Available online: July 05, 2018

Commun. Fac. Sci. Univ. Ank. Ser. A1 Math. Stat.
Volume 68, Number 1, Pages 602–618 (2019)
DOI: 10.31801/cfsuasmas.441096
ISSN 1303–5991 E-ISSN 2618-6470

http://communications.science.ankara.edu.tr/index.php?series=A1

ROBUST BAYESIAN REGRESSION ANALYSIS USING RAMSAY-NOVICK DISTRIBUTED ERRORS WITH STUDENT-T PRIOR

MUTLU KAYA, EMEL ÇANKAYA, AND OLCAY ARSLAN

Abstract. This paper investigates bayesian treatment of regression modelling with Ramsay - Novick (RN) distribution specifically developed for robust inferential procedures. It falls into the category of the so-called heavy-tailed distributions generally accepted as outlier resistant densities. RN is obtained by coverting the usual form of a non-robust density to a robust likelihood through the modification of its unbounded influence function. The resulting distributional form is quite complicated which is the reason for its limited applications in bayesian analyses of real problems. With the help of innovative Markov Chain Monte Carlo (MCMC) methods and softwares currently available, here we first suggested a random number generator for RN distribution. Then, we developed a robust bayesian modelling with RN distributed errors and Student-t prior. The prior with heavy-tailed properties is here chosen to provide a built-in protection against the misspecification of conflicting expert knowledge (i.e. prior robustness). This is particularly useful to avoid accusations of too much subjective bias in the prior specification. A simulation study conducted for performance assessment and a real-data application on the famously known "stack loss" data demonstrated that robust bayesian estimates with RN likelihood and heavy-tailed prior are robust against outliers in all directions and inaccurately specified priors.

1. Introduction

Development of robust estimation procedures has been largely devoted to non-Bayesian estimation framework due to the opinion about Bayesian approaches as being inherently robust because they accommodate uncertainties within a joint posterior probability distribution. However, every good Bayesian practice involves the study of sensitivity of posterior distribution to the major ingredients of Bayesian

Received by the editors: December 27, 2017; Accepted: March 03, 2018.
2010 *Mathematics Subject Classification.* Primary 05C38, 15A15; Secondary 05A15, 15A18.
Key words and phrases. Robust bayesian regression, Ramsay-Novick, heavy-tailed distribution, Student-t prior, prior robustness.

analysis, which are typically the sampling model (or likelihood) and the prior specifications. The traditional Bayesian regression modelling based on normal errors with conjugate structures tries to resolve ill effects resulted for departures from normality due to outliers, or conflictions between the likelihood and the prior, by centering the posterior at some position of large aggreement. However, this might not be reasonable solution when such effects becomes increasingly extreme. Bayesian modelling with heavy-tailed distributions has been suggested as more effective way of conflict resolution, typically by favouring one source of information consistent with the majority over the other conflicting with the rest ([29]). Using alternative error distributions with thicker tails than Normal have revealed a variety of robust models in the literature of both classical and Bayesian analyses (see [12]; [33]; [24]; [41]; [3]; [25]; [38] etc.). Although, heavy-tailed distributions are not restricted to the class of t distributions, many Bayesian analysis of real problems employed Student-t as a natural choice for the reason that the tail thickness can be controlled by suitably chosen degrees of freedom. Besides, its scale mixture of Normal representation provides computational ease for the evaluation of posteriors.

To achieve Bayesian robustness, [33] proposed a procedure fully different in style. They first measured the influence of a single observation on a function which shows the rate of change of the sampling model density with respect to the observation. This is in fact a function of a specific quantity, as they named the influence function of the sampling model. They applied a modification on the unbounded influence function of a non-robust density within a certain symmetric family of distributions so that it would be bounded. By deriving the modified influence function backwards, a new family of distributions, namely Ramsay-Novick (RN) distribution, with robustness properties was obtained ([33]). Following this process, they also examined the concept of robustness under three essential ingredients (prior, likelihood and utility function) of a point estimate from a Bayesian perspective. Since then, [41] implemented a similar idea on Bayesian regression and [15] mentioned the limitations of this new robust family. [31] explained how to obtain bayesian estimates that are robust to outliers and based their comparative study on the same real-world data used in the work of [33]. For the present study, RN distribution within the class of heavy-tailed distributions is thus considered to be an intriguing choice for random errors to capture departures from the usual assumption of normality.

Prior robustness comes into consideration when it is desired to receive information from different sources for the model parameters. Misspecification of priors for parameters of some events may cause the prior to be in confliction with (far from) the reliable data, influencing the posterior. Whether a prior is robust or not depends on the rate at which the influence of the prior decreases. The influence of such priors could also be bounded by the choice of either a heavy-tailed prior density as well as non-informative, flat or reference priors which are naturally robust. Student-t distribution as a prior is chosen here to built-in protection against accusations of too much subjective bias in probability judgements of the available prior information.

The main goal of this paper is therefore to propose theoretical evaluation of robust bayesian estimators of a linear regression model with RN distributed errors and Student-t prior. The price to be paid for utilization of such inherently robust procedure is computational: analytically intractable form of RN distribution causes the posterior to be too complex, which is the reason for the avoidance of practising with this distribution family. Simulation-based Markov Chain Monte Carlo (MCMC) algorithms are here used to obtain the realizations from posterior functionals and a random sample generator for RN distribution was developed for the first time.

The rest of the article is structured as follows. In Section 2, we first describe the properties of RN distribution. We build the framework of robust analysis of Bayesian regression model with RN likelihood and Student-t prior in Section 3. In Section 4, we simplify complicated forms of full conditional posterior densities via a series expansion and employ MCMC sampling method for drawing samples from those. Since an approximate posterior distribution is used as a proposal density, we utilize the Metropolis-Hastings-within-Gibbs (MHWG) algorithm ([20]) to correctly estimate the true target posterior density. Section 5 and 6 present a simulation study and real-data application for the performance comparison, respectively. Finally, some conclusions are drawn and presented in Section 7.

2. Ramsay-Novick (RN) Distribution

A new family of distribution having bounded influence functions was proposed by [33].

$$f(x \mid v, a, b) = r(x)\ A(v)\ s(v)\ \exp\left(-\ \eta_{ab}\left(d(v,x)\right)\right) \quad , \quad a>0, b>0 \tag{2.1}$$

where $\eta_{ab}\left(d(v,x)\right) = \left(ba^{2/b}\right)^{-1}\gamma\left(2/b, ad\left(v,x\right)^{b}\right)$, $\gamma\left(.\right)$ is the lower incomplete gamma function, v is the location parameter, $A\left(v\right)$ is a normalizing constant that does not depend on x and $d\left(v,x\right)$ is a measure of distance of x from v. The constants a and b are the robustness tuning constants ([33]). The normal distribution is obtained for $a \to 0$. Therefore, small values of this parameter are usually considered. RN distribution belongs to the elliptical family with density generator ([31]).

$$g(u) = \exp\left\{-\left(ba^{2/b}\right)^{-1}\int_{0}^{a(2u)^{b/2}} \exp\left(-t\right)t^{2/b-1}dt\right\}$$

If the measure of distance of x from the location $v=\mu$ scaled by σ is set as $d\left(v,x\right) = \left|\frac{x-\mu}{\sigma}\right|$, probability density function (p.d.f) of RN distribution in (2.1) can

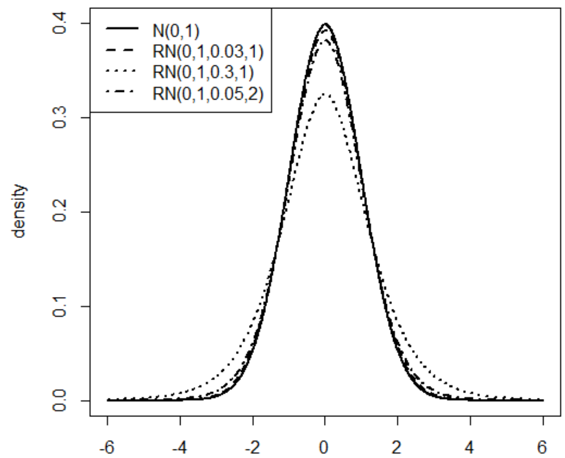

FIGURE 1. Influence of the robustness parameter on the tails of RN density function comparatively with Standard Normal

be obtained as follows

$$f(x \mid \mu, \sigma, a, b) \propto \exp\left(-\left(ba^{2/b}\right)^{-1} \gamma\left(\frac{2}{b}, a\left|\frac{x-\mu}{\sigma}\right|^{b}\right)\right) \tag{2.2}$$

and expressed as $X \sim RN(\mu, \sigma, a, b)$. The modified influence function of this variable becomes

$$MIF(x) = \frac{(x-\mu)}{\sigma^2} \exp\left(-a\left|\frac{x-\mu}{\sigma}\right|^{b}\right) \tag{2.3}$$

Figure 1 exemplifies the distributional forms of RN with differing values of robustness parameters which influence the tail thickness of the standard RN(0, 1, a, b). Figure 2 displays the impact of modification expressed in (2.3) on the unbounded influence function of non-robust Normal density. Influence of single observation (x) for Normal case ($\mu = 0, \sigma = 1$) can be seen here as a straight line passing through the origin. However, such an influence becomes bounded on both sides of extreme observations for RN cases. Reminding that this density has two robustness parameters, the speed of influence function approaching to zero seems to be controlled by parameter b. Note that the particular values of the parameters used here are the same preferences of Ramsay and Novick in their work ([33]). It can be seen from Figure 2, bounding effect of RN$(0, 1, 0.05, 2)$ is more significant than the others.

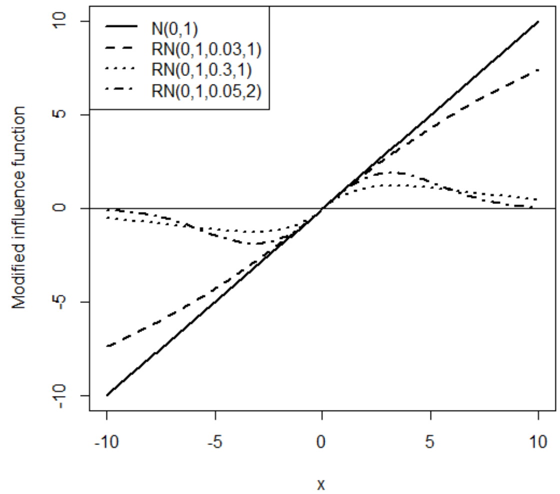

FIGURE 2. Modified influence function plots based on $d(v,x) = \left|\frac{x-\mu}{\sigma}\right|$

3. Robust Bayesian Regression Model

We here consider Bayesian analysis of a multiple linear regression model of the form :

$$y_i = x_i^T \beta + \varepsilon_i \quad , \quad i = 1, 2, ..., n \tag{3.1}$$

where $y_i \in R$ is the response variable, $x_i^T = (x_{i1}, x_{i2}, ..., x_{ip})$ is the $p-$dimensional regression predictor, $\beta = \left(\beta_0, \beta_1, ..., \beta_p\right)^T$ is the vector of unknown regression coefficients and ε_i's are the independently identically distributed (iid) random error terms. Bayes rule requires specification of prior density for the model parameters and the sampling distribution to obtain the joint posterior probability distribution factorized as

$$\text{Posterior} \qquad \propto \qquad \text{Prior} \qquad \times \qquad \text{Likelihood}$$

$$P(\beta_0, \beta_1, ..., \beta_p, \sigma^2 \mid y, x) \propto P(\beta_0, \beta_1, ..., \beta_p, \sigma^2) \times P(y \mid x, \beta_0, \beta_1, ..., \beta_p, \sigma^2) \tag{3.2}$$

In Bayesian regression modelling, the usual preferrence for these is Normal-Normal under the assumption of the Normal random errors. When the validity of this assumption is in doubt, likelihood component of Bayesian analysis needs to be suitable chosen. It is well known that the main source of deviations from the usually assumed Normal model is existence of outliers in the data. One way to achieve robust inferences is to utilize a unimodal heavy-tailed distribution. Student-t, Laplace, Slash and Exponential Power distributions are some examples of heavy-tailed distributions amongst many others and widely applied within robust regression framework in the literature ([14], [26], [17], [27], [39]).

Major criticisim of a Bayesian approach occur if the probabilistic statements for the model parameters were priori obtained by pooling information from multiple or dissimilar studies or sources (subjective beliefs). In such cases, multiple experts judgements may conflict, which can hardly be represented by a unique prior. Prior elicitation of this form may also reveal discrepancies with the sampling information, causing the posterior summaries to be highly affected. Thus, prior robustness has been developed mainly to cope with influences of inaccurately specified priors as a result of conflicting information sources. This issue was throughly discussed by [8] , [9] and [10], in which the prior robustness was based on the choice of classes of priors based on $\varepsilon-$contaminations. Heavy-tailed distribution family also appears as an alternative class of robust priors which can downweight the influence of expert's opinions conflicting with the majority, in other words, occuring in the tails.

Although RN distribution appears within the class of heavy-tailed distributions, its use for robustness purposes is scarce due to its complicated distributional form. Application of this distribution has so far been limited to the real world data and, to the best of our knowledge, the evaluation of its robustness properties within a Bayesian framework has not been performed. As an alternative to Normal-Normal model, it is therefore of interest to develop a Bayesian regression analysis with the RN distributed errors and Student-t prior for the model parameters.

We now turn to model (3.1). Let $\varepsilon_i \sim RN(0, \sigma, a, b)$ and β has multivariate student-t distribution with "0" location, $\sigma^2 D_\tau$ scatter matrix and v degrees of freedom. To simplify bayesian calculations, the Student-t prior of β was included to the analysis in the form of a scale mixture of normal (SMN) distributions with the mixing density being an inverse gamma distribution ([4], [16]). Resulting prior for β_j's;

$$P(\beta \mid \sigma^2) = \prod_{j=1}^{p} \int_0^\infty N(\beta_j \mid 0, \sigma^2 \tau_j^2) IG(\tau_j^2 \mid v/2, v/2) d\tau_j^2 \tag{3.3}$$

Moreover, an invariant non-informative (or improper) prior was assumed for σ^2 parameter, expressed proportionally as $P(\sigma^2) \propto \frac{1}{\sigma^2}$. Hierarchical representation of the robust bayesian regression full model can be given as follows

$$\begin{gathered} y_i \mid x, \beta, \sigma^2, a, b \sim RN(x_i^T \beta, \sigma^2, a, b) \\ \beta \mid \tau_1^2, ..., \tau_p^2, \sigma^2 \sim N_p(0_p, \sigma^2 D_\tau) \quad , \quad D_\tau = \operatorname{diag}(\tau_1^2, ..., \tau_p^2) \\ \tau_1^2, \tau_2^2, ..., \tau_p^2 \mid v \sim IG(v/2, v/2) \quad , \tau_1^2, \tau_2^2, ..., \tau_p^2 > 0 \\ \sigma^2 \sim P(\sigma^2) \end{gathered} \tag{3.4}$$

with $\tau_1^2, \tau_2^2, ..., \tau_p^2$ and σ^2 independent. The likelihood function for a random sample under this full model is given by

$$f(y \mid x, \beta, \sigma^2, a, b) \propto \sigma^{-n} \exp\left\{ -\sum_{i=1}^{n} \eta_{ab}(d_i) \right\} \tag{3.5}$$

where

$$d_i = \frac{\left| y_i - x_i^T \beta \right|}{\sigma}$$

$$\eta_{ab}(d_i) = \left(b a^{2/b} \right)^{-1} \gamma\left(2/b, a d_i^b \right)$$

The joint posterior density for all model parameters can be written as follows

$$P(\beta, \sigma^2, \tau^2 \mid y, x, a, b) \;\propto\; f(y \mid x, \beta, \sigma^2, a, b) \;\times\; P(\sigma^2) \;\times\; P(\beta \mid \sigma^2)$$

$$\propto \left[\sigma^{-n} \exp\left(-\sum_{i=1}^{n} \left(b a^{2/b} \right)^{-1} \gamma(2/b, a d_i^b) \right) \right] \times \sigma^{-2} \times$$

$$\left[\prod_{j=1}^{p} \left(\frac{1}{\sigma \sqrt{\tau_j^2}} \exp\left(-\frac{1}{2\sigma^2 \tau_j^2} \beta_j^2 \right) \right) \left(\left(\tau_j^2 \right)^{-\frac{v}{2}-1} \exp\left(-\frac{v}{2\tau_j^2} \right) \right) \right] \tag{3.6}$$

It can be clearly seen that this density has complex structure with summations and multiplications of many expressions. Major part of complication stems from the lower incomplete gamma function and we here suggest to express it as a series expansion given below ([1], [6], [40]).

$$\gamma(2/b, a d_i^b) = \left(\frac{2}{b} - 1 \right)! \left[1 - \left(\exp(-x) \sum_{k=0}^{\frac{2}{b}-1} \frac{\left(a \left| \frac{y_i - x_i^T \beta}{\sigma} \right|^b \right)^k}{k!} \right) \right] \tag{3.7}$$

The joint posterior density for β, σ^2 and τ^2 now becomes as

$$\sigma^{-n-p-2} \times \exp\{ -\left(b a^{2/b} \right)^{-1} \left(\frac{2}{b} - 1 \right)! \sum_{i=1}^{n} [1 - \exp(-\frac{a}{\sigma^b} \left| y_i - x_i^T \beta \right|^b)$$

$$\sum_{k=0}^{\frac{2}{b}-1} \frac{\left(\frac{a}{\sigma^b} \left| y_i - x_i^T \beta \right|^b \right)^k}{k!}] - \sum_{j=1}^{p} \left(\frac{\beta_j^2}{2\sigma^2 \tau_j^2} + \frac{v \tau_j^2}{2} \right) \} \times \left(\prod_{j=1}^{p} \left(\tau_j^2 \right)^{\frac{v+1}{2}} \right) \tag{3.8}$$

This analytically intractable form leads us to the innovative MCMC methods which require the definition of full conditional posterior distributions for each parameter. This can be achieved by ignoring all terms that are constant with respect to the parameter from the joint posterior density. Sometimes these distributions are well known distributions such as Normal, Gamma or inverse Gamma, enabling the Gibbs sampler for implementation([21]). When the full conditionals of all the parameters of interest are not readily available probability density functions, then Metropolis-Hasting method needs to be incorporated to the Gibbs sampler, which is known as MHWG sampling technique ([20]).

The following section provide the framework for MHWG sampling technique for the current problem.

4. Markov Chain Monte Carlo (MCMC) Approach

The full conditional posterior density for β is proportionally obtained as

$$P(\beta \mid \sigma^2, \tau^2, y, x, a, b) \propto \exp\left[\begin{array}{c} -\left(ba^{2/b}\right)^{-1}\left(\frac{2}{b}-1\right)! \\ \times \sum_{i=1}^{n}\left(\exp\left(-\frac{a}{\sigma^b}\left|y_i - x_i^T\beta\right|^b\right)\sum_{k=0}^{\frac{2}{b}-1}\frac{\left(\frac{a}{\sigma^b}\left|y_i-x_i^T\beta\right|^b\right)^k}{k!}\right) \\ -\frac{1}{2\sigma^2}\sum_{j=1}^{p}\frac{\beta_j^2}{\tau_j^2} \end{array}\right] \tag{4.1}$$

Similarly, the full conditional densities for σ^2 and latent variable τ_j^2 $(j = 1, 2, ..., p)$ are respectively: $P(\sigma^2 \mid \beta, \tau^2, y, x, a, b)$

$$\propto \sigma^{-n-p-2}\exp\left[\begin{array}{c} -\left(ba^{2/b}\right)^{-1}\left(\frac{2}{b}-1\right)! \\ \times \sum_{i=1}^{n}\left(\exp\left(-\frac{a}{\sigma^b}\left|y_i - x_i^T\beta\right|^b\right)\sum_{k=0}^{\frac{2}{b}-1}\frac{\left(\frac{a}{\sigma^b}\left|y_i-x_i^T\beta\right|^b\right)^k}{k!}\right) \\ -\frac{1}{2\sigma^2}\sum_{j=1}^{p}\frac{\beta_j^2}{\tau_j^2} \end{array}\right] \tag{4.2}$$

$$P(\tau_j^2 \mid \beta, \sigma^2, y, x, a, b) \propto \left(\tau_j^2\right)^{\frac{v+1}{2}}\exp\left(-\frac{1}{2}\left(\frac{\beta_j^2}{\sigma^2\tau_j^2} + v\tau_j^2\right)\right) \tag{4.3}$$

It can be deduced that the conditional distribution of latent variable τ_j^2 is Generalized Inverse Gaussian distribution with the parameters

$$\tau_j^2 \mid \beta, \sigma^2, y, x, a, b \sim GIG\left(v, \frac{\widehat{\beta_j}^2}{\widehat{\sigma}^2}, \frac{v+3}{2}\right) \tag{4.4}$$

Therefore, Bayesian estimator of τ_j^2 is as follows

$$\widehat{\tau_j^2} = \frac{\left|\frac{\widehat{\beta_j}}{\widehat{\sigma}}\right| K_{\frac{v+5}{2}}}{\sqrt{v}K_{\frac{v+3}{2}}} \tag{4.5}$$

where K_p is a modified Bessel function of the second kind. Since the form of the full conditional posterior density of the latent variable resembles the density of a known distribution, the marginal posterior distribution of this variable is achieved by the Gibbs sampling method.

On the other hand, full conditionals for the rest of the parameters do not belong to a known distribution family and cannot be directly simulated in an easy way.

For each of those, a Metropolis Hasting algorithm is incorporated to Gibbs sampler, which uses a proposal density that closely matches to the full conditional (i.e. target density) and the resulting procedure becomes a MHWG sampling process as described before.

5. Simulation Study

5.1. **Generating samples from RN distribution.** Random sample generation from RN distribution is currently not possible via either available statistical package programs or libraries. We therefore utilized an independent Metropolis Hastings algorithm ([28], [22]), which corresponds to the acceptance -rejection method so as to fulfill this gap. Target distribution is here the standard RN distribution with $a = 0.05$, $b = 2$. Finite mixture of two Normal distributions were considered to be the most appropriate as a proposal distribution with p.d.f;

$$g(x; \lambda, \sigma_1^2, \sigma_2^2) = (1 - \lambda)\, N(0, \sigma_1^2) + \lambda N(0, \sigma_2^2) \tag{5.1}$$

Values of parameters were chosen as $1 - \lambda = 0.85$, $\sigma_1^2 = 1$ and $\sigma_2^2 = 10$ such that proposal distribution matches to $N(0, 1)$ in terms of first and third quartiles while having tails heavier than the target distribution of $RN(0, 1, 0.05, 2)$. Then the metropolis algorithm is:

1) A realisation of a first-order Markov process is generated, $x^{(1)}, x^{(2)}, ..., x^{(t)}$, starting from some initial state $x^{(0)}$.

2) At step $(t-1)$, current state is $x^{(t-1)}$. Generate a candidate state, x^*, from the proposal distribution: $x^* \sim g(x^*)$. This was achieved by using *rnormmix* function of {mixtools} library in R

3) Calculate the ratio of two states:

$$\alpha_t \left(x^{(t-1)}, x^*\right) = \min\left\{1, \frac{f(x^*; a, b)}{f(x^{(t-1)}; a, b)} \frac{g(x^{(t-l)} \mid x^*)}{g(x^* \mid x^{(t-l)})}\right\} \tag{5.2}$$

as the proposal distribution $g(x)$ is symmetric, the ratio reduces to $\frac{f(x^*;a,b)}{f(x^{(t-1)};a,b)}$.

4) Generate u form $U(0, 1)$

5) If $u \leq \alpha_t \left(x^{(t-1)}, x^*\right) \Longrightarrow$ set $x^{(t)} = x^*$ else set $x^{(t)} = x^{(t-1)}$

6) Repeat steps (2)-(5) N times (# of iteration)

Running the chain 15000 iterations with a burn-in period of 5000, we obtained the acceptance rate as 0.5295 which is plausible as a desirable value for this is stated to be between 0.2 and 0.7 ([36]). The chain produces no diagnostic problems such as autocorrelation. Figure 3 presentation proves that the simulated sample follows $RN(0, 1, 0.05, 2)$ distribution.

5.2. **Simulation Settings.** For the simulation study conducted, the response variable (y) is assumed to follow the linear model below

$$yi = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \varepsilon_i \quad , \quad i = 1, 2, ..., n \tag{5.3}$$

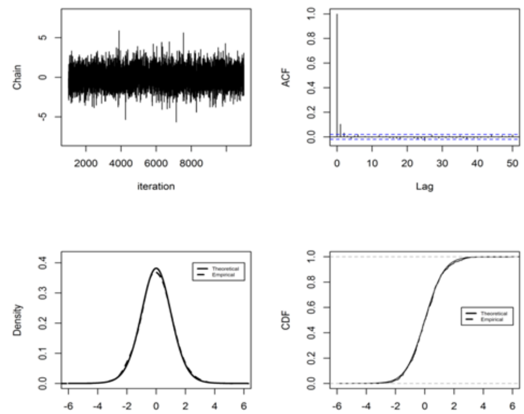

FIGURE 3. Comparison of distribution functions of theoretical RN and simulated sample

where all regressors were generated from $U(-5, 5)$ and regression coefficients were set to be "1". The sample size is taken as $n = 50$. For the random errors (ε), the heavy-tailed $RN(0,\,1,\,0.05,\,2)$ distribution with common variance (σ) was considered. Outlier addition to the samples was achieved by altering the observations with a constant value of "100" at the percentages of 2 and 5 in the $x-$, $y-$ and $x-y$ direction. Note that these rates correspond to "1" and "3" outliers respectively for the sample size of 50.

The prior information for regression coefficients was inserted to the analysis in postulations of: no priori knowledge, informative knowledge and conflicted expert opinions. The first implied a flat prior which was achieved by a Normal density with a very big variance. The usual Normal prior with smaller variance produced the second form. We also examined how Student-t prior behaved in this case. Notation for Student-t will be used as $t(\mu, \sigma, v)$, where μ =location, σ =scale, v =degrees of freedom, in the following sections. Finally, we assumed that conflicted beliefs introduced a subjective bias to the basic summaries of prior density. A positive bias was reflected in the quantification of mean of prior by an arbitrarily chosen value, leaving the scale component unquestioned. Then, the light and heavy-tailed priors, i.e. Normal and Student-t ($v = 3$) respectively were assigned to the parameters to represent such inconsistency between the prior and the data. It must be noted that a flat prior is used for the intercept parameter of the model throughout the whole analyses in this work.

Under each simulation settings, posterior functionals and Metropolis Hastings algorithms were constructed by means of R2WinBUGS library in Rv.3.2.5 Software

([34]). Posterior estimates of parameters of interest were computed from the chains of full conditionals. Repeating this process 100 times (number of simulations), mean of posterior estimates and root mean square errors (RMSE) were obtained as presented in Tables $1-3$. All computations and simulations were performed within the R platform.

5.3. **Simulation Results.** Posterior expectations of the simulated model parameters under different prior settings are listed in Table 1-3. A flat non-informative prior represented by a Normal distribution with a very big variance produces robust estimates against outliers of all directions if it is accompanied by RN distributed errors. It should be noted that the classical regression modelling with RN error distribution reveals maximum likelihood estimates resistant to outliers only in $y-$direction (the article of authors in evaluation). Bayesian analysis incorporated flat priors here extends the robustness to all directions. It is well known that assignment of flat, non-informative, reference etc. priors is the easiest way of building robustness to the analysis in cases of no prior knowledge.

TABLE 1 Robust bayesian parameter estimates(RMSE) with non-informative prior $(N(0,10^4))$, $\varepsilon_i \sim RN(0,\sigma,0.05,2)$

Number of Outliers	Outlier Directions	$\widehat{\beta}_0$	$\widehat{\beta}_1$	$\widehat{\beta}_2$	$\widehat{\beta}_3$	$\widehat{\sigma}$
1	x-	0.995	1.001	0.998	0.999	0.919
		(0.13)	(0.05)	(0.06)	(0.12)	(0.12)
	y-	1.021	0.986	0.998	1.003	0.978
		(0.13)	(0.06)	(0.05)	(0.06)	(0.10)
	x-y	0.995	0.993	0.987	1.008	0.931
		(0.15)	(0.06)	(0.07)	(0.06)	(0.12)
3	x-	1.006	0.991	0.989	1.005	0.942
		(0.13)	(0.06)	(0.06)	(0.05)	(0.11)
	y-	1.025	0.989	1.003	0.999	0.947
		(0.13)	(0.06)	(0.06)	(0.05)	(0.11)
	x-y	0.979	0.998	0.988	1.006	0.903
		(0.14)	(0.05)	(0.05)	(0.05)	(0.14)

RMSE is the root mean square error.

Influence of outliers particularly in the $x-$ or $x-y$ direction becomes apparent when an informative prior consistent with the data is represented through conjugate settings with smaller variance. However, estimators with Student-t prior rather than Normal are observed to tolerate such influences to an arbitrarily large extent.

TABLE 2 Robust bayesian parameter estimates (RMSE) with informative prior prior $\varepsilon_i \sim RN(0, \sigma, 0.05, 2)$

Prior of β	Number of Outliers	Outlier Directions	$\widehat{\beta}_0$	$\widehat{\beta}_1$	$\widehat{\beta}_2$	$\widehat{\beta}_3$	$\widehat{\sigma}$
$N(0,1)$	1	x-	0.955	0.907	1.043	1.017	0.979
			(0.18)	(0.12)	(0.07)	(0.06)	(0.06)
		y-	0.991	0.951	0.960	1.033	0.899
			(0.26)	(0.05)	(0.09)	(0.07)	(0.10)
		x-y	0.893	0.983	1.062	0.957	0.800
			(0.14)	(0.10)	(0.07)	(0.09)	(0.21)
	3	x-	1.022	0.989	0.987	0.945	0.877
			(0.12)	(0.08)	(0.05)	(0.06)	(0.12)
		y-	1.032	1.010	1.053	1.090	0.910
			(0.03)	(0.04)	(0.06)	(0.11)	(0.20)
		x-y	1.076	0.955	0.978	1.017	0.874
			(0.09)	(0.04)	(0.02)	(0.09)	(0.18)
$t(0,1,3)$	1	x-	1.023	0.986	0.989	0.993	0.975
			(0.13)	(0.06)	(0.05)	(0.06)	(0.11)
		y-	1.002	0.989	0.997	0.994	0.968
			(0.13)	(0.06)	(0.05)	(0.06)	(0.11)
		x-y	0.952	0.983	0.996	0.995	0.940
			(0.14)	(0.06)	(0.05)	(0.05)	(0.13)
	3	x-	0.996	0.982	0.982	0.995	0.956
			(0.15)	(0.06)	(0.06)	(0.05)	(0.12)
		y-	0.973	0.983	0.987	0.989	0.945
			(0.15)	(0.06)	(0.06)	(0.05)	(0.12)
		x-y	0.983	0.988	0.984	0.991	0.923
			(0.14)	(0.05)	(0.06)	(0.06)	(0.15)

RMSE is the root mean square error.

When there is likely but not easily detectable discrepancies between the sampling and prior information, the inaccurately specified prior knowledge via Normal distribution distorts the posterior expectations for all outlier settings (Table 3). On the other hand, representation of the misspecified knowledge with a heavy-tailed prior downweights such influences, revealing robust estimates in all directions.

TABLE 3 Robust bayesian parameter estimates (RMSE) with informative but conflicting prior $\varepsilon_i \sim RN(0, \sigma, 0.05, 2)$

Prior of β	Number of Outliers	Outlier Directions	$\widehat{\beta}_0$	$\widehat{\beta}_1$	$\widehat{\beta}_2$	$\widehat{\beta}_3$	$\widehat{\sigma}$
$N(5,1)$	1	x-	1.106 (0.20)	1.183 (0.18)	0.984 (0.15)	1.169 (0.17)	0.946 (0.08)
		y-	0.992 (0.01)	1.118 (0.12)	1.064 (0.07)	1.246 (0.31)	1.017 (0.11)
		x-y	0.891 (0.11)	1.048 (0.10)	0.958 (0.16)	1.010 (0.24)	0.925 (0.09)
	3	x-	1.056 (0.30)	0.986 (0.02)	0.962 (0.04)	0.993 (0.06)	0.824 (0.18)
		y-	0.873 (0.13)	1.027 (0.03)	1.042 (0.04)	1.079 (0.09)	0.999 (0.03)
		x-y	0.948 (0.12)	0.989 (0.02)	0.948 (0.14)	0.962 (0.04)	0.902 (0.15)
$t(5,1,3)$	1	x-	0.993 (0.12)	0.990 (0.06)	0.997 (0.06)	0.989 (0.05)	0.952 (0.10)
		y-	1.005 (0.14)	0.989 (0.05)	0.985 (0.06)	1.004 (0.06)	0.972 (0.09)
		x-y	0.988 (0.14)	0.993 (0.06)	0.994 (0.05)	1.007 (0.05)	0.943 (0.12)
	3	x-	1.006 (0.13)	0.993 (0.06)	0.998 (0.06)	0.991 (0.06)	0.934 (0.12)
		y-	0.972 (0.14)	0.992 (0.05)	1.002 (0.06)	0.998 (0.05)	0.926 (0.12)
		x-y	0.965 (0.14)	0.995 (0.06)	1.008 (0.05)	0.996 (0.05)	0.914 (0.13)

RMSE is the root mean square error.

6. Real World Application

For an illustrative example, the data set famously known in the literature as Brownlee's Stackloss Plant Data, is chosen here. It contains observations of "21" days of "operation of a plant for the oxidation of ammonia (NH3) to nitric acid (HNO3)" with three explanatory variables. The dependent variable is the stackloss which is 10 times the percentage of the ingoing ammonia to the plant that escapes from the absorption column unabsorbed. The predictor variables are Air Flow, the rate of operation of the plant. Water Temperature, the temperature of cooling water circulated through coils in the absorption tower; and Acid Concentration, the concentration of the acid circulating. This data set contains four outliers, one of which is in $x-$ and, three of which are in $y-$direction. This data set was used as an example to illustrate certain computational procedures of multiple regression using the least squares method by Brownlee ([13]). Since then it has been the subject of robust procedures in at least 90 distinct papers of multiple linear regression ([19], [18], [7], [32], [2], [37], [5], [35], [11], [30], [23]).

The linear regression model assuming normally distributed errors was estimated by [13] as follows

$$\widehat{y}_{loss} = -39.919 + 0.7156x_1 + 1.2953x_2 - 0.1521x_3$$

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

Bayesian regression modelling of the available data was performed under the assumption of RN distributed errors. RN likelihood is here expected to provide a built-in protection against the possible heavy-tailed disturbances due to the apparent outliers. Parameters of regressors were assigned informative priors via standard Normal and $t(0, 1, 3)$ densities as well as non-informative priors. Although the most of the information about unknown parameters come from the same RN likelihood, we here calculated Deviance Information Criterion (DIC) to present indications of model performances under different prior settings. As a guidance for our achievements, we here also present one of the estimated models of this data by means of Huber M-estimate method in the literature. The study of [11] was chosen for this purpose and the estimated model is

$$\widehat{y}_{\text{Huber}} = -41.170 + 0.8133x_1 + 1.000x_2 - 0.1324x_3$$

Table 4 presents bayesian parameter estimates of the model under the above mentioned settings. In comparison with Huber estimates, RN estimates obtained by non-informative prior here appear resistant to the outliers of this data set that were in both $x-$ and $y-$ direction. Bayesian modelling with Student-t prior instead of flat prior gives slightly more robust estimates as implied by DIC criteria. Besides, informative Normal prior resulted in posterior estimates that change in sign (see $\widehat{\beta}_3$) and magnitude (see $\widehat{\beta}_0$) in comparison to Huber model.

TABLE 4 Robust bayesian estimates of parameters by using stackloss data set

	Prior of β	$\widehat{\beta}_0$	$\widehat{\beta}_1$	$\widehat{\beta}_2$	$\widehat{\beta}_3$	$\widehat{\sigma}$	DIC
Non Informative	$N(0, 10^4)$	-41.770	0.778	1.094	-0.124	2.92	420073
Informative	$N(0, 1)$	-52.410	0.782	1.028	0.012	3.01	420074
	$t(0, 1, 3)$	-39.960	0.765	1.089	-0.135	2.90	420072

7. Concluding Remarks

Bayesian regression modelling depends heavily on the specification of a sampling model and the prior distribution. Therefore, it is always a good practice to investigate whether the obtained posterior is sensitive to reasonable variations of the statistical model, of the selected prior, or both. Traditional bayesian models based on Normal or other conjugate distributions are known to be inadequate to cope with the conflicting information in either individual observations (i.e. outliers) or in the prior beliefs. In such cases, we ought to consider reasonable alternative models based on more robust distributional assumptions, for example, sampling distributions with heavier tails. Bayesian modelling with heavy-tailed distributions automatically downweights the influences of observations that are extremely distant, resulting in robust bayesian inferential procedures.

The primary aim of this work was to stimulate the continuing development of bayesian heavy-tailed models by employing RN distribution for the sampling model. This distribution appears also within the class of heavy-tailed distributions, however its complicated distributional form has so far limited its applications to real problems. Besides, all the work involving this distribution have been conducted on the same real world data and there has been a lack of its performance assesment under different circumstances. We here fulfilled this gap by proposing a random number generation algorithm from RN distribution, which would serve for simulation studies of its performance assesments in robustness studies.

A major challenge in the application of bayesian methods occurs when it is necessary to address the emprical knowledge of people's conflicting beliefs. The use of subjective opinions in the form of informative priors may result in inaccurately specified prior densities, affecting the posterior summaries. This is the main criticism of bayesian approaches and as a solution, the use of a heavy-tailed prior density is again suggested for the protection against such influences. Developing prior robustness by means of a heavy-tailed distribution was therefore of our secondary aim and Student-t prior was chosen for this purpose. Theoretical evaluation of bayesian estimates with RN likelihood and Student-t prior revealed analytically intractable form for posterior functionals. With the help of a series expansion and simulation based MCMC techniques, we managed to conduct a simulation study by creating conflicts between data (in $x-$, $y-$ and $x-y$ direction) or between the data and the prior. Results from simulation and real data application indicated that RN sampling model along with Student-t prior provide protection against distortions caused by outliers in all sides. Besides, handling of conflictions between data and prior can be automatically incorporated into the bayesian inference procedure through the heavy-tailed Student-t prior. Overall conclusion contributes to the use of heavy-tailed distributions to built robustness in the Bayesian analyses and RN distribution is a good candidate for this purpose.

References

[1] Abramowitz, M. and Stegun, I.A. (Eds), Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables, 9th printing, New York: Dover 260 1972.

[2] Aitkin, M. and Wilson, G.T., Mixture models, outliers and the EM algorithm, *Techno* (1980), 22, 325-331.

[3] Albert, J., Delampady, M. and Polasek, W., A class of distributions for robustness studies, *Journal of Statistical Planning and Inference* (1991), 28, 291-304.

[4] Andrews, D. R. and Mallows, C. L., Scale mixtures of normal distributions, *J. R. Statist. Soc. B* (1974),36, 99-102.

[5] Antoch, J. and Jureckova, J., Trimmed least squares estimator resistant to leverage points, *CSQ* (1985), 4, 329-339.

[6] Arfken, G., The Incomplete Gamma Function and Related Functions, *Mathematical Methods for Physicists,* 3rd ed. Orlanto, FL: Academic Press (1985), 565-572.

[7] Atkinson, A. C., Two graphical displays for outlying and influential observations in regression, *Bimetrika* (1981), 68(1), 13-20.

[8] Berger, J.O., Robust Bayesian analysis: sensitivity to the prior, *J. Statist. Plann. Inference,* (1990),25, 303-328.

[9] Berger, J.O., An overview of robust Bayesian analysis, *Technical Report* (1994), 5-124.

[10] Berger, J.O.,Rios INSUA, D. and Ruggeri, F., Bayesian robustness. In Robust Bayesian Analysis (D.Rios Insua and F.Ruggeri, eds) New York: Springer-Verlag 2000.

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

[12] Box, G. E. P. and Tiao, G. C., A further look at robustness via Bayes's theorem, *Biometrika* (1962), 49, 419-432.

[13] Brownlee, K. A., Statistical Theory and Methodology in Science and Engineering, J. Wiley New York, 491-500, 1960.

[14] Chaturvedi, A., Robust Bayesian analysis of the linear regression model, *Journal of Statistical Planning and Inference* (1996), 50(2), 175-186.

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

[16] Choy, S. T. B. and Smith, A. F. M., Hierarchical models with scale mixtures of normal distributions, *TEST* (1997), 6(1), 205-221.

[17] Choy, S. T. B. and Walker, S. G., The extended exponential power distribution and Bayesian robustness, *Statistics & Probability Letters* (2003), 65, 227-232.

[18] Denby, L. and Mallows, C. L., Two diagnostic displays for robust regression analysis, *Techno* (1977), 19, 1-13.

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

[20] Gelman, A., Carlin, J. B., Stern, H. S. and Rubin, D. B., Bayesian data analysis (2nd ed.) Boca Raton: Chapman and Hall / CRC 2004.

[21] Gilks, W. R., Richardson, S. and Spiegelhalter, D. J., Markov Chain Monte Carlo In Practice, Chapman & Hall / CRC 1996.

[22] Hastings, W. K., Monte Carlo Sampling Methods Using Markov Chains and Their Applications, *Biometrika* (1970), 57(1), 97-109.

[23] Hettmansperger, T. p. and Mc Kean, J. W., A robust alternative based on ranks to least squares in analyzing linear models, *Techno* (1977), 19, 275-284.

[24] Lange, K. L., Little, R. J. A. and Taylor, J. M. G., Robust statistical modelling using the t distribution, *Journal of the American Statistical Association* (1989), 84, 881-896.

[25] Lange, K. and Sinsheimer, J. S., Normal/Independent Distributions and Their Applications in Robust Regression, *Journal of Computational and Graphical Statistics* (1993), 2, 175-198.
[26] Liu, C., Bayesian Robust Multivariate Linear Regression with Incomplete Data, *Journal of the American Statistical Association* (1996), 91(435), 1219-1227.
[27] Mendoza, M. and Pena, E. G., Some thoughts on the Bayesian robustness of location-scale models, *Chilean Journal of Statistics* (2010), 1(1), 35-58.
[28] Metropolis, N., Rosenbluth, A.W., Rosenbluth, M. N., Teller, A. H. and Teller, E., Equations of State Calculations by Fast Computing Machines, *Journal of Chemical Physics* (1953), 21, 1087-1092.
[29] O'Hagan, A. and Pericchi, L., Bayesian heavy-tailed models and conflict resolution: A review, *Brazilian Journal of Probability and Statistics* (2012), 26 (4), 372-401.
[30] Osborne, M. R., Finite Algorithms in Optimization and Data Analysis, J. Wiley New York, 267-270, 1985.
[31] Passarin, K., Robust Bayesian Estimation, Department of Economics of the University of Insubria, Varese, Italy. 2004.
[32] Preece, D. A., Illustrative examples: illustrative of what?, *Statistician* (1986), 35, 33.
[33] Ramsay, J. O. and Novick, M. R., PLU robust Bayesian decision theory: point estimation, *Journal of the American Statistical Association* (1980), 75 (372), 901-907.
[34] R Development Core Team R (http://www.R-project.org.), A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-000051-07-0, 2013.
[35] Rey, W. j. j., M-Estimators in Robust Regression, a case study, M. B. L. E. Researc. Lab. Brussels 1977.
[36] Roberts, G. O. and Rosenthal, J. S., Optimal Scaling for Various Metropolis-Hastings Algorithms, *Statist. Science,* (2001), 16, 351-367.
[37] Ruppert, D. and Carroll, R. J., Trimmed least squares estimation in the linear model, *J. Americ. Statist. Assoc.* (1980), 75, 828-838.
[38] Shi, J., Chen, K. and Song, W., Robust errors-in-variables linear regression via Laplace distribution, *Statistics and Probability Letters* (2014), 84, 113-120.
[39] Vallejos, C. A. and Steel, M. F. J., On posterior propriety for the Student-t linear regression model under Jeffreys priors, arXiv: 1311.1454v2 [stat. ME] 2013.
[40] Wall, H. S. Analytic Theory of Continued Fractions, New York: Chelsea, 1948.
[41] West, M., Outlier Models and Prior Distributions in Bayesian Linear Regression, *Journal Royal Statistical Society* (1984), 46(3), 431-439.

Current address: Mutlu Kaya: Sinop University, Faculty of Art and Science, Department of Statistics 57000 Osmaniye köyü, Merkez SİNOP TURKEY

E-mail address: maltuntas@sinop.edu.tr

ORCID Address: http://orcid.org/0000-0001-9997-3459

Current address: Emel Çankaya:Sinop University, Faculty of Art and Science, Department of Statistics 57000 Osmaniye köyü, Merkez SİNOP TURKEY.

E-mail address: ecankaya@sinop.edu.tr

ORCID Address: http://orcid.org/0000-0002-2892-2520

Current address: Olcay Arslan: Ankara University, Faculty of Science, Department of Statistics 06100 Besevler ANKARA TURKEY.

E-mail address: oarslan@ankara.edu.tr

ORCID Address: http://orcid.org/0000-0002-7067-4997