manuscript No.
(will be inserted by the editor)

Breakdown Point and Computation of Trimmed Likelihood Estimators in Generalized Linear Models

Neyko M. Neykov[1] ⋆**, Christine H. Müller**[2]

[1] National Inst. of Meteorology and Hydrology, Bulgarian Academy of Sciences, 66 Tsarigradsko chaussee, 1784 Sofia, Bulgaria
[2] Carl von Ossietzky University of Oldenburg, Dept. of Mathematics, Postfach 2503, D - 26111 Oldenburg, Germany

Received: date / Revised version: date

Abstract A review of the studies concerning the finite sample breakdown point (BP) of the trimmed likelihood (TL) and related estimators based on the d–fullness technique of Vandev [28], and Vandev and Neykov [31] is made. In particular, the BP of these estimators in the frame of the generalized linear models (GLMs) depends on the trimming proportion and the quantity $\mathcal{N}(X)$ introduced by Müller [17]. A faster iterative algorithm based on resampling techniques for derivation of the TLE is developed. Examples of real and artificial data in the context of grouped logistic and log-linear regression models are used to illustrate the properties of the TLE.

1 Introduction

The Weighted Trimmed Likelihood (WTL) estimators are defined by Hadi and Luceño [9], and Vandev and Neykov [31] as

$$\mathrm{WTL}_k(y_1,\ldots,y_n) := \arg\min_{\theta\in\Theta^p} \sum_{i=1}^{k} w_i f(y_{\nu(i)},\theta), \tag{1}$$

Send offprint requests to: Neyko M. Neykov

⋆ The authors would like to thank the Editor and the referees for their helpful comments. The research of N. Neykov was partialy funded by Deutsche Forschungsgemeinschaft while he was Visiting Fellow of the Inst. of Math. Stochastics, Univ. of Götingen during May-June 1999. The Austrian Ministry of Science fully supported the participation of N. Neykov at the International Conference on Robust Statistics, 23-27 July 2001, Vorau, Austria. He acknowledged gratefully the financial supports.

where $f(y_{\nu(i)},\theta) \leq f(y_{\nu(i+1)},\theta)$, $f(y_i,\theta) = -log\ \varphi(y_i,\theta)$, $y_i \in \mathcal{Y} \subset R^q$ for $i = 1,\ldots,n$ are iid observations with probability density $\varphi(y,\theta)$, which depends on an unknown parameter $\theta \in \Theta^p \subset R^p$, $\nu = (\nu(1),\ldots,\nu(n))$ is the corresponding permutation of the indices, which may depend on θ, k is the trimming parameter and $w_i \geq 0$ are known weights such that $w_k > 0$.

The WTL estimators reduce to TLE if $w_i = 1$ for $i = 1,\ldots,k$. In the case of normal regression and appropriate choice of the weights, the WTLE reduce to the LMS and LTS estimators of Rousseeuw [24] and Rousseeuw and Leroy [25]. Similarly, the WTLE coincide with the MVE and MCD estimators of the multivariate location and scatter considered by Rousseeuw and Leroy [25] in the multivariate normal case (see Vandev and Neykov [30]). The Fisher consistency, asymptotic normality and compact differentiability of the TLE for normal distributions with unknown variance are derived by Bednarski and Clarke [3].

The BP (i.e. the smallest fraction of contamination that can cause the estimator to take arbitrary large values) properties of the WTLE were studied by Vandev and Neykov [31] using the d–fullness technique developed by Vandev [28]. It was proved that the BP of the WTLE is not less than $(n-k)/n$ if the set $F = \{f(y_i,\theta), i = 1,\ldots,n\}$ is d-full, $n \geq 3d$ and $(n+d)/2 \leq k \leq n-d$. We remind that, according to Vandev [28], a finite set F of n functions is called d-full if for each 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.

Vandev and Neykov [30], and Vandev and Marincheva [29] determined the value of d for the multivariate normal and general elliptical family of distributions, respectively. Vandev and Neykov [31] did the same about some linear and logistic regression models under the restriction that the observations are in general position. Similarly, the fullness parameters for the Lognormal, Poisson, Gamma, Geometric and Logarithmic series distributions were derived by Atanasov[1], and the BPs of the WTLE of the corresponding GLMs were characterized (see Atanasov and Neykov [2]).

There are approaches on robust and in particular high BP estimators for logistic regression and other nonlinear models given by Copas [4], Carroll and Pederson [5], Christmann [6], Christmann and Rousseeuw [7], Hubert [12], Künsch et al.[13], Markatou et al. [15], Stromberg and Ruppert[27], to name a few, but these approaches do not concern TLE.

The BP of the LMS, LTS and related regression estimators were derived by Rousseeuw [24], Rousseeuw and Leroy [25], and Hössjer [11], assuming that the observations are in general position. Müller [17], and Mili and Coakley [16] omitted this restriction and showed that then the BP of these estimators is determined by $\mathcal{N}(X) := \max_{0\neq\beta\in R^p} \text{card}\left\{i \in \{1,\ldots,n\};\ x_i^\top \beta = 0\right\}$, where $X := (x_i^\top)$ is the data matrix of the explanatory variables $x_i \in R^p$. If x_i are in general position then $\mathcal{N}(X) = p-1$ whereas in other cases, e.g., ANOVA models or designed experiments $\mathcal{N}(X)$ is much higher.

Müller and Neykov [19] relaxed the compactness condition in the above definition assuming only that the set $L_{g(\theta)}(C)$ is contained in a compact set. However, the meaning of the term subcompact function is retained since if the function $g(\theta)$ is continuous or has at most countable many discontinuities then $L_{g(\theta)}(C)$ is a compact set. The following theorem characterizes the BPs of any estimator S defined by $S(y) := \arg\min_{\theta\in\Theta} s(y,\theta)$, where $s(y,\theta)$ can be estimated by $f(y_{\nu(k)},\theta)$ and satisfies the conditions $\alpha\, f(y_{\nu(k)},\theta) \le s(y,\theta) \le \beta\, f(y_{\nu(k)},\theta)$ for some constants $\alpha \neq 0$ and β, and therefore of the WTLE in particular.

Theorem 1 *If* $\{f(y_i,\theta);\ i=1,\dots,n\}$ *is d-full, then the BP of the estimator* S *is not less than* $\frac{1}{n}\min\{n-k+1, n-d+1\}$.

This theorem is an extension of Theorem 1 of Vandev and Neykov [31] and provides the lower bound of the BP without additional assumptions on k and n (see Müller and Neykov [19]).

Thus, if one wants to study the BP of the WTL and related S estimators for a particular distribution, one has to find the fullness parameter d for the corresponding set of log likelihoods and then the BP can be exemplified by the range of values of k by Theorem 1.

An application of this technique is made, by Müller and Neykov [19], for the general linear exponential families of distributions (known dispersion parameter) of y_i depending on unknown vector parameter $\beta \in R^p$ and known $x_i \in R^p$ for $i=1,\dots,n$. The log likelihoods of these families are $f(y_i,x_i,\beta) = -T(y_i)^\top g(x_i^\top\beta) - c(x_i^\top\beta) - h(y_i)$ for suitably defined vectors and functions. The following theorem holds.

Theorem 2 *The set* $\{f(y_i,x_i,\beta);\ i=1,\dots,n\}$ *is* $\mathcal{N}(X)+1$*–full if the function* $\gamma_z(\theta) = -T(z)^\top g(\theta) - c(\theta) - h(z)$ *is subcompact in* θ *for all* $z \in \mathcal{Y}$ *and arbitrary* $x_i \in R^p$.

For the particular cases of normal, logistic and log-linear regression models Müller and Neykov [19] show that the corresponding $\gamma_z(\theta)$ are subcompact. Therefore, according to Theorem 1 and some additional arguments it is shown that the BP of the WTL estimators is $\frac{1}{n}\min\{n-k+1, k-\mathcal{N}(X)\}$. If k satisfies $\lfloor (n+\mathcal{N}(X)+1)/2\rfloor \le k \le \lfloor (n+\mathcal{N}(X)+2)/2\rfloor$ this BP is maximized and equal to $\frac{1}{n}\lfloor (n-\mathcal{N}(X)+1)/2\rfloor$, where $\lfloor r\rfloor := \max\{n \in N; n \le r\}$. As a consequence, the results of Müller [17] and [18], Vandev and Neykov [31], and Atanasov [1] for these models are derived.

In this way, a unifying theory for the BP of the WTL and related estimators is developed.

2 The FAST-TLE algorithm

From the definition of the WTLE it follows that its minima are achieved over a subsample of size k. The objective function (1) is continuous, but non differentiable and possesses many local minima. Therefore one need

nonsmooth and/or combinatorial optimization in general. In the univariate case Hadi and Luceño [9] developed several algorithms for TL estimation.

Neykov and Neytchev [21] considered an iterative approximate algorithm for finding the TLE which is based on the resampling technique proposed by Rousseeuw and Leroy [25]. Many subsets of k different observations out of n are drawn at random and the MLE is calculated for any one. The estimate with the lowest TL objective function (1) is retained. There is no guarantee that the achieved estimate will be the global minimizer of (1) but one can hope that it would be a close approximation to it.

In this paper we offer a more efficient TLE algorithm called the FAST-TLE as it reduces to the FAST-LTS algorithm developed by Rousseeuw and van Driessen [26] in the normal linear regression case. The corner stone of this algorithm is an analog of the so called *C–step* procedure proposed by these authors. We shall follow closely the terminology and exposition of their paper in order to present the algorithm in a more readable form to those who are acquainted with it.

So as to make sure that there always exists a solution to the optimization problem (1), we assume that the set F is d–full and $k \geq d$(see Neykov [20]). Then the idea behind the FAST-TLE algorithm can be described as follows.

Given $H^{old} = \{y_{j_1}, \ldots, y_{j_k}\} \subset \{y_1, \ldots, y_n\}$ then:
- take $\widehat{\theta}^{old}$ to be either arbitrary or compute $\widehat{\theta}^{old} := MLE$ based on H^{old};
- define $Q^{old} := \sum_{i=1}^{k} f(y_{j_i}, \widehat{\theta}^{old})$;
- sort $f(y_i, \widehat{\theta}^{old})$ for $i = 1, \ldots, n$ in ascending order, $f(y_{\nu(i)}, \widehat{\theta}^{old}) \leq f(y_{\nu(i+1)}, \widehat{\theta}^{old})$, and get the permutation $\nu = (\nu(1), \ldots, \nu(n))$;
- put $H^{new} := \{y_{\nu(1)}, \ldots, y_{\nu(k)}\}$;
- compute $\widehat{\theta}^{new} := MLE$ based on H^{new};
- define $Q^{new} := \sum_{i=1}^{k} f(y_{\nu(i)}, \widehat{\theta}^{new})$.

Proposition. On the basis of the above statements $Q^{new} \leq Q^{old}$.
Proof. From the definition of H^{new} it follows that $\sum_{i=1}^{k} f(y_{\nu(i)}, \widehat{\theta}^{old}) \leq \sum_{i=1}^{k} f(y_{j_i}, \widehat{\theta}^{old}) = Q^{old}$. Since $\widehat{\theta}^{new}$ is the MLE based on H^{new} then $Q^{new} = \sum_{i=1}^{k} f(y_{\nu(i)}, \widehat{\theta}^{new}) \leq \sum_{i=1}^{k} f(y_{\nu(i)}, \widehat{\theta}^{old}) \leq Q^{old}$.

We call this step in our algorithm *C–step* just like Rousseeuw and van Driessen [26] where C is reserved for 'concentration' since H^{new} is more concentrated (has a lower sum of negative log likelihoods) than H^{old}.

Clearly, repeating *C–step* yields an iterative process. When $Q^{new} = Q^{old}$ the process terminates; otherwise we need more *C–steps*. In this way a nonnegative monotonically decreasing sequence $Q^1 \geq Q^2 \geq Q^3 \geq \ldots$ is defined, which by a classical theorem in analysis is always convergent. Moreover, the convergence is guaranteed after a finite number of steps since there are only finitely many k–subsets out of $n!/(k!(n-k)!)$ in all. Finally, we note that this is only a necessary condition for a global minimum of the TL objective function. This gives us a hint as to how to implement an algorithm. Actually, we will be using the suggestion made by Rousseeuw

and Van Driessen [26] *"Take many initial choices of H^{old} and apply C–steps to each until convergence, and keep the solution with lowest value of"* (1).

However, this would not be of much use unless we can tell: how to generate different sets H^{old} to start the algorithm; the necessary number of H^{old} sets; how to avoid duplication of work since several H^{old} may yield the same solution; is it possible to reduce the number of *C–steps*.

Unfortunately, at this stage we cannot provide reasonable answer to all these issue alike Rousseeuw and Van Driessen [26] as the structure of the data in GLMs beyond the linear regression case is usually more complicated. However, it is worth to discuss some of these aspects based on the experience concerning the grouped binary linear logistic and Poisson regression cases.

First, we consider the possibilities for the sample sizes of H^{old}. Since the parameter of fullness of the GLMs is given explicitly by $\mathcal{N}(X)$ then any k within the bounds $\mathcal{N}(X)+1 \leq k \leq n$ can be chosen to draw a random k–subset in order to compute $\widehat{\theta}^{old}$. A recommendable choice of k is $\lfloor (n+\mathcal{N}(X)+1)/2 \rfloor$ as the BP of the TLE is maximized. However, following the same reasoning as Rousseeuw and Van Driessen [26] and because $\widehat{\theta}^{old}$ can be arbitrary, one should draw subsamples with a smaller $k^* := \mathcal{N}(X)+1$ size as the chance to get at least one outlier free subsample is larger. In practice, for the case of initial choices of H^{old}, we draw finitely many random subsamples of size k^*, calculate ML estimate $\widehat{\theta}^{old}$ for any one, and keep those 10 different subsamples of size k whose TL values evaluated at $\widehat{\theta}^{old}$ are lowest. In this way the resampling process would guarantee better initial choice of H^{old} sets. The recommendable choice of k^* and k could be used as defaults in a software implementation. If the expected percentage of outliers in data is low then a larger value of k can be chosen by the user in order to increase the efficiency of the TL estimator.

Second, as the regression models we consider belong to the linear exponential families an iteratively reweighted least squares algorithm discussed by Green [8] for obtaining the MLE can be used. Therefore any modern Gauss-Newton nonlinear regression program can be used to carry out the computations as the iteratively reweighted Gauss-Newton, Fisher scoring and Newton-Raphson algorithms are identical to these families, see Jennrich and Moore [14]. In all the applications of the MLE handled by such a program called NLR, see Neytchev at al. [22], convergence to $\widehat{\theta}^{old}$ discussed in the previous paragraph is reached in about 6 iterations starting from an arbitrary value $\theta^o := (0, \ldots, 0)$.

Third, each *C–step* calculates MLE based on k observations, and the corresponding log likelihoods for all n observations. In practice, we need 4 or 5 *C–steps* at most to reach convergence starting from $\widehat{\theta}^{old}$ at the first *C–step*, which leads to a faster convergence at the remaining *C–steps*.

A combination of the above elements yields the basis of our algorithm.

If the data set is large one can apply partitioning and nestings in a similar way as in FAST-LTS of Rousseeuw and Van Driessen [26], i.e., the entire data is partitioned in a representative way to several data subsets

Table 1 Subset of data set 28 of Hand et al.[10]

t_i	s_i	Area	Needle	x_{1i}	x_{2i}	x_{3i}	x_{4i}
228	223	1	1	1	-1	-1	-1
221	210	1	2	1	-1	-1	1
230	218	1	1	1	-1	-1	-1
221	181	2	1	1	1	-1	-1
213	158	2	2	1	1	-1	1
200	160	2	1	1	1	-1	-1
223	198	3	1	1	0	2	-1
228	189	3	2	1	0	2	1
216	177	3	1	1	0	2	-1

with smaller size. Applying the above algorithm to any subset the best 10 estimates $\widehat{\theta}^{old}_{sub}$ can be calculated. The process continues by making *C–steps* over the merged set, which is composed by pooling the subsets. In this way the best 10 estimates $\widehat{\theta}^{old}_{merged}$ can be obtained, and at last to find the best solution $\widehat{\theta}_{full}$.

When the data set is small all possible subsets with the default size k can be considered for calculation of the TLE skipping the *C–steps* procedure.

The above algorithm can be implemented easely using the environment of the software packages such as GLIM, S-PLUS, SAS, etc.

3 Applications

We illustrate our theory and algorithm by three examples. As a first one we analysed a subset of the data set 28 of Hand et al. [10] concerning the vaccination successes in three different areas (1=Staffordshire, 2=Cardiff, 3=Sheffield) by using two types of needles (1=fixed, 2=detachable). In the original data set an additional factor, the vaccine batch, was given. This factor was dropped since it had no significant influence and reduces the model's low degree of freedom once more. So a subset of the data with design matrix $X = (x_i^\top)$ with $x_i \in R^4$ is given in Table 1. The logistic regression model $\text{logit}(p/(1-p)) = x_i^\top \beta$ is used. As $\mathcal{N}(X) = 6$ the maximum BP attained by any TLE with $k = 8$ is $\frac{2}{9}$. We obtained $TL_8(y, X) = (2.05, -0.92, -0.12, -0.21)^\top$ and $ML(y, X) = (2.01, -0.92, -0.17, -0.15)$ for β. The mean absolute difference between these estimates is less than 0.04. It seems that there are no large influential outliers in the sample. To study the behavior of the estimators in the presence of one outlier we replaced s_1 and t_1 by $s_1 = 0$ and $t_1 = u$, respectively, where u attains several large values. For a study with two outliers we additionally replaced s_9 and t_9 by $s_9 = 0$ and $t_9 = u$. Table 2 provides the mean absolute difference between the estimators at the original and the contaminated samples.

Table 2 Mean absolute differences

Estimator	u	10	20	50	100	200	500	1000
MLE	1 outlier	0.15	0.23	0.39	0.54	0.71	0.94	1.11
TLE_8	1 outlier	0.07	0.07	0.07	0.07	0.07	0.07	0.07
MLE	2 outliers	0.15	0.24	0.41	0.57	0.77	1.06	1.28
TLE_8	2 outliers	0.07	0.11	0.19	0.28	0.40	0.55	0.67

These results show clearly that the TLE is stable in the presence of one outlier and breaks down (explodes) in the presence of two outliers. However, the explosion of it and the MLE is not linear in u, it is more logarithmical.

The second example is about a toxicological experiment conducted at the University of Waterloo, Canada, and discussed in O'Hara Hines and Carter [23] with $n = 48$ observations. A logistic regression model is fitted to the data with covariates for water hardening (WH), and for a linear and quadratic term in log concentration (C) of toxicant

$$\text{logit}(p/(1-p)) = \beta_1 + \beta_2 * WH + \beta_3 * \log_{10}(C) + \beta_4 * \log_{10}(C^2), \quad (2)$$

where $\beta_1, \beta_2, \beta_3$, and β_4 are unknown parameters.

Based on all observations the MLE is $(10.28, 0.03, -11.4, 2.50)^\top$. O' Hara Hines and Carter [23] pinpoint the observations 38, 39 and 26 as possible outliers. They also reported that Pregibon's influence diagnostics indicated the observations 38 and 39 as potential outliers. The MLE without the cases 38 and 39 is $(15.40, 0.27, -15.53, 3.26)^\top$ whereas without the cases 26, 38 and 39 is $(14.04, 0.32, -14.64, 3.11)^\top$.

Markatou et al. [15] analyzed the same data. They identified the observations 38 and 39 as potential outliers, whilst their methods gave a weight nearly 1 to observations 26 by means of the negative exponential RAF (Residual Adjustment Function) downweight function. When the Hellinger RAF was used for the construction of the weights, observations 13, 32, 40, 43 and 44 received a weight of 0. They reported that examination of those observations revealed that observations 32 and 40 had a 0 response, while observations 43 and 44 had the lowest mortality at concentration levels 720 and 1440, respectively, at the same water-hardening level. The MLE without the observations 13, 32, 40, 43 and 44 is $(6.49, -0.23, -8.42, 1.97)^\top$.

We dropped the observations 32 and 40 in TLE analysis as subcomactness can not be prooved because of zero response according to Müller and Neykov [19], and Vandev and Neykov [31]. Since 24 observations satisfy WH=1, we have $\mathcal{N}(X) = 24$. Hence, the maximum breakdown point is 11/46 and is attained by any TL estimator with $k = 35$ or $k = 36$. Using the TLE algorithm we obtained $TL_{36} = (7.36, -0.12, -9.29, 2.16)^\top$. The trimmed observations are 13, 14, 20, 21, 38, 39, 41, 42, 43, 44. The Pearson residuals diagnostic calculated by the TL_{36} estimate indicate these observations as potential outliers (see Fig. 1). As a bench-mark, the value of 3 is considered. Hence there is some coincidence with the results of Markatou et al. [15] with respect to the estimate and the trimmed observations.

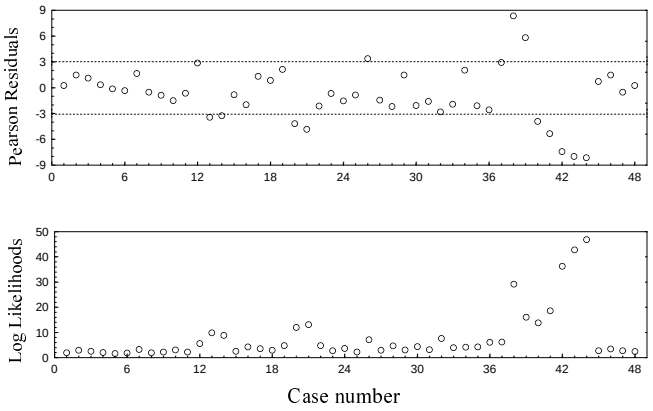

Fig. 1 O' Hara Hines and Carter [23] data: Index plot associated with Pearson residuals and log likelihood values based on TLE.

Table 3 Data set 340 of Hand et al. [10]

i	1	2	3	4	5	6	7	8	9
	10	11	12	13	14	15	16	17	
t_i	376	347	322	104	103	98	96	85	82
	63	44	40	5	5	0	0	0	
y_i	8	5	8	4	6	4	8	6	4
	2	7	4	3	2	4	3	2	

Next example is about the data set 340 of Hand et al. [10], given in Table 3, concerning the amount of newspaper and TV publicity t_i following $i = 17$ murder-suicides through deliberate crashing of private aircraft and the number y_i of fatal crashes during the week immediately following.

Since fatal crashes are rare events a log-linear model can be assumed where the amount t_i of publicity is the explanatory variable. For simplicity we assume a linear influence of t_i, i.e., $x_i = (1, t_i)^\top$. Then the maximum BP is 7/17 is attained by any TLE with $k = 10$ or $k = 11$ as $\mathcal{N}(X) = 3$. We obtained $TL_{11}(y, X) = (1.086, 0.002)^\top$ and $ML(y, X) = (1.310, 0.002)^\top$ for $\beta = (\beta_1, \beta_2)^\top$. Both estimators provides a very small estimate of the slope of the regression line but they differ with respect to the estimated intercept. This difference is caused by the fact that the TLE trims the highest numbers of crashes at i=1, 3, 5, 7, 8, 11. A scatter plot of this two-dimentional data

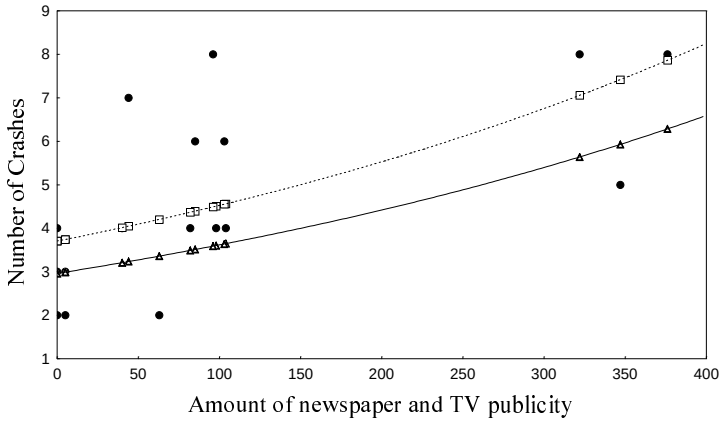

Fig. 2 Scateplot of 340 data set of Hand et al. [10] with MLE (dashed) and TLE (solid) curves.

set is given in Fig. 2, along with the MLE (squares) and TLE (triangles) fits.

References

1. Atanasov, D. V. About the finite sample breakdown point of the WTL estimators, MSc. Thesis, Faculty of Mathematics, Sofia Univ., (in Bulgarian),(1998)
2. Atanasov, D. V. and Neykov, N. M. On the finite sample BP of the WTL estimators and the d–fullness of a set of continous functions. CDAM conference, Minsk, Belarus, (2001) (accepted)
3. Bednarski, T. and and Clarke, B. R. Trimmed likelihood estimation of location and scale of the normal distribution. Astral. J. Statist. **35,** (1993) 141–153
4. Copas, J. B. Binary regression models for contaminated data (with discussion). J. R. Statist. Soc. B. **50,** (1988) 225–265
5. Carroll, R.J. and Pederson, S. On robustness in the logistic regression model. J. R. Statist. Soc. B. **55,** (1993) 693–706
6. Christmann, A. Least median of weighted squares in logistic regression with large strata. Biometrika. **81,** (1994) 413–417
7. Christmann, A. and Rousseeuw, P.J. Measuring overlap in logistic regression. (1999) (Submitted)
8. Green,P.J. Iteratively reweighted least squares for maximum likelihood estimation, and some robust and resistant alternatives, J. Roy. Statist. Soc. Ser. B **46,** (1984) 149–192

9. Hadi, A. and Luceño, A. Maximum Trimmed Likelihood Estimators: A Unified Approach, Examples and Algorithms, Comput. Statist. and Data Analysis. **25,** (1997) 251–272
10. Hand, D.J., Daly, F., Lunn, A.D., Mc Conway, K.J. and Ostrowski, E. *A Handbook of Small Data Sets* (Chapman & Hall, London, 1994)
11. Hössjer, O. Rank-based estimates in the linear model with high breakdown point. J. Amer. Statist. Assoc. **89,** (1994) 149–158
12. Hubert, M. The breakdown value of the L_1 estimator in contingency tables. Statistics and Probability Letters. **33,** (1997) 419–425
13. Künsch, H. R., Stefanski, L. A. and Carroll, R. J. Conditionally unbiased bounded influence estimation in general regression models, with applications to generalized linear models. J. Amer. Statist. Assoc. **84,** (1989) 460–466
14. Jennrich, R. I. and Moore, R.H. Maximum likelihood estimation by means of nonlinear least squares. Proc. of the Statistical Computing Section of the Amer. Statist. Assoc. (1975) 57–65
15. Markatou, M., Basu, A. and Lindsay, B. Weighted likelihood estimating equations: The discrete case with applications to logistic regression. J. Statist. Plann. Inference. **57,** (1997) 215–232
16. Mili, L. and Coakley, C. W. Robust estimation in structured linear regression. Ann. Statist. **15,** (1996) 2593–2607
17. Müller, Ch. H. Breakdown points for designed experiments. J. Statist. Plann. Inference. **45,** (1995) 413–427
18. Müller, Ch. H. *Robust Planning and Analysis of Experiments* (Springer, New York, 1997)
19. Müller, Ch. H. and Neykov, N. M. Breakdown points of trimmed likelihood estimators and related estimators in generalized linear models. J. Statist. Plann. Inference, (2000) (submitted)
20. Neykov, N. M. *Robust methods with high breakdown point in the multivariate statistical analysis* Ph.D. Thesis, Faculty of Mathematics, Sofia University, (in Bulgarian), (1995)
21. Neykov, N. M. and Neytchev, P. N. A robust alternative of the maximum likelihood estimator. Short communications of COMPSTAT, Dubrovnik (1990), 99–100
22. Neytchev, P.N. Neykov,N.M. and Todorov,V.K. User's manual of REGRESS PC program system for fitting models to data. TR of NIMH, Sofia, (1994)
23. O'Hara Hines, R.J. and Carter, E.M. Improved added variable and partial residual plots for the detection of influential observations in generalized linear models. Appl. Statist. **42,** (1993) 3–20
24. Rousseeuw, P. J. Least median of squares regression. J. Amer. Statist. Assoc. **79,** (1984) 851–857
25. Rousseeuw, P. J. and Leroy, A. M. *Robust Regression and Outlier Detection* (Wiley, New York, 1987)
26. Rousseeuw, P. J. and Van Driessen, K. Computing LTS Regression for Large Data Sets. (2000) (submitted)
27. Stromberg, A. J. and Ruppert, D. Breakdown in nonlinear regression. J. Amer. Statist. Assoc. **87,** (1992) 991–997
28. 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
29. Vandev, D.L. and Marincheva, M.The BP of the WLT estimators in the general elliptic family of distribution. In: *Proc. of Statistical Dada Analysis*, Vandev, D.L. (ed.) Varna, (1996) 25–31

30. Vandev, D. L. and Neykov, N. M. Robust maximum likelihood in the Gaussian case, In: *New Directions in Data Analysis and Robustness* Morgenthaler, S., Ronchetti, E. and Stahel, W. A. (eds.), (Birkhäuser Verlag, Basel, 1993) 259–264
31. Vandev, D. L. and Neykov, N. M. About regression estimators with high breakdown point. Statistics **32,** (1998) 111–129