A robust adaptive modified maximum likelihood estimator for the linear regression model

Sukru Acitas, Peter Filzmoser and Birdal Senoglu

Department of Statistics, Eskisehir Technical University, Eskisehir, Turkey; Institute of Statistics and Mathematical Methods in Economics, Vienna University of Technology, Vienna, Austria; Department of Statistics, Ankara University, Ankara, Turkey

ABSTRACT
Robust estimators are widely used in regression analysis when the normality assumption is not satisfied. One example of robust estimators for regression is adaptive modified maximum likelihood (AMML) estimators [Donmez A. Adaptive estimation and hypothesis testing methods [dissertation]. Ankara: METU; 2010]. However, they are not robust to outliers, so-called leverage points. In this study, we propose a new estimator called robust AMML (RAMML) which is not only robust to outliers but also to leverage points. A simulation study is carried out to compare the performance of the RAMML estimators with some existing robust estimators. The results show that the RAMML estimators are preferable in most of the settings according to the mean squared error (MSE) criterion. Two data sets taken from the literature are also analyzed to show the implementation of the RAMML estimation methodology.

1. Introduction
Regression analysis is a widely used statistical tool to analyse the relationships between a dependent variable and predictors. The least squares (LS) method is usually utilized to estimate the unknown model parameters. The LS estimator is popular because of its simple computation and good statistical properties, i.e. it is the best linear unbiased estimator under the normality assumption.

The LS estimator loses its efficiency when the normality assumption is not satisfied. Therefore, some alternative estimators are employed to accommodate the problems arising in case of non-normality which is caused by outliers. Indeed, there are two types of outliers in the context of regression analysis: (i) vertical outliers, and (ii) leverage points. These outliers refer to outlyingness in the space of the response and the predictors, respectively. The LS estimator is sensitive to both types of outliers and bad leverage points can even distort the estimator and lead to a non-sense model [1]. For this reason, many robust estimators which reduce the effects of outliers have been proposed in the literature. For
example, the M estimator [2], the S estimator [3], the least trimmed squares (LTS) and the least median of squares (LMS) of Rousseeuw [4], and the MM estimator [5] are well-known and widely used in the context of robust regression. We refer to Maronna et al. [6] or Gschwandtner and Filzmoser [7] in which descriptions of these estimators and related literature information are briefly given. See also Yu and Yao [8] for an extensive review of robust linear regression estimators.

In the literature, the modified maximum likelihood (MML) estimator is also used to reduce the effect of outlying observations. As its name indicates, the MML estimator is obtained applying a particular modification of the maximum likelihood (ML) method [9,10]. This modification allows to obtain explicit forms of the estimator under the assumption of a non-normal error distribution. A long-tailed symmetric distribution is frequently used in this context since it is useful for modelling outliers occurring in the direction of long tails [11]. For example, Tiku et al. [12] obtain the MML estimator of the unknown parameters of the simple linear regression model when the distribution of the error terms is long-tailed symmetric. The MML estimator is also obtained for the multiple linear regression model [13] with long-tailed symmetric error distributions. The MML estimator has good properties, and it is easy to compute. Furthermore, it is not only robust to outliers but also asymptotically equivalent to the ML estimator. However, it has two drawbacks: (i) the shape parameter of the long-tailed symmetric distribution is assumed to be known and (ii) it is only robust to vertical outliers but not to leverage points.

The first drawback is not a new issue in the related literature. For example, Lucas [14] consider the Student’s $t$ distribution as an alternative to normal distribution for estimating the location parameter under the assumption of known degrees of freedom $\nu$. This is because of the fact that the ML estimator of the other parameters is no longer robust when it is estimated jointly with $\nu$. Therefore, it is treated as a robustness tuning constant – see also Arslan and Genc [15], Acitas et al. [16,17] in which similar discussions are available. Tiku and Surucu [18] also consider this issue and propose a new version of the MML estimator. This version is then called adaptive MML (AMML) by Donmez [19]. In AMML, the main idea of the MML methodology is preserved but the assumption of a known shape parameter on the estimation process is weakened.

To the best of our knowledge, the second drawback is still an open problem. In other words, there is no previous study on the MML and AMML estimators dealing with leverage points as far as we know. Therefore, the motivation of this study is to suggest a new regression estimator which is robust to both types of outliers based on the AMML estimators. This estimator is called robust AMML (RAMML) and obtained by introducing new weights assigned to the predictors in the AMML estimation process [20]. The RAMML is not only robust to both types of outliers but also efficient under the normality assumption. A further advantage of the RAMML estimator is that it is easy to compute since the explicit form of the estimator is available.

The rest of the paper is organized as follows. Section 2 reviews the AMML methodology. Section 3 is reserved to the RAMML estimator. The results of a Monte-Carlo simulation study are given in Section 4. Section 5 includes two real life data taken from the literature to show the implementation of the RAMML estimation methodology. The paper is finalized with some concluding remarks.
2. Review of the AMML estimator

We consider a multiple linear regression model with \( n \) observations and \( m \) predictors,

\[
y_i = \beta_0 + \mathbf{x}_i' \mathbf{\beta} + \epsilon_i, \quad i = 1, 2, \ldots, n
\]  

(1)

where \( y_i \) is the dependent variable, \( \beta_0 \) is the intercept, \( \mathbf{\beta} = (\beta_1 \beta_2 \cdots \beta_m)' \) is the vector of unknown regression coefficients, \( \mathbf{x}_i \) is the vector of predictors defined by \( \mathbf{x}_i = (x_{i1} \ x_{i2} \cdots \ x_{im})' \) and \( \epsilon_i \) is the random error term.

Assume that the distribution of the error terms in model (1) is long-tailed symmetric with the probability density function (pdf) given by

\[
f(\epsilon; p, \sigma) = \frac{1}{\sqrt{qB(0.5, p - 0.5)}\sigma} \left(1 + \frac{\epsilon^2}{q\sigma^2}\right)^{-p}, \quad q = 2p - 3, \ -\infty < \epsilon < \infty
\]  

(2)

where \( p \) is the shape parameter which is assumed to be \( p \geq 2 \), \( \sigma \) is the scale parameter and \( B(\cdot, \cdot) \) is the beta function. The kurtosis value of the long-tailed symmetric distribution is greater than 3. It tends to normal distribution as \( p \) tends to \( \infty \). Therefore, it has been considered as a good alternative to the normal distribution [11].

The ML estimators of the model parameters are the solutions of the following likelihood equations:

\[
\frac{\partial \log L}{\partial \beta_0} = \frac{2p}{q\sigma} \sum_{i=1}^{n} g(z_i) = 0,
\]  

(3)

\[
\frac{\partial \log L}{\partial \beta_j} = \frac{2p}{q\sigma} \sum_{i=1}^{n} g(z_i)x_{ij} = 0, \quad j = 1, 2, \ldots, m
\]  

(4)

\[
\frac{\partial \log L}{\partial \sigma} = -\frac{n}{\sigma} + \frac{2p}{q\sigma} \sum_{i=1}^{n} g(z_i)z_i = 0,
\]  

(5)

where

\[
g(z_i) = \frac{z_i}{\sqrt{1 + \frac{z_i^2}{q}}}, \quad z_i = \frac{y_i - \beta_0 - \mathbf{x}_i' \mathbf{\beta}}{\sigma}, \quad i = 1, 2, \ldots, n.
\]

Since \( g(\cdot) \) is a nonlinear function of the unknown model parameters, the explicit solutions of Equations (3)–(5) cannot be obtained, and numerical methods need to be used instead. However, using numerical methods has some drawbacks such as wrong convergence, multiple roots and non-convergence of iterations, see e.g. Puthenpura and Sinha [21], Vaughan [22]. The modified maximum likelihood (MML) methodology proposed in Tiku [9,10] overcomes these problems and allows to obtain the explicit forms of the estimators. The idea underlying the MML methodology is linearization of the intractable term (i.e. the function \( g(\cdot) \)) in the likelihood equations. This is done using the first two terms of a Taylor series expansion. For more details, we refer to Islam and Tiku [13] in which MML estimators for the multiple linear regression model are obtained under a long-tailed symmetric error distribution assumption.
The MML estimators are explicitly formulated, thus no computational efforts are required. They have also good statistical properties, i.e. they are asymptotically equivalent to the ML estimators and therefore they are minimum variance bound estimators [23]. However, the MML methodology works under the assumption of a known shape parameter \( p \). In some studies, \( p \) is identified using the profile likelihood method, see for example Acitas et al. [16,17]. Tiku and Surucu [18] suggested a new version of the MML methodology in which the known shape parameter assumption is weakened. Tiku and Surucu [18] obtain estimates of the location and the scale parameters using this new version and show that they are as good as or better than M-estimators. Donmez [19] generalize this methodology to the simple and multiple linear regression model and call this methodology as AMML. The steps of the AMML are explained based on Donmez [19] as follows:

**Step 1.** The function \( g(\cdot) \) is linearized around

\[
t_i = \frac{y_i - \beta_0 - x_i'\beta}{\sigma}, \quad i = 1, 2, \ldots, n
\]

using the first two terms of a Taylor series expansion,

\[
g(z_i) \cong \alpha_i + \delta_i z_i, \quad i = 1, 2, \ldots, n
\]

where

\[
\alpha_i = \frac{(1/q)t_i}{(1 + (1/q)t_i^2)^2}, \quad \delta_i = \frac{1}{(1 + (1/q)t_i^2)^2}, \quad i = 1, 2, \ldots, n.
\]

**Step 2.** The linearized version of the function \( g(z_{i(j)}) \) is incorporated into the likelihood equation and the so-called modified likelihood equations are obtained as follows:

\[
\frac{\partial \log L^*}{\partial \beta_0} = \frac{2p}{q\sigma} \sum_{i=1}^{n} (\alpha_i + \delta_i z_i) = 0
\]

\[
\frac{\partial \log L^*}{\partial \beta_j} = \frac{2p}{q\sigma} \sum_{i=1}^{n} (\alpha_i + \delta_i z_i) x_{ij} = 0, \quad j = 1, 2, \ldots, m
\]

\[
\frac{\partial \log L^*}{\partial \sigma} = -\frac{n}{\sigma} + \frac{2p}{q\sigma} \sum_{i=1}^{n} (\alpha_i + \delta_i z_i) z_i = 0
\]

where \( \log L^* \) stands for the modified log-likelihood function.

**Step 3.** The solutions of the Equations (9)–(11) are the AMML estimators. They are formulated as follows:

\[
\hat{\beta}_0 = \tilde{y}_{[1]} - \tilde{x}_{[1]} \hat{\beta}
\]

\[
\hat{\beta} = K + L\hat{\sigma}
\]

\[
\hat{\sigma} = \frac{B + \sqrt{B^2 + 4nC}}{2\sqrt{n(n-m-1)}},
\]

where

\[
\tilde{y}_{[1]} = \frac{1}{n} \sum_{i=1}^{n} y_i, \quad \tilde{x}_{[1]} = \frac{1}{n} \sum_{i=1}^{n} x_i, \quad B = \sum_{i=1}^{n} (\alpha_i + \delta_i z_i) x_{i1}, \quad C = \sum_{i=1}^{n} (\alpha_i + \delta_i z_i) x_{i2},
\]

and the other quantities are defined as in the original AMML methodology.
where

\[
K = (M^\prime \delta M)^{-1} (M^\prime \delta y) = (K_j)_{m \times 1}, \quad L = (M^\prime \delta M)^{-1} (M^\prime \alpha) = (L_j)_{m \times 1}
\]

\[
B = \frac{2p}{q} \sum_{i=1}^{n} \alpha_i \left( y_i - \bar{y}_{[.]} - \sum_{j=1}^{m} K_j M_{ij} \right),
\]

\[
C = \frac{2p}{q} \sum_{i=1}^{n} \delta_i \left( y_i - \bar{y}_{[.]} - \sum_{j=1}^{m} K_j M_{ij} \right)^2
\]

\[
\delta = \text{diag}(\delta_i) \quad \text{and} \quad \alpha = \text{diag}(\alpha_i),
\]

\[
\bar{y}_{[.]} = \frac{1}{n} \sum_{i=1}^{n} \delta_i y_i, \quad \bar{x}_{[.]j} = \frac{1}{n} \sum_{i=1}^{n} \delta_i x_{ij}
\]

\[
\bar{x}_{[.]} = [\bar{x}_{[.]1} \bar{x}_{[.]2} \cdots \bar{x}_{[.]m}].
\]

Here, \( \mathbf{1} \) stands for the vector of ones with \( n \) entries.

It is clear that AMML estimators are also explicitly formulated. They preserve the properties of the MML estimators [19]. It can easily be shown that the AMML estimators are regression, scale and affine equivariant. The proof is not given here for the sake of brevity, but it can be provided upon request.

As it is clear, the values \( t_i \), given in Step 1, should be computed at the beginning of the methodology using some initial estimates. The computational details are explained in the following subsection [19].

### 2.1. Computation of the AMML estimators

The computation of the AMML estimators is obtained in two iterations. An initial estimate of \( t_i \) is required in the first iteration. The initial estimate of \( t_i \), proposed by [19], is obtained based on a reparametrization of the model (1). The reparametrized model is constructed assuming that all \( \beta_j \) coefficients are equal in model (1). This is because of the fact that there is no reason to believe that one predictor is more important than others at the beginning. The reparametrized model is formulated by

\[
y_i = \beta_0 + \theta v_i + \epsilon_i, \quad v_i = \sum_{j=1}^{m} x_{ij}, \quad i = 1, 2, \ldots, n.
\]

Then, the initial estimate of \( t_i \) is given as

\[
\tilde{t}_i = \frac{y_i - T_0}{T_1 v_i} \quad \text{with} \quad T_0 = \sum_{i=1}^{n} \bar{x}_{[.]}, \quad T_1 = \sum_{i=1}^{n} \bar{y}_{[.]}, \quad \epsilon_i = y_i - \beta_0 - \theta v_i.
\]
where the initial estimators of $\beta_0$, $\theta$ and $\sigma$ are given by

$$T_0 = \text{median}\{y_i - T_1v_i\}, \quad T_1 = \text{median}\left\{\frac{y_{\ell+1} - y_{\ell}}{\nu_{\ell+1} - \nu_{\ell}}\right\}$$  \hspace{1cm} (17)$$

and

$$S_0 = 1.483 \cdot \text{median}\{|y_i - T_0 - T_1v_i|\}, \quad i = 1, 2, \ldots, n,$$  \hspace{1cm} (18)$$

respectively. After getting the initial estimate of $t_i$, initial AMML estimators denoted by $\tilde{\beta}_0$, $\tilde{\beta}_j$ and $\tilde{\sigma}$ are obtained following steps 1-3 given in Section 2.

In the second iteration, $t_i$ is updated using the initial AMML estimators ($\tilde{\beta}_0$, $\tilde{\beta}_j$ and $\tilde{\sigma}$) as follows:

$$t_i = \frac{y_i - \tilde{\beta}_0 - \sum_{j=1}^{n} x_{ij}\tilde{\beta}_j}{\tilde{\sigma}}$$  \hspace{1cm} (19)$$

and steps 1–3 are utilized one more time. The final AMML estimates are obtained at the end of step 3 using Equations (12)–(14).

It is clear that the shape parameter of the long-tailed symmetric distribution $p$ appears in the formulas. It is considered as robustness tuning constant and taken to be 16.5 so that the resulting AMML estimators are efficient when the distribution of the error terms is normal [18,19].

Using different simulation schemes, Donmez [19] shows that AMML estimators are robust to $y$—outliers. On the other hand, there are no available results whether they are resistant to $x$—outliers. Indeed, the simulation schemes used in Donmez [19] do not include any $x$—outliers. It should also be mentioned that Donmez [19] use standardized versions of predictors to reduce the effect of $x$—outliers, see also Akkaya and Tiku [24]. However, the standardization is done using mean and standard deviation which are not robust to outliers. Standardization is indeed required when working with the reparametrized model (see Equation (15)) in which the sum of the predictors is used. If the predictors are measured under different scales and/or units, this sum is not remarkable. In our proposed methodology, we do not need any standardization for the predictors and the reparametrized model. Therefore, these AMML estimators are not considered in the rest of the paper.

3. Robust AMML estimators

In this section, we propose RAMML estimators which are robust to $x$—outliers unlike the AMML. This is done by downweighting the $x$—outliers using extra weights for predictors. The steps for the RAMML methodology are similar to those given for AMML, i.e. step 1 is the same while steps 2 and 3 have to be revised. New versions of step 2 and 3 are shown by Step 2* and 3*, respectively, and explained below.

Step 2*. Consider the modified likelihood Equations (9)–(11). We introduce new weights, denoted by $\delta_i^x$, for the predictors to accommodate the effects of $x$—outliers in these
equations:

$$\frac{\partial \log L^*}{\partial \beta_0} = \frac{2p}{q\sigma} \sum_{i=1}^{n} \delta_i^x (\alpha_i + \delta_i z_i) = 0$$  \hspace{1cm} (20)$$

$$\frac{\partial \log L^*}{\partial \beta_j} = \frac{2p}{q\sigma} \sum_{i=1}^{n} \delta_i^x (\alpha_i + \delta_i z_i) x_{ij} = 0, \hspace{0.5cm} j = 1, 2, \ldots, m$$  \hspace{1cm} (21)$$

$$\frac{\partial \log L^*}{\partial \sigma} = -\frac{n}{\sigma} + \frac{2p}{q\sigma} \sum_{i=1}^{n} \delta_i^x (\alpha_i + \delta_i z_i) z_i = 0,$$  \hspace{1cm} (22)$$

where

$$\delta_i^x = \frac{1}{(1 + (1/q)\tilde{x}_i^2)^4} \hspace{1cm} \text{and} \hspace{1cm} \tilde{x}_i = \frac{||x_i - \text{med}_{L_1}(X)||}{\text{median}_i ||x_i - \text{med}_{L_1}(X)||}, \hspace{1cm} i = 1, 2, \ldots, n.$$  \hspace{1cm} (23)$$

Here, \(\text{med}_{L_1}(\cdot)\) denotes the \(L_1\)-median which is also known as geometric median or spatial median. It is a highly robust estimator for the multivariate location, i.e. its breakdown point is 0.5 [25]. We refer to Croux et al. [26] and Fritz et al. [27] in which detailed information is available about \(L_1\)-median.

Step 3*. The RAMML estimators are obtained as solutions of Equations (20)–(22). They are formulated as given in Equations (12)–(14). However, there are some modifications due to the new weighting scheme. They are given as follows:

$$\hat{\beta}_0 = \bar{y}_{[-|]} - \bar{x}_{[-|]} \hat{\beta} + \frac{\Delta}{w} \hat{\sigma}$$  \hspace{1cm} (24)$$

where

$$\bar{y}_{[-|]} = \frac{\sum_{i=1}^{n} \delta_i \delta_i^x y_i}{w}, \hspace{0.5cm} \bar{x}_{[-|]} = \frac{\sum_{i=1}^{n} \delta_i \delta_i^x x_{ij}}{w}, \hspace{0.5cm} \bar{x}'_{[-|]} = [\bar{x}_{[-|]1} \bar{x}_{[-|]2} \cdots \bar{x}_{[-|]m}],$$

$$w = \sum_{i=1}^{n} \delta_i \delta_i^x, \hspace{0.5cm} \Delta = \sum_{i=1}^{n} \alpha_i \delta_i^x.$$

$$\delta = \text{diag}(\delta_i \delta_i^x) \hspace{0.5cm} \text{and} \hspace{0.5cm} \alpha = \text{diag}(\alpha_i \delta_i^x), \hspace{1cm} i = 1, 2, \ldots, n.$$  

The remaining expressions, i.e. \(\hat{\beta}, \hat{\sigma}, K, L, B \) and \(C\) are formulated in the same way as given in the previous section, but using the weights and formulations as introduced above.

It should be mentioned that there is no direct link of the weights \(\delta_i^x\) with the function \(g(\cdot)\) and its linearized version but there is an analogy with \(\delta_i\) and \(\delta_i^x\). Indeed, \(\delta_i^x\) is obtained by modifying the denominator of \(\delta_i\) for the sake of more robustness. Both \(\delta_i\) and \(\delta_i^x\) give small weights to the large observations as expected. The choice of \(\delta_i^x\) can be changed, i.e. some other weight functions or modifications can also be employed, see for example Acitas et al. [28]. It should also be noted that RAMML estimators are defined similar to generalized \(M\) (GM) estimators which are robust to both types of outliers, see Maronna et al. [6] for further details.
3.1. Computation of RAMML estimators

The RAMML estimators are obtained after two iterations. In the first iteration, an initial estimate of $t_i$ is required. Unlike Donmez [19], we do not consider a reparametrization of model (1). We continue with model (1) and suggest to use well-known robust regression estimators such as LTS or S to initialize the iterations. In other words, the initial estimate of $t_i$ is proposed as

$$t_i = \frac{y_i - T_0 - \sum_{j=1}^{n} x_{ij} T_j}{S_0},$$

where $T_0, T_j (j = 1, 2, \ldots, m)$ and $S_0$ are obtained from LTS or S regression. After getting the initial estimate of $t_i$, initial RAMML estimators denoted by $\tilde{\beta}_0, \tilde{\beta}_j$ and $\tilde{\sigma}$ are obtained following Steps 1, 2* and 3*.

In the second iteration, $t_i$ is updated with the initial RAMML estimators obtained in the previous iteration as follows:

$$t_i = \frac{y_i - \tilde{\beta}_0 - \sum_{j=1}^{n} x_{ij} \tilde{\beta}_j}{\tilde{\sigma}}$$

and Steps 1, 2* and 3* are utilized. The final RAMML estimators are obtained at the end of Step 3* using Equations (13), (14) and (24).

The resulting estimators obtained using LTS and S regression are denoted by RAMML$_1$ and RAMML$_2$, respectively. It is clear that RAMML estimators are also explicitly formulated and they can be computed after two simple iterations. Therefore, they are computationally straightforward.

If the weights $\delta_i$ are removed from the estimating equations, RAMML$_1$ and RAMML$_2$ will reduce to the AMML$_1$ and AMML$_2$, respectively. It should be noted that these AMML$_1$ and AMML$_2$ estimators differ from that of Donmez [19] since the initial $t_i$s are defined differently.

It should be noted that different initial estimates can be preferred for finding initial $t_i$s. The important issue here is that the initial estimators should be robust so that the resulting estimator is robust. In this study, we consider two well-known robust estimators as initials, i.e. LTS and S. We also consider the LMS estimator as an initial estimator and obtain more or less the same results, see Acitas et al. [20]. However, we do not consider it here because it is not computationally efficient.

4. Simulation study

In this section, we conduct a simulation study to compare the performances of the proposed estimators RAMML$_1$ and RAMML$_2$ along with some known robust estimators such as MM, LTS and S estimators under different simulation scenarios. AMML$_1$ and AMML$_2$ are also taken into account during the performance comparisons. We also include exponential-type kernel function based robust regression (ETKRR) estimators proposed by Carvalho et al. [29]. These estimators are denoted by ETKRR$_{(S1)}$, ETKRR$_{(S2)}$.
and ETKRR\(_{(S3)}\). ETKRR\(_{(S2)}\) and ETKRR\(_{(S3)}\) can be preferred for data sets including \(x\) and \(y\)-outliers, respectively. See Carvalho et al. [29] for further details.

The simulation setup and contamination schemes are considered similar to those used by Gschwandtner and Filzmoser [7]. Model (1) is used during the simulation study where \(\beta_0\) is assumed to be 0 without loss of generality. The data are generated as follows:

\[
x_i \sim N_m(0, I_m), \quad i = 1, 2, \ldots, n
\]

\[
\beta_j = 1/\sqrt{m}, \quad j = 1, 2, \ldots, m \quad \text{and thus } \|\beta\| = 1.
\]

Here, \(N_m\) denotes the \(m\)-dimensional multivariate normal distribution, \(0\) stands for the vector of zeros and \(I_m\) is the identity matrix.

Five different error distributions are considered in model (1): (i) \(N(0, 1)\), (ii) Laplace, (iii) \(t_5\), (iv) \(t_1\) (Cauchy) and (v) slash distribution. The leverage points are generated replacing the first \(n_{\text{out}}\) observations with the following versions:

\[
x_i = \begin{bmatrix} \ell & \ell & \cdots & \ell \end{bmatrix}' \quad \text{and} \quad y_i = x_i^\prime a \quad \text{for} \quad i = 1, 2, \ldots, n
\]

where \(a\) is a normalized unit vector defined by \(a = v - (v'\beta)\beta\). Here, \(v\) is a vector having entries \((-1)^j\). If \(m = 1\), \(a\) is taken as -1. We consider two different versions for leverage points, i.e. \(\ell = 5\) and \(\ell = 10\). This is because of the fact that we would like to see how the performances of the estimators are affected as the magnitude of the leverage point increases. The sample size is taken to be \(n = 50\) and 200. The number of the predictors \(m\) taken from the set \(\{1, 5, 10, 20\}\). \(n_{\text{out}}\) is the number of leverage points and determined according to two contamination levels, i.e. 10% and 20%.

The mean squared error (MSE) formulated as

\[
\text{MSE} = \frac{1}{n_{\text{rep}}} \sum_{r=1}^{n_{\text{rep}}} \left\| \hat{\beta}^r - \beta \right\|^2
\]

is calculated to evaluate the performances of the estimators based on \(n_{\text{rep}} = 500\) Monte-Carlo runs. All computations are carried out using the software environment R. The results are tabulated in Tables 1 and 2. It should be mentioned that LS estimators are not considered in these tables since they are not robust to outliers.

It can be seen from Table 1 that the MSEs of the RAMML\(_1\) and RAMML\(_2\) are close to each other in the majority of the cases. The results of Table 1 are interpreted according to the contamination levels as follows.

**10% level of contamination**

- The RAMML\(_2\) has the minimum MSE value for \(m \in \{1, 5, 10, 20\}\) and \(\ell = 5\) when the error distribution is normal, Laplace, \(t_5\), \(t_1\) and slash.
- For the simple linear regression (\(m = 1\)), the MM estimator is the best for normal distribution and \(\ell = 10\). The RAMML\(_2\) outperforms its rivals for the other error distributions,
- The RAMML\(_2\) is superior in most of the cases when \(\ell = 10\). However, the RAMML\(_1\) is the best for \(m = 5\) for normal and \(t_5\) distributions, and also for Laplace and \(t_5\) distributions when \(m = 10\).
• The LTS, AMML1, S, ETKRR(S2) and ETKRR(S3) estimators do not provide a satisfactory performance.

20% level of contamination

• The RAMML2 performs substantially better than the other estimators for $m \in \{1, 5, 10\}$, $\ell = 5$ and all error distributions.

Table 1. The simulated MSEs of MM, LTS, AMML1, RAMML1, S, AMML2, RAMML2, ETKRR(S2) and ETKRR(S3) estimators for regression coefficients: $n = 50$.

| Estimator     | Normal | Laplace | $t_5$ | $t_1$ | Slash | Normal | Laplace | $t_5$ | $t_1$ | Slash |
|---------------|--------|---------|-------|-------|-------|--------|---------|-------|-------|-------|
|               | $\ell = 5$ |         |       |       |       | $\ell = 10$ |         |       |       |       |
| MM            | 0.0313 | 0.1328  | 0.0533 | 1.0575 | 2.0005 | 0.0262 | 0.0755  | 0.0511 | 0.3476 | 1.5410 |
| LTS           | 0.0828 | 0.1908  | 0.1618 | 0.3246 | 1.2368 | 0.1062 | 0.1138  | 0.1623 | 0.3644 | 1.5432 |
| AMML1         | 0.7650 | 1.2987  | 1.1882 | 2.1031 | 2.2239 | 0.2647 | 0.7729  | 0.6324 | 2.6792 | 3.3200 |
| RAMML1        | 0.0291 | 0.0581  | 0.0425 | 0.1764 | 0.2752 | 0.0304 | 0.0442  | 0.0381 | 0.1331 | 0.3079 |
| S             | 0.0702 | 0.1283  | 0.0804 | 0.3791 | 1.5208 | 0.0587 | 0.0848  | 0.0773 | 0.3360 | 1.5945 |
| AMML2         | 0.6740 | 1.1063  | 1.0060 | 2.0382 | 2.2122 | 0.1885 | 0.5288  | 0.4271 | 2.4115 | 3.2409 |
| RAMML2        | 0.0290 | 0.0563  | 0.0411 | 0.1670 | 0.2668 | 0.0304 | 0.0439  | 0.0378 | 0.1253 | 0.2921 |
| ETKRR(S2)     | 0.2488 | 0.6357  | 0.4987 | 2.0228 | 2.4789 | 3.8727 | 3.8502  | 3.8748 | 3.6588 | 3.7066 |
| ETKRR(S3)     | 0.9286 | 1.7501  | 1.5125 | 2.4996 | 2.4881 | 3.8164 | 3.7874  | 3.8105 | 3.5152 | 3.5308 |
|               | $m = 1$ |         |       |       |       | $m = 5$ |         |       |       |       |
| MM            | 0.6487 | 0.8233  | 0.8181 | 1.3874 | 1.8251 | 0.6918 | 0.8392  | 0.8366 | 1.4375 | 1.8164 |
| LTS           | 1.1673 | 1.2959  | 1.3244 | 1.7131 | 2.3378 | 1.1952 | 1.3313  | 1.3133 | 1.6490 | 2.5022 |
| AMML1         | 0.9560 | 1.1126  | 1.0560 | 1.6168 | 2.1107 | 0.8754 | 1.0540  | 1.0337 | 1.7361 | 2.1027 |
| RAMML1        | 0.3513 | 0.4802  | 0.4400 | 1.0311 | 1.5751 | 0.1426 | 0.2208  | 0.1930 | 0.7843 | 1.2528 |
| S             | 0.9489 | 1.1128  | 1.2521 | 1.8018 | 2.6589 | 0.9952 | 1.2243  | 1.1610 | 1.8808 | 2.7233 |
| AMML2         | 0.8573 | 1.0602  | 0.9975 | 1.5805 | 2.0582 | 0.6936 | 0.9383  | 0.8749 | 1.6962 | 2.0557 |
| RAMML2        | 0.3001 | 0.4368  | 0.4068 | 0.9970 | 1.5274 | 0.1438 | 0.2205  | 0.1938 | 0.7467 | 1.1970 |
| ETKRR(S2)     | 1.6485 | 1.7350  | 1.7395 | 1.9059 | 2.5091 | 1.6884 | 1.7899  | 1.7483 | 1.9243 | 2.5051 |
| ETKRR(S3)     | 1.4977 | 1.5059  | 1.5412 | 3.4043 | 5.3924 | 1.5748 | 1.5649  | 1.5592 | 2.1126 | 3.1411 |
|               | $m = 10$ |        |       |       |       | $m = 20$ |        |       |       |       |
| MM            | 1.2271 | 1.6813  | 1.5306 | 2.6522 | 3.8120 | 1.2813 | 1.6019  | 1.5725 | 2.7797 | 3.6446 |
| LTS           | 2.4659 | 2.9114  | 2.8500 | 4.1742 | 6.5135 | 2.4292 | 3.1187  | 3.0412 | 4.6464 | 6.1273 |
| AMML1         | 1.3497 | 1.6391  | 1.5305 | 2.9643 | 4.3020 | 1.3164 | 1.5878  | 1.5558 | 3.1651 | 4.3269 |
| RAMML1        | 0.9283 | 1.2108  | 1.1128 | 2.5337 | 3.8298 | 0.3161 | 0.5279  | 0.4740 | 1.9944 | 3.3376 |
| S             | 1.9879 | 2.6548  | 2.4624 | 4.4660 | 6.6034 | 2.0252 | 2.7817  | 2.5944 | 4.5463 | 6.4925 |
| AMML2         | 1.1236 | 1.5657  | 1.4103 | 2.8953 | 4.1502 | 1.0110 | 1.4136  | 1.3418 | 3.0401 | 4.1766 |
| RAMML2        | 0.7595 | 1.1247  | 0.9908 | 2.4773 | 3.6900 | 0.3183 | 0.5369  | 0.4744 | 1.8566 | 3.1814 |
| ETKRR(S2)     | 2.4326 | 2.8025  | 2.6886 | 3.5837 | 11.9171 | 2.4621 | 2.8164 | 2.8251 | 3.8155 | 5.0635 |
| ETKRR(S3)     | 2.2901 | 2.5879  | 2.4866 | 6.4574 | 45.2858 | 2.3364 | 2.5950 | 2.5334 | 9.2147 | 9.0672 |
|               | $\text{Continued}$ |        |       |       |       | $\text{Continued}$ |        |       |       |       |

(continued)
If $m = 20$ and $\ell = 5$, the AMML2 is more preferable for normal and Laplace distributions. However, ETKRR$_{(S2)}$ is more preferable for normal and slash error distributions.

The RAMML2 has a satisfactory performance for $m \in \{1, 5, 10\}$, $\ell = 10$ and normal, Laplace, $t_5$, $t_1$ and slash error distributions. It should be mentioned that the RAMML1 is almost the best and closely followed by the RAMML2.

ETKRR$_{(S3)}$ estimator is the best when $m = 20$ and $\ell = 10$ for heavy-tailed distributions. The RAMML2 is more preferable for the remaining error distributions.
When the sample size is increased to 200, see Table 2, we can draw similar conclusions to those obtained from Table 1. The results of Table 2 are again interpreted according to the contamination levels as follows.

**10% level of contamination**

- Here, the MM estimator gains efficiency and performs better than the other estimators for some cases. For example, for simple linear regression with normal, Laplace and

| Table 2. The simulated MSEs of MM, LTS, AMML₁, RAMML₁, S, AMML₂, RAMML₂, ETKRR(S₂) and ETKRR(S₃) estimators for regression coefficients: \( n = 200 \). |
|---------------------------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| Estimator | Normal | Laplace | \( t_5 \) | \( t_1 \) | Slash | Normal | Laplace | \( t_5 \) | \( t_1 \) | Slash |
| \( \ell = 5 \) | \( \ell = 10 \) | \( \ell = 20 \) |
| MM | 0.0058 | 0.0081 | 0.0075 | 0.4511 | 2.3904 | 0.0059 | 0.0085 | 0.0079 | 0.0197 | 1.0725 |
| LTS | 0.0066 | 0.0082 | 0.0087 | 0.0199 | 0.8371 | 0.0074 | 0.0088 | 0.0099 | 0.0159 | 1.1393 |
| AMML₁ | 0.6921 | 1.2477 | 1.1388 | 2.0903 | 2.2375 | 0.992 | 0.5857 | 0.4051 | 2.9032 | 3.3729 |
| RAMML₁ | 0.0069 | 0.0114 | 0.0104 | 0.0520 | 0.1072 | 0.0068 | 0.0101 | 0.0099 | 0.0304 | 0.0580 |
| S | 0.0151 | 0.0094 | 0.0135 | 0.0460 | 2.0932 | 0.0157 | 0.0090 | 0.0139 | 0.0128 | 0.1365 |
| AMML₂ | 0.5742 | 0.9894 | 0.9266 | 2.0126 | 2.2340 | 0.0740 | 0.3089 | 0.2582 | 2.5357 | 3.3124 |
| RAMML₂ | 0.0069 | 0.0108 | 0.0101 | 0.0454 | 0.1022 | 0.0069 | 0.0100 | 0.0099 | 0.0287 | 0.0562 |
| ETKRR(S₂) | 0.0100 | 0.0648 | 0.0666 | 2.0589 | 2.4916 | 3.8724 | 3.8714 | 3.8714 | 3.6769 | 3.6664 |
| ETKRR(S₃) | 0.2591 | 1.4868 | 1.0113 | 2.2495 | 2.2342 | 3.8051 | 3.7969 | 3.8082 | 3.4014 | 3.4352 |
| MM | 0.2513 | 0.3201 | 0.4169 | 0.8913 | 1.0756 | 0.2926 | 0.3880 | 0.4650 | 0.9805 | 1.1520 |
| LTS | 0.6970 | 0.5505 | 0.7987 | 0.8580 | 1.1411 | 0.7532 | 0.8656 | 0.9077 | 1.1089 | 1.2089 |
| AMML₁ | 0.8454 | 0.9126 | 0.9015 | 1.0277 | 1.1226 | 0.0992 | 0.5587 | 0.4074 | 1.1089 | 1.2089 |
| S | 0.3076 | 0.3525 | 0.4863 | 0.9477 | 0.6259 | 0.3118 | 0.4064 | 0.5078 | 1.0308 | 1.1359 |
| AMML₂ | 0.7188 | 0.9114 | 0.8367 | 1.0196 | 1.2292 | 1.2275 | 1.2478 | 1.2641 | 1.1473 | 1.2961 |
| RAMML₂ | 0.0991 | 0.1525 | 0.1694 | 0.4771 | 0.6163 | 0.0991 | 0.1491 | 0.1407 | 0.1337 | 0.2348 |
| ETKRR(S₂) | 1.2288 | 1.2351 | 1.2292 | 1.1359 | 1.1928 | 1.1113 | 1.1073 | 1.1170 | 1.4246 | 1.7707 |
| ETKRR(S₃) | 1.0975 | 1.0741 | 1.0762 | 1.4961 | 1.4616 | 1.2739 | 1.2793 | 1.2897 | 1.4246 | 1.7707 |
| MM | 0.5630 | 0.9695 | 0.9293 | 1.6385 | 1.9927 | 0.5260 | 0.9683 | 0.9375 | 1.6197 | 2.0479 |
| LTS | 1.3574 | 1.4794 | 1.5557 | 1.8926 | 2.5032 | 1.3713 | 1.4510 | 1.5391 | 1.9300 | 2.5240 |
| AMML₁ | 1.0659 | 1.2160 | 1.2228 | 1.7784 | 2.2702 | 1.0577 | 1.1402 | 1.2039 | 1.7778 | 2.2854 |
| RAMML₁ | 0.8490 | 0.9533 | 0.9825 | 1.5356 | 2.0338 | 0.1309 | 0.2193 | 0.1873 | 0.6853 | 1.2031 |
| S | 1.2260 | 1.8168 | 1.7462 | 2.4758 | 3.3239 | 1.1123 | 1.6759 | 1.6985 | 2.4514 | 3.3286 |
| AMML₂ | 0.7747 | 1.1502 | 1.0984 | 1.7352 | 2.1971 | 0.6258 | 1.0308 | 1.0092 | 1.7460 | 2.2158 |
| RAMML₂ | 0.5841 | 0.9048 | 0.8580 | 1.5229 | 1.9707 | 0.1328 | 0.2190 | 0.1884 | 0.6389 | 1.1336 |
| ETKRR(S₂) | 1.9465 | 2.1111 | 2.0800 | 2.0826 | 2.5287 | 1.9915 | 2.1066 | 2.0991 | 2.2929 | 2.9169 |
| ETKRR(S₃) | 1.7059 | 1.7186 | 1.7413 | 4.0025 | 5.0611 | 1.7438 | 1.7149 | 1.7450 | 4.2339 | 13.2511 |

(continued)
Table 2. Continued.

| Estimator | Normal Laplace | \( t_5 \) | \( t_1 \) | Slash | Normal Laplace | \( t_5 \) | \( t_1 \) | Slash |
|-----------|----------------|-------|-------|------|----------------|-------|-------|------|
|           | \( \ell = 5 \) |       |       |      | \( \ell = 10 \) |       |       |      |
| MM        | 0.4902         | 0.9801 | 1.0486 | 3.0287 | 3.1436         | 0.4986 | 1.0506 | 1.0557 | 3.5622 | 3.7584 |
| LTS       | 2.1264         | 2.9629 | 2.5365 | 3.1159 | 3.3428         | 2.2550 | 2.1266 | 2.6937 | 3.5192 | 3.8134 |
| AMML1     | 2.7408         | 2.8410 | 2.8538 | 2.9591 | 2.9803         | 2.7746 | 2.3717 | 3.3047 | 3.6922 | 3.7018 |
| RAMML1    | 0.1600         | 0.2571 | 0.2638 | 0.6833 | 0.8259         | 0.0082 | 0.0126 | 0.0108 | 0.0389 | 0.0688 |
| S         | 0.5702         | 1.1349 | 1.1829 | 3.4744 | 3.6280         | 0.5209 | 1.1193 | 1.0857 | 3.7057 | 3.9000 |
| AMML2     | 2.4515         | 2.7209 | 2.6982 | 2.9720 | 2.9941         | 1.5742 | 2.6495 | 2.5251 | 3.6936 | 3.7067 |
| RAMML2    | 0.0591         | 0.1709 | 0.1504 | 0.6733 | 0.8259         | 0.0080 | 0.0123 | 0.0105 | 0.0374 | 0.0673 |
| ETKRR\((S)\) | 3.6204       | 3.6157 | 3.5900 | 3.3877 | 3.3878         | 3.8968 | 3.8929 | 3.8803 | 3.8284 | 3.8231 |
| ETKRR\((S)\) | 3.5756       | 3.5609 | 3.5428 | 3.0287 | 3.0287         | 3.8868 | 3.8846 | 3.8731 | 3.6945 | 3.6916 |

|           | \( m = 1 \)  |       |       |      | \( m = 5 \)  |       |       |      | \( m = 10 \) |       |       |      | \( m = 20 \) |       |       |      |
| MM        | 1.0202         | 1.0357 | 1.0297 | 1.0765 | 1.1582         | 1.0608 | 1.0691 | 1.0729 | 1.1105 | 1.1824 |
| LTS       | 1.2873         | 1.3031 | 1.3136 | 1.2726 | 1.4884         | 1.3053 | 1.3281 | 1.3151 | 1.3219 | 1.5089 |
| AMML1     | 0.9911         | 1.0073 | 1.0036 | 1.0981 | 1.1744         | 1.0410 | 1.0533 | 1.0512 | 1.1263 | 1.2194 |
| RAMML1    | 0.7644         | 0.7879 | 0.7807 | 0.8980 | 0.9796         | 0.0421 | 0.0661 | 0.0591 | 0.1892 | 0.3366 |
| S         | 1.3074         | 1.3279 | 1.3527 | 1.3232 | 1.5698         | 1.3093 | 1.3381 | 1.3393 | 1.3402 | 1.5969 |
| AMML2     | 0.9922         | 1.0084 | 1.0047 | 1.0913 | 1.1654         | 1.0419 | 1.0536 | 1.0519 | 1.1200 | 1.2076 |
| RAMML2    | 0.7705         | 0.7936 | 0.7867 | 0.8957 | 0.9756         | 0.0439 | 0.0677 | 0.0611 | 0.1814 | 0.3242 |
| ETKRR\((S)\) | 1.2759       | 1.2885 | 1.3144 | 1.1803 | 1.3184         | 1.2657 | 1.3145 | 1.3110 | 1.8113 | 1.3627 |
| ETKRR\((S)\) | 1.1350       | 1.1190 | 1.1316 | 1.3762 | 1.6970         | 1.1461 | 1.1310 | 1.1520 | 2.2100 | 1.7224 |

|           | \( m = 10 \) |       |       |      | \( m = 10 \) |       |       |      | \( m = 20 \) |       |       |      | \( m = 20 \) |       |       |      |
| MM        | 1.1569         | 1.1906 | 1.1764 | 1.2884 | 1.4731         | 1.1713 | 1.2100 | 1.2037 | 1.3168 | 1.4956 |
| LTS       | 1.5579         | 1.6596 | 1.6157 | 1.7952 | 2.2259         | 1.5552 | 1.6277 | 1.6226 | 1.8255 | 2.2981 |
| AMML1     | 1.0935         | 1.1414 | 1.1145 | 1.3210 | 1.5377         | 1.1187 | 1.1596 | 1.1466 | 1.3605 | 1.5639 |
| RAMML1    | 0.9694         | 1.0189 | 0.9921 | 1.2019 | 1.4122         | 0.0918 | 0.1621 | 0.1439 | 0.4684 | 0.7511 |
| S         | 1.6731         | 1.7579 | 1.7419 | 1.8644 | 2.4330         | 1.6895 | 1.7740 | 1.7650 | 1.9145 | 2.4825 |
| AMML2     | 1.0965         | 1.1426 | 1.1165 | 1.3020 | 1.5112         | 1.1208 | 1.1611 | 1.1490 | 1.3420 | 1.5365 |
| RAMML2    | 0.9764         | 1.0237 | 0.9979 | 1.1875 | 1.3912         | 0.0985 | 0.1698 | 0.1531 | 0.4504 | 0.7274 |
| ETKRR\((S)\) | 1.5563       | 1.6466 | 1.6141 | 1.4450 | 1.7810         | 1.5734 | 1.6446 | 1.6357 | 1.5406 | 1.9102 |
| ETKRR\((S)\) | 1.3764       | 1.3605 | 1.3805 | 2.3463 | 2.5760         | 1.3748 | 1.3718 | 1.3922 | 2.4663 | 2.9316 |

\( t_5 \) error distributions and \( \ell = 5 \) and 10, it has the minimum MSE value, followed by the RAMML1. The LTS, S and RAMML2 are more preferable for heavy tailed error distributions.

- If \( m \) is increased from 5 to 20, the RAMML2 is mostly the best estimator for all error distributions. The RAMML1 also outperforms the other estimators in some cases, see for example \( \ell = 5 \), \( m = 10 \), \( t_1 \) distribution and \( \ell = 10 \), \( m \in \{10,20\} \), normal and \( t_5 \) distributions.
The MM estimator has the minimum MSE value for $\ell = 5$, $m = 20$ and normal distribution.

20% level of contamination

The RAMML1 and/or RAMML2 outperform the others in the majority of the cases regardless of the value of $\ell$. Indeed, the RAMML2 is the best for heavy tailed distributions while the RAMML1 is more preferable for normal, Laplace and $t_5$ distributions for $m \in \{5, 10, 20\}$. The RAMML2 has a remarkable performance for $m = 1$.

The MM estimator has also a promising performance, but not enough to be the best.

Table 3. The simulated MSEs of MM, LTS, AMML1, RAMML1, S, AMML2, RAMML2 estimators for scale parameter when the error distribution is $N(0, 1)$.

|               | $\ell = 5$ | $\ell = 10$ |
|---------------|------------|-------------|
|               | $m = 1$    | $m = 5$     | $m = 10$    | $m = 25$ |
| $n = 50$      |               |             |
| Contamination level = 10% |   |             |
| MM            | 0.0460     | 0.0233      | 0.0308      | 0.0860 |
| LTS           | 0.0778     | 0.0619      | 0.0576      | 0.1296 |
| AMML1         | 0.3110     | 0.0777      | 0.0819      | 0.0689 |
| RAMML1        | 0.0382     | 0.0194      | 0.0301      | 0.0451 |
| S             | 0.0460     | 0.0252      | 0.0329      | 0.0713 |
| AMML2         | 0.2623     | 0.0660      | 0.0626      | 0.0673 |
| RAMML2        | 0.0390     | 0.0251      | 0.0283      | 0.0378 |
|               |             |             |             |         |
| Contamination level = 20% |   |             |
| MM            | 0.1366     | 0.0618      | 0.1897      | 0.3920 |
| LTS           | 0.3193     | 0.0786      | 0.2786      | 0.7529 |
| AMML1         | 0.6533     | 0.0418      | 0.0421      | 0.6658 |
| RAMML1        | 0.0411     | 0.0183      | 0.0503      | 0.6979 |
| S             | 0.1366     | 0.0570      | 0.1498      | 0.2782 |
| AMML2         | 0.6240     | 0.0433      | 0.0303      | 0.0890 |
| RAMML2        | 0.0367     | 0.0188      | 0.0231      | 0.1093 |
|               |             |             |             |         |
| $n = 200$     |               |             |
| Contamination level = 10% |   |             |
| MM            | 0.0282     | 0.0211      | 0.0181      | 0.0129 |
| LTS           | 0.0571     | 0.0751      | 0.0717      | 0.0565 |
| AMML1         | 0.3020     | 0.0771      | 0.0704      | 0.0689 |
| RAMML1        | 0.0294     | 0.0093      | 0.0175      | 0.0256 |
| S             | 0.0282     | 0.0219      | 0.0186      | 0.0113 |
| AMML2         | 0.2522     | 0.0686      | 0.0500      | 0.0389 |
| RAMML2        | 0.0303     | 0.0160      | 0.0196      | 0.0227 |
|               |             |             |             |         |
| Contamination level = 20% |   |             |
| MM            | 0.1393     | 0.0115      | 0.0236      | 0.0768 |
| LTS           | 0.4019     | 0.0182      | 0.0110      | 0.0295 |
| AMML1         | 0.6431     | 0.0375      | 0.0381      | 0.0310 |
| RAMML1        | 0.0133     | 0.0071      | 0.0106      | 0.0093 |
| S             | 0.1393     | 0.0115      | 0.0230      | 0.0753 |
| AMML2         | 0.3177     | 0.0339      | 0.0331      | 0.0238 |
| RAMML2        | 0.0164     | 0.0061      | 0.0086      | 0.0067 |

n = 50

Contamination level = 20%

|               | $\ell = 5$ | $\ell = 10$ |
|---------------|------------|-------------|
|               | $m = 1$    | $m = 5$     | $m = 10$    | $m = 25$ |
| Contamination level = 10% |   |             |
| MM            | 0.0282     | 0.0211      | 0.0181      | 0.0129 |
| LTS           | 0.0571     | 0.0751      | 0.0717      | 0.0565 |
| AMML1         | 0.3020     | 0.0771      | 0.0704      | 0.0689 |
| RAMML1        | 0.0294     | 0.0093      | 0.0175      | 0.0256 |
| S             | 0.0282     | 0.0219      | 0.0186      | 0.0113 |
| AMML2         | 0.2522     | 0.0686      | 0.0500      | 0.0389 |
| RAMML2        | 0.0303     | 0.0160      | 0.0196      | 0.0227 |
|               |             |             |             |         |
| Contamination level = 20% |   |             |
| MM            | 0.1393     | 0.0115      | 0.0236      | 0.0768 |
| LTS           | 0.4019     | 0.0182      | 0.0110      | 0.0295 |
| AMML1         | 0.6431     | 0.0375      | 0.0381      | 0.0310 |
| RAMML1        | 0.0133     | 0.0071      | 0.0106      | 0.0093 |
| S             | 0.1393     | 0.0115      | 0.0230      | 0.0753 |
| AMML2         | 0.3177     | 0.0339      | 0.0331      | 0.0238 |
| RAMML2        | 0.0164     | 0.0061      | 0.0086      | 0.0067 |
Table 4. The simulated MSEs of MM, LTS, AMML1, RAMML1, S, AMML2, RAMML2, ETKRR\textsubscript{(S1)} estimators for regression coefficients when the distribution of error terms is $N(0,1)$ with zero contamination.

| Estimator | $n = 50$ |        |        |        |        |        |        |        |        |        |        |
|-----------|---------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
|           | $m = 1$ | $m = 5$ | $m = 10$ | $m = 20$ | $m = 1$ | $m = 5$ | $m = 10$ | $m = 20$ |
| LS        | 0.0215  | 0.1176 | 0.2578 | 0.7177  | 0.0050  | 0.0260  | 0.0529  | 0.1124  |
| MM        | 0.0226  | 0.1281 | 0.2840 | 0.9829  | 0.0052  | 0.0282  | 0.0573  | 0.1303  |
| LTS       | 0.0340  | 0.2329 | 0.5952 | 1.7939  | 0.0071  | 0.0397  | 0.0801  | 0.1776  |
| AMML\textsubscript{1} | 0.0217  | 0.1190 | 0.2597 | 0.7347  | 0.0050  | 0.0266  | 0.0538  | 0.1147  |
| RAMML\textsubscript{1} | 0.0272  | 0.1202 | 0.2617 | 0.7401  | 0.0060  | 0.0271  | 0.0543  | 0.1152  |
| S         | 0.0725  | 0.4062 | 0.8739 | 2.1882  | 0.0179  | 0.0967  | 0.2091  | 0.4627  |
| AMML\textsubscript{2} | 0.0218  | 0.1194 | 0.2607 | 0.7472  | 0.0050  | 0.0267  | 0.0540  | 0.1153  |
| RAMML\textsubscript{2} | 0.0272  | 0.1208 | 0.2632 | 0.7560  | 0.0060  | 0.0272  | 0.0545  | 0.1161  |
| ETKRR\textsubscript{(S1)} | 0.0284  | 0.1478 | 0.3283 | 0.9600  | 0.0061  | 0.0317  | 0.0637  | 0.1371  |

In this part of the simulation study, we just focus on the MSEs of the estimators for the scale parameter $\sigma$ when the distribution of the error terms is $N(0,1)$. The MSE of $\hat{\sigma}$ is calculated using (27) in which the symbols $\beta$ are replaced by $\sigma$. The results are tabulated in Table 3. The MSEs of RAMML\textsubscript{1} and RAMML\textsubscript{2} are more or less the same. RAMML\textsubscript{1} and/or RAMML\textsubscript{2} has the minimum MSE value in most of the cases.

It is well-known that the trade-off between robustness and efficiency is important. We therefore explore the efficiencies of RAMML\textsubscript{1} and RAMML\textsubscript{2} under normality in this part of the simulation study. For this purpose, the performances of all estimators considered in this study are compared with the LS estimators when the error distribution is $N(0,1)$ with zero contamination. The results are given in Table 4. It is clear from this table that the MSE of LS is closely followed by RAMML\textsubscript{1} and RAMML\textsubscript{2}. This conclusion implies that RAMML\textsubscript{1} and RAMML\textsubscript{2} are not only robust to both types of outliers but also efficient under normality.

Finally, we explore the performances of the estimators when the predictors are correlated. Therefore, we make a simple modification in the simulation scenario given at the beginning of this section. Here, the $x_i$s are generated from $N_m(0, V)$ where $V = [v_{jk}]$ with $v_{jj} = 1$ and $v_{jk} = \rho$ for $j \neq k$ and $k = 1, 2, \ldots, m$. In Table 5, we report the MSEs of the estimators for $\rho = 0.90$, $\ell = 10$ and contamination levels 10%. This table shows that the MM estimator is almost the best one for normal, Laplace and $t_5$ distributions. The AMML\textsubscript{1} has also a promising performance for normally distributed error terms. The LTS and S estimators are more preferrable than the others for $t_1$ and slash distributions. In most of the cases, these estimators are mostly followed by RAMML\textsubscript{1} and RAMML\textsubscript{2}. The ETKRR\textsubscript{(S2)} and ETKRR\textsubscript{(S3)} do not provide a satisfactory performance.

Overall, the MSEs of RAMML\textsubscript{1} and RAMML\textsubscript{2} are close to each other and they outperform their competitors when there are $x-$outliers. Among these two estimators, we see that RAMML\textsubscript{2} is preferable.

5. Application

In this part of the study, we provide two data sets which are widely used in the literature to show the implementation of the proposed estimators. The full data sets can be found in the R package ‘robustbase’. 
Table 5. The simulated MSEs of MM, LTS, AMML\(_1\), RAMML\(_1\), S, AMML\(_2\), RAMML\(_2\), ETKRR\(_{2}(S)\), and ETKRR\(_{3}(S)\) estimators for regression coefficients with correlated predictors: \(n = 200\), \(\ell = 10\) and contamination level = 10%.

| Estimator | Normal | Laplace | \(t_5\) | \(t_1\) | Slash |
|-----------|--------|---------|--------|--------|--------|
| \(m = 5\) |        |         |        |        |        |
| MM        | 0.2392 | 0.3602  | 0.3321 | 0.8311 | 2.1349 |
| LTS       | 0.2792 | 0.3631  | 0.3708 | 0.6970 | 2.9565 |
| AMML\(_1\) | 0.2597 | 0.5551  | 0.4648 | 2.8906 | 4.3981 |
| RAMML\(_1\) | 0.2438 | 0.3913  | 0.3459 | 1.2014 | 2.2416 |
| S         | 0.6424 | 0.4025  | 0.6308 | 0.5794 | 2.6981 |
| AMML\(_2\) | 0.2544 | 0.4702  | 0.4172 | 2.5308 | 4.1810 |
| RAMML\(_2\) | 0.2444 | 0.3851  | 0.3445 | 1.1139 | 2.1071 |
| ETKRR\(_{2}(S)\) | 7.2821 | 7.8668  | 7.7776 | 6.0316 | 6.9662 |
| ETKRR\(_{3}(S)\) | 4.5945 | 4.5348  | 4.6522 | 6.7940 | 9.2656 |
| \(m = 10\) |        |         |        |        |        |
| MM        | 0.5674 | 0.8377  | 0.7832 | 2.0226 | 3.9043 |
| LTS       | 0.6580 | 0.8844  | 0.8837 | 1.7884 | 4.0895 |
| AMML\(_1\) | 0.5549 | 0.9385  | 0.8266 | 5.0717 | 10.0665 |
| RAMML\(_1\) | 0.5871 | 0.9014  | 0.8217 | 2.9135 | 5.2663 |
| S         | 1.5876 | 1.0566  | 1.4456 | 1.4868 | 5.0806 |
| AMML\(_2\) | 0.5564 | 0.9051  | 0.8126 | 4.0906 | 9.1097 |
| RAMML\(_2\) | 0.5896 | 0.8863  | 0.8166 | 2.7153 | 4.9672 |
| ETKRR\(_{2}(S)\) | 25.8928 | 27.0434 | 26.9280 | 27.9979 | 22.5105 |
| ETKRR\(_{3}(S)\) | 16.3730 | 17.4470 | 17.3349 | 51.4907 | 73.7406 |
| \(m = 20\) |        |         |        |        |        |
| MM        | 1.2880 | 1.9905  | 1.7159 | 4.9469 | 9.9462 |
| LTS       | 1.5128 | 2.1419  | 1.9733 | 4.5512 | 9.5219 |
| AMML\(_1\) | 1.2456 | 2.0598  | 1.7371 | 8.9865 | 22.0419 |
| RAMML\(_1\) | 1.2937 | 2.0654  | 1.7741 | 6.9106 | 12.3960 |
| S         | 4.2434 | 3.4726  | 5.6245 | 24.8742 | 51.2499 |
| AMML\(_2\) | 1.3681 | 2.1094  | 2.1667 | 12.9807 | 25.1939 |
| RAMML\(_2\) | 1.2996 | 2.0376  | 1.7658 | 7.0920 | 12.6017 |
| ETKRR\(_{2}(S)\) | 88.9463 | 97.4898 | 95.0580 | 77.3990 | 82.0869 |
| ETKRR\(_{3}(S)\) | 69.4723 | 75.0280 | 72.5969 | 51.4907 | 73.7406 |

5.1. Hertzsprung-Russell diagram data of star cluster CYG OB1

This data set is shortly denoted as starsCYC and includes 47 observations with two variables named \(\log(\text{Te})\) and \(\log(\text{light})\). The former variable is the logarithm of the effective temperature at the surface of the star, while the latter one is the logarithm of its light intensity, see Rousseeuw and Leroy [1]. It is clear from Figure 1 that the starsCYC data contains outliers, i.e. four stars in the left upper corner are far away from bulk of the data, and thus they are leverage points. Rousseeuw and Leroy [1] employ the LMS method and obtain more reliable results than for the LS method. Different from them, in this study, we use the RAMML\(_1\) and RAMML\(_2\) estimation methods along with the other methods considered in the previous section to estimate the regression coefficients. To evaluate the performances of the methods, we use the standard error of prediction (SEP) criterion formulated as

\[
\text{SEP} = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (y_i - \hat{y}_i - \text{bias})^2}, \quad \text{bias} = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)
\]
where, $\hat{y}_i$ denotes the fitted value for the $i$-th observation. It should be noted that using SEP may give misleading results in case of outliers since each observation has the same contribution in the formula of SEP. Therefore, a robust version of the SEP is required. We therefore compute the trimmed version of SEP by excluding 10% of unusually large residuals in terms of their absolute values. The robust performance criterion will lead to a contrary picture since the influence of the largest contributions is reduced. There is no general rule for the trimming value and thus it is subjective. One can take lower or higher values according to the data quality, see for example Liebmann et al. [30], Filzmoser and Todorov [31].

The estimated values of the regression coefficients and the scale parameter and also SEP and SEP$_{trim}$ values are tabulated in Table 6. The SEP$_{trim}$ values suggest that RAMML$_1$ and RAMML$_2$ are preferable to MM, LTS, AMML$_1$, S, AMML$_2$, ETKRR$_{(S2)}$ and ETKRR$_{(S3)}$. Furthermore, the scale estimates obtained based on RAMML$_1$ and RAMML$_2$ are much smaller than those of other estimation methods. This conclusion implies that estimates obtained by RAMML$_1$ and RAMML$_2$ give more reliable results. The scatter plot of the starsCYG data and the fitted regression lines are also illustrated in Figure 1. The fitted regression lines obtained from the RAMML$_1$ and RAMML$_2$ estimates are also satisfactory.

The weights obtained at the final step of the robust estimation procedures MM, RAMML$_1$ and RAMML$_2$ are plotted against the scaled residuals in Figure 2. The weights obtained from RAMML$_1$ and RAMML$_2$ are smaller than those of MM. This is because the weights given by the RAMML methodology allow more flexibility to the data compared to those from the MM, see also Figure 1.
Table 6. The estimated values of the regression coefficients and the scale parameter and also SEP and SEP$_{\text{trim}}$ values for the starsCYG data.

| Method  | $\beta_0$  | $\beta_1$  | $\sigma$ | SEP  | SEP$_{\text{trim}}$ |
|---------|------------|------------|----------|------|------------------|
| MM      | -4.9694    | 2.2532     | 0.4715   | 0.9556 | 0.3282           |
| LTS     | -8.5001    | 3.0462     | 0.4562   | 1.1507 | 0.3266           |
| AMML$_1$| 3.4697     | 0.3409     | 0.5239   | 0.6000 | 0.3994           |
| RAMML$_1$| -8.0822   | 2.9523     | 0.3249   | 1.1269 | 0.3252           |
| S       | -9.5708    | 3.2904     | 0.4715   | 1.2133 | 0.3325           |
| AMML$_2$| 3.3016     | 0.3790     | 0.5214   | 0.6041 | 0.3972           |
| RAMML$_2$| -8.0907   | 2.9553     | 0.3249   | 1.1277 | 0.3252           |
| ETKRR$_{(S2)}$| 8.3265 | -0.6978 | 0.5646 | 0.5646 | 0.4310 |
| ETKRR$_{(S3)}$| 8.2885 | -0.6997 | 0.5646 | 0.5646 | 0.4311 |

Figure 2. Weights versus scaled residuals for the estimators MM, RAMML$_1$ and RAMML$_2$, based on the starsCYG data.

5.2. Aircraft data

This data set consisting of 23 observations on single-engine aircrafts is obtained from the Office of Naval Research over the years 1947–1979, see Gray [32] and Rousseeuw and Leroy [1]. It contains four predictors and a response variable whose descriptions are given below:

\[
y = \text{cost (in units of$\$100,000)} , \quad x_1 = \text{aspect ratio} , \quad x_2 = \text{lift} - \text{to} - \text{drag ratio} , \\
x_3 = \text{weight of plane (in pounds)} \quad \text{and} \quad x_4 = \text{maximal thrust}.
\]

We use the methods MM, LTS, AMML$_1$, RAMML$_1$, S, AMML$_2$ RAMML$_2$, ETKRR$_{(S2)}$ and ETKRR$_{(S3)}$ to estimate the unknown parameters. Robust diagnostics obtained based on LTS show that there are two outliers in the data. The estimated values of the regression coefficients and the scale parameter along with performance criteria are shown in Table 7. The minimum SEP$_{\text{trim}}$ value is obtained for ETKRR$_{(S3)}$ which is followed by ETKRR$_{(S2)}$, LTS, MM, RAMML$_1$ and RAMML$_2$. However, the scale estimates obtained from RAMML$_1$
Table 7. The estimated regression coefficients, scale parameter and values of the performance criteria for the Aircraft data.

| Method       | $\beta_0$ | $\beta_1$ | $\beta_2$ | $\beta_3$ | $\beta_4$ | $\sigma$ | SEP | SEPtrim |
|--------------|-----------|-----------|-----------|-----------|-----------|---------|-----|---------|
| MM           | 6.1396    | -3.2306   | 1.6713    | 0.0019    | -0.0009   | 5.8932  | 10.5866 | 3.4636  |
| LTS          | 9.5007    | -3.0488   | 1.2100    | 0.0014    | -0.0006   | 5.6927  | 12.4080 | 3.4110  |
| AMML$_1$     | 2.2997    | -3.4300   | 2.0052    | 0.0025    | -0.0013   | 6.5780  | 8.9124  | 3.8884  |
| RAMML$_1$    | 7.0264    | -3.2411   | 1.6622    | 0.0018    | -0.0009   | 3.9639  | 11.1015 | 3.4651  |
| S            | 13.3733   | -4.0220   | 1.5413    | 0.0017    | -0.0010   | 5.8932  | 11.8666 | 3.5001  |
| AMML$_2$     | 1.8030    | -3.4535   | 2.0451    | 0.0026    | -0.0014   | 6.7367  | 8.7372  | 3.9778  |
| RAMML$_2$    | 6.9195    | -3.2381   | 1.6674    | 0.0018    | -0.0009   | 3.9821  | 11.0630 | 3.4699  |
| ETKRR$_{(2)}$| 7.5061    | -3.2304   | 1.5284    | 0.0018    | -0.0008   | 11.1169 | 3.4328  |
| ETKRR$_{(3)}$| 10.6596   | -3.3825   | 1.3365    | 0.0015    | -0.0007   | 12.0868 | 3.3540  |

Figure 3. Weights versus scaled residuals for the Aircraft data.

and RAMML$_2$ are smaller than those of LTS and MM. Therefore, RAMML$_1$ and RAMML$_2$ are also promising and can be considered as good alternative estimators.

In Figure 3, the weights obtained at the last step of the robust estimation procedures MM, RAMML$_1$ and RAMML$_2$ are plotted against the scaled residuals. Also in this example, the RAMML weights are smaller than those of the MM estimator; this is because the RAMML weights are more flexible.

6. Conclusions

In this study, we propose RAMML estimators which are robust to both $y$ and $x$ outliers. The RAMML estimators are obtained by downweighting $x$—outliers using an extra weight function in the AMML estimation method.

The RAMML estimators have some advantage over existing robust estimators: they are (i) explicitly formulated and thus easy to compute, (ii) efficient under the normality assumption and (iii) more preferable according to the MSE and SEP criteria. These properties make them good alternatives to existing robust estimators. The simulation studies
also show that the RAMML estimators are more preferable to the other considered robust estimators especially for high contamination level, large number of explanatory variables and heavy-tailed error distributions. Thus, their advantages get more pronounced in more difficult data situations. Another interesting feature of the RAMML estimators is that the estimated residual variance is smaller than for the alternatives in many situation. Together with the other properties, this is a useful feature which is based on a more flexible weighting scheme during the computation of the estimates.

Although the new robust estimators show good performance, there are still some points that should be considered. One of them is the choice of the weights $\delta_i^x$ assigned to the predictors. Different choices of $\delta_i^x$ will affect the efficiencies of the RAMML estimators. In this study, the weights $\delta_i^x$ are obtained similar to the weights $\delta_i$. In the future studies, alternative weight functions such as Huber, Fair, etc. will be considered.

**Acknowledgments**

This study is supported by ‘The Scientific and Technological Research Council of Turkey (TUBITAK)’ as part of the ‘2219 – International Postdoctoral Research Scholarship Programme’.

**Disclosure statement**

No potential conflict of interest was reported by the author(s).

**Funding**

This study is supported by ‘The Scientific and Technological Research Council of Turkey (TUBITAK)’ as part of the ‘2219 – International Postdoctoral Research Scholarship Programme’.

**ORCID**

Sukru Acitas  
http://orcid.org/0000-0002-4131-0086  
Peter Filzmoser  
http://orcid.org/0000-0002-8014-4682  
Birdal Senoglu  
http://orcid.org/0000-0003-3707-2393

**References**

[1] Rousseeuw PJ, Leroy AM. Robust regression and outlier detection. New York: Wiley; 1987.
[2] Huber PJ. Robust regression: asymptotics, conjectures, and monte carlo. Ann Stat. 1973;1:799–821.
[3] Rousseeuw P, Yohai V. Robust regression by means of S-estimators. Springer; 1984. p. 256–272. (Robust nonlinear time Ser Anal.)
[4] Rousseeuw PJ. Least median of squares regression. J Am Stat Assoc. 1984;79:871–880.
[5] Yohai VJ. High breakdown-point and high efficiency robust estimates for regression. Ann Stat. 1987;15:642–656.
[6] Maronna RA, Martin RD, Yohai VJ. Robust statistics: theory and methods. Chichester: Wiley; 2006.
[7] Gschwandtner M, Filzmoser P. Computing robust regression estimators: developments since Dutter 1977. Austrian J Stat. 2012;41:45–58.
[8] Yu C, Yao W. Robust linear regression: a review and comparison. Commun. Stat. Simul. Comput. 2017;46:6261–6282.
[9] Tiku ML. Estimating the mean and standard deviation from a censored normal sample. Biometrika. 1967;54:155–165.
[10] Tiku ML. Estimating the parameters of normal and logistic distributions from a censored normal sample. Aust J Stat. 1968;10:64–74.

[11] Tiku ML, Suresh RP. A new method of estimation for location and scale parameters. J Stat Plan Inference. 1992;30:281–292.

[12] Tiku ML, Islam MQ, Selcuk AS. Nonnormal regression. II. symmetric distributions. Commun Stat – Theor Meth. 2001;30:1021–1045.

[13] Islam MQ, Tiku ML. Multiple linear regression model under nonnormality. Commun Stat – Theor Meth. 2005;33:2443–2467.

[14] Lucas A. Robustness of the student t based M-estimator. Commun Stat – Theor Meth. 1997;26:1165–1182.

[15] Arslan O, Genc AI. The skew generalized t distribution as the scale mixture of a skew exponential power distribution and its applications in robust estimation. Stat (Ber). 2009;43:481–498.

[16] Acitas S, Kasap P, Senoglu B, et al. Robust estimation with the skew t2 distribution. Pakistan J Stat. 2013;29:409–430.

[17] Acitas S, Kasap P, Senoglu B, et al. One-step M-estimators: Jones and Faddy’s Skewed t-distribution. J Appl Stat. 2013;40:1545–1560.

[18] Tiku ML, Surucu B. MMLEs are as good as M-estimators or better. Stat Probab Lett. 2009;79:984–989.

[19] Donmez A. 2010. Adaptive estimation and hypothesis testing methods [dissertation]. Ankara: METU; 2010.

[20] Acitas S, Filzmoser P, Senoglu B. A robust adaptive modified maximum likelihood estimator for the linear regression model. Olomoucian days appl. math. 2019; Olomouc, Czech Republic; 2019.

[21] Puthenpura S, Sinha NK. Modified maximum likelihood method for the robust estimation of system parameters from very noisy data. Automatica. 1986;22:231–235.

[22] Vaughan DC. On the Tiku–Suresh method of estimation. Commun Stat – Theor Meth. 1992;21:329–340.

[23] Bhattacharyya GK. The asymptotics of maximum likelihood and related estimators based on type ii censored data. J Am Stat Assoc. 1985;80:398–404.

[24] Akkaya AD, Tiku ML. Robust estimation in multiple linear regression model with non-Gaussian noise. Automatica. 2008;44:407–417.

[25] Lopuhaa HP, Rousseeuw PJ. Breakdown points of affine equivariant estimators of multivariate location and covariance matrices. Ann Stat. 1991;19:229–248.

[26] Croux C, Filzmoser P, Oliveira MR. Algorithms for projection-Pursuit robust principal component analysis. Chemom Intell Lab Syst. 2007;87:218-225.

[27] Fritz H, Filzmoser P, Croux C. A comparison of algorithms for the multivariate $L_1$-median. Comput Stat. 2012;27:393–410.

[28] Acitas S, Filzmoser P, Senoglu B. A new partial robust adaptive modified maximum likelihood estimator. Chemom Intell Lab Syst. 2020;204:104068.

[29] De Carvalho FDA, Neto EDAL, Ferreira MR. A robust regression method based on exponential-type kernel functions. Neurocomputing. 2017;234:58–74.

[30] Liebmann B, Filzmoser P, Varmuza K. Robust and classical PLS regression compared. J Chemom. 2010;24:111–120.

[31] Filzmoser P, Todorov V. Review of robust multivariate statistical methods in high dimension. Anal Chim Acta. 2011;705:2–14.

[32] Gray JB. 1985. Graphics for Regression Diagnostics. American Statistical Association Proceedings of the Statistical Computing Section, ASA; Washington, DC: 1985