}G + \sqrt{G}\mathbf{Z},$$

for independent $G \sim \Gamma(s, 1)$ and $\mathbf{Z} \sim N_d(\mathbf{0}, \boldsymbol{\Sigma})$ (the d -dimensional multivariate normal distribution with mean vector $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}$).

If the matrix $\boldsymbol{\Sigma}$ is positive-definite, then the distribution is non-degenerate, that is truly d -dimensional, with PDF

$$p(\mathbf{x}) = \frac{2 \exp(\boldsymbol{\theta}^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}))}{(2\pi)^{d/2} \Gamma(s) |\boldsymbol{\Sigma}|^{1/2}} \left(\frac{Q(\mathbf{x}, \boldsymbol{\mu})}{C(\boldsymbol{\Sigma}, \boldsymbol{\theta})} \right)^{s-d/2} \times K_{s-d/2}(Q(\mathbf{x}, \boldsymbol{\mu})C(\boldsymbol{\Sigma}, \boldsymbol{\theta})), \quad \mathbf{x} \in \mathbb{R}^d,$$

where $Q(\mathbf{x}, \boldsymbol{\mu}) = ((\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}))^{1/2}$ and $C(\boldsymbol{\Sigma}, \boldsymbol{\theta}) = (2 + \boldsymbol{\theta}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\theta})^{1/2}$.

2.5.3 Products and quotients of VG random variables

Let $(X_i)_{1 \leq i \leq n} \sim \text{VG}(r_i, 0, \sigma_i, 0)$ be independent, and set $Z_n = \prod_{i=1}^n X_i$ and $\sigma = \sigma_1 \cdots \sigma_n$. A formula for the PDF of the product Z_n was obtained by Gaunt, Mijoule and Swan [49]: for $x \in \mathbb{R}$,

$$f_{Z_n}(x) = \frac{1}{2^n \pi^{n/2} \sigma} \left\{ \prod_{j=1}^n \frac{1}{\Gamma(r_j/2)} \right\} \times G_{0,2n}^{2n,0} \left(\frac{x^2}{2^{2n} \sigma^2} \left| \frac{r_1 - 1}{2}, \dots, \frac{r_n - 1}{2}, 0, \dots, 0 \right. \right),$$

where $G_{0,k}^{k,0}(x|)$ is a Meijer G -function (see Chapter 16 of Olver et al. [99] for a definition and basic properties). An exact formula for the product of two independent $\text{VG}(r, \theta, \sigma, 0)$ random variables with $\theta \neq 0$ has recently been obtained by Gaunt and Li [47].

Suppose $X \sim \text{VG}(r, \theta_1, \sigma_1, 0)$ and $Y \sim \text{VG}(s, \theta_2, \sigma_2, 0)$ are independent, where $r, s > 0$, $\theta_i \in \mathbb{R}$, $\sigma_i > 0$, $i = 1, 2$. An exact formula for the PDF of ratio $Z = X/Y$ in terms of an infinite series involving Gaussian hypergeometric functions (See Chapter 15 of Olver et al. [99] for a definition and basic properties) was obtained by Gaunt and Li [46]. Previously, the following simpler exact formula, expressed as a single Gaussian hypergeometric function, for the case $\theta_1 = \theta_2 = 0$ and $\sigma_1 = \sigma_2 = 1$ was obtained by Nadarajah and Kotz [90]. For $x \in \mathbb{R}$,

$$f_Z(x) = \frac{2|x|^{-s-1} \Gamma((r+1)/2) \Gamma((s+1)/2)}{\pi(r+s) \Gamma(r/2) \Gamma(s/2)} \times {}_2F_1 \left(\frac{r+s}{2}, \frac{s+1}{2}; \frac{r+s}{2} + 1; 1 - x^{-2} \right).$$

We refer the reader to the two aforementioned works for further properties of the VG ratio distribution, such as formulas for the PDF expressed in terms of elementary functions for particular parameter values, a formula for

the CDF, fractional moments, and asymptotic behaviour of the PDF and tail probabilities, from which it is seen that the mean does not exist.

2.6 Stein characterisations

The following Stein characterisation of the VG distribution was given by Gaunt [36]. Let W be a real-valued random variable. Then $W \sim \text{VG}(r, \theta, \sigma, \mu)$ if and only if

$$(25) \quad \mathbb{E}[\sigma^2(W - \mu)g''(W) + (\sigma^2 r + 2\theta(W - \mu)) \times g'(W) + (r\theta - (W - \mu))g(W)] = 0$$

for all twice differentiable $g: \mathbb{R} \rightarrow \mathbb{R}$ for which the expectations $\mathbb{E}|g(X)|$, $\mathbb{E}|Xg(X)|$, $\mathbb{E}|g'(X)|$, $\mathbb{E}|Xg'(X)|$ and $\mathbb{E}|Xg''(X)|$ are finite for $X \sim \text{VG}(r, \theta, \sigma, \mu)$.

Here we provide a sketch of the argument. We start with necessity. The PDF (1) of the $\text{VG}(r, \theta, \sigma, \mu)$ distribution satisfies the ordinary differential equation (ODE)

$$(26) \quad \mathcal{L}p(x) := \sigma^2 x p''(x) - (\sigma^2(r - 2) + 2\theta x)p'(x) + (\theta(r - 2) - x)p(x) = 0,$$

which can be read off from a more general ODE satisfied by the PDF of the GH distribution given in Gaunt [38, Corollary 3.3]. From (26), $\int_{-\infty}^{\infty} g(x)\mathcal{L}p(x) dx = 0$ and integrating by parts twice yields (25). To prove sufficiency, suppose W is a real-valued random variable. For ease of exposition, we set $\mu = 0$, with the general case following by a simple translation. Taking $g(x) = e^{itx}$ (a twice differentiable and bounded function) in (25) and letting $\varphi(t) = \mathbb{E}[e^{itW}]$ leads to the ODE

$$(27) \quad (\sigma^2 t^2 - 2i\theta t + 1)\varphi'(t) + (i\sigma^2 r t + r\theta)\varphi(t) = 0.$$

Here we applied the characterising equation to the real and imaginary parts of $g(x) = e^{itx}$. Solving (27) subject to the condition $\varphi(0) = 1$ gives that $\varphi(t) = (1 - 2i\theta t + \sigma^2 t^2)^{-r/2}$, in agreement with the characteristic function (12) when $\mu = 0$, and hence $W \sim \text{VG}(r, \theta, \sigma, 0)$.

Stein characterisations are most commonly used in Stein's method (Stein [118]) to bound the distance between two probability distributions with respect to a probability metric; however, one can also use them to establish distributional properties. Indeed, letting $Y \sim \text{VG}(r, \theta, \sigma, 0)$ and substituting $g_1(x) = x^k$ and $g_2(x) = (x - \mathbb{E}[Y])^k = (x - r\theta)^k$ into (25) (with $\mu = 0$) leads to the following recursions for the k -th raw moment $\mu'_k = \mathbb{E}[Y^k]$ and the k -th central moment $\mu_k = \mathbb{E}[(Y - \mathbb{E}[Y])^k]$:

$$(28) \quad \begin{aligned} \mu'_{k+1} &= \theta(2k + r)\mu'_k \\ &+ \sigma^2 k(r + k - 1)\mu'_{k-1}, \quad k \geq 1, \end{aligned}$$

$$(29) \quad \begin{aligned} \mu_{k+1} &= 2k\theta\mu_k + k(\sigma^2(r + k - 1) + 2\theta^2 r)\mu_{k-1} \\ &+ k(k - 1)r\theta\sigma^2\mu_{k-2}, \quad k \geq 2. \end{aligned}$$

Lower order raw (central) moments can be efficiently computed using the recurrences given just the first raw (first and second central) moments, along with the basic condition $\mu_0 = 1$.

2.7 Moments and cumulants

The mean and variance of $X \sim \text{VG}(r, \theta, \sigma, \mu)$ are readily calculated via the standard method of obtaining lower order moments from moment generating functions (using (11)). We have

$$(30) \quad \mathbb{E}[X] = \mu + r\theta,$$

$$(31) \quad \text{Var}(X) = r(\sigma^2 + 2\theta^2).$$

These formulas are also straightforward consequences of (18). Now, let $Y \sim \text{VG}(r, \theta, \sigma, 0)$. Lower order raw and central moments of Y can be efficiently obtained from the recurrences (28) and (29). The central moments of $X \sim \text{VG}(r, \theta, \sigma, \mu)$ are equal to the central moments of Y , and raw moments for X are readily obtained through the formula $\mathbb{E}[X^k] = \sum_{j=0}^k \binom{k}{j} \mu^j \mathbb{E}[Y^{k-j}]$. The first four raw moments of $Y \sim \text{VG}(r, \theta, \sigma, 0)$ are

$$\mu'_1 = r\theta, \quad \mu'_2 = r(\sigma^2 + (r + 2)\theta^2),$$

$$\mu'_3 = r(r + 2)\theta(3\sigma^2 + (r + 4)\theta^2),$$

$$\mu'_4 = r(r + 2)(3\sigma^4 + 6(r + 4)\theta^2\sigma^2 + (r + 4)(r + 6)\theta^4),$$

whilst the first four central moments are

$$(32) \quad \mu_1 = 0,$$

$$(33) \quad \mu_2 = r(\sigma^2 + 2\theta^2),$$

$$(34) \quad \mu_3 = 2r\theta(3\sigma^2 + 4\theta^2),$$

$$(35) \quad \begin{aligned} \mu_4 &= 3r((r + 2)\sigma^4 + (4r + 16)\theta^2\sigma^2 \\ &+ (4r + 16)\theta^4). \end{aligned}$$

The skewness $\gamma_1 = \mu_3/\mu_2^{3/2}$ and kurtosis $\beta_2 = \mu_4/\mu_2^2$ are thus given by

$$\gamma_1 = \frac{2\theta(3\sigma^2 + 4\theta^2)}{\sqrt{r}(\sigma^2 + 2\theta^2)^{3/2}},$$

$$\beta_2 = \frac{3((r + 2)\sigma^4 + (4r + 16)\theta^2\sigma^2 + (4r + 16)\theta^4)}{r(\sigma^2 + 2\theta^2)^2},$$

and the excess kurtosis $\gamma_2 = \mu_4/\mu_2^2 - 3$ is equal to

$$\gamma_2 = \frac{6(\sigma^4 + 8\theta^2\sigma^2 + 8\theta^4)}{r(\sigma^2 + 2\theta^2)^2}.$$

Formulas for the first four central moments of the VG distribution, as well as the skewness and kurtosis, were given by Seneta [115].

Higher order moments can be calculated using the representation (20) (with $\mu = 0$) of the $\text{VG}(r, \theta, \sigma, 0)$ distribution and the formula $\mathbb{E}[V^k] = \lambda^{-k}\Gamma(r/2 + k)/\Gamma(r/2)$, $k \geq 1$, where $V \sim \Gamma(r/2, \lambda)$. For $k \geq 1$,

$$\begin{aligned} \mu'_k &= \frac{\sigma^{2k}}{(\Gamma(r/2))^2} \sum_{j=0}^k \binom{k}{j} (-\lambda_-)^j \lambda_+^{k-j} \Gamma\left(\frac{r}{2} + j\right) \\ &\times \Gamma\left(\frac{r}{2} + k - j\right). \end{aligned}$$

We now return to the general case $r > 0$, for which the following expressions, involving the hypergeometric function, for the raw and absolute moments of $Y \sim \text{VG}(r, \theta, \sigma, 0)$ were obtained by Gaunt [43]. Let $k \in \mathbb{Z}^+$, and define $\ell := \lceil k/2 \rceil + 1/2$ and $m := k \bmod 2$. Then

$$\mu'_k = \frac{2^{k+m} \theta^m \sigma^{r+2k}}{\sqrt{\pi} (\theta^2 + \sigma^2)^{(r+k+m)/2} \Gamma(r/2)} \Gamma\left(\frac{r-1}{2} + \ell\right) \Gamma(\ell) \\ \times {}_2F_1\left(\ell, \frac{r-1}{2} + \ell; \frac{1}{2} + m; \frac{\theta^2}{\theta^2 + \sigma^2}\right),$$

whilst, for $k > k_* = k_*(r) := \max\{-1, -r\}$,

$$\mathbb{E}[|Y|^k] = \frac{2^k \sigma^{r+2k}}{\sqrt{\pi} (\theta^2 + \sigma^2)^{(r+k)/2} \Gamma(r/2)} \Gamma\left(\frac{r+k}{2}\right) \\ \times \Gamma\left(\frac{k+1}{2}\right) {}_2F_1\left(\frac{k+1}{2}, \frac{r+k}{2}; \frac{1}{2}; \frac{\theta^2}{\theta^2 + \sigma^2}\right).$$

As ${}_2F_1(a, b; c; 0) = 1$, we observe that in the case $\theta = 0$ the raw moments simplify to

$$\mu'_k = \begin{cases} \frac{2^k \sigma^k \Gamma((r+k)/2) \Gamma((k+1)/2)}{\sqrt{\pi} \Gamma(r/2)}, & \text{if } k \text{ is even,} \\ 0, & \text{if } k \text{ is odd.} \end{cases}$$

Central moments of general order of the $\text{VG}(r, \theta, \sigma, 0)$ distribution can be calculated using the representation (20) and the formula $\mathbb{E}[(V - \mathbb{E}[V])^k] = \lambda^{-k} U(-k, 1 - k - r/2, -r/2)$, where $V \sim \Gamma(r/2, \lambda)$ and $U(a, b, x)$ is a confluent hypergeometric function of the second kind. This formula is given for the case that $V \sim \chi_r^2$ by Weisstein [126], and the generalisation to the gamma distribution is obvious. We observe that $U(a, b, x)$ is a polynomial when $a = -m$, $m = 0, 1, 2, \dots$: $U(-m, b, x) = (-1)^m \sum_{j=0}^m \binom{m}{j} (b+j)_{m-j} (-x)^j$. Here the Pochhammer symbol is given by $(a)_0 = 0$ and $(a)_j = a(a+1)(a+2) \cdots (a+j-1)$, $j \geq 1$ (see Olver et al. [99, Section 13.2(i)]). For $k \geq 1$,

$$\mu_k = \sigma^{2k} \sum_{j=0}^k \binom{k}{j} (-\lambda_-)^j \lambda_+^{k-j} U\left(-j, 1-j-\frac{r}{2}, -\frac{r}{2}\right) \\ \times U\left(j-k, 1+j-k-\frac{r}{2}, -\frac{r}{2}\right).$$

Other formulas for the raw and central moments of the VG distribution are given by Scott [112].

The cumulants of the $\text{VG}(r, \theta, \sigma, \mu)$ distribution can be derived by Taylor expanding the logarithms in (13) or using the representation (20). The latter approach involves using the facts that, for independent random variables S_1 and S_2 and constant c , the k -th cumulant enjoys the properties $\kappa_k(cS_1) = c^k \kappa_k(S_1)$ and $\kappa_k(S_1 + S_2) = \kappa_k(S_1) + \kappa_k(S_2)$, and $\kappa_k(V) = \lambda^{-k} (k-1)! r$ for $V \sim \Gamma(r/2, \lambda)$.

For $k \geq 1$,

$$\kappa_k = \mu \delta_{k,1} + \frac{(k-1)! r}{2} \left[(\theta + \sqrt{\theta^2 + \sigma^2})^k \right. \\ \left. + (\theta - \sqrt{\theta^2 + \sigma^2})^k \right],$$

where $\delta_{i,j}$ is the Kronecker delta (see McKay [87]). In particular,

$$\kappa_1 = \mu + r\theta, \quad \kappa_2 = r(\sigma^2 + 2\theta^2), \quad \kappa_3 = 2r\theta(3\sigma^2 + 4\theta^2), \\ \kappa_4 = 6r(\sigma^4 + 8\theta^2\sigma^2 + 8\theta^4), \\ \kappa_5 = 24r\theta(5\sigma^4 + 20\theta^2\sigma^2 + 16\theta^4), \\ \kappa_6 = 120r(\sigma^6 + 18\theta^2\sigma^4 + 48\theta^4\sigma^2 + 32\theta^6).$$

2.8 Mode and median

A detailed study of the mode and median of the GH and VG distributions was undertaken by Gaunt and Merkle [48]; here we provide a summary for the VG distribution.

The $\text{VG}(r, \theta, \sigma, \mu)$ distribution is unimodal. This follows since the distribution is self-decomposable and self-decomposable distributions are unimodal (Yamazato [128]). Denote the mode of the $\text{VG}(r, \theta, \sigma, \mu)$ distribution by M_r . For $0 < r \leq 2$, $\theta \in \mathbb{R}$, $\sigma > 0$, or $r > 0$, $\theta = 0$, $\sigma > 0$, we have $M_r = \mu$. This is clear when $\theta = 0$, because the $\text{VG}(r, 0, \sigma, \mu)$ distribution is symmetric about μ ; for $0 < r \leq 2$ and $\theta \in \mathbb{R}$ this follows because, for $0 < \nu \leq 1/2$ and $|\beta| < 1$, the function $x \mapsto e^{\beta x} |x|^\nu K_\nu(|x|)$ is increasing on $(-\infty, 0)$ and decreasing on $(0, \infty)$ (this is easily inferred from part (i) of Lemma 5.1 of Gaunt [41]).

Suppose now that $r > 2$, $\theta \in \mathbb{R}$, $\sigma > 0$, $\mu \in \mathbb{R}$. Using (67), we obtain that $M_r = \mu + \text{sgn}(\theta) \cdot x^*$, where x^* is the unique positive solution of the equation (36)

$$K_{\frac{r-3}{2}}\left(\frac{\sqrt{\theta^2 + \sigma^2}}{\sigma^2} x\right) = \frac{|\theta|}{\sqrt{\theta^2 + \sigma^2}} K_{\frac{r-1}{2}}\left(\frac{\sqrt{\theta^2 + \sigma^2}}{\sigma^2} x\right).$$

When $r = 4$ and $r = 6$, applying (64) to (36) leads to simple algebraic equations for x^* which when solved give

$$M_4 = \mu + \theta \left(1 + \frac{1}{\sqrt{1+\kappa}}\right), \quad M_6 = \mu + \frac{\theta}{2} \times \\ \times \left(1 + \frac{1}{\sqrt{1+\kappa}}\right) \left(3 - \sqrt{1+\kappa} + \sqrt{6\sqrt{1+\kappa} + \kappa - 2}\right),$$

where $\kappa = \sigma^2/\theta^2$. For $r = 8$ and $r = 10$, we can apply (64) to (36) to obtain cubic and quartic equations for x^* ; however, their solutions M_8 and M_{10} are too complicated to be worth reporting. For other values of r , exact formulas for M_r are not available, except for the case $\theta = 0$ in which we have already noted that the mode is equal to μ .

Exact formulas for M_r are not available for general $r > 2$; however, simple and accurate lower and upper bounds can be obtained. Fix $\theta > 0$; bounds for $\theta < 0$ follow immediately as if $X \sim \text{VG}(r, \theta, \sigma, \mu)$ then $-X \sim$

$\text{VG}(r, -\theta, \sigma, -\mu)$ (see (8)). Applying the lower and upper bounds of (71) to (36) yields simple algebraic inequalities for x^* . Solving these inequalities and using that $M_r > \mu$, for $r > 2$ and $\theta > 0$, leads to the following double inequality:

$$(37) \quad \theta(r-3)_+ < M_r - \mu < \theta(r-2), \quad r > 2,$$

where $x_+ = \max\{0, x\}$. Note that there is equality in the upper bound when $r = 2$. Applying inequality (72) to equation (36) gives that, for $r > 4$,

$$(38) \quad M_r > \mu + \frac{\theta}{2} \left[r - 2 + \sqrt{\frac{\theta^2(r-2)^2 + \sigma^2(r-4)^2}{\theta^2 + \sigma^2}} \right],$$

with equality if and only if $r = 4$, and the inequality is reversed if $2 < r < 4$. Inequality (38) improves on the lower bound in (37) in its range of validity $r > 4$. The reversed inequality (38) is more accurate than the upper bound in (37) for $3 < r < 4$, but the reverse is true for $2 < r < 3$.

Let $X \sim \text{VG}(r, \theta, \sigma, \mu)$. We note that exact formulas for the median are available for certain parameter values: $\text{Med}(X) = \mu$ if $\theta = 0$, whilst, for $r = 2$,

$$(39) \quad \begin{aligned} \text{Med}(X) &= \mu + \text{sgn}(\theta) \cdot (|\theta| + \sqrt{\theta^2 + \sigma^2}) \\ &\quad \times \log \left(1 + \frac{|\theta|}{\sqrt{\theta^2 + \sigma^2}} \right), \end{aligned}$$

which follows from the formula for the median of the asymmetric Laplace distribution (see Kozubowski and Podgórski [66]). For other parameter values, an exact closed-form formula is not available. Moreover, in contrast to the mode, accurate lower and upper bounds for the median are not available in the literature. Gaunt and Merkle [48] have conjectured accurate bounds for the median, which we present below. The numerical results of Table 1, which are taken from Gaunt and Merkle [48], support the conjectured bounds, since each entry in the table lies between the conjectured lower and upper bounds.

CONJECTURE 2.1. *Let $X \sim \text{VG}(r, \theta, \sigma, \mu)$ and suppose $r, \theta, \sigma > 0$. Then it is conjectured that, for $r > 0$,*

$$(r-1)\theta < \text{Med}(X) - \mu < r\theta e^{-2/3r} < \left(r - \frac{2}{3} + \frac{2}{9r} \right) \theta,$$

and

$$\text{Med}(X) \leq \mu + (r + 2 \log 2 - 2)\theta, \quad r \geq 2.$$

Since $\mathbb{E}[X_r] = \mu + r\theta$ for $X_r \sim \text{VG}(r, \theta, \sigma, \mu)$, it follows from (37) and the fact that $M_r = \mu$ for $0 < r \leq 2$ that, if $\theta > 0$, then $M_r \leq \mathbb{E}[X_r]$ for all $r > 0$. Moreover, if the conjectured median bounds hold, then it would follow that $M_r \leq \text{Med}(X_r) \leq \mathbb{E}[X_r]$ for all $r, \theta, \sigma > 0$, meaning that the $\text{VG}(r, \theta, \sigma, \mu)$ distribution would satisfy the mean-median-mode inequality (Groeneveld and Meeden [51] and van Zwet [125]).

TABLE 1
Median of the $\text{VG}(r, 1, \sigma, 0)$ distribution.

$r \backslash \sigma$	0.1	0.3	1	3	10	30
0.5	0.0863	0.0798	0.0502	0.0195	0.00582	0.00192
1	0.454	0.444	0.380	0.276	0.198	0.157
2.5	1.872	1.861	1.775	1.621	1.531	1.507
5	4.350	4.338	4.246	4.084	4.012	4.001
10	9.340	9.328	9.233	9.071	9.009	9.001

3. PARAMETER ESTIMATION

Due to its numerous applications, parameter estimation of the VG distribution is of much interest. In this section, we give an overview of available techniques from the literature. In the following, we let $\mathbf{X} = (X_1, \dots, X_n)$ be an i.i.d. sample of $\text{VG}(r, \theta, \sigma, \mu)$ random variables from which we wish to estimate the four unknown parameters r, θ, σ and μ .

3.1 Method of moments estimation

Let $m'_j = \mathbb{E}[X^j]$, $j = 1, \dots, 4$, with $X \sim \text{VG}(r, \theta, \sigma, \mu)$ be the first four raw moments of the VG distribution. Moreover, let $m_j = \mathbb{E}[(X - m_1)^j]$, $j = 2, 3, 4$, be the central moments. Moment estimation consists of estimating the raw or central moments through the arithmetic means, i.e. $\hat{m}'_j = n^{-1} \sum_{i=1}^n X_i^j$, $j = 1, \dots, 4$, and $\hat{m}_j = n^{-1} \sum_{i=1}^n (X_i - \hat{m}'_1)^j$, $j = 2, 3, 4$. As noted by Finlay and Seneta [32], in the four parameter case, it is not possible to resolve $\hat{m}_j = m_j$, $j = 1, \dots, 4$, for (r, θ, σ, μ) explicitly. However, in the symmetric case $\theta = 0$ and in the case $\mu = 0$, this is possible. To illustrate this for $\theta = 0$, we will work in the parameterisation (2) of Madan and Seneta [83]. Using the formulas for the first, second and fourth central moments of the VG distribution (32), (33) and (35), one readily obtains the following explicit consistent estimators, which is also unbiased in the case of the estimator for μ (Madan and Seneta [83]):

$$\begin{aligned} \hat{\mu} &= \frac{1}{n} \sum_{i=1}^n X_i, \quad \hat{\sigma}_0^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2, \\ \hat{\nu} &= \frac{1}{3\hat{\sigma}_0^4 n} \left\{ \sum_{i=1}^n (X_i - \bar{X})^4 \right\} - 1. \end{aligned}$$

In Finlay and Seneta [32], the moment estimates are obtained by minimising the quantity $\sum_{j=1}^4 ((\hat{m}'_j - m'_j)/\hat{m}'_j)^2$ with respect to (r, θ, σ, μ) . In Seneta [115], explicit moment estimators are obtained for the parametrisation (2) by assuming that expressions of the form θ_0^k , $k \geq 2$ are negligible since the true value of θ_0 is often small in financial modelling applications. An advantage of the moment estimators is the easy numerical implementation. A competitive simulation study in order to evaluate the performance of the moment estimators is carried out in Finlay and Seneta [32].

3.2 Maximum-likelihood estimation

We maximise the log-likelihood function

$$\ell(\mathbf{X}, r, \theta, \sigma, \mu) = \sum_{i=1}^n \log p(X_i, r, \theta, \sigma, \mu)$$

with respect to (r, θ, σ, μ) , where p is the VG density (1). Due to the complexity of the density involving the modified Bessel function of the second kind, there is no analytic solution to this optimisation problem and one relies on numerical methods in order to determine the global maximum. There exist implementations of numerical procedures for maximum-likelihood estimation, such as the *VarianceGamma* R package of Scott and Dong [111] through the *vgFit*-function. Madan and Seneta [83] proposed a transformed maximum likelihood method which permits to perform optimisation with respect to a more tractable density function (see also Madan and Seneta [82]). In Madan et al. [79], MLE is used in order to estimate the parameters for a real financial data example. In Bee, Dickson and Santi [14], the *VarianceGamma* package is compared to the *ghyp* R package of Breymann and Lüthi [16], which implements the more general GH distribution, and to the ECM algorithm that we describe in Section 3.3. In Cervellera and Tucci [19], the authors investigate three different numerical optimisation procedures (including the *VarianceGamma* R package, *Matlab* and *Ezgrad*) and point out several difficulties that can be encountered when using MLE. More precisely, the authors exhibit that the likelihood function is sensitive to the removal of a single observation and that in some cases the algorithms produce different output for different starting values.

To demonstrate the difficulties which can occur with the MLE, we generated an i.i.d. sample of $\text{VG}_2(1/0.9, 0, 4, 0)$ distributed random variables of size $n = 100$. We then used the *vgFit*-function with different starting values in order to estimate the parameters. For this purpose we chose two options that are implemented in the *VarianceGamma* package through the parameter *startValues* of the *vgFit*-function: option *SL* (starting values are based on a fitted skew Laplace distribution and option *MoM* (starting values are based on the method of moments). As can be seen in Figure 2, the estimated distributions based on these two different starting values differ significantly around the modes of the densities.

One reason why MLE is complicated is that the density function is unbounded for $r \leq 1$. This implies that the MLE does not exist since the joint likelihood of i.i.d. observations x_1, \dots, x_n is singular with respect to μ at all data points x_i if $r \leq 1$. Let us clarify the latter issue with a concrete example. Suppose we have three realisations of independent $\text{VG}(r, \theta, \sigma, \mu)$ distributed random variables $x_1 = -0.2, x_2 = 0, x_3 = 0.1$ and let $r = 0.4, \theta = 0, \sigma = 1$

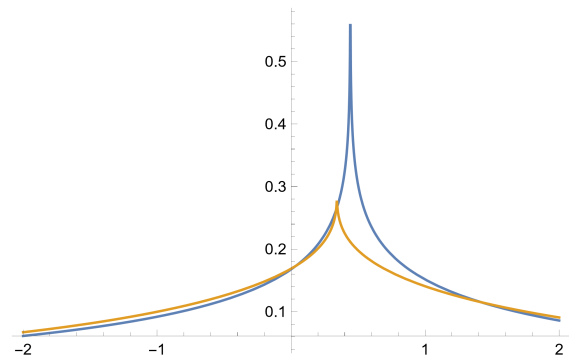


FIG 2. PDFs with estimated parameters with two different starting values: ● starting values based on a fitted skew Laplace distribution and ● based on the method of moments.

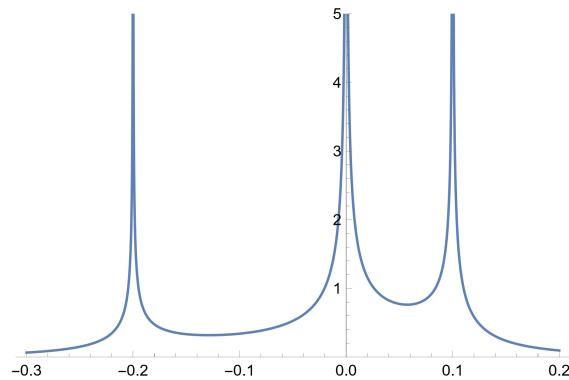


FIG 3. The joint likelihood of x_1, x_2, x_3 with respect to the parameter μ .

be fixed. In Figure 3, the joint likelihood regarding the observations x_1, x_2, x_3 is plotted with respect to μ . One observes that the function is singular at x_1, x_2, x_3 and therefore a MLE does not exist. Thus, numerical optimisation algorithms in order to maximise the function ℓ may get trapped in one of these spikes.

Moreover, the density p is not differentiable at μ if $r \leq 2$. Therefore, the VG distribution does not fulfil the classical regularity assumptions that are required to apply the standard asymptotic theory of maximum likelihood estimators. However, in special cases in which certain parameters are known, the asymptotic theory has already been studied (see, for example, Kotz et al. [64] for the Laplace and asymmetric Laplace distributions).

3.3 ECM algorithms

Motivated by the complexity of the optimisation of the likelihood function, several approaches using ECM (Expectation-Conditional Maximisation) algorithms have been studied recently. The algorithm we present here was first introduced by Nitithumbundit and Chan [93] and McNicholas, McNicholas and Browne [89] for the multivariate VG model and mixtures of the latter. We restrict

ourselves to the one-dimensional case and present the algorithm described in detail in Bee et al. [14]. Here we will consider the VG model under the parametrisation (2). ECM algorithms are considered as iterative procedures based on the maximisation of the likelihood function with missing data. Letting $S \sim \Gamma(\alpha, \alpha)$, $T \sim N(0, 1)$ be independent and using (18) we know that

$$X = \mu + \theta_0 S + \sigma_0 \sqrt{S} T \sim \text{VG}_2(\alpha, \theta_0, \sigma_0, \mu).$$

We now treat $\mathbf{S} = (S_1, \dots, S_n)$ as missing data because the conditional distribution of X on S is $N(\mu + \theta_0 S, \sigma_0^2 S)$, which means we can write the respective joint log-likelihood function as a sum of a conditional normal and a gamma density. More precisely,

$$\begin{aligned} \ell_c(\mathbf{X}, \mathbf{S}, \alpha, \theta_0, \sigma_0, \mu) &= -\frac{n}{2} \log(\sigma_0^2) \\ &\quad - \frac{1}{2} \sum_{i=1}^n \frac{1}{S_i} \frac{(X_i - \mu - S_i \theta_0)^2}{\sigma_0^2} + n\alpha \log(\alpha) \\ &\quad - n \log(\Gamma(\alpha)) + (\alpha - 1) \sum_{i=1}^n \log S_i - \alpha \sum_{i=1}^n S_i \\ &= \ell_c^{(1)}(\mathbf{X}, \mathbf{S}, \theta_0, \sigma_0, \mu) + \ell_c^{(2)}(\mathbf{X}, \mathbf{S}, \alpha). \end{aligned}$$

ECM-type algorithms iterate an expectation (E)-step and a conditional maximisation (CM)-step.

E-step: Here one computes the conditional expectation $\mathbb{E}[\ell_c(\mathbf{X}, \mathbf{S}, \alpha, \theta_0, \sigma_0, \mu) | \mathbf{X}]$, which breaks down to the computation of $\mathbb{E}[S_i | \mathbf{X}]$, $\mathbb{E}[1/S_i | \mathbf{X}]$ and $\mathbb{E}[\log S_i | \mathbf{X}]$. Embrechts [29] showed that the conditional distribution of S on X is generalized inverse Gaussian and hence we can estimate S_i , $1/S_i$ and $\log S_i$ by the corresponding conditional expectations

$$(40) \quad \hat{S}_i = \mathbb{E}[S_i | \mathbf{X}] = \frac{\delta_i K_{\alpha+1/2}(\phi \delta_i)}{\phi K_{\alpha-1/2}(\phi \delta_i)},$$

$$(41) \quad \widehat{1/S_i} = \mathbb{E}[1/S_i | \mathbf{X}] = \frac{\phi K_{\alpha-3/2}(\phi \delta_i)}{\delta_i K_{\alpha-1/2}(\phi \delta_i)},$$

$$(42) \quad \begin{aligned} \widehat{\log S_i} &= \mathbb{E}[\log S_i | \mathbf{X}] = \log \left(\frac{\delta_i}{\phi} \right) \\ &\quad + \frac{\hat{K}_{\alpha-1/2}^{(1,0)}(\phi \delta_i)}{K_{\alpha-1/2}(\phi \delta_i)}, \end{aligned}$$

where $\phi = \sqrt{2\alpha + \theta_0^2/\sigma_0^2}$, $\delta_i = |X_i - \mu|/\sigma_0$, and $\hat{K}_\alpha^{(1,0)}(z) = (K_{\alpha+h}(z) - K_{\alpha-h}(z))/(2h)$, for some small constant $h > 0$, is an approximation of $\partial K_\alpha(z)/\partial \tau|_{\tau=\alpha}$ (see Bee et al. [14, p. 76]).

CM-step: One maximises the conditional expectation of the log-likelihood function $\mathbb{E}[\ell_c(\mathbf{X}, \mathbf{S}, \alpha, \theta_0, \sigma_0, \mu) | \mathbf{X}]$ with respect to $(\alpha, \theta_0, \sigma_0, \mu)$. One can maximise the conditional expectations $\mathbb{E}[\ell_c^{(1)}(\mathbf{X}, \mathbf{S}, \theta_0, \sigma_0, \mu) | \mathbf{X}]$ and

$\mathbb{E}[\ell_c^{(2)}(\mathbf{X}, \mathbf{S}, \alpha) | \mathbf{X}]$ separately. The first optimisation problem can be solved analytically by

$$(43) \quad \hat{\mu} = \frac{\sum X_i \widehat{1/S_i} \sum \hat{S}_i - n \sum X_i}{\sum \widehat{1/S_i} \sum \hat{S}_i - n^2},$$

$$(44) \quad \hat{\theta}_0 = \frac{\sum X_i - n \hat{\mu}}{\sum \hat{S}_i},$$

$$(45) \quad \hat{\sigma}_0^2 = \frac{1}{n} \sum \widehat{1/S_i} (X_i - \hat{\mu})^2 - \frac{1}{n} \hat{\theta}_0^2 \sum \hat{S}_i.$$

For optimisation with respect to α , no analytic solution is available. It is suggested to optimise either $\mathbb{E}[\ell_c^{(2)}(\mathbf{X}, \mathbf{S}, \alpha) | \mathbf{X}]$ or the likelihood of the VG distribution with updated parameters $\ell(\mathbf{X}, \mathbf{S}, \alpha, \hat{\theta}_0, \hat{\sigma}_0, \hat{\mu})$. Bee et al. [14] carried out a competitive simulation study, finding the latter of the two options to be preferable. Given a vector of starting values $(\hat{\alpha}^{(0)}, \hat{\theta}_0^{(0)}, \hat{\sigma}_0^{(0)}, \hat{\mu}^{(0)})$, a pseudocode for the t -th iteration of the algorithm is:

- **E-step 1:** Compute $\hat{S}_i^{(t-1/2)}$ and $\widehat{1/S_i}^{(t)}$ using (40) and (41) and the current estimates $\hat{\alpha}^{(t-1)}$, $\hat{\theta}_0^{(t-1)}$, $\hat{\sigma}_0^{(t-1)}$ and $\hat{\mu}^{(t-1)}$.
- **CM-step 1:** Compute $\hat{\theta}_0^{(t)}$, $\hat{\sigma}_0^{(t)}$ and $\hat{\mu}^{(t)}$ using (43), (44) and (45).
- **E-step 2:** Compute $\hat{S}_i^{(t)}$ and $\widehat{\log S_i}^{(t)}$ through (40) and (42) using the updated parameters $\hat{\theta}_0^{(t)}$, $\hat{\sigma}_0^{(t)}$ and $\hat{\mu}^{(t)}$ as well as $\hat{\alpha}^{(t-1)}$.
- **CM-step 2:** Compute $\hat{\alpha}^{(t)}$ with a numerical optimisation method using the updated expectations $\hat{S}_i^{(t)}$ and $\widehat{\log S_i}^{(t)}$.

The algorithm encounters difficulties during the E-step when X_i is very close to μ due to the singularities of the density function. In this case, it is recommended to replace δ_i by $\delta_i^* = \Delta$ if $\delta_i < \Delta$. Simulations for different values of Δ are available in Nitithumbundit and Chan [94], where the authors also provide an improved version of the algorithm in the multivariate case.

3.4 Further estimation techniques

For the sake of completeness, we mention several other estimation methods in the literature. In Loregian, Mercuri and Rroji [74], a Gaussian quadrature is used to approximate an integral representation of the VG density by a mixture of Gaussian densities, and then an ECM algorithm with respect to the mixing variable is used to estimate the parameters. Nzokem [98] uses a fractional Fourier transform in order to approximate the VG density. In Finaly and Seneta [32], a minimum χ^2 -approach is studied, which compares theoretical and empirical quantiles (see also Finlay and Seneta [31] and Tjetjep and Seneta [122]). Finlay and Seneta [32] further present an

estimation procedure based on the characteristic function. It is worth noting that this approach is still valid when there is a certain dependency structure in the data. Finlay and Seneta [32] describe a Bayesian procedure, and a Markov Chain Monte Carlo method is used in order to sample from the posterior distribution. We also refer to Rathgeber, Johannes and Stöckl [105], in which a number of the estimation techniques described in this section are compared on the basis of real financial data.

3.5 Simulation study

We conducted a competitive simulation study in R in order to assess the finite sample performance of the estimators. We included the MLE as implemented in the *VarianceGamma* package, the ECM algorithm described in Section 3.3 as well as the method of moments estimator (MoM) from Section 3.1. We used the parametrisation $VG'_2 = (\alpha', \theta_0, \sigma_0, \mu)$, where $\alpha' = 1/\alpha$. The MoM estimator is implemented exactly as described at the beginning of Section 3.1, whereby we expressed the central moments in the parametrisation VG'_2 and used the Nelder-Mead algorithm with starting value $(1, 0, 1, 0)$ as implemented in the R function `optim` to calculate the estimator. Concerning the MLE, we employed the *vgFit* function of the *VarianceGamma* package with the Nelder-Mead algorithm and the option *SL* for the *startvalues* parameter. The ECM algorithm is implemented as described in Section 3.3 with $h = 10^{-5}$ and $\Delta = 10^{-10}$ and starting point $(1, 0, 1, 0)$. The algorithm stops if the Euclidean distance between two consecutive parameter values is less than 10^{-4} or after 100 repetitions. Results are reported in Table 2. Since quantities such as bias or mean squared error heavily depend on the parametrisation, we decided to report the average of the \mathcal{L}^1 -distance between the density with true and estimated parameters. The corresponding values can be found in column \mathcal{L}^1 in Table 2. The column *NE* corresponds to *not existent* and reports the estimated relative frequency (out of 100 samples) that an estimator does not exist. For the MoM and the MLE we considered an estimator to not exist whenever the algorithm threw an error and for the ECM algorithm in addition when the maximum repetition number of 100 repetitions was reached.

As can be observed in column *NE*, the MoM and *vgFit*-function encountered numerical difficulties for several parameter constellations whereby the ECM algorithm seems to be stable for all parameter values. Also regarding the \mathcal{L}^1 distance, the *vgFit*-function and MoM are outperformed by ECM throughout all parameter constellations that were considered in our simulation.

4. APPLICATIONS

4.1 Exact distribution in mathematical statistics

4.1.1 Connection to sample correlation Let (X_i, Y_i) , $1 \leq i \leq n$, be i.i.d. bivariate normal random vectors with

TABLE 2
Simulation results for 10,000 Monte Carlo repetitions and a sample size of $n = 1000$.

	\mathcal{L}^1			NE			
	VG'_2	MoM	<i>vgFit</i>	ECM	MoM	<i>vgFit</i>	ECM
$(1, 0, 1, 0)$		0.05	0.331	0.035	0	0	0
$(2.2, 0, 1, 0)$		0.064	0.122	0.034	0	0	0
$(0.15, 0, 1, 0)$		0.04	14.6	0.036	8	1	1
$(5, 0, 3, 0)$		0.479	0.767	0.103	19	0	0
$(0.5, 0.7, 0.5, 0)$		0.032	0.999	0.023	0	0	0
$(1, 1, 0.5, 0)$		0.071	104	0.034	3	44	0
$(0.7, 0.2, 1, 0)$		0.045	24.2	0.036	0	2	0
$(0.4, -1.8, 1.2, 0)$		0.079	0.119	0.055	6	0	0
$(0.3, 0, 0.4, 0)$		0.06	0.015	0.014	16	0	0
$(2, 2, 3, 0)$		0.383	0.238	0.131	4	0	0

mean vector (μ_X, μ_Y) , variances (σ_X^2, σ_Y^2) and correlation coefficient ρ . It was shown by Pearson, Jefferey and Elderton [100] that the product-moment coefficient

$$p_n = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})$$

has a VG distribution (they referred to it as the Bessel function distribution). An alternative proof is given by Kotz et al. [64, Proposition 4.1.5]. In our parametrisation,

$$p_n \sim \text{VG}(n-1, \rho\sigma_X\sigma_Y/n, \sigma_X\sigma_Y\sqrt{1-\rho^2}/n, 0).$$

Here we applied the scaling property (8) of the VG distribution to deal with the case of general variances $\sigma_X^2, \sigma_Y^2 > 0$; Kotz et al. [64] proved the result for the case $\sigma_X = \sigma_Y = 1$. As observed by Kotz et al. [64], when $n = 3$ the product-moment coefficient p_3 has an asymmetric Laplace distribution (see part 1 of Section 2.5.1). Moreover, as in part 4 of Section 2.5.1, let $Z = UV$, where (U, V) is a bivariate normal random vector with zero mean vector, variances (σ_X^2, σ_Y^2) and correlation coefficient ρ . Consider also the mean $\bar{Z}_{n-1} = (n-1)^{-1}(Z_1 + Z_2 + \dots + Z_{n-1})$, where Z_1, Z_2, \dots, Z_{n-1} are independent copies of Z . Then, by the relation (22) and the scaling property (8), we have that, for $n \geq 2$,

$$(46) \quad p_n =_d (1 - 1/n)\bar{Z}_{n-1}.$$

In particular, $p_2 =_d Z/2$. It follows from (46) and the central limit theorem that, as $n \rightarrow \infty$, $\sqrt{n}(p_n - \rho\sigma_X\sigma_Y) \rightarrow_d N(0, \sigma_X^2\sigma_Y^2(1 + \rho^2))$.

4.1.2 Connection to the Wishart distribution Let N be a $p \times n$ matrix, whose n columns are independent p -dimensional multivariate normal random vectors with zero mean vector and positive definite covariance matrix

$V = (v_{ij})$. Then the Wishart distribution is the probability distribution of the $p \times p$ random matrix $X = NN^\top$. We denote such a random matrix by $W_p(V, n)$. The positive integer $n \geq 1$ is referred to as the number of degrees of freedom, and when $V = p = 1$, the Wishart distribution reduces to the χ_n^2 distribution. The Wishart distribution is widely used in multivariate statistics, Bayesian analysis and random matrix theory.

A fundamental distributional property of the Wishart distribution is the marginal distribution of its entries. It is well-known that the diagonal entries are suitably scaled chi-square random variables: $X_{ii} \sim v_{ii}\chi_n^2$, $1 \leq i \leq p$ (see, for example, Kollo [61, Corollary 2.4.2.2]). It is perhaps less well-known that the off-diagonal entries are VG distributed (see Pearson et al. [100] for the two-dimensional case). More precisely,

$$(47) \quad X_{ij} \sim \text{VG}(n, v_{ij}, (v_{ii}v_{jj} - v_{ij}^2)^{1/2}, 0), \quad i \neq j.$$

As V is a covariance matrix, it is natural to set $v_{ii} = \sigma_i^2$ and $v_{ij} = \rho_{ij}\sigma_i\sigma_j$, $i \neq j$, where σ_i^2 is the variance of the i -th component of a p -dimensional $N_p(\mathbf{0}, V)$ multivariate normal random vector, whilst ρ_{ij} is the correlation coefficient of the i -th and j -th components. With this notation, we have that $X_{ij} \sim \text{VG}(n, \rho_{ij}\sigma_i\sigma_j, \sigma_i\sigma_j(1 - \rho_{ij}^2)^{1/2}, 0)$, $i \neq j$. From (22), it therefore follows that X_{ij} has a neat representation as a sum of independent copies of correlated zero mean normal random variables. Let Z_1, \dots, Z_n be independent $\text{PN}(\rho_{ij}, \sigma_i, \sigma_j)$ random variables (recall that this notation was introduced in part 4 of Section 2.5.1). Then

$$X_{ij} = {}_d Z_1 + \dots + Z_n, \quad i \neq j.$$

That the off diagonal entries are VG distributed can be deduced from the Bartlett decomposition of the Wishart distribution (see Anderson [4]) and the representation (18) of the VG distribution in terms of independent normal and chi-square random variables. The Bartlett decomposition of $X \sim W_p(V, n)$ is

$$(48) \quad X = {}_d LAA^\top L^\top,$$

where L is the Cholesky factor of V and A is a lower triangular matrix with diagonal entries $S_i \sim \chi_{n-i+1}$ (so that $S_i^2 \sim \chi_{n-i+1}^2$), $1 \leq i \leq p$, and off-diagonal entries 0 for $j < i$, and $N_{ij} \sim N(0, 1)$ for $i > j$. All entries in the matrix A are mutually independent. From the Bartlett decomposition (48) we obtain that

$$(49) \quad X_{12} = {}_d v_{12}S_1^2 + (v_{11}v_{22} - v_{12}^2)^{1/2}S_1N_{12},$$

where we used that $L_{11} = \sqrt{v_{11}}$, $L_{12} = v_{12}/\sqrt{v_{11}}$ and $L_{22} = (v_{22} - v_{12}^2/v_{11})^{1/2}$. It therefore follows from the representation (18) of the VG distribution that $X_{12} \sim \text{VG}(n, v_{12}, (v_{11}v_{22} - v_{12}^2)^{1/2}, 0)$. We can use the Bartlett decomposition (48) to obtain similar representations of

the off-diagonal elements X_{ij} , $i \neq j$, in terms of independent normal and chi-square random variables, although these representations are more complicated and it is harder to infer that the elements are indeed VG distributed through this approach. However, the fact that the off-diagonal entries are VG distributed with parameters given as in (47) follows from (49) and symmetry considerations. This can be seen analytically, for a general off-diagonal element X_{ij} , by suitably reordering the elements in the Wishart matrix X and making the corresponding reordering in the matrix V , and then obtaining the Bartlett decomposition of the new Wishart matrix X' . The entry X_{ij} (entry X'_{12} in the re-ordered Wishart matrix X') then can be seen to have a representation of the form $X_{ij} = {}_d v_{ij}S_1^2 + (v_{ii}v_{jj} - v_{ij}^2)^{1/2}S_1N_{12}$, and the claim follows.

4.1.3 Connection to Matérn covariance Consider a real-valued Gaussian process $\{X(\mathbf{t}) \in \mathbb{R} : \mathbf{t} \in \mathcal{D} \subset \mathbb{R}^n\}$ on a d -dimensional domain \mathcal{D} with a covariance function $C(\mathbf{h}) = \text{Cov}(X(\mathbf{t}), X(\mathbf{t} + \mathbf{h}))$ that is solely a function of the increment \mathbf{h} . Suppose that $X(\cdot)$ possesses a constant mean function. Then the Gaussian process $X(\cdot)$ is completely determined by its covariance function.

The Matérn covariance function (Matérn [86]) is widely used as a kernel in Gaussian processes, with application areas including spatial statistics (Stein [119]) and uncertainty quantification and machine learning (Gu, Wang and Berger [52], Guttorp and Gneiting [53]). The Matérn covariance function is given by

$$(50) \quad C_M(h) = \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{\sqrt{2\nu}}{\phi} h \right)^\nu K_\nu \left(\frac{\sqrt{2\nu}}{\phi} h \right),$$

where $\phi > 0$ is a range parameter, $\nu > 0$ is a smoothness parameter, and $h = |\mathbf{h}|$. We observe that the Matérn covariance function is equal (up to a constant) to the PDF of a $\text{VG}(r, 0, \sigma, 0)$ distribution with support restricted to the positive real line.

When $\nu \in \{1/2, 3/2, 5/2, \dots\}$ the Matérn covariance takes a simple-closed form (recall that the VG PDF simplifies in this case; see equation (7)). For example, when $\nu = 1/2$, we recover the power exponential covariance function $C_{PE}(h) = \exp(-h/\phi)$, whilst in the limit $\nu \rightarrow \infty$ the covariance function (50) reduces to squared exponential covariance function (known as the radial basis function kernel in the machine learning literature) $C_{RBF}(h) = \exp(-h^2/(2\phi^2))$ (recall that the VG PDF tends to the normal PDF in this limit; see Section 2.5.1, part 3). A key reason for the popularity of the Matérn covariance is that the parameter ν allows for precise control of the smoothness of the process; indeed, a Gaussian process with Matérn covariance function is exactly $\lfloor \nu \rfloor - 1$ times differentiable. On account of this flexible local behaviour, Stein [119] has recommended the Matérn covariance for modelling in spatial statistics. For large h ,

the Matérn covariance function has an exponential decay $C_M(h) \sim h^{\nu-1/2} \exp(-\sqrt{2\nu}h/\phi)$, $h \rightarrow \infty$ (apply (66) to (50)). Recently, Ma and Bhadra [78] have proposed a new class of covariance functions that retains the attractive local behaviour of the Matérn covariance whilst introducing the feature of polynomial decay at large distances which is suitable for modelling processes with polynomially decaying dependence.

4.2 Variance-gamma process

In Section 2.3, we saw that the VG distribution is infinitely divisible. As a consequence of the theory of infinitely divisible distributions and processes (see Ferguson and Klass [30]), we can define a Lévy process with VG distributed increments. The resulting Lévy process is known as the VG process, which, as we shall discuss in Section 4.3, has found application in financial modelling. The terminology *Laplace motion* is also often employed; see, for example, Kotz et al. [64] and Kozubowski and Podgórski [67]. The first complete presentation of the VG model was given by Madan and Seneta [83] for the case of symmetric VG distributed increments ($\theta = 0$), which was extended to incorporate skewness ($\theta \in \mathbb{R}$) by Madan and Milne [80] and Madan et al. [79]. A few years before the seminal paper of Madan and Seneta [83], the same authors published two papers (Madan and Seneta [81, 82]) which contained some basic properties of the VG process; see Seneta [116] for an account of the early years of the VG process.

In this section, we shall review some of the most fundamental properties of the VG process and briefly discuss why these properties are desirable in the context of financial modelling. We shall only consider the univariate VG process; for an account of the multivariate VG process, which was introduced by Madan and Seneta [83], we refer the reader to Luciano and Schoutens [75], Semeraro [114], Luciano and Semeraro [76] and Luciano, Marena and Semeraro [77], and Linders and Stassen [73] for applications to basket options calibration on the DJIA index. We also do not discuss methods for simulating VG processes, and refer the reader to Avramidis and L'Ecuyer [6], Fu [34] and Korn, Korn and Kroisandt [63, Section 7.3.3] for algorithms.

Madan et al. [79] constructed the VG process as a subordinated Brownian motion with a time change by a gamma process. Consider a Brownian motion with drift θ and volatility σ given by $b(t; \theta, \sigma) = \theta t + \sigma B(t)$, where $B(t)$ is standard Brownian motion. Recall that the gamma process $\gamma(t; \mu, \nu)$ with mean rate μ and variance rate ν is a Lévy process with increment $\tau_h = \gamma(t+h; \mu, \nu) - \gamma(t; \mu, \nu)$ following the $\Gamma(\mu^2 h/\nu, \mu/\nu)$ distribution. Suppose that the processes $B(t)$ and $\gamma(t; \mu, \nu)$ are independent. Then, the VG process $X(t) = X(t; \sigma, \nu, \theta)$, with parameters $\nu > 0$, $\theta \in \mathbb{R}$ and $\sigma > 0$, is defined as

$$(51) \quad X(t) = b(\gamma(t; 1, \nu); \theta, \sigma).$$

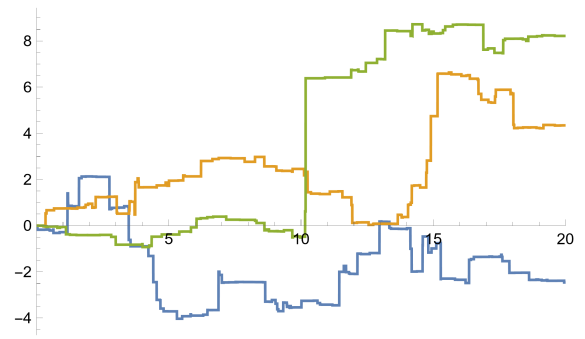


FIG 4. Three realisations of the VG process $X(t; 1, 1, 0)$ (top image) and $X(t; 1, 2, 0)$ (bottom image).

As we shall see, control over skewness and kurtosis can be attained via the parameters θ and ν , respectively. This feature is important in financial modelling. As noted by Madan and Seneta [83, p. 517], the random gamma time change in (51) has an interesting economic interpretation of supposing that economically relevant time is random in that the market (price process $X(t)$) moves at different speeds on different days, and the process $\gamma(t; 1, \nu)$ is a measure of this speed. Samples paths showing the effect of varying the parameter ν in the symmetric VG process are given in Figure 4.

To see that the increments of the VG process (51) are VG distributed, we observe that

$$\begin{aligned} X(t+h) - X(t) &= \theta\tau_h + \sigma(B(\gamma(t+h; 1, \nu)) \\ &\quad - B(\gamma(t; 1, \nu))) =_d \theta\tau_h + \sigma\tau_h^{1/2}B(1). \end{aligned}$$

As $B(1) \sim N(0, 1)$ and $\tau_h \sim \Gamma(\mu^2 h/\nu, \mu/\nu)$ are independent, it follows from (18) that

$$(52) \quad X(t+h) - X(t) \sim \text{VG}(2h/\nu, \theta\nu/2, \sigma\sqrt{\nu/2}, 0).$$

In particular, since $X(0) = 0$,

$$(53) \quad X(t) \sim \text{VG}(2t/\nu, \theta\nu/2, \sigma\sqrt{\nu/2}, 0).$$

As we shall see shortly, the VG process is of finite variation, so can be expressed as the difference of two independent increasing processes. These increasing processes are in fact gamma processes, and we have the representation (see Madan et al. [79])

$$(54) \quad X(t) =_d \gamma_p(t; \mu_p, \nu_p) - \gamma_n(t; \mu_n, \nu_n),$$

where

$$\begin{aligned} \mu_p &= \frac{1}{2}\sqrt{\theta^2 + \frac{2\sigma^2}{\nu}} + \frac{\theta}{2}, & \mu_n &= \frac{1}{2}\sqrt{\theta^2 + \frac{2\sigma^2}{\nu}} - \frac{\theta}{2}, \\ \nu_p &= \mu_p^2\nu, & \nu_n &= \mu_n^2\nu. \end{aligned}$$

Here $\gamma_p(t; \mu_p, \nu_p)$ plays the role of a returns process, whilst $\gamma_n(t; \mu_n, \nu_n)$ is a losses process. The representation (54) of the VG process as a difference of two independent gamma processes follows from the representation (20) and the fact that $\gamma(t; \mu, \nu) \sim \Gamma(\mu^2 t/\nu, \mu/\nu)$.

From (53) and the moments formulas (30) and (33)–(35) we obtain that

$$\mathbb{E}[X(t)] = \theta t,$$

$$\mathbb{E}[(X(t) - \mathbb{E}[X(t)])^2] = (\theta^2\nu + \sigma^2)t,$$

$$\mathbb{E}[(X(t) - \mathbb{E}[X(t)])^3] = (2\theta^3\nu^2 + 3\sigma^2\theta\nu)t,$$

$$\begin{aligned} \mathbb{E}[(X(t) - \mathbb{E}[X(t)])^4] &= (3\sigma^4\nu + 12\sigma^2\theta^2\nu^2 + 6\theta^4\nu^3)t \\ &\quad + (3\sigma^4 + 6\sigma^2\theta^2\nu + 3\theta^4\nu^2)t^2 \end{aligned}$$

(see Madan et al. [79]). We see that the parameter θ does not directly control skewness, although skewness is zero when $\theta = 0$, and positive when $\theta > 0$. When $\theta = 0$, the kurtosis is $3(1 + \nu/t)$, and so ν is the percentage of excess kurtosis for a unit interval of time $t = 1$. As t increases, the kurtosis decreases to 3, the kurtosis of the standard normal distribution. As noted by Madan and Seneta [83], this is consistent with the empirical evidence that daily returns are heavy tailed, whilst monthly returns are normally distributed.

From (53) and (12), we have that the characteristic function of the VG process is given by

$$(55) \quad \varphi_{X(t)}(u) = (1 - i\theta\nu u + (\sigma^2\nu/2)u^2)^{-t/\nu}$$

(see Madan et al. [79]). That the characteristic function (55) takes a simple form is important in option pricing, since, for example, the fast fourier transform approach from the classic paper of Carr and Madan [21] can be applied to numerically determine option values. From (15) and (53) we have that the Lévy-Hinchin representation of the VG process is given by $\varphi_{X(t)}(u) = \exp(t \int_{\mathbb{R}} (e^{iux} - 1)\nu_{VG}(x) dx)$, with Lévy density

$$\nu_{VG}(x) = \frac{\mu_n^2}{\nu_n|x|} e^{-\mu_n|x|/\nu_n} \mathbf{1}_{x<0} + \frac{\mu_p^2}{\nu_p x} e^{-\mu_p x/\nu_p} \mathbf{1}_{x>0}.$$

From its Lévy-Hinchin representation, it follows that the VG process is a pure jump process (there is no diffusion component). The pure jump nature of the VG process represented a departure from much of the literature on option pricing; for example, the Black-Scholes model is a pure diffusion model. It was noted by Bakshi, Chen and Cao [9] that a jump component is important in option pricing, as pure diffusion models have difficulties in explaining smile effects, in particular in short-dated option prices. Moreover, as asserted by Madan et al. [79], the Black-Scholes model is a parametric special case of the VG model (recall from Section 2.5.1 that the normal distribution is a limiting case of the VG distribution), meaning high activity is already accounted for and so it is not necessary to include a diffusion component in addition. Indeed, due to the factor of $|x|^{-1}$, the Lévy density integrates to infinity, and so the VG process is of infinite activity. Also, since $|x|$ is integrable with respect to the

Lévy density, the process is of finite variation. As noted by Carr et al. [20], finite variation processes may be more useful than infinite variation processes in explaining the measure change from the statistical to the risk neutral process, since they allow more flexibility between the local characteristic of the martingale components under the two measures. Finally, again as noted by Carr et al. [20], the VG process has a completely monotone Lévy density. In particular, the derivative of the Lévy density is positive for negative jumps and negative for positive jumps, so that large jumps arrive less frequently than small jumps, which is consistent with what we observe in price movements. Indeed, most jumps are of a vanishingly small size; for a compound Poisson approximation and a detailed account of the fine structure of the jumps see Madan and Seneta [83] and Kotz et al. [64, Section 4.2].

4.3 Financial modelling

The Black-Scholes model is the paradigm model in mathematical finance. Under this model, the asset price $S(t)$ at time t obeys geometric Brownian motion

$$(56) \quad S(t) = S(0) \exp((\mu - \sigma^2/2)t + \sigma B(t)),$$

where $\mu \in \mathbb{R}$ and $\sigma > 0$ are constants denoting drift and volatility, and $B(t)$ is standard Brownian motion. Whilst widely used in financial modelling, a number of deficiencies have been noted. For example, as detailed by Heyde [55], under geometric Brownian motion, the log returns $Y(t) = \log S(t) - \log S(t-1)$ are i.i.d. normal random variables, which is often in contrast to empirical log returns data; see Eberlein and Keller [26] and Leonenko, Petherick and Sikorskii [72] for empirical studies. In particular, the tails of the normal distribution are not heavy enough for modelling of daily returns; this was one of the original motivations of Madan and Seneta [83] to introduce the VG process with an additional parameter to control kurtosis to allow for heavier tails. On account of the flexibility to control skewness and kurtosis to allow for a better empirical fit to financial data, together with other desirable properties such as those discussed in Section 4.2, the VG model has found application in financial modelling; a comparison of the fit of the VG distribution and the normal distribution to real financial data is given in Figure 5. In this section, we give a brief introduction to an extensive literature.

The statistical and risk neutral dynamics of a stock price in terms of the VG process was given by Madan et al. [79]. The statistical stock price is obtained by replacing the role of Brownian motion in the geometric Brownian motion model (56) by the VG process:

$$(57) \quad S(t) = S(0) \exp(mt + X(t; \sigma_S, \nu_S, \theta_S) + \omega_S t),$$

where $X(\cdot)$ is the VG process, m is the mean rate of return on the stock under the statistical probability measure and the subscript S in the VG parameters stresses that they are the statistical parameters. The

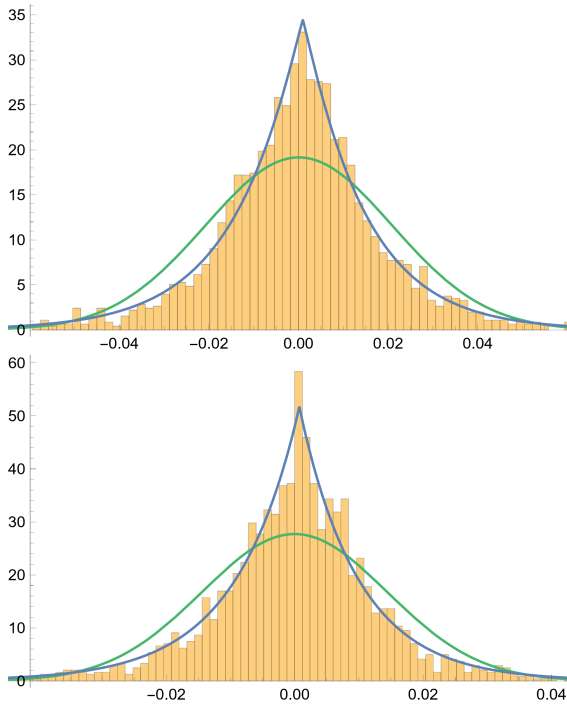


FIG 5. A real data example: We see the daily log returns of the Mercedes Benz Group AG (top picture, 01.01.2013 – 01.01.2023, 2517 observations) and the Global X DAX Germany ETF (bottom picture, 01.01.2015 – 01.01.2023, 2013 observations) fitted to a VG distribution \bullet and a normal distribution \bullet . We used maximum likelihood estimation in order to fit the VG distribution.

value $\omega_S = \nu_S^{-1} \log(1 - \theta_S \nu_S - \sigma_S^2 \nu_S / 2)$ is determined by the non-arbitrage condition to ensure that $\mathbb{E}[S(t)] = S(0) \exp(mt)$, that is $\mathbb{E}[\exp(X(t))] = \exp(-\omega_S t)$, with $\mathbb{E}[\exp(X(t))]$ evaluated using (53) and the moment generating function formula (11).

Under the risk neutral process, stock prices discounted at the risk free interest rate are martingales. Therefore the mean rate of return on the stock under this probability measure is the continuously compounded interest rate r . The risk neutral process is

$$(58) \quad S(t) = S(0) \exp(rt + X(t; \sigma_{RN}, \nu_{RN}, \theta_{RN}) + \omega_{RN} t),$$

where the subscripts RN indicate that these are the risk free parameters, and $\omega_{RN} = \nu_{RN}^{-1} \log(1 - \theta_{RN} \nu_{RN} - \sigma_{RN}^2 \nu_{RN} / 2)$, derived using the same considerations as for the statistical stock price. As discussed in detail by Madan et al. [83], unlike for diffusion based price processes, the statistical and risk free parameters do not need to be equal, and are often quite different.

As noted by Madan et al. [83], the log stock price relative $\log(S(t)/S(0))$ when prices follow the VG process dynamics of (57) is VG distributed (this follows from (8) and (53)):

$$\log(S(t)/S(0)) \sim \text{VG}(2t/\nu, \theta\nu/2, \sigma\sqrt{\nu/2}, (m + \omega)t),$$

where $\omega = \nu^{-1} \log(1 - \theta\nu - \sigma^2\nu/2)$ and we have dropped the subscript S from the parameters. This time using (52), we see that the log returns are also VG distributed:

$$Y(t) = \log S(t) - \log S(t-1) \\ \sim \text{VG}(2/\nu, \theta\nu/2, \sigma\sqrt{\nu/2}, m + \omega);$$

see Finlay and Seneta [31], who also give formulas to describe the covariance structure of the log returns process $\{Y(t)\}$.

Madan et al. [83] derived a closed-form formula for the price of a European call option when the risk neutral dynamics of the stock price follows the VG model (58). A standard result states that the price of a European call option $c(S(0); K, t)$, for a strike K and maturity t , is given by

$$(59) \quad c(S(0); K, t) = e^{-rt} \mathbb{E}[\max(S(t) - K, 0)].$$

Building on the approach of Madan and Milne [80], Madan et al. [79] evaluated the expectation in (59) by first conditioning on the random gamma process time change, under which the VG process $X(t)$ is normally distributed (see part 1 of Section 2.4) and the option value is given by a Black-Scholes type formula. Finally, the European option price for VG risk neutral dynamics is attained on integrating the resulting conditional Black-Scholes formula with respect to the gamma density. The formula is

$$c(S(0); K, t) = S(0) \Psi \left(d \sqrt{\frac{1-c_1}{\nu}}, (\alpha + s) \sqrt{\frac{\nu}{1-c_1}}, \frac{t}{\nu} \right) \\ - K e^{-rt} \Psi \left(d \sqrt{\frac{1-c_2}{\nu}}, \alpha s \sqrt{\frac{\nu}{1-c_2}}, \frac{t}{\nu} \right), \quad (60)$$

where $\alpha = -\theta/\sqrt{\sigma^2 + \theta^2\nu/2}$, $c_1 = \nu(\alpha + s)^2/2$, $c_2 = \nu\alpha^2/2$, and

$$d = \frac{1}{s} \left[\log \left(\frac{S(0)}{K} \right) + rt + \frac{t}{\nu} \log \left(\frac{1-c_1}{1-c_2} \right) \right],$$

and Ψ is expressed explicitly in terms of the modified Bessel function of the second kind and the degenerate hypergeometric function of two variables; see Madan et al. [79] for the formula. As noted by Madan et al. [79], on letting $\nu \rightarrow 0$ in (60), we recover the classical Black-Scholes option pricing formula.

Madan et al. [79] compared the pricing performance on S&P 500 option data of the pricing formula (60) with the Black-Scholes model and the symmetric special case ($\theta = 0$). They found that the performance was superior over the Black-Scholes model, and provided evidence that whilst the statistical density of the underlying risk is typically symmetric, the risk neutral density implied option data is negatively skewed, demonstrating the importance of considering the general non-symmetric VG model.

Beyond the work of Madan et al. [79], the VG model has been successfully tested on real market data and has

been found to outperform that Black-Scholes model and Jump-Diffusion models in a number of other situations, such as European-style options on the HSI index by Lam, Chang and Lee [71] and currency options by Daal and Madan [23]. Closed-form pricing formulas have also been obtained for various path-independent payoffs under the VG model; see, for example, Aguilar [3] and Ivanov [58]. Typically, however, closed-form option formulas are unavailable; see, for example, Hirska and Madan [56], in which a partial integro-differential equation is derived for pricing American options under the VG process. In such cases, numerical methods are required, and we refer the reader to the thesis of Fiorani [33] for an extensive account of methods, as well as the recent article of Aguilar [3] for an overview of some popular numerical methods.

4.4 Time series modelling

Recall from Section 4.3, that the VG distribution is a useful alternative to the normal distribution in financial modelling on account of its heavier tails and capacity to incorporate skewness. This applies equally well to innovations in time series modelling. As discussed in detail by Johannesson [60], VG distributed innovations are a natural alternative to the widely used normally distributed innovations, on account of a natural interpretation of conditionally normally distributed innovations with gamma distributed variance (which results in heavier tails), whilst retaining some analytic tractability and are relatively simple to simulate from. In this section, we provide a brief overview of a growing literature on time series models with VG distributions for innovations.

In Tomy and Jose [123], an autoregression of order 1 with generalized Gaussian-Laplace innovations was studied:

$$(61) \quad X_t = aX_{t-1} + Z_t + V_t,$$

where a is a model parameter and $Z_t \sim N(0, \rho^2)$ and $V_t \sim \text{VG}(r, \theta, \sigma, \mu)$ are independent. Here the distribution for the innovations is more general than the VG distribution: it is the sum of independent VG and normal random variables (centered, without loss of generality). Tomy and Jose [123] fitted parameters using the classic method of moments, by computing the first six moments; however, the problem of goodness-of-fit with this method was not discussed. Nitithumbundit and Chan [94] considered autoregression of order 1 in a slightly different setting than in (61). They only had a VG component V_t (no Z_t), but the setting was multivariate. The authors successfully used ECM (expectation-conditional maximisation) to fit the model even in the case of unbounded VG density of innovations. They applied their methods to financial data, and reported a superior fit to the data than analogous models with multivariate normally distributed innovations. More general ARMA(p, q) models driven

by asymmetric Laplace (the $\text{VG}(2, \theta, \sigma, \mu)$ distribution) noise and ARMA(p, q) models driven by GARCH asymmetric Laplace noise were studied by Trindade, Zhu and Andrews [124]. Similarly to Nitithumbundit and Chan [94], they applied their models to financial data, again reporting a better fit to the data than for corresponding models with normal innovations.

Johannesson et al. [60] considered a more complicated time series model in the context of road topography. For innovations, a symmetric VG random variable $Y = \mu + \sigma\sqrt{S}T$ is taken from (18) (with asymmetry parameter $\theta = 0$), replacing T with T_t , an autoregressive process of order 1 with Gaussian innovations, and S with S_t , a (nonlinear) autoregressive process of order 1 constructed so that S_t has a gamma distribution for each t and is stationary. They presented two methods to fit the autoregressive parameters, and fitted the model to real-life road topology data, showing that it satisfactorily retrieves the distributions of the length, average slope and height of hills.

Finally, we mention that there has been recent interest in the study of non-Gaussian random fields built on the VG distribution, and we refer the reader to Åberg, Podgórski and Rychlik [2], Åberg and Podgórski [1], Bolin [18], Baxevani, Podgórski and Wegener [13] and the survey of Kozubowski and Podgórski [67] for an overview, as well as Podgórski and Wegener [102] for a discussion of estimation methods, and Bogsjö, Podgórski and Rychlik [17] for an application to the modelling of road surface irregularities.

4.5 Approximation on Wiener space

In Section 2.3, we saw that the VG distribution is infinitely divisible and in Section 2.4 we saw that the VG distribution has simple representations in terms of independent standard normal and gamma random variables. In virtue of these properties, the VG distribution is a natural candidate as a limiting distribution. Indeed, there has been recent interest in VG approximation on Wiener space. VG approximations for double Wiener-Itô integrals have been studied by Azmoodeh, Eichelsbacher and Thäle [7], Eichelsbacher and Thäle [28] and Gaunt [41] via Stein's method for VG approximation, which was first developed by Gaunt [36], for which the starting point is the Stein characterisation (25).

4.5.1 A six moment theorem for double Wiener-Itô integrals Let \mathfrak{H} be a real separable Hilbert space and denote by $\mathfrak{H}^{\odot 2}$ the second symmetric tensor product of \mathfrak{H} , that is $\mathfrak{H}^{\odot 2} = \{v_1 \otimes v_2 + v_2 \otimes v_1 : v_1, v_2 \in \mathfrak{H}\}$, the set of all symmetrized tensor products of two vectors from \mathfrak{H} . For $f \in \mathfrak{H}^{\odot 2}$, we denote the double Wiener-Itô integral by $I_2(f)$ (see Nourdin and Peccati [95, Section 2.7] for a definition and fundamental properties). If $f \in L^2([0, T]^2, dt)$

is symmetric then

$$I_2(f) = \int_{[0,T]^2} f(t_1, t_2) dB_{t_1} dB_{t_2},$$

where $B = (B_t)_{t \in [0,T]}$ is a standard Brownian motion (see Nourdin and Peccati [95, Exercise 2.7.6]). Consider the Wasserstein $d_W(F, G)$ and smooth Wasserstein distance $d_2(F, G)$ between the distributions of two random elements F and G , defined by

$$d_{\mathcal{H}}(F, G) := \sup_{h \in \mathcal{H}} |\mathbb{E}[h(F)] - \mathbb{E}[h(G)]|,$$

where the supremum is taken over the function classes

$$\mathcal{H}_W = \{h : \mathbb{R} \rightarrow \mathbb{R} \mid h' \text{ is Lipschitz, } \|h'\|_{\infty} \leq 1\},$$

$$\mathcal{H}_2 = \{h : \mathbb{R} \rightarrow \mathbb{R} \mid h \in \mathcal{H}_W \text{ and } \|h''\|_{\infty} \leq 1\},$$

respectively, where the supremum norm of $g : \mathbb{R} \rightarrow \mathbb{R}$ is defined by $\|g\|_{\infty} = \sup_{x \in \mathbb{R}} |g(x)|$. Clearly, $d_2(F, G) \leq d_W(F, G)$ for any random elements F and G for which $d_W(F, G)$ is well-defined.

Let $F_n = I_2(f_n)$ with $f_n \in \mathfrak{H}^{\odot 2}$, $n \geq 1$. Following Eichelsbacher and Thäle [28], we write $\text{VG}_c(r, \theta, \sigma)$ for $\text{VG}(r, \theta, \sigma, -r\theta)$, the VG distribution with zero mean. Then, Theorem 5.8 of Eichelsbacher and Thäle [28] gives that, as $n \rightarrow \infty$, the sequence $(F_n)_{n \geq 1}$ converges in distribution to $Y \sim \text{VG}_c(r, \theta, \sigma)$ if and only if $\kappa_i(F_n) \rightarrow \kappa_i(Y)$, $i = 2, 3, 4, 5, 6$. The cumulants $\kappa_i(Y)$ are readily obtained through the formulas of Section 2.7 and the formula $\kappa_j(S + c) = \kappa_j(S)$, $j \geq 2$, for any $c \in \mathbb{R}$. This ‘‘six moment’’ theorem implies that the convergence of a sequence of double Wiener-Itô integrals to the $\text{VG}_c(r, \theta, \sigma)$ distribution is governed solely by the behaviour of the first six cumulants (equivalently, first six moments). This is a VG analogue of the celebrated ‘‘fourth moment’’ theorem of Nualart and Peccati [97] for normal approximation of multiple Wiener-Itô integrals.

Furthermore, quantitative ‘‘six moment’’ theorems are also available. Let

$$\mathbf{M}(F_n) = \max\{|\kappa_i(F_n) - \kappa_i(Y)| : i = 2, 3, 4, 5, 6\}.$$

Then, there exists a universal constant $C > 0$, which only depends on r, θ, σ , such that

$$(62) \quad d_W(F_n, Y) \leq C \sqrt{\mathbf{M}(F_n)}$$

(Eichelsbacher and Thäle [28] and Gaunt [41]), and, in the weaker d_2 metric, it was shown by Azmoodeh et al. [7] that there exist universal constants $C_1, C_2 > 0$ such that

$$(63) \quad C_1 \mathbf{M}(F_n) \leq d_2(F_n, Y) \leq C_2 \mathbf{M}(F_n).$$

The rate of convergence in (63) is optimal, and thus constitutes an analogue of the optimal fourth moment theorem of Nourdin and Peccati [96] for normal approximation.

4.5.2 The generalized Rosenblatt process at extreme critical exponent We now review how Azmoodeh et al. [7] and Gaunt [41] used the bounds (62) and (63) to derive bounds on the rate of convergence for a remarkable limit theorem of Bai and Taqqu [8]. Consider the generalized Rosenblatt process $Z_{\gamma_1, \gamma_2}(t)$, which was introduced by Maejima and Tudor [84] as the double Wiener-Itô integral

$$Z_{\gamma_1, \gamma_2}(t) = \int_{\mathbb{R}^2} \left(\int_0^t (s-x_1)_+^{\gamma_1} (s-x_2)_+^{\gamma_2} ds \right) dB_{x_1} dB_{x_2},$$

where B_x is standard Brownian motion, $\gamma_i \in (-1, -1/2)$, $i = 1, 2$, and $\gamma_1 + \gamma_2 > -3/2$, and the prime $'$ indicates exclusion of the diagonals $x_1 = x_2$ in the stochastic integral. The Rosenblatt process is $Z_{\gamma}(t) = Z_{\gamma, \gamma}(t)$, $-3/4 < \gamma < -1/2$ (see Taqqu [120]). We observe that the generalized Rosenblatt process is a member of the second Wiener chaos, since it can be represented as a double Wiener-Itô integral (see Nourdin and Peccati [95, Section 2.7]). As $Z_{\gamma_1, \gamma_2}(t) = {}_d t^{2+\gamma_1+\gamma_2} Z_{\gamma_1, \gamma_2}(1)$, we will consider the random variable $Z_{\gamma_1, \gamma_2}(1)$; results for general $t > 0$ follow from rescaling. For $\rho \in (0, 1)$, define the random variable Y_{ρ} by

$$Y_{\rho} = \frac{a_{\rho}}{\sqrt{2}}(X_1 - 1) - \frac{b_{\rho}}{\sqrt{2}}(X_2 - 1),$$

where X_1 and X_2 are independent χ_1^2 random variables and

$$a_{\rho} = \frac{(2\sqrt{\rho})^{-1} + (\rho + 1)^{-1}}{\sqrt{(2\rho)^{-1} + 2(\rho + 1)^{-2}}},$$

$$b_{\rho} = \frac{(2\sqrt{\rho})^{-1} - (\rho + 1)^{-1}}{\sqrt{(2\rho)^{-1} + 2(\rho + 1)^{-2}}}.$$

Assume that $\gamma_1 \geq \gamma_2$ and $\gamma_2 = (\gamma_1 + 1/2)/\rho - 1/2$. Then, Bai and Taqqu [8] showed that $Z_{\gamma_1, \gamma_2}(1) \rightarrow_d Y_{\rho}$ as $\gamma_1 \rightarrow -1/2$. (Note that if $\gamma_1 \rightarrow -1/2$, then $\gamma_2 \rightarrow -1/2$.)

By the representation (20), we have that $Y_{\rho} \sim \text{VG}_c(1, (a_{\rho} - b_{\rho})/\sqrt{2}, \sqrt{2a_{\rho}b_{\rho}})$. It was shown by Arras et al. [5] that, for any $i \geq 2$, $\kappa_i(Z_{\gamma_1, \gamma_2}(1)) = \kappa_i(Y_{\rho}) + O(-\gamma_1 - 1/2)$, as $\gamma_1 \rightarrow -1/2$, and combining with (62) and (63) yields that, as $\gamma_1 \rightarrow -1/2$,

$$d_W(Z_{\gamma_1, \gamma_2}(1), Y_{\rho}) \leq C_{\rho} \sqrt{-\gamma_1 - \frac{1}{2}},$$

$$d_2(Z_{\gamma_1, \gamma_2}(1), Y_{\rho}) \leq C'_{\rho} \left(-\gamma_1 - \frac{1}{2} \right).$$

where C_{ρ} and C'_{ρ} only depend on ρ . By employing a standard smoothing technique to the above d_2 metric bound, Gaunt and Li [45] deduced the Kolmogorov distance bound

$$\sup_{z \in \mathbb{R}} |\mathbb{P}(Z_{\gamma_1, \gamma_2}(1) \leq z) - \mathbb{P}(Y_{\rho} \leq z)|$$

$$\leq C''_{\rho} \left(-\gamma_1 - \frac{1}{2} \right)^{1/3} \log \left(\frac{1}{-\gamma_1 - 1/2} \right), \quad \text{as } \gamma_1 \rightarrow -\frac{1}{2}.$$

APPENDIX A: THE MODIFIED BESSEL FUNCTION OF THE SECOND KIND

In this appendix, we state some basic properties of the modified Bessel function of the second kind that are used in this review. Unless otherwise stated, these properties are given in Olver et al. [99].

The modified Bessel function of the second kind $K_\nu(x)$ is defined for $\nu \in \mathbb{R}$ by

$$K_\nu(x) = \int_0^\infty e^{-x \cosh(t)} \cosh(\nu t) dt, \quad x > 0.$$

For $x > 0$, the function $K_\nu(x)$ is positive for all $\nu \in \mathbb{R}$. For $\nu = m + 1/2$, $m = 0, 1, 2, \dots$,

$$(64) \quad K_{m+1/2}(x) = \sqrt{\frac{\pi}{2x}} \sum_{j=0}^m \frac{(m+j)!}{(m-j)!j!} (2x)^{-j} e^{-x}.$$

The modified Bessel function $K_\nu(x)$ has the following asymptotic behaviour:

$$(65) \quad K_\nu(x) \sim \begin{cases} 2^{|\nu|-1} \Gamma(|\nu|) x^{-|\nu|}, & x \downarrow 0, \nu \neq 0, \\ -\log x, & x \downarrow 0, \nu = 0, \end{cases}$$

$$(66) \quad K_\nu(x) \sim \sqrt{\frac{\pi}{2x}} e^{-x}, \quad x \rightarrow \infty, \nu \in \mathbb{R}.$$

The following differentiation formula holds:

$$(67) \quad \frac{d}{dx} (x^\nu K_\nu(x)) = -x^\nu K_{\nu-1}(x).$$

An integral formula of Gradshteyn and Ryzhik [50] states that for $a, \alpha > 0$ and $\nu > -1/2$,

$$(68) \quad \int_0^a t^\nu K_\nu(\alpha t) dt = \frac{\sqrt{\pi} 2^{\nu-1} \Gamma(\nu + 1/2)}{\alpha^\nu} \times \\ \times a (K_\nu(\alpha a) \mathbf{L}_{\nu-1}(\alpha a) + K_{\nu-1}(\alpha a) \mathbf{L}_\nu(\alpha a)),$$

It is immediate from the PDF (3) that, for $\nu > -1/2$ and $0 \leq |\beta| < \alpha$,

$$(69) \quad \int_{-\infty}^\infty |t|^\nu e^{\beta t} K_\nu(\alpha |t|) dt = \frac{\sqrt{\pi} (2\alpha)^\nu \Gamma(\nu + 1/2)}{(\alpha^2 - \beta^2)^{\nu+1/2}}.$$

The following asymptotic approximation is deduced from a rescaling of the limiting form (2.13) in Gaunt [39]. For $\beta < \alpha$ and $\nu > -1/2$, we have, as $x \rightarrow \infty$,

$$(70) \quad \int_x^\infty e^{\beta t} t^\nu K_\nu(\alpha t) dt \sim \sqrt{\frac{\pi}{2\alpha}} \frac{1}{(\alpha - \beta)} x^{\nu-1/2} e^{-(\alpha-\beta)x}.$$

The ratio $K_{\nu-1}(x)/K_\nu(x)$ satisfies the following bounds. Let $x > 0$. Then, for $\nu > 1/2$,

$$(71) \quad \frac{x}{\nu - 1/2 + \sqrt{(\nu - 1/2)^2 + x^2}} < \frac{K_{\nu-1}(x)}{K_\nu(x)} \\ < \frac{x}{\nu - 1 + \sqrt{(\nu - 1)^2 + x^2}},$$

and, for $\nu \geq 3/2$,

$$(72) \quad \frac{K_{\nu-1}(x)}{K_\nu(x)} \leq \frac{x}{\nu - 1/2 + \sqrt{(\nu - 3/2)^2 + x^2}},$$

with equality in (72) if and only if $\nu = 3/2$. The lower and upper bounds in (71) were derived by Segura [113] and Ruiz-Antolín and Segura [108], respectively. Inequality (72) is given in Gaunt and Merkle [48].

ACKNOWLEDGMENTS

The authors would like to thank the anonymous referee for their constructive comments that improved the quality of this paper.

FUNDING

The first author was funded in part by ARC Consolidator grant from ULB and FNRS Grant CDR/OL J.0197.20 as well as EPSRC Grant EP/T018445/1.

The second author was funded in part by EPSRC grant EP/Y008650/1.

REFERENCES

- [1] Åberg, S. and Podgórski, K. A class of non-Gaussian second order random fields. *Extremes* **14** (2011), 187–222.
- [2] Åberg, S., Podgórski, K. and Rychlik, I. Fatigue damage assessment for a spectral model of non-Gaussian random loads. *Probabilist. Eng. Mech.* **24** (2009), 608–617.
- [3] Aguilar, J. P. Some pricing tools for the Variance Gamma model. *Int. J. Theor. Appl. Financ.* **23** (2020), 2050025.
- [4] Anderson, T. W. *An Introduction to Multivariate Statistical Analysis* (3rd ed.). Hoboken, N. J.: Wiley Interscience, 2003.
- [5] Arras, B., Azmoodeh, E., Poly, G. and Swan, Y. A bound on the 2-Wasserstein distance between linear combinations of independent random variables. *Stoch. Proc. Appl.* **129** (2019), 2341–2375.
- [6] Avramidis, A. N. and L'Ecyer, P. Efficient Monte Carlo and quasi-Monte Carlo option pricing under the Variance-Gamma model. *Manage. Sci.* **52** (2006), 1930–1944.
- [7] Azmoodeh, E., Eichelsbacher, P. and Thäle, C. Optimal Variance-Gamma approximation on the second Wiener chaos. *J. Funct. Anal.* **282** (2022), Art. 109450.
- [8] Bai, S. and Taqqu, M. S. Behavior of the generalized Rosenblatt process at extreme critical exponent values. *Ann. Probab.* **45** (2017), 1278–1324.
- [9] Bakshi, G., Chen, C. and Cao, Z. Empirical Performance of Alternative Option Pricing Models. *J. Financ.* **52** (1997), 2003–2049.
- [10] Barndorff-Nielsen, O. E. Exponentially decreasing distributions for the logarithm of particle size. *Proc. Roy. Soc. London A* **353** (1977), 401–419.
- [11] Barndorff-Nielsen, O. E. Hyperbolic distributions and distributions on hyperbolae. *Scand. J. Stat.* **5** (1978), 151–157.
- [12] Barndorff-Nielsen, O. E., Kent, J. and Sørensen, M. Normal Variance-Mean Mixtures and z -Distributions. *Int. Stat. Rev.* **50** (1982), 145–159.
- [13] Baxevani, A., Podgórski, K. and Wegener, J. Sample Path Asymmetries in Non-Gaussian Random Processes. *Scand. J. Stat.* **41** (2014), 1102–1123.

- [14] Bee, M., Dickson, M. M. and Santi, F. Likelihood-based risk estimation for variance-gamma models. *Stat. Method. Appl.* **27** (2018), 69–89.
- [15] Bibby, B. M. and Sørensen, M. Hyperbolic Processes in Finance. In: Rachev, S. (Ed.), *Handbook of Heavy Tailed Distributions in Finance*. Elsevier Science, Amsterdam (2003), 211–248.
- [16] Breyermann, W. and Lüthi, D. ghyp: A package on generalized hyperbolic distributions. Manual for R Package ghyp, 2013.
- [17] Bogsjö, K., Podgórski, K. and Rychlik, I. Models for road surface roughness. *Veh. Syst. Dyn.* **50** (2012), 725–747.
- [18] Bolin, D. Spatial Matérn Fields Driven by Non-Gaussian Noise. *Scand. J. Stat.* **41** (2014), 557–579.
- [19] Cervellera, G. P. and Tucci, M. P. A note on the Estimation of a Gamma-Variance Process: Learning from a Failure. *Comput. Econ.* **49** (2017), 363–385.
- [20] Carr, P., Geman, H., Madan, D. and Yor, M. The Fine Structure of Asset Returns: An Empirical Investigation. *J. Bus.* **75** (2002), 305–332.
- [21] Carr, P. and Madan, D. Option valuation using the fast Fourier transform. *J. Comput. Financ.* **2** (1999), 61–73.
- [22] Cui, G., Yu, X., Iommelli, S. and Kong, L. Exact Distribution for the Product of Two Correlated Gaussian Random Variables. *IEEE Signal Process. Lett.* **23** (2016), 1662–1666.
- [23] Daal, E. A. and Madan, D. B. An Empirical Examination of the Variance-Gamma Model for Foreign Currency Options. *J. Bus.* **78** (2005), 2121–2152.
- [24] Devroye, L. *Non-Uniform Random Variate Generation*. New York: Springer-Verlag, 1986.
- [25] Eberlein, E. and Hammerstein E. Generalized Hyperbolic and Inverse Gaussian Distributions: Limiting Cases and Approximation of Processes. In: Dalang, R. C. Dozzi, M. Russo, F. (Eds.), *Seminar on Stochastic Analysis, Random Fields and Applications IV*, in: *Progress in Probability* **58** Birkhäuser Verlag, (2004), 105–153.
- [26] Eberlein, E. and Keller, U. Hyperbolic distributions in finance. *Bernoulli* **1** (1995), 281–299.
- [27] Eberlein, E. and Prause, K. The generalized hyperbolic model: financial derivatives and risk measures. In: Geman, H., Madan, D. B., Pliska, S. and Vorst, T. Eds., *Mathematical finance – Bachelier Congress 2000* Springer, Berlin, (2001), 245–267.
- [28] Eichelsbacher, P. and Thäle, C. Malliavin-Stein method for Variance-Gamma approximation on Wiener space. *Electron. J. Probab.* **20** no. 123 (2015), 1–28.
- [29] Embrechts, P. A property of the generalized inverse Gaussian distribution with some applications. *J. Appl. Probab.* **20** (1983), 537–544.
- [30] Ferguson, T. S. and Klass, M. J. A representation of independent increment processes without Gaussian components. *Ann. Math. Statist.* **43** (1972), 1634–1643.
- [31] Finlay, R. and Seneta, E. Stationary-Increment Student and Variance-Gamma Processes. *J. Appl. Probab.* **43** (2006), 441–453.
- [32] Finlay, R. and Seneta, E. Stationary-Increment Variance-Gamma and t Models: Simulation and Parameter Estimation. *Int. Stat. Rev.* **76** (2008), 167–186.
- [33] Fiorani, F. *Option Pricing Under the Variance Gamma Process*. PhD thesis, Università Degli Studi Di Trieste, 2004.
- [34] Fu, M. C. Variance-Gamma and Monte Carlo. In: Fu, M. C. Jarrow, R. A., Yen, J.-Y. J. and Elliot, R. J. (Eds.), in: *Advances in Mathematical Finance* Birkhäuser Boston, (2007), 21–34.
- [35] Gaunt, R. E. *Rates of Convergence of Variance-Gamma Approximations via Stein’s Method*. DPhil thesis, University of Oxford, 2013.
- [36] Gaunt, R. E. Variance-Gamma approximation via Stein’s method. *Electron. J. Probab.* **19** no. 38 (2014), 1–33.
- [37] Gaunt, R. E. Inequalities for modified Bessel functions and their integrals. *J. Math. Anal. Appl.* **420** (2014), 373–386.
- [38] Gaunt, R. E. A Stein characterisation of the generalized hyperbolic distribution. *ESAIM: Probab. Stat.* **21** (2017), 303–316.
- [39] Gaunt, R. E. Inequalities for integrals of modified Bessel functions and expressions involving them. *J. Math. Anal. Appl.* **462** (2018), 172–190.
- [40] Gaunt, R. E. A note on the distribution of the product of zero mean correlated normal random variables. *Stat. Neerl.* **73** (2019), 176–179.
- [41] Gaunt, R. E. Stein factors for variance-gamma approximation in the Wasserstein and Kolmogorov distances. *J. Math. Anal. Appl.* **514** (2022), Art. 126274.
- [42] Gaunt, R. E. The basic distributional theory for the product of zero mean correlated normal random variables. *Stat. Neerl.* **76** (2022), 450–470.
- [43] Gaunt, R. E. On the moments of the variance-gamma distribution. *Stat. Probabil. Lett.* **201** (2023), Art. 109884.
- [44] Gaunt, R. E. On the cumulative distribution function of the variance-gamma distribution. To appear in *B. Aust. Math. Soc.*, 2023+.
- [45] Gaunt, R. E. and Li, S. Bounding Kolmogorov distances through Wasserstein and related integral probability metrics. *J. Math. Anal. Appl.* **522** (2023), Art. 126985.
- [46] Gaunt, R. E. and Li, S. The variance-gamma ratio distribution. *C. R. Math.* **361** (2023), 1151–1161.
- [47] Gaunt, R. E. and Li, S. The distribution of the product of independent variance-gamma random variables. arXiv:2401.17446, 2024.
- [48] Gaunt, R. E. and Merkle, M. On bounds for the mode and median of the generalized hyperbolic and related distributions. *J. Math. Anal. Appl.* **493** (2021), Art. 124508, 1–19.
- [49] Gaunt, R. E., Mijoule, G. and Swan Y. An algebra of Stein operators. *J. Math. Anal. Appl.* **469** (2019), 260–279.
- [50] Gradshteyn, I. S. and Ryzhik, I. M. *Table of Integrals, Series and Products*, 7th ed. Academic Press, 2007.
- [51] Groeneveld, R. A. and Meeden, G. The Mode, Median, and Mean Inequality. *Amer. Stat.* **31** (1977), 120–121.
- [52] Gu, M., Wang, X. and Berger, J. O. Robust Gaussian stochastic process emulation. *Ann. Stat.* **46** (2018), 3038–3066.
- [53] Guttorp, P. and Gneiting, T. Studies in the history of probability and statistics XLIX on the Matérn correlation family. *Biometrika* **93** (2006), 989–995.
- [54] Hammerstein, E. A. V. *Generalized hyperbolic distributions: Theory and applications to CDO pricing*. PhD thesis, University of Freiburg, 2010.
- [55] Heyde, C. C. A Risky Asset Model with Strong Dependence through Fractal Activity Time. *J. Appl. Probab.* **36** (1999), 1234–1239.
- [56] Hirska, A. and Madan, D. B. Pricing American options under variance gamma. *J. Comput. Financ.* **7** (2004), 63–80.
- [57] Holm, H. and Alouini, M.-S. Sum and Difference of two squared correlated Nakagami variates with the McKay distribution. *IEEE T. Commun.* **52** (2004), 1367–1376.
- [58] Ivanov, R. V. Option pricing in the variance-gamma model under the drift jump. *Int. J. Theor. Appl. Financ.* **21** (2018), 1–19.
- [59] Jankov Maširević, D. and Pogány, T. K. On new formulae for cumulative distribution function for McKay Bessel distribution. *Commun. Stat. Theory* **50** (2021), 143–160.
- [60] Johannesson, P., Podgórski, K., Rychlik, I. and Shariati, N. AR(1) time series with autoregressive gamma variance for road topography modeling. *Probabilist. Eng. Mech.* **43** (2016), 106–116.

- [61] Kollo, T. *Advanced Multivariate Statistics with Matrices*. Springer, 2005.
- [62] Koponen, I. (1995) Analytic approach to the problem of convergence of truncated Lévy flights towards the Gaussian stochastic process. *Phys. Rev. E* **52**, 1197–1199.
- [63] Korn, R., Korn, E. and Kroisandt, G. *Monte Carlo Methods and Models in Finance and Insurance*. Boca Raton, Fla.: Chapman and Hall/CRC, 2010
- [64] Kotz, S., Kozubowski, T. J. and Podgórski, K. *The Laplace Distribution and Generalizations: A Revisit with Applications to Communications, Economics, Engineering, and Finance*. Springer, 2001.
- [65] Kozubowski, T. J., Mazur, S. and Podgórski, K. Matrix Variate Generalized Laplace Distributions. Working Papers 2022:7, Örebro University, School of Business.
- [66] Kozubowski, T. J. and Podgórski, K. Asymmetric Laplace Laws and Modeling Financial Data. *Math. Comput. Model.* **34** (2001), 1003–1021.
- [67] Kozubowski, T. J. and Podgórski, K. Laplace probability distributions and related stochastic processes. In: Shmaliy, Y., Ed. *Probability: Interpretation, Theory and Applications*. New York: Nova Science Publishers, Inc., (2012), pp. 105–145.
- [68] Kozubowski, T. J., Podgórski, K. and Rychlik, I. Multivariate generalized Laplace distribution and related random fields. *J. Multivariate Anal.* (2013), 59–72.
- [69] Küchler, U. and Tappe, S. Tempered stable distributions and processes. *Stoch. Proc. Appl.* **123** (2013), 4256–4293.
- [70] Kyprianou, A. E. *Fluctuations of Lévy Processes with Applications: Introductory Lectures*. Springer Science & Business Media, 2014.
- [71] Lam, K., Chang, E. and Lee, M. C. Option pricing in the variance-gamma model under the drift jump. *Pac.-Basin Financ. J.* **10** (2002), 267–285
- [72] Leonenko, N. N., Petherick, S. and Sikorskii, A. The Student Subordinator Model with Dependence for Risky Asset Returns. *Commun. Stat. Theory* **40** (2011), 3509–3523.
- [73] Linders, D. and Stassen, B. The multivariate Variance Gamma model: basket option pricing and calibration. *Quant. Financ.* **16** (2016), 555–572.
- [74] Loregian, A., Mercuri, L. and Rroji, E. Approximation of the variance gamma model with a finite mixture of normals. *Stat. Probabil. Lett.* **82** (2012), 217–224.
- [75] Luciano, E. and Schoutens, W. A multivariate jump-driven financial asset model. *Quant. Financ.* **6** (2006), 385–402.
- [76] Luciano, E. and Semeraro, P. Multivariate time changes for Lévy asset models: Characterization and calibration. *J. Comput. Appl. Math.* **233** (2010), 1937–1953.
- [77] Luciano, E., Marena, M. and Semeraro, P. Dependence calibration and portfolio fit with factor-based subordinators. *Quant. Financ.* **16** (2016), 1037–1052.
- [78] Ma, P. and Bhadra, A. Beyond Matérn: On A Class of Interpretable Confluent Hypergeometric Covariance Functions. *J. Am. Stat. Assoc.* **118** (2023), 2045–2058.
- [79] Madan, D. B., Carr, P. and Chang, E. C. The variance gamma process and option pricing. *Eur. Finance Rev.* **2** (1998), 74–105.
- [80] Madan, D. B. and Milne, F. Option pricing with V.G. martingale components. *Math. Financ.* **1** (1991), 39–55.
- [81] Madan, D. B. and Seneta, E. Simulation of Estimates Using the Empirical Characteristic Function. *Int. Stat. Rev.* **55** (1987), 153–161.
- [82] Madan, D. B. and Seneta, E. Chebyshev Polynomial Approximation and Characteristic Function Estimation. *J. Roy. Stat. Soc. B* **49** (1987), 163–169.
- [83] Madan, D. B. and Seneta, E. The Variance Gamma (V.G.) Model for Share Market Returns. *J. Bus.* **63** (1990), 511–524.
- [84] Maejima, M. and Tudor, C. A. Selfsimilar processes with stationary increments in the second Wiener chaos. *Probab. Math. Stat.-Pol.* **32** (2012), 167–186.
- [85] Mangilli, A., Plaszczyński, S. and Tristram, M. (2015). Large-scale cosmic microwave background temperature and polarization cross-spectra likelihoods. *Mon. Not. R. Astron. Soc.* **453** (2015), 3174–3189.
- [86] Matérn, B. *Spatial Variation*, Meddelanden fran Statens Skogsforskningsinstitut, 49, 5. Second ed. (1986), Lecture Notes in Statistics 36, New York: Springer, 1960.
- [87] McKay, A. T. A Bessel function distribution. *Biometrika* **24** (1932), 39–44.
- [88] McLeish, D. L. A robust alternative to the normal distribution. *Can. J. Statist.* **10** (1982), 89–102.
- [89] McNicholas, S. M., McNicholas, P. D. and Browne, R. P. Mixtures of variance-gamma distributions. arXiv:1309.2695, 2013.
- [90] Nadarajah, S. and Kotz, S. The Bessel ratio distribution. *C. R. Acad. Sci. Paris, Ser. I* **343** (2006) 531–534.
- [91] Nadarajah, S. and Pogány, T. K. On the distribution of the product of correlated normal random variables. *C.R. Acad. Sci. Paris, Ser. I* **354** (2016), 201–204.
- [92] Nadarajah, S., Srivastava, H. M. and Gupta, A. K. Skewed Bessel function distributions with application to rainfall data. *Statistics* **41** (2007), 333–344.
- [93] Nitithumbundit, T. and Chan, J. S. An ECM algorithm for skewed multivariate variance gamma distribution in normal mean-variance representation. arXiv:1504.01239, 2015.
- [94] Nitithumbundit, T. and Chan, J. S. ECM algorithm for autoregressive multivariate skewed variance gamma model with unbounded density. *Methodol. Comput. Appl.* **22** (2020), 1169–1191.
- [95] Nourdin, I. and Peccati, G. Normal approximations with Malliavin calculus: from Stein’s method to universality. Vol. 192. Cambridge University Press, 2012.
- [96] Nourdin, I. and Peccati, G. The optimal fourth moment theorem. *Proc. Amer. Math. Soc.* **143** (2015), 3123–3133.
- [97] Nualart, D. and Peccati, G. Central limit theorems for sequences of multiple stochastic integrals. *Ann. Probab.* **33** (2005), 177–193.
- [98] Nzokem, A. H. Gamma variance model: Fractional fourier transform (FRFT). In *Journal of Physics: Conference Series* (Vol. 2090, No. 1, p. 012094). IOP Publishing, 2021.
- [99] Olver, F. W. J., Lozier, D. W., Boisvert, R. F. and Clark, C. W. *NIST Handbook of Mathematical Functions*. Cambridge University Press, 2010.
- [100] Pearson, K., Jefferey, G. B. and Elderton, E. M. On the distribution of the first product moment-coefficient, in samples drawn from an indefinitely large normal population. *Biometrika* **21** (1929), 164–193.
- [101] Pearson, K., Stouffer, S. A. and David, F. N. Further applications in statistics of the $T_m(x)$ Bessel function. *Biometrika* **24** (1932), 293–350.
- [102] Podgórski, K. and Wegener, J. Estimation for Stochastic Models Driven by Laplace Motion. *Commun. Stat. Theory* **40** (2011), 3281–3302
- [103] Press, S. J. On the sample covariance from a bivariate normal distribution, *Ann. Inst. Statist. Math.* **19** (1967), 355–361.
- [104] Ramachandran, B. On geometric stable laws, a related property of stable processes, and stable densities. *Ann. Inst. Statist. Math.* **49** (1997), 299–313.
- [105] Rathgeber, A. W., Johannes, S. and Stöckl, S. Modeling share returns - an empirical study on the Variance Gamma model. *J. Econ. Finance* **40** (2016), 653–682.

- [106] Reed, W. J. The Normal-Laplace Distribution and Its Relatives. *Advances in Distribution Theory, Order Statistics, and Inference* (2006). 61–74.
- [107] Reed, W. J. Brownian–Laplace Motion and Its Use in Financial Modelling. *Commun. Stat. Theory* **36** (2007), 473–484.
- [108] Ruiz-Antolín, D. and Segura, J. A new type of sharp bounds for ratios of modified Bessel functions. *J. Math. Anal. Appl.* **443** (2016), 1232–1246.
- [109] Rydberg, T. H. Generalized hyperbolic diffusions with applications towards finance. *Math. Financ.* **9**, 183–201.
- [110] Sato, K. *Lévy Processes and Infinitely Divisible Distributions*. Cambridge: Cambridge University Press, 1999.
- [111] Scott, D. and Dong, C. Y. Package ‘VarianceGamma’, 2018.
- [112] Scott, D. J., Würtz, D., Dong, C. and Tran, T. T. Moments of the generalized hyperbolic distribution. *Computation. Stat.* **26** (2011), 459–476.
- [113] Segura, J. Bounds for ratios of modified Bessel functions and associated Turán-type inequalities. *J. Math. Anal. Appl.* **374** (2011), 516–528.
- [114] Semeraro, P. A multivariate variance gamma model for financial applications. *Int. J. Theor. Appl. Financ.* **11** (2008), 1–18.
- [115] Seneta, E. Fitting the Variance-Gamma Model to Financial Data. *J. Appl. Probab.* **41** (2004), 177–187.
- [116] Seneta, E. The early years of the Variance-Gamma process. In: Fu, M. C., Jarrow, R. A., Yen, J. J. and Elliott, R. J. (Eds.), *Advances in Mathematical Finance*. (2007), Birkhauser, Boston, 3–19.
- [117] Sichel, H. S. Statistical valuation of diamondiferous deposits. *J. S. Afr. Inst. Min. Metall.* **73** (1973), 235–243.
- [118] Stein, C. A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. In *Proc. Sixth Berkeley Symp. Math. Statist. Prob.* (1972), vol. 2, Univ. California Press, Berkeley, 583–602.
- [119] Stein, M. L. *Interpolation of Spatial Data: Some Theory for Kriging*. Springer Science & Business Media, New York, NY, 1999.
- [120] Taqqu, M. S. Weak convergence to fractional Brownian motion and to the Rosenblatt process. *Probab. Theory Rel.* **31** (1975), 287–302.
- [121] Teichrow, D. The mixture of normal distributions with different variances. *Ann. Math. Statist.* **28** (1957), 510–512.
- [122] Tjetjep, A. and Seneta, E. Skewed normal variance-mean models for asset pricing and the method of moments. *Int. Stat. Rev.* **74** (2008), 109–126.
- [123] Tomy, L. and Jose, K. K. Generalized normal-Laplace AR process. *Statist. Probabil. Lett.* **79** (2009), 1615–1620.
- [124] Trindade, A. A., Zhu, Y. and Andrews, B. Time series models with asymmetric Laplace innovations. *J. Stat. Compu. Sim.* **80** (2010), 1317–1333.
- [125] van Zwet, W. R. Mean, median, mode II. *Stat. Neerl.* **33** (1979), 1–5.
- [126] Weisstein, E. W. Chi-Squared Distribution. From MathWorld—A Wolfram Web Resource, <https://mathworld.wolfram.com/Chi-SquaredDistribution.html>. Last visited on 09/03/2023.
- [127] Wu, F. *Applications of The Normal Laplace and Generalized Normal Laplace Distributions*. MSc thesis, University of Victoria, 2008.
- [128] Yamazato, M. Unimodality of infinitely divisible distribution functions of class L. *Ann. Probab.* **6** (1978), 523–531.