Scientific
Research
Publishing

Likelihood and Quadratic Distance Methods for the Generalized Asymmetric Laplace Distribution for Financial Data

Andrew Luong

École d'actuariat, Université Laval, Canada
Email: andrew.luong@act.ulaval.ca

Abstract

Maximum likelihood (ML) estimation for the generalized asymmetric Laplace (GAL) distribution also known as Variance gamma using simplex direct search algorithms is investigated. In this paper, we use numerical direct search techniques for maximizing the log-likelihood to obtain ML estimators instead of using the traditional EM algorithm. The density function of the GAL is only continuous but not differentiable with respect to the parameters and the appearance of the Bessel function in the density make it difficult to obtain the asymptotic covariance matrix for the entire GAL family. Using M-estimation theory, the properties of the ML estimators are investigated in this paper. The ML estimators are shown to be consistent for the GAL family and their asymptotic normality can only be guaranteed for the asymmetric Laplace (AL) family. The asymptotic covariance matrix is obtained for the AL family and it completes the results obtained previously in the literature. For the general GAL model, alternative methods of inferences based on quadratic distances (QD) are proposed. The QD methods appear to be overall more efficient than likelihood methods infinite samples using sample sizes $n \leq 5000$ and the range of parameters often encountered for financial data. The proposed methods only require that the moment generating function of the parametric model exists and has a closed form expression and can be used for other models.

Keywords

M-Estimators, Cumulant Generating Function, Chi-Square Tests, Generalized Hyperbolic Distribution, Simplex Pattern Search, Variance Gamma, Minimum Distance, Value at Risk, Entropic Value at Risk, European Call Option

1. Introduction

1.1. Generalized Asymmetric Laplace (GAL) Distribution

The generalized asymmetric Laplace distribution (GAL) is a four parameters infinitely divisible continuous distribution with four parameters given by

$$\boldsymbol{\beta} = (\theta, \mu, \sigma, \tau)'. \tag{1}$$

The parameter θ is a location parameter and σ is a scale parameter. The parameter μ can be viewed as the asymmetry parameter of the distribution and τ is the shape parameter which controls the thickness of the tail of the distribution. If $\mu = 0$, the distribution is symmetric around θ, see Kotz et $al.$ ([1], p. 180). It is flexible and can be used as an alternative to the four parameters stable distribution. The GAL distribution has a thicker tail than the normal distribution but unlike the stable distribution where even the first positive moment might not exist, all the positive integer moments exist. Its moment generating function is

$$M(s) = \frac{e^{\theta s}}{\left(1 - \mu s - \frac{1}{2}\sigma^2 s^2\right)^{\tau}}, \sigma, \tau \geq 0, \boldsymbol{\beta} = (\theta, \mu, \sigma, \tau)', \tag{2}$$

s must satisfy the inequality

$$1 - \frac{1}{2}\sigma^2 s^2 - \mu s > 0. \tag{3}$$

The GAL distribution is also known as variance gamma (VG) distribution. It was introduced by Madan and Senata [2], Madan et $al.$ [3]. For the GAL distribution, we adopt the parameterizations used by Kotz et $al.$ [1]. It is not difficult to relate them to the original parameterization, see Senata [4]. The commonly used parameterisations will be discussed in Section (1.2).

From the moment generating function, it is easy to see that the first four cumulants of the GAL distribution are given by

$$c_1 = \theta + \tau\mu, \quad c_2 = \tau\sigma^2 + \tau\mu^2, \tag{4}$$

$$c_3 = 3\tau\sigma^2\mu + 2\tau\mu^3, \quad c_4 = 6\tau\mu^4 + 12\tau\mu^2\sigma^2 + 3\tau\sigma^4. \tag{5}$$

Note that $c_3 = 0$ if $\mu = 0$ and c_3 can be positive or negative depending on values of the parameters .Therefore, the GAL distribution can be symmetric or asymmetric. Furthermore, with $c_4 > 0$, the tail of the GAL distribution is thicker than the normal distribution. These characteristics make the GAL distribution useful for modelling asset returns, see Senata [4] for further discussions on financial modelling using the GAL distribution.

The moments can be obtained based on cumulants and they are given below,

$$E(X) = c_1,$$

$$E(X - E(X))^2 = c_2,$$

$$E(X - E(X))^3 = c_3,$$

$$E(X - E(X))^4 = 6\tau\mu^4 + 12\tau\mu^2\sigma^2 + 3\tau\sigma^4 + 3\tau^2\sigma^4 + 6\sigma^2\tau^2\mu^2 + 3\mu^4\tau^2.$$

The GAL distribution belongs to the class of normal mean-variance mixture distributions where the mixture variable follows agamma distribution with shape parameter τ and scale parameter equal to 1, *i.e.*, with density function

$$f_w(w) = \frac{1}{\Gamma(\tau)} w^{\tau-1} \mathrm{e}^{-w}, w, \tau > 0, \quad \Gamma(.) \text{ is the commonly used gamma func-}$$

tion.

This leads to the following representation in distribution using Expression (4.1.10) in Kotz *et al.* ([3], p. 183),

$$X =^d \theta + \mu Y + \sigma\sqrt{Y}Z \quad \text{where} \tag{6}$$

1) $Z \sim N(0,1)$,
2) $Y \sim G(\tau,1)$ as given by expression (8) and independent of Z
3) $\theta, \mu, \sigma, \tau$ are parameters with $\sigma, \tau > 0$ 。

The representation given by expression (6) is useful for simulating samples from a GAL distribution. Note that despite the simple closed form expression for the moment generating function, the density function is rather complicated as it depends on the modified Bessel function of the third kind with real index λ, *i.e.*, $K_\lambda(u)$ see Kotz *et al.* ([1], p. 315) for various representations for the function $K_\lambda(.)$. The density function will be introduced in Section (1.2). Using the moment generating function of the GAL distribution, it is easy to see that the distribution is related to a Lévy process, see Podgorski and Wegener [5] for GAL processes.

The GAL parametric family can be introduced as a limit case of the generalized hyperbolic (GH) family where the mixing random variable belongs to the generalized inverse Gaussian family, see Mc Neil *et al.* [6] for properties of the GH family. Note that the GAL family is nested within the bilateral gamma family as the GAL random variable can be represented in distribution as

$$X =^d \theta + \frac{\sigma}{\sqrt{2}}\left(\frac{1}{\kappa}G_1 - \kappa G_2\right), \tag{7}$$

G_1 and G_2 are independent random variables with common gamma distribution. The common mgf of the gamma distribution is given by

$$M_G(s) = \frac{1}{(1-s)^\alpha}, \text{ see expression (4.1.1) given by Kotz } et\ al.\ ([1], \text{p. 183}).$$

If we introduce κ using $\mu = \frac{\sigma}{\sqrt{2}}\left(\frac{1}{\kappa} - \kappa\right)$, the GAL distribution can also be parameterised using the four equivalent parameters, *i.e.*, with $\theta, \sigma, \kappa, \tau$.

Moment estimation for the GAL family has been given by Podgorski and Wegener [5]. Maximum likelihood estimation for the GH family by fixing the parameter τ within some bounds has been given by Protassov [7], McNeil *et al.* ([6], p. 80). For ML estimation, they implicitly assumed that the mixing random variable $Y \sim \text{Gamma}\left(\tau, \frac{\phi}{2}\right)$ which implies the following form of the moment generating function for X,

$$M(s) = \frac{\mathrm{e}^{\theta s}}{\left(1 - \frac{\phi}{2}\left(\mu s + \frac{1}{2}\sigma^2 s^2\right)\right)^\tau}, \phi > 0.$$

From the above expression, it is easy to see that the parameter $\phi > 0$ is redundant and the parameterisation using five parameters will introduce instability in the estimation process. It appears to be simpler to use the parameterisation given by Kotz *et al.* [1] or the parametrisation used by Madan and Senata [2], Madan *et al.* [3], Senata [4] with only four parameters by letting $\phi = 2$.

Hu [8] advocated fitting the GAL distribution using the EM algorithm but the drawback of this approach is the difficulty to obtain the information matrix using the method of Louis [9], see McLachlan and Krishnan [10] for a comprehensive review of the EM algorithm. The lack of a closed form asymptotic covariance matrix for the estimators might create difficulties for hypotheses testing.

1.2. Some Properties of the GAL Distribution and Parameterisations

In this subsection, we first review a few parameterisations which are commonly used for the GAL distribution.

Definition 1 (GAL density)

From the GH density, the density function for the GAL distribution can be obtained and it can be expressed as

$$f\left(x;\theta,\sigma,\mu,\tau\right)=\frac{\sqrt{2}\mathrm{e}^{\frac{\mu}{\sigma}\left(\frac{X-\theta}{\sigma}\right)}}{\sigma\sqrt{\pi}\Gamma\left(\tau\right)}\left(\frac{1}{\sqrt{2+\left(\frac{\mu}{\sigma}\right)^2}}\right)^{\tau-0.5}\left(\frac{|X-\theta|}{\sigma}\right)^{\tau-0.5}$$
$$\cdot K_{\tau-0.5}\left(\sqrt{2+\left(\frac{\mu}{\sigma}\right)^2}\frac{|X-\theta|}{\sigma}\right). \tag{8}$$

The vector of parameters is $\boldsymbol{\beta}=\left(\theta,\mu,\sigma,\tau\right)'$ and we shall call this parametrisation parameterisation1.The density can be derived using thenormal mean variance mixture representation given by expression (6). See expression (3.30) given by Mc Neil *et al.* ([6], p. 78).

Alternatively, by letting $\mu=\frac{\sigma}{\sqrt{2}}\left(\frac{1}{\kappa}-\kappa\right)$ and keeping other parameters as in parametrisation 1, we obtain the following expression for the density of a GAL distribution

$$f\left(x;\theta,\sigma,\kappa,\tau\right)=\frac{\sqrt{2}\mathrm{e}^{\frac{\sqrt{2}}{2}\left(\frac{1}{\kappa}-\kappa\right)\left(\frac{X-\theta}{\sigma}\right)}}{\sigma\sqrt{\pi}\Gamma\left(\tau\right)}\left(\frac{\sqrt{2}}{\frac{1}{\kappa}+\kappa}\right)^{\tau-0.5}\left(\frac{|X-\theta|}{\sigma}\right)^{\tau-0.5}$$
$$\cdot K_{\tau-0.5}\left(\frac{1}{\sqrt{2}}\left(\frac{1}{\kappa}+\kappa\right)\frac{|X-\theta|}{\sigma}\right). \tag{9}$$

with the vector of parameters given by $\boldsymbol{\beta}=\left(\theta,\kappa,\sigma,\tau\right)'$. We shall call this parameterisation, parameterisation 2 which is used by Kotz *et al.* ([1], p. 184).

Note that θ,σ are respectively the location and scale parameter with either parameterisation 1 or 2. Setting $\theta=0,\sigma=1$, the standardized GAL density with parameterisation 2 will have only two parameters and it is given by

$$f_\varepsilon\left(x;\kappa,\tau\right)=\frac{\sqrt{2}\mathrm{e}^{\frac{\sqrt{2}}{2}\left(\frac{1}{\kappa}-\kappa\right)x}}{\sqrt{\pi}\Gamma\left(\tau\right)}\left(\frac{\sqrt{2}}{\frac{1}{\kappa}+\kappa}\right)^{\tau-0.5}\left(|x|\right)^{\tau-0.5}K_{\tau-0.5}\left(\frac{1}{\sqrt{2}}\left(\frac{1}{\kappa}+\kappa\right)|x|\right)$$

or equivalently by using parametrisation1,

$$f_\varepsilon\left(x;\mu,\tau\right)=\frac{\sqrt{2}\mathrm{e}^{\mu x}}{\sqrt{\pi}\Gamma\left(\tau\right)}\left(\frac{1}{\sqrt{2+\left(\mu\right)^2}}\right)^{\tau-0.5}\left(|x|\right)^{\tau-0.5}K_{\tau-0.5}\left(\sqrt{2+\left(\mu\right)^2}\,|x|\right).$$

Following Kotz *et al.* [1] we only use these two parametrisations but it is easy to see their relationships with parametrisation 3 used by Madan *et al.* [3] and Senata [4]. With parametrisation 3, the mgf of the GAL distribution is

$$M\left(s\right)=\frac{\mathrm{e}^{cs}}{\left(1-\theta'\nu s-\frac{1}{2}\nu\sigma'^2 s^2\right)^{1/\nu}}, \tag{10}$$

the parameters are θ',σ',ν,c with

$$\theta'=\frac{\mu}{\nu},\sigma'^2=\frac{\sigma^2}{\nu},\nu=1/\tau,c=\theta.$$

The first four moments using parameterisation 3 as given by Senata ([4], p. 181) aregiven below,

$$\begin{aligned}
&E\left(X\right)=c+\theta',V\left(X\right)=\sigma'^2+\theta'^2\nu,\\
&E\left(X-E\left(X\right)\right)^3=2\theta'^3\nu^2+3\sigma'^2\theta'\nu,\\
&E\left(X-E\left(X\right)\right)^4=3\sigma'^4\nu+12\sigma'^2\theta'^2\nu^2+6\theta'^4\nu^3+3\sigma'^4+6\sigma'^2\theta'^2\nu+3\theta'^4\nu^2.
\end{aligned}$$

The GAL random variable can also be expressed as the difference of two independent gamma random variables, the GAL random variable is nested inside the class of bilateral gamma random variable Y which can be represented as

$$Y=^d\theta+G_1-G_2 \tag{11}$$

with G_1, G_2 are independent gamma random variables with the mgf's given respectively by $M_{G_1}\left(s\right)=\frac{1}{\left(1-\beta_1 s\right)^\alpha}$ and $M_{G_2}\left(s\right)=\frac{1}{\left(1-\beta_2 s\right)^\alpha}$. We obtain the GAL random variable by letting $\beta_1=\frac{\sigma}{\kappa\sqrt{2}}$ and $\beta_2=\frac{\kappa\sigma}{\sqrt{2}}$.

The class of bilateral gamma distribution was introduced by Küchler and Tappe [11] and they have shown that the Esscher transform of a bilateral gamma distribution remains within this class of distribution .More specifically, let Y^E be the random variable with mgf given by $M_{Y^E}\left(s\right)=\frac{M_Y\left(s+h\right)}{M_Y\left(s\right)}$. I is easy to see that $Y^E=^d\theta+\bar{G}_1-\bar{G}_2$, $\bar{G}_1$ and $\bar{G}_2$ are independent gamma random variables with common shape parameter $\alpha\left(\alpha=1\right)$ and scale parameters given respectively by $\bar{\beta}_1=\frac{\beta_1}{1-\beta_1 h}$ and $\bar{\beta}_2=\frac{\beta_2}{1+\beta_2 h}$.

For option pricing with the risk neutral approach, this property is useful as it

is easy to simulate samples from a bilateral gamma distribution. The use of Esscher transform to find risk neutral parameters for option pricing in financeisdue to the seminal works of Gerber and Shiu [12]. The Esscher transform risk neutral parameters can also be interpreted as minimum entropy risk neutral parameters. See Miyahara [13] for this interpretation, see section 4 for more discussions on financial applications.

For numerical methods to find estimators, Nelder-Mead simplex method and related derivative free simplex methods are recommended. Derivative free simplex direct search methods are well described in chapter 16 of the book by Bierlaire [14].

The paper is organized as follows. In Section 2, some submodels of the GAL family are introduced to highlight the difficulty on obtaining the asymptotic covariance matrix using classical likelihood theory. Asymptotic properties of the ML estimators are investigated in section (3). The ML estimators for the GAL family are shown to be consistent for $\tau > \frac{1}{2}$. For the special case with $\tau = 1$, this corresponds to the asymmetric Laplace (AL) model, we obtain the asymptotic covariance matrix in closed form using the approach based on M-estimation theory as given by Huber [15] which completes the missing components of expression (2) given by Kotz *et al.* ([16], p. 818). As an alternative to ML estimation, QD estimation based on matching cumulant generating functions is developed in section (4) for the entire GAL family. The QD estimators are shown to be consistent and follow an asymptotic normal distribution. The asymptotic covariance matrix can be obtained in closed form for the entire GAL family using QD methods which makes testing for parameters easy to implement. Chi-square goodness of fit tests statistics can also be constructed based on the distance function used to obtained QD estimators. The methods are also general and can be applied to other models. Numerical issues and simulations illustrations are discussed in Section (5). A limited simulation study shows that the proposed QD estimators perform better than ML estimators overall for sample sizes $n \le 5000$ using parameters values often encountered for financial data. Some applications drawn from finance are discussed in Section (6).

We shall consider first a few submodels of the GAL model to show the difficulties encountered when likelihood theory is used to obtain the asymptotic covariance matrix for ML estimators.

The difficulties are mainly due to the score functions when viewed as functions of the parameters have a discontinuity point and fail to be differentiable. If the asymptotic covariance matrix for the ML estimators is derived based on likelihood theory, it will have missing components. This is the problem of expression (2) given by Kotz *et al.* ([16], p. 818) for the AL family, a subfamily of the GAL family. M-estimation theory will be used to replace likelihood theory for deriving the asymptotic covariance matrix.

2. Some Subfamilies of the GAL Family

Example 1

Let $\mu = 0, \tau = 1, \sigma = 1, \tau = 1$ and the only parameter is the location parameter and the family is symmetric around θ. Using the result $K_{\frac{1}{2}}(u) = \sqrt{\dfrac{\pi}{2u}} e^{-u}$, the density function is reduced to

$$f(x, \theta) = \frac{1}{2s_0} e^{-\frac{|x-\theta|}{s_0}}, s_0 = \frac{1}{\sqrt{2}}.$$

Equivalently,

$$f(x, \theta) = f_0(x - \theta), f_0(x) = \frac{1}{\sqrt{2}} e^{-\sqrt{2}|x|}, -\infty < x < \infty.$$

This is the well known double exponential distribution, the maximum likelihood estimator for θ is the sample median. There is no Fisher information matrix available as the score function is discontinuous with respect to the parameter θ. The asymptotic variance of the sample median can be found by using M-estimation theory, see Huber [17], Huber [15], also see Amemiya ([18], p. 148-154) on the least absolute deviations (LAD) estimator. We shall use the same approach to derive the asymptotic covariance matrix for the ML estimators for the GAL distribution with $\tau = 1$. The GAL distribution when $\tau = 1$ is the asymmetric Laplace (AL) distribution. The AL distribution will be introduced below.

Example 2

Using the density of the GAL distribution and setting $\tau = 1$, we obtain the AL distribution with only 3 parameters. The location and scale parameters are given respectively by θ, σ and the asymmetry parameter μ. If parameterisation 2 is used, the density function $g(x; \theta, \sigma, \kappa)$ of the AL distribution is based on the standardized AL density as given by expression (4.1.31) in Kotz et al. ([1], p. 189) with

$$g(x; \theta, \sigma, \kappa) = \frac{\sqrt{2}}{\sigma\gamma} \exp\left(\frac{\sqrt{2}}{2} \delta\left(\frac{X - \theta}{\sigma}\right)\right) \exp\left(-\frac{\sqrt{2}}{2} \gamma\left(\frac{|X - \theta|}{\sigma}\right)\right),$$

$$\gamma = \kappa + \frac{1}{k}, \delta = \kappa - \frac{1}{k}.$$

The AL family can be considered as a subfamily of the GAL family and the score functions for this model are again discontinuous. We shall derive the asymptotic covariance matrix using M-estimation theory in Section (3.2) and complete the expression (2) of Kotz et al. ([16], p. 818). The expression derived by the authors has missing components as it is derived based on likelihood theory. Kotz et al. [16] used a different parametrisation but it is equivalent to the one used in Kotz et al. ([1], p. 189) and it is not difficult to establish the links between these 2 parameterisations.

3. Maximum Likelihood Estimation for the GAL Family

3.1. Maximum Likelihood Estimation for the GAL Distribution

For consistency of the MLE, the following Theorem which is Theorem 2.5 given

by Newey and McFadden ([19], p. 2131) is useful. We make the basic assumption that we have a random sample which consists of n iid observations $X_1, \cdots, X_n$ drawn from the GAL parametric family with density $f(x; \boldsymbol{\beta})$ where $\boldsymbol{\beta}_0$ is the vector of the true parameters.

Theorem (Consistency)

Assume that:

1) If $\boldsymbol{\beta}_1 \neq \boldsymbol{\beta}_2$ then $f(x; \boldsymbol{\beta}_1) \neq f(x; \boldsymbol{\beta}_2)$.

2) The parameter space $\boldsymbol{\Theta}$ is compact, $\boldsymbol{\beta}_0 \in \boldsymbol{\Theta}$.

3) $f(x; \boldsymbol{\beta})$ is a continuous with respect to $\boldsymbol{\beta}$.

4) $E\left(\sup_{\boldsymbol{\beta} \in \boldsymbol{\Theta}} \left| \ln f(x; \boldsymbol{\beta}) \right| \right) < \infty$.

Under the conditons stated, the ML estimators (MLE) given by the vector $\hat{\boldsymbol{\beta}}$ is obtained by maximizing the log of the likelihood function

$\ln L(\boldsymbol{\beta}) = \sum_{i=1}^{n} \ln f(x; \boldsymbol{\beta})$ is conssent, $\hat{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta}_0$.

One can see that the conditions for consistency are mild, the condition d) will be satisfied for the GAL family if $\tau > \frac{1}{2}$ as the density function remains bounded. For $\tau \leq \frac{1}{2}$, the density functions with $\theta = 0, \mu = 0$ tend to infinity as $x \to 0_+$, see Theorem 4.1.2 given by Kotz *et al.* ([1], p. 190-192).

It might be possible to prove consistency using the approach to obtain results of Theorem 4 by Broniatowski *et al.* ([20], p. 2578).

For asymptotic normality, it is more complicated as standard theory often requires that the function $\ln L(\boldsymbol{\beta})$ being twice differentiable with respect to $\boldsymbol{\beta}$. The appearance of the Bessel function creates further complications. It makes it very difficult to establish asymptotic properties even with the use of M estimation theory.

For the special case with $\tau = 1$ which corresponds to the AL distribution, the density function can be expressed without the use of the Bessel function and M-estimation theory can be used to find the asymptotic covariance matrix for the ML estimators. Asymptotic normality has been shown by Kotz *et al.* ([1], p. 158-174) but the asymptotic covariance matrix of the ML estimators is still incomplete.

The formula (2.2) given by Kotz *et al.* ([16], p. 818) does not give the correct asymptotic covariance for the ML estimators. The complete formula for the asymptotic covariance matrix of the ML estimators can be obtained using M-estimation theory. An example is given at the end of section (3.2) which shows that one cannot recover the common asymptotic variance of the sample median using results in Kotz *et al.* ([16], p. 818).

M-estimation theory allows the score functions when viewed as functions of the parameters to have a few points of discontinuities and full differentiability with respect to β can be replaced by one side differentiability accordingly. Amemiya ([18], p. 151) uses this approach. For establishing asymptotic normality for the sample median, the sample median is viewed as a root given by a solution of the estimating equation

$$\frac{1}{n}\sum_{i=1}^{n}\psi(x_i,\theta)=0,$$

using the indicator function $I[.]$,

$\psi(x,\theta)=-I[\theta<x]$ and $\psi(x,\theta)=I[\theta>x], \psi(x,\theta)=0$ if $x=\theta$.

The function $\psi(x,\theta)$ is simply the one side derivative and we adopt the notation

$\psi(x,\theta)=\dfrac{\partial|x-\theta|}{\partial\theta}$ with the meaning of one side derivative, also see Hogg *et al.* ([21], p. 538) on estimating equations based on the sign test. The probability of the existence of such a root tend to 1 as $n\to\infty$.

Another M estimator for the location parameter θ has been proposed by Huber ([17], p. 232-233). It consists of estimating θ by solving

$\dfrac{1}{n}\sum_{i=1}^{n}\psi(x_i,\theta)=0$ with

$\psi(x,\theta)=x-\theta$ if $|x-\theta|\le k$, k is chosen.

$\psi(x,\theta)=k$, if $|x-\theta|>k$.

For M-estimators based on $\psi(x,\boldsymbol{\beta})$, where $\boldsymbol{\beta}$ is a vector of parameters, Huber [17], Huber [15] has generalized and relaxed conditions for the classical Taylor expansion. The technical details can be found in his seminal paper and in Huber [15]. It can be summarized as follows. Suppose that the M-estimators given by $\hat{\boldsymbol{\beta}}$, given as the roots of the following estimating functions

$$\frac{1}{n}\sum_{i=1}^{n}\psi(x,\boldsymbol{\beta})=0. \tag{12}$$

Under the following main conditions:

a) $\dfrac{1}{n}\sum_{i=1}^{n}\psi(x_i,\hat{\boldsymbol{\beta}})\overset{p}{\longrightarrow}0$, assuming $\hat{\boldsymbol{\beta}}\overset{p}{\longrightarrow}\boldsymbol{\beta}_0$ has been shown,

b) $\lambda(\boldsymbol{\beta}_0)=E_{\boldsymbol{\beta}_0}(\boldsymbol{\beta}_0)=0, \lambda(\boldsymbol{\beta})=E_{\boldsymbol{\beta}_0}(\psi(x,\boldsymbol{\beta}))$,

with assumption *N*-3 given by Huber ([15], p. 132) and $\lambda(\boldsymbol{\beta})$ is differentiable with respect to $\boldsymbol{\beta}$, then we have the following representation:

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi(x,\boldsymbol{\beta}_0)=-\Lambda(\boldsymbol{\beta}_0)\sqrt{n}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)+o_p(1),$$

$\Lambda(\boldsymbol{\beta}_0)=\left.\dfrac{\partial\lambda(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}'}\right|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}$ and $o_p(1)$ is a term converging to 0 in probability.

When we compare with the usual Taylor expansion, we only require $\lambda(\boldsymbol{\beta})=E_{\boldsymbol{\beta}_0}(\psi(x,\boldsymbol{\beta}))$ to be differentiable with respect to $\boldsymbol{\beta}$. This differentiability condition is satisfied for the AL family. Note that if indeed the score functions are differentiable then $-\Lambda(\boldsymbol{\beta}_0)$ is the Fisher information matrix.

For the technical details on how to verify the conditions *N*-3, see Hinkley and Revankar ([22], p. 7). The condition 1) is usually verified by making use of the Lebesgue dominated convergence theorem (LDGT) as given by Rudin ([23], p. 321). It can become every technical to construct integrable functions to bound the score functions in order to check the sufficient conditions for the LDGT but they are expected to hold for the AL distribution with the existence of all integer

positive moments and the parameters space is assumed to be compact. Essentially, we need to show that the condition 2) is met by showing the convergence in probability of the integrals

$$\int_{-\infty}^{\infty} \psi\left(x, \hat{\boldsymbol{\beta}}\right) \mathrm{d} F_n\left(x\right) \overset{P}{\longrightarrow} \int_{-\infty}^{\infty} \psi\left(x, \boldsymbol{\beta}_0\right) \mathrm{d} F_{\boldsymbol{\beta}_0}\left(x\right) = E_{\boldsymbol{\beta}_0}\left(\boldsymbol{\beta}_0\right) = 0, \quad F_n\left(x\right)$$ is the

sample distribution function, the score functions are given by expressions (14)-(16).

From the above representation, we then have

$$\sqrt{n}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right) \overset{L}{\longrightarrow} N\left(0, \left(\left[\Lambda\left(\boldsymbol{\beta}_0\right)\right]^{-1}\right) V_{\boldsymbol{\beta}_0}\left(\psi\left(x, \boldsymbol{\beta}_0\right)\right)\left(\left[\Lambda\left(\boldsymbol{\beta}_0\right)\right]^{-1}\right)'\right).$$

The asymptotic covariance matrix of $\hat{\boldsymbol{\beta}}$ is given by

$$V\left(\hat{\boldsymbol{\beta}}\right) = \frac{1}{n}\left(\left[\Lambda\left(\boldsymbol{\beta}_0\right)\right]^{-1}\right) V_{\boldsymbol{\beta}_0}\left(\psi\left(x, \boldsymbol{\beta}_0\right)\right)\left(\left[\Lambda\left(\boldsymbol{\beta}_0\right)\right]^{-1}\right)', \qquad (13)$$

$V_{\boldsymbol{\beta}_0}\left(\psi\left(x, \boldsymbol{\beta}_0\right)\right)$ is the covariance matrix of the vector $\psi\left(x, \boldsymbol{\beta}_0\right)$, $\psi\left(x, \boldsymbol{\beta}_0\right)$ is the vector of the true score functions or quasi score functions if a proxy density function is used to replace the true density function.

Now based on M-estimation theory, we proceed to find $\Lambda\left(\boldsymbol{\beta}_0\right)$ and $V_{\boldsymbol{\beta}_0}\left(\psi\left(x, \boldsymbol{\beta}_0\right)\right)$ for the AL distribution to obtain the asymptotic covariance matrix of the ML estimators in the following section.

3.2. Asymptotic Covariance Matrix for the AL Family

Kotz *et al.* [1], Kotz *et al.* [16] have shown that the ML estimators for the AL distribution have an asymptotic normal distribution but their asymptotic covariance matrix given by expression (3.5.1) of Kotz ([1], p. 158) which is identical to expression (2) given by Kotz *et al.* ([16], p. 818) is still incomplete. If M-estimation theory is used then the asymptotic covariance matrix should be based on Corollary (3.2) as given by Huber ([15], p. 133), also see expression (12.18) given by Woolridge ([24], p. 407).

Since

$$\ln g\left(x; \theta, \sigma, \kappa\right) = \ln \sqrt{2} - \ln \sigma - \ln \gamma + \frac{\delta\sqrt{2}}{2}\frac{\left(x - \theta\right)}{\sigma} - \frac{\gamma\sqrt{2}}{2}\frac{\left|x - \theta\right|}{\sigma},$$

the following derivatives are the score functions of the AL distribution,

$$\psi_1\left(x; \theta, \sigma, \kappa\right) = \frac{\partial \ln g\left(x; \theta, \sigma, \kappa\right)}{\partial \theta} = -\frac{\sqrt{2}\delta}{2\sigma} - \frac{\sqrt{2}\gamma}{2\sigma}v\left(x; \theta\right)$$

$$\text{with } v\left(x; \theta\right) = -I\left[x > \theta\right] + I\left[x < \theta\right], v\left(x; \theta\right) = 0 \text{ if } x = \theta. \qquad (14)$$

$$\psi_2\left(x; \theta, \sigma, \kappa\right) = \frac{\partial \ln g\left(x; \theta, \sigma, \kappa\right)}{\partial \sigma} = -\frac{1}{\sigma} - \frac{\sqrt{2}\delta}{2}\frac{\left(x - \theta\right)}{\sigma^2} + \frac{\sqrt{2}\gamma}{2}\frac{\left|x - \theta\right|}{\sigma^2}, \qquad (15)$$

$$\psi_3\left(x; \theta, \sigma, \kappa\right) = \frac{\partial \ln g\left(x; \theta, \sigma, \kappa\right)}{\partial \kappa} = -\frac{\frac{\partial \gamma}{\partial \kappa}}{\gamma} + \frac{\partial \delta}{\partial \kappa}\frac{\sqrt{2}\left(x - \theta\right)}{\sigma} - \frac{\sqrt{2}}{2}\frac{\partial \gamma}{\partial \kappa}\frac{\left|x - \theta\right|}{\sigma}. \qquad (16)$$

Let $\boldsymbol{\beta} = \left(\theta, \sigma, \kappa\right)'$ and $\boldsymbol{\beta}_0$ the vector of the true parameters we need to find first the vector

$$\lambda(\boldsymbol{\beta}) = \left(\lambda_1(\boldsymbol{\beta}), \lambda_2(\boldsymbol{\beta}), \lambda_3(\boldsymbol{\beta})\right)',$$

$$\lambda_i(\boldsymbol{\beta}) = E_{\boldsymbol{\beta}_0}\left(\psi_i(x;\theta,\sigma,\kappa)\right), i = 1, 2, 3.$$

Subsequently, we need to find the derivatives of these expressions with respect to $\boldsymbol{\beta}$ then evaluated at $\boldsymbol{\beta} = \boldsymbol{\beta}_0$ to obtain the matrix $-\Lambda(\boldsymbol{\beta}_0)$. The matrix $-\Lambda(\boldsymbol{\beta}_0)$ generalizes the Fisher information matrix.

It will be reduced to this matrix if the score functions $\psi_i(x;\boldsymbol{\beta}), i = 1, 2, 3$, are differentiable with respect to $\boldsymbol{\beta}$. It is clear that the elements of $\Lambda(\boldsymbol{\beta}_0)$ will have closed form expressions but are lengthy to display. To obtain $E_{\boldsymbol{\beta}_0}\left(\psi_i(x;\theta,\sigma,\kappa)\right), i = 1, 2, 3$, note that we have a location and scale parameter. Consequently, it appears to be simpler to define first the standardized AL density as the AL density with $\theta = 0, \sigma = 1$, *i.e.*,

$$g_\varepsilon(x;\kappa) = \frac{\sqrt{2}}{\gamma}\exp\left(\frac{\sqrt{2}}{2}\delta x\right)\exp\left(-\frac{\sqrt{2}}{2}\gamma(|x|)\right) \quad \text{and the AL density with three}$$

parameters as

$$g(x;\theta,\sigma,\kappa) = \frac{1}{\sigma}g_\varepsilon\left(\frac{x-\theta}{\sigma};\kappa\right).$$

Making use of $g_\varepsilon(x;\kappa)$,

$$E_{\boldsymbol{\beta}_0}\left(v(x;\theta)\right) = \int_\theta^\infty \frac{1}{\sigma_0}g_\varepsilon\left(\frac{x-\theta_0}{\sigma_0};\kappa_0\right)dx - \int_{-\infty}^\theta \frac{1}{\sigma_0}g_\varepsilon\left(\frac{x-\theta_0}{\sigma_0};\kappa_0\right)dx,$$

or

$$E_{\boldsymbol{\beta}_0}\left(v(x;\theta)\right) = 1 - 2G_\varepsilon\left(\frac{\theta-\theta_0}{\sigma_0};\kappa_0\right), \tag{17}$$

$G_\varepsilon(x;\kappa)$ is the distribution function with density function $g_\varepsilon(x;\kappa)$.

Similarly,

$$E_{\boldsymbol{\beta}_0}\left(|x-\theta|\right) = \int_\theta^\infty (x-\theta)g_\varepsilon\left(\frac{x-\theta_0}{\sigma_0};\kappa_0\right)dx + \int_{-\infty}^\theta (\theta-x)\frac{1}{\sigma_0}g_\varepsilon\left(\frac{x-\theta_0}{\sigma_0};\kappa_0\right)dx. \tag{18}$$

Therefore, $\dfrac{\partial E_{\boldsymbol{\beta}_0}\left(|x-\theta|\right)}{\partial\theta}$ can be obtained by first evaluating the term

$$\frac{\partial}{\partial\theta}\int_\theta^\infty (x-\theta)\frac{1}{\sigma_0}g_\varepsilon\left(\frac{x-\theta_0}{\sigma_0};\kappa_0\right)dx = -\int_\theta^\infty \frac{1}{\sigma_0}g_\varepsilon\left(\frac{x-\theta_0}{\sigma_0};\kappa_0\right)dx$$

$$= -\left[1 - G_\varepsilon\left(\frac{\theta-\theta_0}{\sigma_0};\kappa_0\right)\right],$$

using Leibnitz's rule which taking into account the lower bound of the interval also depends on θ then subsequently evaluate using Leibnitz's rule the expression

$$\frac{\partial}{\partial\theta}\int_{-\infty}^\theta (\theta-x)\frac{1}{\sigma_0}g_\varepsilon\left(\frac{x-\theta_0}{\sigma_0};\kappa_0\right)dx = G_\varepsilon\left(\frac{\theta-\theta_0}{\sigma_0};\kappa_0\right).$$

Consequently,

$$\frac{\partial E_{\boldsymbol{\beta}_0}\left(|x-\theta|\right)}{\partial \theta} = -1 + 2G_\varepsilon\left(\frac{\theta-\theta_0}{\sigma_0}; \kappa_0\right).$$

The elements of $\Lambda(\boldsymbol{\beta}_0)$ can be found subsequently by first forming

$$\lambda_1\left(\boldsymbol{\beta}; \boldsymbol{\beta}_0\right) = \int_{-\infty}^{\infty} \psi_1\left(x, \boldsymbol{\beta}\right) g\left(x; \boldsymbol{\beta}_0\right) \mathrm{d}x = -\frac{\sqrt{2}}{2}\frac{\delta}{\sigma} - \frac{\sqrt{2}}{2}\frac{\gamma}{\sigma}E_{\boldsymbol{\beta}_0}\left(v\left(x; \theta\right)\right),$$

$E_{\boldsymbol{\beta}_0}\left(v\left(x; \theta\right)\right)$ is as given by expression (17). Also,

$$\lambda_2\left(\boldsymbol{\beta}; \boldsymbol{\beta}_0\right) = \int_{-\infty}^{\infty} \psi_2\left(x, \boldsymbol{\beta}\right) g\left(x; \boldsymbol{\beta}_0\right) \mathrm{d}x$$

$$= -\frac{1}{\sigma} - \frac{\sqrt{2}}{2}\frac{\delta}{\sigma^2}\left(E_{\boldsymbol{\beta}_0}\left(x\right) - \theta\right) + \frac{\sqrt{2}}{2}\frac{\gamma}{\sigma^2}E_{\boldsymbol{\beta}_0}\left(|x-\theta|\right),$$

$E_{\boldsymbol{\beta}_0}\left(|x-\theta|\right)$ is as given by expression (18), $E_{\boldsymbol{\beta}_0}\left(x\right) = \theta_0 + \tau_0\mu_0$ with $\tau_0 = 1$ using expression (4). With

$$\lambda_3\left(\boldsymbol{\beta}; \boldsymbol{\beta}_0\right) = \int_{-\infty}^{\infty} \psi_3\left(x, \boldsymbol{\beta}\right) g\left(x; \boldsymbol{\beta}_0\right) \mathrm{d}x \quad \text{or equivalently,}$$

$$\lambda_3\left(\boldsymbol{\beta}; \boldsymbol{\beta}_0\right) = -\frac{\frac{\partial\gamma}{\partial\kappa}}{\gamma} + \frac{\partial\gamma}{\partial\kappa}\frac{\sqrt{2}}{\sigma}\left(E_{\boldsymbol{\beta}_0}\left(x\right) - \theta\right) - \frac{\sqrt{2}}{2}\frac{\partial\gamma}{\partial\kappa}E_{\boldsymbol{\beta}_0}\left(|x-\theta|\right),$$

then the matrix $\Lambda(\boldsymbol{\beta}_0)$ can be obtained by differentiating with respect to $\boldsymbol{\beta}$ the vector

$$\lambda\left(\boldsymbol{\beta}; \boldsymbol{\beta}_0\right) = \left(\lambda_1\left(\boldsymbol{\beta}; \boldsymbol{\beta}_0\right), \lambda_2\left(\boldsymbol{\beta}; \boldsymbol{\beta}_0\right), \lambda_3\left(\boldsymbol{\beta}; \boldsymbol{\beta}_0\right)\right)' \quad \text{and set} \quad \boldsymbol{\beta} = \boldsymbol{\beta}_0, \text{ i.e.,}$$

$$\Lambda(\boldsymbol{\beta}_0) = \left.\frac{\partial\lambda\left(\boldsymbol{\beta}; \boldsymbol{\beta}_0\right)}{\partial\boldsymbol{\beta}'}\right|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}.$$

Clearly, the elements of the matrix $\Lambda(\boldsymbol{\beta}_0)$ have closed form expressions but are lengthy to display. Packages like MATLAB or Mathematica can handle symbolic derivatives and can be used to obtain these elements. Substituting $\boldsymbol{\beta}_0$ by the ML estimator $\hat{\boldsymbol{\beta}}$ in $\Lambda(\boldsymbol{\beta}_0)$ yields an estimate for the matrix $\Lambda(\boldsymbol{\beta}_0)$.

Now we turn our attention to the matrix Σ which is the covariance matrix of the vector of score functions $\psi\left(x, \boldsymbol{\beta}_0\right) = \left(\psi_1\left(x, \boldsymbol{\beta}_0\right), \psi_2\left(x, \boldsymbol{\beta}_0\right), \psi_3\left(x, \boldsymbol{\beta}_0\right)\right)'$. Using a different but equivalent parameterisation, this matrix has been obtained by Kotz *et al.* ([16], p. 818), Kotz *et al.* ([1], p. 158) but its inverse does not give the asymptotic covariance matrix of the ML estimators as claimed in their paper. It is not difficult to establish the relationships between the parameterisation used in example 2 and the one used in the paper by Kotz *et al.* ([1], p. 818).

Note that the inverse of Σ is not the asymptotic covariance matrix of the ML estimators is due to $\Lambda(\boldsymbol{\beta}_0) = \left.\dfrac{\partial\lambda\left(\boldsymbol{\beta}; \boldsymbol{\beta}_0\right)}{\partial\boldsymbol{\beta}'}\right|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}$ is not equal to $-\Sigma^{-1}$ if the differentiability assumptions for the score functions do not hold, see corollary (3.2) and proposition (3.3) given by Huber ([15], p. 133).

The matrix Σ can also be estimated by the following estimator

$$\frac{1}{n}\sum_{i=1}^{n}\left[\psi\left(x_i, \hat{\boldsymbol{\beta}}\right)\right]\left[\psi\left(x_i, \hat{\boldsymbol{\beta}}\right)\right]'.$$

Let us consider the following location model with known σ_0 and check the expression (2.2) as given by Kotz *et al.* ([16], p. 818) has missing components. The density function is given by

$$f(x;\theta) = \frac{1}{\sigma_0\sqrt{2}} e^{-\frac{\sqrt{2}}{\sigma_0}|x-\theta|}, \text{ or alternatively the density can also be expressed as}$$

$$f(x;\theta) = f_0(x-\theta), f_0(x) = \frac{1}{\sigma_0\sqrt{2}} e^{-\frac{\sqrt{2}|x|}{\sigma_0}}.$$

This subfamily will correspond to their parametrisation with $\kappa = 1$ in their paper. The sample median $\hat{\theta}$ is the ML estimator for θ, using their result it will lead to conclude that the asymptotic variance is given by $\left(E\left(\frac{\partial \ln f}{\partial \theta}\right)^2\right)^{-1} = \frac{\sigma_0^2}{2}$, as indicated by case1 in the table of their paper. On the other hand, it is known that the asymptotic distribution of the sample median is given by

$$\sqrt{n}\left(\hat{\theta} - \theta_0\right) \xrightarrow{L} N\left(0, \frac{1}{4\left(f_0(0)\right)^2}\right), \text{ see expression (2.4.19) given by Lehmann}$$

([25], p. 81) for example. For the location model being considered, we have $\frac{1}{4\left(f_0(0)\right)^2} = \frac{1}{8\sigma_0^2}$. Clearly, $\frac{1}{8\sigma_0^2} \neq \frac{\sigma_0^2}{2}$ but the correct asymptotic variance can be obtained using expression (13).

For the general GAL distribution with four parameters, alternative methods of estimation based on quadratic distances (QD) which make use of the empirical cumulant generating function will be introduced in the next section. The QD methods are developed based on empirical findings which show that the ML methods for finite sample sizes as large as $n = 5000$ do not give good estimates for the shape parameter τ and the scale parameter σ but ML methods give good estimates for the other two parameters. Howewer, the overall efficiency of ML methods lags behind QD methods in finite samples. Also, QD methods beside giving better estimates for σ and τ, the methods can be used for parameter testing since the asymptotic covariance matrix for the QD estimators can be obtained explicitly for the entire GAL family. The methods also provide a chi-square test statistics of goodness-of-fit for the model being used. Therefore, it might be of interests to consider using QD methods whenever ML methods might have deficiencies.

4. Quadratic Distance Methods

General Quadratic distance (QD) theory has been developed in Luong and Thompson [26]. Howewer, if it is used for estimating parameters of the GAL distribution we need to specify a distance which can generate estimators with good efficiencies. For applied works, it is also preferable to have methods which are relatively simple to implement numerically.

For financial data, observations are recorded as percentages so they are small

in magnitude, we recommend minimizing the following distance based on matching the empirical cumulant generating function $K_n(t)$ with its model counterpart $K_{\boldsymbol{\beta}}(t)$ using the following points

$$t_j, j = 1, \cdots, m = 20$$

with

$$t_1 = 0.01, t_2 = 0.02, \cdots, t_{10} = 0.1, t_{11} = -0.01, t_{12} = -0.02, \cdots, t_{20} = -0.1. \quad (19)$$

The choice of points as given above is suggested based on empirical findings that overall, the QD estimators are more efficient than the ML estimators for the range of parameters often encountered for modelling financial data using finite sample sizes as large as $n = 5000$. Note that the set of points chosen does not include the origin 0.

The empirical moment generating function, empirical cumulant generating function are given respectively by

$$M_n(s) = \frac{1}{n} \sum_{i=1}^{n} e^{sX_i} \quad \text{and} \quad K_n(s) = \log M_n(s).$$

The model cumulant generating function is $K_{\boldsymbol{\beta}}(t)$, $K_{\boldsymbol{\beta}}(t) = \log M_{\boldsymbol{\beta}}(t)$ with $M_{\beta}(s)$ being the model moment generating function as defined by expression (1). The proposed QD estimators given by the vector $\tilde{\boldsymbol{\beta}}$ is obtained by minimizing with respect to $\boldsymbol{\beta}$ the following specific QD distance given by

$$D(\boldsymbol{\beta}) = \sum_{j=1}^{20} \left(K_n(s_j) - K_{\boldsymbol{\beta}}(s_j) \right)^2. \quad (20)$$

Once the estimates are obtained, goodness of fit test statistics with an asymptotic chi-square distribution with $r = 16$ degree of feedom can also be constructed. General QD distances theory can be used to derive the asymptotic covariance matrix of the QD estimators and the chi-square goodness of fit test statistics. They will be given at the end of this section. Having the asymptotic covariance matrix of the QD estimators in closed form for the GAL family is useful for parameter testing.

For notations, let us define the vector based on observations

$$z_n = \left(K_n(t_1), \cdots, K_n(t_m) \right)', m = 20.$$

Its model counterpart is the vector

$$z_{\boldsymbol{\beta}} = \left(K_{\boldsymbol{\beta}}(t_1), \cdots, K_{\boldsymbol{\beta}}(t_m) \right)'.$$

Therefore,

$$D(\boldsymbol{\beta}) = \left(z_n - z_{\boldsymbol{\beta}} \right)' \left(z_n - z_{\boldsymbol{\beta}} \right).$$

Observe that the elements of the covariance matrix $\boldsymbol{V}_M$ for the vector $\sqrt{n} \left(M_n(t_1), \cdots, M_n(t_m) \right)'$ are given by

$$V_M(i, j) = M_{\boldsymbol{\beta}}(t_i + t_j) - M_{\boldsymbol{\beta}}(t_i) M_{\boldsymbol{\beta}}(t_j), i = 1, \cdots, 20, j = 1, \cdots, 20.$$

The elements of the approximate covariance matrix based on the differential method or delta method for

$$\sqrt{n}\left(K_n\left(t_1\right),\cdots,K_n\left(t_m\right)\right)'$$

are given by

$$V_K\left(i,j\right)=\left(M_{\boldsymbol{\beta}}\left(t_i+t_j\right)-M_{\boldsymbol{\beta}}\left(t_i\right)M_{\boldsymbol{\beta}}\left(t_j\right)\right)\Big/\left(M_{\boldsymbol{\beta}}\left(t_i\right)M_{\boldsymbol{\beta}}\left(t_j\right)\right),i=1,\cdots,20,j=1,\cdots,20.$$

Under the regularity conditions given by Lemma (3.4.1) of Luong and Thompson ([26], p. 244), the QD estimators given by the vector $\tilde{\boldsymbol{\beta}}$ are consistent. Clearly, we need to assume that on the restricted parameter space the model moment generating function and the covariance matrix $\boldsymbol{V}_M$ given by expression (21) are well defined. Some modifications might be necessary if the methods are applied to other models. The conditions are met in general for the GAL distribution when used for modeling financial data. We then have

$$\sqrt{n}\left(\tilde{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\right)\xrightarrow{\;L\;}N\left(0,\boldsymbol{V}\right),$$

$$\boldsymbol{V}=\left(\boldsymbol{S}'\boldsymbol{S}\right)^{-1}\boldsymbol{S}'V_K\boldsymbol{S}\left(\boldsymbol{S}'\boldsymbol{S}\right)^{-1}.$$

The asymptotic covariance for the QD estimators is simply $\dfrac{1}{n}\boldsymbol{V}$.

All the expressions which form V as given above are evaluated under the true vector of parameters $\boldsymbol{\beta}_0$, $\boldsymbol{\beta}=\left(\theta,\mu,\sigma,\tau\right)'=\left(\beta_1,\beta_2,\beta_3,\beta_4\right)'$ and

$$\boldsymbol{S}=\begin{bmatrix}\dfrac{\partial K_{\boldsymbol{\beta}}\left(t_1\right)}{\partial\beta_1} & \cdots & \dfrac{\partial K_{\boldsymbol{\beta}}\left(t_1\right)}{\partial\beta_4}\\[2mm]\vdots & \ddots & \vdots\\[2mm]\dfrac{\partial K_{\boldsymbol{\beta}}\left(t_m\right)}{\partial\beta_1} & \cdots & \dfrac{\partial K_{\boldsymbol{\beta}}\left(t_m\right)}{\partial\beta_4}\end{bmatrix},\ \boldsymbol{S}'\text{ is the transpose of }\boldsymbol{S}.$$

We also use $\boldsymbol{S}=\boldsymbol{S}\left(\boldsymbol{\beta}_0\right),\boldsymbol{V}_K=\boldsymbol{V}\left(\boldsymbol{\beta}_0\right),\boldsymbol{\Sigma}_2=\boldsymbol{\Sigma}_2\left(\boldsymbol{\beta}_0\right)$ to emphasize that these matrices depend on β_0. The matrix $\boldsymbol{\Sigma}_2$ is derived below. For constructing test statistics with chi-square limiting distribution, use expression (3.4.2) given by Luong and Thompson ([26], p. 248) to obtain

$\sqrt{n}\left(z_n-z_{\tilde{\boldsymbol{\beta}}}\right)\xrightarrow{\;L\;}N\left(0,\boldsymbol{\Sigma}_2\right)$ with $\boldsymbol{\Sigma}_2$, a covariance matrix which depends on $\boldsymbol{\beta}_0$ and

$$\boldsymbol{\Sigma}_2=\left[\boldsymbol{I}-\boldsymbol{S}\left(\boldsymbol{S}'\boldsymbol{S}\right)^{-1}\boldsymbol{S}'\right]V_K\left[\boldsymbol{I}-\boldsymbol{S}\left(\boldsymbol{S}'\boldsymbol{S}\right)^{-1}\boldsymbol{S}'\right]. \tag{21}$$

In practice, $\boldsymbol{\beta}_0$ needs te be replaced by $\tilde{\boldsymbol{\beta}}$ so that an estimate of $\boldsymbol{\Sigma}_2$ can be defined as

$$\tilde{\boldsymbol{\Sigma}}_2=\boldsymbol{\Sigma}_2\left(\tilde{\boldsymbol{\beta}}\right).$$

We need to find the Moore-Penrose (MP) generalized inverse for $\tilde{\boldsymbol{\Sigma}}_2$ to constructa chi-square statistics. The quadratic form constructed with the MP inverse will follow a chi-square distribution asymptotically. Many computer packages provide prewritten functions to find the Moore-Penrose inverse of a matrix. It can also be computed easily using the spectral decomposition of $\tilde{\boldsymbol{\Sigma}}_2$, *i.e.*, using the representation $\tilde{\boldsymbol{\Sigma}}_2=\boldsymbol{PDP}'$. The columns of the matrix $\boldsymbol{P}$ are the eigenvectors of $\tilde{\boldsymbol{\Sigma}}_2$ and $\boldsymbol{D}$ is a diagonal matrix with the diagonal elements being the corresponding eigenvalues of $\tilde{\boldsymbol{\Sigma}}_2$ given respectively by $\lambda_i\geq0,i=1,\cdots,m$.

The matrix P is orthonormal with the property $PP' = I$.

The Moore Penrose inverse $\tilde{\Sigma}_2^{MP}$ can be obtained as

$\tilde{\Sigma}_2^{MP} = PD^-P'$ with

D^- being the diagonal matrix constructed based on the diagonal elements $\lambda_i, i = 1, \cdots, 20$ of D. The diagonal elements of D^- are given as

$$\lambda_i^- = \frac{1}{\lambda_i} \text{ if } \lambda_i > 0 \text{ and } \lambda_i^- = 0 \text{ if } \lambda_i = 0.$$

For discussions on property of the Moore Penrose generalized inverse, see Theil ([27], p 273-274), also see expressions (4.3 - 4.6) given by Harville ([28], p 504). For numerical computations using R, see section 8.3 given by Fieller ([29], p. 123-133). The chi-square test statistics for testing the null hypothesis which specifies that observations are drawn from the GAL family can be based on the criterion function

$$Q(\boldsymbol{\beta}) = n\left(z_n - z_{\boldsymbol{\beta}}\right)'\left(PD^-P'\right)\left(z_n - z_{\boldsymbol{\beta}}\right), \tag{22}$$

$$Q(\tilde{\boldsymbol{\beta}}) = n\left(z_n - z_{\tilde{\boldsymbol{\beta}}}\right)'\left(PD^-P'\right)\left(z_n - z_{\tilde{\boldsymbol{\beta}}}\right) \xrightarrow{L} \chi^2(16). \tag{23}$$

The limiting distribution of the test statistics is chi-square with $r = 16$, based on Theorem 3.4.1 of Luong and Thompson ([26], p. 248). The test statistics can also be viewed as a generalized Pearson test statistics. The criterion function $Q(\boldsymbol{\beta})$ can also be used to find a good starting vector to initialize the algorithms for finding the QD estimators, see section (3) given by Andrews ([30], p. 917-922) for more discussions and section (5.2) of this paper.

5. Numerical Issues

5.1. Simplemoment Estimators

The simple approximate moment estimate proposed by Senata [4] can be found explicitly and can be used as starting points for numerical optimization to find QDE or MLE. Let the first four moments be denoted by $\hat{\mu}_{sj}, j = 1, 2, 3, 4$ with $\hat{\mu}_j = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^j, j = 1, \cdots, 4$ and equalizing with the model counterparts and neglecting all the terms with $\theta'^j, j = 2, 3, 4$ yields the following system of estimating equation for moment estimation,

$\hat{\mu}_1 = c + \theta', \hat{\mu}_2 = \sigma'^2, \hat{\mu}_3 = 3\sigma'^2\theta'v, \hat{\mu}_4 = 3\sigma'^4 v + 3\sigma'^4$. The moment estimators are

$$(\bar{\sigma}')^2 = \hat{\mu}_2, \bar{v} = \frac{\frac{\hat{\mu}_4}{3} - (\bar{\sigma}')^4}{(\bar{\sigma}')^4}, \bar{\theta}' = \frac{\hat{\mu}_3}{3\bar{v}(\bar{\sigma}')^2}, \bar{c} = \hat{\mu}_1 - \bar{\theta}'.$$ When converted to the

parameterization given by Kotz et al. [16], the approximate moment estimators for $\tau, \sigma^2, \mu, \theta$ are given respectively as $\bar{\tau} = \frac{1}{\bar{v}}, \bar{\sigma}^2 = \bar{\sigma}'^2\bar{v}, \bar{\mu} = \bar{\theta}'\bar{v}, \bar{\theta} = \bar{c}$. The

approximate moment estimators are not efficient but they are simple and given explicitly .Therefore, they can be used as starting points for the numerical algorithms to implement QD or ML estimation. Moment estimators can also be veri-

fied to see whether they are appropriate as starting points. This will be discussed in the next section.

5.2. The Choice of an Initial Vector

Most of the algorithms will return a local minimizer and the vector which gives the estimators is defined to be the global minimizer. Due to this limitation, some cares are needed to ensure that we can identify the global minimizer. In practice, it is important to test the algorithm with various starting vectors, see Andrews [30]. Andrews [30] has suggested that it is preferable to have the starting vector $\boldsymbol{\beta}^{(0)} = \left(\theta^{(0)}, \mu^{(0)}, \sigma^{(0)}, \tau^{(0)} \right)'$ close to the vector of the estimators given by $\tilde{\boldsymbol{\beta}}$ which globally minimizes the objective function. We might look for a different starting vector if the vector of moment estimators cannot be used as a starting vector to initialize the numerical algorithm.

The criterion function $Q(\boldsymbol{\beta})$ given by expression (22) which is used to construct goodness of fit test can also be used to select a good starting vector. The starting vector $\boldsymbol{\beta}^{(0)}$ is subject to the screening test by checking whether

$$Q\left(\boldsymbol{\beta}^{(0)}\right) \le \chi^2_{0.95}(16),$$

$\chi^2_{0.95}(16)$ is the 95th percentile of the chi-square distribution with 16 degree of freedom to be qualified as a suitable starting vector, see expression (3.5) given by Andrews ([30], p. 919). If $\boldsymbol{\beta}^{(0)}$ passes the screening test then one might consider to use $\boldsymbol{\beta}^{(0)}$ as the vector of starting points for the numerical algorithm used to find the vector of estimators, otherwise look for another one.

5.3. A limited Simulation Study

For financial data, observations are recorded as percentages so they are small in magnitude. We are in the situation of modeling with values for θ and μ are near 0. The plausible values for τ and $\sigma, 0 < \tau \le 10, 0 < \sigma \le 0.1$. For parameters with these ranges we observe that the ML estimators for τ and σ do not perform well for sample size as large as $n = 5000$. For comparisons between QD methods vs ML methods, the ratio of total Mean square errors is used as a measure for the overall relative efficiency. Due to the limited capacity on computing as we only have access to a laptop computer, we can only use $M = 100$ samples with each sample is of a size $n = 5000$.

The overall relative efficiency for comparisons is defined as the ratio

$$\frac{\text{TMSE}(QD)}{\text{TMSE}(ML)} = \frac{\text{MSE}(\tilde{\theta}) + \text{MSE}(\tilde{\mu}) + \text{MSE}(\tilde{\sigma}) + \text{MSE}(\tilde{\tau})}{\text{MSE}(\hat{\theta}) + \text{MSE}(\hat{\mu}) + \text{MSE}(\hat{\sigma}) + \text{MSE}(\hat{\tau})}.$$

The expressions for MSE and TMSE which appear in Table 1 are estimated using simulated samples. The results of the simulation study are lengthy. We only extract the key findings, which is summarized using Table 1.

The study seems to indicate that overall ML methods are less efficient than QD methods but ML methods are more efficient for estimating the first two

Table 1. Illustrations of simulation results.

(a)

Parameters values	$\theta = 0$	$\mu = -0.0001$	$\sigma = 0.001$	$\tau = 0.5$
MSE(ML)	5.376×10^{-12}	2.960×10^{-10}	2014.178×10^{-13}	$534911.573 \times 10^{-11}$
MSE(QD)	6118.876×10^{-12}	104.562×10^{-10}	1.099×10^{-13}	5.083×10^{-11}
MSE(QD)/ MSE(ML)	1138.031	35.316	5.457×10^{-4}	9.502×10^{-6}

Overall relative efficiency: $\dfrac{\mathrm{TMSE(QD)}}{\mathrm{TMSE(ML)}} = 0.003$.

(b)

Parameters values	$\theta = 0$	$\mu = -0.0001$	$\sigma = 0.001$	$\tau = 1.0$
MSE (ML)	3.327×10^{-12}	5.788×10^{-11}	9.634×10^{-10}	2.596×10^{-5}
MSE (QD)	$263740.061 \times 10^{-12}$	$27671.291 \times 10^{-11}$	0.105×10^{-10}	0.000×10^{-5}
MSE (QD)/ MSE(ML)	79,268.320	4771.434	0.0109	1.967×10^{-6}

Overall relative efficiency: $\dfrac{\mathrm{TMSE(QD)}}{\mathrm{TMSE(ML)}} = 0.0207$.

(c)

Parameters values	$\theta = 0$	$\mu = -0.0001$	$\sigma = 0.001$	$\tau = 2.0$
MSE(ML)	1.420×10^{-8}	4.860×10^{-8}	8882.744×10^{-9}	5.6984×10^{-4}
MSE(QD)	3.518×10^{-8}	11.500×10^{-8}	0.001×10^{-9}	0.000×10^{-4}
MSE(QD)/ MSE(ML)	2.5265	2.3658	0.0001	4.19341×10^{-7}

Overall relative efficiency: $\dfrac{\mathrm{TMSE(QD)}}{\mathrm{TMSE(ML)}} = 8.357 \times 10^{-5}$.

parameters namely θ, μ for the AL family and for the entire GAL family in finite samples where little is known about the asymptotic distributions of the ML estimators.

6. Financial Applications

6.1. Option Pricing and Risk Neutral Parameters

For options as they are tradable, risk neutral parameters are used for pricing. Risk neutral parameters are related to the physical parameters which can be estimated using historical data. A set of risk neutral parameters can be obtained by using the Esscher transform change of measure, see Schoutens ([31], p 77) based on the seminal works of Gerber and Shiu [12]. They can also be viewed as minimum entropy risk neutral parameters, see Miyahara [13]. We keep the four historical parameters of the GAL distribution as risk neutral parameters but introduce an extra parameter h^* which is given by the following equation with h being the unknown variable and r is the known risk free rate,

$$r = \log M(h+1) - \log M(h)$$

where $M(s)$ is the moment generating function as given by expression (2).

Therefore, the risk neutral parameters are given by the vector

$$\beta_0^N = \left(h^*, \theta_0, \sigma_0, \mu_0, \tau_0\right)'.$$

The price of the asset is modeled as $S_T = S_0 e^{X_T}$ where:

a) S_0 is the initial asset price at time $t = 0$,

b) $X_T = \sum_{i=1}^{T} R_i$,

c) the log returns $R_i = \log(S_{i+1}) - \log(S_i), i = 1, \cdots, T$ are i.i.d as $R \sim GAL(\beta)$ with mgf $M_\beta(s)$.

We also assume $T \geq 1$ and T is a positive integer.

For pricing an European call option with the initial price S_0, strike price K and interest rate r, the price of the European call option is $e^{-rT} E\left((S_T - K)_+\right)$ where $(S_T - K)_+ = \max\left((S_T - K), 0\right)$ and the expectation is under risk neural parameters. Therefore, it is possible use simulated samples from a bilateral gamma distribution to obtain an estimate for $E\left((S_T - K)_+\right)$ and price the option.

Senata ([4], p. 182-184) has illustrated the use of the GAL family, moment and ML methods to analyze historical data from the Dow Jones industrial average and other indexes. It is not difficult to see that QD methods can be considered as alternative methods for analyzing financial data.

Beside option pricing, measures of risks are used in finance and actuarial sciences. These measures will depend on the underlying distribution which is specified by a set of parameters. We briefly discuss these notions below. The inferences techniques can also be applied to estimate the parameters using historical data and quantify the level of risks incurred.

6.2. VaR, CVar, EvaR Using the GAL Distribution

The Value at Risk at confidence level $1 - \alpha$ of a continuous loss random variable X with distribution function $F(x)$ and density function $f(x)$ is defined as

$VaR_{1-\alpha}(L) = F^{-1}(1-\alpha)$ is the quantile of the loss $X = -R$ specified by

$P\left(X > VaR_{1-\alpha}(X)\right) = \alpha$, the probability of the potential loss encountered by the holder of a financial assetfor one unit of time. The conditional value at risk

$CVaR_{1-\alpha}(X) = \frac{1}{\alpha} \int_{VaR_{1-\alpha}}^{\infty} xf(x)\mathrm{d}x$, see Rockafellar and Uriyasev [32] for this measure of risk. If the log return random variable R follows a $GAL(\theta, \sigma, \mu, \tau)$, $R \sim GAL(\theta, \sigma, \mu, \tau)$, then the loss random variable is $X \sim GAL(-\theta, \sigma, -\mu, \tau)$.

Ahmadi-Javid [33] proposed a coherent measure of risk, the entropic value-at risk (EVaR) using the Chernoff bound, see the seminal paper by Chernoff [34] for the bound. EVaR is defined implicitly using of the moment generating function $M(z)$. Since the moment generating function of the GAL distribution is relatively simple and does not involve the Bessel function, Evar can also be computed easily. For more discussions on estimation and risk measures, see

Toma and Dedu [35].

7. Conclusion

As we can see in finite samples, ML methods only offer good estimators for two of the four parameters for the GAL family. Asymptotic normality can only be guaranteed for the AL family and the lack of a covariance matrix in closed form prevents hypotheses testing for the GAL family. Due to these restrictions, QD methods are developed as complementary methods to ML methods. The methods appear to be suitable for estimation and for parameter testing. The methods also produce a criterion function when evaluated at the values taken by the QD estimators gives a chi-square goodness-of-fit test statistics for the GAL model. The criterion function can be used to select a starting vector which is close to the vector of the QD estimators to start a numerical search algorithm. These last two features are not shared directly by ML methods and appear to be useful for applications.

Acknowledgements

References

[1] Kotz, S., Kozubowski, T.J. and Podgorski, K. (2001) The Laplace Distribution and Generalizations. Birkhauser, Boston. https://doi.org/10.1007/978-1-4612-0173-1

[2] Madan, D.P. and Senata, E. (1990) The Variance Gamma (VG) Model for Share Market. *Journal of Business*, **63**, 551-524. https://doi.org/10.1086/296519

[3] Madan, D.P., Carr, P. and Chang, E.C. (1998) The Variance Gamma Process and Option Pricing. *European Finance Review*, **2**, 79-105. https://doi.org/10.1086/296519

[4] Seneta, E. (2004) Fitting the Variance Gamma Model to Financial Data. *Journal of Applied Probability*, **41**, 177-187. https://doi.org/10.1017/S0021900200112288

[5] Podgorski, K. and Wegener, J. (2011) Estimation for Stochastic Models Driven by Laplace Motion. *Communications in Statistics, Theory and Methods*, **40**, 3281-3302. https://doi.org/10.1080/03610926.2010.499051

[6] McNeil, A.J., Frey, R. and Embrechts, P. (2005) Quantitative Risk Management. Princeton University Press, Princeton.

[7] Protassov, R.S. (2004) EM-Based Maximum Likelihood Parameter Estimation for Multivariate Generalized Hyperbolic Distributions with Fixed λ. *Statistics and Computing*, **14**, 67-77. https://doi.org/10.1023/B:STCO.0000009419.12588.da

[8] Hu, W. (2005) Calibration of Multivariate Generalized Hyperbolic Distributions Using the EM Algorithm with Applications in Risk Management, Portfolio Optimization and Portfolio Credit Risk. Unpublished PHD Thesis, Department of Mathematics, The Florida State University, Tallahassee.

[9] Louis, T.A. (1982) Finding the Observed Information Using the EM Algorithm. *Journal of the Royal Statistical Society Series B*, **44**, 98-130.

[10] McLachlan, G.J. and Krishnan, T. (2008) The EM Algorithm and Extensions. 2nd Edition, Wiley, New York. https://doi.org/10.1002/9780470191613

[11] Küchler, U. and Tappe, S. (2008) Bilateral Gamma Distributions and Processes in Fi-nancial Mathematics. *Stochastic Processes and Their Applications*, **118**, 261-283.

[12] Gerber, H.U. and Shiu, E.S.W. (1994) Option Pricing by Esscher Transforms. *Transactions of the Society of Actuaries*, **46**, 99-191.

[13] Miyahara, Y. (2012) Option Pricing in Incomplete Markets: Modeling Based on Geometric Lévy Processes and Minimal Entropy Martingales Measures. Imperial College Press, London.

[14] Bierlaire, M. (2006) Introduction à l'optimisation différentiable. Presses Polytechniques et Universités Romandes, Lausanne.

[15] Huber, P. (1981) Robust Statistics. Wiley, New York. https://doi.org/10.1002/0471725250

[16] Kotz, S., Kozubowski, T.J. and Podgorski, K. (2002) Maximum Likelihood Estimation of Asymmetric Laplace Parameters. *Annals of Institute of Statistical Mathematics*, **54**, 816-826. https://doi.org/10.1023/A:1022467519537

[17] Huber, P. (1967) The Behaviour of Maximum Likelihood Estimates under Nonstandard Conditions. In: *Proceeding* 5*th Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, University of California Press, Berkeley.

[18] Amemiya, T. (1985) Advanced Econometrics. Harvard University Press, Cambridge.

[19] Newey, W.K. and McFadden, D. (1994) Large Sample Estimation and Hypothesis Testing. In: Engle, R.F. and McFadden, D., Eds., *Handbook of Econometrics*, Vol. 4, North Holland, Amsterdam.

[20] Broniatowski, M., Toma, A. and Vajda, I. (2012) Decomposable Pseudodistances and Application in Statistical Estimation. *Journal of Statistical and Planning Inference*, **142**, 2574-2585.

[21] Hogg, R., McKean, J.W. and Craig, A.T. (2013) Introduction to Mathematical Statistics. 7th Edition, Pearson, New York.

[22] Hinkley, D.V. and Revankar, N.S. (1977) Estimation of the Pareto Law form Underreported Data: A Further Analysis. *Journal of Econometrics*, **5**, 1-11.

[23] Rudin, W. (1976) Principles of Mathematical Analysis. McGraw Hill, New York.

[24] Woolridge, J.M. (2010) Econometric Analysis of Cross Section and Panel Data. 2nd Edition, MIT Press, Cambridge.

[25] Lehmann, E.L. (1999) Elements of Large Sample Theory. Springer, New York. https://doi.org/10.1007/b98855

[26] Luong, A. and Thompson, M.E. (1987) Minimum Distance Methods Based on Quadratic Distances for Transforms. *Canadian Journal of Statistics*, **15**, 239-251. https://doi.org/10.2307/3314914

[27] Theil, H. (1971) Principle of Econometrics. Wiley, New York.

[28] Harville, D.A. (1997) Matrix Algebra from a Statistician's Perspective. Springer, New York. https://doi.org/10.1007/b98818

[29] Fieller, N. (2016) Basics of Matrix Algebra with R. Chapman and Hall, New York.

[30] Andrews, D.W.K. (1997) A Stopping Rule for the Computation of the Generalized Method of Moments Estimators. *Econometrica*, **65**, 913-931. https://doi.org/10.2307/2171944

[31] Schoutens, W. (2003) Lévy Processes in Finance: Pricing Financial Derivatives. Wiley, New York. https://doi.org/10.1002/0470870230

[32] Rockafellar, R.T. and Uriyasev, S. (2002) Conditional Value-at-Risk for General Loss Distribution. *Journal of Banking and Finance*, **26**, 1443-1471.

[33] Ahmadi-Javid, A. (2012) Entropic Value-At Risk: A New Coherent Risk Measure. *Journal of Optimization: Theory and Applications*, **155**, 1105-1123. https://doi.org/10.1007/s10957-011-9968-2

[34] Chernoff, H. (1952) A Measure of Asymptotic Efficiency for Tests of a Hypothesis Based on a Sum of Observations. *Annals of Mathematical Statistics*, **23**, 497-507. https://doi.org/10.1214/aoms/1177729330

[35] Toma, A. and Dedu, S. (2014) Quantitative Techniques for Financial Risk Assessment: A Comparative Approach Using Different Risk Measures and Estimation Methods. *Procedia Economics and Finance*, **8**, 712-719.