ESKİŞEHİR TEKNİK ÜNİVERSİTESİ BİLİM VE TEKNOLOJİ DERGİSİ
B- TEORİK BİLİMLER

Eskişehir Technical University Journal of Science and Technology B- Theoretical Sciences

2022, 10(1), pp. 35-58, DOI: 10.20290/estubtdb.887201

RESEARCH ARTICLE

ALTERNATIVE ROBUST ESTIMATORS FOR PARAMETERS OF THE LINEAR REGRESSION MODEL

Mutlu ALTUNTAŞ [1,*], Emel ÇANKAYA [2], Olcay ARSLAN [3]

[1] Department of Statistics, Faculty of Arts and Science, Sinop University, Sinop, Turkey
[2] Department of Statistics, Faculty of Arts and Science, Sinop University, Sinop, Turkey
[3] Department of Statistics, Faculty of Science, Ankara University, Ankara, Turkey

ABSTRACT

This paper considers parameter estimation of the linear regression model with Ramsay-Novick (RN) distributed errors, focusing on its use to aid robustness. Positioning within the class of heavy-tailed distributions, RN distribution can be defined as the modification of unbounded influence function of a non-robust density so that it has more resistance to outliers. Potential use of this robust density has been assessed in Bayesian settings on real data examples and there is a lack of performance assessment for finite samples in the classical approach. Therefore, this study explores its robustness properties when used as error distribution compared to normal and other alternating heavy-tailed distributions like Laplace and Student-t. An extensive simulation study was conducted for this purpose under different settings of sample size, model parameters and outlier percentages. An efficient data generation of RN distribution through random-walk Metropolis algorithm is here also suggested. The results were supported by a real world application on famously known as Brownlee's stack loss plant data.

Keywords: Heavy-tailed distribution, Modified influence function, Ramsay-Novick, Random walk metropolis, Robust regression

1. INTRODUCTION

Robustness is a desirable property of statistical estimators so as to make reliable inferences from the data without being too sensitive to the underlying statistical assumptions. Departures from the assumptions of regression have thus long been an intriguing subject in this respect. In most regression modelling scenarios, inference depends much more heavily on the measurement error distribution with a long-tail due to either few of many outlying observations. This brings about a major source of departure from the assumed normal data generating model as the tails of such distributions decrease zero more slowly than the normal case. Robustness to long tails is crucial and achieved by specifying the sampling model, or likelihood, within a class of family of symmetric or skewed distributions enriched with robustness parameters. Attempts to using alternative error distributions, heavy-tailed relative to the normal, have revealed a variety of robust models in the literature [1-4]. A typical example is furnished by univariate Student-t family distributions, a special case of elliptical distributions and robustness achieved by choosing a proper degrees of freedom parameter. This family has been applied widely for robust least-squares fitting of multiple regression [1,5-8]. Slash and contaminated normal families serve as alternating in accommodating outlying measurements of the modelling applications [3,9]. Besides, Laplace distributed errors were also considered for linear and mixture linear regression models [2], having been confirmed that the least absolute deviation (LAD) regression estimator is superior to ordinary least square (OLS) estimator in small samples under the Laplace error model [10]. [4] uncovered extended power distribution family and investigated its robust properties in the context of location parameter estimation.

*Corresponding Author: maltuntas@sinop.edu.tr
Received: 02.03.2021 Published:25.02.2022

In aiding robustness, [11] proposed a procedure completely different in style for the same purpose. They first measured a single observation's influence as a function that shows the rate of change of the sampling model density with respect to the observation. For a density within a certain symmetric family of distributions, they applied a modification on the density's unbounded influence function so that it would be bounded. By deriving the modified influence function backwards, they ended up with a new family of distributions, namely RN distribution family, with robustness properties. By means of this procedure, they also performed a robust Bayesian analysis of linear modelling in their work. Since then, [12] applied a similar idea on Bayesian regression and [13] addressed this new robust family's limitations. [14] compared the estimates and also relative local influence measures of the regression coefficients under different sampling models including RN, again from a Bayesian point of view. They based their comparative study on the same real-world data employed by [11]. An empirical performance assessment through a comprehensive simulation study, for the first time, is presented in a very recent work [15] that considers likelihood robustness via RN distribution and prior robustness via Student-t distribution. They presented theoretical evaluation of robust estimators with these settings as well as an algorithm to generate samples from RN distribution by means of the independent Metropolis Hastings method. All the above mentioned studies concentrate the role of RN distribution in achieving the Bayesian robustness for regression modelling. To the best of our knowledge, there is no attempt to evaluate the resistance of RN error distribution to the outliers comparatively with other heavy-tailed distributions within the classical regression modelling framework.

The main objective of the present study is therefore to propose new robust estimators for a regression model with RN distributed errors when the data have outliers. It is also of our interest to compare the estimators of regression with error distributions as Student-t, Laplace and Ramsay-Novick. The secondary aim of this study is to develop a new algorithm to generate samples from RN distribution. The very first attempt in this respect was made by means of independent Metropolis Hastings algorithm in [15] whereas we here utilized random walk Metropolis algorithm. Therefore, this study enlarges the number of alternative data generation processes from RN distribution. In addition, this study also offers an approach to determine the optimal value for tuning parameter of RN distribution as opposed to the other studies in the literature that assume particular values for the parameters.

In the next section, we first present the procedure to derive the distributional form of RN from a non-robust normal density. Then, we theoretically evaluate the robust estimators for the unknown quantities of regression model when RN distribution is attained to the errors. Section 4 presents the connection between the proposed new robust estimators and the previous robust Bayesian estimators through the representation of our beliefs about the regression parameters by a non-informative prior. Motivated by a performance comparison, RN, Student-t, Laplace and Normal distributions were used for the sampling specification of the regression model. The finite sample performances of these error distributions are evaluated by a simulation study. In addition, the determination of the robustness tuning parameters for RN distribution was performed by the cross-validation as a data-driven method with empirical justification. This is followed by a real-world application on the data famously known as "stack loss data" and widely used in the literature [16-28] and many others. Then the paper is finalized by the results and discussion section. Rv3.2. software [29] is used for all the computations necessary throughout this study.

2. RAMSAY-NOVICK (RN) DISTRIBUTION

[11] raised the issue of robustness by modifying the influence function of a non robust density so that it is bounded and the resulting density yields modified robust influence function. This modification procedure can be applied on a family of densities including Gaussian, the multivariate Gaussian, inverse Gaussian, lognormal and log-odds densities.

A general expression for these densities can be given in the following way:

$$f(x \mid v) = r(x)\, s(v)\, exp\left[-\frac{1}{2} d^2(v,x)\right] \tag{1}$$

where x and v may be vector valued, v is the location parameter, $d(v,x)$ is a measure of distance of x from v. Densities of this form are non-robust having unbounded influence functions. As a measure of robustness to small deviations from the reference model (1), the influence function of x is given by (IF);

$$-\frac{d\, log\ f(x \mid v)}{dx} = d(v,x)\frac{d(d(v,x))}{dx} - \frac{d\, log\, r(x)}{dx} \tag{2}$$

Modified influence function (MIF) can be expressed as

$$-\frac{d\, log\ f(x \mid v)}{dx} = d(v,x)\frac{d(d(v,x))}{dx}\ exp(-ad^b) - \frac{d\, log\, r(x)}{dx} \tag{3}$$

noting that the factor $exp(-ad^b)$ has little effect on the influence function for small values of d. The constants a and b are the robustness tuning constants [11]. Thus, the robust version of a density in Eq. (1) can be defined as

$$f(x \mid v,a,b) = r(x)\, A(v)\, s(v)\, exp(-\eta_{ab}(d)) \quad ; \quad a>0\,, b>0 \tag{4}$$

where $\eta_{ab}(d) = \left(ba^{2/b}\right)^{-1}\gamma\left(2/b, ad^b\right)$, γ is the incomplete gamma function and $A(v)$ is a normalizing constant that does not depend on x [11]. If a random variable X follows the distributional form of Eq. (4), then it is said to have a RN distribution with the parameters a and b.

Suppose, this modification procedure is applied to a random variable X having N(μ,σ) density. Rephrasing the density function as in Eq. (1) produces $d(v,x) = (x-\mu)/\sigma$ and modified influence function of this variable, as defined in Eq. (3), becomes

$$MIF(x) = \frac{(x-\mu)}{\sigma^2} exp\left(-a\left|\frac{x-\mu}{\sigma}\right|^b\right)$$

This modified influence function with the details given in Appendix produces a RN distribution with the probability density function (p.d.f) as follows:

$$f(x \mid v) \propto \exp(-(ba^{2/b})^{-1}\gamma(\frac{2}{b}, a\left|\frac{x-\mu}{\sigma}\right|^b)) \tag{5}$$

and briefly shown as $X \sim RN(\mu,\sigma,a,b)$, here μ and σ are the location and scale parameters of the distribution, respectively. Figure 1 presents examples of this density with some differing values of robustness parameters which influence the thickness of tails of the standard RN distribution $(\mu = 0$ and $\sigma = 1)$.

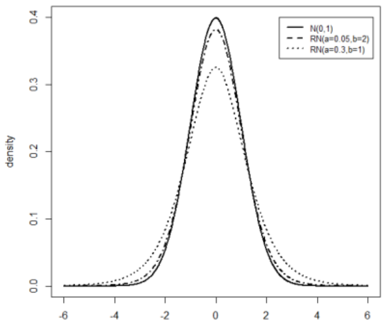

Figure 1. Influence of the robustness parameter on the tails of RN density functions comparatively with Standard Normal.

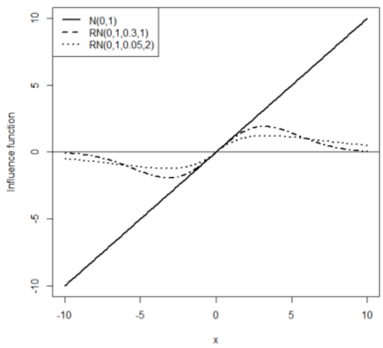

Figure 2. Plots of three influence functions from the Gaussian density

Figure 2 shows this procedure's achievement for converting a non-robust Gaussian density, having an unbounded IF, to a robust density with a bounded IF. It can also be seen here parameter b controls the speed of IF approaching to zero.

3. ROBUST REGRESSION BASED ON RN DISTRIBUTION

Consider the following linear regression model,

$$y_i = \boldsymbol{x_i}'\boldsymbol{\beta} + \varepsilon_i \tag{6}$$

where y_i is a dependent variable, $\boldsymbol{x_i}' = \left(x_{i1},...,x_{ip}\right)$ is a set of regressors, $\boldsymbol{\beta} = \left(\beta_0, \beta_1,...,\beta_p\right)'$ is a vector of unknown model parameters and ε_i are i.i.d. random errors. Assuming the validity of normal data generating model, maximum likelihood (ML) and also ordinary least square (OLS) estimates of $\boldsymbol{\beta}$ are $\hat{\boldsymbol{\beta}}_{\boldsymbol{OLS}} = \left(\boldsymbol{X'X}\right)^{-1}\left(\boldsymbol{X'y}\right)$. The ML estimate for σ^2 is $\hat{\sigma}^2_{ML} = \left(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}\right)'\left(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}\right)/n$ [30-33].

These estimates are known to be highly sensitive to the departures from the assumption of normality mainly due to the presence of outliers in the data set. One suggested way of achieving inferences more robust to outliers is to employ a unimodal heavy-tailed error distributions. Student-t and Laplace distributions are examples of heavy-tailed distributions and are widely applied within the regression framework in the literature [34-38]. We here tried to place the position of RN distribution within those applications as it also has a bounded influence function. Leaving aside the Bayesian perspective, which was the main consideration of the work by [11], we assume that the errors in model (6) have modified RN distribution $\left(\varepsilon_i \sim RN(0,\sigma,a,b)\right)$ and carry on estimation of the parameters of interest using the ML method with this assumption. We have

$$f\left(y_i \mid \mathbf{x}_i, \boldsymbol{\beta}, \sigma, a, b\right) = \left(A_i\right)^{-1} exp\left[-\rho\right] \quad \text{for} \quad i = 1, 2, ..., n \tag{7}$$

where

$$\rho = \eta_{ab}\left(d_i\right) = \left(ba^{2/b}\right)^{-1} \gamma\left(2/b, ad_i^{\,b}\right), \quad A_i = \int_{u_1}^{u_2} \exp\left(-\eta_{ab}(d_i(\mathrm{u}))\right) du, \quad d_i = \left|y_i - \mathbf{x}_i{}'\boldsymbol{\beta}\right| / \sigma$$

Note that the quantity A_i is a normalizing constant, u_1 and u_2 are finite fixed limits of integration. The choice of u_1 and u_2 is not important as long as they are adequately far away from any dependent variable observation to provide robustness. It is assumed that the parameters a and b are fixed and parameters are regarded as the robustness tuning constants like '*k*' in Huber's ρ function. In Eq. (7), $\rho = \eta_{ab}\left(d\right)$ term behaves like the ρ function for M estimator and assures the influence function to be bounded [11]. Plot of ρ function can be seen in Figure 3.

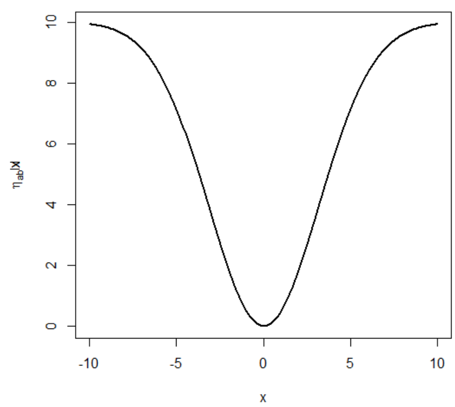

Figure 3. Plot of $\rho = \eta_{ab}\left(|x|\right)$ function with $a = 0.05$ and $b = 2$

Figure 4 reveals a visual comparison of $\rho = x^2/2$ used for OLS estimation and $\rho = \eta_{ab}\left(|x|\right)$ for robust estimation. It is here clearly seen that $\rho = \eta_{ab}\left(|x|\right)$ obtained from RN distribution is bounded and robust to outlying observations of any given data.

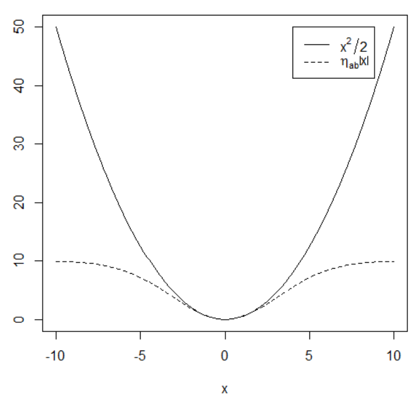

Figure 4. Plots of $\rho = x^2/2$ and $\rho = \eta_{ab}\left(|x|\right)$ function

To obtain estimates of the model parameters, log-likelihood function can be defined as

$$ln\,L\left(\boldsymbol{y};\boldsymbol{\beta},\sigma\right)=ln\prod_{i=1}^{n} f\left(y_i \mid \boldsymbol{x_i}\,,\boldsymbol{\beta},\sigma,a,b\right) = \left(-ln\left(\prod_{i=1}^{n} A_i\right)-\sum_{i=1}^{n}\left(ba^{2/b}\right)^{-1}\gamma\left(2/b,\,ad_i^{\,b}\right)\right) \tag{8}$$

Robust estimator of $\boldsymbol{\beta}$ are the solutions of the equations:

$$\left.\frac{\partial\, ln\,L(\,\boldsymbol{y};\boldsymbol{\beta},\sigma\,)}{\partial\boldsymbol{\beta}}\right|_{\substack{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}_{RN}\\ \sigma=\hat{\sigma}_{RN}}}=0$$

$$\left(-\sum_{i=1}^{n}\left(\partial A_i/\partial\boldsymbol{\beta}\right)\big/A_i\right)-\left(ba^{2/b}\right)^{-1}\sum_{i=1}^{n}\left.\frac{\partial^2\gamma\left(\frac{2}{b},ad_i^{\,b}\right)}{\partial\left(ad_i^{\,b}\right)\partial\boldsymbol{\beta}}\right|_{\substack{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}_{RN}\\ \sigma=\hat{\sigma}_{RN}}}=0 \tag{9}$$

The derivatives of these equations are a computational problem since it requires numerical integration for each observation. However, it is shown that the quantities $\left(\partial A_i/\partial\hat{\boldsymbol{\beta}}_{\boldsymbol{RN}}\right)\big/A_i$ approach zero in the limit [11]. The partial derivative of lower incomplete gamma function in Eq. (9) w.r.t $\boldsymbol{\beta}$ can be obtained as follows:

$$\left.\frac{\partial^2\gamma\left(\frac{2}{b},ad_i^{\,b}\right)}{\partial\left(ad_i^{\,b}\right)\partial\boldsymbol{\beta}}\right|_{\substack{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}_{\boldsymbol{RN}}\\ \sigma=\hat{\sigma}_{\boldsymbol{RN}}}}=a^{2/b}b\left(\hat{\sigma}_{RN}^{-2}\right)\left(y_i-\boldsymbol{x_i}'\hat{\boldsymbol{\beta}}_{\boldsymbol{RN}}\right)\left(-\boldsymbol{x_i}'\right)\hat{w}_{i_{(RN)}}$$

where $\hat{w}_{i_{(RN)}}=exp\left(-a\left(\left|y_i-\boldsymbol{x_i}'\hat{\boldsymbol{\beta}}_{\boldsymbol{RN}}\right|\big/\hat{\sigma}_{RN}^{b}\right)^{b}\right)$. In this respect, derivative of log-likelihood function for $\boldsymbol{\beta}$ is obtained as

$$\left.\frac{\partial\, ln\,L(\,\boldsymbol{y};\boldsymbol{\beta},\boldsymbol{\sigma}\,)}{\partial\boldsymbol{\beta}}\right|_{\substack{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}_{RN}\\ \sigma=\hat{\sigma}_{RN}}}=\left(\hat{\boldsymbol{\sigma}}_{RN}^{-2}\right)\sum_{i=1}^{n}\left[\hat{w}_{i_{(RN)}}x_i\left(y_i-\boldsymbol{x_i}'\hat{\boldsymbol{\beta}}_{RN}\right)\right]=0$$

Then, robust estimator of $\boldsymbol{\beta}$ is defined as follows:

$$\hat{\boldsymbol{\beta}}_{\boldsymbol{RN}}=\left(\sum_{i=1}^{n}\hat{w}_{i_{(RN)}}\boldsymbol{x_i}\boldsymbol{x_i}'\right)^{-1}\left(\sum_{i=1}^{n}\hat{w}_{i_{(RN)}}\boldsymbol{x_i}y_i\right)$$

In matrix form as

$$\hat{\boldsymbol{\beta}}_{\boldsymbol{RN}}=\left(\boldsymbol{X'WX}\right)^{-1}\left(\boldsymbol{X'Wy}\right) \tag{10}$$

where $\boldsymbol{y}$ is a column vector of n elements, $\boldsymbol{X}$ is an $n\times(p+1)$ design matrix and $\boldsymbol{\beta}$ is a column vector of the $(p+1)$ elements. The matrix $\boldsymbol{W}=diag(\hat{\boldsymbol{w}}_{\boldsymbol{1}},\hat{\boldsymbol{w}}_{\boldsymbol{2}},...,\hat{\boldsymbol{w}}_{\boldsymbol{n}})$ with the scalars $\hat{w}_i$. Similarly, robust estimator of σ^2 is obtained from the following equation:

$$\left.\frac{\partial \ln L(\boldsymbol{y};\boldsymbol{\beta},\sigma)}{\partial\sigma}\right|_{\substack{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}_{RN}\\ \sigma=\hat{\sigma}_{RN}}}=0 \tag{11}$$

The partial derivative of lower incomplete gamma function in Eq. (8) w.r.t σ can be obtained as follows:

$$\left.\frac{\partial^2\gamma\left(\frac{2}{b},a\mathrm{d}_{\mathrm{i}}^{\,b}\right)}{\partial\left(a\mathrm{d}_{\mathrm{i}}^{\,b}\right)\partial\boldsymbol{\sigma}}\right|_{\substack{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}_{RN}\\ \sigma=\hat{\sigma}_{RN}}}=a^{2/b}\,\hat{\boldsymbol{\sigma}}_{RN}^{-3}(-b)\,\hat{w}_{i_{(RN)}}\left(y_i-\boldsymbol{x}_{\boldsymbol{i}}'\hat{\boldsymbol{\beta}}_{RN}\right)^2$$

Derivative of log-likelihood function for σ is obtained as

$$\begin{aligned}\left.\frac{\partial \ln L(\boldsymbol{y};\boldsymbol{\beta},\sigma)}{\partial\sigma}\right|_{\substack{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}_{\boldsymbol{RN}}\\ \sigma=\hat{\sigma}_{RN}}}&=\left(-\sum_{i=1}^{n}\left(\partial A_i/\partial\hat{\sigma}_{RN}\right)/A_i\right)+\left(ba^{2/b}\right)^{-1}\\&\times\sum_{i=1}^{n}\left[a^{2/b}\left(\hat{\sigma}_{RN}^{-3}\right)b\,\hat{w}_{i_{(RN)}}\left(y_i-\boldsymbol{x}_{\boldsymbol{i}}'\hat{\boldsymbol{\beta}}_{\boldsymbol{RN}}\right)^2\right]=0\end{aligned}$$

Here the quantities $(\partial A_i/\partial\hat{\sigma}_{RN})/A_i$ converge in the limit to a value that does not depend on i if it is assumed that the interval of integration includes all observations [11], thus can be ignored in this equation. In this case, robust estimator of σ^2 is defined as follows:

$$\hat{\sigma}_{RN}^2=\left(\hat{\sigma}_{RN}\right)^{-1}\sum_{i=1}^{n}\hat{w}_{i_{(RN)}}\left(y_i-\boldsymbol{x}_{\boldsymbol{i}}'\hat{\boldsymbol{\beta}}_{RN}\right)^2 \tag{12}$$

Numerical methods are necessary in obtaining these estimators because of $\mathbf{W}$ matrix includes both $\boldsymbol{\beta}$ and σ.

4. BAYESIAN INTERPRETATION OF RN ESTIMATES

Consider the standard multiple linear regression model as in Eq. (6). Assuming that the errors are independent and RN distributed random variables, the likelihood function is:

$$p\left(\boldsymbol{y}\mid x,\boldsymbol{\beta},\sigma,a,b\right)\propto exp\left[-\sum_{i=1}^{n}\left(ba^{2/b}\right)^{-1}\gamma\left(2/b,a\frac{\left|y_i-x_i'\beta\right|^b}{\sigma^b}\right)\right] \tag{13}$$

A vague prior jointly for $\boldsymbol{\beta}$ and σ is

$$p(\boldsymbol{\beta},\sigma)\propto 1/\sigma$$

Resulting posterior distribution is

$$p(\boldsymbol{\beta},\sigma \mid y, x, a, b) \propto p(y \mid x, \boldsymbol{\beta}, \sigma, a, b) \times p(\boldsymbol{\beta},\sigma)$$
$$\propto \sigma^{-1} exp\left[-\sum_{i=1}^{n}\left(ba^{2/b}\right)^{-1}\gamma\left(2/b, a\frac{\left|y_i - x_i'\beta\right|^b}{\sigma^b}\right)\right]$$

The log-posterior function is defined as

$$ln\, p(\boldsymbol{\beta},\sigma \mid y, x, a, b) \propto (-ln\sigma) - \sum_{i=1}^{n}\left(ba^{2/b}\right)^{-1}\gamma\left(2/b, a\frac{\left|y_i - x_i'\beta\right|^b}{\sigma^b}\right) \tag{14}$$

The partial derivative of the log-posterior w.r.t. $\boldsymbol{\beta}$ produces the following equations:

$$\left.\frac{\partial\, ln\, p(\boldsymbol{\beta},\sigma \mid \boldsymbol{y},\boldsymbol{x},a,b)}{\partial\boldsymbol{\beta}}\right|_{\substack{\beta=\hat{\beta}_{RN_{bayes}} \\ \sigma=\hat{\sigma}_{RN_{bayes}}}} = 0$$
$$-\left(ba^{2/b}\right)^{-1}\sum_{i=1}^{n}a^{2/b}b\left(\hat{\sigma}_{RN_{bayes}}^{-2}\right)\left(y_i - \boldsymbol{x_i}'\hat{\boldsymbol{\beta}}_{\boldsymbol{RN_{bayes}}}\right)\left(-\boldsymbol{x_i}'\right) \tag{15}$$
$$\times exp\left(-a\frac{\left|y_i - \boldsymbol{x_i}'\hat{\boldsymbol{\beta}}_{\boldsymbol{RN_{bayes}}}\right|^b}{\hat{\sigma}_{RN_{bayes}}^b}\right) = 0$$

where $\hat{w}_{i_{(RN_{bayes})}} = exp\left(-a\left(\left|y_i - \boldsymbol{x_i}'\hat{\boldsymbol{\beta}}_{\boldsymbol{RN_{bayes}}}\right| \Big/ \hat{\sigma}_{RN_{bayes}}^b\right)^b\right)$. In this case, Bayesian estimator of $\boldsymbol{\beta}$ is obtained as follows:

$$\hat{\boldsymbol{\beta}}_{\boldsymbol{RN_{bayes}}} = \left(\sum_{i=1}^{n}\hat{w}_{i_{RN_{bayes}}}\boldsymbol{x_i}\boldsymbol{x_i}'\right)^{-1}\left(\sum_{i=1}^{n}\hat{w}_{i_{RN_{bayes}}}\boldsymbol{x_i}y_i\right)$$

In matrix form as

$$\hat{\boldsymbol{\beta}}_{\boldsymbol{RN_{bayes}}} = \left(\boldsymbol{X'WX}\right)^{-1}\left(\boldsymbol{X'Wy}\right) \tag{16}$$

which is exact the same give in Eq. (10). Similarly, Bayesian estimate of σ^2 can be obtained from the solution of the following Eq. (17).

$$\left.\frac{\partial\, ln\, p(\boldsymbol{\beta},\sigma \mid \boldsymbol{y},\boldsymbol{x},a,b)}{\partial\sigma}\right|_{\substack{\beta=\hat{\beta}_{RN_{bayes}} \\ \sigma=\hat{\sigma}_{RN_{bayes}}}} = 0$$

$$-\frac{1}{\hat{\sigma}_{RN_{bayes}}}-\left(ba^{2/b}\right)^{-1}\sum_{i=1}^{n}a^{2/b}\left(\hat{\sigma}_{RN_{bayes}}^{-3}\right)(-b)$$

$$\times exp\left(-a\frac{\left|y_i-\boldsymbol{x_i}'\hat{\boldsymbol{\beta}}_{\boldsymbol{RN_{bayes}}}\right|^{b}}{\hat{\sigma}_{RN_{bayes}}^{b}}\right)\left(y_i-\boldsymbol{x_i}'\hat{\boldsymbol{\beta}}_{\boldsymbol{RN_{bayes}}}\right)^{2}=0$$

Arranging this equation we get the following $\hat{\sigma}_{RN_{bayes}}^{2}$.

$$\hat{\sigma}_{RN_{bayes}}^{2}=\sum_{i=1}^{n}\hat{w}_{i_{\left(RN_{bayes}\right)}}\left(y_i-\boldsymbol{x_i}'\hat{\boldsymbol{\beta}}_{\boldsymbol{RN_{bayes}}}\right)^{2} \tag{17}$$

Note that this weights have same forms given in Eq. (10). In addition, numerical methods are necessary for solution of Eq. (16) and Eq. (17).

5. GENERATING SAMPLES FROM RN DISTRIBUTION

Due to the unusual functional form of RN, direct sampling of this distribution is not straightforward. We therefore utilized random walk Metropolis Hastings algorithm to generate samples from this distribution [39-40]. Target distribution is here the standard RN with density function (d.f.);

$$f\left(x;a,b\right)\propto exp\left(-\left(\left(b\,a^{\frac{2}{b}}\right)^{-1}\gamma\left(\frac{2}{b},a|x|^{b}\right)\right)\right)\quad ;\ a,b>0\ ;\ \gamma=\text{ incomplete gamma function}$$

For the generation of candidate values of the chain, the proposal distribution $g\left(x\right)$ is chosen as Normal with a mean determined by the chain's current state (i.e. random walk), and with a scale defining the size of the walk. Therefore, the scale of the proposal was chosen big enough to ensure tail thickness and good mixing properties. A realization of a first-order Markov process, $x^{(1)},x^{(2)},\ldots,x^{(t)}$, can then be generated by the following steps:

1) Set the parameter values for a and b. Then initialize the algorithm by an arbitrary value $x^{(0)}$.
2) At step ($t-1$), current state is $x^{(t-1)}$. Generate a candidate state, x^{*}, using a random walk as;

$$x^{*}=x_{t-1}+\varepsilon\quad\text{where }\varepsilon\sim N\left(0,3.4\right)$$

3) Calculate the ratio of two states:

$$\alpha_t\left(x^{(t-1)},x^{*}\right)=min\left\{1,\frac{f\left(x^{*};a,b\right)}{f\left(x^{(t-1)};a,b\right)}\frac{g\left(x^{(t-l)}\middle|x^{*}\right)}{g\left(x^{*}\middle|x^{(t-l)}\right)}\right\}$$

as the proposal density is symmetric, the ratio reduces to $\dfrac{f\left(x^{*};a,b\right)}{f\left(x^{(t-1)};a,b\right)}$.

4) Generate a random value of u from U(0,1)
5) If $u \le \alpha_t\left(x^{(t-1)},x^{*}\right) \Rightarrow$ set $x^{(t)} = x^{*}$ else set $x^{(t)} = x^{(t-1)}$
6) Repeat steps (2)-(5) N times (# of iteration)

An initial run of the chain with 15000 iterations and a burn-in period of 5000 produced the acceptance rate as 0.398 which is plausible as the desirable value for an acceptance rate is stated to be between 0.2 and 0.7 [41]. In the process of generating samples, we run the chain longer and recorded every 5th value in order to minimize the autocorrelation. Figure 5 presents the resulting chain and its autocorrelation function.

A sample with the size of 500, randomly chosen from the whole chain, was also plotted here for a comparison of theoretical and empirical densities, which indicated that the generated sample follows a RN distribution.

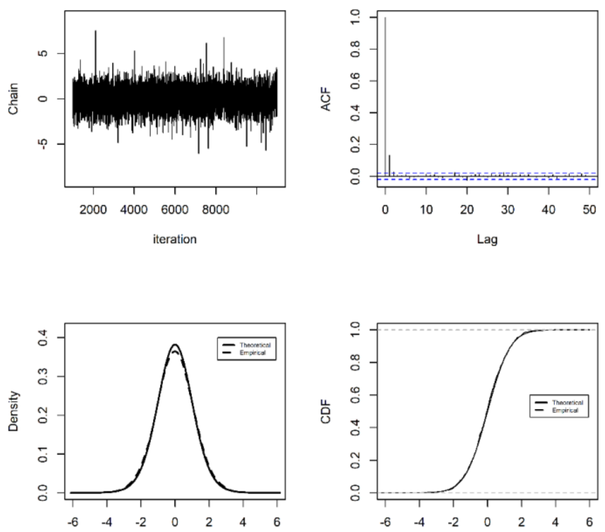

Figure 5. Comparison of density functions of theoretical RN(a = 0.05, b = 2) and generated sample

6. SIMULATION STUDY

We consider the following model

$$y_i = \beta_0 + \beta_1 x_{i1} + ... + \beta_p x_{ip} + \varepsilon_i \quad , \quad i = 1,2,...,n$$

with $p = 5$. All β_j coefficients are set to the value of 1. The sample size are taken as n = 20, 50 and 100. The regressors are generated from the standard normal distribution. For the error, we consider the following distributions: N(0,1), Student-t $(v = 3)$, standard Laplace, standard RN (0, 1, a = 0.05, b = 2). Tail behaviours of these alternative error distributions are given in Figure 6. To assess the influence of outliers on the proposed estimators, we consider adding outliers to the data in $y-$ direction with the amount of $n*k\%$ were k is taken as 5 and 10. The outliers are generated from N(100,1) and added to the data. With these settings; four error distributions, three sample sizes, an outlier case with a percentage of 5 or 10 and no outlier case altogether made up to 108 scenarios in total for this simulation study. We generated 1000 Monte Carlo replicates for each settings.

Parameter estimates for Normal model were obtained by means of "*lm"* function in {stats} library, for Laplace model via "l1fit" function as well as "*rlaplace*" function to generate Laplace distributed errors in {vgam} library and for Student-t model by means of "*tlm*" function in {hett} library of Rv.3.2 software [29]. Iteratively Reweighted Least Squares (IRLS) numerical solution of the Eq. (10) and Eq. (12) were applied for the same purpose when the errors of the model are assumed to follow RN distribution. Functions of our own were used to calculate the RN estimates. We computed the empirical mean, standard error (SE), bias and root mean square error (RMSE) of the parameter estimators and the results are presented in Tables 1-5. All necessary computations were performed within R platform.

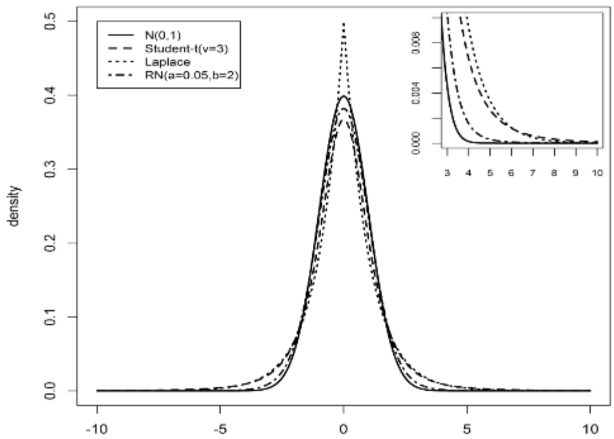

Figure 6. Visual comparison of tails of the considered error distributions

6.1. Simulation Results

Regression modelling of the Normal generating data without outliers revealed that the results for robustified version of Normal density namely Ramsay-Novick are similar with the estimates of OLS estimators at the expence of some computational cost. RMSE's of RN estimates appears slightly larger than Normal, however this difference diminishes as the sample size enlarges. When there is no departures from normality assumption, Student-t and Laplace regression estimators also perform well except for scale parameters evaluated incorrectly with larger RMSE's (Table 1). Assumption of the Student-t data generation process without outliers produced performances, in the decreasing order, as RN, Laplace and Normal followed by naturally the best estimates of Student-t. It must be noted that the scale parameter was better estimated by Student-t and Laplace estimators compared to others. Student-t fit of the model out performs even though the errors were generated from Laplace distribution for both considered number of parameters. RN estimates except for scale parameter also appears reasonably well. If a RN distribution family is assumed to fit best to the data, performances of both OLS and RN estimates appear almost the same. Estimates of Student-t and Laplace follow those comparatively with respect to the RMSE measure (Table 1).

Table 1. Robust and non-robust estimates of the regression model without outliers

n	Estimators	$\hat{\beta}_0$ (RMSE)	$\hat{\beta}_1$ (RMSE)	$\hat{\beta}_2$ (RMSE)	$\hat{\beta}_3$ (RMSE)	$\hat{\beta}_4$ (RMSE)	$\hat{\beta}_5$ (RMSE)	$\hat{\sigma}$ (RMSE)
$\varepsilon_i \sim$ Normal								
20	Normal	0.994(0.26)	1.005(0.28)	0.999(0.28)	1.019(0.28)	0.994(0.27)	0.993(0.28)	0.816(0.24)
	t(3)	0.991(0.29)	1.006(0.31)	0.999(0.30)	1.019(0.31)	0.994(0.31)	0.997(0.31)	0.588(0.43)
	Laplace	0.991(0.32)	1.013(0.34)	0.996(0.33)	1.027(0.34)	0.987(0.34)	0.997(0.34)	0.603(0.42)
	RN(0.05,2)	0.993(0.26)	1.006(0.28)	0.999(0.28)	1.019(0.27)	0.996(0.28)	0.993(0.28)	0.748(0.29)
50	Normal	0.994(0.14)	1.001(0.15)	0.997(0.15)	0.998(0.15)	0.998(0.15)	0.999(0.15)	0.929(0.12)
	t(3)	0.993(0.16)	1.003(0.16)	0.998(0.16)	0.995(0.16)	0.998(0.16)	0.995(0.16)	0.720(0.29)
	Laplace	0.990(0.18)	1.005(0.19)	0.999(0.18)	0.991(0.18)	1.004(0.18)	0.988(0.18)	0.721(0.29)
	RN(0.05,2)	0.993(0.15)	1.001(0.15)	0.997(0.15)	0.997(0.15)	0.998(0.15)	0.998(0.15)	0.854(0.17)
100	Normal	1.002(0.09)	0.999(0.10)	1.007(0.10)	0.999(0.10)	1.005(0.10)	1.001(0.10)	0.961(0.08)
	t(3)	1.003(0.10)	0.999(0.11)	1.007(0.11)	0.999(0.11)	1.006(0.11)	1.003(0.11)	0.753(0.25)
	Laplace	1.001(0.12)	0.999(0.13)	1.006(0.13)	0.996(0.12)	1.004(0.12)	0.999(0.13)	0.755(0.25)
	RN(0.05,2)	1.002(0.09)	0.999(0.10)	1.007(0.10)	0.999(0.10)	1.005(0.10)	1.002(0.10)	0.884(0.13)
$\varepsilon_i \sim t(3)$								
20	Normal	1.001(0.25)	1.005(0.27)	1.005(0.28)	0.995(0.27)	1.001(0.27)	1.007(0.27)	0.829(0.23)
	t(3)	1.009(0.28)	1.005(0.30)	1.002(0.31)	0.997(0.30)	1.005(0.31)	1.004(0.30)	0.598(0.42)
	Laplace	1.017(0.31)	1.007(0.33)	1.000(0.34)	0.989(0.33)	1.007(0.34)	1.005(0.34)	0.609(0.41)
	RN(0.05,2)	1.002(0.25)	1.005(0.27)	1.004(0.28)	0.995(0.27)	1.001(0.28)	1.007(0.27)	0.761(0.28)
50	Normal	0.995(0.14)	1.009(0.15)	0.998(0.15)	1.004(0.15)	0.998(0.15)	0.995(0.15)	0.930(0.12)
	t(3)	0.993(0.16)	1.006(0.16)	1.002(0.17)	1.008(0.16)	1.000(0.17)	0.994(0.17)	0.719(0.29)
	Laplace	0.991(0.18)	1.004(0.19)	1.000(0.19)	1.006(0.19)	1.000(0.19)	0.992(0.19)	0.721(0.29)
	RN(0.05,2)	0.995(0.14)	1.008(0.15)	0.998(0.15)	1.005(0.15)	0.999(0.15)	0.995(0.15)	0.855(0.17)
100	Normal	0.999(0.10)	1.001(0.10)	0.995(0.10)	1.002(0.10)	1.009(0.10)	0.995(0.10)	0.965(0.08)
	t(3)	0.999(0.10)	1.002(0.11)	0.993(0.11)	1.004(0.11)	1.010(0.17)	0.994(0.11)	0.756(0.25)
	Laplace	0.995(0.12)	1.003(0.13)	0.990(0.12)	1.004(0.12)	1.009(0.12)	0.993(0.13)	0.759(0.25)
	RN(0.05,2)	0.999(0.10)	1.001(0.10)	0.995(0.10)	1.002(0.10)	1.009(0.10)	0.995(0.10)	0.887(0.13)
$\varepsilon_i \sim$ Laplace								
20	Normal	1.003(0.26)	0.993(0.28)	1.003(0.28)	1.011(0.26)	1.005(0.27)	1.005(0.28)	0.817(0.24)
	t(3)	1.001(0.29)	0.988(0.30)	1.009(0.32)	1.012(0.30)	1.007(0.30)	1.009(0.32)	0.591(0.43)
	Laplace	1.006(0.32)	0.987(0.33)	1.013(0.34)	1.006(0.32)	1.003(0.33)	1.006(0.35)	0.610(0.42)
	RN(0.05,2)	1.002(0.26)	0.992(0.28)	1.005(0.28)	1.012(0.27)	1.006(0.27)	1.006(0.29)	0.750(0.29)
50	Normal	0.996(0.15)	0.998(0.14)	0.998(0.15)	1.003(0.16)	0.994(0.15)	1.006(0.15)	0.935(0.12)
	t(3)	0.998(0.16)	0.998(0.16)	0.998(0.16)	1.003(0.17)	0.995(0.17)	1.007(0.16)	0.722(0.29)
	Laplace	1.001(0.18)	0.999(0.18)	0.999(0.19)	1.001(0.19)	0.993(0.19)	1.008(0.19)	0.724(0.29)
	RN(0.05,2)	0.997(0.15)	0.998(0.14)	0.998(0.15)	1.002(0.16)	0.994(0.16)	1.007(0.15)	0.859(0.17)
100	Normal	0.999(0.10)	1.000(0.09)	1.005(0.11)	0.999(0.10)	0.997(0.10)	1.003(0.10)	0.965(0.08)
	t(3)	0.999(0.11)	1.002(0.10)	1.007(0.11)	0.997(0.11)	0.997(0.11)	1.003(0.11)	0.754(0.25)
	Laplace	0.999(0.13)	1.002(0.12)	1.008(0.13)	0.994(0.13)	0.995(0.13)	1.002(0.13)	0.757(0.25)
	RN(0.05,2)	0.999(0.10)	1.000(0.10)	1.005(0.11)	0.999(0.10)	0.997(0.10)	1.004(0.11)	0.886(0.13)
$\varepsilon_i \sim$ RN(0.05,2)								
20	Normal	1.011(0.29)	1.003(0.30)	1.008(0.29)	0.992(0.29)	1.002(0.30)	0.992(0.31)	0.893(0.21)
	t(3)	1.008(0.32)	1.005(0.33)	1.014(0.33)	0.983(0.32)	1.000(0.33)	0.992(0.34)	0.635(0.39)
	Laplace	1.003(0.35)	1.006(0.36)	1.018(0.36)	0.985(0.36)	1.001(0.37)	0.994(0.36)	0.655(0.38)
	RN(0.05,2)	1.009(0.29)	1.003(0.30)	1.009(0.29)	0.991(0.29)	1.003(0.30)	0.994(0.31)	0.813(0.25)
50	Normal	1.003(0.16)	0.995(0.16)	0.999(0.17)	0.989(0.16)	1.005(0.15)	0.998(0.16)	0.981(0.11)
	t(3)	1.005(0.17)	0.996(0.17)	0.997(0.18)	0.987(0.17)	1.005(0.16)	1.006(0.17)	0.754(0.26)
	Laplace	1.007(0.20)	0.996(0.20)	0.992(0.21)	0.990(0.19)	1.009(0.19)	1.002(0.19)	0.757(0.26)
	RN(0.05,2)	1.002(0.16)	0.995(0.16)	0.999(0.17)	0.988(0.16)	1.005(0.15)	0.998(0.16)	0.898(0.14)
100	Normal	1.012(0.11)	1.001(0.11)	1.002(0.11)	0.994(0.10)	0.998(0.10)	1.001(0.11)	1.023(0.09)
	t(3)	1.000(0.12)	1.000(0.12)	0.999(0.12)	0.994(0.11)	0.998(0.11)	1.001(0.12)	0.791(0.22)
	Laplace	0.999(0.14)	0.999(0.13)	0.999(0.13)	0.997(0.13)	1.001(0.13)	1.002(0.14)	0.797(0.21)
	RN(0.05,2)	1.007(0.11)	1.001(0.11)	1.001(0.11)	0.994(0.11)	0.998(0.10)	1.001(0.11)	0.933(0.09)

Existence of outliers in $y-$ direction however caused large departures in the parameter estimates when Normal distribution was adapted for the errors. It can also be seen that the estimates are very vulnerable to increasing number of outlying observations. On the other hand, RN estimators can tolerate such an influence without being affected regardless of the outliers percentage level with respect to the sample size. This robust likelihood correctly assigned the parameter values with RMSE’s getting smaller as the sample size increases. It must be noted that under all settings Student-t estimators also behave well with very slightly larger variances (Table 2).

Table 2. Robust estimates for the regression model with outliers in $y-$ direction ($\varepsilon_i \sim N(0,1)$)

n	Outlier (%)	Estimators	$\hat{\beta}_0$ (RMSE)	$\hat{\beta}_1$ (RMSE)	$\hat{\beta}_2$ (RMSE)	$\hat{\beta}_3$ (RMSE)	$\hat{\beta}_4$ (RMSE)	$\hat{\beta}_5$ (RMSE)	$\hat{\sigma}$ (RMSE)
20	5	Normal	3.636(3.11)	1.354(3.46)	1.064(2.64)	0.801(3.31)	1.186(3.00)	1.134(3.03)	9.144(8.19)
		t(3)	1.034(0.28)	1.059(0.34)	0.990(0.32)	0.999(0.34)	0.995(0.35)	1.019(0.33)	0.722(0.31)
		Laplace	1.179(0.39)	1.066(0.44)	1.005(0.39)	0.990(0.41)	1.001(0.44)	1.032(0.41)	2.991(1.99)
		RN(0.05,2)	1.019(0.26)	1.052(0.33)	0.999(0.30)	1.010(0.29)	0.990(0.32)	1.007(0.30)	0.738(0.29)
	10	Normal	6.261(5.65)	1.395(3.87)	0.923(3.82)	1.387(4.16)	1.369(3.96)	1.406(4.02)	12.752(11.79)
		t(3)	1.019(0.31)	1.051(0.33)	0.999(0.30)	1.071(0.34)	1.022(0.31)	1.029(0.29)	0.870(0.26)
		Laplace	1.169(0.43)	1.082(0.46)	0.974(0.39)	1.082(0.46)	1.025(0.43)	1.073(0.38)	5.384(4.39)
		RN(0.05,2)	1.010(0.29)	1.047(0.31)	1.002(0.29)	1.069(0.34)	1.017(0.29)	1.034(0.29)	0.694(0.35)
50	5	Normal	3.944(3.00)	0.739(1.86)	1.030(1.76)	1.223(1.65)	0.716(1.61)	0.873(1.74)	11.128(10.14)
		t(3)	0.992(0.18)	0.990(0.16)	1.003(0.16)	1.013(0.18)	0.973(0.17)	0.987(0.15)	0.858(0.18)
		Laplace	1.077(0.21)	0.984(0.20)	1.006(0.19)	1.008(0.22)	0.965(0.20)	0.966(0.19)	3.599(2.60)
		RN(0.05,2)	0.986(0.16)	0.992(0.15)	0.999(0.15)	1.016(0.18)	0.976(0.17)	0.989(0.14)	0.812(0.21)
	10	Normal	6.015(5.06)	0.999(2.09)	0.727(2.27)	1.295(2.29)	0.886(2.35)	0.438(2.39)	13.978(12.99)
		t(3)	1.018(0.18)	0.998(0.16)	1.005(0.17)	0.988(0.16)	1.020(0.17)	1.028(0.18)	1.003(0.12)
		Laplace	1.158(0.25)	0.999(0.19)	1.001(0.21)	1.006(0.21)	1.031(0.24)	1.030(0.23)	5.531(4.53)
		RN(0.05,2)	1.008(0.17)	0.999(0.16)	1.003(0.17)	0.994(0.16)	1.019(0.16)	1.026(0.17)	0.791(0.23)
100	5	Normal	3.430(2.45)	0.949(0.99)	0.999(1.11)	0.976(1.18)	0.922(1.24)	1.004(1.07)	10.474(9.48)
		t(3)	1.012(0.11)	0.999(0.11)	0.993(0.12)	1.004(0.13)	0.987(0.11)	1.010(0.12)	0.889(0.13)
		Laplace	1.073(0.16)	0.988(0.13)	1.001(0.14)	1.008(0.16)	1.003(0.13)	1.004(0.14)	3.176(2.18)
		RN(0.05,2)	1.009(0.10)	1.001(0.12)	0.988(0.11)	1.005(0.12)	0.986(0.11)	1.009(0.11)	0.868(0.15)
	10	Normal	5.958(4.97)	1.036(1.48)	1.035(1.48)	0.978(1.51)	1.196(1.51)	0.728(1.49)	14.375(13.38)
		t(3)	1.036(0.11)	1.010(0.11)	1.006(0.10)	1.020(0.11)	1.006(0.11)	1.022(0.11)	1.057(0.10)
		Laplace	1.171(0.21)	1.016(0.15)	1.013(0.13)	1.015(0.13)	1.019(0.14)	1.029(0.16)	5.559(4.56)
		RN(0.05,2)	1.022(0.11)	1.007(0.11)	1.005(0.10)	1.020(0.11)	1.004(0.10)	1.021(0.11)	0.828(0.18)

It is challenging to observe that RN estimates perform almost equivalently well as the Student-t estimates when the errors were allowed to have tails of Student-t and also included outliers. LS estimates are again vulnarable to the abnormal observations but its robustified version shows great performance in obtaining real values of parameters (Table 3). When the errors were generated from Laplace distribution, Student-t estimates were observed to be better than Laplace estimates. The performance of RN estimators follow the Student-t and as the sample size gets larger they compete with Student-t estimates (Table 4).

Table 3. Robust estimates for the regression model with outliers in $y-$ direction ($\varepsilon_i \sim t\ (3)$)

n	Outlier (%)	Estimators	$\hat{\beta}_0$ (RMSE)	$\hat{\beta}_1$ (RMSE)	$\hat{\beta}_2$ (RMSE)	$\hat{\beta}_3$ (RMSE)	$\hat{\beta}_4$ (RMSE)	$\hat{\beta}_5$ (RMSE)	$\hat{\sigma}$ (RMSE)
20	5	Normal	3.429(2.86)	1.052(2.99)	0.873(2.87)	0.889(2.98)	1.057(2.99)	1.338(2.99)	9.149(8.21)
		t(3)	0.985(0.27)	0.988(0.30)	1.039(0.27)	1.004(0.31)	1.003(0.29)	0.999 (0.36)	0.701(0.34)
		Laplace	1.094(0.36)	1.001(0.37)	1.042(0.33)	0.993(0.36)	0.979(0.40)	0.997(0.42)	2.989(1.99)
		RN(0.05,2)	0.984(0.27)	0.986(0.30)	1.035(0.24)	1.004(0.28)	1.014(0.26)	1.012(0.32)	0.725(0.31)
	10	Normal	5.648(5.11)	0.706(4.45)	0.409(4.26)	0.954(4.52)	0.664(4.67)	0.791(4.47)	12.517(11.61)
		t(3)	1.046(0.31)	0.987(0.30)	0.999(0.28)	0.989(0.29)	0.988(0.32)	0.993(0.34)	0.849(0.25)
		Laplace	1.224(0.46)	0.974(0.42)	0.965(0.39)	0.981(0.46)	0.980(0.43)	0.951 (0.46)	5.419(4.42)
		RN(0.05,2)	1.037(0.30)	0.991(0.29)	0.993(0.27)	0.995(0.28)	0.990(0.31)	0.997(0.33)	0.685(0.35)
50	5	Normal	3.967(3.02)	1.101(1.85)	0.868(1.73)	0.584(1.85)	0.867(1.74)	0.799(1.95)	11.084(10.09)
		t(3)	0.996(0.15)	1.019(0.18)	1.001(0.17)	1.006(0.16)	0.994(0.16)	1.019(0.18)	0.881(0.17)
		Laplace	1.079(0.19)	1.028(0.24)	0.993(0.19)	1.001(0.21)	0.985(0.20)	1.015(0.22)	3.622(2.62)
		RN(0.05,2)	0.998(0.15)	1.021(0.16)	1.001(0.16)	1.015(0.15)	0.994(0.14)	1.019(0.17)	0.833(0.19)
	10	Normal	5.786(4.86)	1.067(2.14)	0.470(2.36)	0.853(2.43)	1.306(2.50)	0.744(2.61)	13.911(12.92)
		t(3)	1.015(0.18)	1.023(0.17)	1.008(0.17)	0.978(0.18)	0.999(0.19)	1.008(0.17)	1.032(0.12)
		Laplace	1.165(0.27)	1.032(0.23)	0.979(0.23)	0.974(0.24)	1.001(0.22)	1.005(0.24)	5.536(4.54)
		RN(0.05,2)	0.999(0.18)	1.019(0.17)	1.008(0.17)	0.979(0.18)	0.996(0.18)	1.006(0.16)	0.815(0.20)
100	5	Normal	3.475(2.49)	0.984(1.07)	0.870(1.04)	0.965(1.06)	1.051(1.28)	0.950(1.17)	10.462(9.47)
		t(3)	1.033(0.12)	0.991(0.12)	1.006(0.12)	0.996(0.11)	0.996(0.11)	0.996(0.12)	0.879(0.14)
		Laplace	1.096(0.16)	1.009(0.14)	1.008(0.14)	0.996(0.14)	0.990(0.14)	0.986(0.14)	3.166(2.17)
		RN(0.05,2)	1.022(0.11)	0.994(0.12)	1.003(0.11)	1.001(0.11)	0.998(0.11)	0.993(0.11)	0.861(0.16)
	10	Normal	5.972(4.99)	0.794(1.64)	0.888(1.57)	1.182(1.49)	1.028(1.48)	0.999(1.40)	14.349(13.35)
		t(3)	1.022(0.12)	0.989(0.12)	1.017(0.13)	0.992(0.12)	0.999(0.12)	0.988(0.13)	1.069(0.11)
		Laplace	1.159(0.21)	0.990(0.16)	1.004(0.17)	0.997(0.16)	0.996(0.14)	0.979(0.16)	5.564(4.56)
		RN(0.05,2)	1.004(0.11)	0.989(0.12)	1.017(0.13)	0.993(0.11)	0.999(0.12)	0.989(0.12)	0.839(0.18)

Table 4. Robust estimates for the regression model with outliers in $y-$ direction ($\varepsilon_i \sim$ Laplace)

n	Outlier (%)	Estimators	$\hat{\beta}_0$ (RMSE)	$\hat{\beta}_1$ (RMSE)	$\hat{\beta}_2$ (RMSE)	$\hat{\beta}_3$ (RMSE)	$\hat{\beta}_4$ (RMSE)	$\hat{\beta}_5$ (RMSE)	$\hat{\sigma}$ (RMSE)
20	5	Normal	3.762(3.19)	0.891(2.93)	1.058(3.20)	1.217(2.82)	0.879(3.38)	0.876(3.11)	9.186(8.25)
		t(3)	1.022(0.26)	1.028(0.32)	1.004(0.32)	0.999(0.25)	1.039(0.39)	1.007(0.32)	0.703(0.33)
		Laplace	1.096(0.33)	1.059(0.42)	1.012(0.36)	0.994(0.33)	1.049(0.42)	1.022(0.38)	3.004(2.01)
		RN(0.05,2)	1.018(0.24)	1.021(0.30)	1.008(0.30)	0.992(0.25)	1.035(0.39)	1.018(0.29)	0.726(0.31)
	10	Normal	5.721(5.08)	0.578(3.81)	0.756(4.64)	0.930(4.09)	1.275(4.14)	1.564(4.53)	12.429(11.5)
		t(3)	0.964(0.33)	0.963(0.30)	0.938(0.38)	0.978(0.29)	1.023(0.31)	1.032(0.33)	0.896(0.22)
		Laplace	1.173(0.45)	0.912(0.45)	0.972(0.54)	0.948(0.51)	1.029(0.46)	1.017(0.49)	5.415(4.42)
		RN(0.05,2)	0.950(0.33)	0.974(0.29)	0.938(0.36)	0.989(0.28)	1.023(0.31)	1.020(0.32)	0.717(0.32)
50	5	Normal	3.997(3.07)	1.229(1.95)	0.809(1.71)	0.823(1.84)	0.584(1.64)	1.112(1.96)	11.083(10.09)
		t(3)	1.001(0.15)	1.009(0.18)	0.990(0.16)	1.001(0.17)	0.993(0.14)	0.985(0.16)	0.859(0.17)
		Laplace	1.089(0.21)	1.013(0.23)	0.996(0.19)	0.992(0.22)	0.988(0.16)	0.980(0.20)	3.599(2.60)
		RN(0.05,2)	0.996(0.14)	1.010(0.18)	0.990(0.15)	1.005(0.16)	0.994(0.14)	0.986(0.15)	0.815(0.20)
	10	Normal	5.881(4.93)	0.800(2.25)	0.957(2.30)	1.002(2.68)	0.596(2.33)	1.084(2.16)	13.936(12.95)
		t(3)	1.005(0.17)	0.979(0.18)	1.019(0.16)	1.015(0.16)	0.989(0.19)	1.001(0.19)	1.032(0.12)
		Laplace	1.151(0.26)	0.979(0.23)	1.016(0.23)	0.995(0.22)	0.985(0.23)	1.003(0.22)	5.544(4.54)
		RN(0.05,2)	0.992(0.16)	0.983(0.18)	1.017(0.16)	1.008(0.16)	0.991(0.18)	1.001(0.18)	0.815(0.20)
100	5	Normal	3.465(2.48)	0.874(1.11)	0.777(1.09)	0.897(1.12)	0.808(1.08)	0.848(1.18)	10.517(9.52)
		t(3)	0.990(0.11)	0.994(0.12)	0.988(0.11)	0.989(0.11)	1.022(0.12)	1.002(0.11)	0.874(0.14)
		Laplace	1.045(0.13)	0.990(0.14)	0.991(0.15)	0.994(0.14)	1.022(0.14)	0.992(0.14)	3.175(2.18)
		RN(0.05,2)	0.986(0.09)	0.995(0.11)	0.988(0.11)	0.987(0.10)	1.015(0.10)	1.002(0.10)	0.857(0.16)
	10	Normal	5.899(4.91)	0.902(1.29)	1.012(1.62)	1.137(1.59)	0.683(1.67)	0.952(1.66)	14.336(13.34)
		t(3)	1.019(0.12)	0.985(0.10)	1.006(0.11)	0.997(0.10)	1.007(0.11)	1.017(0.12)	1.044(0.09)
		Laplace	1.158(0.21)	0.990(0.13)	1.012(0.14)	1.010(0.15)	0.996(0.14)	1.030(0.15)	5.553(4.55)
		RN(0.05,2)	1.007(0.11)	0.986(0.10)	1.007(0.11)	0.993(0.09)	1.007(0.11)	1.013(0.11)	0.819(0.19)

Robustness properties of RN estimators become more evident when the errors follow RN distribution and data include outliers (Table 5). In all cases, RN estimates perform better than the others. However, the difference between the performances of RN and Student-t diminishes as the sample size increases. Laplace estimator follows those performances with larger RMSE values. Besides, all results indicate that least squares estimates behave badly when the errors are heavy-tailed and contaminated by outliers.

Table 5. Robust estimates for the regression model with outliers in $y-$ direction ($\varepsilon_i \sim RN(0.05, 2)$)

n	Outlier (%)	Estimators	$\hat{\beta}_0$ (RMSE)	$\hat{\beta}_1$ (RMSE)	$\hat{\beta}_2$ (RMSE)	$\hat{\beta}_3$ (RMSE)	$\hat{\beta}_4$ (RMSE)	$\hat{\beta}_5$ (RMSE)	$\hat{\sigma}$ (RMSE)
20	5	Normal	3.633(2.98)	0.991(2.71)	1.040(3.37)	0.849(2.87)	0.213(3.16)	0.919(3.19)	9.298(8.35)
		t(3)	1.057(0.36)	1.020(0.34)	1.035(0.35)	1.004(0.33)	1.007(0.29)	0.957(0.29)	0.735(0.31)
		Laplace	1.116(0.43)	1.026(0.42)	1.020(0.39)	0.991(0.40)	0.999(0.40)	0.957(0.37)	3.046(2.05)
		RN(0.05,2)	1.045(0.33)	1.012(0.31)	1.019(0.33)	1.009(0.30)	0.996(0.29)	0.945(0.28)	0.762(0.28)
	10	Normal	5.907(5.55)	0.901(4.20)	0.590(4.24)	0.877(4.46)	0.429(3.65)	1.176(4.15)	12.726(11.78)
		t(3)	0.988(0.32)	0.980(0.33)	0.991(0.35)	1.004(0.31)	0.991(0.35)	1.028(0.33)	0.904(0.24)
		Laplace	1.223(0.48)	0.988(0.44)	0.986(0.45)	1.014(0.41)	1.010(0.48)	1.018(0.40)	5.465(4.48)
		RN(0.05,2)	0.975(0.31)	0.988(0.30)	0.997(0.34)	1.001(0.30)	0.988(0.35)	1.031(0.32)	0.730(0.32)
50	5	Normal	3.842(2.93)	1.289(1.75)	0.883(1.87)	0.984(2.10)	0.821(1.84)	0.990(1.74)	11.039(10.05)
		t(3)	0.973(0.15)	1.012(0.16)	1.019(0.17)	0.975(0.19)	0.972(0.17)	1.009(0.17)	0.913(0.14)
		Laplace	1.053(0.20)	1.010(0.18)	1.023(0.22)	0.975(0.23)	0.964(0.22)	1.014(0.21)	3.649(2.65)
		RN(0.05,2)	0.966(0.14)	1.004(0.16)	1.017(0.16)	0.978(0.18)	0.970(0.16)	1.004(0.17)	0.872(0.16)
	10	Normal	5.823(4.88)	1.346(2.42)	0.815(2.39)	0.579(2.37)	0.549(2.18)	1.215(2.41)	13.916(12.93)
		t(3)	1.014(0.18)	1.014(0.18)	0.989(0.18)	1.032(0.19)	0.996(0.17)	0.992(0.16)	1.074(0.15)
		Laplace	1.153(0.25)	1.012(0.24)	0.986(0.24)	1.020(0.25)	0.971(0.23)	0.996(0.20)	5.562(4.56)
		RN(0.05,2)	1.002(0.18)	1.016(0.18)	0.989(0.17)	1.035(0.19)	0.999(0.17)	0.991(0.15)	0.850(0.18)
100	5	Normal	3.438(2.45)	0.898(1.22)	1.152(1.08)	0.977(1.03)	0.660(1.19)	1.009(1.23)	10.458(9.46)
		t(3)	0.995(0.12)	0.996(0.12)	1.022(0.12)	0.997(0.11)	1.019(0.15)	0.996(0.12)	0.931(0.10)
		Laplace	1.058(0.15)	1.004(0.15)	1.027(0.14)	1.001(0.14)	1.008(0.17)	0.993(0.14)	3.215(2.22)
		RN(0.05,2)	0.991(0.12)	0.993(0.12)	1.019(0.11)	0.999(0.10)	1.018(0.14)	0.994(0.12)	0.919(0.11)
	10	Normal	5.915(4.93)	0.548(1.61)	0.949(1.73)	0.678(1.56)	0.789(1.54)	0.879(1.60)	14.379(13.38)
		t(3)	1.035(0.12)	1.005(0.13)	0.982(0.12)	1.009(0.12)	0.979(0.12)	1.016(0.11)	1.117(0.15)
		Laplace	1.179(0.23)	0.996(0.17)	0.980(0.15)	1.000(0.14)	0.993(0.16)	1.004(0.15)	5.622(4.62)
		RN(0.05,2)	1.019(0.12)	1.004(0.13)	0.982(0.11)	1.010(0.12)	0.977(0.11)	1.017(0.10)	0.882(0.14)

6.2. Choosing Optimal Tuning Parameters

RN distribution involves two robustness tuning parameters which control the degree of boundedness of the IF. So far we have assumed that the values of these parameters are known and set the values as a =0.05 and b =2 the most preferable in the literature. It is now of our interest for real world applications to determine the optimal tuning parameters via a data-driven method. We here preferred the k-fold cross-validation (CV) method that has been widely used for the evaluation of model accuracies. In k-fold CV procedure, the original data set is randomly divided into "k" equal size subsets (or folds). One of the "k" subsets is used as validation data to test the model constructed by the remaining "k-1" subset known as training data [42-45]. For choosing optimal tuning parameters, it is necessary to impose boundedness condition on the IF to achieve a comparable trade-off. We here followed the statement of [11] "*The robustness parameters (a, b) should be chosen so that (a) the robust density has the same shape in the central region as the target density and (b) the influence function (3) attains its bounds at approximately the point at which the target density (1) begins to assign negligible probability to y. These bounds are attained at* $\pm$ $(ab)^{-1/b}$. *In the case of a Gaussian target density, the choices (.3, 1.0) and (.05, 2.0) position the bounds at about three scale units from the location*". Therefore, we started with the value of 0.3 for the parameter a when b =1 and decreased its value with a grid of 0.1.

Similarly, we began with a =0.05 when b =2 and decreased its value with a grid of 0.01. Note that the optimality search within those values put the boundedness condition at the values of "3 or more" as the standard RN distribution scale is one. In order to show an empirical justification, we again simulated a model with 5 regressors using the same settings in section 6.1 except for the sample size taken as 1000. Pure (no outlier) and contaminated models (5 or 10 outliers created by adding a constant, C, to $y-$ direction) were tested. 10-fold CV is performed for the models with the trial values of tuning parameters above mentioned. Prediction values of compared models ($\hat{y}_i^{-k}$) are obtained and the residuals are evaluated as

$$e_{i,k} = (y_i - \hat{y}_i^{-k}) \quad , \quad i \in \text{ validation set}$$

Then the mean square error (MSE) is calculated as

$$MSE = \frac{\sum_{k=1}^{10} \sum_{e_{i,k}>3} (e_{i,k})^2}{(\# \, of \;\; e_{i,k} > 3)}$$

To achieve a more precise estimate of MSE, this 10-fold cross validation process was repeated "100" times and the averages of the MSEs were calculated and the model with the smallest estimate is chosen as the "optimal" model. The results for the pure model are presented in Table 6. It appears that CV method achieved to discover the optimal values of tuning parameters which are a = 0.05 and b =2 as we set. However, inclusion of outliers affects the issue of selecting "optimum" tuning parameters (see Table 7). In case of moderate outliers (C=10), the optimal model still appears as RN(0.05, 2). The optimum for increasing values of outliers (C=25, 50 and 100) is highlighted as RN(0.3,1) with however very close MSE values to those of RN(0.05,2).

Table 6. The average MSE values for the model without outliers (n =1000)

RN distribution	Average MSE
RN(0,0)	12.4099
RN(0.1,1)	12.4095
RN(0.2,1)	12.4093
RN(0.3,1)	12.4094
RN(0.01,2)	12.4096
RN(0.02,2)	12.4094
RN(0.03,2)	12.4091
RN(0.04,2)	124090
RN(0.05,2)	**12.4088**

Table 7. The average MSE values for the model with outliers achieved by adding the constant, C = 10, 25, 50, 100; (n =1000)

RN distribution	Number of outliers							
	5				10			
	C constant							
	10	25	50	100	10	25	50	100
RN(0,0)	12.3852	12.4617	12.4704	12.4347	12.3959	12.4067	12.4858	13.2494
RN(0.1,1)	12.3393	12.1106	10.9064	7.4676	12.2863	11.5102	8.7284	4.0858
RN(0.2,1)	12.3155	12.0591	10.8933	7.4635	12.2208	11.3043	8.6826	4.0845
RN(0.3,1)	**12.3069**	**12.0547**	**10.8930**	**7.4593**	**12.1955**	**11.2985**	**8.6814**	**4.0834**
RN(0.01,2)	12.3405	12.0626	10.8938	7.4696	12.2977	11.3042	8.6843	4.0862
RN(0.02,2)	12.3167	12.0599	10.8936	7.4683	12.2285	11.2997	8.6839	4.0859
RN(0.03,2)	12.3074	12.0586	10.8935	7.4670	12.2002	11.2993	8.6837	4.0857
RN(0.04,2)	12.3042	12.0574	10.8934	7.4657	12.1908	11.2989	8.6832	4.0855
RN(0.05,2)	**12.3030**	**12.0560**	**10.8933**	**7.4643**	**12.1876**	**11.2987**	**8.6830**	**4.0854**

7. REAL DATA APPLICATION

The dataset, famously known as "Brownlee's stack loss plant data" in the literature, here is chosen to illustrate the performance of RN distribution comparatively with Normal, Student-t and Laplace distributions. This data set has been first introduced as an example to apply procedures of multiple regression using the least squares method by [46] and since then it has been used in, at least, 90 distinct papers and books for various applications of linear modelling [47]. Four observations were detected as outliers by many authors [48-54], which caused it to be entitled as "a real data set with four outliers". Estimated regression models by different robust methods as well as OLS are listed in Table 8 with corresponding references.

Table 8. Literature review for the methods of regression modelling for stack loss data

Method	Estimated Regression Models
Ordinary Least Square (OLS)	$\hat{y} = -39.9 + 0.716x_1 + 1.295x_2 - 0.152x_3$ ([46])
Huber's M-estimate with outliers	$\hat{y} = -41.0 + 0.83x_1 + 0.91x_2 - 0.13x_3$ ([49]) $\hat{y} = -39.21 + 0.8297x_1 + 0.7512x_2 - 0.1088x_3$ ([55]) $\hat{y} = -41.19 + 0.811x_1 + 1.010x_2 - 0.133x_3$ ([53]) $\hat{y} = -39.33 + 0.8288x_1 + 0.7590x_2 - 0.1087x_3$ ([56]) $\hat{y} = -41.17 + 0.8133x_1 + 1.000x_2 - 0.1324x_3$ ([52])
Huber's M-estimate without outliers	$\hat{y} = -37.35 + 0.8301x_1 + 0.4918x_2 - 0.0711x_3$ ([55])
Least Median of Squares (LMS) with outliers	$\hat{y} = -39.25 + 0.75x_1 - 0.50x_2 + 0.0x_3$ ([57])
LMS without outliers	$\hat{y} = -35.9 + 0.82x_1 + 0.43x_2 - 0.07x_3$ ([58])
Tukey's Biweight	$\hat{y} = -40.56 + 0.7668x_1 + 1.129x_2 - 0.1392x_3$ ([55])
L-Estimate	$\hat{y} = -40.37 + 0.72x_1 + 0.96x_2 - 0.07x_3$ ([49]) $\hat{y} = -40.79 + 0.851x_1 + 0.869x_2 - 0.129x_3$ ([59])

y : Stack Loss , x_1 : Air Flow , x_2 : Cooling Water Inlet Temperature , x_3 : Acid Concentration

This study was first of interest to investigate the effect of considered heavy-tailed error distributions by excluding all outliers from the data. Then we repeated the whole analysis on the data with four outliers one of which is in $x-$ and three of which are in $y-$ directions. For choosing optimal tuning parameters of RN distribution, we applied 10-fold CV method to stack loss data set. This method suggested the RN (0.05.2) model with the smallest MSE.

For the sake of consistency with our simulation study, we also created an artificial case by altering the outliers of *just* $y-$ direction with a constant value of 50 so as to put more emphasis on the influence of $y-$ direction outliers. CV method in this case again suggested RN(0.3, 1) as an alternative to RN(0.05,2).

All the models fitted produced parameter estimates as presented in Table 9. Modelling the data without outliers produced parameter estimates almost the same under both error distributions, while information criteria implied a slightly better fit of Normal than RN model to the data. Inclusion of four original outliers however had lowering effect on the performances of all models in terms of AICs. The least affected results belong to the Laplace estimators, having the lowest AIC. Model including Student-t estimates follow this performance. This inconsistency with the simulation results might be caused by the outlier existence which is in both $x-$ and $y-$ direction. Enlarging the outlier existence of $y-$ direction in some amount as described above, however, highlighted the robustness of both RN and Student-t estimators. It must be noted that resistance to outliers achieved by RN is now more evident when the influence of the outlying observations becomes more immense.

Table 9. Estimated model parameters for the stack loss data

Outliers	Error distributions	Regression Coefficients				Scale	Akaike Information Criterion (AIC)	
		$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\sigma}$	– ln L	AIC
without outliers	Normal	– 37.652	0.798	0.577	– 0.067	1.095	25.67	61.34
	Student-t(3)	– 37.429	0.833	0.472	– 0.067	0.836	26.16	62.32
	Laplace	– 35.941	0.822	0.438	– 0.070	0.829	30.66	71.33
	RN(0.05, 2)	– 37.384	0.808	0.546	– 0.069	1.008	25.75	61.50
with original outliers	Normal	– 39.920	0.716	1.295	– 0.152	2.912	52.29	114.58
	Student-t(3)	– 39.124	0.854	0.657	– 0.104	1.755	51.07	112.14
	Laplace	– 39.689	0.832	0.574	– 0.061	2.004	50.15	110.31
	RN(0.05, 2)	– 40.554	0.769	1.119	– 0.138	2.612	52.18	114.36
with altered outliers	Normal	– 80.259	1.354	3.283	– 0.535	16.459	88.61	187.23
	Student-t(3)	– 36.804	0.770	0.659	– 0.078	2.013	76.11	162.22
	Laplace	– 39.689	0.832	0.574	– 0.061	9.147	82.04	174.08
	RN(0.05, 2)	– 37.265	0.807	0.547	– 0.070	0.911	52.18	114.36
	RN(0.3, 1)	– 40.319	0.825	0.892	– 0.125	1.973	53.50	117.01

8. CONCLUSION

Robust methods of estimation in regression modelling have been a fundamental issue because linear least squares estimates behave badly, particularly when the errors are heavy-tailed than normal. Some contributions for the protection against such ill effects hypothesize that the inference drawn by means of heavy-tailed error distributions would be resistant to outliers. Robust modelling can be therefore based on measurement distributions having fatter tails than the normal distribution.

In this study we concentrated on robustness with respect to sampling specifications in the way [11] proposed, that is to look at the rate of change of the sampling model density with respect to an observation value. For a robust analysis, the observation's influence should reach a peak and then begin to decay as the observation becomes more outlying. It is well known that such influence for normal likelihood is unbounded. [11] provided a means of accommodating outliers by converting the usual form of a non-robust density such as Normal to a robust likelihood by modifying the density's unbounded influence function. The resulting distributional forms created a family of RN distribution having additional parameters that control the tail thickness. We here illustrated all the details of this procedure by drawing parallels with its pioneer Bayesian evaluation and also with the major classical robust approach of Huber. We showed how to derive robust estimators of a regression model quantities using RN distributed errors. Random data generation of this distribution through random-walk Metropolis algorithm is here also suggested.

Simulation study indicated that RN produce the same parameter estimates as OLS at the expense of slightly lower efficiency when the true error distribution is Normal. Allowing heavy-tailed errors via Student-t or Laplace, correct values of model quantities were best recovered by Student-t estimators, however RN estimator competes with those, especially when the sample size gets larger. When the tail tickness was achieved by RN distribution, it was observed that RN estimates were much more accurate than the others. In all cases of heavy-tailed errors, OLS produced estimates with larger RMSEs. When the model includes more regressors, results do not change except for the errors of estimates enlarged to some extent.

Finite sample performance assesment also indicated that RN estimators are robust to the outliers of y-direction embedded in the Normal or RN distributed errors. However, Student-t estimators' robustness appears better than the others, especially when the true error distribution Student-t or Laplace again includes outliers. It must be noted that RN estimates' performance closely follow those in both cases of no matter what the sample sizes are. Determination of "optimal" tuning parameters was also performed via the cross-validation method. Empirical results suggested that the optimum value of robustness tuning parameter could be achieved for pure model (without outliers) or a contaminated model with moderate outliers. When b =1, a =0.3 appears as alternative best values for the model with RN(0.05, 2) in the cases of more influential outliers.

In the results of real world application, the estimators of the modified distribution i.e. RN appeared to behave almost the same as OLS estimators, which were badly influenced by the outliers appeared in both $x-$ and $y-$ direction. For this case, Laplace distribution presented better performance in modeling fit. Considering the empirical results that RN estimators are robust to the outliers of $y-$ direction, we highlighted the robustness of RN by enlarging the outliers influence in that direction. In this case, RN appeared to be superior in tolerating more immense outliers for even small samples. This result is consistent with the study of choosing optimal tuning parameters where MSE's of the model get smaller as the amount of the contamination (C) gets larger.

In comparison with the selected heavy-tailed distributions (Student-t and Laplace), the robustness of RN becomes better when the influence of outlying observations appears more extensive. Although RN distribution serve as a robust alternative not only to unbounded normal model but also to the heavy-tailed Student-t distribution, there is a price to be paid for the utilization of this procedure, which is analytical: it involves derivatives that are difficult to compute apart from a particular family of distributions. Due to this limitation, the procedure suggested by [11] has been largely avoided in the literature. A generalization of the procedure for the distributions other than normal could however be developed with the advanced analytical tools available at the moment, and suggested here as a future work.

APPENDIX

Here we show how the modification process is applied on Gaussian distribution so as to evaluate the probability density function of a RN distribution. Suppose that a random variable X follows $N(\mu,\sigma)$ with the non-robust probability density function defined as in the form of Eq. (1);

$$f(x\mid v)=r(x)\,s(v)\,exp\left[-\frac{1}{2}d^2(v,x)\right] \tag{18}$$

where $d(v,x)=\left|\frac{x-\mu}{\sigma}\right|$. Modified influence function of X as in Eq. (3) can be expressed as

$$-\frac{d\,log\,f(x\mid v)}{dx} = d(v,x)\frac{d(d(v,x))}{dx}exp\left(-a\left(d(v,x)^b\right)\right)-\frac{d\,log\,r(x)}{dx}$$

$$\frac{d\,log\,f(x\mid v)}{dx} = -\left|\frac{x-\mu}{\sigma}\right|\frac{d\left(\left|\frac{x-\mu}{\sigma}\right|\right)}{dx}exp\left[-a\left|\frac{x-\mu}{\sigma}\right|^b\right]$$

$$\frac{d\,log\,f(x\mid v)}{dx}=-\left|\frac{x-\mu}{\sigma}\right|\frac{1}{\sigma}sgn\left(\frac{x-\mu}{\sigma}\right)exp\left[-a\left|\frac{x-\mu}{\sigma}\right|^b\right] \tag{19}$$

Using the relation ($|x|sgn(x) = x$) and integrating the both sides gives

$$log\,f(x\mid v)=-\int_{-\infty}^{\infty}\frac{x-\mu}{\sigma^2}exp\left(-a\left|\frac{x-\mu}{\sigma}\right|^b\right)$$

$$f(x\mid v)=exp\left(-\frac{1}{\sigma^2}\int_{-\infty}^{\infty}(x-\mu)exp\left(-a\left|\frac{x-\mu}{\sigma}\right|^b\right)dx\right) \tag{20}$$

This is the robust version of the density in Eq. (18). The integral in the exponential function here can be defined as

$$\int_{-\infty}^{\infty}(x-\mu)exp\left(-a\left|\frac{x-\mu}{\sigma}\right|^b\right)dx = \int_{u=0}^{a\left|\frac{x-\mu}{\sigma}\right|^b}-\left(\frac{u\sigma^b}{a}\right)^{1/b}exp(-u)\left(-\frac{1}{b}\right)\left(\frac{u\sigma^b}{a}\right)^{\frac{1}{b}-1}\frac{\sigma^b}{a}du$$

$$= \frac{1}{b}\,\frac{\sigma^2}{a^{2/b}}\int_{u=0}^{a\left|\frac{x-\mu}{\sigma}\right|^b}u^{\frac{2}{b}-1}\,exp(-u)\,du$$

where $u = a\left|\frac{x-\mu}{\sigma}\right|^b$.

Inserting this result to Eq. (20) gives the density as

$$f\left(x \mid v\right) = exp\left(-\frac{1}{\sigma^2}\left[\frac{1}{b}\frac{\sigma^2}{a^{2/b}}\gamma\left(\frac{2}{b},a\left|\frac{x-\mu}{\sigma}\right|^b\right)\right]\right)$$

$$= exp\left(-\left[\left(ba^{2/b}\right)^{-1}\gamma\left(\frac{2}{b},a\left|\frac{x-\mu}{\sigma}\right|^b\right)\right]\right) \quad (21)$$

where γ is the lower incomplete gamma function.

ACKNOWLEDGEMENTS

The authors thank the authors which listed in References for many useful support.

CONFLICT OF INTEREST

The authors stated that there are no conflicts of interest regarding the publication of this article.

REFERENCES

[1] Lange KL, Little RJA, Taylor JMG. Robust statistical modelling using the t distribution. J Americ Statist Assoc 1989; 84: 881-896.

[2] Shi J, Chen K, Song W. Robust errors-in-variables linear regression via Laplace distribution. Statist Prob Lett 2013; 84: 113-120.

[3] Lange K, Sinsheimer JS. Normal/Independent Distributions and Their Applications in Robust Regression. J Comput Graph Statist 1993; 2: 175-198.

[4] Albert J, Delampady M, Polasek W. A class of distributions for robustness studies. J Statist. Plann Inf 1991; 28: 291-304.

[5] Butler RJ, McDonald JB, Nelson RD, White SB. Robust and partially adaptive estimation of regression models. Rev. Econom. Statist. 1990; 72: 321-327.

[6] Taylor J, Verbyla A. Joint modelling of location and scale parameters of the t distribution. Stat. Model 2004; 4: 91-112.

[7] Bolfarine H, Arellano-Valle RB. Robust modelling in measurement error models using the t-distribution. REBRAPE 1994; 8: 67-84.

[8] Tiku ML, Islam MQ, Selcuk AS. Non-normal regression, II: Symmetric distributions. Comm Stat Theor Meth 2001; 30: 1021-1045.

[9] Alcantara IC, Cysneiros FJA. Linear regression models with slash-elliptical errors. Comput. Stat. Data Anal. 2013; 64: 153-164.

[10] Smith VK, Hall TH. A comparison of maximum likelihood versus BLUE estimators. Rev. Econ. Stat. 1972; 54: 186-190.

[11] Ramsay JO, Novick MR. PLU robust Bayesian decision theory: point estimation. J Americ Statist Assoc 1980; 75: 901-907.

[12] West M. Outlier Models and Prior Distributions in Bayesian Linear Regression. J Royal Statist Soc 1984; 46: 431-439.

[13] Chen Y, Fournier D. Impacts of atypical data on Bayesian inference and robust Bayesian approach in fisheries. Can J Fish Aquat Sci 1999; 56: 1525-1533.

[14] Passarin K. Robust Bayesian Estimation. Varese: Italy, 2004.

[15] Kaya M, Çankaya E, Arslan O. Robust bayesian regression analysis using Ramsay-Novick distributed errors with Student-t prior. Commun Fac Sci Univ Ank Ser A1 Math Stat 2019; 68: 602-618.

[16] Draper NR, Smith H. Applied Regression Analysis. J Wiley New York 1996; 337: 204-215.

[17] Daniel C, Wood FS. Fitting Equations to Data. 2nd ed. New York, Wiley, 1971.

[18] Andrews DF. A robust method for multiple linear regression. Technometrics 1974; 16: 523-531.

[19] Denby L, Mallows CL. Two diagnostic displays for robust regression analysis. Technometrics 1977; 19: 1-13.

[20] Andrews DF, Pregibon D. Finding the outliers that matter. J Roy Statist Soc B 1978; 40: 85-93.

[21] Cook RD. Influential observations in linear regression. J Americ Statist Assoc 1979; 74: 169-174.

[22] Dempster AP, Gasko-Green M. New tools for residuals analysis. Ann Statist. 1981; 9: 945-959.

[23] Atkinson AC. Regression diagnostics, transformations and constructed variables. J Roy Statist Soc B 1982a; 44: 1-36.

[24] Atkinson AC. Robust and diagnostic regression analyses. Comm Stat Theor Meth 1982b; 11:2559-2571.

[25] Rey WJJ. Introduction to Robust and Quasi-Robust Statistical Methods. Springer-Verlag Berlin, 1983.

[26] Carroll RJ, Ruppert D. Transformations in regression: A robust analysis. Technometrics 1985; 27: 1-12.

[27] Li G. Robust regression. In: David C, Hoaglin FM, John WT, editors. Exploring Data Tables, Trends, and Shapes. New York: J Wiley, 1985, pp. 281-343.

[28] Preece DA. Illustrative examples: illustrative of what?. Stat 1986; 33:35.

[29] R Development Core Team R. A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2013.

[30] Rao CR. Unified theory of least squares. Comm in Statist 1973; 1:1-8.

[31] Zellner A. Bayesian and non-Bayesian analysis on the regression model with multivariate Student-t error terms. J Americ Statist Assoc, 1976; 71:400-405.

[32] Amemiya T. Advanced econometrics. Cambridge, Massachusetts, 1985.

[33] Geer S. Least squares estimators. Encyclop Statististics in the Behaviour Science Wiley, 2005.

[34] Rousseeuw PJ, Leroy AM. Robust Regression & Outlier Detection. J Wiley New York 1987; 76-78.

[35] Huber PJ. Robust statistics. J Wiley New York, 1981.

[36] Maronna RA, Yohai VJ. Asymptotic behavior of M-estimates for the linear model. Ann Statist 1979; 7: 258-268.

[37] Bloomfield P, Staiger W. Least Absolute Deviations Theory-Applications and Algorithms. Birkhauser Boston Mass, 1983.

[38] Yohai VJ. High breakdown point and high efficiency robust estimates for regression Univ. Washington, 1985.

[39] Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. Equation of state calculations by fast computing machines. J Chem Phys 1953; 21: 1087-1092.

[40] Hastings WK. Monte Carlo sampling methods using Markov chains and their applications. Biometrika 1970; 57: 97-109.

[41] Roberts GO, Rosenthal JS. Optimal Scaling for Various Metropolis-Hastings Algorithms. Statist Scienc 2001; 16: 351–367.

[42] Kohavi R. A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. Appears in the International Joint Conference on Artificial Intelligence(IJCAI), 1995.

[43] Hastie T, Tibshirani R, Friedman J. The Elements of Statistical Learning. New York, NY, USA: 1.10. Springer series in statistics, 2001.

[44] Arlot S, Celisse A. A survey of cross-validation procedures for model selection. Stat. Surv. 2010; 4: 40-79.

[45] James G, Witten D, Hastie T, Tibshirani R. An Introduction to Statistical Learning. Springer, 2013.

[46] Brownlee KA. Statistical Theory and Methodology in Science and Engineering. J Wiley New York, 1960. pp. 491-500.

[47] Dodge Y. The guinea pig of multiple regression. Robust Statistics, Data Analysis, and Computer Intensive Methods (Lecture Notes in Statistics) Springer-Verlag New York 1996; 109: 91-118.

[48] Aitkin M, Wilson GT. Mixture models, outliers, and the EM algorithm. Technometrics 1980; 22: 325-331.

[49] Ruppert D, Carroll RJ. Trimmed least squares estimation in the linear model. J Americ Statist Assoc, 1980; 75: 828-838.

[50] Antoch J, Jureckova J. Trimmed least squares estimator resistant to leverage points. CSQ 1985; 4:329-339.

[51] Rey WJJ. M-Estimators in Robust Regression, a Case Study. M.B.L.E. Researc Lab Brussels, 1977.

[52] Birkes D, Dodge Y. Alternative Methods of Regression. J Wiley New York 1993; 177-179.

[53] Osborne MR. Finite Algorithms in Optimization and Data Analysis. J Wiley New York, 1985; 267-270.

[54] Hettmansperger TP, McKean JW. A robust alternative based on ranks to least squares in analyzing linear models. Technometrics 1977; 19: 275-284.

[55] Chambers RL, Heathcote CR. On the estimation of slope and the identification of outliers in linear regression. Biometrika 1981; 68: 21-33.

[56] Wiens DP. A note of the computation of robust, bounded influence estimates and test statistics in regression. Comput Statist Data Analy 1992; 13: 211-220.

[57] Tichavsky P. Algorithms for and geometrical characterization of solutions in the LMS and the LTS linear regression. CSQ 1991; 2: 139-151.

[58] Moberg TF, Ramberg JS, Randles RH. An adaptative multiple regression procedure based on Mestimators. Technometrics 1980; 22: 213-224.

[59] Welsh AH. The trimmed mean in the linear model. Ann Statist 1987; 15: 20-36.