UNIVERSITÀ DI SIENA
1240

QUADERNI DEL DIPARTIMENTO
DI ECONOMIA POLITICA E STATISTICA

Gian P. Cervellera
Marco P. Tucci

A note on the estimation of a Gamma-Variance process: Learning from a failure

n. 702 – Ottobre 2014

A note on the estimation of a Gamma-Variance process: Learning from a failure

Gian P. Cervellera, Marco P. Tucci

Dipartimento di Economia Politica e Statistica, Università di Siena, Piazza S. Francesco, 7 - 53100 Siena, Italy

Abstract

This paper confirms that, as originally reported in Seneta (2004, p. 183), it is impossible to replicate Madan *et al.*'s (1998) results using log daily returns on S&P 500 Index from January 1992 to September 1994. This failure leads to a close investigation of the computational problems associated with finding maximum likelihood estimates of the parameters of the popular VG model. Both standard econometric software, such as *R*, and non-standard optimization software, such as *Ezgrad* described in Tucci (2002), are used. The complexity of the log-likelihood function is studied. It is shown that it looks very complicated, with many local optima, and may be incredibly sensitive to very small changes in the sample used. Adding or removing a single observation may cause huge changes both in the maximum of the log-likelihood function and in the estimated parameter values.

Key words: Variance-Gamma, log stock returns, maximum likelihood estimation, globally optimizing procedures.

JEL Classification: C58, C61, C63.

1 Introduction

Since Black and Scholes (1973) seminal article on option pricing, there has been a(n exponentially) growing literature on modeling both asset returns (log price increments) and option prices. Black and Scholes model implies

Email addresses: cervellera@unisi.it (Gian P. Cervellera), tucci@unisi.it (Marco P. Tucci).

that log price returns are identically and independently distributed as a normal random variable. When used for option pricing, given the assumption of constant parameters in the model, "a minimal prediction of the Black-Scholes formula is that all options on the same underlying asset with the same time-to-expiration but with different striking prices should have the same implied volatility" (Rubinstein, 1994, p. 4).

Both implications appear to be contradicted by the empirical evidence. Daily log-returns are typically characterized by distributions with kurtosis greater than three. In other words, higher peaks above the mean and thicker tails, or more frequent extreme events, than a normal distribution. Usually, there is little or no autocorrelation in returns, at least for one or two lags, but it is present a long range dependence structure in squared and absolute returns, violating the independence assumption. Then, returns often show periods of low variation followed by periods of higher variation. Namely, they display varying volatility, or heteroscedasticity.[1] In the options market, at the end of the 1980s the graph of implied volatility as a function of striking price, for otherwise identical options, begins to depart from a horizontal line, maybe as a consequence of the stock market crash of October 1987.[2] Since then volatility smiles and/or volatility smirks (either forward skew or reverse skew) are common across option markets as remarked in a countless number of studies.

To overcome these problems, many models have been proposed. The Variance-Gamma (VG) process is clearly one of the most popular. Introduced as a model for asset returns in Madan and Seneta (1990), it has then been generalized to a non-symmetric process in Madan *et al.* (1998) and its use extended to option pricing. It belongs to the family of Lévy processes of infinite activity and "like the Poisson process, the VG process is pure jump ... and ... it can be expressed in terms of its Lévy density" (Fu, 2007, p. 23).[3] It is obtained by evaluating a Brownian motion at a random time given by a gamma process. Random time that allows to model the flow of "economically relevant time", or "market activity time", reflecting the random speedups and slowdowns in real-time economic and business activity. In other words, "the more share trades that occur, or the more information released to the market on a given day, the faster 'time' progresses" (Finlay, 2009, pp. 10-11).[4] "Under this process, the unit period continuously compounded return is normally distributed, conditional on the realization of a random time" (Madan *et al.*, 1998, p. 80), with condi-

[1] See, e.g., Heyde and Liu (2001), and the references therein cited. Finlay (2009, pp. 2-3) provides a nice summary of the main points.
[2] See, e.g., Rubinstein (1994).
[3] See the apppendix in Fu (2007) for a short review of the basic definitions of the Wiener, Poisson, Gamma and Lévy processes. Process of infinite activity indicates that the paths jump infinitely many times, for each finite interval and jumps that are larger than a given quantity occur only a finite number of times.
[4] See also Madan and Seneta (1990, p. 517) for some additional intuition.

tional variance given by a gamma random variable, from here the name VG. [5] The resulting stochastic process has three (four when the location parameter is included) parameters. In addition to the volatility parameter characterizing the Brownian motion, there is a parameter that determines the percentage excess kurtosis in the log return distribution (i.e. a measure of the symmetric increase in the left and right tail probabilities of the distribution compared to the normal distribution) and one for skewness that allows for asymmetry of the left and right tails of that density. Very conveniently, for testing purposes, it nests the lognormal density and the Black-Scholes formula as a parametric special case.

The VG model has been reported to perform better than the Black-Scholes, or geometric Brownian motion (GBM), model in a number of empirical studies. Madan *et al.* (1998) show that, once calibrated to the market prices, it captures volatility smile and fat-tailness of the asset return distribution. Among the others, Daal and Madan (2005) use it in pricing foreign currency options, Fiorani (2003) in pricing European and American options on S&P 500, Fiorani and Luciano (2006) and Hurd (2007) in credit risk modeling. Given that analytical solutions are available only for European-style options, a lot of attention has been devoted to implementing efficient numerical methods to evaluate other kinds of options. [6] The progress in this direction and the nice properties of the VG model have led to its implementation in the Bloomberg system through the function SKEW as described in Stein *et al.* (2007). Recently, VG correlated models have been introduced in Madan and Khanna (2009) and Eberlein and Madan (2010) and versions of the multivariate VG model applied in a large scale application (Wallmeier and Diethelm, 2012). New applications of the VG model include its use to study the magnitude of the overpricing of the so-called reverse convertibles in Deng *et al.* (2013) and, in a regime-switching framework, to value a new type of insurance contract for near to retirement population in Fard and Rong (2014).

So far, the estimation problem has not received much attention. Some papers present only few estimated parameters for a small, selected empirical database. Others, see e.g. Finlay and Seneta (2008), use simulated data. Only recently, a few works report estimates based on a broad data set. Rathgeber *et al.* (2013) consider daily data on the US companies listed in the Dow Jones for the period 01/01/1991-12/31/2011. After applying a regime switching model in order to identify normal and turbulent times within the data set at hand, they estimate the VG model for log returns for the various periods with six alternative estimation methods. In another study, Le Courtois and Walter (2014) estimate

[5] See Madan *et al.* (1998, pp. 80-81) for a summary of the main differences between Brownian motion and a VG process.
[6] See, e.g., Ribeiro and Webber (2003), Almendral and Oosterlee (2007), Kaishev and Dimitrova (2009) and the references therein cited.

the VG parameters on the returns of top French companies and of the French CAC40 Index, over the period 01/03/01-04/15/09, by maximum likelihood (ML). In both cases the authors, consistent with most of the literature on this topic, do not report any computational problems. Does this mean that usually, apart from the occasional need to distinguish between normal and turbulent periods, the estimation of parameters in a VG model for daily log returns is computationally straightforward? And that, as reported in Figueroa-López *et al.* (2012, p. 19), some problems may be encountered only in finding ML estimates for a VG model applied to very high frequency data (i.e. observations distanced 30 minutes or less)? Can practitioners use a generally available, well documented econometric software, instead of having to write to one of the cited authors for their individually implemented algorithms, to estimate their own VG model?

Some hints of caution about the 'computational straightforward*ness*' can be found in Seneta (2004, pp. 180-183). First, he points out that care has to be taken in computing ML estimates with packages such as Matlab® which use implicit functions. Then, in discussing a couple of empirical examples dealing with log daily price differences, he states that "when the original data... were used ... attempts to find ... (the ML estimates of the parameters) were unsuccessful." Moreover, since April 2012, it is possible to estimate the parameters of a VG model by using the, generally available, 'VarianceGamma' *R* package documented in Scott and Yang Dong (2012). The interested users can compute the parameters of a VG model simply calling the *vgFit* function incorporated in this package for their data set. The goal of this paper is to shed some more light on the computational complexity of estimating a VG model for daily log returns. This is done by reporting the author's experience in trying to replicate the ML estimates presented in Madan *et al.* (1998, p. 90, Table 1).

The discussion is organized as follows. After a brief introduction of the VG model (Section 2), the problems encountered in replicating Madan *et al.*'s (1998) results using daily returns on S&P 500 Index from January 1992 to September 1994 are described in Section 3. At this stage both standard econometric software, such as the *vgFit* command available in the *R* package, and non-standard optimization software, such as *Ezgrad* described in Tucci (2002), are used to find the optimum of the log-likelihood function. As in the original reference, the increments in prices are assumed independent so classical statistical procedures for estimation and hypothesis testing are readily applicable. Section 4 describes how sensitive the parameter estimates can be to very small changes of the estimation sample. Then the performances of several procedures using *R* functions are compared on a big data set (Section 5). The best performing procedure is applied to Madan *et al.*'s (1998) sample in Section 6. The main conclusions are summarized in Section 7.

2 The VG model

The standard model for asset price movements, namely the Black-Scholes or GBM model, assumes that the price of an asset at time $t \geq 0$ is given by

$$P_t = P_0 \exp\left(\theta t + \sigma B\left(t\right)\right) \tag{1}$$

for $\theta \in \mathbb{R}$ and $\sigma > 0$ with $\{B(t)\}$ a standard Brownian motion, or Wiener process. Here θ and σ correspond to the drift and the volatility coefficient of the Brownian motion, respectively. Log price increments, i.e. continuously compounded returns, are then given by

$$X_t = \ln(P_t) - \ln(P_{t-1}) = \theta + \sigma\left(B\left(t\right) - B\left(t-1\right)\right) \tag{2}$$

which implies that returns are independent and identically distributed (iid) normal random variables.

Alternatively, the VG model assumes that the price of an asset at time $t \geq 0$ "is obtained by evaluating Brownian motion (with constant drift and volatility) at a random time given by a gamma process" (Madan *et al.*, 1998, p. 82) independent of $\{B(t)\}$.[7] This random time is the "market activity time", denoted here by $\{T_t\}$, introduced in the previous section. In this case, log price increments, i.e. continuously compounded returns, are given by

$$X_t = \ln(P_t) - \ln(P_{t-1}) = \theta\tau_t + \sigma\left(B\left(T_t\right) - B\left(T_{t-1}\right)\right) \tag{3}$$

where the increments $\tau_t = T_t - T_{t-1}$ are independently, gamma distributed random variables over non-overlapping intervals of time, i.e.

$$f_\Gamma\left(\tau_t; \alpha, \lambda\right) = \frac{1}{\lambda^\alpha \Gamma\left(\alpha\right)} \tau_t^{\alpha-1} e^{-\lambda\tau_t} \tag{4}$$

for $\tau_t > 0$, $\Gamma\left(\bullet\right)$ the gamma function and $\alpha = 1/\lambda = 1/\nu$ then $E\left(\tau_t\right) = 1$ and $Var\left(\tau_t\right) = \nu$. Setting $E\tau_t = 1$ so as "to make the expected activity time change over unit calendar time equal to one unit" (Seneta, 2004, p. 177) does not imply any loss of generality. As pointed out in the just cited work, as long as the increments have a finite mean, any scaling can be absorbed into the parameters θ and σ given that

$$\sigma\left(B\left(T_t\right) - B\left(T_{t-1}\right)\right) \overset{D}{=} \sigma\tau_t^{1/2} B\left(1\right). \tag{5}$$

[7] This is sometimes referred to as a subordinated GBM process. The idea of subordination was developed by Bochner (1955) and subordinated processes were first considered for stock prices by Clark (1973). See, e.g., Finlay (2009, p. 12) for more details.

Generally, asset prices include a component associated with the calendar time t and Eqt. (3) is rewritten as [8]

$$
\begin{aligned}
X_t = \ln(P_t) - \ln(P_{t-1}) &= \mu + \theta\tau_t + \sigma\left(B\left(T_t\right) - B\left(T_{t-1}\right)\right) \\
&\stackrel{D}{=} \mu + \theta\tau_t + \sigma\tau_t^{1/2} B\left(1\right)
\end{aligned} \tag{6}
$$

with $\mu \in \mathbb{R}$. [9] It follows that $EX_t = \mu + \theta$ and, given that the conditional distribution of X_t given τ_t is $N(\mu + \theta\tau_t, \sigma^2\tau_t)$, [10]

$$E\left(X_t - EX_t\right)^2 = \sigma^2 + \theta^2\nu \tag{7}$$

$$E\left(X_t - EX_t\right)^3 = 3\theta\sigma^2\nu + 2\theta^3\nu^2 \tag{8}$$

and

$$E\left(X_t - EX_t\right)^4 = 3\sigma^4\left(1+\nu\right) + 6\theta^2\sigma^2\left(\nu + 2\nu^2\right) + 3\theta^4\left(\nu^2 + 2\nu^3\right) \tag{9}$$

with skewness and kurtosis of the return distribution given by

$$\beta = \frac{3\theta\sigma^2\nu + 2\theta^3\nu^2}{\left(\sigma^2 + \theta^2\nu\right)^{3/2}} \tag{10}$$

and

$$\kappa = \frac{3\sigma^4\left(1+\nu\right) + 6\theta^2\sigma^2\left(\nu + 2\nu^2\right) + 3\theta^4\left(\nu^2 + 2\nu^3\right)}{\left(\sigma^2 + \theta^2\nu\right)^2}, \tag{11}$$

respectively.

Combining the distribution for τ_t with Eqt. (6) results in X_t having the marginal (skew) VG distribution (Madan *et al.*, 1998) with probability density function (pdf)

$$
\begin{aligned}
f_{VG}\left(X_t\right) = \sqrt{\frac{2}{\pi}}\frac{e^{\theta(X_t-\mu)/\sigma^2}}{\sigma\nu^{1/\nu}\Gamma\left(\frac{1}{\nu}\right)}\left[\frac{\left|X_t - \mu\right|}{\sqrt{\theta^2 + 2\left(\sigma^2/\nu\right)}}\right]^{\frac{1}{\nu}-\frac{1}{2}} \\
\times K_{\frac{1}{\nu}-\frac{1}{2}}\left(\frac{\left|X_t - \mu\right|\sqrt{\theta^2 + 2\left(\sigma^2/\nu\right)}}{\sigma^2}\right)
\end{aligned} \tag{12}
$$

[8] The VG process can be viewed as a special case of the Generalised Hyperbolic distribution (Barndorff-Nielsen and Halgreen, 1977), because the Gamma distribution is a special case of the Generalised Inverse Gaussian distribution (Finlay, 2009, p. 21).
[9] The parameter μ in Eqt. (6) corresponds to m in Madan *et al.* (1998) and c in Seneta (2004).
[10] See, e.g., Seneta (2004). These formulae are slightly different in Madan *et al.* (1998) where the log price increments are relative to time 0.

for $X_t \in \mathbb{R}$ and characteristic function[11]

$$\phi_X(u) = e^{i\mu u}\left[1 - i\theta\nu u + \left(\sigma^2 \nu u^2/2\right)\right]^{-1/\nu}. \tag{13}$$

In Eqt. (12), the function $K_\eta(\omega)$ for $\eta \in \mathbb{R}$ and $\omega > 0$ given by

$$K_\eta(\omega) = \frac{1}{2}\int_0^\infty z^{\eta-1} e^{-\frac{\omega}{2}\left(z+\frac{1}{z}\right)} dz \tag{14}$$

is a modified Bessel function of the third kind (Erdélyi *et al.*, 1953) with index η, and the VG is sometimes also known as the Bessel K-function distribution (see, e.g., Johnson *et al.*, 1994, pp. 50–51).[12]

The VG distribution is sometimes denoted by $VG(\mu, \theta, \sigma, \nu)$. The four parameters (three when returns are centered) determining the VG process are: the location parameter μ, the volatility σ of the Brownian motion, the variance ν of the gamma distributed time and the drift θ of the time-changed Brownian motion with drift. The parameter θ measures the degree of skewness of the distribution and ν controls the excess of kurtosis with respect to the normal distribution. As noticed in Madan *et al.* (1998, p. 86), "$\theta = 0$ does indeed imply that there is no skewness, and furthermore the sign of the skewness is that of θ. Furthermore ... when $\theta = 0$, the fourth central moment divided by the squared second central moment is $3(1+\nu)$ and so ν is the percentage excess kurtosis in the distribution." In other words, ν measures the degree of "peakedness" with respect to the normal distribution. A large value of ν results in fat tails, which is observed in the empirical log-returns.

It is apparent that when $T_t = t$ the VG model boils down to the classical GBM model with a log normal distribution for prices and a normal distribution for returns. Equivalently, when $\theta = 0$ and $\nu \to 0+$ the time change is close to the linear time change and the pdf in Eqt. (12) reduces to that of the standard normal distribution.[13]

3 Computational and estimation problems: The failure

In the paper introducing the non-symmetric VG model, Madan *et al.* (1998, pp. 89-90) present their ML estimated parameter values for log daily returns on the S&P 500 Index from January 1992 to September 1994, for a total of

[11] See, e.g., Seneta (2004, p. 180).
[12] As observed in Seneta (2004, p. 180) there is some ambiguity in the terminology associated with this function. In some works it is referred to as a modified Bessel function of the second kind.
[13] See, e.g., Seneta (2004, p. 181).

691 observations (second column of Table 1).[14] This section confirms that, as originally reported in Seneta (2004, p. 183), it is impossible to replicate Madan *et al.*'s (1998) results.

Table 1
ML estimates of the VG parameters for daily returns on S&P 500 Index

Parameter	**Madan *et al.***	**various**	*vgFit*	**results***	**Best**
	(1998)	**(1)**	**(2)**	**(3)**	*Ezgrad*
μ	0.0591	0.0056	−0.0001	−0.0001	0.0003
σ	0.1172	0.099	7.E+14	2.201	0.0072
ν	0.002	0.1184	89.27	15.24	0.8771
θ	0.0048	−0.007	129	140.7	−0.0016
ln $\mathcal{L}$	2569.78	994.60	910.82	2023.55	2890.98
No. of obs.	691	691	691	691	691

* Column (1) is obtained with "default *vgFit*", i.e. no options selected, columns (2) and (3) with the optimization algorithms nlm and the BFGS, respectively, selected (see Scott and Yang Dong, 2012).

When the *vgFit* function available in *R* is used with no options selected (call it "default *vgFit*"), to estimate the parameters of a VG model for the above sample, it yields the results reported in column labelled (1) of Table 1. They greatly differ from the original ones. To check if this is due to a 'bad sample' for the default optimization algorithm, the alternative algorithms made available in *vgFit* are used and the associated results shown in columns labelled (2) and (3) of Table 1.[15] It is apparent that the three 'optimization versions' of *vgFit* give results very different from the original ones and different among themselves. What is more disturbing is that columns labelled (1) and (2) present fairly close log-likelihood values but very different values for the parameters. Similarly the best *vgFit* in terms of log-likelihood value present estimates very different from the original ones.

[14] The parameters μ and θ are called m and α, respectively, in Madan *et al.* (1998). The standard deviations of parameters are omitted from Tab. 1. At this stage, the focus is on the estimated parameter values and associated likelihood obtained with the various methods rather than on how significantly different from zero a certain parameter is.

[15] As documented in Scott and Yang Dong (2012, pp. 16-18), it is possible to choose different sets of starting values, namely user-supplied, based on a fitted skew-Laplace distribution or derived from the method of moments, for each optimization method selected in the *vgFit* function. It is worth it to point out that each optimization method used, i.e. BFGS, Nelder-Mead and nlm, ends up to the same optimum regardless of the selected set of starting values.

For this reason a non-standard optimization software, such as *Ezgrad* described in Tucci (2002) is also tried. The log-likelihood of the VG model is maximized over the parameter space

$$-0.50 < \theta < 0.50,\ 0.0001 < \sigma < 1.00,\ 0.0001 < \nu < 1.00,\ \ -0.10 < \mu < 0.10$$

with the critical parameter NDPTH set to 14 which means that a function of 2^{14} local optima is computed. *Ezgrad* finds the following candidate for the global optimum

$$\mu = 0.00029,\ \sigma = 0.02378,\ \nu = 0.75426,\ \theta = -0.00247$$

with log-Likelihood equal to 2177.76. When the multidimensional parameter space is modified (narrowed in some sense) to

$$-0.30 < \theta < 0.30,\ 0.0001 < \sigma < 0.30,\ 0.50 < \nu < 1.00,\ \ -0.10 < \mu < 0.10$$

with NDPTH=14 the result is

$$\mu = 0.00029,\ \sigma = 0.0072,\ \nu = 0.87712,\ \theta = -0.00158$$

with log-likelihood equal to 2890.98 reported in the last column of Table 1.

Simply plugging Madan *et al.*'s (1998, pp. 89-90) set of estimates into *vgFit* and *Ezgrad* (a Fortran code) to compute the associated log-likelihood value does not work. Both programs fail to go through the computation. Afraid that the problem is due to the number precision in *R* and to a poor implementation of Bessel and Gamma functions in *Ezgrad*, based on the numerical recipes in Press *et al.* (2007, 1997), the authors decide to independently code the VG log-likelihood function in Matlab®, where the two functions can be simply called as implicit functions. Again when Madan *et al.* (1998) estimates are used, the two new Matlab® codes explode. The good news is that these new codes, *R* and *Ezgrad* produce exactly the same numbers at double precision. This confirms that it is impossible to replicate Madan *et al.*'s (1998) results using log daily returns data as originally noticed in Seneta (2004, p. 183) where it is stated "we do not know the precise methodology ... used ... to obtain the ML estimates."

4 Sensitivity of the parameter estimates of a VG model to one single observation

At this point, one thing is crystal clear. The different results obtained, for the same data set, with the various optimization techniques and, sometimes, with different sets of starting points or parameter spaces are a clear indication that the likelihood function of a VG density function is highly complicated and it is easy to land on a local optimum. To see if the computational problems suffered in the previous section by "default *vgFit*", *Ezgrad* and the two Matlab® codes using implicit functions, are specific to the particular sample used the attention is now turned to a big data set. In this way it is more likely to incorporate different types of subperiods, for instance quiet and turbulent periods, to reflect a wide variety of real situations facing the researcher. The S&P 500 Index from Jan. 2nd 1992 to Aug. 20th 2012, for a total of 5200 observations of log daily returns seems an appropriate data set in this sense. It is used to compute ML estimates of the parameters for a moving window of 1000 observations. Namely, ML parameter estimates are computed for 4201 windows with the first one covering observations 1 through 1000, the second one observations 2 through 1001 and so on and so forth until the last one from observation 4201 to 5200. Even though the width of the moving window is arbitrary, only 1000 observations, it seems appropriate to find reasonably good estimates of the four parameters (three when the location parameter μ is excluded) characterizing a VG model regardless of the good or bad luck of the estimating researcher.

Before checking the whole set of results, it may be useful to briefly focus on the first two moving windows which differ only for the first observation being replaced by the 1001^{st} observation in the second window. Namely, the second window starts one day later and ends one day later, as all the 'following' windows in this exercise. The maximum value of the log-likelihood function drops from 4500 for the first window to 1500.[16] Similarly the estimates of the parameters change remarkably considering that the two samples differ by only one observation. That associated with the location parameter μ goes from -4.53 (1^{st} window) to -6.27, σ from .31 to .23 and θ from 9.7E-03 to 5.1E-03. Strangely enough, the estimate of the excess kurtosis parameter ν stays constant at 1 in both windows.

The same qualitative behavior can be observed when the analysis is extended to a few more sequential subsamples. The maximum value of the log-likelihood function associated with the first 27 moving windows obtained with the "default *vgFit*" are reported in Fig. 1. Both the maximum of the log-likelihood function and the parameter estimates considerably change when replacing a single observation out of 1000.In general, estimates vary up to 30/50% (Fig. 2). Strangely enough, that for ν stays constant at 1 for the initial 27 moving windows. At this point it is unclear if the results are due to a poorly implemented "default *vgFit*" function, from now on simply *vgFit,* or to the computational complexity of the VG model which translates into its extreme sensitivity to the estimation sample. For this reason several procedures using

[16] The selected optimization algorithm is Nelder-Mead in this and the following sections.

Fig. 1. Maximum value of the log-likelihood function with "default *vgFit*" (27 moving windows of 1000 obs.)

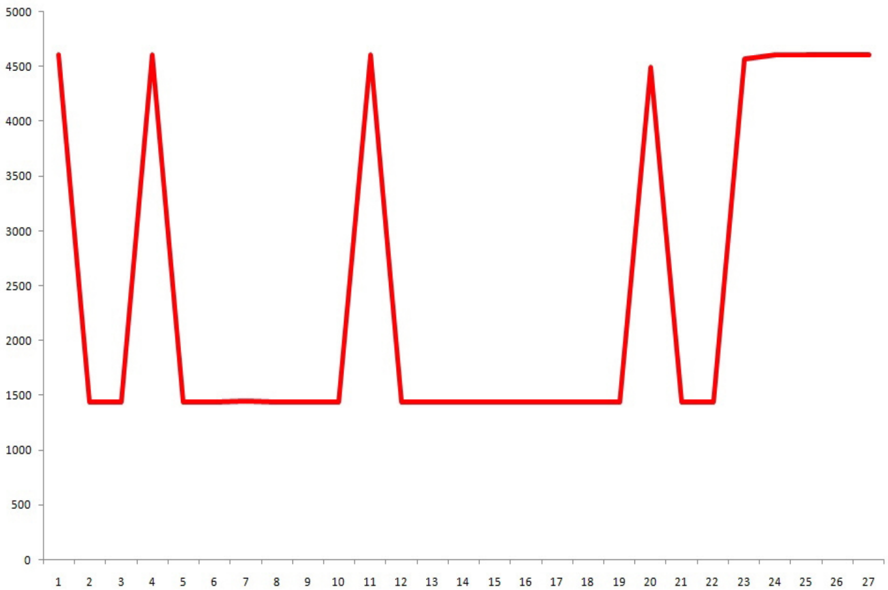
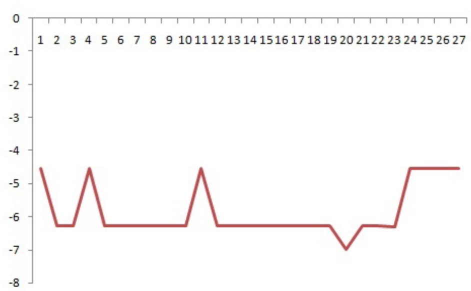
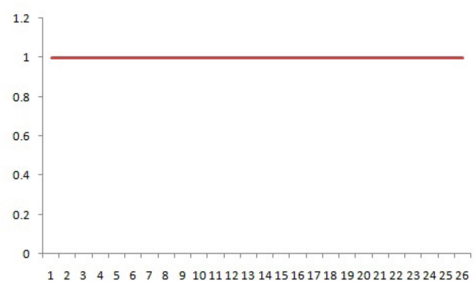


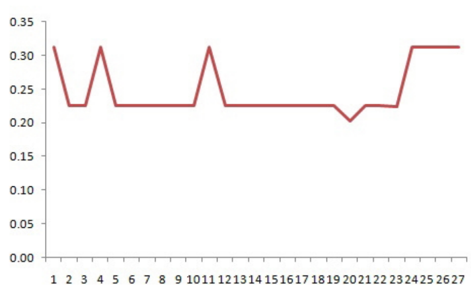
Fig. 2. ML estimates of the VG parameters with "default *vgFit*" (27 moving windows of 1000 obs.)



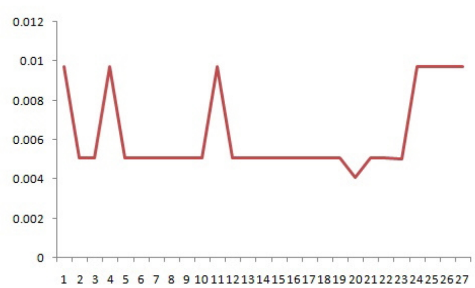
parameter μ



parameter ν



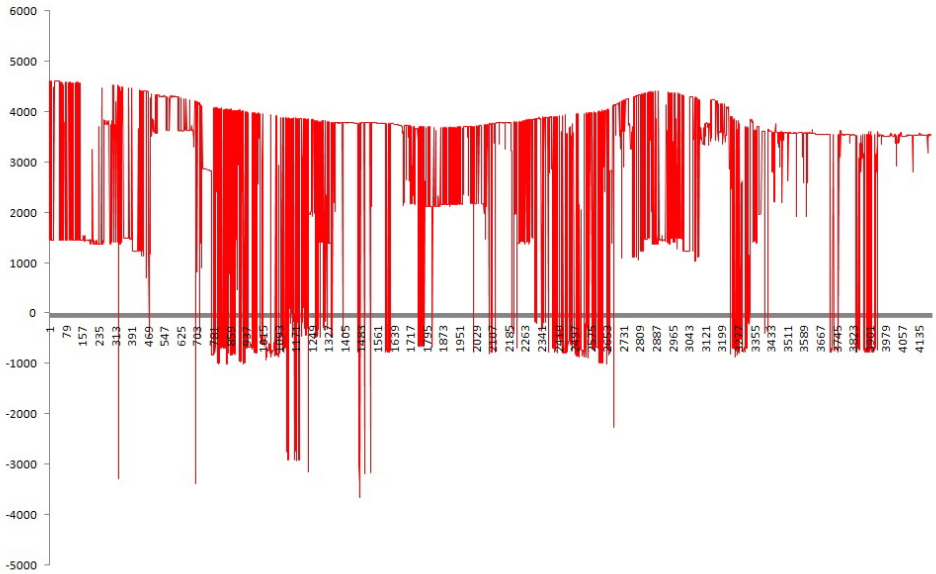
parameter σ



parameter θ

the generally available *R* software are implemented and compared in the next section.

Fig. 3. Maximum value of the log-likelihood function with *vgFit* (moving windows of 1000 obs. for the period 1/3/1992-8/20/2012)



5 Comparing alternative procedures using *R*

The results associated with *vgFit* for the whole set of windows are reported on Fig. 3. One can observe that in some cases the maximum value of the log-likelihood function is more volatile than what seen in the previous section. Due to the scale of graph, what looks like a band with a variable width is indeed a sequence of jumps and the width represents the highest jump suffered by the variable on the y-axis. So, for instance, the band for the first 150-160 moving windows indicates that the value of the log-likelihood function keeps jumping between 4500 and 1500 with the simple change of one observation. Then the wider width, for instance between 4000 and -3000 around moving window 320, denotes a more dramatic change in the maximum value of the log-likelihood function when the moving window goes forward one observation.

Given these disappointing results, it is decided to exploit *R* internal capabilities to construct a number of 'homemade' alternatives to *vgFit*.[17] The first candidate, call it procedure *vgFitMom*, is based on the *optim* function in *R* with starting values for the parameters generated by *vgFitStart* with the option *startValues* set to *MoM*.[18] The second candidate, procedure *MadanMom*, exploits the same *optim* function with starting values for the parameters ob-

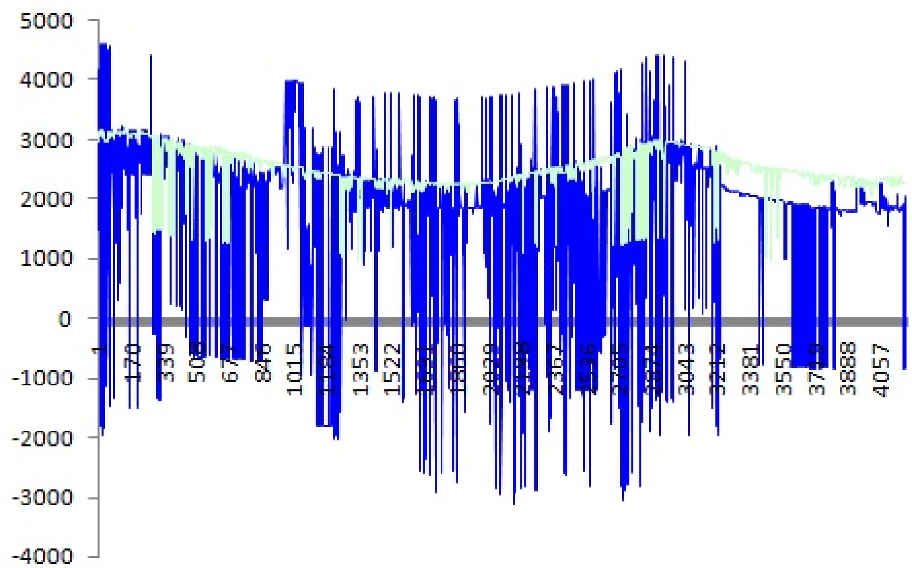
[17] This is in the spirit of the algorithm that can be downloaded from http://stats.stackexchange.com/questions/30054/variance-gamma-distribution-parameter-estimation.

[18] These starting values are based on Barndorff-Nielsen *et al.* (1985) as documented in Scott and Yang Dong (2012, p. 19).

tained with the method of moments as described in Madan *et al.* (1998 fn. 16). Namely, the solution to the moment equations discussed in Section 2 is used as starting point. The third candidate, procedure *adhoc*, calls the *optim* function with starting values for the parameters μ, σ, θ and ν set equal to 0, 0.5, 0 and 0.5, respectively.[19] The last candidate, call it procedure *3optim*, consists of three sequential calls to the *optim* function. The first call maximizes the likelihood of a normal distribution.[20] Then the ML parameter estimates obtained, together with $\nu = 0.5$, are used in the second call as starting values in the optimization procedure for a symmetric VG. Finally, the ML estimates for a symmetric VG, together with $\theta = 0$, are used as starting values in the third call to the *optim* function for the non-symmetric VG.

The likelihood results for procedures *vgFitMom* and *MadanMom* are reported in Figure 4. The maximum value of the log-likelihood function obtained with

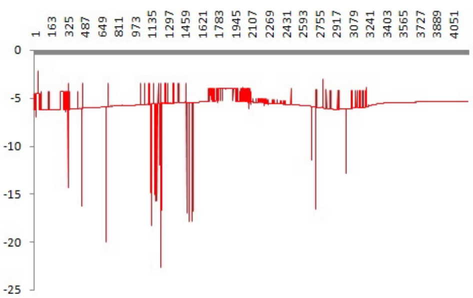
Fig. 4. Maximum value of the log-likelihood function with *vgFitMom* (dark line) and *MadanMom* (moving windows of 1000 obs. for the period 1/3/1992-8/20/2012)



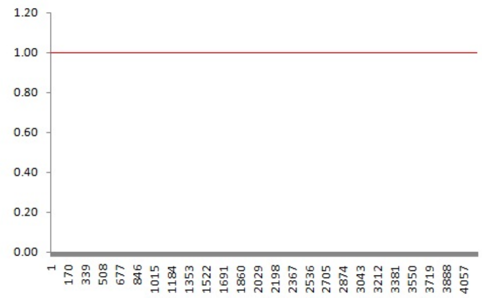
the former procedure appears to be highly sensitive to the data set. Dramatic changes occur when only one new observation is replaced. On the other hand *MadanMom*, based on the method of moments described in Madan *et al.* (1998), is clearly less sensitive than the competitor. However, it is worth it to point out that the density of the VG distribution is not defined in a meaningful number of cases, when these optimization procedures are used. In these circumstances the starting values of the estimates are constrained to be in the range suggested in Scott and Yang Dong (2012, pp. 22–23), i.e. μ and θ in the

[19] This set of parameters is used fairly often in Scott and Yang Dong (2012, p. 19).
[20] This call can be bypassed simply computing the sample mean and standard deviation of the data set at hand.

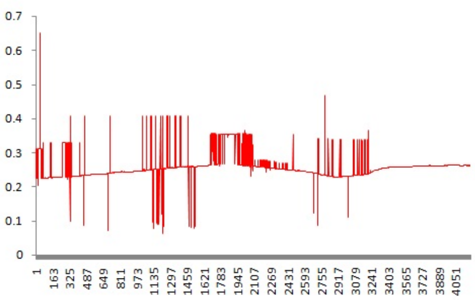
Fig. 5. ML estimates of the VG parameters with *vgFit* (moving windows of 1000 obs. for the period 1/3/1992-8/20/2012)



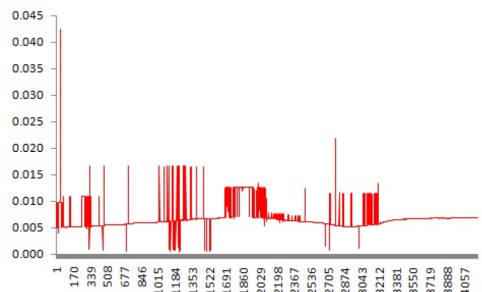
parameter μ



parameter ν



parameter σ



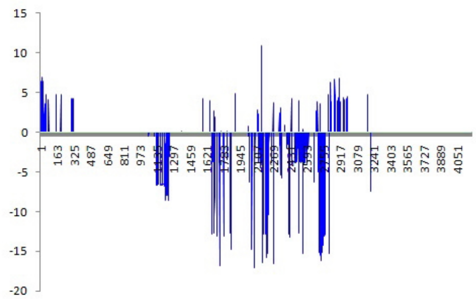
parameter θ

interval $(-4.0, 4.0)$ and σ and ν in the interval $(0.25, 4.0)$. Therefore, it is hard to tell how much of the observed difference between these two procedures is due to the fact that one of them yields more often good starting values, in the sense that they lead to better results, or to the fact that it leads more often to bad starting values, i.e. outside the feasible range. In the latter case, similar results may be the side effect of using the same constrained starting point in successive windows.

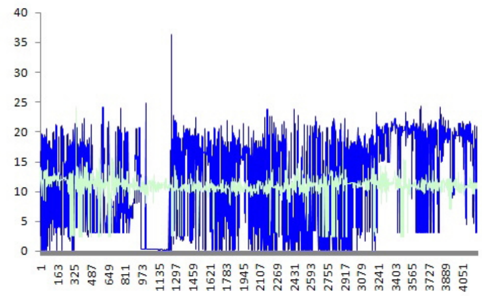
The big changes in the maximum value of the log-likelihood function characterizing *vgFit*, *vgFitMom* and *MadanMom*, when adding or removing one observation, are associated with pronounced changes of the estimated parameter values as well (Fig. 5 and 6). As it is apparent from Fig. 5, the estimate for μ ranges between -2.0 and -22.0, that for σ between .05 and .4, with one isolated case at .6, for θ between 0 and .02, with one isolated case at .04, and for ν is always 1. The procedure labelled *vgFitMom* shows an even more extreme volatility in the estimates as denoted by the graph scales of Fig. 6.

Undoubtedly, *MadanMom* performs much better than *vgFitMom* in terms of parameter volatility. All parameters look fairly stable throughout the whole set of moving windows. However it is worth it to point out that this is mainly due to the scale of the graphs. Indeed, *MadanMom* performs worst than *vgFit*, except for the estimate of the location parameter. Its estimates range between

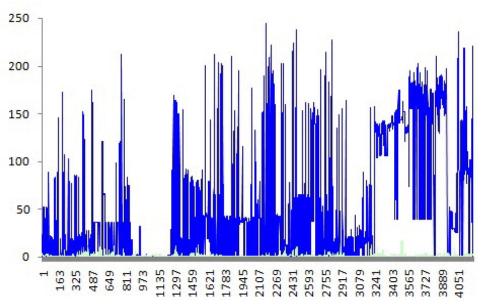
Fig. 6. ML estimates of the VG parameters with *vgFitMom* (dark line) and *MadanMom* (moving windows of 1000 obs. for the period 1/3/1992-8/20/2012)



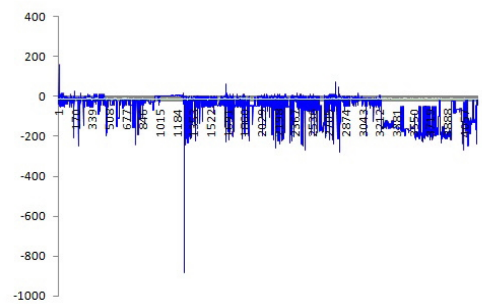
parameter μ



parameter ν



parameter σ



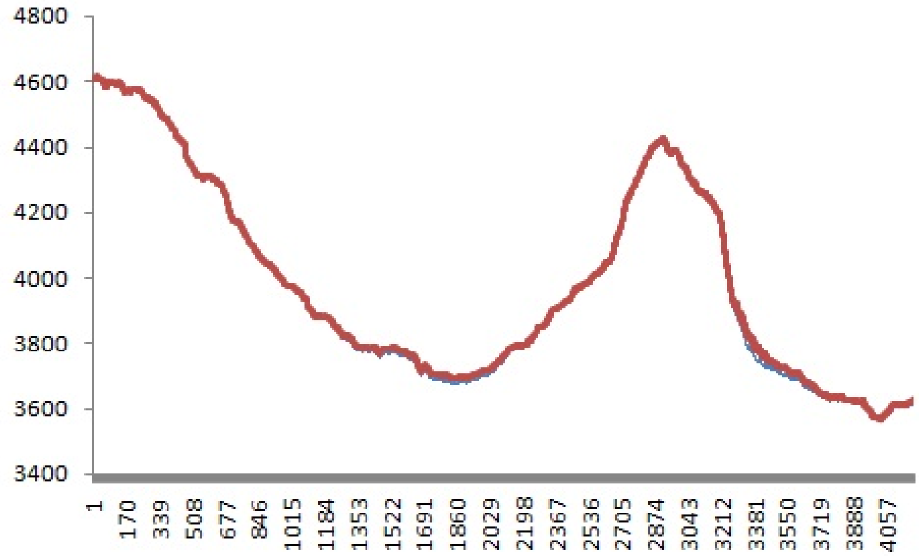
parameter θ

-3.0E-03 and 5.3E-03 for μ , 2.4 and 16 with one instance at 25 for ν, 0.3 and 18 for σ and -7 to 18 for θ. Again, as for the optimum of the log-likelihood function, it is unclear how heavily the relative performance of these two 'moments' procedures is affected by the number of cases in which the estimated starting values of the parameters need to be constrained for computational reasons.

On the other hand, procedures *adhoc* and *3optim* seem quite insensitive to the data set used. As apparent from Figure 7, the maximum of the log-likelihood function changes very smoothly and it never jumps when one observation is changed. Indeed the latter procedure yields almost always the highest likelihood. Out of a total of 4201 likelihoods only in 118 occasions it does not outperform *adhoc* and the difference is always incredibly small as shown in the graph.

The smooth changes in the maximum of the log-likelihood function characterizing procedures *adhoc* and *3optim* are associated with smooth and small changes in their parameter estimates as shown in Figure 8. Moreover, in this case the graph scales are smaller than all the others seen so far. Even more surprising is the fact that when *adhoc* and *3optim* are compared the latter shows incredibly smooth changes in the estimates over time. Incredibly, because most of the time the two procedures yield very close results in terms of

Fig. 7. Maximum value of the log-likelihood function with *3optim* (red line) and *adhoc* (moving windows of 1000 obs. for the period 1/3/1992-8/20/2012)



maximum value of the log-likelihood function. Only in the last 1000 windows, approximately the last four years of trading activity, the estimates show some noticeable variations but are incredibly smaller than those associated with *adhoc* and are fully consistent with the historical events affecting financial markets. Even though it is not fully clear why this procedure behaves so well in this context, it clearly represents a Computationally Great Procedure (*CGP*) for the VG model at hand.

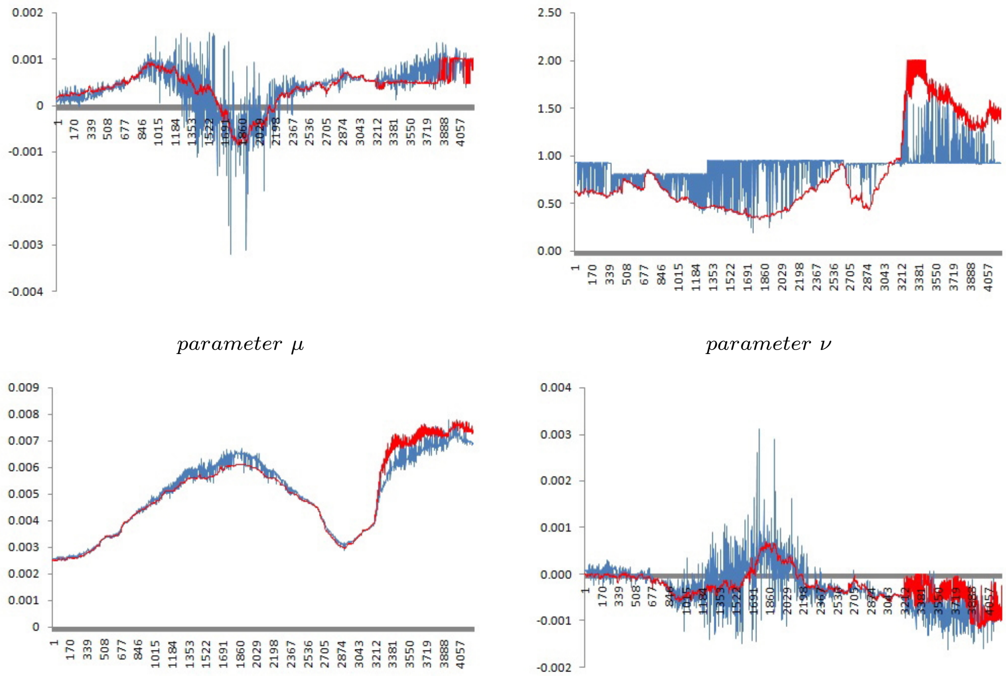
6 Back to the starting line

The main conclusion of the previous section is that the procedure dubbed *CGP*, i.e. three sequential calls to the *optim* function in *R*, seems the most reliable procedure for this class of models. When it is applied to the data set in Madan *et al.* (1998, pp. 89-90) it yields the results reported in Table 2. They look much better than those originally produced with the *R* package and are even better than those generated by *Ezgrad*.

On a chi-squared test (taking twice the difference in log likelihoods given in the fifth row of the first two columns of the various tables) the lognormal model is easily rejected in favor of the symmetric VG with a χ^2_1 statistic of over 36 and the symmetric VG is not rejected in favor of the non symmetric VG.[21] These results are qualitatively very similar to those originally reported in Madan *et*

[21] The critical value at 99% is $\chi^2_{1,99\%} = 6.635$ and at 99.5% is $\chi^2_{1,99.5\%} = 7.879$.

Fig. 8. ML estimates of the VG parameters with *3optim* (red line) and *adhoc* (moving windows of 1000 obs. for the period 1/3/1992-8/20/2012)



parameter μ *parameter* ν

parameter σ *parameter* θ

Table 2
ML estimates of the VG parameters for daily returns on S&P 500 Index

Parameter	**GBM**	**Symm. VG***	**VG****
μ	0.0001	0.0001	0.0001 (3.2E-4)
σ	0.0025	0.0026	0.0026 (4.5E-5)
ν	-	0.5833	0.5850 (0.1204)
θ	-	-	0 (3.5E-4)
ln $\mathcal{L}$	3147.58	3165.62	3165.62
No. of obs.	691	691	691

* The starting values used for the parameters are the GBM optimum values and $\nu = 0.5$ because this parameter has to be positive.** The starting values used for the parameters are the Symm. VG optimum values and $\theta = 0.0$.

al.'s (1998, p. 90). In both cases the skewness parameter θ is not significantly different from zero and the GBM model is resoundingly rejected in favor of the Symmetric VG model. However the parameter estimates seem very different with volatility around 50 times smaller than the original estimate and the excess kurtosis parameter 250 times, approximately, larger.

7 Conclusion

This paper confirms that, as originally reported in Seneta (2004, p. 183), it is impossible to replicate Madan *et al.*'s (1998) results using log daily returns on S&P 500 Index from January 1992 to September 1994. This failure has led to a close investigation of the likelihood function associated with the popular VG model. Both standard econometric software, such as *R*, and non-standard optimization software, such as *Ezgrad* described in Tucci (2002), have been used to study the complexity of the log-likelihood function. It has been shown that the likelihood function is very complicated, with many local optima, and may be incredibly sensitive to very small changes in the estimation sample. Adding or removing a single observation may cause huge changes in the maximum of the log-likelihood function as well as in the estimated parameter values. The technique which seems to perform better in terms of robustness to small changes in the sample, dubbed *CGP* in this study, is the successive optimization of the GBM model, symmetric and non-symmetric VG model with the optimum of the simpler model used as starting point for the optimization of the more complicated model. However, it is unclear if this good performance is sample specific, problem specific or is the outcome of a computationally efficient general principle (namely to optimize an n parameter function sequentially with the optimum of the $n-1$ parameter nested function used as starting point together with the $n-th$ parameter set equal to 0 or positive, as required) which may prove relevant also in other contexts of global optimization.

A Appendix: CODE FOR *MadanMom*

```
library(VarianceGamma)
dataVector1<-read.table(file="Directory");
typeof(dataVector1)
dataVector<-as.matrix(unlist(dataVector1))
typeof(dataVector)
str (dataVector)
mean(dataVector)
## Data set on daily log returns on the S&P 500 Index from 01/03/92 to
## 8/20/2012
ii=1
for (i in 0:4200){
ii=i
nf=999
f=ii+nf
data=dataVector[ii:f]
m1=mean(data)
```

```
m2=mean((data-m1)^2)
m3=mean((data-m1)^3)
m4=mean((data-m1)^4)
a=0
b=0.5
c=0
d=0.5
e1=a+c
e2=b^2+c^2*d
e3=2*c^3*d^2+3*b^2*c*d
e4=3*b^4*d+12*b^2*c^2*d^2+6*c^4*d^3+3*b^4+6*b^2*c^2*d+3*c^4*d^2
ll = function(par){
(m1-(par[1]+par[3]))^(2)+(m2-(par[2]^(2)+par[3]^(2)*par[4]))^(2)
+(m3-(2*par[3]^(3)*par[4]^(2)+3*par[2]^(2)*par[3]*par[4]))^(2)
+(m4-(3*par[2]^(4)*par[4]+12*par[2]^(2)*par[3]^(2)*par[4]^(2)
+6*par[3]^(4)*par[4]^(3)+3*par[2]^(4)+6*par[2]^(2)*par[3]^(2)*par[4]
+3*par[3]^(4)*par[4]^(2)))^2
}
Ris=optim(c(a,b,c,d),ll, gr=NULL, method = "Nelder-Mead")
print(Ris)
## Constraining the pmts to the Scott and Yang Dong's (2012, pp. 22-23)
## range
if (b<0.25) {
b=0.25
} else
b=b
if (b>4) {
b=4
} else
b=b
if (d<0.25) {
d=0.25
} else
d=d
if (d>4) {
d=4
} else
d=d
if (c<-4) {
c=-4
} else
c=c
if (c>4) {
c=4
} else
```

```
c=c
if (a>4) {
a=4
} else
a=a
if (a<-4) {
a=-4
} else
a=a
print(a)
print(b)
print(c)
print(d)
## Computing −log-likelihood with the Variance-Gamma package
ll = function(par){
if(par[2]>0&par[4]>0) return( - sum(log(dvg(data, vgC = par[1], sigma=par[2],
theta=par[3], nu = par[4]) )))
else return(Inf)}
## Direct maximisation/minimisation with optim command
Ris=optim(c(a,b,c,d),ll, gr=NULL, method = "Nelder-Mead")
## Printing and saving final results
print(Ris)
r1=Ris[[1]][1]
r2=Ris[[1]][2]
r3=Ris[[1]][3]
r4=Ris[[1]][4]
rr=Ris[2]
c=paste(r1,r2,r3,r4,rr)
print(c)
name=paste("Directory",i,".txt")
write.table(c,name)
}
```

B Appendix: CODE FOR *vgFitMom*

```
library(VarianceGamma)
dataVector1<-read.table(file="Directory");
typeof(dataVector1)
dataVector<-as.matrix(unlist(dataVector1))
typeof(dataVector)
str (dataVector)
mean(dataVector)
```

```
## Data set on daily log returns on the S&P 500 Index from 01/03/92 to
## 8/20/2012
ii=1
for (i in 0:0){
ii=i
nf=999
f=ii+nf
data=dataVector[ii:f]
Ris<-vgFitStart(data, startValues="MoM")
print(Ris)
a=Ris[[1]][1]
a=a[[1]][1]
b=Ris[[1]][2]
b=b[[1]][1]
b=exp(b)
c=Ris[[1]][3]
c=c[[1]][1]
d=Ris[[1]][4]
d=c[[1]][1]
d=exp(d)
aa=a
bb=b
cc=c
dd=d
## Constraining the pmts to the Scott and Yang Dong's (2012, pp. 22-23)
## range
...
## Computing −log-likelihood with the Variance-Gamma package
...
## Direct maximisation/minimisation with optim command
Ris=optim(c(a,b,c,d),ll, gr=NULL, method = "Nelder-Mead")
## Printing and saving final results
...
}
```

C Appendix: CODE FOR *adhoc*

```
library(VarianceGamma)
dataVector1<-read.table(file="Directory");
typeof(dataVector1)
dataVector<-as.matrix(unlist(dataVector1))
typeof(dataVector)
```

```
str (dataVector)
mean(dataVector)
## data set on daily log returns on the S&P 500 Index from 01/03/92 to
## 8/20/2012
ii=1
for (i in 0:4200){
ii=i
nf=999
f=ii+nf
data=dataVector[ii:f]
## Initial values for the four parameters
a=0
b=0.5
c=0
d=0.5
print(a)
print(b)
print(c)
print(d)
# Computing −log-likelihood with the Variance-Gamma package
ll = function(par){
if(par[2]>0&par[4]>0) return( - sum(log(dvg(data, vgC = par[1], sigma=par[2],
theta=par[3], nu = par[4]) )))
else return(Inf)}
# Direct maximisation/minimisation with optim command
Ris=optim(c(a,b,c,d),ll, gr=NULL, method = "Nelder-Mead")
## Printing and saving final results
...
}
```

C Appendix: CODE FOR *3optim*

```
library(VarianceGamma)
dataVector1<-read.table(file="Directory");
typeof(dataVector1)
dataVector<-as.matrix(unlist(dataVector1))
typeof(dataVector)
str (dataVector)
mean(dataVector)
## Data set on daily log returns on the S&P 500 Index from 01/03/92 to
## 8/20/2012
ii=1
```

```
for (i in 0:4200){
ii=i
nf=999
f=ii+nf
data=dataVector[ii:f]
## Initial values for the mean and std. dev. of the data. Alternatively
## the sample mean and std. dev. can be used directly in the
## second call to optim in the place of r1 and r2.
a=0
b=0.5
## Computing −log-likelihood with dnorm command
## for the first call to optim
ll = function(par){
if(par[2]>0) return( - sum(log(dnorm(data, par[1],
par[2]) )))
else return(Inf)}
## Direct maximisation/minimisation with optim command (first call)
Ris=optim(c(a,b),ll, gr=NULL, method = "Nelder-Mead")
print(Ris)
r1=Ris[[1]][1]
r2=Ris[[1]][2]
rr=Ris[2]
c=paste(r1,r2,rr)
print(c)
name=paste("Directory",i,".txt")
write.table(c,name)
a=r1
b=r2
c=0
d=0.5
## Computing −log-likelihood with the Variance-Gamma package
## for the second call to optim
ll = function(par){
if(par[2]>0&par[4]>0) return( - sum(log(dvg(data, vgC = par[1], sigma=par[2],
theta=0, nu = par[4]) )))
else return(Inf)}
## Direct maximisation/minimisation with optim command (second call)
Ris=optim(c(a,b,0,d),ll, gr=NULL, method = "Nelder-Mead")
print(Ris)
r1=Ris[[1]][1]
r2=Ris[[1]][2]
r3=0
r4=Ris[[1]][4]
rr=Ris[2]
c=paste(r1,r2,r3,r4,rr)
```

```
print(c)
name=paste("Directory",i,".txt")
write.table(c,name)
a=r1
b=r2
c=0
d=r4
## Computing −log-likelihood with the Variance-Gamma package
## for the third call to optim
ll = function(par){
if(par[2]>0&par[4]>0) return( - sum(log(dvg(data, vgC = par[1], sigma=par[2],
theta=par[3], nu = par[4]) )))
else return(Inf)}
## Direct maximisation/minimisation with optim command (third call)
Ris=optim(c(a,b,c,d),ll, gr=NULL, method = "Nelder-Mead")
## Printing and saving final results
...
}
```

Acknowledgements

An earlier version of the present paper, entitled "Modeling log stock returns: Are parameters stable?", was presented at the 19th International Conference on Computing in Economics and Finance held in Vancouver, CA, from July 10 through 12, 2013. The authors would like to thank the partecipants to that session for their numerous and interesting comments, some of which are reflected in the present version. Moreover they would like to express their deep gratitude to Alessandro Giorgione, PhD student at the University of Sienna, for his valuable research assistance and Dott. Alberto Montesi for his highly professional librarian assistence.

References

Almendral, A. and Oosterlee, C. W.: 2007, On american options under the variance gamma process, *Applied Mathematical Finance* **14**, 131–152.

Barndorff-Nielsen, O. E. and Halgreen, C.: 1977, Infinite divisibility of the hyperbolic and generalized inverse gaussian distributions, *Zeitschrift für Wahrschein-lichkeitstheorie und verwandte Gebiete* **38**, 309–311.

Black, F. and Scholes, M. S.: 1973, The pricing of options and corporate liabilities, *Journal of Political Economy* **81**, 637–654.

Bochner, S.: 1955, *Harmonic Analysis and the Theory of Probability*, University of California Press, Berkeley, California.

Clark, P.: 1973, A subordinated stochastic process model with finite variance for speculative prices, *Econometrica* **41**, 135–156.

Daal, E. and Madan, D. B.: 2005, An empirical examination of the variance-gamma model for foreign currency options, *Journal of Business* **78**, 2121–2152.

Deng, G., Dulaney, T. and McCann, C.: 2013, Valuation of reverse convertibles in the variance gamma economy, *Journal of Derivatives & Hedge Funds* **19**, 244–258.

Eberlein, E. and Madan, D. B.: 2010, On correlating lévy processes, *Journal of Risk* **13**, 3–16.

Erdélyi, A., Magnus, W., Oberhettinger, F. and Tricomi, F. G.: 1953, *Bateman Manuscript Project: Tables of Integral Transforms*, Vol. 13, McGraw-Hill, New York, New York.

Fard, F. A. and Rong, N.: 2014, Pricing and managing risks of ruin contingent life annuities under regime switching variance gamma process, *Ann. Finance* **10**, 315–332.

Figueroa-López, J. E., Lancette, S. R., Lee, K. and Mi, Y.: 2012, Estimation of nig and vg models for high frequency financial data, *in* F. Viens, M. C. Mariani and I. Florescu (eds), *Handbook of Modeling High-Frequency Data in Finance*, John Wiley, New York, New York, pp. 1–23.

Finlay, R.: 2009, *The Variance Gamma (VG) Model with Long Range Dependence*, PhD thesis, University of Sydney, School of Mathematics and Statistics, Sydney, Australia.

Finlay, R. and Seneta, E.: 2008, Stationary-increment variance-gamma and t models: simulation and parameter estimation, *International Statistical Review* **76**, 167–186.

Fiorani, F.: 2003, The variance-gamma process for option pricing. Paper presented at the IV Workshop di Finanza Quantitativa, International Centre for Economic Research (ICER), Torino, Italy, available at http//www.icer.it/workshop/Fiorani.pdf.

Fiorani, F. and Luciano, E.: 2006, Credit risk in pure jump structural models, *Working Paper 6*, International Center for Economic Research (ICER), Torino, Italy.

Fu, M. C.: 2007, Variance-gamma and monte carlo, *in* M. C. Fu, R. A. Jarrow, J. J. Yen and R. J. Elliott (eds), *Advances in Mathematical Finance*, Birkhäuser, Boston, Massachussets, pp. 21–35.

Heyde, C. C. and Liu, S.: 2001, Empirical realities for a minimal description risky asset model. the need for fractal features, *Journal of the Korean Mathematical Society* **38**, 1047–1059.

Hurd, T. R.: 2007, Credit risk modelling using time-changed brownian motion. Paper presented at the Further Developments in Quantitative Finance Workshop, International Centre for Mathematical Sciences (ICMS), Edinburgh, UK, available at http://www.math.mcmaster.ca/tom/ HurdTCBM.pdf.

Johnson, N. L., Kotz, S. and Balakrishnan, N.: 1994, *Continuous Univariate Distributions*, Vol. 1, 2nd edn, John Wiley, New York, New York.

Kaishev, V. K. and Dimitrova, D. S.: 2009, Dirichlet bridge sampling for the vg process: Pricing path-dependent options, *Management Science* **55**, 483–496.

Le Courtois, O. and Walter, C. P.: 2014, The computation of risk budgets under the lévy process assumption, *Finance* **35**. forthcoming.

Madan, D. B., Carr, P. and Chang, E.: 1998, The variance gamma process and option pricing, *European Finance Review* **2**, 135–156.

Madan, D. B. and Khanna, A.: 2009, Non gaussian models of dependence in returns, *Working Paper RHS-06-112*, Robert H. Smith School of Business, University of Maryland at College Park, MD, USA.

Madan, D. B. and Seneta, E.: 1990, The variance gamma (vg) model for share market returns, *Journal of Business* **63**, 511–524.

Press, W., Teukolsky, S., Vetterling, W. and Flannery, B.: 1997, *Numerical Recipes in FORTRAN 77: The Art of Scientific Computing*, Vol. 1 of Fortran numerical recipes, 2nd edn, Cambridge University Press, New York, New York. Reprinted with corrections.

Press, W., Teukolsky, S., Vetterling, W. and Flannery, B.: 2007, *Numerical Recipes: The Art of Scientific Computing*, 3rd edn, Cambridge University Press, New York, New York.

Rathgeber, A., Stadler, J. and Stöckl, S.: 2013, Modeling share returns: an empirical study on the variance gamma model. Draft downloaded from http://www.efmaefm.org/0EFMAMEETINGS/EFMAapers/EFMA2013_0201_fullpaper.pdf on July 31, 2014.

Ribeiro, C. and Webber, N.: 2003, Valuing path-dependent options in the variance-gamma model by monte carlo with a gamma bridge, *Journal of Computational Finance* **7**, 81–100.

Rubinstein, M.: 1994, Implied binomial trees, *Journal of Finance* **49**, 771–818.

Scott, D. and Dong, C. Y.: 2012, Package 'variance-gamma': The variance gamma distribution, *available at http://www.r-project.org*, Version 0.3-1.

Seneta, E.: 2004, Fitting the variance-gamma model to financial data: Stochastic methods and their applications, *Journal of Applied Probability* **Special Volume 41(A)**, 177–187.

Stein, H., Hoganand, A. and Carr, P.: 2007, Time for a change: The variance-gamma model and option pricing, *Working paper*, Bloomberg Financial Markets, New York. available at http://ssrn. com/abstract=956625.

Tucci, M. P.: 2002, A note on global optimization in adaptive control. econometrics and macroeconomics, *Journal of Economic Dynamics and Control* **26**, 1739–1764.

Wallmeier, M. and Diethelm, M.: 2012, Multivariate downside risk: Normal versus variance gamma, *The Journal of Futures Markets* **32**, 431–458.