Supplementary materials for this article are available online. Please go to http://www.tandfonline.com/r/TECH

Robust Estimators of the Generalized Log-Gamma Distribution

Claudio AGOSTINELLI

Dipartimento di Scienze Ambientali
Informatica e Statistica
Università Cà Foscari
Venezia, Italy
(*claudio@unive.it*)

Alfio MARAZZI

Institute of social and preventive medicine
Lausanne University Hospital
Switzerland
(*alfio.marazzi@chuv.ch*)

Victor J. YOHAI

Departamento de Matematicas
Facultad de Ciencias Exactas y Naturales
University of Buenos Aires, Argentina
(*vyohai@dm.uba.ar*)

We propose robust estimators of the generalized log-gamma distribution and, more generally, of location-shape-scale families of distributions. A (weighted) Qτ estimator minimizes a τ scale of the differences between empirical and theoretical quantiles. It is $n^{1/2}$ consistent; unfortunately, it is not asymptotically normal and, therefore, inconvenient for inference. However, it is a convenient starting point for a one-step weighted likelihood estimator, where the weights are based on a disparity measure between the model density and a kernel density estimate. The one-step weighted likelihood estimator is asymptotically normal and fully efficient under the model. It is also highly robust under outlier contamination. Supplementary materials are available online.

KEY WORDS: Minimum quantile distance estimators; τ Estimators; Weighted likelihood estimators.

1. INTRODUCTION

Generalized log-gamma distributions are used to model highly skewed positive data on a logarithmic scale. This very flexible three-parameter family was introduced by Stacy (1962) and further studied by Prentice (1974) and Lawless (1980). This family includes the widely used log-exponential, log Weibull, and log-gamma distributions and even the normal distribution as special cases. Applications include modeling of reliability data (Meeker and Escobar 1998), speech processing data (Shin, Chang, and Kim 2005; Almpanidis and Kotropoulos 2008), chemical data (Barkauskas et al. 2009), drought data (Nadarajaha and Gupta 2007), and health expenditures (Manning, Basu, and Mullahy 2005). Applications in different branches of electrical engineering are described in Nadarajah (2008).

Usually, the parameters are estimated by means of the maximum likelihood (ML) principle, which provides fully efficient estimators when the observations follow the model. Unfortunately, the ML estimator is extremely sensitive to the presence of outliers in the sample and, therefore, it is not robust. There are several proposals of robust estimators for models involving a shape parameter (Field and Smith 1994; Cowell and Victoria-Feser 1996; Dupuis and Mills 1998; Marazzi and Ruffieux 1999; Dornheim and Brazauskas, 2007; Clarke, McKinnon, and Riley 2012; Ruckdeschel and Horbenko 2012). However, there are no specific proposals of robust estimators for three-parameter distribution families where the parameters are location, scale, and shape.

This article proposes two families of robust estimators: the (weighted) Qτ estimators and the one-step weighted likelihood (1SWL) estimators. A Qτ estimator minimizes a τ scale (Yohai and Zamar 1988) of the differences between empirical and theoretical quantiles. A one-step weighted likelihood (1SWL) estimator approximately solves a weighted likelihood equation, where the weights are based on a disparity measure between the model density and a kernel density estimate.

A (weighted) Qτ estimator is a version of quantile distance estimator (La Riccia 1982). A similar robust estimator based on regressing empirical and theoretical quantiles has been proposed by Boudt, Caliskan, and Croux (2011) for the two parameters of the gamma distribution. The Qτ estimator is highly robust (its breakdown point is 50%) and $n^{1/2}$ consistent. Unfortunately it is not asymptotically normal and it is, therefore, inconvenient for inference. It is, however, a convenient starting point of a 1SWL estimator, which maintains the high degree of robustness of the Qτ estimator, while improving its efficiency under the model. Moreover, the 1SWL estimator is asymptotically normal.

The estimation procedures proposed here for the generalized log-gamma family can be applied to other families of distributions with three parameters characterizing location, scale, and shape, such as the three-parameter log Weibull family. Other examples can be found in Lawless (2003, Section 1.3.6).

Section 2 introduces the generalized log-gamma family of distributions and the corresponding ML estimator. Section 3

© 2014 American Statistical Association and
the American Society for Quality
TECHNOMETRICS, FEBRUARY 2014, VOL. 56, NO. 1
DOI: 10.1080/00401706.2013.818578

describes the family of Qτ and weighted Qτ estimators and Section 4 the family of weighted likelihood estimators including the 1SWL estimators. Section 5 reports the results of a Monte Carlo study on the robustness and efficiency of the studied estimators. Section 6 describes an example with real data. Conclusions are provided in Section 7.

A separate document of supplemental materials (referred as supplementary materials in the following) provides the mathematical justifications of the proposed methods and additional results of the simulation experiments.

2. THE GENERALIZED LOG-GAMMA FAMILY OF DISTRIBUTIONS

The *generalized log-gamma* family of distributions depends on three parameters μ, σ, and λ. We consider the parameterization given by Prentice (1974) and denote the family by LG(μ, σ, λ), $\mu \in \mathbb{R}$, $\sigma > 0$, $\lambda \in \mathbb{R}$. We say that a random variable y has log-gamma distribution LG(μ, σ, λ) if

$$y = \mu + \sigma u \tag{1}$$

and u has density

$$f_\lambda(u) = \begin{cases} \dfrac{|\lambda|}{\Gamma(\lambda^{-2})}(\lambda^{-2})^{\lambda^{-2}} \exp((\lambda^{-2})(\lambda u - e^{\lambda u})) & \text{if } \lambda \neq 0, \\ \dfrac{1}{\sqrt{2\pi}} \exp\left(-\dfrac{u^2}{2}\right) & \text{if } \lambda = 0, \end{cases} \tag{2}$$

where Γ denotes the Gamma function. This family includes many common models, such as the normal model ($\lambda = 0$), the log Weibull model ($\lambda = 1$), the log-exponential model ($\lambda = 1$ and $\sigma = 1$), and the log-gamma model ($\sigma = \lambda$). Naturally, the family of distributions of $t = \exp(y)$ is called the *generalized gamma* family. The density of y is

$$f_{\boldsymbol{\theta}}(y) = \frac{1}{\sigma} f_\lambda\left(\frac{y-\mu}{\sigma}\right),$$

where $\boldsymbol{\theta} = (\mu, \sigma, \lambda)$. The score functions are

$$\begin{aligned} -\frac{d}{d\mu} \log f_{\boldsymbol{\theta}}(y) &= \frac{1}{\sigma}\xi_\lambda(u), \\ -\frac{d}{d\sigma} \log f_{\boldsymbol{\theta}}(y) &= \frac{1}{\sigma}(\xi_\lambda(u)u + 1), \\ -\frac{d}{d\lambda} \log f_{\boldsymbol{\theta}}(y) &= \psi_\lambda(u), \end{aligned}$$

where $u = (y - \mu)/\sigma$,

$$\begin{aligned} \xi_\lambda(u) &= \frac{f'_\lambda(u)}{f_\lambda(u)} = \frac{(1 - e^{\lambda u})}{\lambda}, \\ \psi_\lambda(u) &= \frac{-\partial \log f_\lambda(u)}{\partial \lambda} \\ &= \frac{1}{\lambda^3}(2\zeta(\lambda) - \lambda^2 + \lambda u - \exp(\lambda u)(2 - \lambda u)), \end{aligned}$$

$\zeta(\lambda) = -2\log(\lambda) - \dot{\Gamma}(\lambda^{-2}) + 1$ and $\dot{\Gamma}$ denotes the Digamma function, that is, the derivative of the log of the Gamma function. Then, the ML estimate of $\boldsymbol{\theta}$ is given by the following system of equations

$$\frac{1}{n}\sum_{j=1}^{n} \mathbf{z}(y_j, \boldsymbol{\theta}) = \mathbf{0}, \tag{3}$$

where $\mathbf{z}(y, \boldsymbol{\theta}) = (\xi_\lambda(u), \xi_\lambda(u)u + 1, \psi_\lambda(u))^{\mathrm{T}}$ is the score function vector.

If we are interested to fit a generalized gamma distribution to positive data, such as durations t, then an important parameter of interest is

$$\eta = E(t) = \delta\Gamma(\alpha + 1/\gamma)/\,\Gamma(\alpha),$$

where $\alpha = \lambda^{-2}, \gamma = \lambda/\sigma, \delta = \exp(\mu + 2\log(\lambda)\sigma/\lambda)$. Note that η depends on μ , σ, and λ, and therefore, no parameter will be treated as a nuisance parameter.

3. THE Qτ ESTIMATORS

The Qτ estimators proposed here are robust estimators of $\boldsymbol{\theta} = (\mu, \sigma, \lambda)$ based on the τ scale introduced by Yohai and Zamar (1988). In this section, we define some robust scales, the associated robust regression estimators, and the new robust estimators.

3.1 Scale Measures

We start this section defining a *scale function*. Let $\mathbf{u} = (u_1, \ldots, u_n)$ be a sample of size n. A function $s(\mathbf{u})$ is a scale if: (i) $s(\mathbf{u}) \geq 0$; (ii) for any scalar γ, $s(\gamma\mathbf{u}) = |\gamma|s(\mathbf{u})$; (iii) $s(u_1, \ldots, u_n) = s(|u_1|, \ldots, |u_n|)$; (iv) if $|u_i| \leq |v_i|$, $1 \leq i \leq n$, then $s(u_1, \ldots, u_n) \leq s(v_1, \ldots, v_n)$. It is easy to show that properties (i)–(iv) imply (v) $s(0, \ldots, 0) = 0$ and that, (vi) given $\varepsilon > 0$, there exists δ such that $|u_i| \leq \delta$ for $1 \leq i \leq n$ imply $s(u_1, .., u_n) < \delta$. Properties (i)–(vi) clearly show that $s(\mathbf{u})$ can be used as a measure of the absolute largeness of the elements of $\mathbf{u}$. Suppose now that $\mathbf{y} = (y_1, \ldots, y_n)$ are observed values and $\mathbf{z} = (z_1, \ldots, z_n)$ are the corresponding fitted values using some statistical model. Then we can evaluate the goodness of the fit by $s(\mathbf{y} - \mathbf{z})$.

The most common scale is the one based on the quadratic function and is given by $s_1(\mathbf{u}) = (\sum_{j=1}^{n} u_j^2/n)^{1/2}$. This scale is clearly nonrobust. Huber (1981) defines a general class of robust scales, called M scales, as follows.

Let ρ be a function satisfying the following properties:

A: (i) $\rho(0) = 0$; (ii) ρ is even; (iii) if $|x_1| < |x_2|$, then $\rho(x_1) \leq \rho(x_2)$; (iv) ρ is bounded; (v) ρ is continuous.

Then, an *M scale* $s_3(\mathbf{u})$ based on ρ is defined by the value s satisfying

$$\frac{1}{n}\sum_{j=1}^{n} \rho\left(\frac{u_j}{s}\right) = b, \tag{4}$$

where b is a given scalar and $0 < b < a = \sup \rho$.

Yohai and Zamar (1988) introduce the family of τ scales. A τ *scale* is based on two functions ρ_1 and ρ_2 satisfying conditions A such that $\rho_2 \leq \rho_1$. To define a τ scale, one considers an M scale $s_3(\mathbf{u})$ defined by (4) with ρ_1 in place of ρ; then, the τ scale

is given by

$$\tau^2(\mathbf{u}) = s_3^2(\mathbf{u})\frac{1}{n}\sum_{j=1}^{n}\rho_2\left(\frac{u_j}{s_3(\mathbf{u})}\right). \quad (5)$$

In the next section we will discuss the advantages of the τ scale with respect to the M scale to define regression estimates.

3.2 τ Estimators for Linear Regression

Let us consider the regression model

$$y_j = \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_j + e_j,\ 1 \le j \le n, \quad (6)$$

where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^{\mathrm{T}}$ and $\mathbf{x}_j = (x_{j1}, \ldots, x_{jp})$. For a given $\boldsymbol{\beta}$, let $r_j(\boldsymbol{\beta}) = y_j - \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_j$ be the corresponding residuals. The scale $\tau^2(r_1(\boldsymbol{\beta}), \ldots, r_n(\boldsymbol{\beta}))$ may be considered as a measure of goodness of fit. Based on this remark, Yohai and Zamar (1988) define robust estimators of the coefficients of a regression model by

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \tau(r_1(\boldsymbol{\beta}), \ldots, r_n(\boldsymbol{\beta})). \quad (7)$$

These estimators are called τ *regression estimators*. If $a/b = 0.5$, the τ estimators have breakdown point (bdp) close to 50% (Yohai and Zamar 1988). Moreover, we note that, if $\rho_2(u) = u^2$, $\tau^2(u_1, \ldots, u_n) = \mathrm{ave}(u_j^2)$ and then the regression τ estimator coincides with the least squares estimator. Therefore, taking as ρ_2 a bounded function close to the quadratic function, the regression τ estimators can be made arbitrarily efficient for normal errors. We should remark that a robust regression estimator could be defined in a similar way using an M scale. However, Hössjer (1992) showed that this estimator cannot simultaneously have high bdp and high efficiency.

If the errors e_j in (6) are heteroscedastic with variances proportional to σ_j^2, the efficiency of $\hat{\boldsymbol{\beta}}$ can be improved by means of a weighted procedure. A *regression weighted* τ *estimator* is given by

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \tau(r_1^*(\boldsymbol{\beta}), \ldots, r_n^*(\boldsymbol{\beta})),$$

where $r_j^*(\boldsymbol{\beta}) = r_j(\boldsymbol{\beta})/\sigma_j$.

Usually, one chooses ρ_1 and ρ_2 in the *Tukey biweight family*

$$\rho^T(u, c) = \begin{cases} 3(u/c)^2 - 3(u/c)^4 + (u/c)^6 & \text{if} \quad |u| \le c, \\ 1 & \text{if} \quad |u| > c, \end{cases} \quad (8)$$

using two values c_1 and c_2 of the "tuning parameter" c. For example, one can take $c_1 = 1.548$ and $c_2 = 6.08$. With $b = 0.5$, these values yield regression estimators with breakdown point 0.5 and normal efficiency of 95%.

3.3 Qτ and Weighted Qτ Estimators

Consider a three-parameter family $F_{\boldsymbol{\theta}}(y) = F^*((y-\mu)/\sigma, \lambda)$ with $\boldsymbol{\theta} = (\mu, \sigma, \lambda)$, such as the LG$(\mu, \sigma, \lambda)$ family. We are now ready to define the Qτ estimators for this family.

For $0 < u < 1$ we define $Q(u, \boldsymbol{\theta})$ as the u-quantile of $F_{\boldsymbol{\theta}}(y)$. Then, $Q(u, \boldsymbol{\theta}) = \sigma\, Q^*(u, \lambda) + \mu$, where $Q^*(u, \lambda) = Q(u, (0, 1, \lambda))$. Let $y_{(1)}, \ldots, y_{(n)}$ be the order statistics of a random sample of size n from $F_{\boldsymbol{\theta}_0}(y)$, where $\boldsymbol{\theta}_0 = (\mu_0, \sigma_0, \lambda_0)$ is the parameter vector to be estimated. Since, $y_{(j)}$ can be considered the quantile $u_{n,j} = (j - 0.5)/n$ of the empirical distribution, it should be close to $\sigma_0 Q^*(u_{n,j}, \lambda_0) + \mu_0$. More precisely, we consider the differences between the empirical and the theoretical quantiles as a function of $\boldsymbol{\theta}$ which are given by

$$r_{n,j}(\boldsymbol{\theta}) = y_{(j)} - \mu - \sigma\, Q^*(u_{n,j}, \lambda).$$

To measure how large are these residuals, we can use a robust scale. We can use an M or a τ scale; however, our preliminary experiments showed that the τ scale performs better than the M scale. This is not new according to what is known for regression (Yohai and Zamar 1988; see also the comment after (7)). Therefore, we will take the scale $\tau^2(r_{n,1}(\boldsymbol{\theta}), \ldots, r_{n,n}(\boldsymbol{\theta}))$ as a measure of goodness of fit of a distribution $F_{\boldsymbol{\theta}}$ with respect to the data. Hence, it is natural to define the $Q\tau$ *estimators* by

$$\tilde{\boldsymbol{\theta}}_n = \arg\min_{\boldsymbol{\theta}} \tau(r_{n,1}(\boldsymbol{\theta}), \ldots, r_{n,n}(\boldsymbol{\theta})). \quad (9)$$

We note that, fixing λ, the value of μ and σ minimizing the τ scale are obtained by a simple regression τ estimate for the responses $y_{(j)}$ and the regressors $Q^*(u_{n,j}, \lambda)$. We also note (Serfling 1980) that $n^{1/2} r_{n,j}(\boldsymbol{\theta}_0)$ is approximately distributed according to $N(0, v^2(\boldsymbol{\theta}_0, u_{n,j}))$, where

$$v^2(\boldsymbol{\theta}_0, u) = \frac{\sigma_0^2 u(1-u)}{f_{\lambda_0}^2(Q^*(u, \lambda_0))}. \quad (10)$$

Then, since we are dealing with a regression with heteroscedastic errors, it is natural to estimate the variances of the errors by

$$\hat{\sigma}_j^2 = v^2(\tilde{\boldsymbol{\theta}}_n, u_{n,j}) \quad (11)$$

and to improve the basic estimator by means of a weighted procedure. We finally define the *weighted* $Q\tau$ (W$Q\tau$) *estimators* of $\boldsymbol{\theta}_0$ by

$$\tilde{\boldsymbol{\theta}}_n^w = \arg\min_{\boldsymbol{\theta}} \tau\left(\frac{r_{n,1}(\boldsymbol{\theta})}{\hat{\sigma}_1}, \ldots, \frac{r_{n,n}(\boldsymbol{\theta})}{\hat{\sigma}_n}\right). \quad (12)$$

3.4 Computation

We briefly discuss how to optimize the τ scales in (9) and (12). For a given λ_0 the values $\mu(\lambda)$ and $\sigma(\lambda)$ minimizing the τ scale can be computed using the algorithm described in Salibian-Barrera, Willems, and Zamar (2008) for τ regression estimates. Therefore, we take a grid $\lambda_1, \ldots, \lambda_k$ of values of λ and, for each value λ_l in the grid, we compute the corresponding values $\mu(\lambda_l)$ and $\sigma(\lambda_l)$ using the algorithm mentioned above. Then, an initial value for λ is obtained by minimizing the τ scales over the triplets $(\mu(\lambda_l), \sigma(\lambda_l), \lambda_l)$ for $1 \le l \le k$. This procedure can be iterated taking local grids around the current approximation until the desired accuracy is obtained.

3.5 Asymptotic and Robustness Properties

Under regularity conditions on the parameter space, the family $F_{\boldsymbol{\theta}}$, and the functions ρ_1 and ρ_2, we can prove that, for $n \to \infty$, the Qτ and the W$Q\tau$ estimators converge almost surely to $\boldsymbol{\theta}_0$ and that the rate of convergence is $n^{1/2}$ (supplementary materials, Theorem 1 and Theorem 2). We can also show that these estimators have maximal breakdown point (bdp) of 50% (see however Section 3.6). Moreover, the empirical results reported

below show that the WQτ estimators are quite efficient when the data are distributed according to the model. Unfortunately, we cannot prove that they are asymptotically normal, and empirical results suggest that they are not. Therefore, they are not convenient for inference. However, a WQτ estimator is a convenient starting point of a combined procedure, where it provides the basis to down weight the outliers. The final estimator is a 1SWL estimator described in Section 4.

3.6 The Finite Sample Distribution Breakdown Point

According to the traditional definitions (Donoho and Huber 1983; Hampel et al. 1986) the bdp of an estimator $\hat{\boldsymbol{\theta}}_n$ is, roughly speaking, the largest proportion ϵ of atypical observations that the data may contain such that $\hat{\boldsymbol{\theta}}_n$ remains bounded. In the case of the generalized log-gamma distribution, this definition is however inadequate for the following reason: we can find a sequence of parameters $\boldsymbol{\theta}^{(k)} = (\mu^{(k)}, \sigma^{(k)}, \lambda^{(k)})$ such that $||\boldsymbol{\theta}^{(k)}|| \to \infty$, but $F_{\boldsymbol{\theta}^{(k)}} \to F_0$, where F_0 is a proper distribution. One example where this occurs is $\boldsymbol{\theta}^{(k)} = (0, 1/k, -k)$; in this case $F_0(y) = 1 - \exp(-y)$. Suppose now that $\mathbf{y}^{(k)}$ $(k \geq 1)$ is a sequence of contaminated samples with a proportion of outliers smaller or equal than ϵ and that $\hat{\boldsymbol{\theta}}_n(\mathbf{y}^{(k)}) = (0, 1/k, -k)$. According to the usual definition, we would say that the bdp of $\hat{\boldsymbol{\theta}}_n$ is smaller or equal than ϵ. However, since the sequence of distribution $F_{\hat{\boldsymbol{\theta}}^{(k)}}$ remains bounded, this contradicts our intuitive notion of bdp. To overcome this contradiction we propose the following new definition of bdp.

Definition. Assume that we have a fixed sample $\mathbf{x} = (x_1, \ldots, x_n)$ with nominal distribution belonging to a family $F_{\boldsymbol{\theta}}$ with domain $\mathbb{R}$ and let

$$\mathcal{M}_m = \left\{ \mathbf{y} = (y_1, \ldots, y_n) \middle| \sum_{i=1}^{n} I(y_i \neq x_i) \leq m \right\}$$

be the set of all contaminated samples with a number of outliers smaller or equal than m. The *finite sample distribution breakdown point* (fsdbdp) of an estimator $\hat{\boldsymbol{\theta}}_n$ at the sample $\mathbf{x}$ is defined as the supremum of all ε such that, if $m < n\varepsilon$, for any $\delta > 0$ there exists K such that, for any sample $\mathbf{y} \in \mathcal{M}_m$, we have that $F_{\hat{\boldsymbol{\theta}}_n(y)}(K) - F_{\hat{\boldsymbol{\theta}}_n(\mathbf{y})}(-K) > 1 - \delta$.

According to this definition, the fsdbdp occurs when the estimated distribution corresponding to contaminated samples assigns a positive mass to the complement of any compact set. It can be shown that the fsdbdp of the Qτ and WQτ estimators of the log-gamma model is larger than or equal 0.5 (the details are provided in Section 4 of the supplementary materials).

4. THE WEIGHTED LIKELIHOOD ESTIMATORS

Let us assume that an initial highly robust and consistent but not necessarily efficient estimate $\tilde{\boldsymbol{\theta}}_n^{(0)}$ of $\boldsymbol{\theta}_0$—for example, a WQτ estimate $\tilde{\boldsymbol{\theta}}_n^{w}$ defined above—is available. Then we can define a one-step weighted likelihood estimator with starting value $\tilde{\boldsymbol{\theta}}_n^{(0)}$. This family of estimators was introduced by Agostinelli and Markatou (1998) as a one-step version of the weighted likelihood (WL) estimators proposed by Lindsay (1994) and also studied by Markatou, Basu, and Lindsay (1997, 1998) and Basu, Shioya, and Park (2011). Under very general conditions, the one-step weighted likelihood estimators are asymptotically normal with asymptotic covariance matrix equal to the inverse of the information matrix, that is, they are fully efficient. Moreover, they inherit the highly robust behavior of $\tilde{\boldsymbol{\theta}}_n^{(0)}$.

The basic WL equation is

$$\frac{1}{n} \sum_{j=1}^{n} w(y_j, \boldsymbol{\theta}) \mathbf{z}(y_j, \boldsymbol{\theta}) = \mathbf{0}, \tag{13}$$

where $\mathbf{z}(y, \boldsymbol{\theta})$ is the score function vector. The weight function $w(y, \boldsymbol{\theta})$ is defined by

$$w(y, \boldsymbol{\theta}) = \min\left(1, \frac{[A(\delta(y, \boldsymbol{\theta})) + 1]^+}{\delta(y, \boldsymbol{\theta}) + 1}\right),$$

where $\delta(y, \boldsymbol{\theta})$ is the *Pearson residual*, measuring the agreement between the data and the assumed model. It is defined as $\delta(y, \boldsymbol{\theta}) = [f^*(y) - f_{\boldsymbol{\theta}}^*(y)]/f_{\boldsymbol{\theta}}^*(y)$, where $f^*(y) = \int k(y, t, h) dF_n(t)$ is a kernel density estimate of $f_{\boldsymbol{\theta}}$ with bandwidth h, $f_{\boldsymbol{\theta}}^*(y) = \int k(y, t, h) f_{\boldsymbol{\theta}}(t) dt$ is the corresponding smoothed model density, F_n is the empirical cumulative distribution function, and $[x]^+ = \max(0, x)$. Comments on the choice of the bandwidth can be found in Section 6 of the supplementary materials. The function $A(.)$ is called *residual adjustment function* (RAF). When $A(\delta(y_j, \boldsymbol{\theta})) = \delta(y_j, \boldsymbol{\theta})$ the weights $w(y_j, \boldsymbol{\theta}) = 1$ and (13) coincides with (3). Generally, the weight function w uses functions A that correspond to minimum disparity problems. In particular, the function

$$A(\delta) = 2 - (2 + \delta)\exp(-\delta) \tag{14}$$

corresponds to the *negative exponential* (NE) *disparity* measure and will be used in all our numerical experiments. Equation (13) can be solved using an iterative algorithm. We call the solution of (13) the *fully iterated weighted likelihood (FIWL) estimator*. We also consider a simpler *one-step weighted likelihood (1SWL) estimator* defined by

$$\hat{\boldsymbol{\theta}}_n = \tilde{\boldsymbol{\theta}}_n^{(0)} - \mathbf{J}^{-1} \sum_{j=1}^{n} w\big(y_j, \tilde{\boldsymbol{\theta}}_n^{(0)}\big) \mathbf{z}\big(y_j, \tilde{\boldsymbol{\theta}}_n^{(0)}\big), \tag{15}$$

where $\mathbf{J} = \sum_{j=1}^{n} w(y_j, \tilde{\boldsymbol{\theta}}_n^{(0)}) \nabla \mathbf{z}(y_j, \tilde{\boldsymbol{\theta}}_n^{(0)})$ and ∇ denotes differentiation with respect to $\boldsymbol{\theta}$. This definition is similar to a one-step Taylor expansion of (13) in the neighborhood of $\tilde{\boldsymbol{\theta}}_n^{(0)}$. The Taylor expansion contains an extra term which is obtained by differentiating the weight with respect to $\boldsymbol{\theta}$. This term, when evaluated at the model, is equal to zero. We note that the estimate of σ_0 provided by (15) can be negative and to avoid this incident we reparameterize the model using $\gamma = \sigma^{0.5}$.

The numerical experiments reported in the next section, indicate that, for finite sample sizes, $\hat{\boldsymbol{\theta}}_n$ is almost as efficient as the fully iterated estimator, but provides a better resistance to outlier contamination. Under regularity conditions we can prove that, for $n \to \infty$,

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{\mathcal{L}} N(\mathbf{0}, \mathbf{I}(\boldsymbol{\theta}_0)^{-1}),$$

where $\mathbf{I}(\boldsymbol{\theta}_0)$ is the Fisher information matrix (supplementary materials, Theorem 3). In addition, the influence function IF of the 1SWL estimator at the model is exactly that of the ML estimator (supplementary materials, Section 4). As such, the 1SWL estimator is fully first order efficient. The IF clearly depends on the score function and it is unbounded whenever

the score is unbounded. Despite this fact, the 1SWL estimator of the generalized log-gamma model is robust and has the same fsdbdp as the starting estimator (supplementary materials, Theorem 5).

5. MONTE CARLO EXPERIMENTS

Several numerical simulations were performed to compare the following estimators: the maximum likelihood estimator (ML); the Qτ estimator (Qτ) given by (9); the weighted Qτ estimator (WQτ) given by (12); the FIWL estimator defined by (13), and the 1SWL estimator defined by (15). To compute the WL estimators, we used a normal kernel with fixed bandwidth $h = 0.3$ in all experiments. FIWL and 1SWL started with WQτ and used (14). For Qτ and WQτ we took ρ_1 and ρ_2 in the bisquare family with $c_1 = 1.548$, $c_2 = 6.08$, and $b = 0.5$.

To compare the performances of the different estimators, we computed the mean absolute error (MAE) of the single parameter estimates with respect to the true values λ_0, μ_0, σ_0, and η_0. In addition, for each set of parameter estimates $\bar{\boldsymbol{\theta}}$ we computed the total variation distance (TVD)

$$\mathrm{TVD}(f_{\bar{\boldsymbol{\theta}}}, f_{\boldsymbol{\theta}_0}) = \int |f_{\bar{\boldsymbol{\theta}}}(y) - f_{\boldsymbol{\theta}_0}(y)| dy$$

between the estimated density $f_{\bar{\boldsymbol{\theta}}}$ and the true underlying density $f_{\boldsymbol{\theta}_0}$. TVD clearly measures the quality of the estimated density over the entire range of y values. We also simulated asymptotic confidence intervals for the parameters and compared their empirical and nominal coverages.

5.1 Monte Carlo Experiments at the Nominal Model

In a first experiment, we generated n observations y_j according to (1) with $\mu = 0$ and $\sigma = 1$, five values of λ, namely 0, 0.5, 1, 1.5, and 2, and four values of n, namely $n = 50$, 100, 400, and 1000. Note that, since all the estimates are location and scale equivariant, this choice of μ and σ is without loss of generality. The number of replications was 1000. The MAE of the single parameter estimates as well as the mean TVD of the estimated densities are reported in the supplementary materials. Figure 1

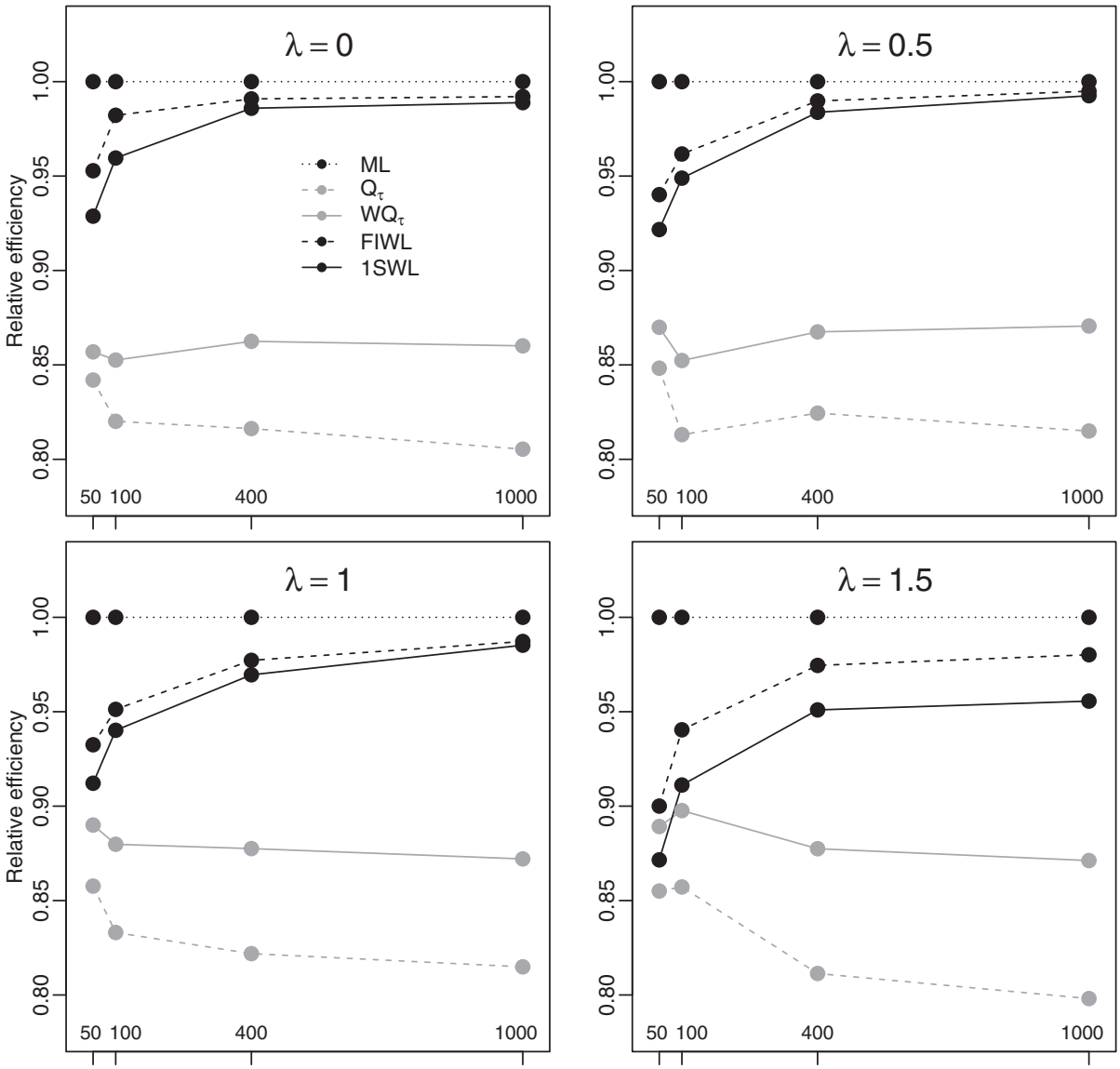

Figure 1. Relative efficiency with respect to the ML estimator for $n = 50$, 100, 400, 1000.

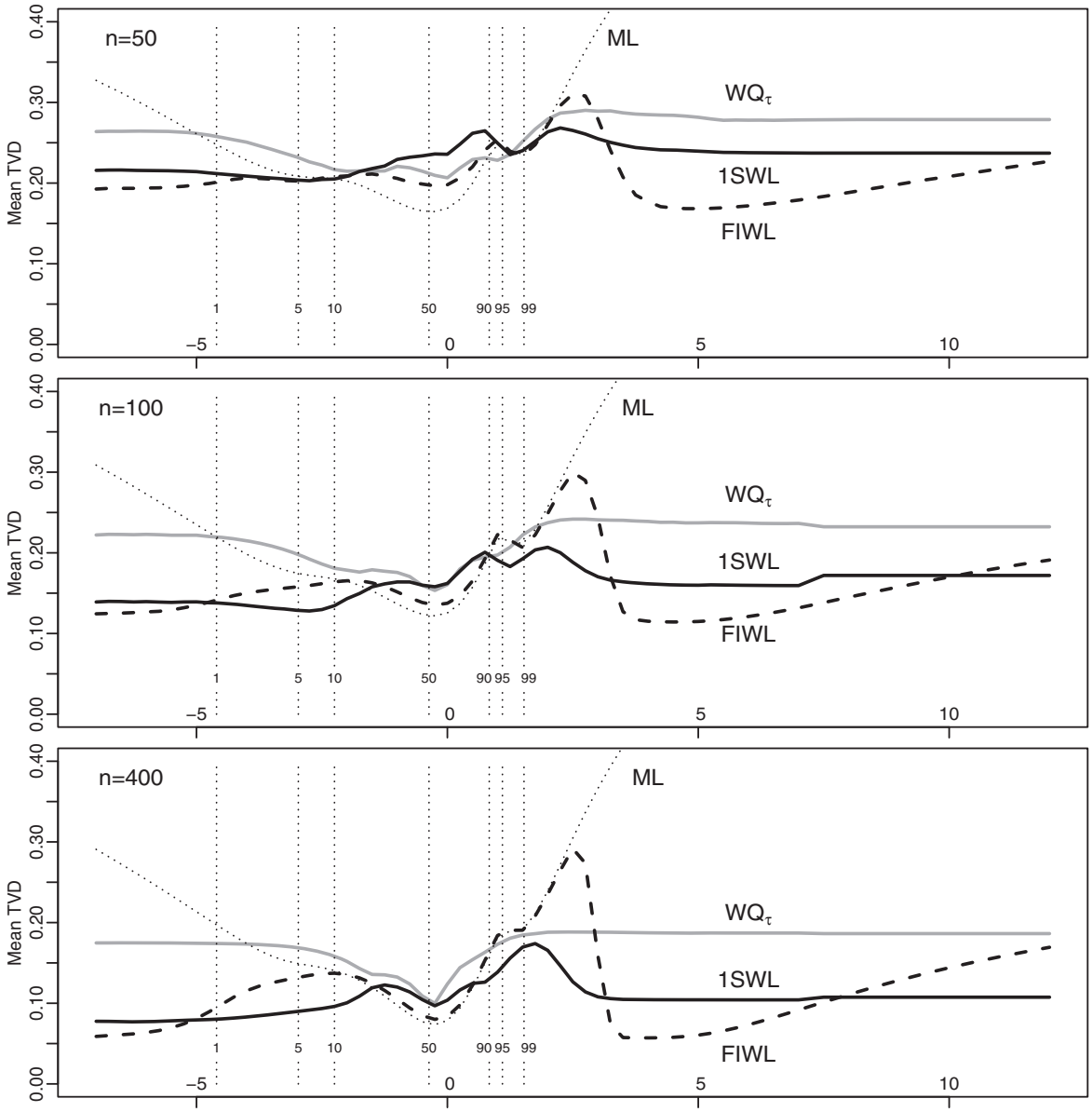

Figure 2. Mean total variation distance under 10% point contamination at different values of the abscissa. Vertical lines indicate the 1%, 5%, 10%, 50%, 90%, 95%, and 99% quantiles of the model.

graphically shows (for $\lambda = 0, 0.5, 1, 1.5$) the mean TVD of the ML estimate divided by the mean TVD of the robust estimate as a function of the sample size. This ratio can be interpreted as a measure of relative efficiency. We note that WQτ provides a noticeable increase of efficiency with respect to Qτ and that FIWL and 1SWL substantially improve the behavior of WQτ. FIWL is the best performing robust estimator. However, the performance of 1SWL is not significantly inferior.

5.2 Monte Carlo Experiments Under Point Contamination

In a second experiment, we compared ML, WQτ, FIWL, and 1SWL with a simulation under point contamination. We generated $n - n_0$ "good" observations y_j according to a log-gamma model (1)–(2) with $\mu = 0$, $\sigma = 1$, $\lambda = 1$, and n_0 "outliers" at the point u for u ranging from -6 to 12, $n = 50$, 100, and 400. This kind of point contamination is clearly unrealistic; however, it is generally the least favorable one and allows evaluation of the maximal bias an estimator can incur. The fraction $\epsilon = n_0/n$ of outliers was 10%, a very high contamination level in real data applications. For each value of u the number of replications was 1000. Figure 2 reports the mean TVD of the estimated densities as a function of u. The MAE of the single parameter estimates as a function of u can be found in the supplementary materials. The results show that the mean TVD (and the MAE) of the WQτ density estimate is remarkably stable. Thus, WQτ is a good starting point for the 1SWL. The maximal TVD of 1SWL is smaller than the maximal TVD of FIWL, indicating that 1SWL provides a better resistance to point contamination. Finally, 1SWL improves the performance of the initial WQτ estimate and provides reliable results for all values of u.

5.3 Coverage of the Asymptotic Confidence Intervals

In a third experiment, we evaluated by Monte Carlo simulation the coverage of the asymptotic confidence intervals of the

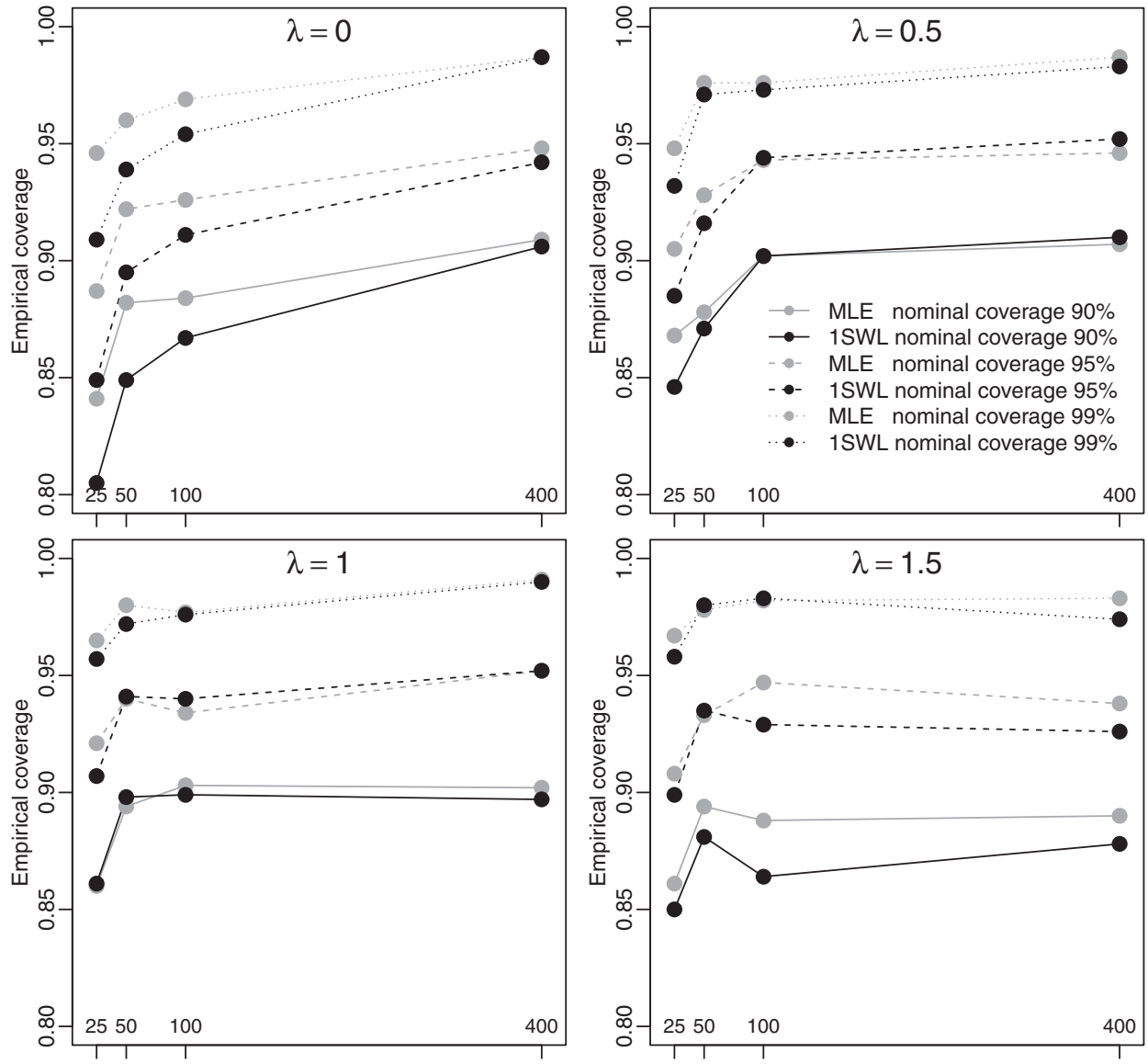

Figure 3. Monte Carlo coverages in the absence of contamination for $n = 25, 50, 100, 400$.

parameters for finite sample size. We considered five values of λ (0, 0.5, 1, 1.5, and 2), four values of n (25, 50, 100, and 400), and three nominal coverage levels (0.9, 0.95, and 0.99). The complete numerical results are available in the supplementary materials. In Figure 3 we graphically report the empirical coverages for $\lambda = 0, 0.5, 1$, and 1.5. We consider that the coverages are satisfactory for practical purposes for all sample sizes, with the exception of $n = 25$. Note that, in this case, all the considered estimators, including ML, yield asymptotic confidence intervals with coverage very different from the nominal level. Even bootstrap experiments with $n = 25$ (not reported here) provided very large and practically noninformative intervals. We believe that —as a popular rule of thumb suggests—the sample size should be somewhat larger than 10 times the number of parameters.

6. EXAMPLE

Clarke, McKinnon, and Riley (2012) describe a simple dataset taken from a large measurement quality assurance (QA) data base of ALCOA aluminium refineries in Western Australia. Under ALCOA's QA program, several thousand bauxite ore samples are routinely submitted to Fourier transform infrared spectroscopy (Eyer and Riley 1999). Part of the quality assurance is the need to automatically highlight unusual spectra and this is obtained with the help of special statistical diagnostics—called representation indicators—derived from the Fourier transform. Three samples of a particular indicator are displayed in Figure 4, both on the original and the logarithmic scales. Sample 3 contains two outliers, which are so far away in the original scale that fall out of the graph limits. However, they are still clearly visible on the logarithmic scale. Modeling the sample distribution of this indicator is potentially an important step of the QA procedure and the two-parameter gamma family of distributions has been used for this purpose in Clarke, McKinnon, and Riley (2012). Due to the presence of sporadic "out of control" spectra in training data (e.g., sample 3) robust estimation has to be used to summarize the samples.

Four estimated models are shown in Figure 4: the ML estimates of the two-parameter gamma and log-gamma models

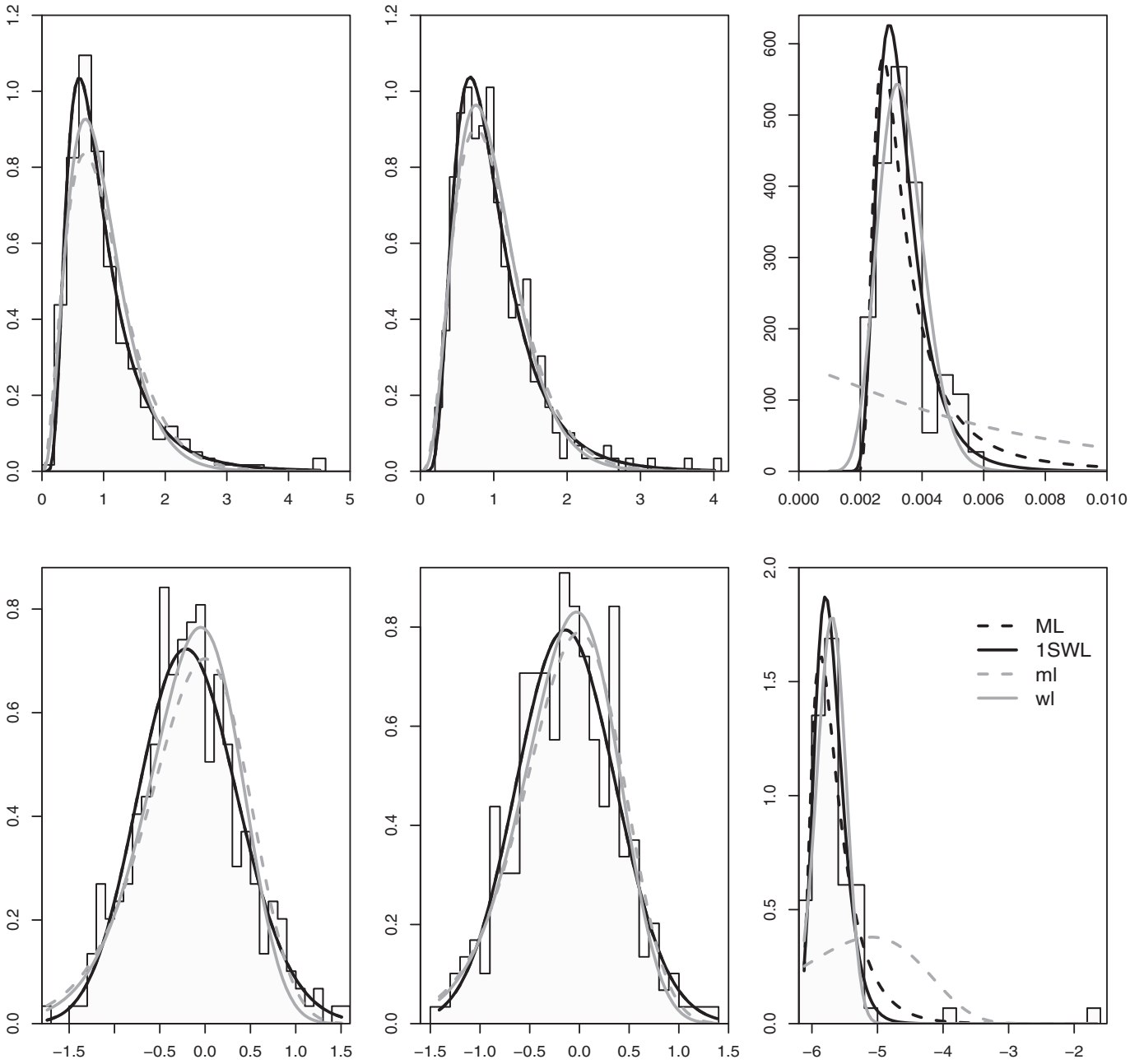

Figure 4. Three samples of a bauxite ore quality indicator and density estimates.

(ml), the ML estimates of the generalized gamma and log-gamma models (ML), the WL estimates of the gamma and log-gamma models as implemented in the R package wle (Agostinelli 2001) (wl), and the 1SWL estimate of the generalized gamma and log-gamma models.

The parameter estimates of the corresponding log-gamma models are provided in Table 1 (note that $\sigma = \lambda$ in the two-parameter models). In this table, two additional estimates are reported from Clarke, McKinnon, and Riley (2012)—the method of moments estimate (mo) and a robust Cramér von Mises estimate (cm). The FIWL estimate is also shown in Table 1.

We immediately see that the classical and robust estimates are very similar for sample 1 and sample 2, whereas the shapes of the classical estimates (ml, mo, and ML) are heavily distorted by the outliers in sample 3. In fact, both FIWL and 1SWL have weights approximately equal to 1 for all the observations in the first two samples, while the two extreme observations on the right tail are heavily downweighted in the third sample. Thus, the weights can be used to identify the outliers. If we remove the two outliers, the ML estimates of μ, σ, and λ become -5.78, 0.21, and -0.59, respectively. These values are very close to those provided by FIWL and 1SWL.

We also note that in the three-parameter models all the 95% confidence intervals for λ include $\lambda = 0$, with the exception of the nonrobust interval based on ML in sample 3. Thus, the lognormal model may be a possible model for these data. Is the two-parameter gamma family another acceptable model? To answer this question, we tested the hypothesis $\lambda = \sigma$ with the help of a weighted version of the log-likelihood ratio test proposed in Agostinelli and Markatou (2001). We used the weights

Table 1. Parameter estimates and standard errors (italic) of models fitted to three samples of a bauxite ore quality indicator

	Sample 1 $n = 297$				Sample 2 $n = 297$				Sample 3 $n = 76$			
	μ	σ	λ	η	μ	σ	λ	η	μ	σ	λ	η
ML	−0.21	0.55	−0.19	1.00	−0.14	0.50	−0.06	1.00	−5.47	0.19	−1.78	0.004
	0.05	*0.02*	*0.14*	*0.04*	*0.05*	*0.02*	*0.14*	*0.03*	*0.05*	*0.03*	*0.44*	*0.000*
FIWL	−0.21	0.55	−0.19	1.00	−0.14	0.50	−0.06	1.00	−5.78	0.20	−0.60	0.003
	0.05	*0.02*	*0.14*	*0.04*	*0.05*	*0.02*	*0.14*	*0.03*	*0.04*	*0.02*	*0.29*	*0.000*
1SWL	−0.21	0.55	−0.19	1.00	−0.14	0.49	−0.07	1.00	−5.79	0.20	−0.75	0.003
	0.05	*0.02*	*0.14*	*0.04*	*0.05*	*0.02*	*0.14*	*0.03*	*0.04*	*0.02*	*0.29*	*0.000*
ml	0.00	0.55	0.55	1.00	0.00	0.49	0.49	1.00	−5.09	0.97	0.97	0.006
	0.03	*0.02*	*0.02*	*0.03*	*0.03*	*0.02*	*0.02*	*0.03*	*0.11*	*0.07*	*0.07*	*0.001*
wl	−0.05	0.51	0.51	0.95	−0.02	0.47	0.47	0.97	−5.69	0.22	0.22	0.003
	0.03	*0.02*	*0.02*	*0.03*	*0.03*	*0.02*	*0.02*	*0.03*	*0.03*	*0.02*	*0.02*	*0.001*
mo	−0.00	0.63	0.63	1.00	−0.01	0.54	0.54	100	−5.10	3.53	3.53	0.06
cm	−0.06	0.51	0.51	0.94	−0.04	0.48	0.48	0.97	−5.53	0.22	0.22	0.003

NOTE: ML: ML of generalized log-gamma; FIWL: fully iterated WL of generalized log-gamma; 1SWL: one-step WL of the generalized log-gamma; ml: ML of two-parameter log-gamma; wl: WL of two-parameter log-gamma; mo: method of moments of two-parameter log-gamma; cm: Cramér von Mises of two-parameter log-gamma.

provided by ML (all weights equal 1, which leads to the classical log-likelihood ratio test), FIWL, and 1SWL, and obtained the following values of the respective test statistics: 27.6, 27.6, and 27.6 in sample 1; 15.13, 15.10, and 14.90 in sample 2; 198.18, 12.7, and 11.6 in sample 3. Removing the two outliers, the classical test statistic becomes 5.96. The hypothesis is rejected in all the cases (all the p-values are extremely small). We note the huge effect of outliers on the classical statistic for sample 3.

7. DISCUSSION

Most literature about robust estimation deals with location-scale models (and their extensions to regression and multivariate scatter). In this article, we consider for the first time distribution models with an additional shape parameter and for these models—in particular, for the generalized log-gamma model—we develop new robust procedures.

The $Q\tau$ estimate is a robust version of the well-known quantile distance estimates described in La Riccia (1982). Our Monte Carlo simulations show that this kind of estimate performs well both in the case that the model is correct and under a corrupted model. These empirical findings are corroborated by a theoretical results showing that the $Q\tau$ estimate has a maximum 50% bdp according to a new definition (the fsdbdp) which is particularly designed to asses the degree of global stability of a distribution estimate.

Unfortunately, the $Q\tau$ estimate is not asymptotically normal and, therefore, inconvenient for inference. Its rates of convergence is, however, of order $n^{1/2}$ and this makes it a good starting point for a two-step procedure which is asymptotically normal and improves the efficiency of the $Q\tau$ estimate while maintaining the same fsdbdp. More precisely, we recommend a fully efficient 1SWL estimate.

The $Q\tau$ estimate may be extended to the case of randomly censored observations. For this purpose, the quantiles of the empirical distribution can be replaced by the quantiles of the Kaplan–Meier distribution corresponding to the noncensored observations (some details are available in the supplementary materials). A similar idea can be applied to the 1SWL estimate, where a kernel density estimates for censored data (Liebscher 2002) can be used to compute the weights. Another idea is to proceed in a similar way as in Locatelli, Marazzi, and Yohai (2010), where a parametric consistent estimator of the underlying distribution has been used. Extensions to accelerated failure time regression with errors distributed according to the generalized log-gamma model are also possible. Both these extensions are matter of further research.

SUPPLEMENTARY MATERIALS

This document provides the mathematical justifications of the proposed methods, additional results of the simulation experiments, and an introduction to the R package *robustloggamma* (Agostinelli et al. 2013) containing code to compute the estimators described in the article. The package also contains the dataset used in the example of Section 6.

ACKNOWLEDGMENTS

This research was partially supported by grants W276 from Universidad of Buenos Aires, PIP 112-2008-01-00216 and 112-2011-01-00339 from CONICET and PICT 2011-0397 from ANPCYT, Argentina. The authors thank A. Randriamiharisoa for the programming help and B.R. Clarke for making available the quality assurance dataset. The authors are grateful to the two anonymous referees and to the editors for their comments that have largely contributed to improve the original manuscript.

[Received December 2011. Revised June 2013.]

REFERENCES

Agostinelli, C. (2001), "Wle: A Package for Robust Statistics Using Weighted Likelihood," available at *http://cran.r-project.org*. [8]

Agostinelli, C., Marazzi, A., Yohai, V. J., and Randriamiharisoa, A. (2013), "robustloggamma: A Package for Robust Estimation of the Generalized Log-Gamma Distribution," available at *http://cran.r-project.org*. [9]

Agostinelli, C., and Markatou, M. (1998), "A One-Step Robust Estimator for Regression Based on the Weighted Likelihood Reweighting Scheme," *Statistics & Probability Letters*, 37, 341–350. [4]

——— (2001), "Test of Hypotheses Based on the Weighted Likelihood Methodology," *Statistica Sinica*, 11, 499–514. [8]

Almpanidis, G., and Kotropoulos, C. (2008), "Phonemic Segmentation Using the Generalised Gamma Distribution and Small Sample Bayesian Information Criterion," *Speech Communication*, 50, 38–55. [1]

Barkauskas, D. A., Kronewitter, S. R., Lebrilla, C. B., and Rocke, D. M. (2009), "Analysis of MALDI FT-ICR Mass Spectrometry Data: A Time Series Approach," *Analytica Chimica Acta*, 648, 207–214. [1]

Basu, A., Shioya, H., and Park, C. (2011), *Statistical Inference: The Minimum Distance Approach*, Boca Raton: Chapman & Hall/CRC. [4]

Boudt, K., Caliskan, D., and Croux, C. (2011), "Robust Explicit Estimators of Weibull Parameters," *Metrika*, 73, 187–209. [1]

Clarke, B. R., McKinnon, P. L., and Riley, G. (2012), "A Fast Robust Method for Fitting Gamma Distributions," *Statistical Papers*, 53, 1001–1014. [1,7,8]

Cowell, F. A., and Victoria-Feser, M.-P. (1996), "Poverty Measurement With Contaminated Data: A Robust Approach," *European Economic Review*, 40, 1761–1771. [1]

Donoho, D. L., and Huber, P. J. (1983), "The Notion of Breakdown Point," in *Festschrift fur Erich L. Lehmann*, eds. P. J. Bickel, K. Doksum, and J. L. Hodges, Jr., Belmont, CA: Wadsworth, pp. 157–184. [4]

Dornheim, H., and Brazauskas, V. (2007), "Robust and Efficient Methods for Credibility When Claims are Approximately Gamma Distributed," *North American Actuarial Journal*, 11, 138–158. [1]

Dupuis, D. J., and Mills, J. E. (1998), "Robust Estimation of the Birnbaum-Saunders Distributions," *IEEE Transactions on Reliability*, 47, 88–95. [1]

Eyer, S., and Riley, G. (1999), "Measurement Quality Assurance in a Production System for Bauxite Analysis by FTIR," North American Chapter of the International Chemometrics Society, Newsl No. 19. [7]

Field, C., and Smith, B. (1994), "Robust Estimation: A Weighted Maximum Likelihood Approach," *International Statistical Review*, 62, 405–424. [1]

Huber, P. J. (1981), *Robust Statistics*, New York: Wiley. [2]

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., and Stahel, W. A. (1986), *Robust Statistics: The Approach Based on Influence Functions*, New York: Wiley. [4]

Hössjer, O. (1992), "On the Optimality of S-Estimators," *Statistics & Probability Letters*, 14, 413–419. [3]

La Riccia, V. N. (1982), "Asymptotic Properties of Weighted L_2 Quantile Distance Estimators," *The Annals of Statistics*, 10, 621–624. [1,9]

Lawless, J. F. (1980), "Inference in the Generalized Gamma and Log-Gamma Distributions," *Technometrics*, 22, 409–419. [1]

——— (2003), *Statistical Models and Methods for Lifetime Data* (2nd ed.), Hoboken: Wiley. [1]

Liebscher, K. (2002), "Kernel Density and Hazard Rate Estimation for Censored Data Under α-Mixing Condition," *Annals of the Institute of Statistical Mathematics*, 54, 19–28. [9]

Lindsay, B. J. (1994), "Efficiency Versus Robustness: The Case of Minimum Hellinger Distance and Related Methods," *The Annals of Statistics*, 22, 1081–1114. [4]

Locatelli, I., Marazzi, A., and Yohai, V. J. (2010), "Robust Accelerated Failure Time Regression," *Computational Statistics and Data Analysis*, 55, 874–887. [9]

Manning, W. G., Basu, A., and Mullahy, J. (2005), "Generalized Modeling Approaches to Risk Adjustment of Skewed Outcomes Data," *Journal of Health Economics*, Amsterdam: Elsevier, 24, 465–488. [1]

Marazzi, A., and Ruffieux, C. (1999), "The Truncated Mean of an Asymmetric Distribution," *Computational Statistics & Data Analysis*, 32, 79–100. [1]

Markatou, M., Basu, A., and Lindsay, B. G. (1997), "Weighted Likelihood Estimating Equations: The Discrete Case With Applications to Logistic Regression," *Journal of Statistical Planning and Inference*, 57, 215–232. [4]

——— (1998), "Weighted Likelihood Equations With Bootstrap Root Search," *Journal of the American Statistical Association*, 93, 740–750. [4]

Meeker, W. Q., and Escobar, L. A. (1998), *Statistical Methods for Reliability Data*, New York: Wiley. [1]

Nadarajah, S. (2008), "On the Use of the Generalised Gamma Distribution," *International Journal of Electronics*, 95, 1029–1032. [1]

Nadarajah, S., and Gupta, A. K. (2007), "A Generalized Gamma Distribution With Application to Drought Data," *Mathematics and Computers in Simulation*, 74, 1–7. [1]

Prentice, R. L. (1974), "A Logamma Model and its Maximum Likelihood Estimation," *Biometrika*, 61, 539–544. [1,2]

Ruckdeschel, P., and Horbenko, N. (2012), "Yet Another Breakdown Point Notion: EFSBP – Illustrated at Scale-Shape Models," *Metrika*, 75, 1025–1047. [1]

Salibian-Barrera, M., Willems, G., and Zamar, R. H. (2008), "The Fast-Tau Estimator for Regression," *Journal of Computational and Graphical Statistics*, 17, 659–682. [3]

Serfling, R. J. (1980), *Approximation Theorems of Mathematical Statistics*, New York: Wiley. [3]

Shin, J. W., Chang, J.-H., and Kim, N. S. (2005), "Statistical Modeling of Speech Signals Based on Generalized Gamma Distribution," *IEEE Signal Processing Letters*, 12, 258–261. [1]

Stacy, E. W. (1962), "A Generalization of the Gamma Distribution," *The Annals of Mathematical Statistics*, 33, 1187–1192. [1]

Yohai, V. J., and Zamar, R. H. (1988), "High Breakdown Estimates of Regression by Means of the Minimization of an Efficient Scale," *Journal of the American Statistical Association*, 83, 406–413. [1,2,3]