# Quantile Least Squares: A Flexible Approach for Robust Estimation and Validation of Location-Scale Families

Mohammed Adjieteh<sup>1</sup>
University of Wisconsin-Milwaukee

Vytaras Brazauskas<sup>2</sup>
University of Wisconsin-Milwaukee

December 15, 2023

In this paper, the problem of robust estimation and validation of locationscale families is revisited. The proposed methods exploit the joint asymptotic normality of sample quantiles (of i.i.d. random variables) to construct the ordinary and generalized least squares estimators of location and scale parameters. These quantile least squares (QLS) estimators are easy to compute because they have explicit expressions, their robustness is achieved by excluding extreme quantiles from the least-squares estimation, and efficiency is boosted by using as many non-extreme quantiles as practically relevant. The influence functions of the QLS estimators are specified and plotted for several location-scale families. They closely resemble the shapes of some well-known influence functions yet those shapes emerge automatically (i.e., do not need to be specified). The joint asymptotic normality of the proposed estimators is established and their finite-sample properties are explored using simulations. Also, computational costs of these estimators, as well as those of MLE, are evaluated for sample sizes  $n = 10^6, 10^7, 10^8, 10^9$ . For model validation, two goodness-of-fit tests are constructed and their performance is studied using simulations and real data. In particular, for the daily stock returns of Google over the last four years, both tests strongly support the logistic distribution assumption and reject other bell-shaped competitors.

Keywords. Goodness-of-Fit; Least Squares; Quantiles; Relative Efficiency; Robustness.

<sup>&</sup>lt;sup>1</sup> Mohammed Adjieteh, ASA, is a Ph.D. candidate and Graduate Teaching Assistant in the Department of Mathematical Sciences, University of Wisconsin-Milwaukee, P.O. Box 413, Milwaukee, WI 53201, USA. *e-mail*: adjieteh@uwm.edu

<sup>&</sup>lt;sup>2</sup> CORRESPONDING AUTHOR: Vytaras Brazauskas, Ph.D., ASA, is a Professor in the Department of Mathematical Sciences, University of Wisconsin-Milwaukee, P.O. Box 413, Milwaukee, WI 53201, USA. *e-mail*: vytaras@uwm.edu

#### 1 Introduction

The problem of robust estimation of location-scale families can be traced back to the seminal works of Tukey (1960), Huber (1964), and Hampel (1968). Since then, numerous robust methods for this problem have been proposed in the literature; they are summarized in the books of Hampel et al. (1986), Maronna et al. (2006), and Huber and Ronchetti (2009). While at first it might seem like the topic is exhausted and fully "solved", in this paper we argue that it is worthwhile revisiting it. In particular, connections with best linear unbiased estimators or BLUE, based on strategically selected order statistics, can be exploited and studied from various theoretical and practical perspectives: robustness, efficiency, model validation (goodness of fit), and computational cost. All of this within the same framework.

The literature on BLUE methods for location-scale families, which are constructed out of a few order statistics, goes back to Mosteller (1946). Since then, numerous papers on parameter estimation, hypothesis testing, optimal spacings, simulations, and applications have been published. A very short list of contributions to this area includes: a first comprehensive review of estimation problems by Sarhan and Greenberg (1962) (and many specialized papers by at least one of these authors); estimation of parameters of the Cauchy distribution by Chan (1970) (and multiple related papers by the same author and his co-authors) and Cane (1974); a relatively recent review of this literature by Ali and Umbach (1998) (and numerous technical papers by at least one of these authors). A common theme in many of the papers in this field is to show that for various distributions, highly-efficient estimators can be constructed using less than ten order statistics. Also, when the number of order statistics is fixed, then the optimal spacings, according to the asymptotic relative efficiency criterion, can be determined. Computational ease and robustness of such estimators are often mentioned, but to the best of our knowledge no formal studies of robustness that specify breakdown points and influence functions have been pursued. Interestingly, those optimal (most efficient) estimators usually include order statistics that are very close to the extreme levels of 0 or 1, making the estimators practically nonrobust.

Xu et al. (2014), focusing on the g-and-h distributional family, did study the breakdown points and influence functions of robust estimators that they derived using the criterion of quantile least squares (QLS). In this paper, we will link the QLS criterion with BLUE methods for location-scale families and thus introduce two types of estimators: ordinary QLS and generalized QLS (Sections 2 and 3). Besides studying small-sample properties of the new estimators under clean and contaminated data scenarios (Section 4), we will evaluate computational costs of these estimators and those of MLE for sample sizes  $n = 10^6, 10^7, 10^8, 10^9$ . In addition, two goodness-of-fit tests will be constructed (Section 3.5) and their performance will be studied using simulations (Section 4.4) and real data (Section 5).

# 2 Quantile Least Squares

In this section, a general formulation of the least squares estimators based on sample quantiles is presented, and their asymptotic robustness and relative efficiency properties are specified.

Suppose a sample of independent and identically distributed (i.i.d.) continuous random variables,  $X_1, \ldots, X_n$ , with the cumulative distribution function (cdf) F, probability density function (pdf) f, and quantile function (qf)  $F^{-1}$ , is observed. Let the cdf, pdf, and qf be given in a parametric form, and suppose that they are indexed by an m-dimensional parameter  $\theta = (\theta_1, \ldots, \theta_m)$ . Further, let  $X_{(1)} \leq \cdots \leq X_{(n)}$  denote the ordered sample values. The empirical estimator of the pth population quantile is the corresponding sample quantile  $X_{(\lceil np \rceil)} = \widehat{F}^{-1}(p)$ , where  $\lceil \cdot \rceil$  denotes the rounding up operation. Also, throughout the paper the notation  $\mathcal{AN}$  stands for "asymptotically normal."

## 2.1 Regression Estimation

To specify a regression framework, we first recall the joint asymptotic normality result of sample quantiles. (The following theorem is slightly edited to match the context of the current paper.)

**Theorem 2.1** [Serfling (2002a, p.80, Theorem B)]

Let  $0 < p_1 < \cdots < p_k < 1$ , and suppose that pdf f is continuous. Then the k-variate vector of empirical quantiles  $(\widehat{F}^{-1}(p_1), \dots, \widehat{F}^{-1}(p_k))$  is  $\mathcal{AN}$  with the mean vector  $(F^{-1}(p_1), \dots, F^{-1}(p_k))$  and  $k \times k$  covariance-variance matrix with elements  $\sigma_{ij}/n$ , where

$$\sigma_{ij} = \frac{p_i(1 - p_j)}{f(F^{-1}(p_i))f(F^{-1}(p_j))} \quad \text{for } i \le j$$
 (2.1)

and  $\sigma_{ij} = \sigma_{ji}$  for i > j.

For large sample size and general F, this result can be interpreted as a nonlinear regression model with normally distributed error terms. That is,

$$\widehat{F}^{-1}(p_i) = F^{-1}(p_i) + \varepsilon_i, \qquad i = 1, \dots, k,$$
 (2.2)

where the error term  $\varepsilon = (\varepsilon_1, \dots, \varepsilon_k)$  is  $\mathcal{AN}(\mathbf{0}, \Sigma/n)$  with the elements of  $\Sigma$  given by (2.1). Since  $F^{-1}(p_i)$  is a function of  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_m)$ , the number of quantiles (k) should be at least as large as the number of parameters (m). Then, the least squares problem can be formulated as follows:

minimize 
$$\sum_{i=1}^{k} \left( \widehat{F}^{-1}(p_i) - F^{-1}(p_i) \right)^2 \quad \text{with respect to } \theta_1, \dots, \theta_m.$$
 (2.3)

In general, (2.3) is a challenging computational problem that requires numerical optimization algorithms. Moreover, the objective function may have many local minima and even the global minimum may produce a biased estimate. But as was demonstrated by Xu et al. (2014, Section 2.1) for the g-and-h distributional family, this problem can be solved with rapidly converging algorithms, and its solution possesses several desirable properties: consistency, asymptotic normality, bounded influence functions, positive breakdown point. We also notice that using similar arguments to those of Xu et al. (2014) the equivalent theoretical properties can be established for other parametric distributions, which will be discussed in Sections 2.2-2.3. Further, it will be shown in Section 3 that for location-scale families and their variants, the nonlinear regression model (2.2) becomes a linear regression model with (approximately) normally distributed error terms whose covariance-variance matrix has a convenient structure. As a result, the latter problem has explicit solutions with known theoretical properties.

## 2.2 Robustness Properties

The quantile least squares (QLS) estimator found by solving (2.3) can be viewed as an *indirect* estimator, robust inferential properties of which are provided by Genton and de Luna (2000) and Genton and Ronchetti (2003). Using those general results and the arguments of Xu *et al.* (2014) several properties of the QLS estimator can be stated. First, as is clear from the choice of quantile confidence levels,

$$0 < a = p_1 < p_2 < \dots < p_{k-1} < p_k = b < 1,$$

the order statistics with the index less than  $\lceil na \rceil$  and more than  $\lceil nb \rceil$  play no role in estimation of the regression model (2.2). This implies that the QLS estimator is globally robust with the (asymptotic) breakdown point equal to:

$$BP = \min\{LBP, UBP\} = \min\{a, 1 - b\} > 0.$$
 (2.4)

Note that when the underlying probability distribution F is not symmetric it makes sense to consider lower (LBP) and upper (UBP) breakdown points separately. For more details on the relevance of LBP and UBP in applications, see Brazauskas and Serfling (2000) and Serfling (2002b). Second, the influence function (IF) of the QLS estimator for  $\theta$  is directly related to the influence functions of "data", i.e., the selected sample quantiles  $\hat{F}^{-1}(p_1), \ldots, \hat{F}^{-1}(p_k)$ :

$$\operatorname{IF}(x, \widehat{F}^{-1}(p_i)) = \frac{p_i - \mathbf{1}\{x \le F^{-1}(p_i)\}}{f(F^{-1}(p_i))}, \quad i = 1, \dots, k,$$

where  $-\infty < x < \infty$  and  $\mathbf{1}\{\cdot\}$  denotes the indicator function. Specifically, the IF of  $\hat{\boldsymbol{\theta}}$  is given by

$$\operatorname{IF}(x,\widehat{\boldsymbol{\theta}}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\left(\operatorname{IF}(x,\widehat{F}^{-1}(p_1)), \dots, \operatorname{IF}(x,\widehat{F}^{-1}(p_k))\right)', \tag{2.5}$$

where  $\mathbf{X} = \left[X_{ij}\right]_{k \times m} = \left[\frac{\partial \widehat{F}^{-1}(p_i)}{\partial \widehat{\theta}_j}\right]\Big|_{\widehat{\boldsymbol{\theta}} = \boldsymbol{\theta}}$ , and is bounded because  $p_1 = a > 0$  and  $p_k = b < 1$ .

## 2.3 Asymptotic Relative Efficiency

To start with, the model assumptions used in Theorem 1 of Xu et al. (2014, Section 2.1) can be broadened to include other parametric models. Then, repeating the arguments used by these authors to prove the theorem, it can be established that in general the QLS estimator is consistent and  $\mathcal{AN}$  with the mean vector  $\boldsymbol{\theta}$  and  $m \times m$  covariance-variance matrix

$$\frac{1}{n} (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \mathbf{\Sigma} \mathbf{X} (\mathbf{X}'\mathbf{X})^{-1}, \tag{2.6}$$

where X is defined as in (2.5) and the elements of  $\Sigma$  are given by (2.1).

Further, under suitable regularity conditions (Serfling, 2002a, Section 4.2.2), the maximum likelihood estimator (MLE) is  $\mathcal{AN}$  with the mean vector  $\boldsymbol{\theta}$  and  $m \times m$  covariance-variance matrix  $\frac{1}{n} \mathbf{I}^{-1}$ , where  $\mathbf{I}$  is the Fisher information matrix. Since MLE is the most efficient  $\mathcal{AN}$  estimator (i.e., its asymptotic variance attains the Cramér-Rao bound), its performance can serve as a benchmark for the QLS estimator. In particular, the following asymptotic relative efficiency (ARE) criterion will be used:

ARE (QLS, MLE) = 
$$\left(\frac{\det\left[\mathbf{I}^{-1}\right]}{\det\left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Sigma}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\right]}\right)^{1/m}$$
, (2.7)

where 'det' stands for the determinant of a square matrix (Serfling, 2002a, Section 4.1).

#### 3 Location-Scale Families

In this section, the proposed methodology is worked out for location-scale families. Several such families are listed in Section 3.1. Two QLS-type estimators are developed in Section 3.2. Further, efficiency and robustness properties of the new estimators are established in Sections 3.3 and 3.4, respectively. Finally, in Section 3.5, we analyze model residuals and explore its goodness-of-fit properties.

#### 3.1 Preliminaries

The pdf f, cdf F, and the qf  $F^{-1}$  of the location-scale family are given by:

$$f(x) = \frac{1}{\sigma} f_* \left( \frac{x - \mu}{\sigma} \right), \qquad F(x) = F_* \left( \frac{x - \mu}{\sigma} \right), \qquad F^{-1}(u) = \mu + \sigma F_*^{-1}(u),$$
 (3.1)

where  $-\infty < x < \infty$  (or depending on the distribution, x can be restricted to some interval), 0 < u < 1,  $-\infty < \mu < \infty$  is the location parameter, and  $\sigma > 0$  is the scale parameter. The functions  $f_*$ ,  $F_*$ ,  $F_*^{-1}$  represent pdf, cdf, qf, respectively, of the standard location-scale family (i.e.,  $\mu = 0$ ,  $\sigma = 1$ ). Choosing  $\mu$  or  $\sigma$  known, the location-scale family is reduced to either the *scale* or *location* family, respectively.

In Table 3.1, we list key facts for several location-scale families. The selected distributions include typical symmetric bell-shaped densities, with domains on all real numbers (e.g., Cauchy, Laplace, Logistic, Normal), as well as few asymmetric densities with varying domains (e.g., Exponential, Gumbel, Lévy). In the latter group, the Gumbel pdf is defined on all real numbers but is slightly skewed; this distribution plays an important role in extreme value theory. Two-parameter Exponential and Lévy densities are highly skewed and have domains  $(\mu, \infty)$ . They represent examples when the aforementioned regularity conditions are not satisfied, due to the presence of  $\mu$ . Both distributions are widely used in applications and have many probabilistic connections. For example, the Lévy distribution is directly related to the following well-known distributions: Inverse Gamma, Stable, and Folded Normal.

Table 3.1. Key probabilistic formulas and information for selected location-scale families.

Probability	Standard PDF	Standard QF	Information Matrix
Distribution	$f_*(z)$	$F_*^{-1}(u)$	$\mathbf{I}_* (= \sigma^2 \times \mathbf{I})$
Cauchy	$\frac{1}{\pi(1+z^2)}$	$\tan(\pi(u-0.5))$	$\begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix}$
Laplace	$0.5  e^{- z }$	$\begin{cases} \ln(2u), & u \le 0.5 \\ -\ln(2(1-u)), & u > 0.5 \end{cases}$	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$
Logistic	$\frac{e^{-z}}{(1+e^{-z})^2}$	$-\ln(1/u-1)$	$\begin{bmatrix} \frac{1}{3} & 0\\ 0 & \frac{3+\pi^2}{9} \end{bmatrix}$
Normal	$\frac{1}{\sqrt{2\pi}} e^{-z^2/2}$	$\Phi^{-1}(u)$	$\begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$
Exponential	$e^{-z}, z > 0$	$-\ln(1-u)$	1 (for $\sigma$ ; $\mu$ is known)
Gumbel	$\exp\{-z - e^{-z}\}$	$-\ln(-\ln(u))$	$\begin{bmatrix} 1 & \gamma - 1 \\ \gamma - 1 & \frac{\pi^2}{6} + (\gamma - 1)^2 \end{bmatrix}$
Lévy	$\frac{1}{\sqrt{2\pi}} z^{-3/2} e^{-(2z)^{-1}}, \ z > 0$	$\left(\Phi^{-1}(1-u/2)\right)^{-2}$	$\frac{1}{2}$ (for $\sigma$ ; $\mu$ is known)

Note:  $\gamma = -\Gamma'(1) \approx 0.5772$  is the Euler-Mascheroni constant.

It is worthwhile mentioning that there exist numerous variants of the location-scale family such as folded distributions (e.g., Folded Normal, Folded Cauchy) or log-location-scale families (e.g., Lognormal, Pareto type I). Since their treatment requires only suitable parameter-free transformation of the data variable, the estimators developed in this paper will work for those distributions as well.

# 3.2 Parameter Estimation

Incorporating expressions (3.1) into the model (2.2) yields a linear regression model:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},\tag{3.2}$$

where  $\mathbf{Y} = (\widehat{F}^{-1}(p_1), \dots, \widehat{F}^{-1}(p_k))'$ ,  $\boldsymbol{\beta} = (\mu, \sigma)'$ , and  $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_k)'$  is  $\mathcal{AN}(\mathbf{0}, \sigma^2 \boldsymbol{\Sigma}_* / n)$ . The entries of  $\boldsymbol{\Sigma}_*$  are defined by (2.1), but now they are completely *known* because f and  $F^{-1}$  are replaced with  $f_*$  and  $F_*^{-1}$ , respectively. The design matrix  $\mathbf{X}$  is defined as in (2.5) and has simple entries:

$$\mathbf{X} = \begin{bmatrix} \frac{\partial F^{-1}(p_1)}{\partial \mu} & \cdot & \cdot & \frac{\partial F^{-1}(p_k)}{\partial \mu} \\ \frac{\partial F^{-1}(p_1)}{\partial \sigma} & \cdot & \cdot & \frac{\partial F^{-1}(p_k)}{\partial \sigma} \end{bmatrix}' = \begin{bmatrix} 1 & \cdot & \cdot & 1 \\ F_*^{-1}(p_1) & \cdot & \cdot & F_*^{-1}(p_k) \end{bmatrix}'.$$
(3.3)

Solving (2.3) for the model (3.2) leads to the *ordinary* least squares estimator

$$\widehat{\boldsymbol{\beta}}_{\text{oQLS}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \tag{3.4}$$

which is  $\mathcal{AN}$  with the mean vector  $\boldsymbol{\beta} = (\mu, \sigma)'$  and  $2 \times 2$  covariance-variance matrix

$$\frac{\sigma^2}{n} (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \mathbf{\Sigma}_* \mathbf{X} (\mathbf{X}'\mathbf{X})^{-1}, \tag{3.5}$$

where X is given by (3.3).

Further, the oQLS solution (3.4) implicitly assumes that  $\Sigma_*$  is the  $k \times k$  identity matrix, which might be a sensible assumption for the non-linear regression model (2.3) because the resulting estimator is consistent while the computational complexity is significantly reduced. In general, however, such a simplification decreases the estimator's efficiency. Since for the linear regression model (3.2),  $\Sigma_*$  is known, ARE of oQLS can be improved by employing the generalized least squares estimator

$$\widehat{\boldsymbol{\beta}}_{gQLS} = (\mathbf{X}'\boldsymbol{\Sigma}_{*}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}_{*}^{-1}\mathbf{Y}$$
(3.6)

which is  $\mathcal{AN}$  with the mean vector  $\boldsymbol{\beta} = (\mu, \sigma)'$  and  $2 \times 2$  covariance-variance matrix

$$\frac{\sigma^2}{n} \left( \mathbf{X}' \mathbf{\Sigma}_*^{-1} \mathbf{X} \right)^{-1}. \tag{3.7}$$

Finally, note that if a one-parameter family – location or scale – needs to be estimated, the formulas (3.4)–(3.7) still remain valid, but the design matrix (3.3) would be a column of 1's (for location) or a column of  $F_*^{-1}(p)$ 's (for scale).

# 3.3 Relative Efficiency Studies

To see how much efficiency is sacrificed when one uses  $\widehat{\boldsymbol{\beta}}_{\text{oQLS}}$  instead of  $\widehat{\boldsymbol{\beta}}_{\text{gQLS}}$ , we will compute (2.7) for several distributions of Table 3.1. In view of (3.5) and (3.7), the ARE formula (2.7) is now given by

ARE (oQLS, MLE) = 
$$\left(\frac{\det\left[\mathbf{I}_{*}^{-1}\right]}{\det\left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}_{*}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\right]}\right)^{1/2}$$

and

$$\mathrm{ARE}\left(\mathrm{gQLS},\,\mathrm{MLE}\right) \;=\; \left(\frac{\det\left[\mathbf{I}_{*}^{-1}\right]}{\det\left[(\mathbf{X}'\boldsymbol{\Sigma}_{*}^{-1}\mathbf{X})^{-1}\right]}\right)^{1/2},$$

where  $I_*$  is specified in Table 3.1. For one-parameter families (location or scale), the covariance-variance matrices in the ARE formulas get reduced to scalars and the exponents become 1.

These ARE expressions are functions of k, the number of selected sample quantiles, therefore the choice of  $p_1, \ldots, p_k$  (with  $k \geq 2$ ) is important. As mentioned earlier, our top priority is estimators' robustness. Thus, to keep the breakdown points positive and influence functions bounded, we first fix  $a = p_1 > 0$  and  $b = p_k < 1$  and then make the remaining  $p_i$ 's equally spaced:

$$p_i = a + \frac{i-1}{k-1}(b-a), \qquad i = 1, \dots, k.$$
 (3.8)

It is clear that choosing larger a and smaller b yields higher robustness, while choosing larger k improves efficiency. But there are practical limits to the efficiency improvement. As can be seen from Figure 3.1, the (pointwise) ARE curves become almost flat for k > 10 making the efficiency gains negligible. This holds true irrespectively of the underlying location-scale family. On the other hand, choosing gQLS over oQLS gives a major boost to ARE, especially for the heavier-tailed distributions such as Gumbel, Laplace, and Cauchy. Also, the seesaw pattern of the ARE curve (for location and to a lesser degree for location-scale) for the Laplace distribution can be explained as follows: for a = 1 - b and k odd, one of the  $p_i$ 's is always equal to 0.50 resulting in the selection of the sample median, which in this case is MLE for  $\mu$  and thus full efficiency is attained.

Further, to demonstrate that increasing k yields no substantial gains in efficiency, in Table 3.2 we list AREs of the generalized QLS estimators for Cauchy, Gumbel, Laplace, Logistic, and Normal distributions, when k = 15, 20, 25. As is evident from the table, choosing k = 25 over k = 15 results in  $\sim 1\%$  improvement of AREs. Similar magnitude improvements can be observed when the extreme quantile levels a and b are changed from (0.02, 0.98) to (0.05, 0.95) to (0.10, 0.90). In view of this discussion and keeping in mind that k should be odd (see the ARE entries for Laplace), we can say that the choice of k = 15 would suffice in most situations. However, in order to not squander much efficiency at some unusual location-scale distributions, k = 25 is a safer choice.

Finally, it is tempting to try the brute force approach for the ordinary QLS estimators with the hope that it would substantially improve ARE. In Table 3.3, we list AREs for the oQLS estimator (of the joint location-scale parameter) when k ranges from 15 to 200. Depending on the distribution and how low the ARE value is at k = 15, some tiny improvements are still possible even for k = 200 (Cauchy) but they are leveling off (Logistic). More interestingly, for the Laplace, Normal, and Gumbel distributions, their

AREs reach the peak at some lower k and then start slowly declining. This behavior is not unexpected because the oQLS estimator, while consistent, is based on an incorrect simplifying assumption that  $\Sigma_*$  is the  $k \times k$  identity matrix. This assumption, combined with the brute force approach, will eventually penalize the estimator's performance.

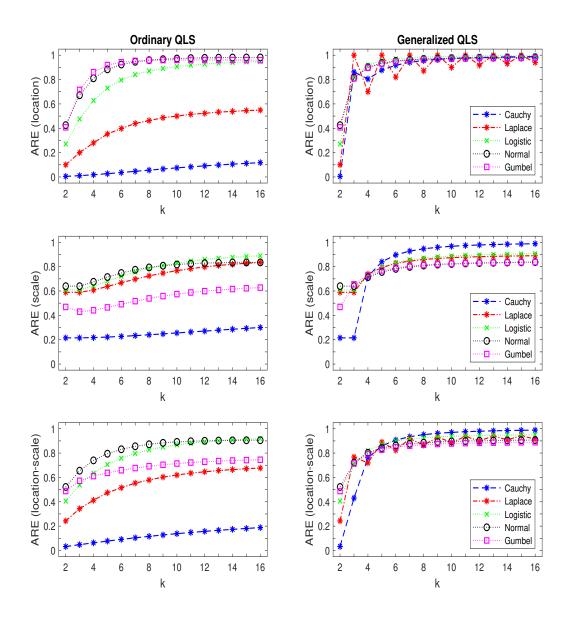


FIGURE 3.1. AREs of the ordinary and generalized QLS estimators of location, scale, and joint location-scale parameters for Cauchy, Gumbel, Laplace, Logistic, and Normal distributions. The quantiles are selected according to (3.8) with (a,b) = (0.05,0.95) and k=2:16.

TABLE 3.2. AREs of the generalized QLS estimators of location, scale, and joint location-scale parameters for Cauchy, Gumbel, Laplace, Logistic, and Normal distributions. The quantiles are selected according to (3.8) with various (a,b) and k=15, 20, 25.

Probability		Location			Scale		Location-Scale		
Distribution	k = 15	k = 20	k = 25	k = 15	k = 20	k = 25	k = 15	k = 20	k = 25
(a, b) = (0.02, 0.98)									
Cauchy	0.986	0.992	0.995	0.985	0.992	0.995	0.985	0.992	0.995
Laplace	1	0.950	1	0.930	0.943	0.949	0.965	0.946	0.974
Logistic	0.996	0.998	0.998	0.938	0.951	0.958	0.966	0.974	0.978
Normal	0.987	0.991	0.992	0.901	0.915	0.922	0.943	0.952	0.957
Gumbel	0.985	0.990	0.991	0.902	0.913	0.918	0.933	0.941	0.946
(a, b) = (0.05)	, 0.95)								
Cauchy	0.988	0.993	0.995	0.987	0.993	0.995	0.987	0.993	0.995
Laplace	1	0.953	1	0.888	0.894	0.896	0.943	0.923	0.947
Logistic	0.996	0.998	0.999	0.904	0.910	0.913	0.949	0.953	0.955
Normal	0.982	0.984	0.985	0.836	0.841	0.843	0.906	0.909	0.911
Gumbel	0.979	0.981	0.982	0.836	0.840	0.842	0.888	0.892	0.893
(a, b) = (0.10)	, 0.90)								
Cauchy	0.981	0.985	0.986	0.989	0.993	0.995	0.985	0.989	0.991
Laplace	1	0.958	1	0.796	0.798	0.799	0.892	0.874	0.894
Logistic	0.995	0.997	0.997	0.814	0.816	0.817	0.900	0.902	0.903
Normal	0.964	0.965	0.965	0.708	0.710	0.711	0.826	0.828	0.828
Gumbel	0.956	0.957	0.957	0.719	0.720	0.721	0.803	0.805	0.805

Table 3.3. AREs of the ordinary QLS estimators of joint location-scale parameters for Cauchy, Gumbel, Laplace, Logistic and Normal distributions, with (a,b) = (0.05,0.95) and various k (see (3.8)).

Probability				k			
Distribution	15	20	25	50	75	100	200
Cauchy	0.181	0.211	0.232	0.282	0.299	0.308	0.321
Laplace	0.742	0.753	0.757	0.759	0.758	0.757	0.756
Logistic	0.672	0.693	0.704	0.718	0.720	0.721	0.722
Normal	0.914	0.930	0.936	0.941	0.940	0.940	0.938
Gumbel	0.905	0.905	0.902	0.890	0.885	0.882	0.877

#### 3.4 Robustness Investigations

To see what kind of shapes the influence functions of  $\widehat{\boldsymbol{\beta}}_{\text{oQLS}}$  and  $\widehat{\boldsymbol{\beta}}_{\text{gQLS}}$  exhibit, we evaluate and plot (2.5) for the symmetric (Figure 3.2) and asymmetric (Figure 3.3) location-scale families of Table 3.1. In view of (3.4), (3.6), and (3.1), the expression (2.5) is now given by

$$\operatorname{IF}\left(x,\widehat{\boldsymbol{\beta}}_{\operatorname{oQLS}}\right) = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\left(\operatorname{IF}\left(x,\widehat{F}^{-1}(p_1)\right), \dots, \operatorname{IF}\left(x,\widehat{F}^{-1}(p_k)\right)\right)'$$

and

$$\operatorname{IF}\left(x,\widehat{\boldsymbol{\beta}}_{\operatorname{gQLS}}\right) = \left(\mathbf{X}'\boldsymbol{\Sigma}_{*}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{\Sigma}_{*}^{-1}\left(\operatorname{IF}\left(x,\widehat{F}^{-1}(p_{1})\right),\ldots,\operatorname{IF}\left(x,\widehat{F}^{-1}(p_{k})\right)\right)',$$

where  $\text{IF}(x, \widehat{F}^{-1}(p_i)) = \sigma \frac{p_i - \mathbf{1}\{x \leq \mu + \sigma F_*^{-1}(p_i)\}}{f_*(F_*^{-1}(p_i))}, i = 1, \dots, k. \text{ In Figures 3.2 and 3.3, } \mu = 0 \text{ and } \sigma = 1.$ 

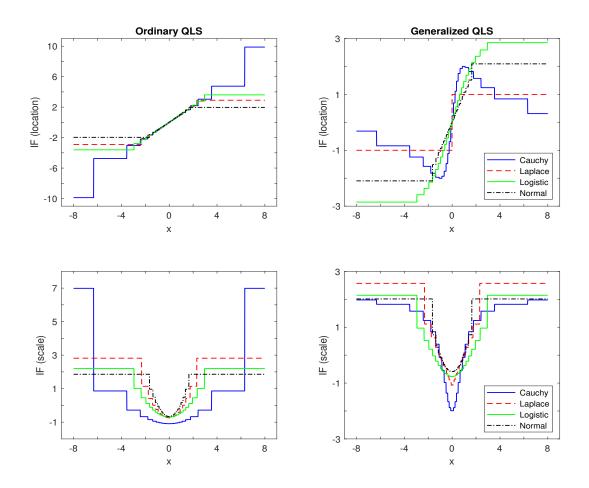


FIGURE 3.2. Influence functions of the ordinary and generalized QLS estimators of location and scale parameters for Cauchy, Laplace, Logistic, and Normal distributions. The quantiles are selected according to (3.8) with (a, b) = (0.05, 0.95) and k = 25.

In Figure 3.2, we see that the IF shapes of the ordinary QLS estimators look familiar. For estimation of  $\mu$ , the estimators act like a stepwise approximation of a trimmed/winsorized mean or Huber estimator (Hampel et al., 1986, Figure 1, p.105). For estimation of  $\sigma$ , they behave like an approximate version of an M-estimator for scale (Hampel et al., 1986, Figure 2, p.123). On the other hand, the generalized QLS estimators demonstrate a remarkable flexibility. For estimation of  $\sigma$ , gQLS shrinks the height of the Cauchy IF and keeps the other curves similar to those of oQLS. But most impressively, it automatically changes the shape of the IF when estimating  $\mu$ : for Normal and Logistic distributions, it acts like a trimmed/winsorized mean; for Laplace, it behaves like a median; and for Cauchy, its shape resembles that of a Tukey's biweight (Hampel et al., 1986, Figure 3, p.151).

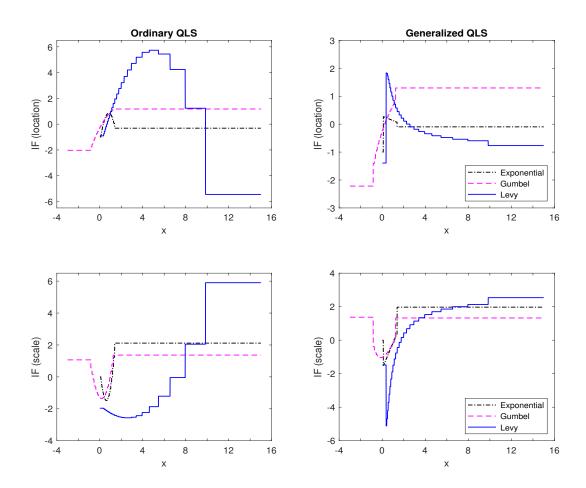


FIGURE 3.3. Influence functions of the ordinary and generalized QLS estimators of location and scale parameters for Exponential, Gumbel, and Lévy distributions. The quantiles are selected according to (3.8) with (a, b) = (0.10, 0.75) and k = 25.

In Figure 3.3, the shapes of IF are predictably non-symmetric. For oQLS and gQLS at Gumbel, they are fairly similar to the IFs of Normal or Logistic distributions. (Note that by choosing  $a = 0.10 \neq$ 

0.25 = 1 - b we made the IF look more symmetric than it actually is.) For Exponential and Lévy, parameter  $\mu$  is not the "center" of the pdf anymore; it is the left boundary of its support. This fact has an effect on the shapes of IFs. For gQLS of  $\mu$  and  $\sigma$ , we see that points near the boundary exhibit most dramatic swings of influence. Overall, these IFs can be seen as a half of a symmetric family IF.

## 3.5 Model Validation

For model validation, we consider two goodness-of-fit tests, both are constructed using  $\widehat{\beta}_{gQLS}$  which possesses more favorable efficiency-robustness properties than  $\widehat{\beta}_{oQLS}$ . The first test is a typical  $\chi^2$  test that is based on a quadratic form in model residuals. This approach will be called "in-sample validation" (Section 3.5.1). The second test is conceptually similar but is based on a combination of the model residuals and additional sample quantiles. The inclusion of quantiles that had not been used for parameter estimation allows us to make a fair comparison among the estimators with different a and b. This approach will be called "out-of-sample validation" (Section 3.5.2).

#### 3.5.1 In-Sample Validation

After the parameter estimation step is completed, the predicted value of  $\mathbf{Y}$  is defined as  $\hat{\mathbf{Y}} = \mathbf{X} \hat{\boldsymbol{\beta}}_{\text{gQLS}}$ . Then the corresponding residuals are

$$\widehat{\boldsymbol{\varepsilon}} \ = \ \mathbf{Y} - \widehat{\mathbf{Y}} \ = \ \mathbf{Y} - \mathbf{X} \widehat{\boldsymbol{\beta}}_{\mathrm{gQLS}} \ = \ \left(\mathbf{I}_{k} - \mathbf{X} (\mathbf{X}'\boldsymbol{\Sigma}_{*}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}_{*}^{-1}\right)\mathbf{Y},$$

where  $I_k$  is the  $k \times k$  identity matrix. Using (3.6), (3.7) and standard statistical inference techniques for linear models (Hogg *et al.*, 2005, Section 12.3) the following properties can be verified:

- Y has a  $\mathcal{AN}\left(\mathbf{X}\boldsymbol{\beta}, \frac{\sigma^2}{n}\boldsymbol{\Sigma}_*\right)$  distribution.
- $\hat{\mathbf{Y}}$  has a  $\mathcal{AN}\left(\mathbf{X}\boldsymbol{\beta}, \frac{\sigma^2}{n}\mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma}_*^{-1}\mathbf{X})^{-1}\mathbf{X}'\right)$  distribution.
- $\widehat{\boldsymbol{\varepsilon}}$  has a  $\mathcal{AN}\left(\mathbf{0}, \frac{\sigma^2}{n}\left(\boldsymbol{\Sigma_*} \mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma_*^{-1}}\mathbf{X})^{-1}\mathbf{X}'\right)\right)$  distribution.
- ullet  $\widehat{\mathbf{Y}}$  and  $\widehat{oldsymbol{arepsilon}}$  are (asymptotically) independent.

Next, these properties can be exploited to construct a diagnostic plot (e.g., predicted values *versus* residuals) and to show that the quadratic form

$$Q = \frac{n}{\sigma^2} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})' \boldsymbol{\Sigma}_*^{-1} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})$$

has the following orthogonal decomposition:

$$Q = Q_1 + Q_2 = \frac{n}{\sigma^2} \left( \mathbf{Y} - \mathbf{X} \widehat{\boldsymbol{\beta}}_{gQLS} \right)' \boldsymbol{\Sigma}_*^{-1} \left( \mathbf{Y} - \mathbf{X} \widehat{\boldsymbol{\beta}}_{gQLS} \right) + \frac{n}{\sigma^2} \left( \widehat{\boldsymbol{\beta}}_{gQLS} - \boldsymbol{\beta} \right)' \mathbf{X}' \boldsymbol{\Sigma}_*^{-1} \mathbf{X} \left( \widehat{\boldsymbol{\beta}}_{gQLS} - \boldsymbol{\beta} \right).$$

Therefore, since asymptotically Q has a  $\chi_k^2$  distribution and  $Q_2$  has a  $\chi_2^2$  distribution, the above decomposition implies that  $Q_1$  has an approximate  $\chi_{k-2}^2$  distribution.

Now, to test the hypotheses

$$\begin{cases} H_0: X_1, \dots, X_n & \text{were generated by a location-scale family } F \\ H_A: X_1, \dots, X_n & \text{were } not \text{ generated by } F, \end{cases}$$

the quadratic form  $Q_1$  can be utilized as follows. Recall that  $\widehat{\boldsymbol{\beta}}_{gQLS} = (\widehat{\mu}_{gQLS}, \widehat{\sigma}_{gQLS})'$  is a consistent estimator of  $\boldsymbol{\beta}$ , thus  $\widehat{\sigma}_{gQLS}^2$  converges in probability to  $\sigma^2$ . Define a test statistic

$$W = \frac{n}{\widehat{\sigma}_{gOLS}^2} \left( \mathbf{Y} - \mathbf{X} \widehat{\boldsymbol{\beta}}_{gQLS} \right)' \boldsymbol{\Sigma}_*^{-1} \left( \mathbf{Y} - \mathbf{X} \widehat{\boldsymbol{\beta}}_{gQLS} \right). \tag{3.9}$$

Since  $W = \frac{\sigma^2}{\widehat{\sigma}_{\rm gQLS}^2} Q_1$  and  $\frac{\sigma^2}{\widehat{\sigma}_{\rm gQLS}^2} \to 1$  (in probability), it follows from Slutsky's Theorem that under  $H_0$  the test statistic W has an approximate  $\chi^2_{k-2}$  distribution. Note that a similar goodness-of-fit test was proposed by Ali and Umbach (1989), but there  $\sigma^2$  was estimated by the sample variance, which requires that F has a finite variance. The test based on (3.9) has wider applicability (e.g., it works for heavy-tailed distributions such as Cauchy) and inherits the robustness properties of  $\widehat{\beta}_{\rm gQLS}$ .

#### 3.5.2 Out-of-Sample Validation

To compare the goodness of fit of location-scale distributions for which  $\widehat{\boldsymbol{\beta}}_{\text{gQLS}}$  are computed using different a and b (i.e.,  $a_1, b_1$  versus  $a_2, b_2$ ), we first fix a universal set of sample quantiles. That is, select  $\mathbf{Y}_{\text{out}} = \left(\widehat{F}(p_1^{\text{out}}), \dots, \widehat{F}(p_r^{\text{out}})\right)'$ , where  $p_1^{\text{out}}, \dots, p_r^{\text{out}}$  can be all different from, partially overlapping with, or completely match  $p_1, \dots, p_k$  (which are used for parameter estimation). Of course, the latter choice simplifies the out-sample-validation test to the test of Section 3.5.1. After this selection is made, we proceed by mimicking the structure of (3.9). The predicted value of  $\mathbf{Y}_{\text{out}}$  is  $\mathbf{X}_{\text{out}}\widehat{\boldsymbol{\beta}}_{\text{gQLS}}$  with

$$\mathbf{X}_{\text{out}} = \begin{bmatrix} 1 & \cdot & \cdot & \cdot & 1 \\ F_*^{-1}(p_1^{\text{out}}) & \cdot & \cdot & \cdot & F_*^{-1}(p_r^{\text{out}}) \end{bmatrix}',$$

but  $\widehat{\boldsymbol{\beta}}_{gQLS} = (\widehat{\mu}_{gQLS}, \widehat{\sigma}_{gQLS})'$  is based on  $\mathbf{Y} = (\widehat{F}(p_1), \dots, \widehat{F}(p_k))'$ . Then the test statistic is

$$W_{\text{out}} = \frac{n}{\widehat{\sigma}_{\text{gQLS}}^2} \left( \mathbf{Y}_{\text{out}} - \mathbf{X}_{\text{out}} \widehat{\boldsymbol{\beta}}_{\text{gQLS}} \right)' \boldsymbol{\Sigma}_{\text{out}}^{-1} \left( \mathbf{Y}_{\text{out}} - \mathbf{X}_{\text{out}} \widehat{\boldsymbol{\beta}}_{\text{gQLS}} \right), \tag{3.10}$$

where the elements of  $\Sigma_{\text{out}}$  are  $\sigma_{ij}^{\text{out}} = \frac{p_i^{\text{out}}(1-p_j^{\text{out}})}{f_*(F_*^{-1}(p_i^{\text{out}}))f_*(F_*^{-1}(p_j^{\text{out}}))}$  for  $i \leq j$  with  $i, j = 1, \ldots, r$ .

Now, unless  $p_1^{\text{out}}, \ldots, p_r^{\text{out}}$  perfectly match  $p_1, \ldots, p_k$  (this case was solved in Section 3.5.1), the theoretical derivation of the distribution of  $W_{\text{out}}$  is a major challenge. Therefore, to estimate the p-value associated with this test statistic, the following bootstrap procedure can be employed.

#### Bootstrap Procedure (for finding the p-value of (3.10))

**Step 1.** Given the original sample,  $X_1, \ldots, X_n$ , the estimates of  $\boldsymbol{\beta}$  and  $W_{\text{out}}$  are obtained. Denote them  $\widehat{\boldsymbol{\beta}}_{\text{gQLS}}^o = (\widehat{\mu}_{\text{gQLS}}^o, \widehat{\sigma}_{\text{gQLS}}^o)'$  and  $\widehat{W}_{\text{out}}^o$ . Remember that  $\widehat{\boldsymbol{\beta}}_{\text{gQLS}}^o$  is computed using the quantile levels  $p_1, \ldots, p_k$ , while  $\widehat{W}_{\text{out}}^o$  is based on  $p_1^{\text{out}}, \ldots, p_r^{\text{out}}$  and  $\widehat{\boldsymbol{\beta}}_{\text{gQLS}}^o$ .

**Step 2.** Generate an *i.i.d.* sample  $X_1^{(b)}, \ldots, X_n^{(b)}$  from F (assumed under  $H_0$ ) using the parameter values  $\widehat{\boldsymbol{\beta}}_{\text{gQLS}}^o = (\widehat{\mu}_{\text{gQLS}}^o, \widehat{\sigma}_{\text{gQLS}}^o)'$ . Based on this sample, compute  $\widehat{\boldsymbol{\beta}}_{\text{gQLS}}^{(b)}$  (using  $p_1, \ldots, p_k$ ) and  $\widehat{W}_{\text{out}}^{(b)}$  (using  $p_1^{\text{out}}, \ldots, p_r^{\text{out}}$  and  $\widehat{\boldsymbol{\beta}}_{\text{gQLS}}^{(b)}$ ).

**Step 3.** Repeat Step 2 a B number of times (e.g., B=1000) and save  $\widehat{W}_{\mathrm{out}}^{(1)},\ldots,\widehat{W}_{\mathrm{out}}^{(B)}$ 

**Step 4.** Estimate the *p*-value of (3.10) by

$$\widehat{p}_{\text{val}} = \frac{1}{B} \sum_{b=1}^{B} \mathbf{1} \left\{ \widehat{W}_{\text{out}}^{(b)} > \widehat{W}_{\text{out}}^{o} \right\}$$

and reject  $H_0$  when  $\hat{p}_{val} \leq \alpha$  (e.g.,  $\alpha = 0.05$ ).

# 4 Simulations

In this section, we conduct a simulation study with the objective to verify and augment the theoretical properties established in Section 3. We start by describing the study design (Section 4.1). Then we explore how the MLE, oQLS and gQLS estimators perform as the sample size n increases (Section 4.2), and when data are contaminated with outliers (Section 4.3). We finish the study by investigating the power of the goodness-of-fit tests against several alternatives (Section 4.4).

#### 4.1 Study Design

The study design is based on the following choices.

For any given distribution, we generate  $M=10^4$  random samples of a specified length n. For each sample, we estimate  $\mu$  and  $\sigma$  of the distribution using the MLE, oQLS, and gQLS estimators. The results are then presented using boxplots and few summarizing statistics.

#### Simulation Design

- Location-scale families ( $F_0$  with  $\mu = 0$  and  $\sigma = 1$ ). Cauchy, Exponential, Gumbel, Laplace, Lévy, Logistic, Normal.
- Estimators. MLE, oQLS (abbreviated 'o') and gQLS (abbreviated 'g'). For 'o' and 'g' estimators, the quantiles are selected according to (3.8) with  $(a_1, b_1) = (0.02, 0.98)$ ,  $(a_2, b_2) = (0.05, 0.95)$ ,  $(a_3, b_3) = (0.10, 0.90)$  and k = 25 (in all cases).
- Contaminating distributions (G in the contamination model  $F_{\varepsilon} = (1 \varepsilon) F_0 + \varepsilon G$ , where  $F_0$  is a location-scale family). Exponential ( $\mu^* = 1, \sigma^* = 3$ ) and Normal ( $\mu^* = 1, \sigma^* = 3$ ).
- Levels of contamination.  $\varepsilon = 0, 0.03, 0.05, 0.08$ .
- Goodness-of-fit tests (at  $\alpha = 0.05$ ). Based on W, given by (3.9), and  $W_{\text{out}}$ , given by (3.10).
- Quantile levels for model validation.  $p_1^{\text{out}} = 0.01, p_2^{\text{out}} = 0.03, \dots, p_{49}^{\text{out}} = 0.97, p_{50}^{\text{out}} = 0.99.$
- Sample sizes.  $n = 10^2, 10^3$  (and  $n = 10^6, 10^7, 10^8, 10^9$  for computational time evaluations).
- Number of bootstrap samples.  $B = 10^3$ .
- Number of Monte Carlo runs.  $M = 10^4$ .

#### 4.2 From Small to Big Data

The boxplots of the estimators under consideration are presented in Figures 4.1 (for Normal and Cauchy) and 4.2 (for Exponential and Lévy). Barring a few exceptions, most estimators are correctly calibrated, i.e., they are centered at  $\mu = 0$  for location and  $\sigma = 1$  for scale, and shrink toward the respective targets according to the rate of  $n^{1/2}$ . The latter statement can be illustrated by reporting the values of the ratio  $\sqrt{\text{MSE}}$  (at n = 100)/ $\sqrt{\text{MSE}}$  (at n = 1000) for each estimator, and few selected distributions.

- For Cauchy, the ratios are: 3.08  $(\widehat{\mu}_{MLE})$ , 13.74  $(\widehat{\mu}_{o1})$ , 3.62  $(\widehat{\mu}_{o2})$ , 3.29  $(\widehat{\mu}_{o3})$ , 3.33  $(\widehat{\mu}_{g1})$ , 3.20  $(\widehat{\mu}_{g2})$ , 3.18  $(\widehat{\mu}_{g3})$ ; and 2.52  $(\widehat{\sigma}_{MLE})$ , 16.83  $(\widehat{\sigma}_{o1})$ , 4.10  $(\widehat{\sigma}_{o2})$ , 3.50  $(\widehat{\sigma}_{o3})$ , 3.33  $(\widehat{\sigma}_{g1})$ , 3.36  $(\widehat{\sigma}_{g2})$ , 3.33  $(\widehat{\sigma}_{g3})$ .
- For Normal, the ratios are: 3.14  $(\widehat{\mu}_{MLE})$ , 3.20  $(\widehat{\mu}_{o1})$ , 3.13  $(\widehat{\mu}_{o2})$ , 3.13  $(\widehat{\mu}_{o3})$ , 3.19  $(\widehat{\mu}_{g1})$ , 3.14  $(\widehat{\mu}_{g2})$ , 3.14  $(\widehat{\mu}_{g3})$ ; and 3.16  $(\widehat{\sigma}_{MLE})$ , 3.15  $(\widehat{\sigma}_{o1})$ , 3.13  $(\widehat{\sigma}_{o2})$ , 3.13  $(\widehat{\sigma}_{o3})$ , 3.14  $(\widehat{\sigma}_{g1})$ , 3.13  $(\widehat{\sigma}_{g2})$ , 3.13  $(\widehat{\sigma}_{g3})$ .

Note that according to the asymptotic results of Section 3.2, these ratios are expected to fall around  $\sqrt{1000/100} \approx 3.16$ . The "incorrect" behavior of o1 (and to a lesser degree of o2) for Cauchy is not surprising and can be attributed to its very poor efficiency properties: ARE  $(\hat{\mu}_{o1}, \hat{\mu}_{MLE}) = 0.029$  and ARE  $(\hat{\sigma}_{o1}, \hat{\sigma}_{MLE}) = 0.107$ . Similar conclusions can be drawn for distributions in Figure 4.2.

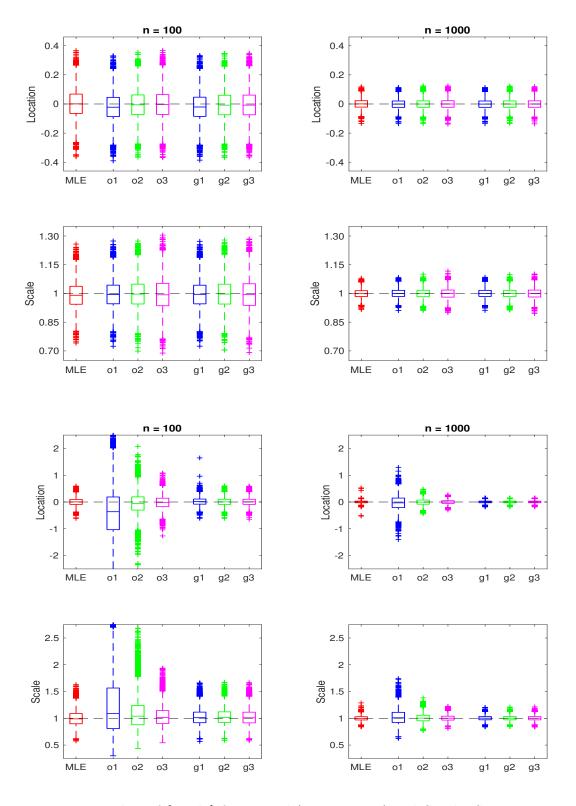


FIGURE 4.1. Boxplots of  $\hat{\mu}$  and  $\hat{\sigma}$  for Normal (top two rows) and Cauchy (bottom two rows) distributions, using MLE, oQLS ('o') and gQLS ('g') estimators, where (a,b) is equal to (0.02,0.98) for o1/g1, (0.05,0.95) for o2/g2, and (0.10,0.90) for o3/g3.

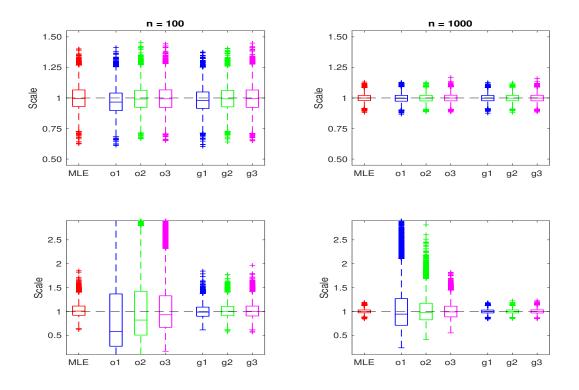


FIGURE 4.2. Boxplots of  $\widehat{\sigma}$  for Exponential (top) and Lévy (bottom) distributions, using MLE, oQLS ('o') and gQLS ('g') estimators, where (a, b) is equal to (0.02, 0.98) for o1/g1, (0.05, 0.95) for o2/g2, and (0.10, 0.90) for o3/g3.

It is also of interest to see how fast these estimators can be computed when sample sizes are very large, which is quite common nowadays. Using MATLAB® R2023a software on a basic laptop (with Apple M2 8-core CPU, RAM 8GB, Mac OS), the MLE, oQLS and gQLS estimators of six location-scale families have been computed for samples of size  $n=10^6,10^7,10^8,10^9$  and their computational times (in seconds) have been recorded in Table 4.1. Note that for all these distributions, oQLS and gQLS have explicit formulas (although they require inversion of medium-sized matrices), and MLE has explicit formulas for four chosen distributions but requires numerical optimization for Cauchy and Logistic. The conclusion is clear: the computational costs for oQLS and gQLS are on a par with the explicit-formula MLEs and at least 10 times less than those of optimization-based MLEs. This computational advantage is highly relevant in situations involving "big data".

Table 4.1. Computational times (in seconds) of MLE, oQLS, and gQLS for large n.

Sample	Estimation		Probability Distribution								
Size	Method	Cauchy	Exponential	Laplace	Lévy	Logistic	Normal				
$n = 10^6$	MLE	0.76	0.02	0.07	0.09	0.76	0.14				
	oQLS	0.05	0.05	0.05	0.10	0.05	0.07				
	gQLS	0.08	0.07	0.06	0.11	0.10	0.10				
$n = 10^7$	MLE	5.32	0.20	0.36	0.51	4.51	0.70				
	oQLS	0.41	0.43	0.32	0.72	0.42	0.52				
	gQLS	0.42	0.48	0.48	0.75	0.47	0.60				
$n = 10^8$	MLE	79.44	1.78	3.19	4.79	84.24	3.52				
	oQLS	5.21	4.14	3.95	6.87	3.84	5.33				
	gQLS	4.58	5.40	4.27	7.36	4.71	5.83				
$n = 10^9$	MLE	**	207	844	555	19716	498				
	oQLS	397	467	585	742	433	788				
	gQLS	406	506	672	732	447	989				

<sup>\*\*</sup> MLE for Cauchy and  $n = 10^9$  did not converge.

## 4.3 Good Data, Bad Data

When the distributional assumption is correct ("clean" or "good" data scenario), the simulated largesample performance of MLE, oQLS, or gQLS is consistent with the asymptotic results of Section 3.2, which was verified in the previous section. When data are contaminated by outliers ("bad" data scenario), all estimators are affected by it, but to a different extent. As is evident from the boxplots of Figure 4.3, the robust QLS-type estimators successfully cope with outliers as long as their breakdown point exceeds the level of contamination  $\varepsilon$ . They work especially well for estimation of  $\mu$  and less so for estimation of  $\sigma$ . Further, for estimation of  $\sigma$  under Normal, MLEs completely miss the target and their variability gets significantly inflated even for the smallest levels of contamination. For Cauchy, which easily accommodates the outliers from Normal ( $\mu^* = 1, \sigma^* = 3$ ), MLEs perform reasonably well. This suggests that to mitigate the effects of potential contamination on MLEs, a prudent approach is to always assume that data follow a heavy-tailed distribution. Of course, if one were to take that route they would have to accept the fact that no mean and other moments exist, and thus all subsequent inference should be based on quantiles. Finally, more simulations have been conducted using Laplace, Logistic, and other distributions. The conclusions were similar to those of Normal and Cauchy: if a light-tailed distribution is assumed, contamination is devastating to MLEs, but a heavier-tailed distribution can correct the MLEs' performance. Those additional studies will not be presented here.

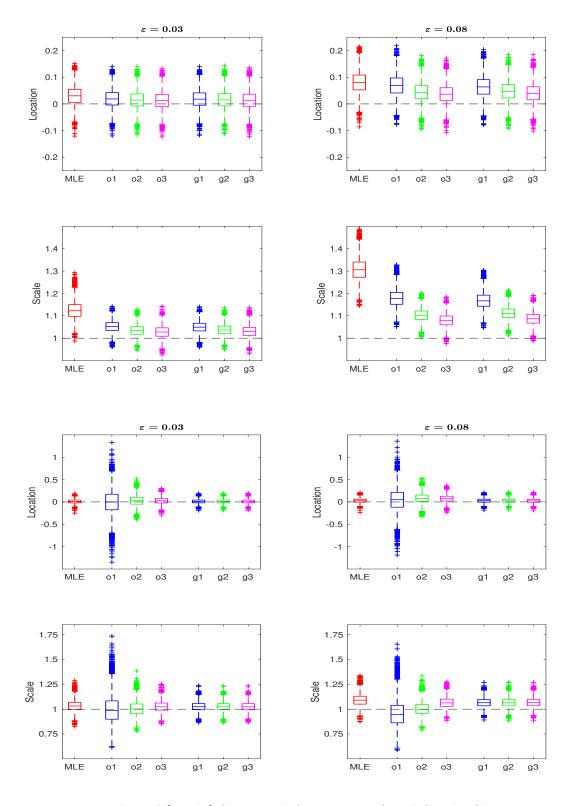


FIGURE 4.3. Boxplots of  $\hat{\mu}$  and  $\hat{\sigma}$  for Normal (top two rows) and Cauchy (bottom two rows) distributions, using MLE, oQLS ('o') and gQLS ('g') estimators, where (a,b) is equal to (0.02,0.98) for o1/g1, (0.05,0.95) for o2/g2, and (0.10,0.90) for o3/g3.

#### 4.4 Goodness of Fit

A simulation study has been conducted to assess the power properties of the goodness-of-fit tests based on statistics W and  $W_{\text{out}}$ . The results are presented in Tables 4.2 (for W) and 4.3 (for  $W_{\text{out}}$ ). The following conclusions emerge from the tables.

- For the test based on W (Table 4.2), the estimated probability of rejecting the null hypothesis approaches 1 as  $n \to 1000$  for most distributions under  $H_0$  and most alternatives. The exceptions are Logistic against Normal and  $F_{0.05}$ , and Normal against Logistic and  $F_{0.05}$ . For n = 100, Cauchy has very low power against all alternatives, and the test designed for Cauchy exceeds its level of 0.05. Comparisons between different levels of (a, b), i.e., (0.02, 0.98) versus (0.05, 0.95) versus (0.10, 0.90), do reveal some patterns. However, recall that choosing one pair of (a, b) versus another means comparing the model fit on two overlapping but different ranges of data.
- For the test based on W<sub>out</sub> (Table 4.3), all model fits are compared on the same set of quantiles, ranging from the level 0.01 to 0.99 (50 quantiles in total). The estimated probability of rejecting the null hypothesis approaches 1 as n → 1000 for most distributions under H<sub>0</sub> and most alternatives. The power of Logistic against Normal and F<sub>0.05</sub> is still low, but higher than that based on W. This time Normal exhibits fairly high power against Logistic and F<sub>0.05</sub>. The patterns among different choices of (a, b) are mixed and depend on H<sub>0</sub> and the alternative distribution. All tests match the significance level of 0.05. Interestingly, for n = 100 the Cauchy-based test has no power at all against any of the selected alternatives.

Table 4.2. Proportions of rejections of  $H_0$  by the goodness-of-fit test (3.9) at  $\alpha = 0.05$  for several distributions under  $H_0$  and  $H_A$ , and varying n. In all cases,  $\mu = 0$  and  $\sigma = 1$ , and  $F_{0.05} = 0.95 \, \text{Normal} \, (\mu = 0, \sigma = 1) + 0.05 \, \text{Normal} \, (\mu^* = 1, \sigma^* = 3)$ .

gQLS	Assumed		I	Data Gener	rated by		
Estimator	Distribution $(H_0)$	Cauchy	Gumbel	Laplace	Logistic	Normal	$F_{0.05}$
Sample Size: $n = 1$	100						
a = 0.02, b = 0.98	Cauchy	0.25	0.08	0.04	0.05	0.07	0.06
	Gumbel	1.00	0.08	0.89	0.66	0.47	0.51
	Laplace	0.96	0.43	0.09	0.21	0.34	0.27
	Logistic	1.00	0.32	0.28	0.09	0.09	0.13
	Normal	1.00	0.46	0.52	0.16	0.08	0.20
a = 0.05, b = 0.95	Cauchy	0.18	0.08	0.04	0.05	0.07	0.06
	Gumbel	0.99	0.07	0.75	0.45	0.32	0.32
	Laplace	0.78	0.37	0.08	0.17	0.26	0.21
	Logistic	0.96	0.29	0.23	0.08	0.08	0.07
	Normal	0.98	0.37	0.38	0.11	0.08	0.09
a = 0.10, b = 0.90	Cauchy	0.14	0.11	0.06	0.08	0.09	0.08
	Gumbel	0.89	0.08	0.56	0.30	0.24	0.23
	Laplace	0.40	0.28	0.09	0.16	0.22	0.18
	Logistic	0.74	0.22	0.20	0.09	0.09	0.08
	Normal	0.82	0.25	0.28	0.10	0.08	0.09
Sample Size: $n = 1$	1000						
a = 0.02, b = 0.98	Cauchy	0.08	1	0.99	1	1	1
	Gumbel	1	0.06	1	1	1	1
	Laplace	1	1	0.06	0.97	1	0.99
	Logistic	1	1	0.98	0.05	0.35	0.18
	Normal	1	1	1	0.57	0.05	0.53
a = 0.05, b = 0.95	Cauchy	0.07	1	0.99	1	1	1
	Gumbel	1	0.05	1	1	1	1
	Laplace	1	1	0.05	0.94	1	1
	Logistic	1	1	0.96	0.05	0.19	0.10
	Normal	1	1	1	0.29	0.05	0.13
a = 0.10, b = 0.90	Cauchy	0.06	1	0.67	1	1	1
	Gumbel	1	0.05	1	1	1	0.99
	Laplace	0.99	1	0.05	0.81	0.98	0.95
	Logistic	1	1	0.86	0.05	0.10	0.07
	Normal	1	1	0.99	0.13	0.05	0.07

Table 4.3. Proportions of rejections of  $H_0$  by the goodness-of-fit test (3.10) at  $\alpha = 0.05$  for several distributions under  $H_0$  and  $H_A$ , and varying n. In all cases,  $\mu = 0$  and  $\sigma = 1$ , and  $F_{0.05} = 0.95 \, \text{Normal} \, (\mu = 0, \sigma = 1) + 0.05 \, \text{Normal} \, (\mu^* = 1, \sigma^* = 3)$ .

gQLS	Assumed		I	Data Gener	rated by		
Estimator	Distribution $(H_0)$	Cauchy	Gumbel	Laplace	Logistic	Normal	$F_{0.05}$
Sample Size: $n = 1$	100						
a = 0.02, b = 0.98	Cauchy	0.05	0	0	0	0	0
	Gumbel	1	0.05	0.83	0.57	0.32	0.52
	Laplace	0.96	0.18	0.05	0.08	0.14	0.15
	Logistic	1	0.12	0.19	0.05	0.04	0.14
	Normal	1	0.25	0.43	0.14	0.05	0.32
a = 0.05, b = 0.95	Cauchy	0.05	0	0	0	0	0
	Gumbel	1	0.05	0.90	0.68	0.41	0.60
	Laplace	0.98	0.13	0.05	0.06	0.10	0.13
	Logistic	1	0.10	0.21	0.05	0.03	0.15
	Normal	1	0.26	0.53	0.19	0.05	0.37
a = 0.10, b = 0.90	Cauchy	0.05	0	0	0	0	0
	Gumbel	1	0.05	0.94	0.75	0.46	0.65
	Laplace	0.99	0.07	0.06	0.03	0.04	0.07
	Logistic	1	0.09	0.28	0.06	0.02	0.13
	Normal	1	0.28	0.66	0.24	0.05	0.38
Sample Size: $n = 1$	1000						
a = 0.02, b = 0.98	Cauchy	0.05	1	0.48	1	1	1
	Gumbel	1	0.05	1	1	1	1
	Laplace	1	1	0.05	0.89	1	0.99
	Logistic	1	1	0.95	0.06	0.24	0.45
	Normal	1	1	1	0.59	0.05	0.87
a = 0.05, b = 0.95	Cauchy	0.05	1	0.47	1	1	1
	Gumbel	1	0.05	1	1	1	1
	Laplace	1	1	0.05	0.87	1	0.99
	Logistic	1	1	0.96	0.05	0.21	0.46
	Normal	1	1	1	0.67	0.05	0.90
a = 0.10, b = 0.90	Cauchy	0.05	1	0.44	0.99	1	1
	Gumbel	1	0.05	1	1	1	1
	Laplace	1	1	0.05	0.82	1	0.98
	Logistic	1	1	0.98	0.05	0.18	0.43
	Normal	1	1	1	0.76	0.05	0.91

# 5 Real Data Examples

To illustrate how our proposed estimators and goodness-of-fit tests work on real data, we will use the daily stock returns of Alphabet Inc., the parent company of *Google*, for the period from January 2, 2020, to November 30, 2023. The stock prices are available at the *Yahoo! Finance* website https://finance.yahoo.com/quote/GOOG/. The daily returns are calculated as the difference between the *closing* prices on two consecutive trading days. Below are summary statistics for these data.

Diagnostic tools such as histogram, quantile-quantile plot, and probability-probability plot were employed and revealed that a symmetric and (approximately) bell-shaped distribution might be a suitable candidate for the data set at hand. In view of this, all the distributions of Section 4.4 were fitted to the daily returns using gQLS and their fits were formally validated with the goodness-of-fit tests (3.9) and (3.10). Table 5.1 summarizes the findings of this analysis.

Table 5.1. Parameter estimates and goodness-of-fit statistics for several location-scale families fitted to the daily returns of the Google stock (1/02/2020 - 11/30/2023).

gQLS Estimator	Assumed	Paramete	er Estimates	Goodness-of-	Fit Statistics
(with $k = 25$ )	Distribution	$\widehat{\mu}$	$\widehat{\sigma}$	W (p-value)	$W_{\mathrm{out}}$ (p-value)
a = 0.02, b = 0.98	Cauchy	0.16	0.95	74.68 (0.00)	84.85 (0.01)
	Gumbel	-0.70	1.61	259.05 (0.00)	324.80 (0.00)
	Laplace	0.15	1.28	50.60 (0.00)	68.11 (0.04)
	Logistic	0.11	0.92	<b>29.06</b> (0.18)	45.21 (0.64)
	Normal	0.08	1.62	50.64 (0.00)	72.13 (0.02)
a = 0.05, b = 0.95	Cauchy	0.15	0.95	71.67 (0.00)	84.87 (0.01)
	Gumbel	-0.64	1.51	165.38 (0.00)	390.07 (0.00)
	Laplace	0.15	1.30	45.26 (0.00)	66.81 (0.05)
	Logistic	0.16	0.91	<b>25.71</b> (0.31)	45.62 (0.64)
	Normal	0.09	1.58	35.06 (0.05)	78.51 (0.01)
a = 0.10, b = 0.90	Cauchy	0.15	0.96	61.54 (0.00)	83.77 (0.02)
	Gumbel	-0.57	1.43	114.11 (0.00)	475.43 (0.00)
	Laplace	0.15	1.33	39.27 (0.02)	64.80 (0.08)
	Logistic	0.12	0.91	24.58 (0.37)	45.93 (0.62)
	Normal	0.10	1.54	<b>29.45</b> (0.17)	87.83 (0.00)

As is evident from the table, the Logistic distribution provides the best fit among the candidate models, with its p-values significantly exceeding the 0.10 level under both tests. Note that the more robust estimators (i.e., those with higher a = 1 - b) achieve better fits according to the chi-square test (3.9). This pattern is particularly evident in the case of the Normal distribution and a = 0.10, b = 0.90, but it is not surprising because as a = 1 - b increases the quantile range for which residuals are calculated shrinks making the fit better. On the other hand, the test (3.10) computes residuals on the universal set of 50 quantile levels (from 0.01 to 0.99) and practically shows no sensitivity to the choice of a and b.

# 6 Concluding Remarks

In this paper, two types of quantile least squares estimators for location-scale families have been introduced: ordinary (denoted oQLS) and generalized (denoted gQLS). Both approaches are robust. While the oQLS estimators are quite effective for more general probability distributions, the gQLS estimators can match the levels of robustness of oQLS yet offer much higher efficiency for estimation of location and/or scale parameters. These properties have been derived analytically (for large n) and verified using simulations (for small and medium-sized samples). In addition, two goodness-of-fit tests have been contructed and their power properties have been investigated via simulations. Also, it has been established that computational times of these estimators are similar to those of explicit-formula MLEs, and are more than 10 times lower when MLEs have to be found using numerical optimization. For example, oQLS and gQLS can be computed for a sample of a billion observations in 7-15 minutes while non-explicit MLEs take more than 5 hours or do not converge at all.

The research presented in this paper can be extended in several directions. The most obvious one is to develop QLS estimators for more general classes of distributions. Another direction (perhaps less obvious) is to follow the literature on L-moments (Hosking, 1990) and trimmed L-moments (see Elamir and Seheult (2003); Hosking (2007)) and construct QLS-type statistics to summarize the shapes of parametric distributions. This line of research is more probabilistic in nature. On the statistical side, direct comparisons with the MTM estimators of Brazauskas  $et\ al.\ (2009)$ , MWM estimators (see Zhao  $et\ al.\ (2018a)$ ; Zhao  $et\ al.\ (2018b)$ ), or even more general trimming methods such as trimmed likelihood estimators of Neykov  $et\ al.\ (2007)$  are also of interest.

# Acknowledgments

Declarations of interest: none.

# References

- Ali, M.M. and Umbach, D. (1989). A Shapiro-Wilk type goodness-of-fit test using a few order statistics. Journal of Statistical Planning and Inference, 22(2), 251–261.
- Ali, M.M. and Umbach, D. (1998). Optimal linear inference using selected order statistics in location-scale models. In *Handbook of Statistics, Vol. 17; Order Statistics: Applications* (N. Balakrishnan and C.R. Rao, eds.); 183–213. Elsevier Science B.V.
- Brazauskas, V., Jones, B., and Zitikis, R. (2009). Robust fitting of claim severity distributions and the method of trimmed moments. *Journal of Statistical Planning and Inference*, **139**(6), 2028–2043.
- Brazauskas, V. and Serfling, R. (2000). Robust and efficient estimation of the tail index of a single-parameter Pareto distribution (with discussion). *North American Actuarial Journal*, **4**(4), 12–27. Discussion: **5**(3), 123–126. Reply: **5**(3), 126–128.
- Cane, G.J. (1974). Linear estimation of parameters of the Cauchy distribution based on sample quantiles. Journal of the American Statistical Association, **69**(345), 243–245.
- Chan, L.K. (1970). Linear estimation of the location and scale parameters of the Cauchy distribution based on sample quantiles. *Journal of the American Statistical Association*, **65**(330), 851–859.
- Elamir, E.A.H. and Seheult, A.H. (2003). Trimmed L-moments. Computational Statistics and Data Analysis, 43(3), 299–314.
- Genton, M.G. and de Luna, X. (2000). Robust simulation-based estimation. *Statistics and Probability Letters*, **48**(3), 253–259.
- Genton, M.G. and Ronchetti, E. (2003). Robust indirect inference. *Journal of the American Statistical Association*, **98**(461), 67–76.
- Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J., and Stahel, W.A. (1986). Robust Statistics: The Approach Based on Influence Functions. Wiley.
- Hampel, F.R. (1968). Contributions to the Theory of Robust Estimation. Ph.D. thesis, University of California, Berkeley.
- Hogg, R.V., McKean, J.W., and Craig, A.T. (2005). *Introduction to Mathematical Statistics*, 6th edition. Pearson Education.
- Hosking, J.R.M. (1990). L-moments: Analysis and estimation of distributions using linear combinations of order statistics. Journal of the Royal Statistical Society: Series B, **52**(1), 105–124.
- Hosking, J.R.M. (2007). Some theory and practical uses of trimmed *L*-moments. *Journal of Statistical Planning and Inference*, **137**(9), 3024–3039.
- Huber, P.J. (1964). Robust estimation of a location parameter. *Annals of Mathematical Statistics*, **35**(1), 73–101.
- Huber, P.J. and Ronchetti, E.M. (2009). Robust Statistics, 2nd edition. Wiley.
- Maronna, R.A., Martin, R.D., and Yohai, V.J. (2006). Robust Statistics: Theory and Methods. Wiley.

- Mosteller, F. (1946). On some useful "inefficient" statistics. *Annals of Mathematical Statistics*, **17**(4), 377–408.
- Neykov, N., Filzmoser, P., Dimova, R., and Neytchev, P. (2007). Robust fitting of mixtures using the trimmed likelihood estimator. *Computational Statistics and Data Analysis*, **52**(1), 299–308.
- Sarhan, A.E. and Greenberg, B.G. (1962). Contributions to Order Statistics. Wiley.
- Serfling, R.J. (2002a). Approximation Theorems of Mathematical Statistics. Wiley.
- Serfling, R. (2002b). Efficient and robust fitting of lognormal distributions (with discussion). *North American Actuarial Journal*, **6**(4), 95–109. Discussion: **7**(3), 112–116. Reply: **7**(3), 116.
- Tukey, J.W. (1960). A survey of sampling from contaminated distributions. In *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling* (I. Olkin *et al.*, eds.), Stanford University Press, 448–485.
- Xu, Y., Iglewicz, B., and Chervoneva, I. (2014). Robust estimation of the parameters of g-and-h distributions, with applications to outlier detection. Computational Statistics and Data Analysis, 75(July 2014), 66–80.
- Zhao, Q., Brazauskas, V., and Ghorai, J. (2018a). Robust and efficient fitting of severity models and the method of Winsorized moments. ASTIN Bulletin, 48(1), 275–309.
- Zhao, Q., Brazauskas, V., and Ghorai, J. (2018b). Small-sample performance of the MTM and MWM estimators for the parameters of log-location-scale families. *Journal of Statistical Computation and Simulation*, 88(4), 808–824.

