PLISKA

STUDIA MATHEMATICA BULGARICA

ПЛИСКА

БЪЛГАРСКИ МАТЕМАТИЧЕСКИ СТУДИИ

Pliska Stud. Math. Bulgar. **17** (2005), 187–200

PLISKA
STUDIA MATHEMATICA
BULGARICA

TRIMMED LIKELIHOOD ESTIMATION OF THE PARAMETERS OF THE GENERALIZED EXTREME VALUE DISTRIBUTIONS: A MONTE-CARLO STUDY

Neyko Neykov[1], Rositsa Dimova[2], Plamen Neytchev

The applicability of the Trimmed Likelihood Estimator (TLE) proposed by Neykov and Neytchev [14] to the extreme value distributions is considered. The effectiveness of the TLE in comparison with the classical MLE in the presence of outliers in various scenarios is illustrated by an extended simulation study. The FAST-TLE algorithm developed by Neykov and Müller [13] is used to get the parameter estimate. The computations are carried out in the R environment using the packages *ismev* originally developed by Coles [5] and ported in R by Stephenson [16].

1. Introduction

The extreme value distributions theory has been intensively developed. The book of Coles [5] provides a useful theoretical background. The Maximum Likelihood is the standard technique for statistical inference in extremes. It is well known that the MLE can be very sensitive to outliers in the data. Indeed, the simulation study of Barão and Tawn [2] shows that in the presence of outliers, the parameter estimates are significantly influenced and thus the return period. Relatively

[1]The work of N. Neykov and P. Neytchev is partially supported by the Ministry of Education, Science and Culture of Republic of Bulgaria, grant MM 1103/2001.

[2]Partialy supported by Pro-ENBIS GTC1 -2001-43031.

2000 *Mathematics Subject Classification:* Primary 62F35; Secondary 62P99

Key words: Generalized extreme value distribution, maximum likelihood estimation, trimmed likelihood estimation, Monte-Carlo simulation.

little attention to robustness has been paid in the context of extreme values. To overcome this problem Dupuis and Field [7], Dupuis and Morgenthaler [8], and Dupuis and Tawn [9] estimate robustly the parameters of various extreme value distributions using the so called B-optimal robust M-estimators of Hampel et al. [10]. It is concluded that these estimators are more efficient than MLE under some model assumptions violation. Unfortunately, these estimators do not possess a high Breakdown Point (BP) and hence are not appropriate for the modeling purposes, as with the increasing the number p of the explanatory variables their BP decreases to zero as $1/p$. (Roughly speaking, the BP is the smallest fraction of contamination that can cause the estimator to take an arbitrarily large value.) In practice, one needs robust estimators that possess a high BP resistant against high percentage of surrogate (aberrant, anomalous) observations in data. For instance, such observations arise when data are collected by different ways.

Several parametric robust alternatives of the ML estimator possessing high BP have been developed, e.g., Choi et al. [3], Markatou et al. [11], Neykov and Neytchev [14], and Windham [20]. To our knowledge, none of these high BP estimators has been used for the purposes of the extreme value modeling. Thus, the main goal of the paper is to develop a robust parametric approach for extreme values statistical modeling based on the TLE proposed by Neykov and Neytchev [14]. The TLE is looking for that sub-sample of k observations out of n, the original data size, with the optimal likelihood. The trimming number of observations can be chosen by the user in appropriate bounds to get a high BP and optimal efficiency. Details about the properties of the TLE can be found in Vandev and Neykov [18], Vandev and Neykov [19], Neykov and Müller [13], Cizek [4], Müller and Neykov [12], and Dimova and Neykov [6]. Because the TLE accommodates the classical MLE, the extreme value methodology, which is based mainly on the MLE, can be adapted and further developed.

In this paper we consider an application of the TLE to the Generalized Extreme Value (GEV) distribution, however, the generalized Pareto distribution or the Poison point approaches for modeling of extreme values can be used instead. A simulation study is performed to illustrate the effectiveness of the TLE in comparison with the MLE.

2. Basic definitions and notions

In the following, the GEV distribution is introduced. It arises as the limiting distribution of the maxima of a series of independent and identically distributed (i.i.d.) observations. The distribution function of the GEV is given by

$$G(x;\mu,\sigma,\xi) = \begin{cases} \exp\left\{-\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\xi}\right\} & \text{if} \quad \xi \neq 0, \\ \exp\left\{-\exp\left[-\left(\frac{x-\mu}{\sigma}\right)\right]\right\} & \text{if} \quad \xi = 0. \end{cases} \tag{1}$$

where $\{x : 1+\xi(x-\mu)/\sigma > 0\}$, $\sigma > 0$, and μ, σ, ξ are location, scale and shape parameters, respectively, see Coles (2001).

The Fréchet and Weibull distributions are obtained for $\xi < 0$ and $\xi > 0$, respectively. The case of $\xi = 0$ is interpreted as the limit of the GEV as $\xi \to 0$, widely known as the Gumbel distribution. The MLE is completely regular if $\xi > -0.5$, it exists but is not completely regular if $-1 < \xi < -0.5$ and it does not exist if $\xi < -1$, according to Smith [15].

We now recall the definition of the Trimmed Likelihood Estimator. Let $x_1, \ldots, x_n$ be i.i.d. observations with density function $f(x,\theta)$, depending on unknown parameter θ and $l(x_i,\theta) = -\log f(x_i,\theta)$.

Definition 1. *The Trimmed Likelihood Estimator (TLE) is defined in Neykov and Neytchev [14] as*

$$\hat{\theta} := \arg\min_{\theta\in\Theta} \sum_{i=1}^{k} l(x_{\nu(i)},\theta), \tag{2}$$

where $l(x_{\nu(1)},\theta) \le l(x_{\nu(2)},\theta) \le \ldots \le l(x_{\nu(n)},\theta)$ *are the ordered values of* $l(x_i,\theta)$ *for* $i = 1,\ldots,n$ *at* θ, $\nu = (\nu(1),\ldots,\nu(n))$ *is the corresponding permutation of the indexes, which depends on* θ *and* k *is the trimming parameter.*

The basic idea behind the trimming in this estimator is in removal of those $n-k$ observations which values would be highly unlikely to occur, had the fitted model been true. The TLE coincides with the MLE if $k = n$. Due to the representation

$$\min_{\theta\in\Theta} \sum_{i=1}^{k} l(x_{\nu(i)},\theta) = \min_{\theta\in\Theta} \min_{I\in I_k} \sum_{i\in I} l(x_i,\theta) = \min_{I\in I_k} \min_{\theta\in\Theta} \sum_{i\in I} l(x_i,\theta)$$

where I_k is the set of all k–subsets of the set $\{1,\ldots,n\}$, it follows that all possible $\binom{n}{k}$ partitions of the data have to be fitted by the MLE. Therefore, the TLE is given by the partition with that MLE fit for which the negative log likelihood is minimal.

General conditions for the existence of a solution of (2) can be found in Dimova and Neykov [6], whereas the consistency is proved in Cizek [4]. The BP of the TLE is studied by Vandev and Neykov [19], Atanasov and Neykov [1], and Müller and Neykov [12] using the d–fullness technique proposed by Vandev [17]. According to Vandev [17], the set $F = \{l(x_i, \theta), i = 1, \ldots, n\}$ is called d-full if for any subset of cardinality d of F, the supremum of this subset is a subcompact function. A real valued function $g(\theta)$ is called subcompact if the sets $L_{g(\theta)}(C) = \{\theta : g(\theta) \leq C\}$ are compact for any constant C. The BP of the TL is not less than $\frac{1}{n} \min\{n - k, k - d\}$ if the corresponding set of negative loglikelihoods is d–full (see, Müller and Neykov [12]). It is easy to show that in case of Gumbel distribution $d = 2$. When the location parameter is a monotone function of a linear predictor, $\mu = h(z_i^T \beta)$, where $\beta \in R^p$ is unknown parameter and $Z := (z_i^\top)$ is the data matrix of rank p of the explanatory variables $z_i \in R^p$, then $d = p + 1$. Determination of the d–fullness parameter for the Fréchet and Weibull distributions is not considered because of the complexity of parameters' domain.

Increasing k, the estimator will possess a BP point less than the highest possible, but it will be more efficient at the same time.

Computation of the TLE is infeasible for large data sets because of its combinatorial nature. To get approximate TLE an algorithm called FAST-TLE is developed in Neykov and Müller [13]. It reduces to the FAST-LTS or FAST-MCD algorithms in the normal regression or multivariate Gaussian cases. The basic idea behind the FAST-TLE algorithm consists of carrying out finitely many times a two-step procedure: a trial step followed by a refinement step. In the trial step a subsample of size k^* is selected randomly from the data sample and then the model is fitted to the subsample to get a trial MLE. The refinement step is based on the so called concentration procedure. The cases with the k smallest negative log likelihoods from the trial fit are found. Fitting the model to these k cases gives an improved fit. Repeat of the improvement step yields an iterative process. Convergence is always guaranteed after a finite number of steps since there are only finitely many k–subsets out of $\binom{n}{k}$ in all. The estimate with the lowest TL objective function is retained. There is no guarantee that this value will be the global minimizer but one can hope that it would be a close approximation to it. The trial subsample size k^* should be greater than or equal to $p + 1$ which is needed for the existence of the MLE but the chance to get at least one outlier free subsample is larger if $k^* = p+1$. Any k within the interval [p+1, n] can be chosen in the refinement step. A recommendable choice of k is $\lfloor (n + p + 1)/2 \rfloor$ because then the BP of the TLE is maximized, where $\lfloor r \rfloor := \max\{n \in N; n \leq r\}$. We note that, if the data set is small, all possible subsets of size k can be considered.

3. Simulation design

We compare the performance of the MLE and the TLE through a simulation study for a range of different situations of GEV generated data sets. The regular data follow the model

$$y_i \sim \text{GEV}(\mu_i = 1 + x_i, \sigma = 1, \xi = 0.3), \text{ where } x \sim N(0, 7).$$

The outliers follow the model

$$y_i \sim \begin{cases} \text{U}(y_{max} + \mu_i, y_{max} + (y_{max} - y_{min}) + \mu_i) & \text{if} \quad x_i \geq \bar{x}, \\ \text{U}(y_{min} - (y_{max} - y_{min}) - \mu_i, y_{min} - \mu_i) & \text{if} \quad x_i < \bar{x}. \end{cases} \tag{3}$$

The regular observations and outliers union comprises the contaminated sample of size $n = 100$. Thus, samples with levels of contamination 0%, 10%, 20%, 30% and 40% are considered. The trimming percentage $\frac{n-k}{n}100\%$ is held fixed at 0%, 5%, ..., 45%. The MLE and TLE are computed over the regular and contaminated data. These estimators are compared using the mean, median, root mean square error and various quantiles criteria over 400 independent replications of the simulation experiment at any contamination level.

All the computations were carried out in the R environment using the *ismev* package originally developed by Coles (2001) [5] in S-Plus and ported in R by Stephenson [16].

4. Simulation results

On all plots in Figures 1-4, the data sets, that constitute the regular observations, are represented by bullets, while the outliers, if any, are represented by triangles. The ML and TL fits are based on the contaminated samples. Exceptions are the upper left plots where the ML fits are based only on the regular observations. The dashed lines describe the generated model, whereas the straight lines describe the ML and TL fits in all of these plots. Due to space limitations, only some selected TL fits are presented. In all other plots, the empty triangles or tiny circles describe the trimmed observations (either regular or outliers). The two numbers in the plots' title represent the (trimmed)log-likelihood value of the current estimate and the (trimmed)log-likelihood value evaluated over the regular data at this estimate.

It is a well known fact that the MLE can easily be influenced by a single bad observation, whereas the TLE is resistant up to $\frac{n-k}{n}100\%$ percentage of outliers. To explore the behavior of the estimators the simulation experiment was replicated more than 400 times. As a consequence, a series of estimates were obtained and their distribution was studied.

The plot panels on Fig. 1-4 represent some of these experiments. Generally, the plots indicate that the MLE becomes completely useless if the percentage of observations that do not follow the model is large, while the TLE gives better fits. However, the quality of the TLE fits depends on the trimming percentage $\frac{n-k}{n}100\%$. As it could be expected, the TL estimates are more stable for those values of k that satisfy the inequality $\frac{n-k}{n}100\% \geq \alpha$, where α is the contamination level. The series of box-plots in the "Intercept", "Slope", "Scale" and "Shape" panels on Figure 5 give a more detailed characterization of the distribution of the GEV parameters' estimates conditional on the different trimming percentages. Any of these box-plots series exhibits some specific properties that could serve as a guide to get an idea about the optimal trimming percentage. One can see that the box-plots variation in any panel becomes more stable by increasing the trimming percentage. A large percentage of trimming exhibits great influence on the scale and shape estimates. This is because the relationship between the trimming percentage and the scale estimate is inversely proportional while it is generally nonlinear for the shape parameter estimate. For instance, even a single outlier can drive the MLE scale estimate to infinity. However, increasing the trimming percentage leads to underestimating of the scale. So, trimming with large percentages outside the location-scale distributions framework should be done with great care. It can be seen that there is a common interval of trimming percentages where the parameters estimates become more stable in all these panels ("Intercept", "Slope", "Scale" and "Shape"). Therefore, an optimal choice of the trimming percentage could be the minimal value of that interval.

Usually, the percentage of outliers in real data is unknown. Therefore, one can proceed by a TLE, based on a decreasing range of values for k, starting with $k = n$. However, the TL estimation procedure must be repeated several times at any particular value of k. When the parameters estimate stabilization occurs, then following the previous recommendations on the trimming choice, not only the unknown GEV parameters but also the outliers percentage in the data can be estimated robustly.

5. Summary and conclusions

The simulation study demonstrates that the TLE is a useful alternative to the MLE in the framework of extreme value modeling. The extreme values data can be analyzed with the TLE methodology just as with the classical MLE, however, over sub-samples. Therefore, the computation can be carried out by a standard MLE procedure for fitting extreme value distributions to data closely following the FAST-TLE algorithm of Neykov and Müller [13]. Such procedures are widely

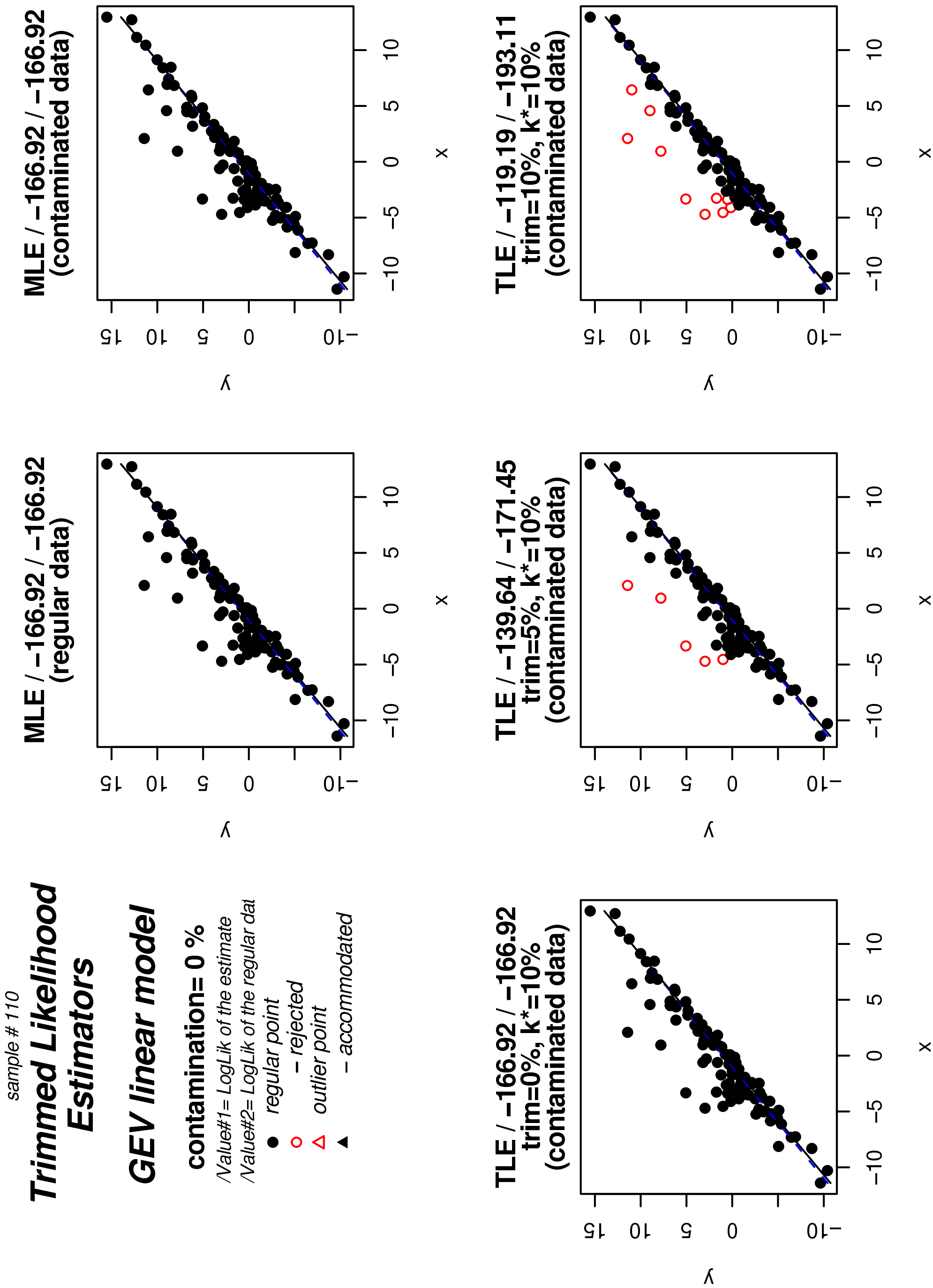

Figure 1: Experiments with zero percentage of contamination.

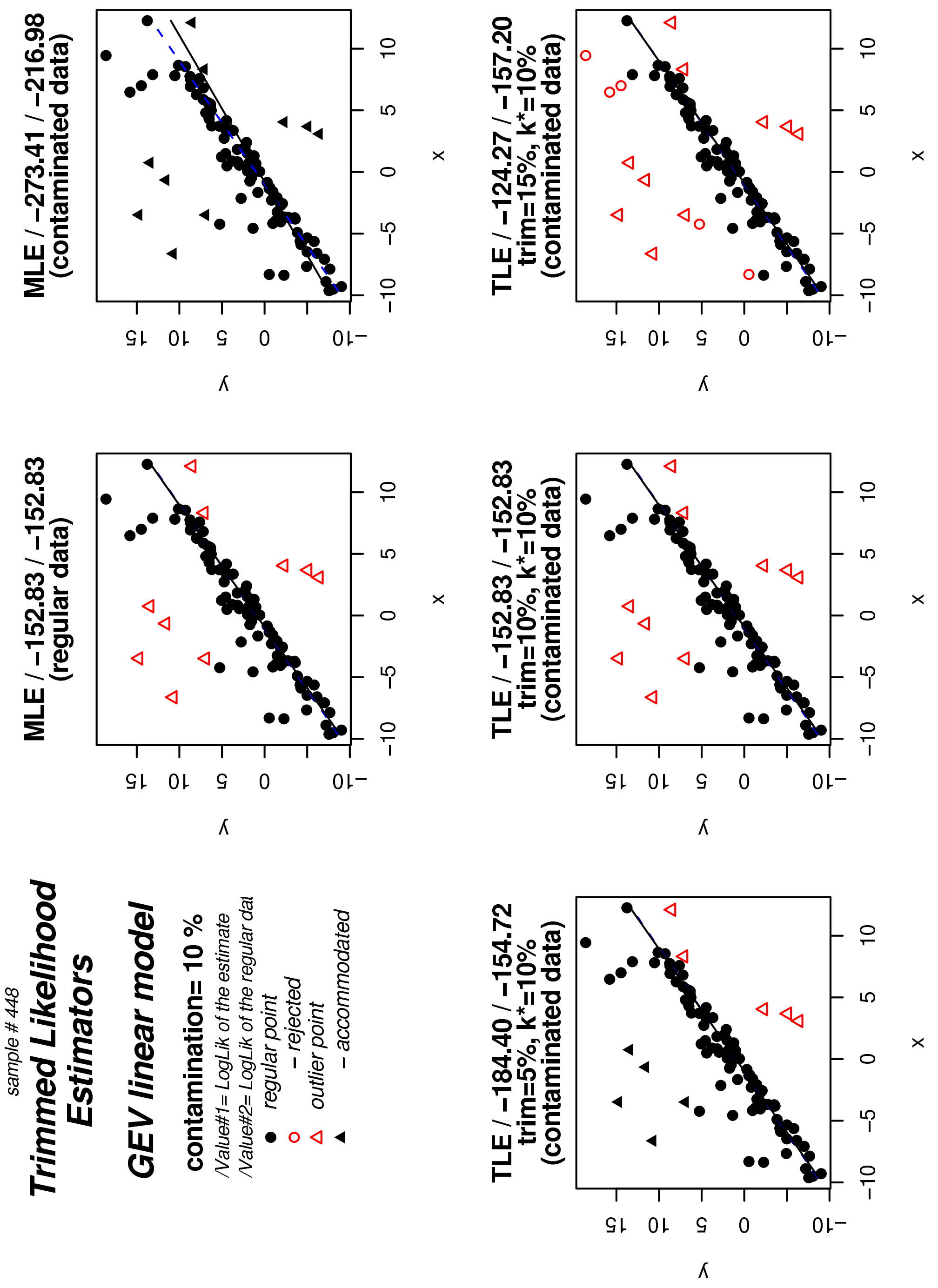

Figure 2: Experiments with 10% of contamination.

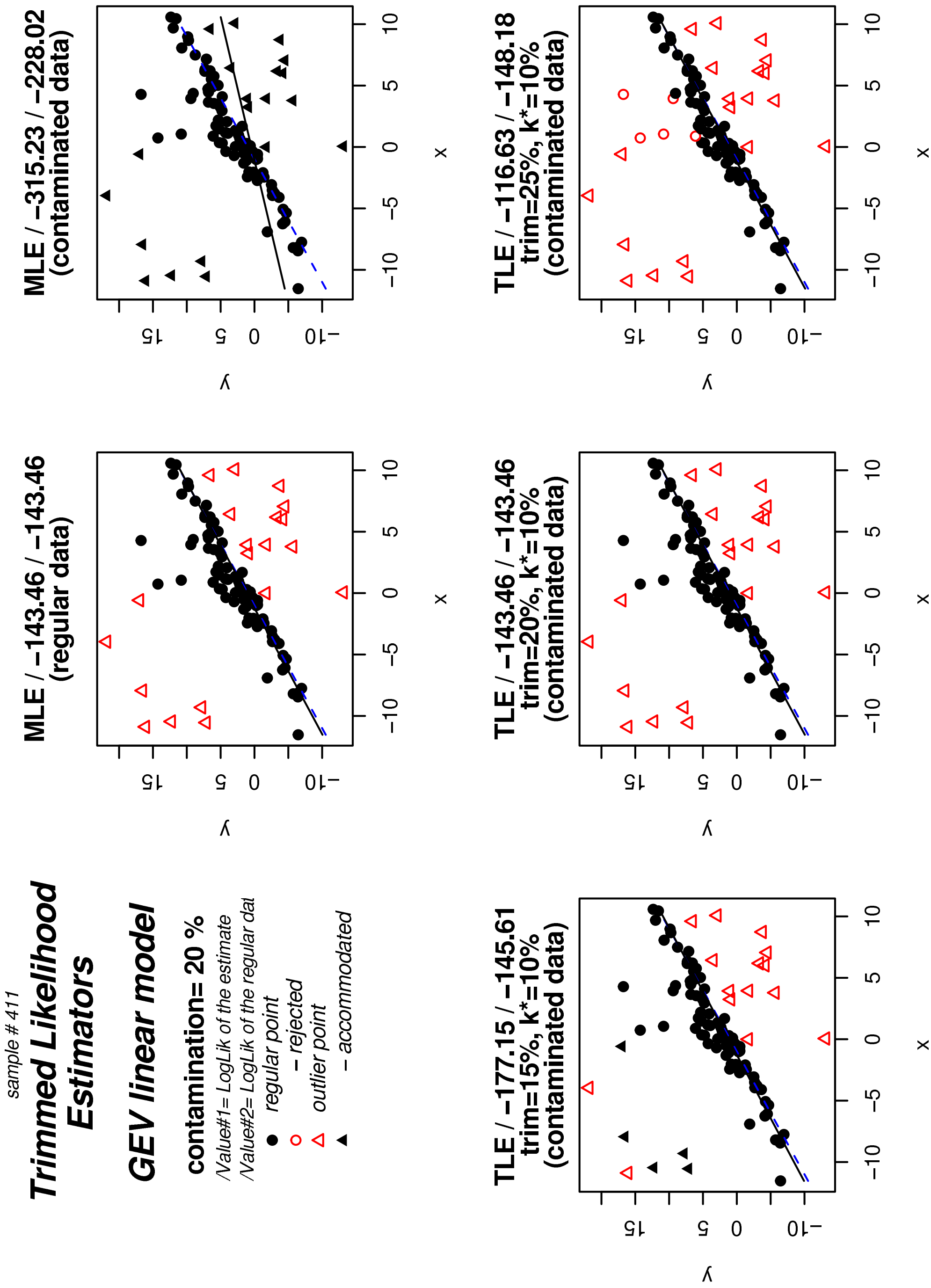

Figure 3: Experiments with 20% of contamination.

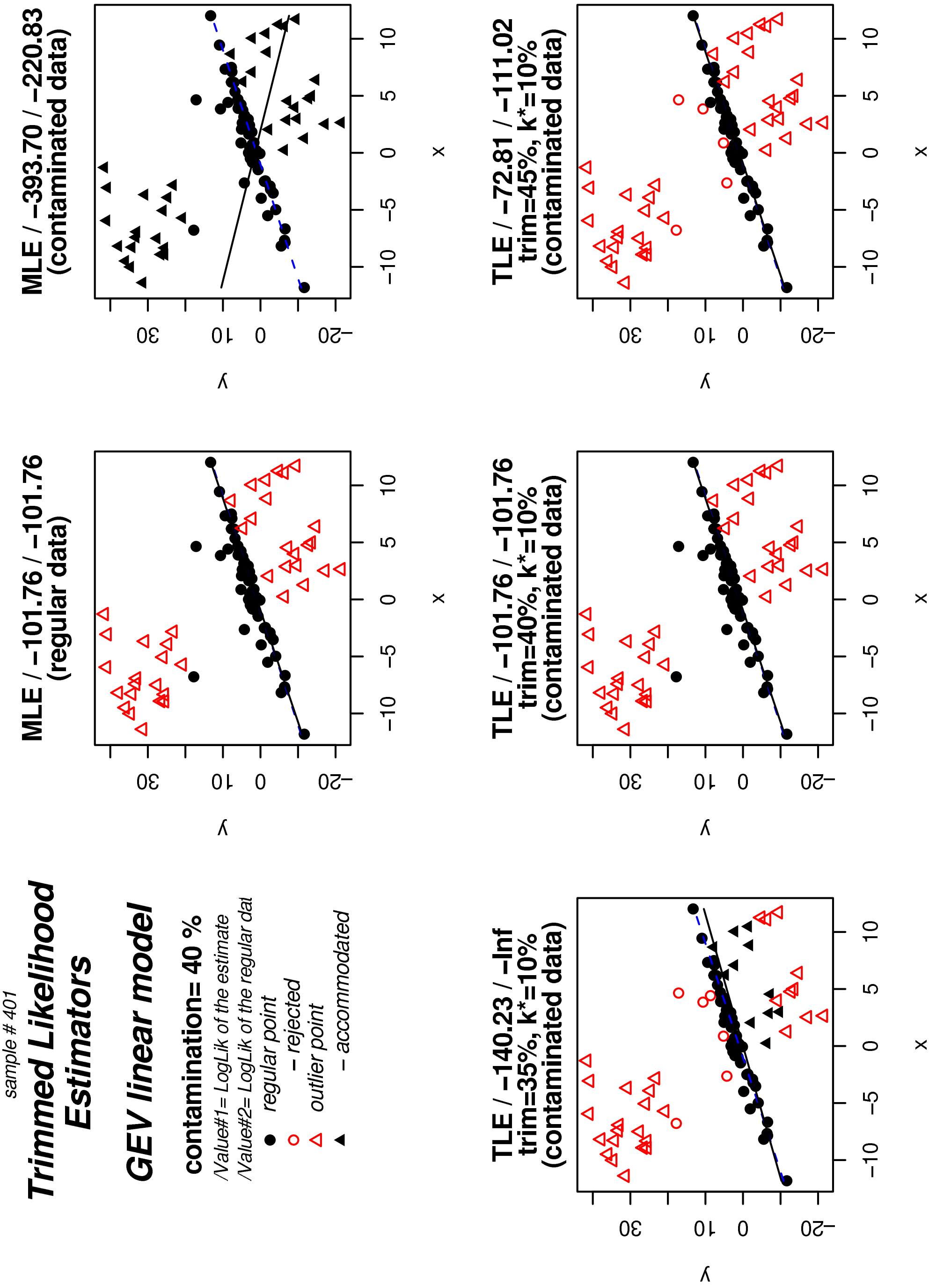

Figure 4: Experiments with 40% of contamination.

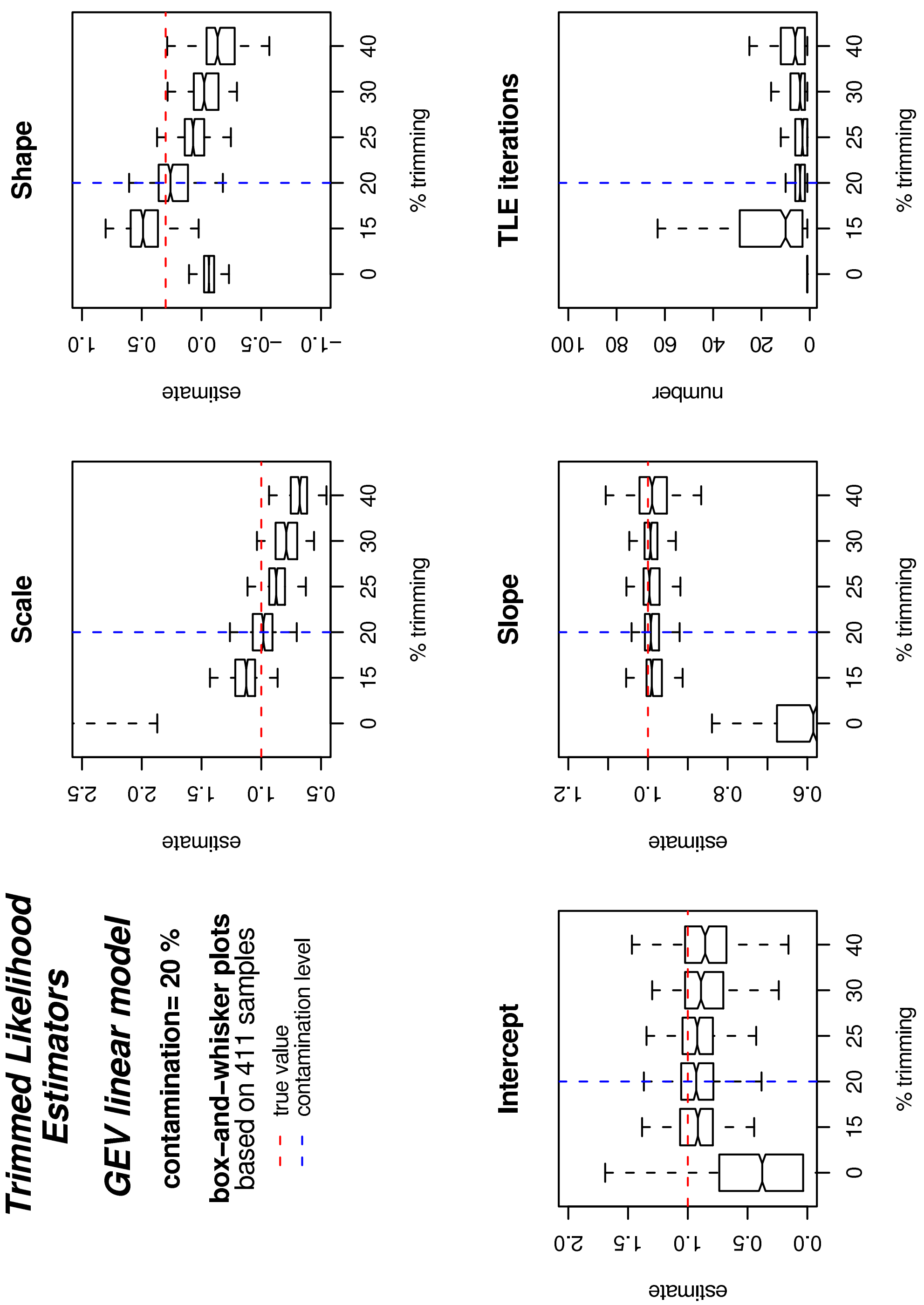

Figure 5: Distribution of the GEV estimates of location (intercept and slope), scale and shape parameters based on 411 experiments.

available in software packages such as S-PLUS, R, SAS. The TLE will lead to greater computational effort, but having in mind the growing power of modern-day processors and memory, one can afford it.

REFERENCES

[1] Atanasov, D.V., N. M. Neykov On the finite sample breakdown point of the weighted trimmed likelihood estimators and the d–fullness of a set of continuous functions. In S. Aivazian et al., editors, *Proc. of the VI Intern. Conf. "Computer Data Analysis and Modelling"*, Vol **1**, 52-57,Minsk, Belarus, 2001.

[2] Barão, M.I., J.A. Tawn Extremal analysis of short series with outliers: Sea-levels and athletic records. *Appl. Statist.* **48** (1999), 469–487.

[3] Choi, E.P., P. Hall and B. Presnell Rendering parametric procedures more robust by empirically tilting the model. *Biometrika* **87** (2000), 453–465.

[4] P. Cizek Robust estimation in nonlinear regression and limited dependent variable models. 2002, http://econpapers.hhs.se/paper/wpawuwpem/0203003.htm

[5] Coles, S.G. An introduction to statistical modeling of extreme values. Springer-Verlag, London, 2001.

[6] Dimova, R.B. and N.M. Neykov Generalized d-fullness Technique fo Breakdown Point Study of the Trimmed Likelihood Estimator with Application. In M. Hubert et al. eds., Theory and Applications of Recent Robust Methods, Birkhaeuser, Basel 2004

[7] Dupuis, D.J., C.A. Field Robust estimation of extremes. *The Canadian J. of Statistics,* **26** (1998), 199–215.

[8] Dupuis, D.J., S. Morgenthaler Robust weighted likelihood estimators with an application to bivariate extreme value problems. *The Canadian J. of Statistics* **30** (2002), 19–31.

[9] Dupuis, D.J., J.A. Tawn Efects of misspecification in bivariate extreme value problems. *Extremes* **4** (2001), 315–330.

[10] Hampel, F. R., E. M. Ronchetti, P. J. Rousseeuw and W. A. Stahel Robust statistics. The approach based on influence functions. John Wiley & Sons, Inc, 1986.

[11] M. Markatou, A. Basu and B. C. Lindsay Weighted likelihood estimating equations: The discrete case with applications to logistic regression. *J. Statist. Plann. Inference.* **57** (1997), 215–232.

[12] Müller, Ch.H., N.M. Neykov Breakdown points of the trimmed likelihood and related estimators in generalized linear models. *J. Statist. Plann. Inference.* **116** (2003), 503–519.

[13] Neykov, N. M., Ch. Müller Breakdown point and computation of trimmed likelihood estimators in generalized linear models. In R. Dutter et al., editors, *Developments in Robust Statistics* Physica-Verlag, Heidelberg, 2002, 277–286.

[14] Neykov, N.M., P.N. Neytchev A robust alternative of the maximum likelihood estimator. In Short communications of COMPSTAT'90, Dubrovnik, 1990, 99–100.

[15] Smith, R.L. Maximum Likelihood estimation in a class of non regular cases. *Biometrika* **72** (1985), 67–90.

[16] Stephenson, A.G. EVD: Extreme Value Distributions. *R-News* **2** (2002), 31–32. URL http://CRAN.R-project.org/doc/Rnews/

[17] Vandev, D.L. A note on breakdown point of the least median squares and least trimmed squares. *Statistics and Probability Letters* **16** (1993), 117–119.

[18] Vandev, D.L., N. M. Neykov Robust maximum likelihood in the Gaussian case. In: New Directions in Data Analysis and Robustness, S. Morgenthaler, E. Ronchetti and W. A. Stahel, (eds.), Birkhäuser Verlag, Basel, 1993, 259–264

[19] Vandev, D. L., N.M. Neykov About regression estimators with high breakdown point. *Statistics* **32** (1998), 111–129.

[20] Windham, M.P. Robustifying model fitting. *J. Roy. Statist. Soc. Ser. B* **57** (1995), 599–609.

Neyko Neykov, Plamen Neytchev
National Institute of
Meteorology and Hydrology,
BAS,
66 Tsarigradsko chaussee, 1784
Sofia, Bulgaria,
e-mail: neyko.neykov@meteo.bg,
plamen.neytchev@meteo.bg

Rositsa Dimova
Sofia University
"St.Kliment Ohridski",
Faculty of Mathematics
and Informatics,
Sofia, Bulgaria,
e-mail: rdimova@fmi.uni-sofia.bg