

Article

Improving the Efficiency of Robust Estimators for the Generalized Linear Model

Alfio Marazzi ^{1,2}¹ Nice Computing SA, 1052 Le Mont-sur-Lausanne, Switzerland; alfio.marazzi@chuv.ch² Faculty of Biology and Medicine, University of Lausanne, 1015 Lausanne, Switzerland

Abstract: The distance constrained maximum likelihood procedure (DCML) optimally combines a robust estimator with the maximum likelihood estimator with the purpose of improving its small sample efficiency while preserving a good robustness level. It has been published for the linear model and is now extended to the GLM. Monte Carlo experiments are used to explore the performance of this extension in the Poisson regression case. Several published robust candidates for the DCML are compared; the modified conditional maximum likelihood estimator starting with a very robust minimum density power divergence estimator is selected as the best candidate. It is shown empirically that the DCML remarkably improves its small sample efficiency without loss of robustness. An example using real hospital length of stay data fitted by the negative binomial regression model is discussed.

Keywords: robust GLM estimators; robust Poisson regression; conditional maximum likelihood estimator; minimum density power divergence estimator; distance constrained maximum likelihood



Citation: Marazzi, A. Improving the Efficiency of Robust Estimators for the Generalized Linear Model. *Stats* **2021**, *4*, 88–107. <https://doi.org/10.3390/stats4010008>

Academic Editor: Marco Riani
Received: 2 December 2020
Accepted: 25 January 2021
Published: 4 February 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

For a long time, early robust parametric procedures were unable to combine a high level of outlier-resistance and a high level of efficiency under the model. Ref. [1] suggested one of the first tools to join robustness and efficiency, i.e., trimming some outliers identified with the aid of least absolute deviations followed by the computation of the least squares estimator from the remaining data. For the linear model, several proposals with the same purpose followed. An adaptive combination of the least squares and the least absolute deviations estimators was introduced by [2]. Their approach could be adapted in order to achieve the minimum asymptotic variance over a scale family of densities. Ref. [3] proposed MM estimators, which have a 50% breakdown point and asymptotic efficiency as close to one as desired. Ref. [4] proposed τ estimates, which combine the same two properties as MM estimators. Ref. [5] introduced adaptive cutoff values based on an initial highly robust and possibly inefficient estimator (such as an S estimator) to define a region Z such that observations out of Z are considered as outliers. The maximum likelihood (ML) estimator is then computed after removal of the outliers. Since, under the model, Z tends to the whole space for increasing sample size, the final estimator is asymptotically fully efficient. A similar idea was used by [6] to compute a highly robust and fully efficient truncated ML estimator for regression with asymmetric errors. Ref. [7] used the adaptive cutoff values to define adaptive weights for a least weighted squares estimator, which is asymptotically efficient and highly robust.

Several robust methods for the generalized linear model (GLM) have also been proposed. Ref. [8] made use of diagnostic tools based on ML residuals. This tool may, however, be affected by the well-known “masking effect” of ML residuals described in [9] (Section 4.3). Other authors introduced different types of M estimators ([10–15], among others) where the trade-off between robustness and efficiency is controlled by means of one or more tuning constants. A fully efficient and highly robust estimator was proposed

in [16] using a conditional ML principle, after removal of outliers not belonging to an adaptive region which tends to the whole space.

Unfortunately, when n is not very large, the finite sample efficiency of most estimators, including the fully efficient ones, may be much smaller than the asymptotic one. To overcome this shortcoming, ref. [17] proposed a regression estimator with the maximum breakdown point and high finite sample efficiency. However, the method suffers from a serious loss of robustness. Ref. [18] introduced a general method to improve the finite sample efficiency of a robust estimator. Details were given only for the cases of the linear model and the multivariate location and scatter of a random vector. The method assumes that an initial robust estimator, not necessarily with high finite sample efficiency, is available. Then, the improved estimator—called the distance constrained maximum likelihood (DCML) estimator—maximizes the likelihood function subject to the estimate being sufficiently close to the initial one. As a distance, the Kullback-Leibler divergence from the maximum likelihood estimated model was chosen. For the linear model, it turns out that the final estimator is a convex combination of the robust and the maximum likelihood estimator. In principle, the procedure may be applied to any parametric model; however, for the GLM, the solution of the constrained optimality problem is more involved.

In this paper, we consider a simple modification of the DCML principle for the GLM. The modification exploits only the most crucial elements of the original proposal: it directly defines the DCML as a convex combination of the robust and the ML estimator and uses a quadratic distance between the coefficients. We explore the performance of the DCML method by means of Monte Carlo experiments considering some well-known robust estimators, for which the public domain R software is available, as initial estimators. To limit our presentation, we focus on the Poisson model. We first compare the efficiency and robustness of the initial estimators in a simple regression case. We then select the best starting estimators and assess the performance of the DCML in critical multiple regression scenarios for $p = 5, 10, 20$, and $n = 5p$, when the covariables follow a spherical normal distribution or a spherical t distribution with four degrees of freedom. We finally describe a bootstrap experiment with real hospital length of stays fitted by a negative binomial regression model.

2. Candidate Estimators and Software for Poisson Regression

We consider the following families of robust estimators for Poisson regression:

- the conditionally unbiased bounded influence estimators [10];
- the robust quasi-likelihood estimators [11];
- the transformed M estimators [13];
- the minimum density power divergence estimators [14];
- the modified conditionally maximum likelihood estimators (MCML) [16].

For short definitions of the CUBIF, RQL, and MT estimators and their tuning parameters, we refer to [9]. CUBIF is computed with the help of the R package “robcbi”; a function to compute the RQL estimators is provided in “robustbase”, and functions to compute the MT estimators can be found in the R package “poissonMT”. There is no public domain software to compute MD estimators, but the loss function defined in [14] can easily be minimized with the help of the R function “optim”. However, since the loss function of MD estimators may have multiple local minima, we use a specialized subsampling algorithm to find the absolute minimum with a high probability.

Particular attention is given to the MCML estimator. The MCML starts with the computation of a highly robust, but possibly inefficient initial estimator. Using randomized quantile residuals [19] based on this estimator, outliers are detected with the help of adaptive cutoff values as in [5,6]. Finally, a conditional ML estimator is computed, where observations are constrained to belong to a region free of detected outliers. Since in the absence of outliers, this region tends to the whole space when the sample size increases, the asymptotic efficiency of the MCML estimator is 100%. For moderate sample sizes, the MCML markedly improves the efficiency of the initial estimator; unfortunately, for

small sample sizes, the MCML efficiency is disappointingly low; nevertheless, the MCML maintains a similar degree of robustness as the initial estimator. It turns out, however, that the MCML is a good candidate for the DCML method. Computational tools to compute the MCML estimator are made available in the R package “robustnegbin”. In the original paper, as well as in the package, the initial step is based on a combination of the maximum rank correlation estimator [20] and CUBIF. Here, we propose a simpler version, where the initial step is replaced by an MD estimator.

3. The DCML for GLM

We consider a GLM:

$$y|x \sim f_{\mu_x, \sigma}(y|x) \tag{1}$$

relating a response y to a p -components explanatory covariate x . The conditional expected response $E(y|x) = \mu_x$ depends on a linear predictor $\beta_0^T x$ through a link function g , i.e., $\beta_0^T x = g(\mu_x)$, where β_0 is a p -component vector of coefficients. In addition, σ is a dispersion parameter. In this paper, we focus on the Poisson regression model with the log-link, i.e., $\beta_0^T x = \log(\mu_x)$ and $\sigma = 1$, which may be extended to the negative binomial (NB) model with $\sigma \neq 1$. We also write $\mu_x = h(\beta_0^T x)$, where h is the inverse of g ; in the Poisson case, $\mu_x = \exp(\beta_0^T x)$. If an intercept term is present, we sometimes use the notations $x^T = (1, x^{*T})$ and $\beta_0^T = (\alpha, \beta_0^{*T})$, where α is the intercept.

Suppose that $\hat{\beta}_K$ is a robust consistent estimator of β_0 , not necessarily efficient, and that $\hat{\beta}_H$ is the ML estimator of β_0 . We measure the distance between $\hat{\beta}_K$ and $\hat{\beta}_H$ by:

$$\Delta(\hat{\beta}_K, \hat{\beta}_H) = (\hat{\beta}_K - \hat{\beta}_H)^T \hat{C}^{-1} (\hat{\beta}_K - \hat{\beta}_H),$$

where \hat{C} is a robust estimate of the covariance matrix of $\hat{\beta}_K - \hat{\beta}_H$. We define the DCML estimator $\hat{\beta}_D$ as the convex combination:

$$\hat{\beta}_D = t\hat{\beta}_H + (1 - t)\hat{\beta}_K \tag{2}$$

such that t minimizes $\Delta(t\hat{\beta}_H + (1 - t)\hat{\beta}_K, \hat{\beta}_H)$ under the constraint:

$$\Delta(t\hat{\beta}_H + (1 - t)\hat{\beta}_K, \hat{\beta}_K) \leq \delta \tag{3}$$

and δ is an adequately chosen constant. It immediately follows that the optimal t is:

$$t_0 = \min \left(1, \sqrt{\frac{\delta}{\Delta(\hat{\beta}_K, \hat{\beta}_H)}} \right) \tag{4}$$

as in Formula (8) in [18].

Under the model and for large n , the distribution of $\Delta(\hat{\beta}_K, \hat{\beta}_H)$ is approximately χ_p^2 . Therefore, in order to control the efficiency of $\hat{\beta}_D$, it seems reasonable to take δ as a quantile of the χ_p^2 distribution such that the efficiency attains a satisfactory level. Monte Carlo simulation shows that a large quantile is usually required because the finite sample distribution of $\Delta(\hat{\beta}_K, \hat{\beta}_H)$ may be much more skewed than χ_p^2 .

We compare a few candidate estimators $\hat{\beta}_K$ for which the R software is easily available. These candidates belong to the M estimator family (see e.g., [9]). Asymptotically, an M estimator $\hat{\beta}_K$ of regression is a functional of the cdf F of (x, y) , satisfying an estimating equation:

$$E_F[\psi(x, y, \hat{\beta}_K)x] = 0, \tag{5}$$

where $\psi(x, y, \beta)$ is a given function that may depend on the model parameters. The influence function of $\hat{\beta}_K$ is:

$$IF_K(x, y, F) = \psi(x, y, \hat{\beta}_K)M_K^{-1}x, \tag{6}$$

where:

$$M_K = E_F[\psi'(x, y, \hat{\beta}_K)xx^T] \tag{7}$$

and $\psi'(x, y, \beta) = \partial/\partial\beta\psi(x, y, \beta)$. The asymptotic covariance matrix of $\hat{\beta}_K$ is given by:

$$V_K = E_F[IF_K(x, y, F)IF_K(x, y, F)^T] = M_K^{-1}Q_KM_K^{-1}, \tag{8}$$

where:

$$Q_K = E_F[\psi(x, y, \hat{\beta}_K)^2xx^T]. \tag{9}$$

The ψ functions and the influence functions of the candidate estimators of Poisson regression are provided in Appendix A. For the ML estimator, we have:

$$\begin{aligned} M_H &= -E[\hat{\mu}_x xx^T], \\ Q_H &= E[(y - \hat{\mu}_x)^2 xx^T], \\ V_H &= E[\hat{\mu}_x xx^T]^{-1}. \end{aligned}$$

The covariance matrix C of $\hat{\beta}_K - \hat{\beta}_H$ is obtained with the help of the influence functions $IF_H(x, y, F)$ and $IF_K(x, y, F)$:

$$C = E_F[(IF_H(x, y, F) - IF_K(x, y, F))(IF_H(x, y, F) - IF_K(x, y, F))^T]$$

i.e.,

$$C = V_H + V_K - (M_H^{-1}K_{HK}M_K^{-1} + M_K^{-1}K_{HK}M_H^{-1}), \tag{10}$$

where:

$$K_{HK} = E_F[(y - \hat{\mu}_x)\psi(x, y, \hat{\beta}_K)xx^T].$$

The distribution F is of the form $F(y|x)G(x)$. To estimate the cdf $F(y|x)$, we use the cdf $F_{\hat{\mu}_x, \hat{\sigma}_K}(y|x)$ where $\hat{\mu}_x = h(\hat{\beta}_K^T x)$ and $\hat{\sigma}_K$ is a robust estimate of σ ($\hat{\sigma}_K = 1$ in the Poisson case). Using the empirical cdf $G_n(x)$ to estimate G , we obtain:

$$\begin{aligned} \hat{Q}_K &= \frac{1}{n} \sum_i \psi(x_i, y_i, \hat{\beta}_K)^2 x_i x_i^T, \\ \hat{M}_K &= \frac{1}{n} \sum_i \psi'(x_i, y_i, \hat{\beta}_K) x_i x_i^T, \\ \hat{K}_{HK} &= \frac{1}{n} \sum_i (y_i - \hat{\mu}_{x_i}) \psi(x_i, y_i, \hat{\beta}_K) x_i x_i^T, \end{aligned}$$

etc. In general, these estimates are not robust. Therefore, as often proposed, we remove all the observations such that $Md(x_i^*)^2 > q_{p-1,0.95}$, where Md is the Mahalanobis distance of the point x_i^* from the center of the covariate distribution based on a very robust (with a maximum 50% breakdown point) covariance such as the minimum volume ellipsoid [21] enclosing 80% of the sample and $q_{p-1,0.95}$ is the 95% quantile of a χ_{p-1}^2 distribution. In the simple regression case, we remove observations such that $|x_i - \text{median}(x_i)| \leq 2 \times \text{MAD}(x_i)$. We then obtain an estimate of \hat{C} of C replacing M_K, V_K, M_H, V_H , and K_{HK} by their robust estimates in (10). Note that, in order to have a positive definite \hat{C} , the same robust estimates $\hat{\beta}_K$ and $\hat{\mu}_x = h(\hat{\beta}_K^T x)$ must be used in the evaluation of \hat{M}_K, \hat{V}_K , as well as \hat{M}_H, \hat{V}_H , and \hat{K}_{HK} .

It is easy to show that, for any fixed weight t , the total variance of $\hat{\beta}_D$ is smaller than the total variance of $\hat{\beta}_K$. Unfortunately $\hat{\beta}_D$ is not robust in this case. Nevertheless, under

the model, $\Delta(\hat{\beta}_K, \hat{\beta}_H)$ is very often smaller than δ , and t_0 equals one; therefore, we may expect that $\hat{\beta}_D$ is more efficient than $\hat{\beta}_K$. This is confirmed by the simulations in Section 6.

The constraint (3) makes the optimal t_0 a data dependent weight. Since x -outliers are removed, the eigenvalues of \hat{C}^{-1} are bounded away from zero. Let $\lambda_0 > 0$ be the smallest eigenvalue. Then,

$$\lambda_0 \left| \hat{\beta}_K - \hat{\beta}_D \right|^2 \leq (\hat{\beta}_K - \hat{\beta}_D)^T \hat{C}^{-1} (\hat{\beta}_K - \hat{\beta}_D) = \Delta(\hat{\beta}_K, \hat{\beta}_D) \leq \delta$$

implies that $\hat{\beta}_D$ cannot be unbounded when $\hat{\beta}_K$ is bounded; in other words, $\hat{\beta}_D$ has the same breakdown point as $\hat{\beta}_K$. A formal proof for the linear model case when $\hat{\beta}_K$ is an MM estimator was provided by [18]. The empirical results in Section 6 show that, for the best candidates $\hat{\beta}_K$, the DCML $\hat{\beta}_D$ is highly resistant to heavy point contaminations.

A method to compute an approximate distribution that can be used in the GLM case was also proposed in [18]. We note that the DCML estimator is a nonlinear function of $\hat{\beta}_H$ and $\hat{\beta}_K$, and its distribution is not necessarily normal.

4. Some Invariance Properties

In general, Poisson regression does not own regression invariance—as in the linear model case—and the Monte Carlo results depend on the chosen coefficients, as well as on the distribution of the predictors. Unfortunately, this makes it impossible to draw very general conclusions. We restrict our attention to spherical and elliptical distributions of x . For these models, the following results may be used to expand the results to a certain extent beyond those provided by the reported experiments.

As usual, we measure the error of $\hat{\beta}_K$ by the mean squared estimation error:

$$\text{MSEE}_F(\hat{\beta}_K) = E_F \left[\left| \hat{\beta}_K - \beta_0 \right|^2 \right]$$

or by the mean squared prediction error:

$$\text{MSPE}_F(\hat{\beta}_K) = E_F \left[\left| \hat{\beta}_H x - \beta_0 x \right|^2 \right].$$

The efficiency of $\hat{\beta}_K$ with respect to the ML estimator $\hat{\beta}_H$ is measured by:

$$\text{Eff}_1(\hat{\beta}_K) = \text{MSEE}_F(\hat{\beta}_H) / \text{MSEE}_F(\hat{\beta}_K)$$

or by:

$$\text{Eff}_2(\hat{\beta}_K) = \text{MSPE}_F(\hat{\beta}_H) / \text{MSPE}_F(\hat{\beta}_K).$$

Note that $\text{MSEE}_F(\hat{\beta}_K) = \text{trace}(\mathbf{V}_K)$, the total variance of $\hat{\beta}_K$.

We consider two models with no intercept and slope vectors β_1, β_2 . We denote by F_{kx} the cdf of (x, y) when β_0 is replaced by β_k ($k = 1, 2$) in (1). Let $T(F)$ be the functional form of a consistent estimator and let $\mathbf{V}(F_{kx})$ its asymptotic covariance under F_{kx} .

Theorem 1. *Suppose that x has a spherical distribution with center $\mathbf{0}$ and that β_1 and β_2 are two slope vectors with the same Euclidean norm. Then, $\text{trace}(\mathbf{V}(F_{1x})) = \text{trace}(\mathbf{V}(F_{2x}))$.*

Proof. Let P be the orthogonal matrix such that $\beta_2 = P\beta_1$ and let $z = Px$. Let F_{kz} be the cdf of (z, y) when the slope vector is β_k . We have $\mathbf{V}(F_{2z}) = P\mathbf{V}(F_{1z})P^T$. Since x is spherically distributed $\mathbf{V}(F_{2z}) = \mathbf{V}(F_{2x}), \mathbf{V}(F_{1z}) = \mathbf{V}(F_{1x})$. Then, $\text{trace}(\mathbf{V}(F_{1x})) = \text{trace}(\mathbf{V}(F_{2x}))$. \square

If an intercept term is present, a similar argument applies to a spherically distributed x^* and the slope vector β^* . It follows that, if x^* has a spherical distribution, the asymptotic efficiency just depends on the intercept and on the norm of the slope vector. This allows us to restrict attention to coefficient vectors of the form $\beta^T = (\alpha, \beta^{*T})$ for a varying intercept α

and $\beta^{*T} = (\gamma, 0, \dots, 0)^T$ for a varying γ . In the Monte Carlo simulations of Section 6, we choose values of α and γ that make the empirical efficiency of the estimators under study very low. We explore the efficiency improvement provided by the DCML procedure in such unfavorable cases.

The case where x^* is not centered in $\mathbf{0}$ can be led back to the centered case by changing the intercept term.

In the case where x has an elliptical distribution such that $x = Az$, where A is a regular $p \times p$ matrix and z has a spherical distribution with $E(z) = \mathbf{0}$ and $E(zz^T) = I_{p-1}$, we define $\gamma_0 = A^T \beta_0$, $\hat{\gamma}_H = A^T \hat{\beta}_H$, $\hat{\gamma}_K = A^T \hat{\beta}_K$. Then, $MSPE_F(\hat{\beta}_K) = MSEE(\hat{\gamma}_K)$ and $Eff_2(\hat{\beta}_K) = MSEE(\hat{\gamma}_H)/MSEE(\hat{\gamma}_K)$. Thus, using the MSPE, the elliptical case can be led back to the spherically symmetric case.

5. Monte Carlo Scenarios for Poisson Regression

We examine the following specific candidate estimators of Poisson regression:

- CUBIF11: CUBIF with tuning parameter $b = 1.1\sqrt{2}$;
- CUBIF18: CUBIF with tuning parameter $b = 1.8\sqrt{2}$;
- RQL01: RQL with tuning parameter $tcc = 0.1$;
- RQL14: RQL with default tuning parameter $tcc = 1.4$;
- MTI: the initial value of MT provided by poissonMT;
- MT23: MT with default tuning parameter $cc = 2.3$;
- MD04: MD with tuning parameter $\tau = 0.4$;
- MD07: MD with tuning parameter $\tau = 0.7$;
- MD10: MD with tuning parameter $\tau = 1.0$;
- MCML04: MCML starting from MD04r;
- MCML07: MCML starting from MD07r;
- MCML10: MCML starting from MD10r.

We consider both the case where leverage points are not removed and the case where they are removed. If necessary, to avoid confusion, a suffix “r” in the estimator name is added in the second case (e.g., MD04r, MD07r, etc.). The estimators CUBIF11, RQL01, MTI, and MD10 are (close to) the most robust, but are inefficient representatives of their families. In principle, they can be used as starting values to compute two stage estimators such as the MCML estimators. The estimators CUBIF18, RQL14, MT23, MD04, MD07, MCML04, MCML07, and MCML10 are considered as robust and efficient estimators.

Any estimator of this list can act as a $\hat{\beta}_K$ candidate in a DCML procedure. However, to illustrate the performance of the DCML, we first reduce the number of candidates on the grounds of their efficiency and robustness levels evaluated with the help of a quick Monte Carlo experiment. For this purpose, we use a simple Poisson regression model:

$$y|x \sim f_{\mu_x}(y|x) = \mu_x \exp(-\mu_x), \tag{11}$$

where $\mu_x = \exp(\beta_0 x)$ and $\beta_0 = (1, 0.25)$, $x = (1, x^*)^T$, $x^* \sim N(0, 1)$.

In a second set of experiments, we explore the DCML procedure starting at selected candidates. We use Poisson regression models with p coefficients:

$$y|x \sim f_{\mu_x}(y|x) = \mu_x^y \exp(-\mu_x), \tag{12}$$

where $\mu_x = \exp(\beta_0 x)$, $\beta_0^T = (\alpha, \beta_0^{*T})$, $\beta_0^{*T} = (\gamma, 0, \dots, 0)^T$, $x^T = (1, x^{*T})$. The intercept α is considered as a nuisance parameter. Simulations are performed for $\alpha = 1$, two values of γ ($\gamma = 0.25$ and 0.50), three values of p ($p = 5, 10, 20$), and two distributions of x^* : $x^* \sim N(\mathbf{0}, I_{p-1})$ and $x^* \sim t_4(\mathbf{0}, I_{p-1})$.

To measure the quality of an estimator $\hat{\beta}$, we use the empirical MSEE:

$$MSEE_n(\hat{\beta}^*) = \text{mean} \left(\left| \hat{\beta}_i^\# - \beta_0^* \right|^2 \right), \tag{13}$$

where $\hat{\beta}_i^\#$ is the estimate of β_0^* based on the i th replication. It turns out that for most robust estimators, a very small fraction of errors $|\hat{\beta}_i^\# - \beta_0^*|^2$ is extremely large. This is a well-known feature of robust estimators, and as [22] noted, “despite its seductive simplicity, the ordinary variance is not an adequate measure of performance of a robust estimate: it is much too sensitive to the irrelevant extreme tail behavior of the estimate”. We therefore use the upper 1% trimmed mean (in place of the usual mean) to evaluate the MSEE (13).

The following “simple” DCML estimators are compared:

- MD07r + D : DCML starting from MD07r
- MD10r + D: DCML starting from MD10r
- MCML07 + D: DCML starting from MCML07
- MCML10 + D: DCML starting from MCML10.

In addition, the following DCML “iterated” estimators are considered:

- MCML+: MCML starting from (MD10r + DCML)
- MCML*: MCML starting from (MCML + DCML)

as well as:

- MCML++: DCML starting from MCML+
- MCML*+: DCML starting from MCML*.

6. Monte Carlo Results for Poisson Regression

6.1. Simulations for the Nominal Simple Regression Model

In a first experiment, we drew 1000 samples of size 100 from Model (11). To compare the candidate estimators, we first computed the finite sample efficiencies Eff_{1n} and the asymptotic efficiencies reported in Table 1. The results of the first column were obtained using the complete (unweighted) data. To compute the second column, leverage points such that $|x_i - \text{median}(x_i)| \leq 2 \times \text{MAD}(x_i)$ were removed, except for the MCML estimators. We observe that the efficiencies of CUBIF, RQL, MT, and MD depend, as expected, on the choice of the tuning parameters; with the full data, they are slightly lower than their asymptotic values and markedly lower when leverage points are removed. By definition, the MCML estimators remove data points only when they are discordant with the fitted model and hence are asymptotically fully efficient. In our experiment with $n = 100$, their finite sample efficiencies are almost independent of the starting estimator MD04, MD07, or MD10. Further experiments reported below show, however, that for smaller n , the efficiency of the MCML is directly related to the efficiency of the starting estimator. In any case, the MCML efficiencies are disappointingly lower than their asymptotic value of one.

Table 1. Efficiencies of the robust Poisson regression candidates: Eff_{1n} is the empirical efficiency based on complete data; $Eff_{1n,r}$ is the empirical efficiency based on cleaned data; Eff_1 is the asymptotic efficiency. CUBIF, conditionally unbiased bounded influence estimator; RQL, robust quasi-likelihood estimator; MT, transformed M estimator; MD, minimum density power divergence estimator; MCML, modified conditionally maximum likelihood estimator.

Estimator	Eff_{1n}	$Eff_{1n,r}$	Eff_1
CUBIF11	0.62	0.52	0.66
CUBIF18	0.89	0.73	0.92
RQL01	0.68	0.54	0.66
RQL14	0.94	0.75	0.93
MT-I	0.85	0.78	—
MT23	0.87	0.70	0.90
MD04	0.84	0.67	0.89
MD07	0.75	0.60	0.79
MD10	0.66	0.54	0.68
MCML04	0.78	—	1.00
MCML07	0.78	—	1.00
MCML10	0.78	—	1.00

6.2. Simulations for the Contaminated Simple Regression Model

In a second experiment, samples from Model (11) were corrupted with point contaminations of the form $((1, x_{out})^T, y_{out})$ where x_{out} varied in the set $\{-3, -2, -1, 0, 1, 2, 3, 4, 5\}$ and y_{out} varied in the set $\{0, 1, 2, 10, 20, 30, 40, 60\}$. This kind of contamination is generally the least favorable one and allows the evaluation of the maximal bias an estimator can incur. We generated 100 samples of size 90; then, for each pair (x_{out}, y_{out}) , we added 10 observations with identical outliers of the form $((1, x_{out})^T, y_{out})$; thus, the contamination level was 10%.

We first compared the four less efficient estimators: CUBIF11r, RQL01r, MTIr, and MD10r. The MSEs of these estimators are displayed in Figure 1 as functions of y_{out} . For $x_{out} = -3, 3, 4, 5, 6$, the MSEs of all estimators are very small because leverage points are removed and the graphs for $x_{out} \geq 3$ are not shown. Important differences are observed for x_{out} between -2 and 2 , and MD10r clearly provides the overall most robust behavior. For this reason, MD10r is considered as a good starting value for the MCML estimator MCML10.

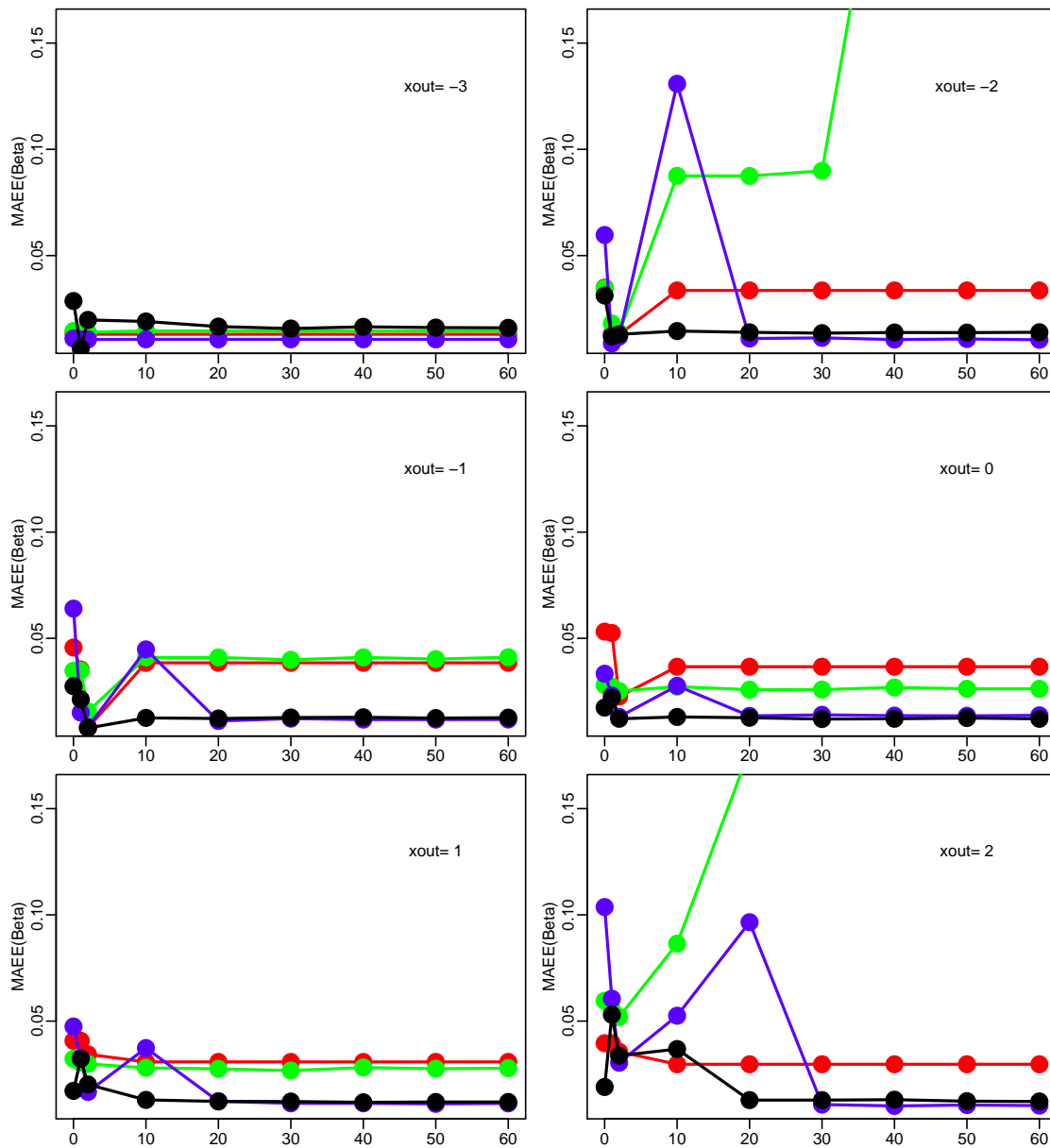


Figure 1. Monte Carlo results for the contaminated simple regression model. Mean square estimation error $MSEE(\hat{\beta})$ for varying y_{out} and x_{out} . CUBIF11r, red; RQL01r, green; MTIr, blue; MD10r, black.

Then, we compared the more efficient estimators RQL14, MT23, MD04, and MCML10. As Figure 2 shows, MCML10 generally provides the lowest MSEE. We conclude that MCML10 has a good robustness level and a full asymptotic efficiency, but a poor finite sample efficiency. It is therefore a natural candidate to study the performance of the DCML procedure.

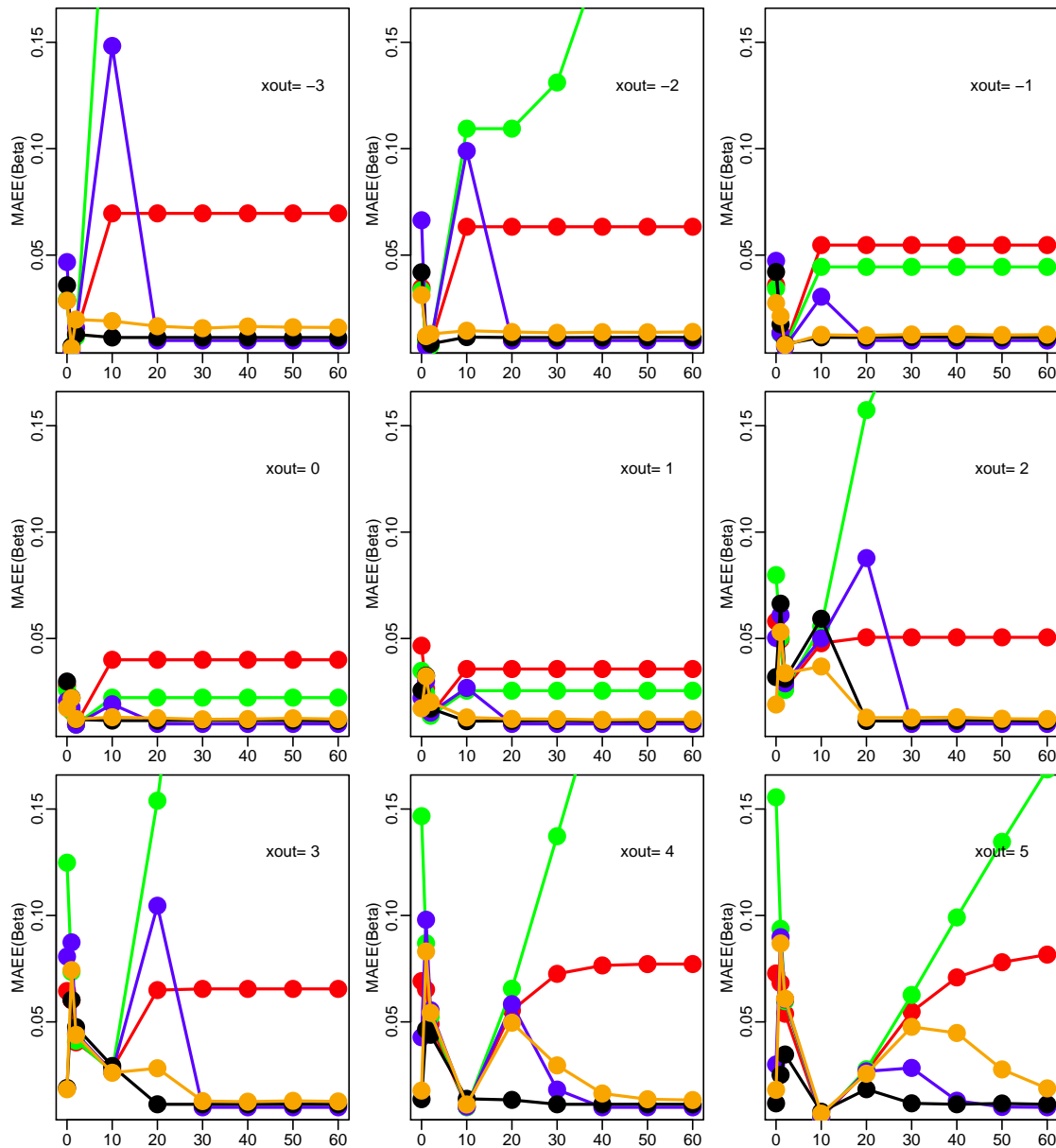


Figure 2. Monte Carlo results for the contaminated simple regression model. Mean square estimation error $MSEE(\hat{\beta})$ for varying y_{out} and x_{out} . RQL14, green; MT23, blue; MD04, orange; MCML10, black.

6.3. Simulations for the Nominal Multiple Regression Model

To compare the DCML estimators, we first used 1000 samples of size $n = 5p$ from Model (12) for $p = 5, 10, 20$ and computed empirical efficiencies Eff_{1n} . It turns out that the efficiency of the MD estimators is a decreasing function of $|\gamma|$; however, its dependence on α is quite weak (e.g., Figure 3). In addition, experiments with $x^* \sim t_4(\mathbf{0}, I_{p-1})$ and γ a bit larger than 0.5 or $\alpha > 1$ gave rise to computational crashes due to the occurrence of extremely high values of $\exp(\hat{\beta}_K x)$. This suggests that $\alpha = 1, \gamma = 0.5, x^* \sim t_4(\mathbf{0}, I_{p-1})$ is already a quite unfavorable scenario. We report the results for this case in Table 2. Results for the more favorable choice $\alpha = 1, \gamma = 0.25, x^* \sim t_4(\mathbf{0}, I_{p-1})$ are shown in Table 3.

As expected, higher efficiencies were obtained with the same values of γ and α , but for $\mathbf{x}^* \sim N(\mathbf{0}, I_{p-1})$; they are not reported.

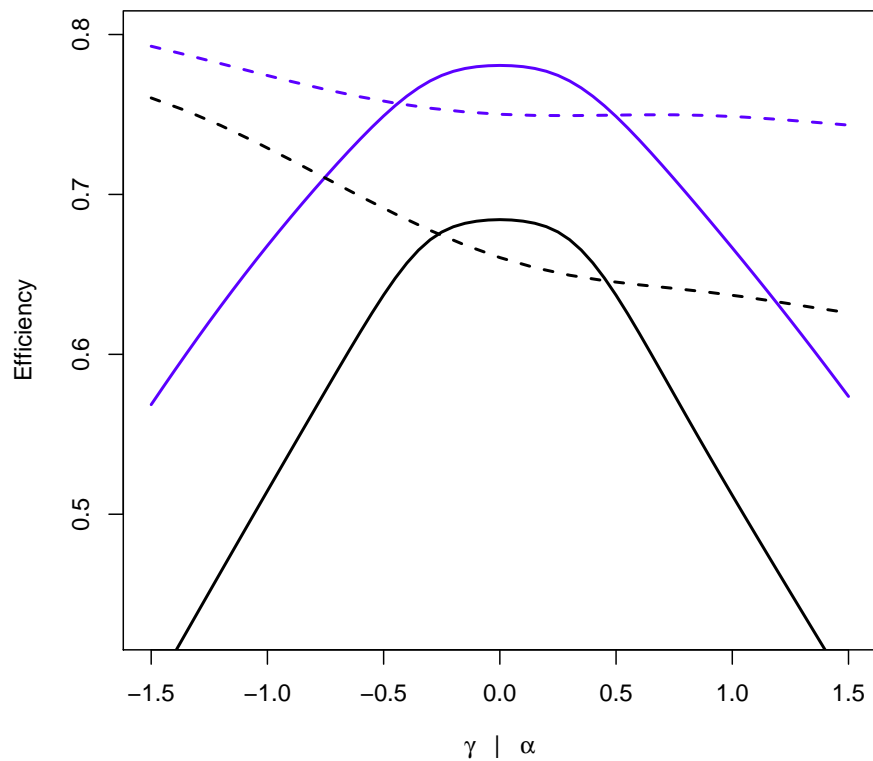


Figure 3. Efficiency of MD07 (blue) and MD10 (black) as a function of γ (solid) and of α (broken): $\mathbf{x}^* \sim N(\mathbf{0}, I_{p-1})$, $n = 5000$, $p = 5$; the solid line is obtained with $\alpha = 1$ and $\gamma \in [-1.5, 1.5]$; the broken line is obtained with $\gamma = 0.5$, and $\alpha \in [-1.5, 1.5]$.

Table 2. Monte Carlo efficiencies (%) of the MDE and MCML estimators of the Poisson regression for $\gamma = 0.5$, $\alpha = 1$, $\mathbf{x}^* \sim t_4(\mathbf{0}, I_{p-1})$.

p	n	δ	MDE07r	MDE10r	MCML07	MCML10	MCML+	MCML*
5	25	$\chi^2_{p,0.00}$	26	22	36	32	54	62
5	25	$\chi^2_{p,0.50}$	61	47	67	59	80	87
5	25	$\chi^2_{p,0.95}$	83	66	83	76	88	93
5	25	$\chi^2_{p,0.99}$	88	73	89	82	91	94
10	50	$\chi^2_{p,0.00}$	22	18	42	38	56	69
10	50	$\chi^2_{p,0.50}$	54	39	72	64	79	90
10	50	$\chi^2_{p,0.95}$	71	51	82	73	85	93
10	50	$\chi^2_{p,0.99}$	77	56	85	77	87	94
20	100	$\chi^2_{p,0.00}$	22	19	57	55	73	84
20	100	$\chi^2_{p,0.50}$	54	45	86	84	94	98
20	100	$\chi^2_{p,0.95}$	68	56	91	89	97	99
20	100	$\chi^2_{p,0.99}$	74	61	93	91	97	99

Table 3. Monte Carlo efficiencies (%) of the MDE and MCML estimators of the Poisson regression for $\gamma = 0.25, \alpha = 1, \mathbf{x}^* \sim t_4(\mathbf{0}, I_{p-1})$.

p	n	δ	MDE07r	MDE10r	MCML07	MCML10	MCML+	MCML*
5	25	$\chi_{p,0.00}^2$	47	42	58	55	78	83
5	25	$\chi_{p,0.50}^2$	73	72	84	81	95	97
5	25	$\chi_{p,0.95}^2$	85	87	93	91	98	99
5	25	$\chi_{p,0.99}^2$	90	91	96	94	99	99
10	50	$\chi_{p,0.00}^2$	40	35	67	65	82	87
10	50	$\chi_{p,0.50}^2$	67	66	91	89	97	99
10	50	$\chi_{p,0.95}^2$	78	79	96	94	99	100
10	50	$\chi_{p,0.99}^2$	83	84	97	95	99	100
20	100	$\chi_{p,0.00}^2$	39	37	79	78	89	93
20	100	$\chi_{p,0.50}^2$	65	69	96	95	99	100
20	100	$\chi_{p,0.95}^2$	73	79	98	97	100	100
20	100	$\chi_{p,0.99}^2$	77	93	98	98	100	100

The DCML procedure was computed for several values of δ , but only the results for $\delta = \chi_{p,0.0}^2 = 0$, $\delta = \chi_{p,0.5}^2$, $\delta = \chi_{p,0.95}^2$, and $\delta = \chi_{p,0.99}^2$ are shown. Note that the results corresponding to $\delta = 0$ are the efficiencies of the starting estimators. We observe that for $p = 5$ ($n = 25$) and $p = 10$ ($n = 50$), MD07r and MD10r are extremely inefficient. MCML07 and MCML10 somehow improve them, but are still very inefficient; they can exceed a 90% efficiency level only for $n = 100$. Good efficiency levels for all values of p (and n) are attained when the DCML procedure is started with MCML+ and MCML*, i.e., with MCML++ and MCML*+.

6.4. Simulations for the Contaminated Multiple Regression Model

In a final experiment, samples from Model (12) with $p = 10, \gamma = 0.5, \alpha = 1$, and $\mathbf{x}^* \sim N(\mathbf{0}, I_{p-1})$ were corrupted with point contaminations of the form $((1, \mathbf{x}_{\text{out}})^T, y_{\text{out}})$, where $\mathbf{x}_{\text{out}} = (0, x_0, 0, \dots, 0)$, x_0 varied in the set $\{-2, -1, 1, 2, 3, 4, 5, 6, 7\}$ and y_{out} in $\{0, 1, 2, 10, 20, 30, 40, 50\}$. We generated 100 samples of size 90; then, for each pair $(\mathbf{x}_{\text{out}}, y_{\text{out}})$, we added 10 observations with identical outliers of the form $((1, \mathbf{x}_{\text{out}})^T, y_{\text{out}})$; thus, the contamination level was 10%. Figure 4 shows the MSEE of the starting estimators as functions of y_{out} ; MCML10 is clearly the overall most robust one. Figure 5 shows the MSEE of the simple DCML estimators with $\delta = \chi_{p,0.99}^2$; here, MCML10 + D has the best behavior. In Figure 6, MCML10 + D is compared to the iterated DCML estimators MCML++ and MCML*+; the three graphs are almost superimposable. The numerical values of the maximum (over y_{out}) MSEE can be found in Table 4. It is worth noting that the robust estimators are quite sensitive to point contaminations in the neighborhood of zero (“excess zeros”), a problem that deserves further research. The DCML procedure lessens this kind of sensitivity.

Table 4. Monte Carlo maximum MSEE (over $y_{out} \in \{0, 1, 2, 10, 20, 30, 40, 50\}$) of the starting and DCML estimators of the Poisson regression under 10% point contamination in $((1, x_{out})^T, y_{out})$, where $x_{out} = (0, x_o, 0, \dots, 0)$ ($p = 10, \gamma = 0.5, \alpha = 1, x^* \sim N(0, I_{p-1})$).

x_0	-2	-1	1	2	3	4	5	6	7
MDE07r	0.12	0.10	0.12	0.14	0.16	0.15	0.16	0.14	0.12
MDE10r	0.12	0.13	0.12	0.15	0.15	0.16	0.17	0.16	0.12
MCML07	0.11	0.09	0.10	0.13	0.14	0.14	0.14	0.13	0.11
MCML10	0.11	0.09	0.10	0.13	0.13	0.14	0.14	0.13	0.11
MDE07r + D	0.11	0.08	0.08	0.10	0.13	0.12	0.14	0.13	0.10
MDE10r + D	0.13	0.10	0.09	0.13	0.12	0.13	0.13	0.13	0.11
MCML07 + D	0.09	0.06	0.07	0.08	0.12	0.11	0.13	0.12	0.09
MCML10 + D	0.08	0.06	0.07	0.09	0.10	0.10	0.11	0.11	0.09
MCML++	0.09	0.06	0.06	0.08	0.11	0.11	0.11	0.12	0.09
MCML*+	0.10	0.06	0.06	0.08	0.12	0.11	0.11	0.12	0.09

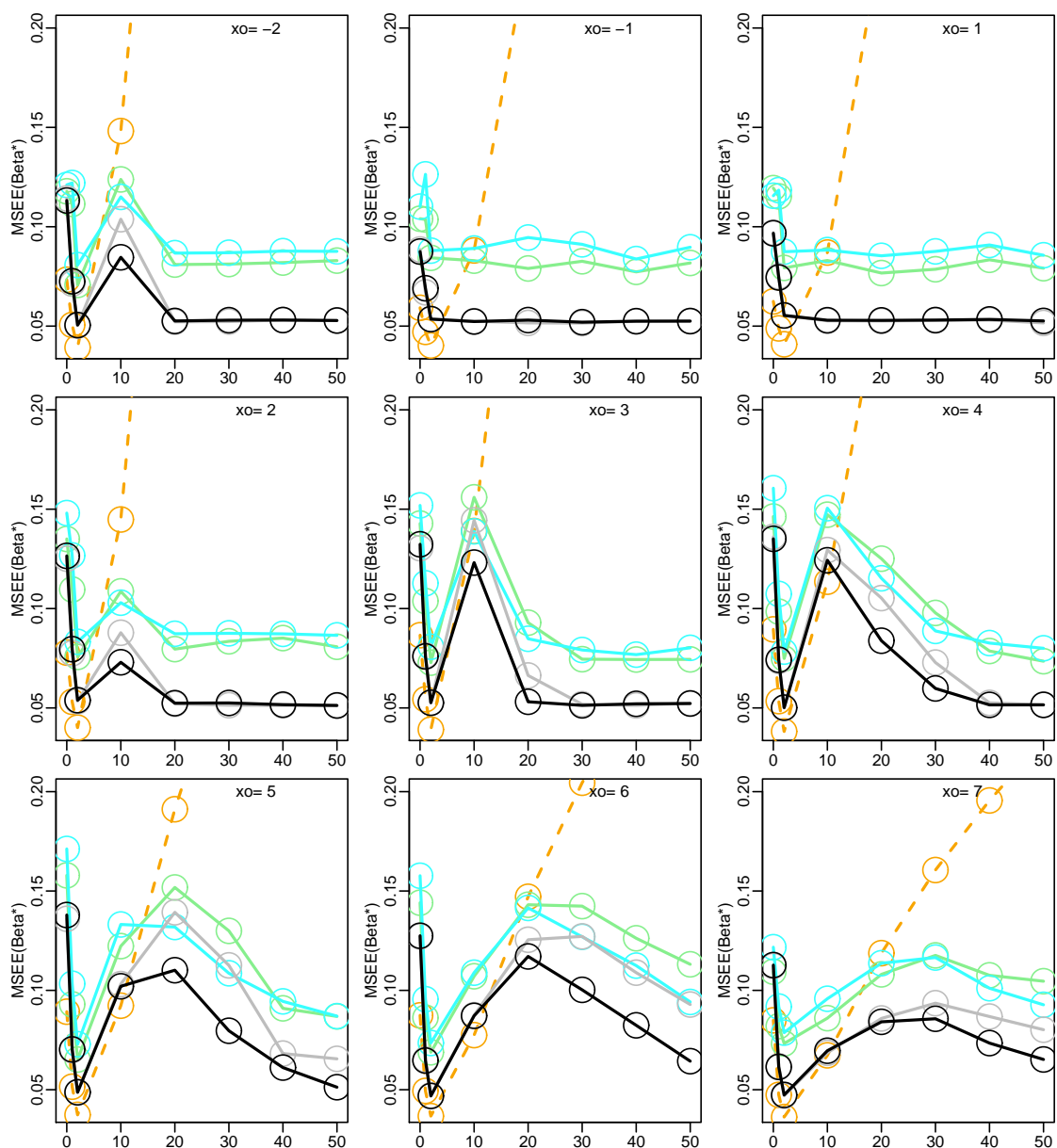


Figure 4. Monte Carlo results for the contaminated multiple regression model. Mean square estimation error $MSEE(\hat{\beta}^*)$ for varying y_{out} and x_{out} . Starting estimates: MD07r, light green; MD10r, cyan; MCML07, grey; MCML10, black; (ML, orange).

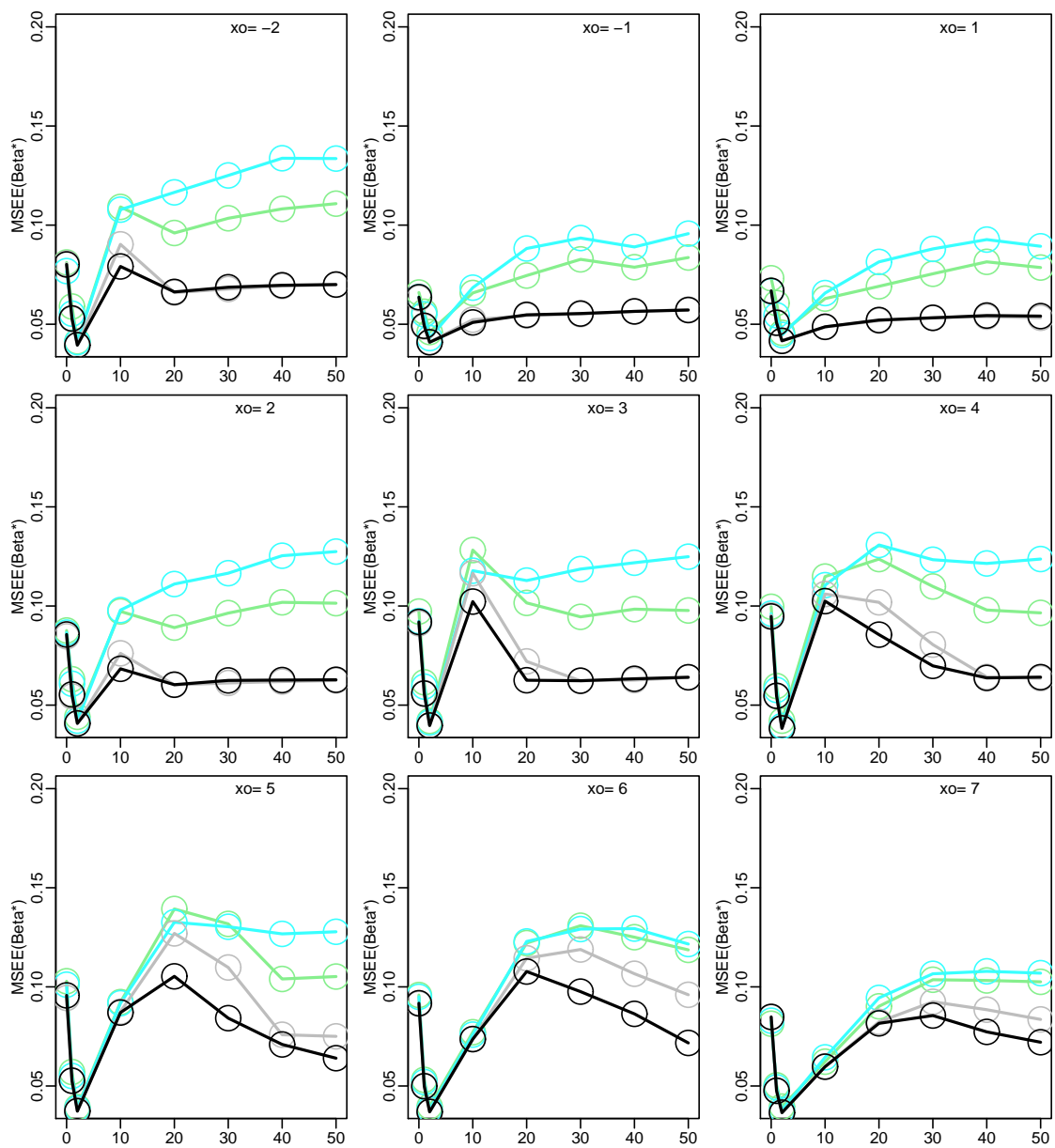


Figure 5. Monte Carlo results for the contaminated multiple regression model. Mean square estimation error $MSEE(\hat{\beta}^*)$ for varying y_{out} and x_{out} . Simple DCML estimates: MD07r + D, light green; MD10r + D, cyan; MCML07 + D, grey; MCML10 + D, black.

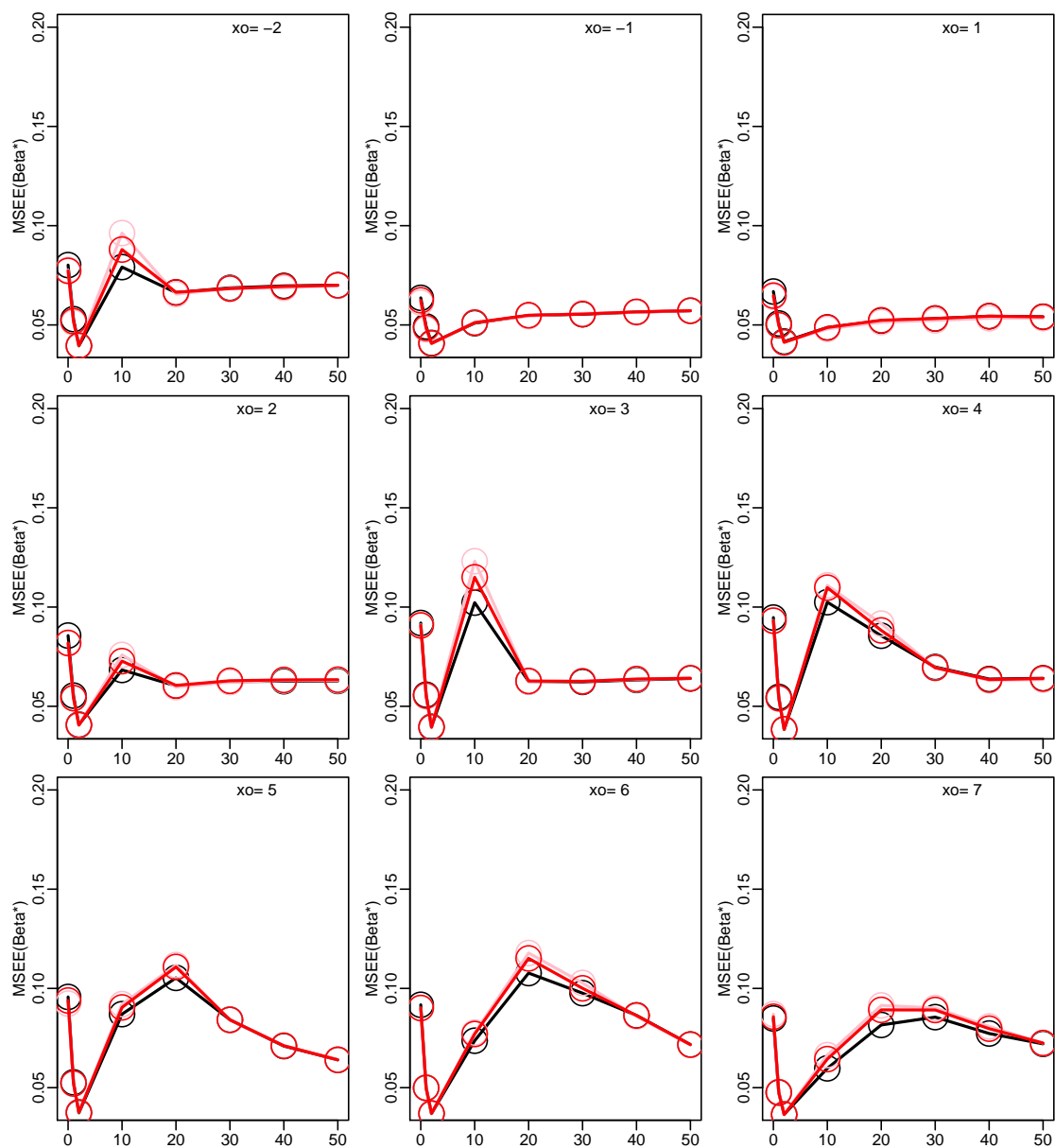


Figure 6. Monte Carlo results for the contaminated multiple regression model. Mean square estimation error $MSEE(\hat{\beta}^*)$ for varying y_{out} and x_{out} . Iterated DCML estimates: MCML++, pink; MCML*+, red; (MCML10 + D, black).

7. Bootstrapping Real Data and NB Fits

We consider 92 hospital stays belonging to the Major Diagnostic Category “poly-trauma” of the SwissDRG system. These data were described in [16], where the relationship between “length of stay” (y in days) and four covariates—“number of diagnosis” (x_1), “number of treatments” (x_2), “age of the patient” (x_3 in years), and “sex of the patient” ($x_4 = 0$ for males and $x_4 = 1$ for females)—was analyzed. Overdispersion is a usual feature of this kind of data, and the NB regression model has often been proposed to take it into account. In [16] an NB model with predictor $\exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_3 x_4)$ was estimated using the MCML procedure. The contamination fraction based on the MCML procedure was estimated at about 10%.

In a first experiment, we drew 1000 samples with a replacement of size 92, and for each sample, we estimated the six coefficients and the dispersion parameter by means of the ML, MD10r, MCML10, and MCML10 + D estimators of NB regression. We then computed the bootstrap coefficient means and standard errors, after 1% trimming of the most extreme

cases. In addition, we also computed the (1% trimmed) total variance of the estimates. The results are reported in Table 5. We observe that the coefficient means are a bit lower than those reported in [16]; the most important difference between ML and MCML10 is the estimated intercept. However, the standard errors are somewhat larger than those based on asymptotic approximations. The standard errors and total variance of MD10r are the largest ones; those of MCML10 and ML are smaller and very similar. As expected, the DCML procedure (with $\delta = \chi_{p,0.99}^2$) makes MCML10 closer to ML; however, it does not improve the variance estimates and does not seem to be very valuable in this case.

Table 5. Bootstrap coefficient means and standard errors of the ML, MDE10r, MCML10, and MCML10 + D estimators of the negative binomial (NB) regression for length of stay data ($n = 92$); var is the total variance; rej is the number of rejected samples.

	β_0	β_1	β_2	β_3	β_3	β_3	α	var	rej
ML	2.07 (0.47)	0.02 (0.03)	0.03 (0.02)	−0.01 (0.01)	0.09 (0.52)	0.00 (0.01)	0.88	0.49	0
MD10r	1.56 (0.54)	0.02 (0.05)	0.06 (0.04)	−0.01 (0.01)	0.42 (0.74)	0.01 (0.02)	0.67	0.85	0
MCML10	1.70 (0.43)	0.02 (0.04)	0.05 (0.04)	−0.01 (0.01)	0.29 (0.57)	0.01 (0.02)	0.48	0.51	1
MCML10 + D	1.95 (0.45)	0.02 (0.03)	0.04 (0.02)	−0.01 (0.01)	0.15 (0.50)	0.01 (0.01)	0.79	0.39	45

In a second experiment, we drew 1000 bootstrap samples of size 50 taken from a subset of the original data. Unfortunately, our estimation procedures of the six parameter model experienced several numerical malfunctions because many subsamples had ill-conditioned model matrices. We therefore used the four parameter predictor $\exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3)$. The results are shown in Table 6. We observe that the robust intercept means are still lower than the ML means. We also notice the awfully large standard errors of the MCML10 coefficient estimates due to their extreme tail behavior shown in the boxplots of Figure 7. Nevertheless, the DCML succeeds in getting rid of these tails and substantially reduces the standard errors and total variance.

Table 6. Bootstrap coefficient means and standard errors of the ML, MDE10r, MCML10, and MCML10 + D estimators of the NB regression for length of stay data ($n = 50$); var is the total variance; rej is the number of rejected samples.

	β_0	β_1	β_2	β_3	α	var	rej
ML	2.04 (0.64)	0.04 (0.04)	0.02 (0.03)	0.00 (0.01)	0.93	0.41	8
MD10r	1.59 (0.99)	0.04 (0.08)	0.02 (0.07)	0.00 (0.02)	0.85	0.99	0
MCML10	1.72 (1.21)	0.04 (0.35)	0.04 (0.31)	0.00 (0.14)	0.52	1.70	2
MCML10 + D	1.92 (0.67)	0.03 (0.05)	0.02 (0.04)	0.00 (0.01)	0.86	0.46	8

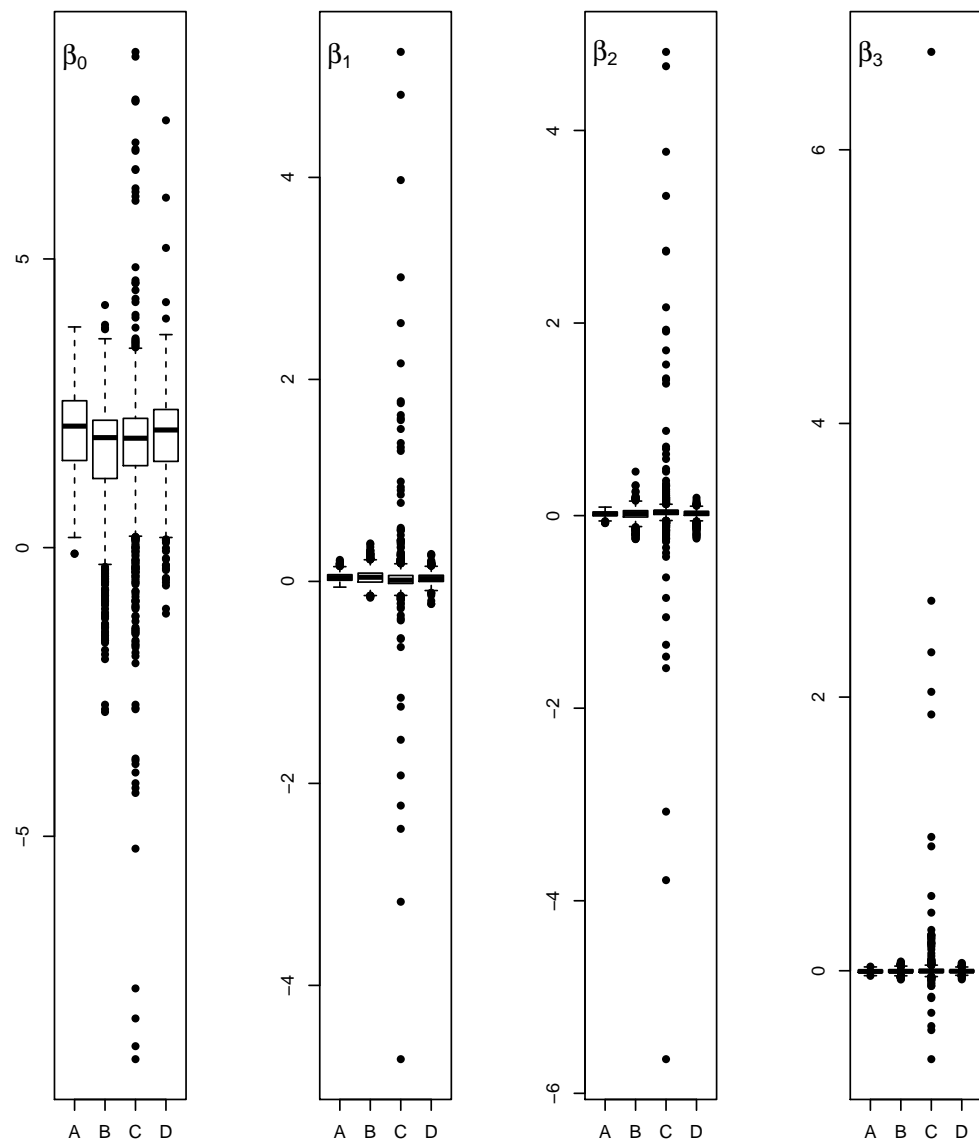


Figure 7. Boxplots of the coefficient estimates $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\beta}_2$, and $\hat{\beta}_3$ of the NB model for length of stays ($n = 50$); A = ML, B = MD10r, C = MCML10, D = MCML + D.

Remark: In our bootstrap experiments, several numerical breakdowns occurred. Both the function “glm.nb” (from the R package “MASS”) and our subsampling algorithm to compute MD10r stopped working in the presence of ill-conditioned model matrices. Other computations failed due to the occurrence of extreme predicted values, and remedies had to be implemented. For $n = 50$, a subsampling algorithm based on the Poisson model was used to compute the initial coefficient estimates followed by the solution of the MD10r equation for the NB dispersion parameter. As noted in Marazzi et al. (2019, Remark 5), although this procedure does not provide consistent initial coefficients under the NB model, the final MCML estimates are consistent. In addition, both for $n = 50$ and $n = 92$, samples associated with extreme predictions (larger than 25,000) were rejected to avoid program breakdowns. In the subsampling procedures, these samples were simply replaced with harmless ones, but we did not touch the “glm.nb” and “optim” functions. For these reasons, some of the results in Tables 5 and 6 are based on less than 1000 samples; the number of rejections is reported.

8. Conclusions

Robust estimators may be very inefficient when the sample size is not large; this shortcoming is also frequent for asymptotically fully efficient estimators. As a remedy,

Ref. [18] introduced the DCML method in the linear model case. In this paper, we introduced a simple modification of the original DCML that can be used for the GLM. The new DCML is simply defined as the convex combination of the robust and the ML estimator that minimizes a quadratic distance from the ML estimator under the constraint that the distance from the robust estimator is smaller than a given bound δ .

With the help of Monte Carlo experiments, we explored the performance of the DCML in the particular case of Poisson regression. For this purpose, we introduced a simplified version of the fully efficient MCML estimator where a very robust MD estimator is used as the initial step. A small Monte Carlo comparison of several published competitors indicated that this MCML was a very good candidate for the DCML procedure. We also introduced two iterated versions of the DCML procedure.

Unfortunately, Poisson regression is not regression invariant. This makes it difficult to infer general conclusions from the Monte Carlo experiments, in particular to find a general rule to set the bound δ . It seems, however, that a large quantile of the χ_p^2 distribution ($\delta \geq \chi_{p,0.95}^2$) is a convenient choice. We restricted our experiments to the case of spherical distributions of the covariates showing that, in this case, the efficiency essentially depends on the length of the slope vector. Monte Carlo experiments were based on a very unfavorable design. We observed, however, that the small sample efficiency of our best starting estimator MCML10 could be remarkably improved—without loss of robustness—with the help of the DCML method. It seems therefore reasonable to assume that the procedure works in practice as has been shown in an example with hospital length of stay fitted with the help of an NB regression model.

Funding: This research received no external funding.

Data Availability Statement: The data used in Section 6 are available in the R package “robustnegbin”.

Acknowledgments: I am very grateful to Ricardo Maronna and Victor Yohai for their critical remarks.

Conflicts of Interest: The author declares no conflict of interest.

Appendix A

The candidate estimators belong to the M estimator family (see e.g., [9]). In general, M estimators are defined by an “estimating equation” (5). The influence function, the matrices M_K and Q_K , and the asymptotic variance V_K of a general M estimator are given by (6)–(9), respectively. The special cases used in the Monte Carlo experiments are considered here.

The CUBIF estimator: The estimating equation of the CUBIF estimator with tuning parameter b is:

$$E_F \left[\varphi_{b/|z|} \left(y - c(\mathbf{x}^T \boldsymbol{\beta}, b/|z|) - h(\mathbf{x}^T \boldsymbol{\beta}) \right) \mathbf{x} \right] = 0,$$

where $\varphi_a(s) = \max[-a, \min(s, a)]$ is Huber’s function, $\mathbf{z} = \mathbf{A}\mathbf{x}$, \mathbf{A} satisfies:

$$\mathbf{A} E_F \left\{ \left[\varphi_{b/|z|} \left(y - c(\mathbf{x}^T \boldsymbol{\beta}, b/|z|) - h(\mathbf{x}^T \boldsymbol{\beta}) \right) \right]^2 \mathbf{x} \mathbf{x}^T \right\} \mathbf{A}^T = \mathbf{I}_p,$$

and the bias correction function $c(\boldsymbol{\theta}, a)$ is defined by:

$$\int [\varphi_a(y - c(\boldsymbol{\theta}, a) - h(\boldsymbol{\theta}))] f_{h(\boldsymbol{\theta})}(y) dy = 0.$$

Thus, the ψ -function of the CUBIF estimator is:

$$\tilde{\psi}(\mathbf{x}, y, \boldsymbol{\beta}) = \varphi_{b/|z|} \left(y - c(\mathbf{x}^T \boldsymbol{\beta}, b/|z|) - h(\mathbf{x}^T \boldsymbol{\beta}) \right).$$

We have:

$$\mathbf{M}_K = E_F \left[\tilde{\psi}(\mathbf{x}, y, \boldsymbol{\beta}) (y - h(\mathbf{x}^T \boldsymbol{\beta})) \mathbf{x} \mathbf{x}^T \right],$$

$$\mathbf{Q}_K = E_F \left[\tilde{\psi}(\mathbf{x}, y, \boldsymbol{\beta})^2 \mathbf{x} \mathbf{x}^T \right].$$

The RQL estimator: The estimating equation of the RQL estimator with tuning parameter c is:

$$E_F \left[\varphi_c \left(\frac{y - \mu_x}{\mu_x^{1/2}} \right) \mu_x^{1/2} \mathbf{x} - \alpha(\boldsymbol{\beta}) \right] = 0,$$

where φ_c is Huber’s function and:

$$\alpha(\boldsymbol{\beta}) = E_F \left[\varphi_c \left(\frac{y - \mu_x}{\mu_x^{1/2}} \right) \right] \mu_x^{1/2} \mathbf{x}.$$

Thus, the ψ -function of the RQL estimator is:

$$\varphi_c \left(\frac{y - \mu_x}{\mu_x^{1/2}} \right) \mu_x^{1/2} \mathbf{x} - \alpha(\boldsymbol{\beta}).$$

The formulae to compute \mathbf{M}_K and \mathbf{Q}_K were given in [10]. The function “glmrob” in “robustbase” implements these formulae.

The MD estimator: The estimating equation of the MD estimator with tuning parameter τ is:

$$E_F \left[\gamma(\mu_x) - (y - \mu_x) f_{\mu_x}(y)^\tau \right] \mathbf{x} = 0,$$

where $\gamma(\mu) = \sum_{k=1}^{\infty} (k - \mu) f_{\mu}(k)^{1+\tau}$. We obtain:

$$\mathbf{M}_K = E \left[\left[\gamma'(\mu_x) \mu_x + \left[\mu_x - \tau(y - \mu_x)^2 \right] f_{\mu_x}(y)^\tau \right] \mathbf{x} \mathbf{x}^T \right],$$

$$\mathbf{Q}_K = E \left[\left(\gamma(\mu_x) - (y - \mu_x) f_{\mu_x}(y)^\tau \right)^2 \mathbf{x} \mathbf{x}^T \right],$$

where:

$$\gamma'(\mu) = \sum_{k=1}^{\infty} q(k, \mu, \tau),$$

and:

$$q(k, \mu, \tau) = [-1 + (k - \mu)(1 + \tau)(k/\mu - 1)] f_{\mu}(k)^{\tau+1}.$$

The MT estimator: The estimating equation of the MT estimator with tuning parameter c is:

$$E \left[\rho'_c \left((\sqrt{y} - m(\mu_x)) m'(\mu_x) \mu_x \right) \mathbf{x} \right] = 0,$$

where ρ_c is Tukey’s biweight:

$$\rho_c(u) = 1 - \max \left(\left(1 - \left(\frac{u}{c} \right)^2 \right)^3, 1 \right),$$

$m(\mu)$ is the consistency correction function:

$$m(\mu) = \arg \min_{\gamma} E \left[\rho_c(\sqrt{y}) - \gamma \right].$$

We have:

$$\mathbf{M}_K = E \left[\left(-\rho''_c(\sqrt{y} - m(\mu_x)) m'(\mu_x)^2 \mu_x^2 + \rho'_c(\sqrt{y} - m(\mathbf{x}^T \boldsymbol{\beta})) (m''(\mu_x) \mu_x^2 + \mu_x) \right) \mathbf{x} \mathbf{x}^T \right],$$

$$\mathbf{Q}_K = E \left[\rho'_c(\sqrt{y} - m(\mu_x))^2 m'(\mu_x)^2 \mu_x^2 \mathbf{x} \mathbf{x}^T \right].$$

The MCML estimator: The MCML estimator is based on the conditional density of y given x and $z \in [a, b]$, where $z \sim U[0, 1]$. This distribution can be written as:

$$p_{\mu_x}(y | z \in [a, b]) = f_{\mu_x}(y)W_{a,b}(\mu_x),$$

where $W_{a,b}(\mu_x)$ is defined in the following. For any c , let:

$$y_x^*(c) = \max\{y : F_{\mu_x}(y) \leq c\}$$

and:

$$t_{c,x} = (F_{\mu_x}(y_x^*(c) + 1) - c) / f_{\mu_x}(y_x^*(c) + 1),$$

where F_{μ} denotes the cdf of f_{μ} . Put:

$$T_{a,x} = y_x^*(a) + 2, \quad T_{b,x} = y_x^*(b),$$

$$A_x = \{y : T_{a,x} \leq y \leq T_{b,x}\},$$

$$Q(\mu_x) = F_{\mu_x}(T_{b,x}) - F_{\mu_x}(T_{a,x} - 1) + f_{\mu_x}(T_{a,x} - 1)t_{a,x} + f_{\mu_x}(T_{b,x} + 1)(1 - t_{b,x}),$$

$$L_x = \{y | y = T_{a,x} - 1\}, \quad A_x = \{y | T_{a,x} \leq y \leq T_{b,x}\}, \quad U_x = \{y | y = T_{b,x} + 1\},$$

$$H_x = L_x \cup A_x \cup U_x.$$

Then:

$$W_{a,b}(\mu_x) = \begin{cases} \frac{1}{Q(\mu_x)} & \text{if } y \in A_x, \\ \frac{t_{a,x}}{Q(\mu_x)} & \text{if } y \in L_x, \\ \frac{1-t_{b,x}}{Q(\mu_x)} & \text{if } y \in U_x, \\ 0 & \text{if elsewhere.} \end{cases}$$

We approximately compute the influence function assuming that a and b are fixed. The estimating equation is:

$$E_F \left[\left(\bar{\psi}(x, y, \mu_x) h'(x^T \beta) \right) x \right] = 0,$$

i.e.:

$$\int \sum_{T_{a,x}-1 \leq k \leq T_{b,x}+1} \bar{\psi}(x, k, \mu_x) h'(x^T \beta) x p(k|x) dG(x) = 0,$$

where:

$$\bar{\psi}(x, y, \mu_x) = \begin{cases} s(y, \mu_x) - D(\mu_x) & \text{if } y \in A_x \\ t_{a,x}[s(y, \mu_x) - D(\mu_x)] & \text{if } y \in L_x \\ (1 - t_{b,x})[s(y, \mu_x) - D(\mu_x)] & \text{if } y \in U_x \\ 0 & \text{otherwise} \end{cases}$$

$$s(y, \mu) = \partial / \partial \mu [\ln f_{\mu}(y)] = \left(\frac{y}{\mu} - 1 \right),$$

$$D(\mu) = \frac{\partial}{\partial \mu} \ln Q(\mu).$$

We obtain:

$$M_K = E_F \left[\left(\bar{\psi}_{\mu}(x, y, \mu_x) h'(x^T \beta) \right)^2 + \psi(x, y, \mu_x) h''(x^T \beta) \right] x x^T,$$

$$Q_K = E_F \left[\left(\bar{\psi}(x, y, \mu_x) h'(x^T \beta) \right)^2 \right] x x^T,$$

where $\psi_{\mu}(y, \mu) = (\partial / \partial \mu) \psi(y, \mu)$.

References

1. Taylor, L.D. Estimation by minimizing the sum of absolute errors. In *Frontiers in Econometrics*; Zarembka, P., Ed.; Academic Press: New York, NY, USA, 1973; pp. 189–190.
2. Dodge, Y.; Jurečková, J. Adaptive combination of least squares and least absolute deviations estimators. In *Statistical Data Analysis Based on L1-Norm and Related Methods*; Dodge, Y., Ed.; Springer: New York, NY, USA, 2000.
3. Yohai, V.J. High breakdown-point and high efficiency robust estimates for regression. *Ann. Stat.* **1987**, *15*, 642–656. [[CrossRef](#)]
4. Yohai, V.J.; Zamar, R.H. High breakdown estimates of regression by means of the minimization of an efficient scale. *J. Am. Stat. Assoc.* **1988**, *83*, 406–413. [[CrossRef](#)]
5. Gervini, D.; Yohai, V.J. A class of robust and fully efficient regression estimators. *Ann. Stat.* **2002**, *30*, 583–616. [[CrossRef](#)]
6. Marazzi, A.; Yohai, V.J. Adaptively truncated maximum likelihood regression with asymmetric errors. *J. Stat. Plan. Inference* **2004**, *122*, 271–291. [[CrossRef](#)]
7. Čížek, P. Semiparametrically weighted robust estimation of regression models. *Comput. Stat. Data Anal.* **2011**, *55*, 774–788. [[CrossRef](#)]
8. Davison, A.C.; Snell, E.J. Residuals and diagnostics. In *Statistical Theory and Modelling: In Honour of Sir David Cox*; Hinkley, D.V., Reid, N., Snell, E.J., Eds.; Chapman and Hall: London, UK, 1991; pp. 83–106.
9. Maronna, R.A.; Martin, R.D.; Yohai, V.J. *Robust Statistics*; Wiley: Hoboken, NJ, USA, 2019.
10. Künsch, H.R.; Stefanski, L.A.; Carroll, R.J. Conditionally unbiased bounded-influence estimation in general regression models, with applications to generalized linear models. *J. Am. Stat. Assoc.* **1989**, *84*, 460–466.
11. Cantoni, E.; Ronchetti, E. Robust inference for generalized linear models. *J. Am. Stat. Assoc.* **2011**, *96*, 1022–1030. [[CrossRef](#)]
12. Neykov, N.M.; Filzmoser, P.; Neytchev, P.N. Robust joint modeling of mean and dispersion through trimming. *Comput. Stat. Data Anal.* **2012**, *56*, 34–48. [[CrossRef](#)]
13. Valdora, M.; Yohai, V.J. Robust estimation in generalized linear models. *J. Stat. Plan. Inference* **2014**, *146*, 31–48. [[CrossRef](#)]
14. Ghosh, A.; Basu, A. Robust Estimation in Generalized Linear Models: The Density Power Divergence Approach. *TEST* **2016**, *25*, 269–290. [[CrossRef](#)]
15. Aeberhard, W.H.; Cantoni, E.; Heritier, S. Robust inference in the negative binomial regression model with an application to falls data. *Biometrics* **2014**, *70*, 920–931. [[CrossRef](#)] [[PubMed](#)]
16. Marazzi, A.; Valdora, M.; Yohai, V.J.; Amiguet, M. A robust conditional maximum likelihood estimator for generalized linear models with a dispersion parameter. *Test* **2019**, *28*, 223–241. [[CrossRef](#)]
17. Bondell, H.D.; Stefanski, L.A. Efficient robust regression via two-stage generalized empirical likelihood. *J. Am. Stat. Assoc.* **2013**, *108*, 644–655. [[CrossRef](#)] [[PubMed](#)]
18. Maronna, R.A.; Yohai, V.J. High finite sample efficiency and robustness based on distance-constrained maximum likelihood. *Comput. Stat. Data Anal.* **2015**, *83*, 262–274. [[CrossRef](#)]
19. Dunn, P.K.; Smyth, G.K. Randomized quantile residuals. *J. Comput. Graph. Stat.* **1996**, *5*, 236–244.
20. Han, A.K. Non-parametric analysis of a generalized regression model: The maximum rank correlation estimator. *J. Econom.* **1987**, *35*, 303–316. [[CrossRef](#)]
21. Rousseeuw, P.J. Multivariate estimation with high breakdown point. In *Mathematical Statistics and Applications*; Grossman, W., Pflug, G., Vincze, I., Wertz, W., Eds.; Reidel Publishing: Dordrecht, The Netherlands, 1985; pp. 283–297.
22. Huber, P.J. The 1972 Wald Lecture, Robust Statistics: A Review. *Ann. Math. Stat.* **1972**, *43*, 1041–1067. [[CrossRef](#)]