Robust estimators of accelerated failure time regression with generalized log-gamma errors

Claudio Agostinelli[a], Isabella Locatelli[b], Alfio Marazzi [1c], Víctor J. Yohai[d]

[a]*Department of Mathematics, University of Trento, Trento, Italy*
[b]*Institute of social and preventive medicine, Lausanne University Hospital, Switzerland*
[c]*Institute of social and preventive medicine, Lausanne University Hospital, Switzerland and Nice Computing SA, Ch. de Maillefer 37, CH-1052 Le Mont/Lausanne*
[d]*Departamento de Matematicas, Facultad de Ciencias Exactas y Naturales, University of Buenos Aires and CONICET*

Abstract

The generalized log-gamma (GLG) model is a very flexible family of distributions to analyze datasets in many different areas of science and technology. Estimators are proposed which are simultaneously highly robust and highly efficient for the parameters of a GLG distribution in the presence of censoring. Estimators with the same properties for accelerated failure time models with censored observations and error distribution belonging to the GLG family are also introduced. It is proven that the proposed estimators are asymptotically fully efficient and the maximum mean square error is examined using Monte Carlo simulations. The simulations confirm that the proposed estimators are highly robust and highly efficient for a finite sample size. Finally, the benefits of the proposed estimators in applications are illustrated with the help two real datasets.

Keywords: Censored data, Quantile distance estimates, τ estimators, Truncated maximum likelihood estimators, Weighted likelihood estimators

1. Introduction

Generalized log-gamma (GLG) regression with censored observations is a large class of Accelerated Failure Time (AFT) models introduced by Lawless

[1]Corresponding author: Alfio Marazzi, Institut universitaire de médecine sociale et préventive, Route de la Corniche 10, 1010 Lausanne - Switzerland, Tel: ++41 21 314 72 72, Fax: ++41 21 314 73 73, Email: Alfio.Marazzi@chuv.ch

(1980). Many models broadly used in the lifetime data analysis – including log-normal, log-gamma, and log-Weibull regression – are specific cases of GLG regression. GLG regression has been widely applied in various areas of survival analysis (e.g. Kim et al., 1993; Sun et al., 1999; Abadi et al., 2012). Procedures to fit the GLG regression model have been added to the capabilities of leading statistical software such as SAS and STATA.

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.

There are two basic strategies to detect outliers in regression models. The first one makes use of diagnostic tools based on ML residuals. Specific proposals for GLG regression are given in Ortega et al. (2003, 2008) and Silva et al. (2010). However, this strategy may fail because the ML estimators could be largely distorted by the outliers. As a result, the corresponding residuals are not necessarily large and, therefore, may not be visible. An improvement of this strategy is the "leave one approach", where the ML residual of one observation is computed without that observation. This strategy may cope with isolated outliers. However, in the case that the sample contains a group of similar outliers, a "masking effect" can occur, that is, the remaining outliers of the group may cause the residual of the observation under study to be small and this observation remains "hidden".

A better strategy, which avoids this shortcoming, is the use of a robust estimator, that is an estimator which is not very sensitive to the presence of outliers. Two families of robust estimators of models with three parameters (location, scale, and shape), including GLG, without censored observations and without covariate information have been introduced by Agostinelli et al. (2014). These families of estimators are: the (weighted) quantile τ (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. It is $n^{1/2}$ consistent but not asymptotically normal. However, it is a convenient starting point to define the 1SWL-estimator.

In this paper, we extend the Qτ estimator proposed in Agostinelli et al. (2014) to GLG regression with right censoring by introducing the trimmed Qτ-estimator (TQτ-estimator); we also extend the truncated maximum likelihood (TML) estimator proposed in Marazzi and Yohai (2004) to GLG regression. To improve the robustness of this estimator without modifying its asymptotic efficiency we also define a one-step version of the TML estimator (1TML-estimator). For the sake of completeness, we also define an extension of the 1SWL estimator which is fully described in the Supplementary Material. However a Monte Carlo study shows that this estimator is much less robust than the TQτ- and 1TML- estimators.

The procedures introduced here for the GLG family can be applied to other location-scale-shape models, such as the three-parameter log-Weibull family.

Section 2 defines the Qτ- and TQτ-estimators for censored observations in the absence of covariables. Section 3 describes the TML estimators. Section 4 extends the estimators to the regression case. Section 5 shows the results of a Monte Carlo study comparing the performance of the proposed methods for finite sample sizes. Section 6 discusses two examples with real data. Section 7 provides concluding remarks. Proofs and complementary technical details can be found in the Supplementary Material.

2. The Qτ- and TQτ-estimators for censored observations without covariables

2.1. The generalized gamma and log-gamma distribution models

According Cox et al. (2007), a positive random survival time T has a *generalized gamma distribution* $GG(\mu, \sigma, \lambda)$ with parameters μ, σ, and λ ($\mu \in \mathbb{R}$, $\sigma > 0$, $\lambda \in \mathbb{R}$), if the cdf of T is

$$G_{\mu,\sigma,\lambda}(t) = \begin{cases} G_{\lambda^{-2}}\left[\lambda^{-2}(e^{-\mu}t)^{\lambda/\sigma}\right] & \text{if} \quad \lambda > 0, \\ 1 - G_{\lambda^{-2}}\left[\lambda^{-2}(e^{-\mu}t)^{\lambda/\sigma}\right] & \text{if} \quad \lambda < 0. \end{cases} \tag{1}$$

Here, $G_\gamma(t) = \int_0^t x^{\gamma-1}e^{-x}dx/\Gamma(\gamma)$ is the cumulative distribution function (cdf) of the gamma distribution with mean and variance equal to $\gamma > 0$ and Γ denotes the Gamma function. The density of T is given by

$$g_{\mu,\sigma,\lambda}(t) = \frac{|\lambda|}{\sigma t\Gamma\left(\lambda^{-2}\right)}\left[\lambda^{-2}(e^{-\mu}t)^{\lambda/\sigma}\right]^{\lambda^{-2}} \exp\left[-\lambda^{-2}(e^{-\mu}t)^{\lambda/\sigma}\right]. \tag{2}$$

We refer to $GG(0, 1, \lambda)$ as the *standard case* with $\mu = 0$, $\sigma = 1$. If $S \sim GG(0, 1, \lambda)$, then $e^\mu S^\sigma \sim GG(\mu, \sigma, \lambda)$. The generalized gamma family of distributions includes - as special cases - many common survival models (e.g. Lawless, 2003) such as the two-parameter gamma distribution ($\lambda = \sigma$) with mean e^μ and variance $\sigma^2 e^{2\mu}$, the Weibull distribution ($\lambda = 1$), the exponential distribution ($\lambda = \sigma = 1$). The limiting case $\lambda = 0$ is the lognormal distribution.

In order to describe our proposed methods, we will use logarithmic survival times. Thus, we consider the random variable $Y = \log(T)$ and say that Y follows a *generalized log-gamma distribution* $GLG(\mu, \sigma, \lambda)$. It can be shown that Y follows a location-scale model such that

$$Y = \mu + \sigma U, \tag{3}$$

where U has density

$$f_\lambda(t) = \begin{cases} \frac{|\lambda|}{\Gamma(\lambda^{-2})}(\lambda^{-2})^{\lambda^{-2}} \exp\left((\lambda^{-2})\left(\lambda t - e^{\lambda t}\right)\right) & \text{if} \quad \lambda \neq 0, \\ \frac{1}{\sqrt{2\pi}} \exp(-\frac{t^2}{2}) & \text{if} \quad \lambda = 0. \end{cases} \tag{4}$$

The distribution of U is a $GLG(0, 1, \lambda)$, which we refer as the *standard case*. The density of Y will be denoted by $f_{\mu,\sigma,\lambda}$ and called a *generalized log-gamma density* with *location* parameter μ, *scale* parameter σ, and *shape* parameter λ. We will use the abbreviations $\boldsymbol{\theta} = (\mu, \sigma, \lambda)$ and $f_{\boldsymbol{\theta}}(t) = f_{\mu,\sigma,\lambda}(t)$. We have

$$f_{\boldsymbol{\theta}}\,(t) = \frac{1}{\sigma} f_\lambda \left(\frac{t - \mu}{\sigma} \right) \tag{5}$$

and, since $T = \exp(Y) \sim GG(\mu, \sigma, \lambda)$, the cdf of Y is given by

$$F_{\boldsymbol{\theta}}(t) = G_{\mu,\sigma,\lambda}(\exp(t)). \tag{6}$$

The generalized log-gamma family of distributions includes the normal model ($\lambda = 0$), the log-Weibull model ($\lambda = 1$), the log-exponential model ($\lambda = \sigma = 1$) as well as the gamma GLM with logarithmic link.

Note. There are other equivalent parametrizations of the gamma and log-gamma distributions (e.g. Johnson et al., 1994). The parametrization defined by (1)-(2) and (3)-(4) is due to Prentice (1974). According to Cox et al. (2007) this is a convenient parametrization to compute the ML estimator. In addition, the location-scale property (3) is a basic requirement of the Qτ estimator proposed below.

We will suppose that $y_1, \ldots, y_n$ are n i.i.d. observations of Y with cdf $F_{\boldsymbol{\theta}_0}$, where $\boldsymbol{\theta}_0 = (\mu_0, \sigma_0, \lambda_0)$ and propose estimators of $\boldsymbol{\theta}_0 = (\mu_0, \sigma_0, \lambda_0)$.

We consider single censoring on the right, where the true values of $y_1, ..., y_n$ are not observed. Instead, the censored observations $y_i^* = \min(y_i, c_i)$ ($i = 1, ..., n$) are observed, where $c_1, ..., c_n$ are i.i.d. censoring "times", which are independent of the y_i's. Note that this type of non-informative censoring also includes the case where censoring is partly due to the end of the study. We define the censoring indicator $\delta_i = 1$ if $y_i^* = y_i$ and $\delta_i = 0$ if $y_i^* = c_i$. Let $\boldsymbol{z}_i = (y_i^*, \delta_i)$ and G_n be the empirical distribution function based on $(\boldsymbol{z}_1, \ldots, \boldsymbol{z}_n)$.

2.2. Score functions and ML estimator

The ML estimator of the parameters of a GLG model under censoring can be easily defined as follows. Let $S_{\boldsymbol{\theta}}(t) = 1 - F_{\boldsymbol{\theta}}(t) = 1 - F_\lambda((t-\mu)/\sigma)$ denotes the survival function. Then, the negative log-likelihood function is

$$-\sum_{i=1}^{n} \left[\delta_i \log f_{\boldsymbol{\theta}}(y_i^*) + (1-\delta_i) \log S_{\boldsymbol{\theta}}(y_i^*)\right]. \tag{7}$$

In the absence of censoring, the score functions $\boldsymbol{d} = (d_1, d_2, d_3)^\top$ are

$$d_1(t, \boldsymbol{\theta}) = -\frac{\partial}{\partial \mu} \log f_{\boldsymbol{\theta}}(t) = \frac{1}{\sigma} \xi_\lambda(u), \tag{8}$$

$$d_2(t, \boldsymbol{\theta}) = -\frac{\partial}{\partial \sigma} \log f_{\boldsymbol{\theta}}(t) = \frac{1}{\sigma} \left(\xi_\lambda(u) u + 1\right), \tag{9}$$

$$d_3(t, \boldsymbol{\theta}) = -\frac{\partial}{\partial \lambda} \log f_{\boldsymbol{\theta}}(t) = \psi_\lambda(u), \tag{10}$$

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

$$\xi_\lambda(u) = \frac{f'_\lambda(u)}{f_\lambda(u)} = \frac{(1 - e^{\lambda u})}{\lambda},$$

$$\psi_\lambda(u) = -\frac{\partial}{\partial \lambda} \log f_\lambda(u) = \frac{1}{\lambda^3}(2\zeta(\lambda) - \lambda^2 + \lambda u - \exp(\lambda u)(2 - \lambda u)),$$

$\zeta(\lambda) = -2\log(\lambda) - \dot{\Gamma}(\lambda^{-2}) + 1$, and $\dot{\Gamma}$ denotes the digamma function. Let

$$s_1(t, \boldsymbol{\theta}) = -\frac{\partial}{\partial \mu} \log S_{\boldsymbol{\theta}}(t) = -\frac{1}{\sigma} \frac{f_\lambda(u)}{S_\lambda(u)}, \tag{11}$$

$$s_2(t, \boldsymbol{\theta}) = -\frac{\partial}{\partial \sigma} \log S_{\boldsymbol{\theta}}(t) = -\frac{1}{\sigma} \frac{f_\lambda(u) u}{S_\lambda(u)}, \tag{12}$$

$$s_3(t, \boldsymbol{\theta}) = -\frac{\partial}{\partial \lambda} \log S_{\boldsymbol{\theta}}(t) = \frac{\dot{F}_\lambda(u)}{S_\lambda(u)}, \tag{13}$$

where $\dot{F}_\lambda(u) = dF_\lambda(u)/d\lambda$. Then, the score functions for the case with censored observations are

$$v_k(t,\delta,\boldsymbol{\theta}) = \delta d_k(t,\boldsymbol{\theta}) + (1-\delta)s_k(t,\boldsymbol{\theta}), \qquad k = 1,2,3. \tag{14}$$

The *ML estimator* of $\boldsymbol{\theta}$ is given by the following system of equations

$$E_{G_n}(\boldsymbol{v}(y,\delta,\boldsymbol{\theta})) = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{v}(y_i^*,\delta_i,\boldsymbol{\theta}) = \mathbf{0}, \tag{15}$$

where $\boldsymbol{v} = (v_1, v_2, v_3)^\top$ is the score function vector. It is easy to show that

$$-\nabla_{\boldsymbol{\theta}} \log S_{\boldsymbol{\theta}}(y_i^*) = E_{\boldsymbol{\theta}}(\boldsymbol{d}(y,\boldsymbol{\theta})|y > y_i^*),$$

where $\nabla_{\boldsymbol{\theta}}$ indicates differentiation w.r.t. $\boldsymbol{\theta}$. Hence, an alternative expression for the likelihood equations is

$$\frac{1}{n}\sum_{i=1}^{n} \delta_i \boldsymbol{d}(y_i^*,\boldsymbol{\theta}) + (1-\delta_i)E_{\boldsymbol{\theta}}(\boldsymbol{d}(y,\boldsymbol{\theta})|y > y_i^*) = \mathbf{0}.$$

Following Locatelli et al. (2010), we define the *semiempirical cdf of* y *for a given* $\boldsymbol{\theta}$ as

$$H_{n,\boldsymbol{\theta}}(t) = \frac{1}{n}\sum_{i=1}^{n} E_{\boldsymbol{\theta}}\left[I\left(y \le t\right) \middle|\, y_i^*, \delta_i\right]$$

or equivalently,

$$H_{n,\boldsymbol{\theta}}\left(t\right) = \frac{1}{n}\sum_{i=1}^{n} \delta_i I\left(y_i \le t\right) + \frac{1}{n}\sum_{i=1}^{n} \left(1-\delta_i\right) \frac{\left[F_{\boldsymbol{\theta}}\left(t\right) - F_{\boldsymbol{\theta}}(y_i^*)\right]^+}{1 - F_{\boldsymbol{\theta}}(y_i^*)}. \tag{16}$$

Thus, when there is no censoring, $H_{n,\boldsymbol{\theta}}(t)$ coincides with the usual empirical cdf. If $\tilde{\boldsymbol{\theta}}$ is a consistent estimator of $\boldsymbol{\theta}_0$, then $H_{n,\tilde{\boldsymbol{\theta}}}(t)$ is a consistent estimator of $F_{\boldsymbol{\theta}_0}(t)$ and, for any measurable function $h(t)$, we have $\lim_{n\to\infty} E_{n,\tilde{\boldsymbol{\theta}}}\left[h\left(y\right)\right] = E_{\boldsymbol{\theta}_0}\left[h\left(y\right)\right]$ a.s., where $E_{n,\boldsymbol{\theta}}$ denotes expectation under $H_{n,\boldsymbol{\theta}}$. Finally, another expression of the likelihood equations is

$$E_{n,\boldsymbol{\theta}}(\boldsymbol{d}(y,\boldsymbol{\theta})) = \mathbf{0}. \tag{17}$$

Let $\boldsymbol{M}(\boldsymbol{\theta}) = E\left(\boldsymbol{v}(y,\delta,\boldsymbol{\theta})\boldsymbol{v}(y,\delta,\boldsymbol{\theta})^\top\right)$ and $\boldsymbol{G}(\boldsymbol{\theta}) = E\left(\nabla_{\boldsymbol{\theta}}\boldsymbol{v}(y,\delta,\boldsymbol{\theta})\right)$ then, the asymptotic covariance matrix of the ML estimator is

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}_0) = \boldsymbol{G}(\boldsymbol{\theta}_0)^{-1}\boldsymbol{M}(\boldsymbol{\theta}_0)\boldsymbol{G}(\boldsymbol{\theta}_0)^{-\top}.$$

2.3. The trimmed Qτ estimator

Agostinelli et al. (2014) define the quantile τ (Qτ) estimator and the weighted Qτ (WQτ) estimator for non-censored i.i.d. observations as follows. For $0 < u < 1$, let $Q(u, \boldsymbol{\theta})$ denote the u-quantile of $F_{\boldsymbol{\theta}}$. Then, $Q(u, \boldsymbol{\theta}) = \sigma Q^*(u, \lambda) + \mu$, where $Q^*(u, \lambda) = Q(u, (0, 1, \lambda))$. Given a sample $y_1, \ldots, y_n$, let F_n denote the empirical cdf of Y. Then, $y_{(1)}, \ldots, y_{(n)}$, the ordered observations, are the quantiles $u_{n,j} = (j - 0.5)/n$ of F_n and should be close to $\sigma_0 Q^*(u_{n,j}, \lambda_0) + \mu_0$ for $j = 1, \ldots, n$. Consider the differences between the empirical and the theoretical quantiles

$$r_{n,j}(\boldsymbol{\theta}) = y_{(j)} - \mu - \sigma Q^*(u_{n,j}, \lambda), \qquad j = 1, \ldots, n.$$

The *Q*τ *estimator* is defined by

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

where τ denotes the τ scale.

The τ scale was introduced by Yohai and Zamar (1988) to define estimators which combine high finite sample breakdown point with high efficiency in the linear model with normal errors. Given a sample $\boldsymbol{u} = (u_1, \ldots, u_n)$, a function $s(\boldsymbol{u})$ is called a *scale* if: (i) $s(\boldsymbol{u}) \geq 0$; (ii) for any scalar γ, $s(\gamma \boldsymbol{u}) = |\gamma| s(\boldsymbol{u})$; (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 follows that (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 $s(u_1, \ldots, u_n) < \varepsilon$. Properties (i)-(vi) clearly show that $s(\boldsymbol{u})$ can be used as a measure of the absolute largeness of the elements of $\boldsymbol{u}$. The most common scale is the one based on the quadratic function and is given by $s_1(\boldsymbol{u}) = \left(\sum_{i=1}^n u_i^2/n\right)^{1/2}$. This scale is clearly non robust. Huber (1981) defines a general class of robust scales, called M scales, as follows. Let ρ be a function satisfying the following properties:

A1 : (i) $\rho(0) = 0$; (ii) ρ is even; (iii) if $|t_1| < |t_2|$, then $\rho(t_1) \leq \rho(t_2)$; (iv) ρ is bounded; (v) ρ is continuous.

Then, an *M scale* $s_2(\boldsymbol{u})$ based on ρ is defined by the value s satisfying

$$\frac{1}{n} \sum_{i=1}^n \rho\left(\frac{u_i}{s}\right) = b, \tag{18}$$

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 **A1** and such that $\rho_2 \leq \rho_1$. One considers an M scale $s_2(\boldsymbol{u})$ defined by (18) with ρ_1 in place of ρ; then, the τ *scale* is given by

$$\tau^2(\boldsymbol{u}) = s_2^2(\boldsymbol{u}) \frac{1}{n} \sum_{i=1}^{n} \rho_2 \left(\frac{u_i}{s_2(\boldsymbol{u})} \right). \tag{19}$$

Usually, ρ, ρ_1 and ρ_2 are selected in the Tukey's bi-weight family given by

$$\rho_c^T(t) = 1 - \max\left(\left(1 - \left(\frac{t}{c}\right)^2\right)^3, 1 \right) \tag{20}$$

for convenient values of c and b (see Section 5).

The Qτ estimator in the case of randomly censored observations is obtained by replacing the quantiles of the empirical distribution by the quantiles of the Kaplan-Meier (KM) distribution corresponding to the non censored observations. More precisely, let $\tilde{F}_n$ denote the KM estimator (Kaplan and Meier, 1958) of $F_{\boldsymbol{\theta}_0}$ and $z_{(1)}, \ldots, z_{(m)}$ the ordered non censored observations. Then, $z_{(1)}, \ldots, z_{(m)}$ are the quantiles $\tilde{u}_{n,i} = \tilde{F}_n(z_{(i)}) - 0.5/n$. The residuals

$$\tilde{r}_{n,i}(\boldsymbol{\theta}) = z_{(i)} - \mu - \sigma Q^*(\tilde{u}_{n,i}, \lambda), \qquad i = 1, \ldots, m$$

are then used to define the *Qτ estimator for censored observations* by

$$\tilde{\boldsymbol{\theta}}_n = \arg\min_{\boldsymbol{\theta}} \tau(\tilde{r}_{n,1}(\boldsymbol{\theta}), \ldots, \tilde{r}_{n,m}(\boldsymbol{\theta})). \tag{21}$$

It is known that the KM estimator distributes the mass of the censored observations among all the observations that are on their right. Some of these observations may be outliers and, therefore, the mass assigned to the outliers by KM may be inflated by the observations on their left. To reduce the influence of outliers we propose a trimmed version of the Qτ estimator as follows. Let $0 < \alpha < 1$ be the fraction of trimming and let $k_\alpha = \max\{h : \tilde{F}_n(z_{(h)}) \leq 1 - \alpha\}$ then, the *α-trimmed Qτ (α-TQτ) estimator* is defined by

$$\tilde{\boldsymbol{\theta}}_n = \arg\min_{\boldsymbol{\theta}} \tau(\tilde{r}_{n,1}(\boldsymbol{\theta}), \ldots, \tilde{r}_{n,k_\alpha}(\boldsymbol{\theta})). \tag{22}$$

Since in all the simulations and examples reported below we will use $\alpha = 0.1$, to simplify the notation in the remaining of the paper, we will write TQτ instead of α-TQτ.

Finally, note that the residuals $\tilde{r}_{n,i}(\boldsymbol{\theta})$ are heteroskedastic and, according to Serfling (1980), their variance can be approximated by

$$\hat{\sigma}_i^2 = v^{*2}_{\tilde{u}_{n,i}} / f^2_{\tilde{\lambda}}(Q^*(\tilde{u}_{n,i}, \tilde{\lambda})), \tag{23}$$

where $v^{*2}_{\tilde{u}_{n,i}}$ is Greenwood's variance estimator of $\tilde{F}_n(z_{(i)})$ (Greenwood, 1926). Then, as in Agostinelli et al. (2014), we might consider the *weighted* TQτ (WTQτ) *estimator*

$$\tilde{\boldsymbol{\theta}}^w_n = \arg\min_{\boldsymbol{\theta}} \tau\left(\frac{\tilde{r}_{n,1}(\boldsymbol{\theta})}{\hat{\sigma}_1}, \ldots, \frac{\tilde{r}_{n,k_\alpha}(\boldsymbol{\theta})}{\hat{\sigma}_{k_\alpha}}\right). \tag{24}$$

However, our Monte Carlo experiments have shown that weighting does not provide an important improvement and, for this reason, we are not going to consider the WTQτ estimator further.

In Theorem 1 of the Supplementary Material we prove that, under general conditions, the TQτ estimator $\tilde{\boldsymbol{\theta}}_n$ is $n^{1/2}$-consistent, that is,

$$n^{1/2}(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) = O_p(1). \tag{25}$$

As in the non censored case, one drawback of the Qτ estimators is that they are not asymptotically normally distributed, making inference difficult. In order to overcome this problem we introduce, in the next Section, the one step and the two steps truncated ML estimators starting at TQτ. These estimators have similar robustness properties as the TQτ estimator; in addition, they have asymptotic normal distribution with the same asymptotic variance as the ML estimator under the model.

3. The truncated maximum likelihood estimators

In order to obtain a robust and highly efficient procedure for estimating the unknown parameter vector we use the truncated maximum likelihood (TML) estimator and the one-step TML (1TML) estimator. Both are based on a weighted form of the likelihood equations. Suppose that $y \sim GLG(\boldsymbol{\theta}_0)$ and consider a sample $(\boldsymbol{z}_1, \ldots, \boldsymbol{z}_n)$, $\boldsymbol{z}_i = (y_i^*, \delta_i)$ $(i = 1, \ldots, n)$. Assume that $\tilde{\boldsymbol{\theta}} = (\tilde{\mu}, \tilde{\sigma}, \tilde{\lambda})$ is an initial consistent estimator of $\boldsymbol{\theta}_0$. A natural robustification of the likelihood equations (17) can be obtained by weighting the equations. More precisely, given a weight function $w(t, \boldsymbol{\theta})$, we consider the equations

$$E_{n,\boldsymbol{\theta}}[w(y, \tilde{\boldsymbol{\theta}})\boldsymbol{d}(y, \boldsymbol{\theta})] = E_{\tilde{\boldsymbol{\theta}}}[w(y, \tilde{\boldsymbol{\theta}})\boldsymbol{d}(y, \tilde{\boldsymbol{\theta}})]. \tag{26}$$

The right hand side mitigates the bias of the estimator and allows to proof asymptotic normality. For increasing sample size its tends to zero. In the next subsections we show how one can define the weight function.

3.1. The outlier rejection rule

We proceed as in Marazzi and Yohai (2004). Let $\tilde{r}_i^* = (y_i^* - \tilde{\mu})/\tilde{\sigma}$ denote the standardized residuals with respect to the initial model and let $l_\lambda(t) = -\log f_\lambda(t)$ be the negative log-likelihood function. We consider the negative log-likelihoods of the residuals $l_i^* = l_{\tilde{\lambda}}(\tilde{r}_i^*)$ $(i = 1, \ldots, n)$. A large l_i^* corresponds to an observation with a small likelihood under the model and suggests that y_i is an outlier. Let M_λ be the cdf of $l_\lambda(y)$ and

$$M_{n,\tilde{\lambda}}(t) = \frac{1}{n}\sum_{i=1}^{n}\left[\delta_i I\left(l_i^* \le t\right) + (1-\delta_i)P_{\tilde{\lambda}}(l_{\tilde{\lambda}}(y) \le t | y > y_i^*)\right],$$

be the semi-empirical cdf of $l_1^*, \ldots, l_n^*$ for $\lambda = \tilde{\lambda}$. One can show that $M_{n,\tilde{\lambda}}$ is a consistent estimator of M_{λ_0}. Let $M_{n,\tilde{\lambda}}^{(\varphi)}$ denote $M_{n,\tilde{\lambda}}$ truncated at φ, i.e.,

$$M_{n,\tilde{\lambda}}^{(\varphi)}(t) = \begin{cases} M_{n,\tilde{\lambda}}(t)/M_{n\tilde{\lambda},}(\varphi) & \text{if } t \le \varphi, \\ 1 & \text{otherwise.} \end{cases} \tag{27}$$

We want to compare the right tail of the truncated empirical distribution $M_{n,\tilde{\lambda}}^{(\varphi)}$ with the right tail of $M_{\tilde{\lambda}}$, which is the theoretical distribution when $\lambda = \tilde{\lambda}$. To specify what we understand by the tail, we take a number ε close to 0, for example $\varepsilon = 0.01$, as the probability of falling in the tail. Then, we define the cutoff point φ^* on the likelihood scale as the largest φ such that $M_{n,\tilde{\lambda}}^{(\varphi)}(t) \geqq M_{\tilde{\lambda}}(t)$ for all $t \ge M_{\tilde{\lambda}}^{-1}(1-\varepsilon)$, i.e.,

$$\varphi^* = \sup\{\varphi | M_{n,\tilde{\lambda}}^{(\varphi)}(t) \geqq M_{\tilde{\lambda}}(t) \text{ for all } t \ge M_{\tilde{\lambda}}^{-1}(1-\varepsilon)\}.$$

Note that φ^* is the minimum value φ such that the tails of $M_{n\tilde{\lambda},}^{(\varphi)}(t)$ and $M_{\tilde{\lambda}}(t)$ are comparable. As in Gervini and Yohai (2002), one can prove that, if the sample does not contain outliers, $\varphi^* \to \infty$ a.s. as $n \to \infty$. Finally, since $l_{\tilde{\lambda}}(t)$ is unimodal, there exist two solutions $\tilde{c}_L$ and $\tilde{c}_U$ of the equation $l_{\tilde{\lambda}}(t) = \varphi^*$. It is immediate that $l_{\tilde{\lambda}}(t) \le \varphi^*$ is equivalent to $\tilde{c}_L \le t \le \tilde{c}_U$. The cutoff points on the data scale are $\tilde{t}_L = \tilde{\mu} + \tilde{\sigma}\tilde{c}_L$ and $\tilde{t}_U = \tilde{\mu} + \tilde{\sigma}\tilde{c}_U$.

3.2. Weight functions

Let $\omega(z)$ be a function, such that

A2 (i) $\omega(t)$ is non-increasing; (ii) $\lim_{z\to-\infty}\omega(t) = 1$; (iii) $\omega(z) = 0$ for $z > 0$.

For example, let $c > 0$ and consider the function

$$\omega(t) = \rho(t, c) \cdot I(t \leq 0), \tag{28}$$

where $\rho(t, c) = \rho^T(t/c)$ is in the bi-weight family as defined in (20) and let $\vartheta^* = 1/\varphi^*$. Then, define the weight function

$$w_0(t, \lambda, \vartheta^*) = \omega\left(l_\lambda(t) - 1/\vartheta^*\right) \tag{29}$$

and, for an observation y_i,

$$w(y_i, \boldsymbol{\theta}, \vartheta^*) = w_0\left(\frac{y_i - \mu}{\sigma}, \lambda, \vartheta^*\right). \tag{30}$$

Note that if we use (28) and $c \to 0$, we have $w(y_i, \boldsymbol{\theta}, \vartheta^*) = 1$ if $y_i \epsilon [\tilde{t}_L, \tilde{t}_U]$ and 0 otherwise; this rule is usually called *hard rejection*.

3.3. The TML estimators

The *TML estimator* $\hat{\boldsymbol{\theta}} = (\hat{\mu}, \hat{\sigma}, \hat{\lambda})$ is the solution of the equations (26), i.e.,

$$\boldsymbol{g}(\boldsymbol{\theta}, \tilde{\boldsymbol{\theta}}, \vartheta^*) = \boldsymbol{h}_{\tilde{\lambda}}, \tag{31}$$

where

$$\boldsymbol{g}(\boldsymbol{\theta}, \tilde{\boldsymbol{\theta}}, \vartheta^*) = E_{n,\tilde{\boldsymbol{\theta}}}\left[w(y, \tilde{\boldsymbol{\theta}}, \vartheta^*)\boldsymbol{d}(y, \boldsymbol{\theta})\right], \tag{32}$$

$$\boldsymbol{h}_{\tilde{\lambda}} = (h_{1,\tilde{\lambda}}, h_{2,\tilde{\lambda}}, h_{3,\tilde{\lambda}})^\top = E_{\tilde{\lambda}}[w_0(u, \tilde{\lambda}, \vartheta^*)\boldsymbol{d}(u, (0, 1, \tilde{\lambda})], \tag{33}$$

and $u \sim f_{0,1,\tilde{\lambda}}$. Note that, when $\vartheta^* \to 0$, we obtain the ML equations.

The *1TML estimator* is obtained by applying one iteration of the Newton-Raphson procedure to equations (31) and it turns out to be

$$\hat{\boldsymbol{\theta}}_1 = \hat{\boldsymbol{\theta}}_0 - \boldsymbol{J}(\hat{\boldsymbol{\theta}}_0, \vartheta^*)^{-1}(\boldsymbol{g}(\hat{\boldsymbol{\theta}}_0, \hat{\boldsymbol{\theta}}_0, \vartheta^*) - \boldsymbol{h}_{\hat{\lambda}_0}), \tag{34}$$

where $\hat{\boldsymbol{\theta}}_0 = \tilde{\boldsymbol{\theta}}$, $(\hat{\lambda}_0 = \tilde{\lambda})$, $\boldsymbol{J}(\hat{\boldsymbol{\theta}}_0, \vartheta^*) = \nabla_{\boldsymbol{\theta}} \boldsymbol{g}(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}_0, \vartheta^*)|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_0}$ is a Jacobian matrix. Similarly, we can define a *two-step TML* (2TML) *estimator* $\hat{\boldsymbol{\theta}}_2$ by replacing $\hat{\boldsymbol{\theta}}_0$, with $\hat{\boldsymbol{\theta}}_1$ in (34).

In Theorem 4 of the Supplementary Material we prove that, under general conditions including $n^{1/2}(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = O_p(1)$, the 1TML estimator satisfies the following asymptotic result:

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_0) \xrightarrow{\mathcal{L}} N(\boldsymbol{0}, \boldsymbol{G}(\boldsymbol{\theta}_0)^{-1}\boldsymbol{M}(\boldsymbol{\theta}_0)\boldsymbol{G}(\boldsymbol{\theta}_0)^{-\top}).$$

A similar result holds for the 2TML estimator. In the Monte Carlo simulations reported in Section 5 we show that, for finite sample sizes, the 1TML and the 2TML estimators have a reasonably robust behavior under outlier contamination and are more efficient than the TQτ estimator (with 10% trimming). Moreover, 2TML improves 1TML. Therefore we propose, as a final procedure to estimate $\boldsymbol{\theta}_0$, the 2TML estimator starting with the TQτ estimator with 10% trimming. Numerical experiments show that further steps do not provide any significant improvement.

4. The case with covariables

We now consider an AFT model for pairs of observations $(\boldsymbol{x}_i, y_i)$, $1 \leq i \leq n$, where $\boldsymbol{x}_i \in \mathbb{R}^p$ and $y_i \in \mathbb{R}$ satisfies

$$y_i = \mu_0 + \boldsymbol{\beta}_0^\top \boldsymbol{x}_i + \sigma_0 u_i, \qquad i = 1, \ldots, n. \tag{35}$$

Usually, y_i represents a duration on the logarithmic scale and $\boldsymbol{x}_i$ the corresponding covariable vector. The slopes $\boldsymbol{\beta}_0 \in \mathbb{R}^p$, the intercept μ_0, and the scale σ_0 are unknown parameters. The errors u_i, $1 \leq i \leq n$, are assumed to be i.i.d. and independent of $\boldsymbol{x}_i$. Moreover, the distribution of the carriers $\boldsymbol{x}_i$ is unknown. We assume that the error density is $f_{(0,1,\lambda_0)}$ according to (5), where λ_0 is an unknown shape parameter. We observe $(y_i^*, \boldsymbol{x}_i, \delta_i)$, where $y_i^* = \min(y_i, c_i)$, $c_1, \ldots, c_n$ are i.i.d. censoring times, which are independent of the u_i's. We put $\delta_i = 1$ if $y_i^* = y_i$ and $\delta_i = 0$ otherwise. We write $\boldsymbol{\gamma}_0 = (\boldsymbol{\theta}_0, \boldsymbol{\beta}_0)$, where $\boldsymbol{\theta}_0 = (\mu_0, \sigma_0, \lambda_0)$.

Let $\boldsymbol{\gamma} = (\boldsymbol{\theta}, \boldsymbol{\beta})$ and $u = (y - \mu - \boldsymbol{\beta}^\top \boldsymbol{x})/\sigma$. Let $\boldsymbol{d}(\boldsymbol{x}, y, \boldsymbol{\gamma})$ and $\boldsymbol{s}(\boldsymbol{x}, y, \boldsymbol{\gamma})$ denote the $(3+p)$-component non-censored and censored regression score function vectors respectively. The first 3 components of $\boldsymbol{d}(\boldsymbol{x}, y, \boldsymbol{\gamma})$ and $\boldsymbol{s}(\boldsymbol{x}, y, \boldsymbol{\gamma})$

are given by the right-hand sides of (8)-(10) and (11)-(13) respectively. In addition, for $k = 1, \ldots, p$, we have

$$d_{3+k}(\boldsymbol{x}, y, \boldsymbol{\gamma}) = \frac{x_k}{\sigma}\xi_\lambda(u),$$
$$s_{3+k}(\boldsymbol{x}, y, \boldsymbol{\gamma}) = -\frac{x_k}{\sigma}\frac{f_\lambda(u)}{S_\lambda(u)}.$$

Then, simple derivations shows that the ML estimator of $\boldsymbol{\gamma}$ is the solution of

$$\frac{1}{n}\sum_{j=1}^{n}\boldsymbol{v}(\boldsymbol{x}_j, y_j^*, \delta_j, \boldsymbol{\gamma}) = \mathbf{0}, \tag{36}$$

where $\boldsymbol{v}(\boldsymbol{x}, y, \delta, \boldsymbol{\gamma}) = \delta\boldsymbol{d}(\boldsymbol{x}, y, \boldsymbol{\gamma}) + (1-\delta)\boldsymbol{s}(\boldsymbol{x}, y, \boldsymbol{\gamma})$. A similar expression as (17) can be obtained, where the *semi-parametric cdf* is defined by

$$\begin{aligned} H_{n,\boldsymbol{\gamma}}(t, \boldsymbol{t}) &= \frac{1}{n}\sum_{i=1}^{n} E_{\boldsymbol{\gamma}}\left[I\left(y \le t\right) \middle| y_i^*, \boldsymbol{x}_i, \delta_i\right] I(\boldsymbol{x}_i \le \boldsymbol{t}) \\ &= \frac{1}{n}\sum \delta_i I\left(y_i \le t\right) I(\boldsymbol{x}_i \le \boldsymbol{t}) \\ &+ \frac{1}{n}\sum \left(1-\delta_i\right) \frac{\left[F_{\boldsymbol{\gamma}}(t) - F_{\boldsymbol{\gamma}}(y_i^*)\right]^+}{1 - F_{\boldsymbol{\gamma}}(y_i^*)} I(\boldsymbol{x}_i \le \boldsymbol{t}). \end{aligned}$$

We denote by $E_{n,\boldsymbol{\gamma}}$ the expectation with respect to $H_{n,\boldsymbol{\gamma}}$.

In the next two subsections we define a robust and efficient procedure for estimating $\boldsymbol{\gamma}$. In a first step an initial highly robust but not necessarily efficient estimator of $\boldsymbol{\gamma}_0$ is computed; the second step uses a highly robust and asymptotically efficient estimator of $\boldsymbol{\gamma}_0$ based on a 1TML procedure.

4.1. The initial regression estimator

The initial estimator of $\boldsymbol{\gamma}_0$ is defined as follows:

1. Let $\bar{\mu}$ and $\bar{\boldsymbol{\beta}}$ be MM-estimators for censored data of μ_0 and $\boldsymbol{\beta}_0$ as proposed in Salibian-Barrera and Yohai (2008).

2. Let $\omega_j = y_j - \bar{\boldsymbol{\beta}}^\top\boldsymbol{x}_j$ and $\omega_j^* = y_j^* - \bar{\boldsymbol{\beta}}^\top\boldsymbol{x}_j = \min(\omega_j, c_j^*)$, where $c_j^* = c_j - \bar{\boldsymbol{\beta}}^\top\boldsymbol{x}_j$. For large n, the distribution of the ω_js is close to $F_{\boldsymbol{\theta}_0}$ for $1 \le j \le n$. We therefore estimate $\boldsymbol{\theta}_0$ using the observations $(\omega_1^*, \delta_1), \ldots, (\omega_n^*, \delta_n)$ by means of a TQτ estimator that will be denoted by $\tilde{\boldsymbol{\theta}} = (\tilde{\mu}, \tilde{\sigma}, \tilde{\lambda})$.

3. The initial regression estimator of $\boldsymbol{\gamma}_0$ is $\tilde{\boldsymbol{\gamma}} = (\tilde{\boldsymbol{\theta}}, \bar{\boldsymbol{\beta}})$. It will be called *MM-TQτ estimator*.

In Theorem 2 of the Supplementary Material we show that, under general conditions, if $\bar{\boldsymbol{\beta}}$ is $n^{1/2}$ consistent, that is if

$$n^{1/2}(\bar{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = O_p(1), \tag{37}$$

then $\tilde{\boldsymbol{\theta}}$ satisfies

$$n^{1/2}(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = O_p(1). \tag{38}$$

The result (37) remains still a conjecture, however Salibian-Barrera and Yohai (2008) provide compelling arguments in favor of this conjecture (see in particular their Theorem 6). Besides, their Monte Carlo study seems to confirm this conjecture.

4.2. The final regression estimator

Let $\tilde{\mu}(\boldsymbol{x}) = \tilde{\mu} + \boldsymbol{x}^\top \bar{\boldsymbol{\beta}}$. Then, the standardized residuals with respect to the initial estimator are $\tilde{r}_i^* = (y_i^* - \tilde{\mu}(\boldsymbol{x}_i))/\tilde{\sigma}$ and can be used to obtain the cutoff point ϑ^* and the weights

$$w(\boldsymbol{x}, y, \tilde{\boldsymbol{\gamma}}, \vartheta) = w_0\left(\frac{y - \tilde{\mu}(\boldsymbol{x})}{\tilde{\sigma}}, \tilde{\lambda}, \vartheta\right),$$

where w_0 is given by (29). The *TML regression estimator* $\hat{\boldsymbol{\gamma}}$ is the solution of the equations

$$\boldsymbol{g}(\boldsymbol{\gamma}, \tilde{\boldsymbol{\gamma}}, \vartheta^*) = \boldsymbol{h}_{\tilde{\lambda}}, \tag{39}$$

where

$$\boldsymbol{g}(\boldsymbol{\gamma}, \tilde{\boldsymbol{\gamma}}, \vartheta^*) = E_{n,\tilde{\boldsymbol{\gamma}}}\left[w(\boldsymbol{x}, y, \tilde{\boldsymbol{\gamma}}, \vartheta^*)\boldsymbol{d}(\boldsymbol{x}, y, \boldsymbol{\gamma})\right],$$

and $\boldsymbol{h}_{\tilde{\lambda}} = (h_{1,\tilde{\lambda}}, h_{2,\tilde{\lambda}}, h_{3,\tilde{\lambda}}, \boldsymbol{h}_{4,\tilde{\lambda}}^\top)^\top$, where $h_{1,\tilde{\lambda}}, h_{2,\tilde{\lambda}}, h_{3,\tilde{\lambda}}$ are defined in (33) and

$$\boldsymbol{h}_{4,\tilde{\lambda}} = \frac{1}{n}\sum_{i=1}^{n} E_{\tilde{\lambda}}\left[w_0\left(u, \tilde{\lambda}, \vartheta^*\right)\xi_{\tilde{\lambda}}(u)\boldsymbol{x}_i\right].$$

The *1TML regression estimator* is obtained by applying one Newton-Raphson iteration to equations (39), i.e.,

$$\hat{\boldsymbol{\gamma}}_1 = \tilde{\boldsymbol{\gamma}} - \boldsymbol{J}(\tilde{\boldsymbol{\gamma}}, \vartheta^*)^{-1}(\boldsymbol{g}(\tilde{\boldsymbol{\gamma}}, \tilde{\boldsymbol{\gamma}}, \vartheta^*) - \boldsymbol{h}_{\tilde{\lambda}}),$$

where $\boldsymbol{J}(\tilde{\boldsymbol{\gamma}}, \vartheta^*) = \nabla_{\boldsymbol{\gamma}} \boldsymbol{g}(\boldsymbol{\gamma}, \tilde{\boldsymbol{\gamma}}, \vartheta^*)|_{\boldsymbol{\gamma}=\tilde{\boldsymbol{\gamma}}}$ is a Jacobian matrix. The *2TML regression estimator* is defined in the obvious way.

In Theorem 3 of the Supplementary Material we prove that, under general conditions including $n^{1/2}(\tilde{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0) = O_p(1)$, the 1TML regression estimator satisfies the following asymptotic result:

$$\sqrt{n}(\hat{\boldsymbol{\gamma}}_1 - \boldsymbol{\gamma}_0) \xrightarrow{\mathcal{L}} N(\mathbf{0}, \boldsymbol{G}(\boldsymbol{\gamma}_0)^{-1}\boldsymbol{M}(\boldsymbol{\gamma}_0)\boldsymbol{G}(\boldsymbol{\gamma}_0)^{-\top}),$$

where $\boldsymbol{M}(\boldsymbol{\gamma}) = E\left(\boldsymbol{v}(\boldsymbol{x}, y, \delta, \boldsymbol{\gamma})\boldsymbol{v}(\boldsymbol{x}, y, \delta, \boldsymbol{\gamma})^\top\right)$ and $\boldsymbol{G}(\boldsymbol{\gamma}) = E\left(\nabla_{\boldsymbol{\gamma}} \boldsymbol{v}(\boldsymbol{x}, y, \delta, \boldsymbol{\gamma})\right)$. The same result can be obtained for the 2TML regression estimator.

5. Monte Carlo experiments

5.1. The case without covariables

A first set of experiments was run in the case without covariables. In general, the results are similar to those described in Agostinelli et al. (2014) for the non censored case and we report here only the most representative cases. We compared the following estimators: ML, TQτ (with 10% trimming), 1TML, 2TML, and the one step WL estimator (1SWL) as defined in Section SM–3-1 of the Supplementary Material.

The TQτ estimator was defined using ρ_1 and ρ_2 in the Tukey's bi-weight family (20) with c equal to 1.548 and 6.08 respectively, and $b = 0.5$. The values 1.548 and $b = 0.5$ have been chosen so that the regression estimator based on the τ-scale has a finite sample breakdown point equal to 0.5. The value 6.08 makes the asymptotic efficiency equal to 0.95 in the case of normal errors. To compute the 1SWL estimator, we used a normal kernel with bandwidth $h = 0.3\tilde{\sigma}$ in all experiments, where the scale estimator $\tilde{\sigma}$ is the scale initial estimate.

To compare the global performances of the different estimators we use the total variation distance (TVD)

$$\mathrm{TVD}(\boldsymbol{\theta}) = \frac{1}{2}\int |f_{\boldsymbol{\theta}}(y) - f_{\boldsymbol{\theta}_0}(y)|\, dy$$

between a given density $f_{\boldsymbol{\theta}}$ and the true underlying density $f_{\boldsymbol{\theta}_0}$. The performance of the estimator $\hat{\boldsymbol{\theta}}$ is measured by the mean value of $\mathrm{TVD}(\hat{\boldsymbol{\theta}})$:

$$\mathrm{MTVD}(\hat{\boldsymbol{\theta}}) = E(\mathrm{TVD}(\hat{\boldsymbol{\theta}})).$$

MTVD($\hat{\boldsymbol{\theta}}$) clearly measures the quality of the estimated density. It is estimated, using the simulated values $\hat{\boldsymbol{\theta}}_k$ $(1 \leq k \leq N)$ of the estimator, by the average TVD:

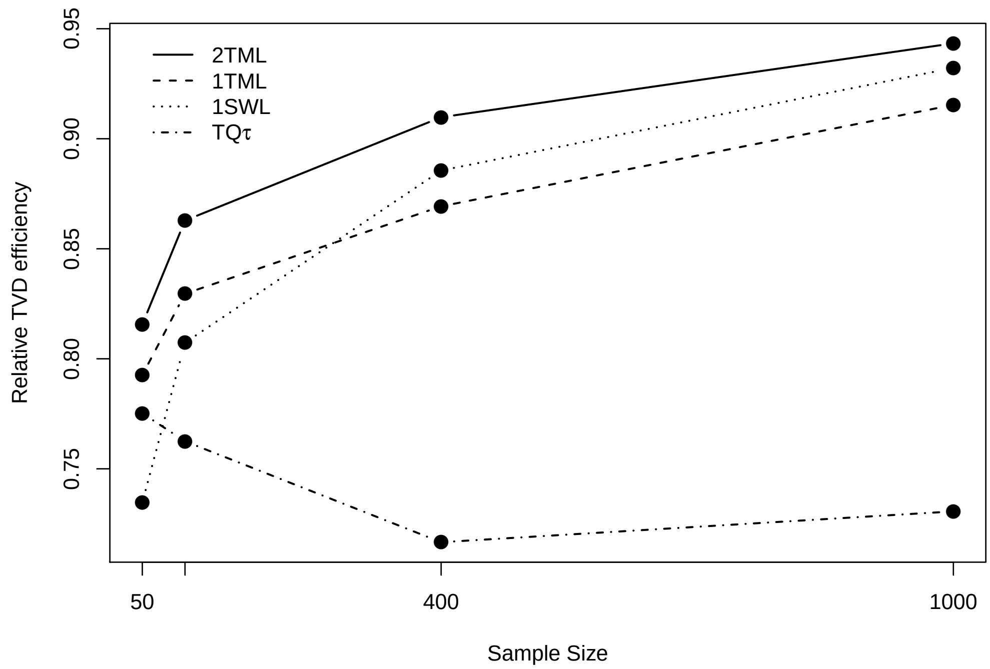
$$\text{ATVD}(\hat{\boldsymbol{\theta}}) = \frac{1}{N} \sum_{k=1}^{N} \text{TVD}(\hat{\boldsymbol{\theta}}_k).$$

5.1.1. Simulation under the nominal model

We studied the efficiency of the estimators under the nominal model for $n = 50, 100, 400, 1000$ and $\lambda_0 = 1$. Without loss of generality, we took $\mu_0 = 0$ and $\sigma_0 = 1$. We considered two censoring proportions: 15% and 50%. Here, we only report the results for the 15% proportion; the results for the 50% proportion show a similar behavior an can be found in the Supplementary Material. The number of replications was 1200. Figure 1 (top) reports the ATVD of the robust estimators divided by the ATVD of the ML estimator as a function of the sample size for $\lambda_0 = 1$. This ratio can be interpreted as a measure of relative efficiency that we call *TVD efficiency*. As expected, the TVD efficiency of 1TML, 2TML and 1SWML is markedly larger than the efficiency of the initial estimator as the sample size grows. Moreover, when n increases the TVD efficiency becomes close to 1, that is, it becomes close to the efficiency of the ML estimator. Similar patterns are observed in the other simulated cases.

5.1.2. Simulation under point mass contamination

In a second Monte Carlo experiment, we compared TQτ, 1TML, 2TML and 1SWL under point mass contamination for $n = 50, 100, 400, 1000$, $\lambda_0 = 1$, $\mu_0 = 0$, $\sigma_0 = 1$, and two censoring proportions: 15% and 50%. (Only the results for the 15% fraction are reported here.) We generated 90% "good" observations y_j according to the GLG model and 10% "outliers" at the point y_0. We then varied the value of y_0 from -10 to 10 with a step of 0.5. This kind of point mass contamination is generally the least favorable one and allows an evaluation of the maximal bias an estimator can incur. For each value of y_0, the number of replications was 1200. Figure 1 (bottom) reports the ATVD of the estimated densities as a function of y_0. The results show that the ATVD of TQτ, 1TML and 2TML are comparable and very stable over the whole range of y_0 values, while 1SWL provides a very high ATVD for positive values of y_0. The ATVD of ML is not shown because it becomes very high and does not fit in the frame.



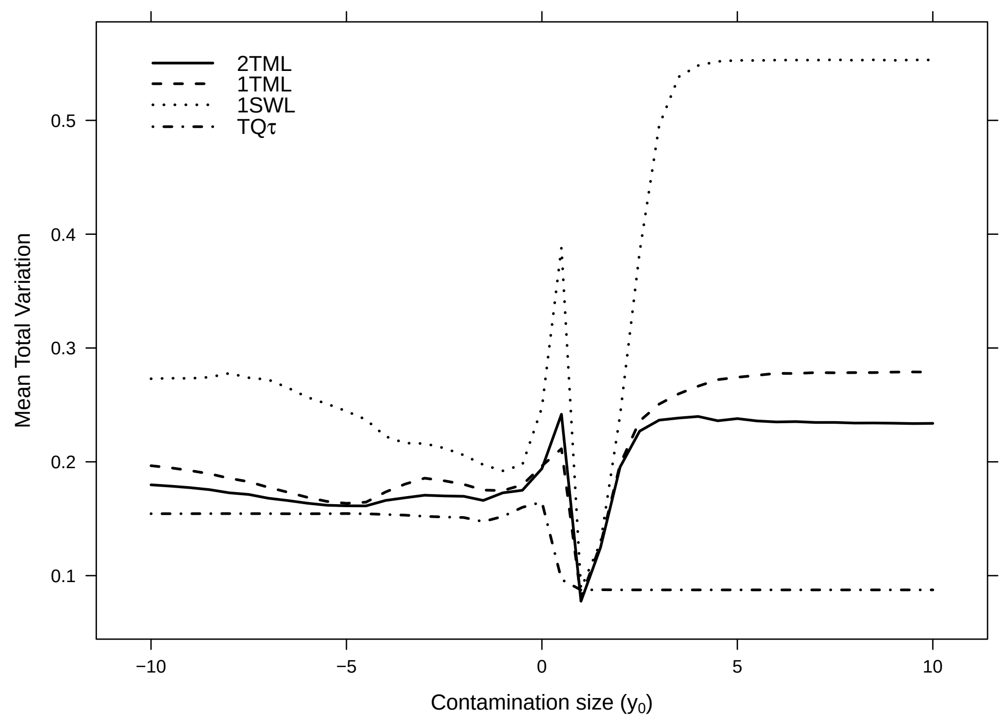


Figure 1: The case without covariables. Top: Estimated relative TVD efficiency versus sample size. Bottom: Average TVD under 10% contamination, 15% censoring proportion, and sample size $n = 100$.

5.2. The case with covariables

A second set of experiments was run to investigate the behavior of the MM-TQτ estimator defined in Section 4.1, the 1TML and 2TML estimators defined in Section 4.2, and the one step weighted likelihood (1SWL) estimator defined in Section SM–3-2 of the Supplementary Material.

In each experiment, n pairs of observations $(\boldsymbol{x}_i, y_i^*)$ were generated according to the regression model

$$
\begin{aligned}
y_i &= \mu_0 + \boldsymbol{\beta}_0^\top \boldsymbol{x}_i + \sigma_0 u_i, && (40)\\
y_i^* &= \min(y_i, c_i), && (41)\\
c_i &= \mu_c + e_i, && (42)\\
\delta_i &= 1 \text{ if } y_i^* = y_i \text{ and } 0 \text{ otherwise}, && (43)
\end{aligned}
$$

with $\mu_0 = 0$, $\sigma_0 = 1$, $\lambda_0 = 1$, $\boldsymbol{\beta}_0^\top = (2, 3)$, $\boldsymbol{x}_i = (x_{i1}, x_{i2})$, where $x_{i1} \sim N(0,1)$, $x_{i2} \sim \mathcal{B}(1, 0.4)$ with x_{i1} and x_{i2} independent and $e_i \sim GLG(\mu_c, 1, 1)$. The parameter μ_c was chosen so that the censoring proportion was 15% and 50%. Only the results for the 15% proportion are reported here; the results for the 50% proportion show a similar behavior and can be found in the Supplementary Material.

To evaluate the performance of an estimator $(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\beta}})$ we defined the total variation distance

$$\mathrm{TVD}(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\beta}}, \boldsymbol{x}) = \frac{1}{2} \int \left| f_{(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\beta}})}(y|\boldsymbol{x}) - f_{(\boldsymbol{\theta}_0, \boldsymbol{\beta}_0)}(y|\boldsymbol{x}) \right| dy,$$

where $f_{(\boldsymbol{\theta},\boldsymbol{\beta})}(y|\boldsymbol{x}) = f_\lambda\left((y - \mu - \boldsymbol{\beta}^\top \boldsymbol{x})/\sigma\right)$. The mean value

$$\mathrm{MTVD}(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\beta}}) = E[\mathrm{TVD}(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\beta}}, \boldsymbol{x})],$$

was estimated using the simulated values $(\hat{\boldsymbol{\theta}}_k, \hat{\boldsymbol{\beta}}_k)$ and $\boldsymbol{x}_{k1}^*, \ldots, \boldsymbol{x}_{ki}^*, \ldots, \boldsymbol{x}_{kn}^*$ $(1 \leq k \leq N)$ by

$$\mathrm{ATVD}(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\beta}}) = \frac{1}{Nn} \sum_{k=1}^{N} \sum_{i=1}^{n} \mathrm{TVD}(\hat{\boldsymbol{\theta}}_k, \hat{\boldsymbol{\beta}}_k, \boldsymbol{x}_{ki}^*).$$

5.2.1. Simulation under the nominal model

We studied the efficiency of the estimators under the nominal model for $n = 50$, 100, 400 and 1000. Figure 2 (top) shows the relative TVD efficiency of MM-TQτ, 1TML, 2TML and 1SWL with respect to ML for $\lambda_0 = 1$ and censoring proportion 15%.

The efficiency of 1TML, 2TML and 1SWL is clearly higher than the efficiency of MM-TQτ. This is an expected result, since 1TML, 2TML and 1SWL are asymptotically fully efficient under the nominal model.

5.2.2. Simulation under point mass contamination

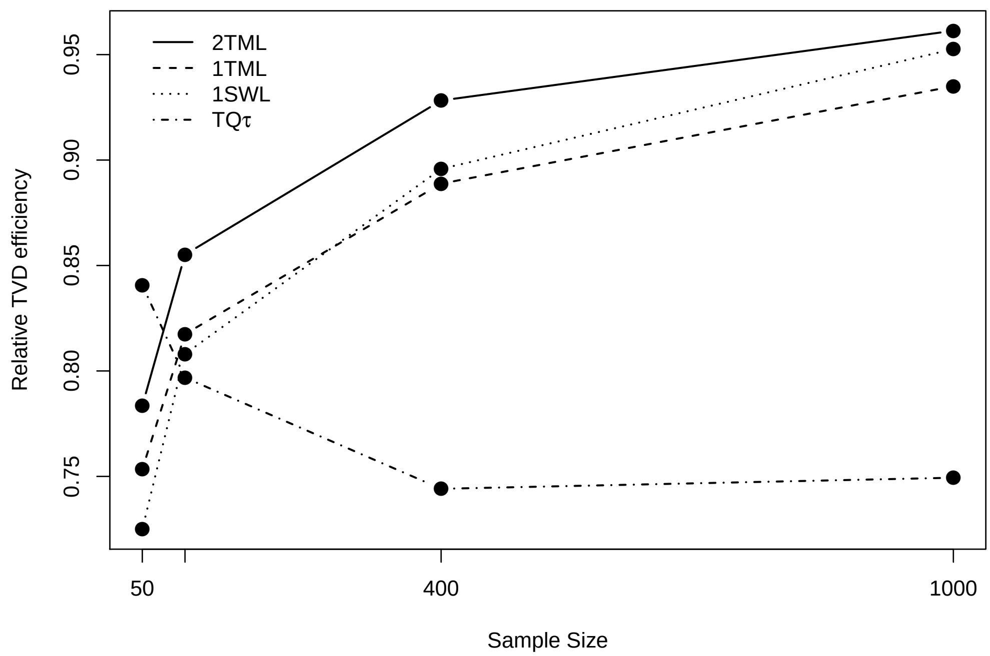
We also studied the behavior of the estimators under point mass contamination for n = 50, 100, 400, 1000. The values of the parameters were the same as in the case of no contamination. We generated n "good" observations $(\boldsymbol{x}_i, y_i^*)$ according to (40)-(43). We then replaced 10% values y_i^* with a value y_0 ranging from -10 to 20. For each value of y_0 the number of replications was 1200. Figure 2 (bottom) shows the ATVD of the estimators as a function of y_0 for sample size $n = 100$, contamination level 10%, and censoring proportion 15%. We observe that MM-TQτ, 1TML, and 2TML are very resistant under point mass contamination, while the 1SWL is highly sensitive to outliers on the right tail of the distribution.

5.2.3. Empirical finite sample breakdown point

We were not able to obtain the theoretical breakdown point of the proposed estimators. To fill this gap, we performed a Monte Carlo simulation to explore the behavior of the maximum mean square error as a function of the contamination level. This provides information about the highest contamination the proposed estimators can cope with and hence about the finite sample breakdown point. We used the same setting as in the previous subsection, $n = 1000$, $\lambda_0 = 1$, and censoring proportion 15%. Several values of the contamination level ϵ in the interval $[0, 0.3]$ and several values y_0 in the interval $[-100, 100]$ were considered. For each pair (ϵ, y_0), we run 100 Monte Carlo replications and computed the maximum MSE (MMSE) of the regression parameters (μ, β) – note that μ is the intercept–, the scale parameter (σ) and the shape parameter (λ). Results for the regression parameters (slopes) are reported in Figure 3; they show that the MMSE starts to increase rapidly around the 20% level. This behavior is consistent with the MMSE of the initial regression parameters provided by the MM non parametric estimator.

6. Illustrations with real data

In modern hospital management, stays are classified into "diagnosis related groups" (DRGs; Fetter et al. (1980)) which are designed to be as ho-



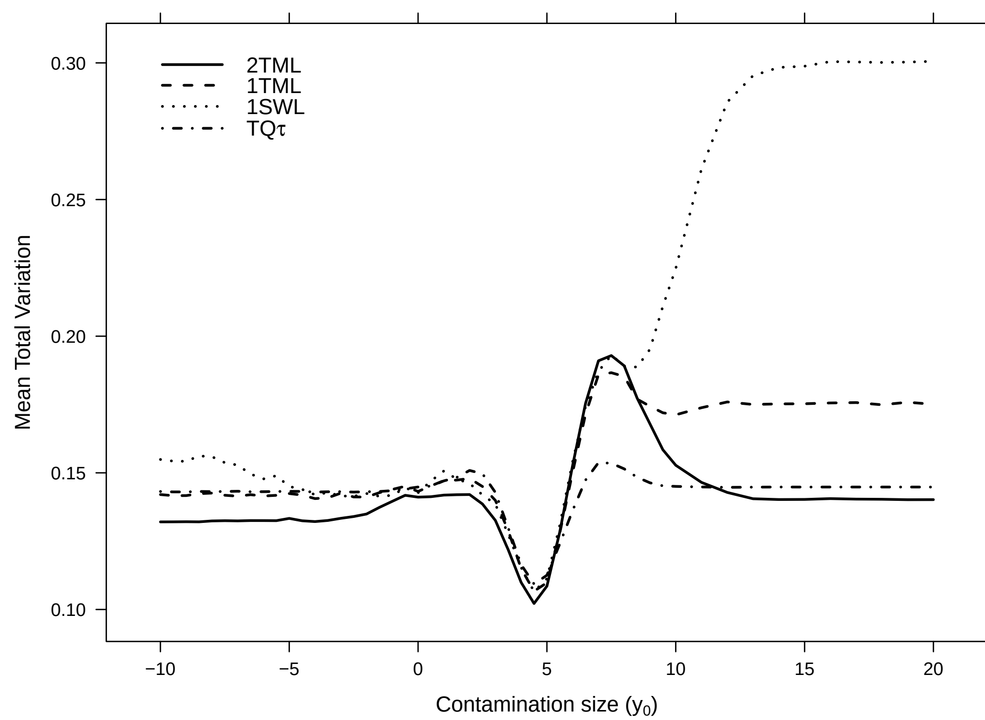


Figure 2: The case with covariables. Top: Estimated relative TVD efficiency versus sample size. Bottom: Average TVD under 10% contamination and sample size $n = 100$.

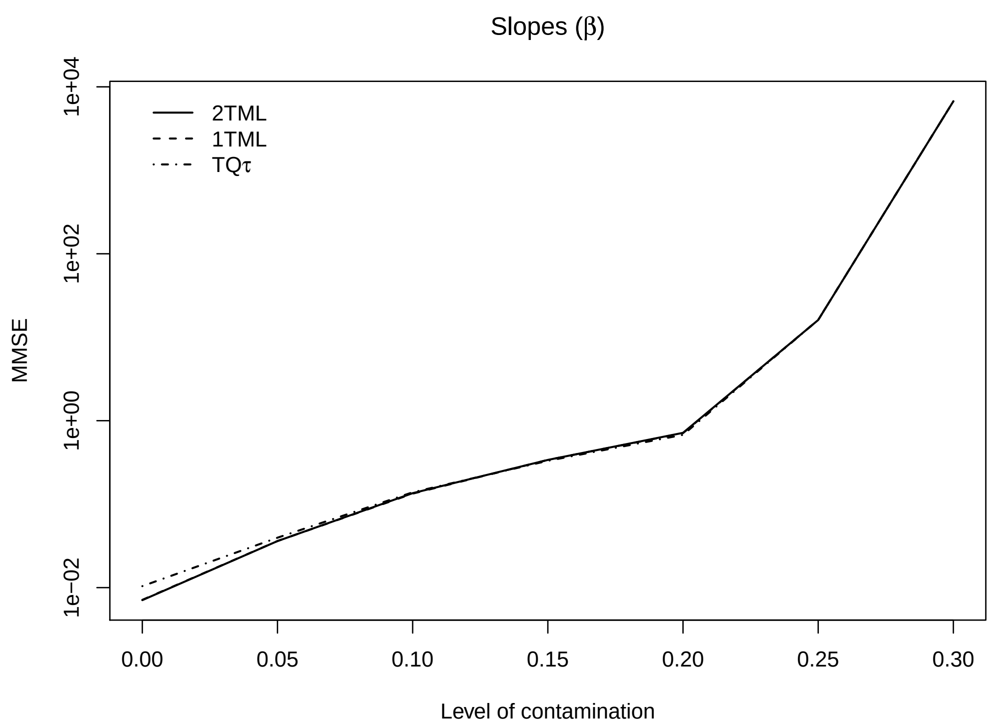


Figure 3: Estimated maximum MSE of the regression parameters (slopes) for MM-TQ$_\tau$ (dashed-dotted line), 1TML (dashed line), and 2TML (solid line). The censoring proportion is 15% and $n = 1000$. The MMSEs of the 1TML and 2TML are almost always overlapping. The y-axis is on log scale.

mogeneous as possible with respect to diagnosis, treatment, and resource consumption. The "mean cost of stay" of each DRGs is periodically estimated with the help of administrative data on a national basis and used to determine "standard prices" for hospital funding and reimbursement. Since it is difficult to measure cost, "length of stay" (LOS) is often used as a proxy. In designing and "refining" the groups, the relationship between LOS and other variables which are usually available on administrative files has to be assessed and taken into account. We discuss two examples in this domain.

6.1. Major cardiovascular interventions

In a first example, we consider a sample of 75 stays in a Swiss hospital and DRG "Major cardiovascular interventions". Of these stays, 45 were censored because the patients were transferred to a different hospital before dismissal. The data – shown in Figure 4 and made available in Marazzi and Muralti (2013) – were first analyzed in Locatelli et al. (2010) and Locatelli and Marazzi (2013). These authors studied the relationship between LOS and two covariables: Age of the patient (x_1) and Admission type ($x_2 = 0$ for planned admissions, $x_2 = 1$ for emergency admissions) with the help of the model $y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \gamma x_1 x_2 + \sigma u$, where $y = \log(\text{LOS})$ and u is following a Gaussian model. They observed that two young patients had exceptionally high non censored LOS and, as a consequence, the ML estimator yielded an unexpected large estimate of the interaction γ. Therefore, they proposed the use of a robust parametric procedure called "weighted maximum likelihood" (WML/G) based on the Gaussian error model (that performed better than log-Weibull). They compared WML/G with other published robust procedures an found that the robust estimates of γ were close to zero.

Here, we assume a GLG error model and consider the ML, 1TML, 2TML, 1SWL and 2SWL regression estimates reported in Table 1. The ML estimates based on GLG were obtained with the help of the R function mle (ML/mle) and the procedure LIFEREG of SAS (ML/LR). However, these data caused problems to these programs: a warning message of LIFEREG put in doubt the convergence of the algorithm and the results are unreliable; mle could not compute three standard errors (error messages were given). For comparison, Table 1 also reports the ML estimate (ML/G) and the WML estimate (WML/G) based on the Gaussian error model given in Locatelli et al. (2010). Some of the regression lines are reported in Figure 4.

Table 1: Estimates of the regression model for length of stay of "Major cardiovascular interventions"

	μ	$10\beta_1$	$10\beta_2$	10γ	σ	λ
1TML	2.16 (0.11)	-1.84 (8.73)	0.08 (0.02)	0.09 (0.12)	0.45 (0.08)	-1.82 (0.14)
2TML	2.16 (0.11)	-1.84 (8.71)	0.08 (0.02)	0.09 (0.12)	0.45 (0.08)	-1.82 (0.14)
1SWL	2.16 (0.16)	-1.84 (8.10)	0.08 (0.03)	0.09 (0.12)	0.45 (0.07)	-1.82 (0.43)
2SWL	2.16 (0.16)	-1.84 (8.12)	0.08 (0.03)	0.09 (0.12)	0.45 (0.07)	-1.82 (0.43)
ML/mle	2.05 (0.16)	22.03 (2.37)	0.07 (0.03)	-0.27 (—)	0.45 (—)	-3.34 (—)
ML/LR	1.95 (1.10)	-4.92 (9.15)	0.05 (0.02)	0.13 (0.14)	0.30 (0.08)	-3.33 (1.05)
ML/G	2.93	23.42	0.11	-0.30	—	—
WML/G	2.44	10.57	0.11	-0.10	—	—

We first notice the good agreement among the robust estimates based on GLG. Not surprisingly, the main differences between the robust procedures based on GLG and those based on the Gaussian model concern the intercept terms (μ and β_1). However, the robust prediction lines based on GLG (Figure 4) seem to provide a better fit to the bulk of the data. This observation is supported by the plots in Figure 5, where three types of distributions of the standardized residuals are displayed: non-parametric KM, semi-parametric (normal and GLG), and parametric (normal and GLG). Note the very large steps of the KM functions corresponding to the two extreme non censored observations. The reason is that KM puts the mass of several censored residuals on these two points. The ML/G survival functions in Figure 4a of Locatelli et al. (2010) (not reported in Figure 5) were strongly affected by these two points. Both the WML/G and 2TML semi-parametric survival functions behave much better: with two exceptions, the robust residuals follow the models very well. However 2TML is better than WML/G in the left tail. Finally, we note that the use of GLG provides a reasonable fit for the two young patients with high non censored LOS: these observations were outliers for the Gaussian model but not for the GLG model. The censored observations corresponding to emergency admissions in the right bottom corner of Figure 4 are considered outliers.

6.2. Minor bladder interventions

In a second example, we consider a sample of 48 stays for DRG "Minor bladder interventions". The data are shown in Figure 6. Six patients were transferred to a different hospital before dismissal. Four young patients have surprisingly large values of LOS. We study the relationship between LOS and

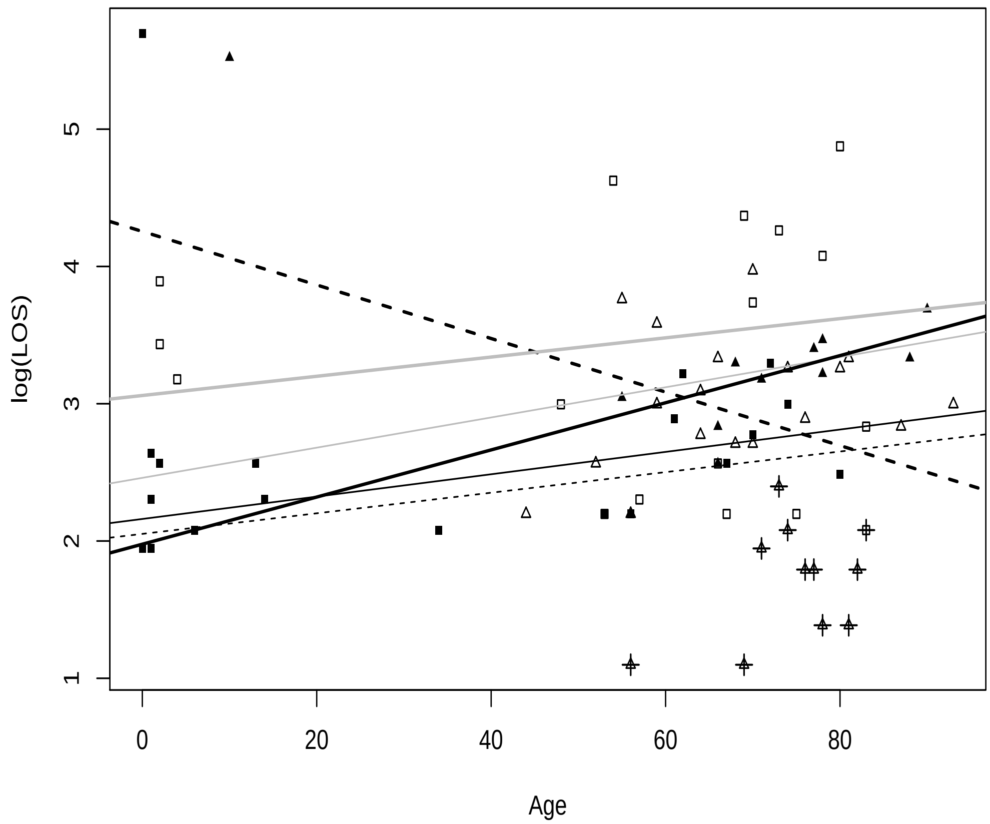


Figure 4: Estimated regression lines for log-LOS of “Major cardiovascular interventions”. Black lines are estimates based on GLG: 2TML is solid, ML/mle is dashed. Grey lines are estimates (WML/G) based on the Gaussian model. Thin lines refer to planned admissions, while thick lines refer to emergency admission. Filled marks represent complete observation; empty marks represent censored observation; square marks are planned admission; triangle are emergency admission. Outliers according to 2TML are marked with crosses.

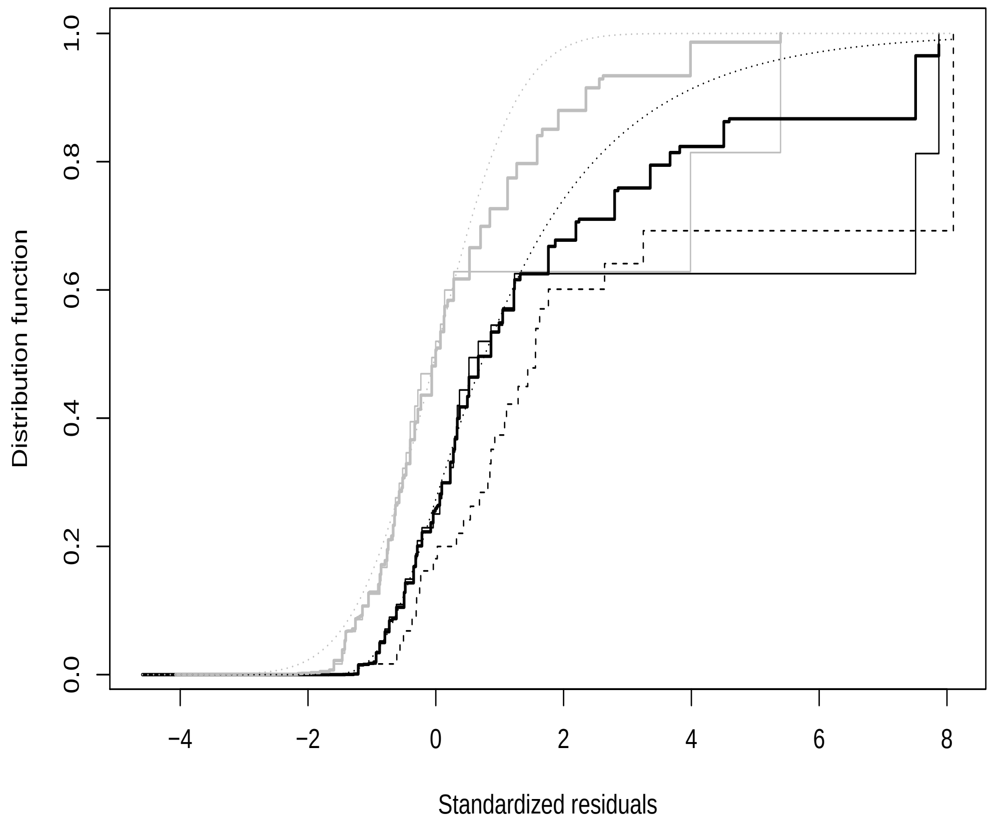


Figure 5: Standardized residuals cdfs based on 2TML, ML/mle (GLG), and WML/G for "Major cardiovascular interventions". Black lines are based on the GLG model. KM of ML residuals: thin-dashed line; KM of 2TML residuals: thin-solid line; semi-parametric cdf based on 2TML residuals: thick-solid line. Grey lines are based on on the Gaussian model. KM of WML/G residuals: thin-solid line; semi-parametric cdf based on WML/G residuals: thick-solid line. The smooth dotted lines are the parametric cdf.

Table 2: Estimated regression models for LOS of "Minor bladder interventions"

Method	μ	10β	σ	λ
1TML	0.992 (0.365)	0.127 (0.056)	0.480 (0.050)	0.489 (0.170)
2TML	0.992 (0.365)	0.127 (0.056)	0.482 (0.047)	0.489 (0.172)
1SWL	0.992 (0.600)	0.130 (0.080)	0.481 (0.063)	0.490 (0.120)
2SWL	0.992 (0.606)	0.131 (0.081)	0.484 (0.064)	0.489 (0.120)
ML/mle	1.597 (0.565)	-0.001 (0.070)	0.677 (0.099)	-1.005 (0.483)
ML/LR	1.566 (0.556)	0.003 (0.071)	0.675 (0.097)	-1.019 (0.444)
ML*/mle	0.530 (0.437)	0.168 (0.063)	0.525 (0.061)	-0.081 (0.422)
ML*/LR	0.528 (0.438)	0.169 (0.063)	0.525 (0.061)	-0.81 (0.420)

Age considering the model $y = \mu + \beta x + \sigma u$, where $x =$Age and $y = \log(\text{LOS})$. As in the first example, ML was computed with the help of the R function mle and the SAS procedure LIFEREG. The ML/mle, ML/LR, 1TML, 2TML, 1SWL, and 2SWL estimates and the estimated standard errors are reported in Table 2. For this dataset, ML/mle and ML/LR coincide. Some of the prediction lines are drawn in Figure 6.

According to ML, log(LOS) does not seem to depend on Age (the p-value is 0.04) and the inverse log-Weibull distribution ($\lambda = -1$) is the adequate error model. The robust estimates are as alike as four peas in a pod and provide a much larger slope (p-value = 0.02). Moreover – as it is expected from these data – they suggest a positive linear relationship and a positively asymmetric error model (the p-value for λ is 0.0024). Clearly, the outliers (those with weights equal to zero in 2TML are marked with crosses in Figure 6) have an important leverage effect on the ML coefficients and shape parameter. Removing the outliers, the ML coefficients (ML*/mle and ML*/LR in Table 2) become similar to the robust ones. In practice, this simple analysis suggests that the possibility of splitting this particular DRG into two groups should be further investigated.

7. Discussion

As mentioned in the introduction, the GLG model is a very flexible family of distributions which is used to describe asymmetrically distributed data in many real applications. In this paper, we considered estimators which are simultaneously robust and efficient for AFT models, when the errors follow a GLG distribution and the data may contain censored observations. The

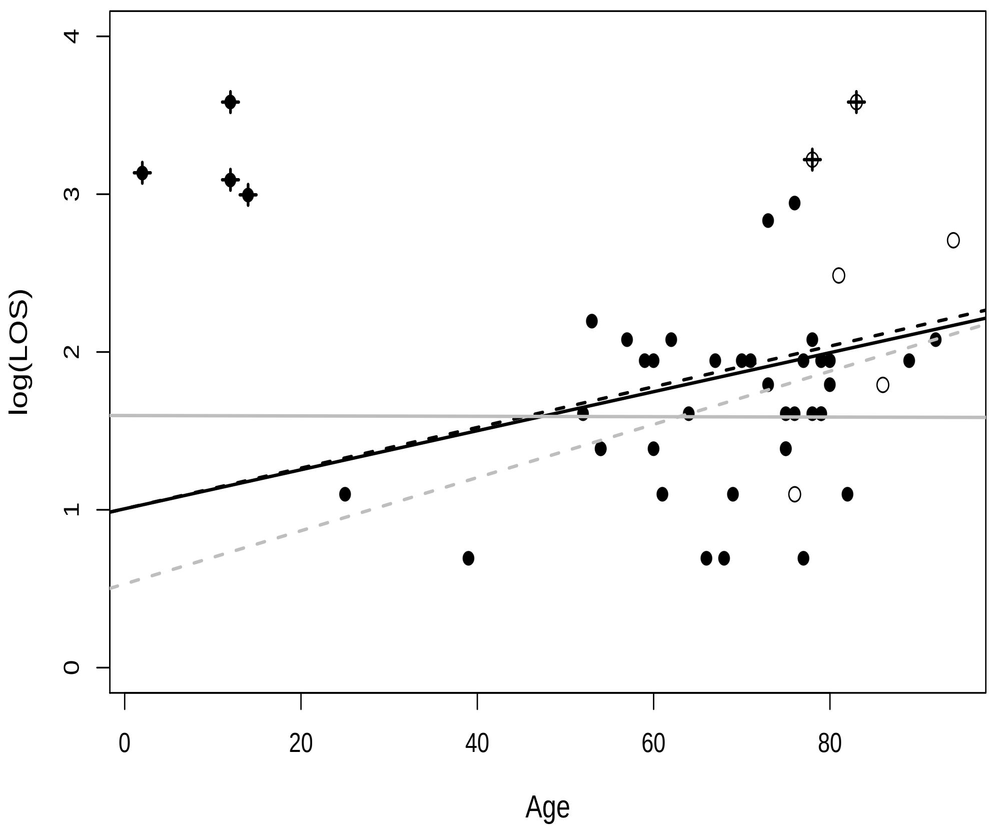


Figure 6: Estimated regression models for LOS of "Minor bladder interventions". 2TML is black and solid; 2SWL is black and dashed; ML is grey and solid; ML without outliers is gray and dashed. Filled marks represent complete observation; empty marks represent censored observation. Outliers according to 2TML are marked with crosses.

estimation procedures have three main components: an initial highly robust but not necessarily efficient estimator; the identification and removal of the outliers (according to an adaptive procedure proposed in Marazzi and Yohai (2004)); a final efficient estimator based on the inliers.

We first considered the case, where no covariables are present and, in this case, we proposed an initial estimator, called TQτ estimator, that minimizes a τ scale of the differences between theoretical an empirical quantiles of order smaller than $(1-\alpha)$, where $0 < \alpha < 1$ is a trimming fraction. The final estimator is a one step weighted likelihood estimator, where the weights penalizing the outliers are derived from the initial estimator.

For the case, where covariables are present, the proposed estimators were derived as follows. In the first step we used a regression MM-estimator as proposed in Salibian-Barrera and Yohai (2008) to obtain initial slope estimates and to compute the corresponding residuals. We then computed an initial estimator of the GLG parameters by applying the procedure for the no covariables case to these residuals. For the final estimator of all the parameters we used a one step truncated ML which removes the outliers and starts with the initial estimator.

We provided asymptotic results and extensive Monte Carlo results showing that the final estimators are highly efficient and maintain the same robustness level as the initial ones.

A possible extension of the proposed methods concerns the case where the sample contains interval censored data, that is, when for some observations y_i, it is only known that $a_i \leq y_i \leq b_i$. In the absence of covariables, the TQτ estimator can be extended by replacing the Kaplan-Meier estimator by another estimator, which is consistent for the interval censoring case. An example is the estimator proposed by Turnbull (1976). For the regression case, the following changes in the definition of the initial MM-estimator would be required: (i) replace the Kaplan-Meier estimator of the residual distribution by another estimator which is consistent for the case of interval censoring; (ii) compute the conditional expected values of the censored observations under the condition that they belong to the censored intervals. Once the MM residuals are computed, the error distribution should also be estimated using a consistent estimator for interval censoring. Double censoring can also be accommodated in a similar way by replacing Kaplan Meier with the double censored version of Turnbull (1974).

A different problem is the case where there is a given cut point for the response. Tobin (1958) introduced a class of regression models where errors have normal distribution and the responses are left censored by 0. Several authors extended this model using different error distributions. For example Martínez-Flórez et al. (2013) replace the normal distribution by an alpha-power family of distributions with density function $\phi_F(z,\alpha) = \alpha f(z)F(z)^{\alpha-1}$, where $\alpha > 0$ and $F(z)$ is a known distribution function with density $f(z)$. It seems difficult to extend the TQτ estimator to these models. The reason is that there is not, in this case, a non-parametric procedure to estimate the error distribution. However, robust estimators can be defined using a totally parametric approach as in Locatelli et al. (2010).

Completing the details of these extensions requires however much further work and may be matter of future research.

8. Supplementary Material

Supplementary Material is available. Section SM–1 provides proofs for consistency and $n^{1/2}$-consistency of the TQτ estimator; Section SM–2 provides a proof of the asymptotic normality of the 1TML estimator; Section SM–3 extends the one step Weighted Likelihood estimator of Agostinelli et al. (2014) to the regression case with censored observations; Section SM–4 reports results of supplementary Monte Carlo experiments.

The procedures proposed in this paper are implemented in the R (R Core Team, 2015) package robustloggamma (Agostinelli et al., 2016), version 0.5 or higher, available at CRAN.

Acknowledgements

This research was partially supported by the Italian-Argentinian project "Metodi robusti per la previsione del costo e della durata della degenza ospedaliera" funded by the collaboration program MINCYT-MAE AR14MO6. Víctor Yohai's research was also partially supported by Grants 20020130100279 from Universidad of Buenos Aires, PIP 112-2008-01-00216 and 112-2011-01-00339 from CONICET and PICT 2011-0397 from ANPCYT. We also thank two anonymous referees for their comments and suggestions which contributed to important improvements of the paper. All statistical analysis were performed on SCSCF (www.dais.unive.it/scscf), a multiprocessor cluster system owned by Ca' Foscari University of Venice running under GNU/Linux.

References

Abadi, A., Amanpour, F., Bajdik, C., Yavari, P., 2012. Breast cancer survival analysis: Applying the generalized gamma distribution under different conditions of the proportional hazards and accelerated failure time assumptions. International Journal of Preventive Medicine 3, 644–651.

Agostinelli, C., Marazzi, A., Yohai, V., 2014. Robust estimators of the generalized log-gamma distribution. Technometrics 56 (1), 92–101.

Agostinelli, C., Marazzi, A., Yohai, V., Randriamiharisoa, A., 2016. Robust estimation of the generalized loggamma model. the r package robustloggamma. Journal of Statistical SoftwareAccepted.

Cox, C., Chu, H., Schneider, M., Muñoz, A., 2007. Parametric survival analysis and taxonomy of hazard functions for the generalized gamma distribution. Statistics in Medicine 26, 4352–4374.

Fetter, R., Shin, Y., Freeman, J., Averill, R., Thompson, J., 1980. Case mix definition by diagnosis-related groups. Medical care 18 (1), 1–53.

Gervini, D., Yohai, V., 2002. A class of robust and fully efficient regression estimators. The Annals of Statistics 30 (2), 583–616.

Greenwood, M., 1926. The natural duration of cancer. Her Majesty's Stationery Office, London.

Huber, P., 1981. Robust statistics. Wiley, New York.

Johnson, L., Kotz, S., Balakrishnan, N., 1994. Continuous univariate distributions. Volume 1, 2nd Edition. Wiley, New York.

Kaplan, E., Meier, P., 1958. Nonparametric estimation from incomplete observations. Journal of the American Statistical Association 53 (282), 457–481.

Kim, M., DeGruttola, V., Lagakos, S., 1993. Analyzing doubly censored data with covariates, with application to aids. Biometrics 49, 13–22.

Lawless, J., 1980. Inference in the generalized gamma and log-gamma distributions. Technometrics 22 (3), 409–419.

Lawless, J., 2003. Statistical Models and Methods for Lifetime Data, 2nd Edition. John Wiley and Sons, Hoboken.

Locatelli, I., Marazzi, A., 2013. Robust parametric indirect estimates of the expected cost of a hospital stay with covariates and censored data. Statistics in Medicine 32 (14), 2457–2466.

Locatelli, I., Marazzi, A., Yohai, V., 2010. Robust accelerated failure time regression. Computational Statistics & Data Analysis 55, 874–887.

Marazzi, A., Muralti, J., 2013. RobustAFT: a package for truncated maximum likelihood fit and robust accelerated failure time regression for gaussian and log-weibull case.
URL `cran.r-project.org/web/packages/RobustAFT/index.html`

Marazzi, A., Yohai, V., 2004. Adaptively truncated maximum likelihood regression with asymmetric errors. Journal of Statistical Planning and Inference 122 (1-2), 271–291.

Martínez-Flórez, G., Bolfarine, H., Gómez, H. W., 2013. Asymmetric regression models with limited responses with an application to antibody response to vaccine. Biometrical Journal 55, 156–172.

Ortega, E., Bolfarine, H., Paula, G., 2003. Influence diagnostics in generalized log-gamma regression models. Computational Statistics & Data Analysis 42, 165–186.

Ortega, E., Paula, G., Bolfarine, H., 2008. Deviance residuals in generalized log-gamma regression models with censored observations. Journal of Statistical Computation and Simulation 78 (8), 747–764.

Prentice, R., 1974. A log-gamma model and its maximum likelihood estimation. Biometrika 61 (3), 539–544.

R Core Team, 2015. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
URL `https://www.R-project.org/`

Salibian-Barrera, M., Yohai, V., 2008. High breakdown point robust regression with censored data. The Annals of Statistics 36, 118–146.

Serfling, R., 1980. Approximation Theorems of Mathematical Statistics. Wiley.

Silva, G., Ortega, E., Cancho, V., 2010. Log-weibull extended regression model: Estimation, sensitivity and residual analysis. Statistical Methodology 7, 614–631.

Sun, J., Liao, Q., Pagano, M., 1999. Regression analysis of doubly interval censored failure time data with applications to aids. Biometrics 55, 909–914.

Tobin, J., 1958. Estimation of relationships for limited dependent variables. Econometrica, 26 (1), 24–36.

Turnbull, B., 1974. Nonparametric estimation of a survival ship function with doubly censored data. Journal of the American Statistical Association 69, 169–173.

Turnbull, B., 1976. The empirical distribution function with arbitrarily grouped, censored, and truncated data. Journal of the Royal Statistical Society, Series B 38, 290–295.

Yohai, V., Zamar, R., 1988. High breakdown estimates of regression by means of the minimization of an efficient scale. Journal of the American Statistical Association 83, 406–413.