Inference for a Leptokurtic Symmetric Family of Distributions Represented by the Difference of Two Gamma Variates

Maciej Augustyniak and Louis G. Doray

Département de mathématiques et de statistique, Université de Montréal, C.P. 6128, Succursale Centre-ville, Montréal, Québec, Canada H3C 3J7

Key Words and Phrases: Empirical characteristic function, quadratic distance, parameter estimation, leptokurtosis, gamma distribution, Laplace distribution, goodness-of-fit, normality test, test of symmetry.

ABSTRACT

We introduce a family of leptokurtic symmetric distributions represented by the difference of two gamma variates. Properties of this family are discussed. The Laplace, sums of Laplace and normal distributions all arise as special cases of this family. We propose a two-step method for fitting data to this family. First, we perform a test of symmetry, and second, we estimate the parameters by minimizing the quadratic distance between the real parts of the empirical and theoretical characteristic functions. The quadratic distance estimator obtained is consistent, robust and asymptotically normally distributed. We develop a statistical test for goodness-of-fit and introduce a test of normality of the data. A simulation study is provided to illustrate the theory.

1 INTRODUCTION

We introduce a family of leptokurtic symmetric distributions by presenting its characteristic function. Consider X and Y to be independent and identically distributed random variables from a gamma distribution with parameters of shape $1/\lambda$ and scale $\sqrt{\lambda\theta}$ (i.e. $X, Y \sim \Gamma(1/\lambda, \sqrt{\lambda\theta})$), where λ and θ are defined on the positive real line. The new family is represented by the random variable Z, where $Z = X - Y$ with characteristic function

$$\phi_Z(t) = \phi_X(t)\phi_{-Y}(t) = \left(\frac{1}{1 - it\sqrt{\lambda\theta}}\right)^{1/\lambda} \left(\frac{1}{1 + it\sqrt{\lambda\theta}}\right)^{1/\lambda} = \left(\frac{1}{1 + t^2\lambda\theta}\right)^{1/\lambda}.$$

In the limit, as $\lambda \to 0$, we have $\phi_Z(t) \to e^{-t^2\theta}$, which is the characteristic function of a normal random variable centered at 0 and with variance 2θ.

Hence, we define this family of symmetric distributions by its characteristic function

$$\phi(t) = \begin{cases} \left(\frac{1}{1+t^2\lambda\theta}\right)^{1/\lambda} & \text{for } \lambda, \theta > 0, \\ e^{-t^2\theta} & \text{for } \lambda = 0, \theta > 0. \end{cases} \tag{1.1}$$

We will use the notation $DGD(\lambda, \theta)$ for the double gamma difference distribution with parameters (λ, θ) and characteristic function given by (1.1). For $\lambda < 0$ or $\theta < 0$, $\phi(t)$ is not a characteristic function because $|\phi(t)|$ is not bounded by 1. Please refer to Lukacs (1970) for more details on the properties of characteristic functions.

When $\lambda = 1$, the family becomes the difference of two independent exponentially distributed variates with mean $\sqrt{\theta}$. Kotz, Kozubowski and Podgórski (2001) proved that a Laplace random variable centered at the origin can be represented as the difference between two independent exponentially distributed variates; the characteristic function of a classical Laplace random variable centered at 0 with scale paramter s is

$$\frac{1}{1+t^2s^2}.$$

Hence, the classical Laplace distribution is a special case of the DGD family with parameters $(\lambda = 1, \theta = s^2)$. When $\lambda = 1/n$, $n \in \mathbb{N}$, the difference of two independent gamma variates can be seen as the sum of n differences of two independent exponential variates which is simply the sum of n independent Laplace variates. In the limit, when $n \to \infty$ (i.e. $\lambda \to 0$), our result is consistent with the Central Limit Theorem, as the sum of n independent Laplace variates converges to the normal distribution. We now list some properties of this family.

Property 1: Odd and even moments

Since its characteristic function is real and even, it is a family of symmetric density functions centered at 0. Thus, the odd moments are 0; the positive even moments can be calculated from the formula in Proposition 1, the proof of which is in Appendix A.

Proposition 1. *Let Z be a $DGD(\lambda, \theta)$ random variable with characteristic function $\phi(t)$ as defined by* (1.1). *Then,*

$$\mathrm{E}[Z^{2k}] = \frac{\theta^k (2k)!}{k!} \prod_{j=0}^{k-1} (1 + j\lambda), \quad k = 1, 2, 3, \ldots.$$

Property 2: Kurtosis

From Proposition 1, we obtain the variance and the kurtosis of a $DGD(\lambda, \theta)$ random variable which are 2θ and $(3 + 3\lambda)$ respectively. Kurtosis is defined as the normalized fourth central moment and it is a measure of peakedness and of heaviness of the tails. Since $\lambda \geq 0$, the kurtosis is always greater or equal to 3. Thus, the family is leptokurtic because the kurtosis is always at least that of the normal distribution.

Property 3: Closure under transformations
Let $Z_1, \ldots, Z_n$ be independent and identically distributed $DGD(n\lambda, \theta/n)$ variates and consider $Z = Z_1 + \ldots + Z_n$, then

$$\phi_Z(t) = \phi_{Z_1}(t) \cdots \phi_{Z_n}(t) = [\phi_{Z_1}(t)]^n = \left(\frac{1}{1 + t^2\lambda\theta}\right)^{1/\lambda}. \tag{1.2}$$

Clearly, Z is a $DGD(\lambda, \theta)$ random variable. Also, if $a \in \mathbb{R}$, then the characteristic function of aZ is $\phi_{aZ}(t)$, where

$$\phi_{aZ}(t) = \phi_Z(at) = \left(\frac{1}{1 + t^2a^2\lambda\theta}\right)^{1/\lambda}.$$

This entails that aZ is a $DGD(\lambda, a^2\theta)$ random variable. Thus, the family is closed under scale and convolution operations but not under the translation operation as the center of symmetry is fixed at the origin. Moreover, from (1.2) we can recognize that its characteristic function is infinitely divisible.

When the family reduces to Lapace or normal random variables, the density function can be expressed in a closed form. For example, when $\lambda = 1$, we obtain the classical Laplace random variable centered at 0 with density function equal to

$$f(z; \lambda = 1, \theta = s^2) = \frac{1}{2s} e^{-|z|/s}.$$

When $\lambda = 1/n$, $n \in \mathbb{N}$, and $\theta = n$, Kotz, Kozubowski and Podgórski (2001) obtained the density function of this sum of n standard classical Laplace variates. However, in the general case the density function does not have a closed form expression.

Since the family consists of symmetric leptokurtic distributions, this suggests that data exhibiting the properties of being symmetric around the origin and of having excess kurtosis can be fitted to this family. In the following section, we develop a two-step method for fitting data to the DGD family. The first step comprises model validation and the second, parameter estimation. In Section 3, goodness-of-fit tests for the simple and composite hypotheses are presented. The test statistics are shown to follow a chi-square distribution asymptotically. In addition, we explain how the parameter λ can be employed to test for distributional assumptions. More precisely, a test of normality of the data is presented. In Section 4, we provide simulation results for the methods developed.

2 FITTING TO THE DGD FAMILY

2.1 Introduction

We suggest a two-step method for fitting data to the DGD family. The first step consists of assessing the compatibility between the data and the family. Since it is a family of leptokurtic symmetric distributions, the data have to exhibit those characteristics. We verify them by performing a test of symmetry and by only considering data with a kurtosis greater than 3. Once we have confirmed that the family is well-suited for the data, we proceed with parameter estimation which is the second step. Parameter estimation is achieved through a minimum-distance method based on the characteristic function. We choose the parameters which minimize the distance between the real parts of the theoretical characteristic function and the empirical characteristic function. The estimators obtained are consistent, robust and asymptotically normal.

2.2 Testing Symmetry

2.2.1 Introduction

Let $x_1, \ldots, x_n$ be n independent observations from a continuous random variable X with distribution function F, density f and known center μ_0. We consider the problem of testing

$$\begin{aligned} &H_0 : F(\mu_0 - x) = 1 - F(\mu_0 + x) \quad \text{against} \\ &H_a : F(\mu_0 - x) \neq 1 - F(\mu_0 + x). \end{aligned}$$

Thus, we are interested in testing whether the density f is symmetric about the known median μ_0, or skewed.

Many tests of symmetry have been described in the literature (see Lehman, 1975; Randles and Wolfe, 1979). McWilliams (1990) and Moddares and Gastwirth (1996) used tests based on a runs statistic. Tajuddin (1994) and Thas, Rayner and Best (2005) used tests based on the Wilcoxon signed rank statistic. Also, Cheng and Balakrishnan (2004) proposed a modified sign test for symmetry.

We suggest using the hybrid test proposed by Moddares and Gastwirth (1998) to test the hypothesis of symmetry around a known median. We favor this test due to its high power and simplicity. Thas, Rayner and Best (2005) performed extensive simulations comparing the power of different tests of symmetry, which revealed that the hybrid test is more powerful than most alternatives considered.

2.2.2 Hybrid Test

The hybrid test is defined in two stages. Stage I consists of the sign test at level $\alpha_1 < \alpha$. If H_0 is accepted in stage I, then the percentile-modified two-sample Wilcoxon test is performed in stage II at level $\alpha_2 < \alpha$. The hybrid procedure is an α-level test, where $\alpha = \alpha_1 + (1 - \alpha_1)\alpha_2$. Suggested values for the levels of the tests are $\alpha_1 = 0.01$ and $\alpha_2 = 0.0404$ yielding an overall level of $\alpha = 0.05$. Please refer to Moddares and Gastwirth (1998) for more details on the hybrid procedure.

The first step of our method involves validating the compatibility between the data and the DGD family. It consists of two elements: the kurtosis of the data has to be greater than 3 and the hybrid test must not reject the hypothesis of symmetry around μ_0. If the data qualify, then we can carry on with the second step, parameter estimation. For the DGD family, μ_0 is conveniently set to 0. However, in the particular case where μ_0 is known and $\mu_0 \neq 0$, then μ_0 must be subtracted from the data and the shifted data can be fitted. If μ_0 is unknown, our model must be extended by adding a third parameter for location. This will be discussed in the conclusion.

2.3 Parameter Estimation

2.3.1 Introduction

We will estimate the parameters through a minimum-distance method based on the characteristic function. There is an extensive literature involving the characteristic function in parameter estimation. For example, it is a widely used method with stable distributions. References include Paulson, Holcomb and Leitch (1975), Feuerverger and McDunnough (1981a), Csörgő (1987), Gürtler and Henze (2000) and Matsui and Takemura (2005a, 2005b). Moreover, Yu (2004) shows how techniques relying on the characteristic function are used in mixtures of normal distributions, in the variance gamma distribution, in stable ARMA processes, and in a diffusion model.

Traditionally, the maximum likelihood approach is widely favored due to its generality and asymptotic efficiency (see Barreto-Souza, Santos and Cordeiro (2010), Tzavelas (2009) or Da Silva, Ferrari and Cribari-Neto (2008) for examples of application of the method to the beta generalized exponential distribution, the three-parameter gamma distribution and the Weibull regression). However, the likelihood function is not always tractable, as is the case with stable laws. When this occurs, the characteristic function might be used. Since the empirical characteristic function retains all the information in the sample, this suggests that estimation and inference via the empirical characteristic function should work as efficiently as the likelihood-based approaches. Feuerverger and McDunnough (1981a) showed that the asymptotic variance-covariance matrix of the parameters estimated using a minimum-distance method based on the characteristic function can be made arbitrarily close to the Cramér-Rao bound so that the method can attain arbitrarily high asymptotic efficiency. Moreover, the estimators obtained are consistent, robust and asymptotically normally distributed. Feuerverger and McDunnough (1981b) noted that the robustness properties for procedures associated with the empirical characteristic function are the result of a bounded influence curve for the estimators. For more details on the influence curve, see Hampel (1974).

2.3.2 The Empirical Characteristic Function

Consider $Z_1, \ldots, Z_n$ to be independent and identically distributed observations from the $DGD(\lambda, \theta)$. Let us define the empirical and theoretical characteristic functions at a specific point t_0 as $\phi_n(t_0)$ and $\phi(t_0)$ respectively, where

$$\phi_n(t_0) = \frac{1}{n}\sum_{j=1}^{n} e^{it_0 Z_j} = \frac{1}{n}\sum_{j=1}^{n}[\cos(t_0 Z_j) + i\sin(t_0 Z_j)]$$

and

$$\phi(t_0) = \left(\frac{1}{1 + t_0^2\lambda\theta}\right)^{1/\lambda}.$$

Thus, $\phi(t_0)$ only has a real part and let us denote the real part of $\phi_n(t_0)$ as $\phi_n^{Re}(t_0)$, where

$$\phi_n^{Re}(t_0) = \frac{1}{n}\sum_{j=1}^{n}\cos(t_0 Z_j). \tag{2.1}$$

For any fixed t_0, $\phi_n(t_0)$ is an average of bounded independent and identically distributed random variables having mean $\phi(t_0)$ and finite variance. Therefore, it follows by the strong law of large numbers that $\phi_n(t_0)$ converges almost surely to $\phi(t_0)$. Furthermore, Feuerverger and Mureika (1977) proved, for fixed $T < \infty$, the convergence of

$$\sup_{|t| \le T} |\phi_n(t) - \phi(t)| \to 0$$

almost surely as $n \to \infty$ and assert that $\phi_n^{Re}(t)$ will become uniformly close to $\phi(t)$ when the underlying distribution is symmetric. This implies that the imaginary part of $\phi_n(t)$, denoted by $\phi_n^{Im}(t)$, is approximately 0 for large n. Thus, any discrepancies observed between $\phi_n^{Im}(t)$ and 0 will be due to sampling error and consequently $\phi_n^{Im}(t)$ will not hold any information about the parameters λ and θ. Hence, since we are only fitting data that are symmetric around the origin, we will only consider the real parts of $\phi_n(t)$ and $\phi(t)$ to estimate the parameters as the imaginary parts will be uninformative.

2.3.3 Quadratic Distance

The method used is a form of nonlinear weighted least squares estimation. It is similar to the $k - L$ procedure introduced by Feuerverger and McDunnough (1981a) and it is a special distance within the class of quadratic distances introduced by Luong and Thompson (1987), where a unified theory for estimation and goodness-of-fit is developed. More precisely, the technique consists in choosing the parameters which minimize the quadratic distance between the real parts of the theoretical characteristic function and the empirical characteristic function.

Let us choose the points $t_1, \ldots, t_k > 0$ and let us define the vectors

$$\mathbf{Z}_n = [\phi_n^{Re}(t_1), \ldots, \phi_n^{Re}(t_k)]'$$
$$\mathbf{Z}(\lambda, \theta) = [\phi(t_1), \ldots, \phi(t_k)]'.$$

The quadratic distance estimator (QDE) based on the characteristic function, denoted by $(\hat{\lambda}, \hat{\theta})$, is defined as the value of (λ, θ) which minimizes the distance

$$d(\lambda, \theta) = [\mathbf{Z}_n - \mathbf{Z}(\lambda, \theta)]' \, \mathbf{Q}(\lambda, \theta) \, [\mathbf{Z}_n - \mathbf{Z}(\lambda, \theta)], \tag{2.2}$$

where $\mathbf{Q}(\lambda, \theta)$ is a positive definite matrix which may depend on (λ, θ). Luong and Thompson (1987) showed that an optimal choice of $\mathbf{Q}(\lambda, \theta)$ in the sense of minimizing the norm of the variance-covariance matrix of the estimated parameters is $\mathbf{Q}(\lambda, \theta) = \mathbf{\Sigma}^{-1}(\lambda, \theta)$, where $\mathbf{\Sigma}(\lambda, \theta)$ is the variance-covariance matrix of $\mathbf{Y}_n(\lambda, \theta) = \sqrt{n}[\mathbf{Z}_n - \mathbf{Z}(\lambda, \theta)]$. With this choice of matrix $\mathbf{Q}(\lambda, \theta)$, the QDE will be denoted by $(\hat{\lambda}^*, \hat{\theta}^*)$

Let $Y_n(t) = \sqrt{n}[\phi_n^{Re}(t) - \phi(t)]$, then $\mathbf{\Sigma}(\lambda, \theta) = (\sigma_{ij})$ is the $k \times k$ symmetric matrix with elements

$$\sigma_{ij} = \mathrm{Cov}[Y_n(t_i), Y_n(t_j)] = \frac{1}{2}[\phi(t_i + t_j) + \phi(t_i - t_j)] - \phi(t_i)\phi(t_j).$$

This result follows because $E[\cos(tZ)] = \phi(t)$ and

$$E[\cos(tZ)\cos(sZ)] = E[\frac{1}{2}(\cos((t+s)Z) + \cos((t-s)Z))] = \frac{1}{2}[\phi(t+s) + \phi(t-s)].$$

Since minimization of $d(\lambda, \theta)$ involves the inverse of the matrix $\mathbf{\Sigma}(\lambda, \theta)$ which depends on the parameters, a simpler procedure would be to replace $\mathbf{\Sigma}(\lambda, \theta)$ by a consistent estimate $\hat{\mathbf{\Sigma}}$ and minimize

$$d'(\lambda, \theta) = [\mathbf{Z}_n - \mathbf{Z}(\lambda, \theta)]' \, \hat{\mathbf{\Sigma}}^{-1} \, [\mathbf{Z}_n - \mathbf{Z}(\lambda, \theta)]. \tag{2.3}$$

Let (λ_0, θ_0) be the true value of (λ, θ) and $\mathbf{\Sigma}(\lambda_0, \theta_0) = \mathbf{\Sigma}$, then, if $\hat{\mathbf{\Sigma}} \xrightarrow{\mathcal{P}} \mathbf{\Sigma}$ (i.e. $\hat{\mathbf{\Sigma}}$ is a consistent estimate of $\mathbf{\Sigma}$), Luong and Doray (2002, 2009) assert that minimization of (2.2) with $\mathbf{Q}(\lambda, \theta) = \mathbf{\Sigma}^{-1}(\lambda, \theta)$, and (2.3) yields asymptotically equivalent estimators. For example, $\mathbf{\Sigma}_n^{Re}$ defined analogously to $\mathbf{\Sigma}$ in terms of $\phi_n^{Re}(t)$ is a consistent estimate of $\mathbf{\Sigma}$. More precisely, $\mathbf{\Sigma}_n^{Re} = (a_{ij})$ is the $k \times k$ matrix with elements

$$a_{ij} = \frac{1}{2}[\phi_n^{Re}(t_i + t_j) + \phi_n^{Re}(t_i - t_j)] - \phi_n^{Re}(t_i)\phi_n^{Re}(t_j).$$

Luong and Doray (2002) suggested an iterative procedure to estimate $(\hat{\lambda}^*, \hat{\theta}^*)$. First, we obtain $(\tilde{\lambda}, \tilde{\theta})$ by choosing $\mathbf{Q}(\lambda, \theta) = \mathbf{I}$, the identity matrix. Despite the fact that $(\tilde{\lambda}, \tilde{\theta})$ is less efficient, it can be used to estimate $\mathbf{\Sigma}$, by letting $\hat{\mathbf{\Sigma}} = \mathbf{\Sigma}(\tilde{\lambda}, \tilde{\theta})$. We then can use $\hat{\mathbf{\Sigma}}$ to obtain the first iteration for $(\hat{\lambda}^*, \hat{\theta}^*)$ and this procedure can be repeated with $\mathbf{\Sigma}$ reestimated at each step; $(\hat{\lambda}^*, \hat{\theta}^*)$ is defined as the convergent vector value of the procedure.

2.3.4 Asymptotic Properties of the Quadratic Distance Estimator

From (2.1), we observe that $\phi_n^{Re}(t)$ is an average of bounded processes and it follows, by means of the multivariate Central Limit Theorem, that $\mathbf{Y}_n(\lambda_0, \theta_0) = \mathbf{Y}_n$ converges in law to a multivariate normal distribution with zero mean and covariance structure $\mathbf{\Sigma}$. Thus, we have

$$\mathbf{Y}_n = \sqrt{n}\,[\mathbf{Z}_n - \mathbf{Z}(\lambda_0, \theta_0)] \xrightarrow{\mathcal{D}} N(0, \mathbf{\Sigma}). \tag{2.4}$$

Let $(\hat{\lambda}^*, \hat{\theta}^*)$ be the estimator obtained by minimizing (2.2) with $\mathbf{Q}(\lambda, \theta) = \mathbf{\Sigma}^{-1}(\lambda, \theta)$. Under the conditions that $d(\lambda, \theta)$ attains its minimum at an interior point of $\mathbf{\Theta} = \{\lambda, \theta \in \mathbb{R}; \lambda \geq 0, \theta > 0\}$ and that $\mathbf{Z}(\lambda, \theta)$ and $\mathbf{Q}(\lambda, \theta)$ are differentiable, the estimator $(\hat{\lambda}^*, \hat{\theta}^*)$ may also be defined implicitly as a root of the 2-dimensional system of estimating equations

$$\frac{\partial}{\partial(\lambda, \theta)} \left\{[\mathbf{Z}_n - \mathbf{Z}(\lambda, \theta)]'\, \mathbf{\Sigma}^{-1}(\lambda, \theta)\, [\mathbf{Z}_n - \mathbf{Z}(\lambda, \theta)]\right\} = 0.$$

Using lemmas (2.4.2) and (3.4.1) in Luong and Thompson (1987), we can conclude that

(i) $(\hat{\lambda}^*, \hat{\theta}^*) \xrightarrow{\mathcal{P}} (\lambda_0, \theta_0)$, i.e. $(\hat{\lambda}^*, \hat{\theta}^*)$ is a consistent estimator of (λ_0, θ_0),

(ii) $(\hat{\lambda}^*, \hat{\theta}^*)$ satisfies $\dfrac{\partial \mathbf{Z}'(\hat{\lambda}^*, \hat{\theta}^*)}{\partial(\lambda, \theta)} \left\{\mathbf{\Sigma}^{-1}(\hat{\lambda}^*, \hat{\theta}^*)\mathbf{Y}_n(\hat{\lambda}^*, \hat{\theta}^*)\right\} + o_p(1) = 0$,

(iii) $\sqrt{n}[(\hat{\lambda}^*, \hat{\theta}^*) - (\lambda_0, \theta_0)] = (\mathbf{S}'\mathbf{\Sigma}^{-1}\mathbf{S})^{-1}\mathbf{S}'\mathbf{\Sigma}^{-1}\mathbf{Y}_n + o_p(1)$,

(iv) $\mathbf{Y}_n(\hat{\lambda}^*, \hat{\theta}^*) = \mathbf{Y}_n - \{\mathbf{S} + o_p(1)\}\sqrt{n}[(\hat{\lambda}^*, \hat{\theta}^*) - (\lambda_0, \theta_0)]$,

(v) $\sqrt{n}[(\hat{\lambda}^*, \hat{\theta}^*) - (\lambda_0, \theta_0)] \xrightarrow{\mathcal{D}} N(0, (\mathbf{S}'\mathbf{\Sigma}^{-1}\mathbf{S})^{-1})$.

The symbol $o_p(1)$ denotes an expression converging to 0 in probability (i.e. $o_p(1) \xrightarrow{\mathcal{P}} 0$), $\mathbf{S}$ is a matrix of dimension $k \times 2$ defined as

$$\mathbf{S} = \begin{pmatrix} \frac{\partial Z_1(\lambda,\theta)}{\partial\lambda} & \frac{\partial Z_1(\lambda,\theta)}{\partial\theta} \\ \vdots & \vdots \\ \frac{\partial Z_k(\lambda,\theta)}{\partial\lambda} & \frac{\partial Z_k\lambda,\theta)}{\partial\theta} \end{pmatrix} = \begin{pmatrix} \frac{\partial\phi(t_1)}{\partial\lambda} & \frac{\partial\phi(t_1)}{\partial\theta} \\ \vdots & \vdots \\ \frac{\partial\phi(t_k)}{\partial\lambda} & \frac{\partial\phi(t_k)}{\partial\theta} \end{pmatrix},$$

where

$$\frac{\partial\phi(t)}{\partial\lambda} = \frac{(1+\lambda\theta t^2)\ln(1+\lambda\theta t^2) - \lambda\theta t^2}{\lambda^2\left(1+\lambda\theta t^2\right)^{1+1/\lambda}} \quad \text{and}$$
$$\frac{\partial\phi(t)}{\partial\theta} = -\frac{t^2}{\left(1+\lambda\theta t^2\right)^{1+1/\lambda}} \quad ,$$

all quantities being evaluated at (λ_0, θ_0). Thus, the estimator $(\hat{\lambda}^*, \hat{\theta}^*)$ is consistent and asymptotically normally distributed with variance-covariance matrix $(\mathbf{S}'\mathbf{\Sigma}^{-1}\mathbf{S})^{-1}$. The same results hold for the estimator obtained by minimizing (2.3).

The choice of points $t_1, \ldots, t_k$ affects $(\mathbf{S}'\mathbf{\Sigma}^{-1}\mathbf{S})^{-1}$ and thus we must choose them with care. Feuerverger and McDunnough (1981a) showed that by using a sufficiently extensive grid $\{t_i\}_{i=1}^k$, $(\mathbf{S}'\mathbf{\Sigma}^{-1}\mathbf{S})^{-1}$ can be made arbitrarily close to the Cramér-Rao bound. However, by choosing more points, the $k \times k$ matrix $\mathbf{\Sigma}$ can become near singular and computational problems may arise. For our simulation study, we will consider sets of points having the general form

$$\{t_i\}_{i=1}^k = \left\{\frac{Mi}{k}\right\}_{i=1}^k = \left\{\frac{M}{k}, \frac{2M}{k}, \ldots, M\right\}, \tag{2.5}$$

where M is an arbitrary number. More precisely, we will use values of $M = 0.01, 0.1, 1, 2$ and 3 and examine the effect on our estimation when $k = 5, 10, 20$ or 30. We will determine the choices of points for which the variances of the estimated parameters are a minimum.

3 HYPOTHESIS TESTING

3.1 Goodness-of-Fit

3.1.1 Introduction

Since we built statistics based on a minimum distance between empirical and theoretical parts, it is natural to use them for testing goodness-of-fit. Luong and Thompson (1987) developed a unified theory for estimation and goodness-of-fit when quadratic distances are employed. They showed that test statistics for goodness-of-fit follow a chi-square distribution asymptotically. Their results generalize the tests based on the characteristic function proposed by Koutrouvelis (1980) and Koutrouvelis and Kellermeir (1981). We now present the test statistics for the simple and composite hypotheses respectively. The following theorem appearing in Luong and Doray (2002) is needed; its proof can be found in Rao (1973).

Theorem 1. *Suppose that the random vector* $\mathbf{Y}_n$ *of dimension* k *is* $N(0, \mathbf{\Sigma})$ *and* $\mathbf{Q}$ *is any* $k \times k$ *symmetric positive semi-definite matrix; then the quadratic form* $\mathbf{Y}_n' \mathbf{Q} \mathbf{Y}_n$ *is chi-square distributed with* ν *degrees of freedom if* $\mathbf{\Sigma Q}$ *is idempotent and* $\text{trace}(\mathbf{\Sigma Q}) = \nu$. *(The same result holds asymptotically if* $\mathbf{Q}$ *is replaced by a consistent estimate* $\hat{\mathbf{Q}}$ *and* $\mathbf{Y}_n \xrightarrow{\mathcal{D}} N(0, \mathbf{\Sigma})$*).*

3.1.2 Simple Hypothesis

To test the simple hypothesis $H_0 : Z_1, \ldots, Z_n$ come from a specified DGD distribution with parameters (λ_0, θ_0), the following test statistic can be used,

$$\begin{aligned} nd(\lambda_0, \theta_0) &= n[\mathbf{Z}_n - \mathbf{Z}(\lambda_0, \theta_0)]' \, \mathbf{\Sigma}_0^{-1} \, [\mathbf{Z}_n - \mathbf{Z}(\lambda_0, \theta_0)] \\ &= \mathbf{Y}_n' \, \mathbf{\Sigma}_0^{-1} \, \mathbf{Y}_n, \end{aligned}$$

where $\mathbf{\Sigma}_0$ equals $\mathbf{\Sigma}$ evaluated at (λ_0, θ_0). It follows from (2.4) and Theorem 1, that $nd(\lambda_0, \theta_0) \xrightarrow{\mathcal{D}} \chi^2_\nu$, where

$$\nu = \text{trace}(\mathbf{\Sigma \Sigma}^{-1}) = \text{trace}(\mathbf{I}_k) = k,$$

and $\mathbf{I}_k$ is the $k \times k$ identity matrix. Thus, the test statistic follows a limiting chi-square distribution with $\nu = k$ degrees of freedom under H_0. To test the hypothesis H_0 at significance level α, compute the value of the test statistic $nd(\lambda_0, \theta_0)$ from the sample. The null hypothesis H_0 should be rejected if $nd(\lambda_0, \theta_0) > \chi^2_{k,1-\alpha}$, where $\chi^2_{k,1-\alpha}$ is the $100(1-\alpha)^{\text{th}}$ quantile of a χ^2 distribution with k degrees of freedom.

3.1.3 Composite Hypothesis

To test the composite hypothesis $H_0 : Z_1, \ldots, Z_n$ come from a DGD distribution where the values of the parameters are not specified, we first calculate the quadratic distance estimator $(\hat{\lambda}^*, \hat{\theta}^*)$ by minimizing (2.2) with $\mathbf{Q}(\lambda, \theta) = \mathbf{\Sigma}^{-1}(\lambda, \theta)$. Luong and Thompson (1987) showed that the test statistic

$$\begin{aligned} nd(\hat{\lambda}^*, \hat{\theta}^*) &= n[\mathbf{Z}_n - \mathbf{Z}(\hat{\lambda}^*, \hat{\theta}^*)]' \, \mathbf{\Sigma}^{-1}(\hat{\lambda}^*, \hat{\theta}^*) \, [\mathbf{Z}_n - \mathbf{Z}(\hat{\lambda}^*, \hat{\theta}^*)] \\ &= \mathbf{Y}_n'(\hat{\lambda}^*, \hat{\theta}^*) \, \mathbf{\Sigma}^{-1}(\hat{\lambda}^*, \hat{\theta}^*) \, \mathbf{Y}_n(\hat{\lambda}^*, \hat{\theta}^*) \end{aligned}$$

follows an asymptotic chi-square distribution with $\nu = k-2$ degrees of freedom under H_0. Again, $\mathbf{\Sigma}(\hat{\lambda}^*, \hat{\theta}^*)$ can be replaced by a consistent estimate $\hat{\mathbf{\Sigma}}$. Analogous to the case for the simple null hypothesis, a significance level α test can be performed to test H_0.

3.2 Test of Normality

In Section 2.3.4, we showed that the estimator $(\hat{\lambda}, \hat{\theta})$ is asymptotically normally distributed with variance-covariance matrix $(\mathbf{S}'\mathbf{\Sigma}^{-1}\mathbf{S})^{-1}$. Thus, we can easily construct individual and joint $(1-\alpha)\%$ confidence intervals for the parameters λ and θ.

Of more practical interest is testing for the parameter λ. In Section 1, we saw that particular values of λ define specific distributions within the DGD distribution family. For example, when $\lambda = 0$ or $\lambda = 1$, we obtain the normal and the Laplace distributions respectively. This suggests using the parameter λ to test distributional assumptions. A test of normality of the data can be constructed by testing

$$H_0 : \lambda = 0 \quad \text{versus} \quad H_a : \lambda > 0. \tag{3.1}$$

In Section 1, we noted that the kurtosis of a $DGD(\lambda, \theta)$ random variable Z is $(3+3\lambda)$. Thus, if we have a sample from Z, $\hat{\beta}_2 = (3+3\hat{\lambda})$ is a consistent estimate of the population kurtosis β_2. Moreover, since β_2 is a linear function of λ, the hypotheses identified in (3.1) are equivalent to

$$H_0 : \beta_2 = 3 \quad \text{versus} \quad H_a : \beta_2 > 3.$$

This implies that in (3.1) we are testing the normal distribution against symmetric distributions with heavier tails. Thus, it would be interesting to compare the power of this test to a normality test based on the sample kurtosis. D'Agostino and Pearson (1973) describe such a test. Moreover, when the alternative is the Laplace distribution, the power of the test can be compared to the likelihood ratio test. Kotz, Kozubowski and Podgórski (2001) assert that the likelihood ratio test is the most powerful scale invariant test for testing the normal against the Laplace when the center of symmetry is known. In Section 4, we provide a simulation study for estimating parameters and testing hypotheses with the methods presented previously.

4 SIMULATION STUDY

4.1 Parameter Estimation

While the expressions for quadratic distance estimators may seem complex, they are relatively simple to implement using a computer software with built-in statistical functions. The quadratic distance estimator can be computed numerically using a nonlinear least squares method. All of our simulations were completed using Maple 11.0.

We first generated 100 random samples from a $DGD(\lambda = 1, \theta = 1)$ random variable of sizes 100, 500 and 1000. For each sample, we estimated the parameters using the method of moments (MOM), ordinary least squares (OLS) (i.e. using (2.2) with $\mathbf{Q}(\lambda, \theta) = \mathbf{I}$, the identity matrix) and weighted least squares (WLS) (i.e. using (2.2) with an appropriate choice of $\mathbf{Q}(\lambda, \theta)$). OLS and WLS methods were implemented using 20 different sets of points, $\{t_i\}_{i=1}^{k}$, in order to determine which are the best choices. All the sets have the general form defined by (2.5). Values of $M = 0.01, 0.1, 1, 2$ and 3 and values of $k = 5, 10, 20$ and 30 were used to define $\{t_i\}_{i=1}^{k}$.

Tables 1, 2, and 3 summarize the pertinent results for sample sizes of 100, 500 and 1000 respectively. Each table provides the mean and the standard error based on 100 random samples of the estimated values of λ and θ using the MOM, OLS and WLS. The WLS estimates were obtained using the iterative procedure to estimate $\mathbf{\Sigma}$ presented in Section 2.3.3. Results for values of $M =$ 0.01 and 0.1 are not presented as the WLS method rarely found an improved estimate over the OLS method. Consequently, we do not recommend using those choices of M. The other values of M all yielded good estimates but we suggest using $M = 3$, as the standard errors of the estimates were generally the lowest for this choice. Moreover, increasing the value of k (i.e. increasing the number of points in the sets) generally improved estimates. However, when using $k = 30$, the estimation process was slow and the improvement over $k = 10$ or 20 is not substantial and not worthy of the additional computation time. Thus, we suggest using values of $k = 10$ or 20 for a fast and efficient estimation. Moreover, the asymptotic standard deviations of the estimators that can be calculated from the results of Section 2.3.4 are very close to the standard errors that were observed in our samples.

We also performed WLS estimation with choices of $\mathbf{Q}(\lambda, \theta) = (\mathbf{\Sigma}_n^{Re})^{-1}$ and $\mathbf{\Sigma}_0^{-1}$. With $(\mathbf{\Sigma}_n^{Re})^{-1}$, we obtained poor estimates and often they did not converge to a solution. The choice of $\mathbf{\Sigma}_0^{-1}$ produced comparable estimates to the ones obtained in the tables under the WLS columns. However, the choice of $\mathbf{\Sigma}_0^{-1}$ is not a viable selection in practice as the true parameters (λ_0, θ_0) are unknown.

4.2 Goodness-of-Fit Testing

We now perform goodness-of-fit testing for the simple hypothesis as presented in Section 3.1.2. First, we wish to determine if the test has a correct size when using critical values from the chi-square distribution for sample sizes of $n = 100, 500$ and 1000. For each sample size n, we generated 5000

Table 1: Estimates based on 100 random samples of size 100

		λ		θ	
True values		1		1	
MOM (s.e.)		0.6853 (0.4099)		1.0287 (0.2184)	
k	M	OLS (s.e.)	WLS (s.e.)	OLS (s.e.)	WLS (s.e.)
5	1	1.1111 (0.5926)	1.0877 (0.5026)	1.0714 (0.2388)	1.0465 (0.2221)
5	2	1.0642 (0.4332)	1.0434 (0.3926)	1.0622 (0.2592)	1.0477 (0.2232)
5	3	1.0371 (0.4332)	1.0684 (0.3477)	1.0563 (0.2928)	1.0697 (0.2358)
10	1	1.1038 (0.6062)	1.0642 (0.4280)	1.0698 (0.2388)	1.0446 (0.2204)
10	2	1.0688 (0.4333)	1.0314 (0.3707)	1.0647 (0.2610)	1.0415 (0.2157)
10	3	1.0342 (0.4362)	1.0270 (0.3414)	1.0556 (0.2913)	1.0419 (0.2185)
20	1	1.0990 (0.6142)	1.0298 (0.3974)	1.0688 (0.2389)	1.0418 (0.2203)
20	2	1.0705 (0.4323)	1.0264 (0.3602)	1.0653 (0.2606)	1.0410 (0.2159)
20	3	1.0340 (0.4363)	1.0297 (0.3387)	1.0550 (0.2882)	1.0421 (0.2180)

Table 2: Estimates based on 100 random samples of size 500

		λ		θ	
True values		1		1	
MOM (s.e.)		0.8841 (0.2745)		1.0112 (0.0991)	
k	M	OLS (s.e.)	WLS (s.e.)	OLS (s.e.)	WLS (s.e.)
5	1	1.0441 (0.2449)	1.0161 (0.2194)	1.0232 (0.1064)	1.0151 (0.1007)
5	2	1.0083 (0.2329)	1.0221 (0.2024)	1.0153 (0.1185)	1.0184 (0.1003)
5	3	1.0294 (0.2095)	1.0314 (0.1981)	1.0263 (0.1260)	1.0244 (0.1108)
10	1	1.0439 (0.2470)	1.0030 (0.2051)	1.0231 (0.1058)	1.0140 (0.1003)
10	2	1.0072 (0.2337)	1.0124 (0.1923)	1.0151 (0.1188)	1.0150 (0.0988)
10	3	1.0255 (0.2132)	1.0198 (0.1841)	1.0244 (0.1268)	1.0163 (0.1006)
20	1	1.0437 (0.2485)	1.0025 (0.1997)	1.0230 (0.1054)	1.0139 (0.0994)
20	2	1.0065 (0.2337)	1.0117 (0.1877)	1.0148 (0.1187)	1.0149 (0.0990)
20	3	1.0240 (0.2145)	1.0188 (0.1780)	1.0236 (0.1259)	1.0160 (0.1004)

Table 3: Estimates based on 100 random samples of size 1000

		λ		θ	
True values		1		1	
MOM (s.e.)		0.9428 (0.2606)		1.0106 (0.0771)	
k	M	OLS (s.e.)	WLS (s.e.)	OLS (s.e.)	WLS (s.e.)
5	1	1.0197 (0.1816)	1.0058 (0.1662)	1.0143 (0.0791)	1.0116 (0.0769)
5	2	1.0180 (0.1664)	1.0173 (0.1395)	1.0162 (0.0979)	1.0144 (0.0779)
5	3	1.0017 (0.1635)	1.0053 (0.1379)	1.0088 (0.0974)	1.0121 (0.0801)
10	1	1.0205 (0.1835)	1.0094 (0.1544)	1.0144 (0.0788)	1.0119 (0.0773)
10	2	1.0190 (0.1655)	1.0087 (0.1367)	1.0166 (0.0966)	1.0118 (0.0771)
10	3	1.0037 (0.1680)	1.0032 (0.1297)	1.0099 (0.1016)	1.0111 (0.0763)
20	1	1.0209 (0.1848)	1.0176 (0.1458)	1.0145 (0.0786)	1.0128 (0.0773)
20	2	1.0192 (0.1652)	1.0106 (0.1335)	1.0166 (0.0958)	1.0120 (0.0768)
20	3	1.0051 (0.1686)	1.0021 (0.1284)	1.0107 (0.1019)	1.0108 (0.0762)

samples from a $DGD(\lambda = 1, \theta = 1)$ random variable and calculated the test statistics $nd(\lambda_0 = 1, \theta_0 = 1)$. We repeated the procedure for samples from a $DGD(\lambda = 2, \theta = 1)$. We were thus able to obtain simulated critical values for a level α test by taking the $100(1-\alpha)^{\text{th}}$ quantiles from the empirical distributions of the test statistics. All of the test statistics were obtained using the set of points defined by (2.5) with values of $M = 3$ and $k = 10$. We present our results in Tables 4 and 5.

Table 4: Actual sizes of the test using $\chi^2_{10,1-\alpha}$ with 5000 simulation runs

		Actual sizes of the test		
α	(λ_0, θ_0)	$n = 100$	$n = 500$	$n = 1000$
0.100	(1,1)	0.1250	0.1242	0.1168
	(2,1)	0.1420	0.1350	0.1194
0.050	(1,1)	0.0966	0.0834	0.0764
	(2,1)	0.1138	0.0898	0.0738
0.025	(1,1)	0.0798	0.0594	0.0544
	(2,1)	0.0930	0.0628	0.0448
0.010	(1,1)	0.0658	0.0406	0.0310
	(2,1)	0.0756	0.0412	0.0264

Table 5: Critical values obtained for various sample sizes

		Critical values			
α	(λ_0, θ_0)	$n = 100$	$n = 500$	$n = 1000$	$\chi^2_{10,1-\alpha}$
0.100	(1,1)	17.9453	17.1660	16.8809	15.9872
	(2,1)	19.6814	17.5147	16.9050	15.9872
0.050	(1,1)	27.2592	21.5835	20.9238	18.3070
	(2,1)	27.6104	21.9962	20.0182	18.3070
0.025	(1,1)	37.0803	27.4926	24.4911	20.4832
	(2,1)	37.6163	25.9949	23.5219	20.4832
0.010	(1,1)	68.5516	39.5960	29.3234	23.2093
	(2,1)	58.1972	32.5478	28.3944	23.2093

The results of Table 4 indicate that the goodness-of-fit test has an incorrect size that is severe enough to warrant a recommendation that the test should not be used without appropriately sized critical values. From Table 5, we remark that the test statistic $nd(\lambda_0, \theta_0)$ converges very slowly to a chi-square random variable. Even for sample sizes of 1000, the approximation is not satisfactory. The real distribution of the test statistic will generally have a heavier right tail than the chi-square distribution, even for large sample sizes, and thus the test will always be oversized when using the critical value $\chi^2_{k,1-\alpha}$.

We will now assess the power of the goodness-of-fit test for the simple hypothesis $H_0 : (\lambda_0 = 1, \theta_0 = 1)$ against alternatives $H_a : (\lambda_a, \theta_a = 1)$, where $\lambda_a = 0, 0.5, 1, 1.5$ and 2. A level $\alpha = 0.05$ and sample sizes of $100, 500$ and 1000 were employed. We determined the power of the test by generating 5000 samples for each of the alternatives considered. Appropriately sized critical values (CV) were calculated by taking the average of the two critical values obtained in Table 5 for each sample size. Results are shown in Table 6.

Table 6: Power of the test ($\alpha = 0.05$) with 5000 simulation runs

		Alternatives (λ_a, θ_a)				
n	CV	$(0,1)$	$(0.5,1)$	$(1,1)$	$(1.5,1)$	$(2,1)$
100	27.4348	0.0626	0.0164	0.0518	0.1070	0.1852
500	21.7899	0.9986	0.2692	0.0516	0.3656	0.9394
1000	20.4710	1.0000	0.8000	0.0572	0.7524	1.0000

The goodness-of-fit test performed poorly in rejecting the selected alternatives for a sample size of 100. When $n = 500$, the test did very well for alternatives of λ_a one unit away of $\lambda_0 = 1$ but not so well when λ_a was half a unit away. For a large sample size of 1000, the test was powerful for all alternatives considered. For sample sizes of 500 and 1000, the recorded powers for alternatives $(0, 1)$ and $(2, 1)$ were close to or equal to 100%. This suggests that for a large enough sample size, the test is well suited for discriminating between the fits of normal, Laplace and heavier tailed symmetric distributions. Moreover, by using adjusted critical values instead of $\chi^2_{k,1-\alpha}$, the tests had an adequate size. The discrepancies between the actual sizes and $\alpha = 0.05$ are due to the precision of the simulated critical values and to the large variability of the test statistic.

5 Conclusion

We have introduced the double gamma difference family, which is a family of leptokurtic symmetric distributions. The Laplace, the sums of Laplace and the normal distributions all arise as special cases of this family. While there is in general no closed form expression for the density function, the characteristic function is simple to work with. Parameters can be estimated through a minimum quadratic distance method based on the characteristic function. The estimators obtained were shown to be consistent, robust and asymptotically normally distributed. Goodness-of-fit tests for the simple and composite hypotheses were presented and the test statistics shown to follow a chi-square distribution asymptotically. Moreoever, we suggested employing the parameter λ to test for distributional assumptions. Simulations revealed that large sample sizes are required to get a reasonable amount of precision for estimating the parameters. Also, the goodness-of-fit tests must be carried out with appropriate simulated critical values for the tests to have a correct size because the convergence to the chi-square distribution is slow.

The family can be extended by adding a third parameter for location μ. The characteristic function $\phi^*(t)$ would then both have real and imaginary parts, where

$$\phi^*(t) = \begin{cases} e^{it\mu} \cdot \left(\frac{1}{1+t^2\lambda\theta}\right)^{1/\lambda} & \text{for } \lambda, \theta > 0, \\ e^{it\mu - t^2\theta} & \text{for } \lambda = 0, \theta > 0. \end{cases}$$

Parameter estimation could still be achieved through a minimum distance method based on the characteristic function. However, both real and imaginary parts would have to be taken into account. For more details on the minimum distance method when the real and imaginary parts are involved, see Feuerverger and McDunnough (1981b). Before fitting data to this family, it is still necessary to verify symmetry. For testing symmetry around the unknown median μ we suggest using the triples test introduced by Randles, Fligner, Policello and Wolfe (1980).

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the financial support of the Natural Sciences and Engineering Research Council of Canada.

References

[1] Barreto-Souza W., Santos A. and Cordeiro G. (2010). The beta generalized exponential distribution. *Journal of Statistical Computation and Simulation*, **80**, 159-172.

[2] Cheng W.H. and Balakrishnan N. (2004). A Modified Sign Test for Symmetry. *Comm. Statist. Simulation Comput.*, **33**, 703-709.

[3] Csörgő S. (1987). Testing for stability. In *Goodness-of-fit (Debrecen, 1984)*, volume 45 of *Colloq. Math. Soc. János Bolyai*. North-Holland, Amsterdam, pp. 101-132.

[4] D'Agostino R. and Pearson E. S. (1973). Tests for departures from normality. Empirical results for the distribution of b_2 and $\sqrt{b_1}$. *Biometrika*, **60**, 613-622.

[5] Da Silva M.F., Ferrari S. and Cribari-Neto F. (2008). Improved likelihood inference for the shape parameter in Weibull regression. *Journal of Statistical Computation and Simulation*, **78**, 789-811.

[6] Feuerverger A. and McDunnough P. (1981a). On the efficiency of empirical characteristic function procedures. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, **43**, 20-27.

[7] Feuerverger A. and McDunnough P. (1981b). On some Fourier methods for inference. *J. Amer. Statist. Assoc.*, **76**, 379-387.

[8] Feuerverger A. and Mureika R.A. (1977). The empirical characteristic function and its applications. *Ann. Statist.*, **1**, 88-97.

[9] Gürtler N. and Henze N. (2000). Goodness-of-fit tests for the Cauchy distribution based on the empirical characteristic function. *Ann. Inst. Statist. Math.*, **52**, 267-286.

[10] Hampel F.R. (1974). The influence curve and its role in robust estimation. *J. Amer. Statist. Assoc.*, **69**, 383-393.

[11] Kotz S. and Kozubowski T.J. and Podgórski K. (2001). *The Laplace distribution and generalizations.* Birkhäuser, Boston.

[12] Koutrouvelis I.A. (1980). A goodness-of-fit test of simple hypotheses based on the empirical characteristic function. *Biometrika*, **67**, 238-240.

[13] Koutrouvelis I.A. and Kellermeier J. (1981). A goodness-of-fit test based on the empirical characteristic function when parameters must be estimated. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, **43**, 173-176.

[14] Lehmann E. (1975). *Nonparametrics: Statistical methods based on ranks.* McGraw-Hill, New York.

[15] Lukacs E. (1970). *Characteristic Functions.* Hafner Publishing Co., New York. Second edition, revised and enlarged.

[16] Luong A. and Thompson M.E. (1987). Minimum-distance methods based on quadratic distances for transforms. *Canad. J. Statist.*, **15**, 239-251.

[17] LUONG A. AND DORAY L.G. (2002). General quadratic distance methods for discrete distributions definable recursively. *Insurance Math. Econom.*, **30**, 255-267.

[18] LUONG A. AND DORAY L.G. (2009). Inference for the positive stable laws based on a special quadratic distance. *Stat. Methodol.*, **6**, 147-156.

[19] MATSUI M. AND TAKEMURA A. (2005a). Empirical characteristic function approach to goodness-of-fit tests for the Cauchy distribution with parameters estimated by MLE or EISE. *Ann. Inst. Statist. Math.*, **57**, 183-199.

[20] MATSUI M. AND TAKEMURA A. (2005b). Goodness-of-fit tests for symmetric stable distributions - empirical characteristic function approach. CIRJE F-Series CIRJE-F-384, CIRJE, Faculty of Economics, University of Tokyo. available at `http://ideas.repec.org/p/tky/fseres/2005cf384.html`.

[21] MCWILLIAMS T. (1990). A distribution-free test for symmetry based on a runs statistic. *J. Amer. Statist. Assoc.*, **85**, 1130-1133.

[22] MODARRES R. AND GASTWIRTH J. (1996). A modified runs test for symmetry. *Statist. Probab. Lett.*, **31**, 107-112.

[23] MODARRES R. AND GASTWIRTH J. (1998). Hybrid test for the hypothesis of symmetry. *J. Appl. Stat.*, **25**, 777-783.

[24] PAULSON A.S. AND HALCOMB E.W. AND LEITCH R.A. (1975). The estimation of the parameters of the stable laws. *Biometrika*, **62**, 163-170.

[25] RANDLES R.H. AND WOLFE D.A. (1979). *Introduction to the Theory of Nonparametric Statistics*. Wiley, New York.

[26] RANDLES R. H., FLIGNER M. A., POLICELLO II G. E. AND WOLFE D. A. (1980). An asymptotically distribution-free test for symmetry versus asymmetry. *J. Amer. Statist. Assoc.*, **75**, 168-172.

[27] RAO C.R. (1973). *Linear Statistical Inference and its Applications*. Wiley, New York.

[28] TAJUDDIN I. (1994). Distribution-free test for symmetry based on Wilcoxon two-sample test. *J. Appl. Stat.*, **21**, 409-416.

[29] THAS O. AND RAYNER J.C.W. AND BEST D.J. (2005). Tests for symmetry based on the one-sample wilcoxon signed rank statistic. *Comm. Statist. Simulation Comput.*, **34**, 957-973.

[30] Tzavelas G (2009). Maximum likelihood parameter estimation in the three-parameter gamma distribution with the use of Mathematica. *Journal of Statistical Computation and Simulation*, **79**, 1457-1466.

[31] Yu J. (2004). Empirical characteristic function estimation and its applications. *Econometric Rev.*, **23**, 93-123.

Appendix A

Proof of Proposition 1:
From the generalized binomial theorem, we obtain the binomial series for $\phi(t)$, where

$$\phi(t) = (1 + t^2\lambda\theta)^{-1/\lambda} = \sum_{k=0}^{\infty} \binom{-1/\lambda}{k} (t^2\lambda\theta)^k$$

From the relationship between moments of a random variable and the derivatives of its characteristic function, we have

$$\mathrm{E}[Z^{2k}] = i^{2k} \left. \frac{d^{2k}}{dt^{2k}}\phi(t)\right|_{t=0} = (-1)^k \phi^{(2k)}(0).$$

$\phi^{(2k)}(0)$ corresponds to the $(k+1)^{\text{th}}$ term from the binomial series for $\phi(t)$ differentiated $2k$ times. Thus,

$$\mathrm{E}[Z^{2k}] = (-1)^k \binom{-1/\lambda}{k} (2k)!\lambda^k\theta^k.$$

Since the binomial coefficients admit the representation

$$\binom{-\frac{1}{\lambda}}{k} = \frac{1}{k!}\prod_{j=0}^{k-1}(-\frac{1}{\lambda} - j) = \frac{(-1)^k}{k!\lambda^k}\prod_{j=0}^{k-1}(1 + j\lambda),$$

we get the following expression:

$$\mathrm{E}[Z^{2k}] = (-1)^k(2k)!\lambda^k\theta^k \left[\frac{(-1)^k}{k!\lambda^k}\prod_{j=0}^{k-1}(1 + j\lambda)\right] = \frac{\theta^k(2k)!}{k!}\prod_{j=0}^{k-1}(1 + j\lambda).$$

We note here that the proof only applies for values of $\lambda > 0$. However, the same result holds for $\lambda = 0$. The expression for $\lambda = 0$ can be derived similarly by using the series expansion for $e^{-t^2\theta}$.