

A robustness study of student- t distributions in regression models with application to infant birth weight data in Indonesia

This content has been downloaded from IOPscience. Please scroll down to see the full text.

2017 IOP Conf. Ser.: Earth Environ. Sci. 58 012013

(<http://iopscience.iop.org/1755-1315/58/1/012013>)

View [the table of contents for this issue](#), or go to the [journal homepage](#) for more

Download details:

IP Address: 202.124.205.249

This content was downloaded on 04/04/2017 at 04:31

Please note that [terms and conditions apply](#).

A robustness study of student- t distributions in regression models with application to infant birth weight data in Indonesia

A Ubaidillah^{1,3}, K A Notodiputro^{1*}, A Kurnia¹, A Fitrianto^{1,4} and I W Mangku²

¹Department of Statistics, Faculty of Mathematics and Natural Sciences, Bogor Agricultural University, Bogor, Indonesia

²Department of Mathematics, Faculty of Mathematics and Natural Sciences, Bogor Agricultural University, Bogor, Indonesia

³Institute of Statistics, Jakarta, Indonesia

⁴Department of Mathematics, Faculty of Science, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia

Email: khairilnotodiputro@gmail.com

Abstract. In regression models, the use of least squares method may not appropriate in modelling the data containing outliers. Many robust statistical methods have been developed to handle such a problem. Lange *et al.* [1] developed robust models based on t distributions and using M-estimation approaches. In this recent article we evaluate the performance of M-estimation as well as investigated the robustness of t distribution models in linear regression by means of simulation. The models are then applied to infant birth-weight data in Indonesia. We show that the t distribution models with small degrees of freedoms have produced better estimates from perspectives of their performance and robustness when compared to other estimates.

1. Introduction

Classical methods in regression analysis, such as least squares regression, are widely used by researchers in many disciplines. In the usage of the least squares regression, the assumptions such as normality and homoscedasticity must be satisfied. Unfortunately, these assumptions are rarely met when analyzing real data. Micceri [2] examined 440 large datasets from the psychological and educational literature. None of the data were normally distributed. The use of least squares regression with violated assumptions can result in the inaccurate computation of p-values, effect sizes, and confidence intervals. This may lead to substantive errors in the interpretation of data.

Violation of the assumptions in the least squares regression can be due to the presence of outliers. The data values which are extremely different to the majority of the dataset are called outliers. Outliers may be correct observations, but they should always be checked for transcription errors. Since 1960, many robust and resistant methods have been developed to be less sensitive to outliers, i.e. robust with respect to outliers and stable with respect to small deviations from the assumed parametric model.

It is well-known that to screen the data, to remove the outliers and then to apply classical inferential procedures in many cases are not obvious. First, in multivariate or highly structured data, it can be



difficult to identify outliers and influential observations. Second, it could be better to down-weight uncertain observations rather than to reject them, although we may wish to reject completely wrong observations. Moreover, rejecting outliers reduces the sample size and could affect the data distribution. The variances could also be underestimated from the cleaned data. Finally, empirical evidence shows that good robust procedures behave quite better than techniques based on the rejection of outliers [3].

Robust statistical procedures focus in estimation procedures, testing hypotheses and in regression models. It can proceed in two ways. First, analyst could design estimators so that a pre-selected behaviour of the influence function is achieved. There exist a great variety of approaches toward the robustness method. Among these, procedures based on M-estimators and high breakdown point estimators employ an important and complementary role. For the references, Huber [4], Hampel *et al.* [5] and Staudte and Sheather [6] are the main theoretical; see also Susanti *et al.* [7] for more practical one. Second, by replacing estimators that are optimal under the assumption of a normal distribution with estimators that are optimal for, or at least derived for, other distributions: for example using the t -distribution or with a mixture of two or more distributions. Lange *et al.* [1] proposed robust statistical modeling using the t distribution for both linear and nonlinear regression models.

This paper presents a comparative study of robustness in regression models. Section 2 reviews on robust method of estimation, and section 3 discusses robust regression using student- t distribution. Section 4 and 5 present the evaluation and application of robust methods. Simulation and implementation to real dataset are presented in these section. Section 6 discusses comparison with other study, advantage and interesting results, and section 7 states conclusions.

2. Review on robust method of estimation

Robust regression is an important tool for analyzing data affected by outliers so that the resulting models are stout against outliers. An anomalous observations may be dealt with by a preliminary of the data screening, but this is not possible with influential observations which can only be detected once the model has been fitted. In the light of the amount of data available nowadays and the automated procedures used to analyze them, robust techniques may be preferable as they automatically take possible deviations into account.

In regression problem, there are two possible sources of errors, the observations y_i , and the corresponding row vector of p regressors \mathbf{x}_i . Most robust methods in regression only consider the first. Consider a regression problem with n cases (y_i, \mathbf{x}_i) from the model

$$y_i = \mathbf{x}_i^T \boldsymbol{\beta} + e_i, \quad i = 1, \dots, n, \quad (1)$$

for a p -variate row vector \mathbf{x} .

2.1. M-estimation

M-estimation is one of the robust regression estimation methods. The letter M indicates an estimation of the maximum likelihood type. Let $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}_n(\mathbf{x}_1, \dots, \mathbf{x}_n)$ then

$$E(\boldsymbol{\beta}_n(\mathbf{x}_1, \dots, \mathbf{x}_n)) = \boldsymbol{\beta}. \quad (2)$$

Equation (2) shows that $\hat{\boldsymbol{\beta}}$ is unbiased and has minimum variance, so M-estimator has the smallest variance estimator.

$$\text{var}(\hat{\beta}_j) \geq \frac{[\hat{\beta}_j^T]^2}{nE\left(\frac{d}{d\boldsymbol{\beta}} \ln f(x_{ij}, \beta_j)\right)^2}, \quad j = 1, \dots, p, \quad (3)$$

where $\hat{\beta}_j$ is other linear and unbiased estimator for β_j .

M-estimation is an extension of the maximum likelihood estimate method that is possible to eliminate some of the data. M-estimation principle is to minimize the function $\rho(\cdot)$:

$$\hat{\beta}_j^M = \min_{\beta_j} \sum_{i=1}^n \rho \left(y_i - \sum_{j=0}^p x_{ij} \beta_j \right). \quad (4)$$

To obtain (4), we have to solve

$$\hat{\beta}_j^M = \min_{\beta_j} \sum_{i=1}^n \rho(u_i) = \min_{\beta_j} \sum_{i=1}^n \rho \left(\frac{e_i}{\sigma} \right) = \min_{\beta_j} \sum_{i=1}^n \rho \left(\frac{1}{\sigma} \left(y_i - \sum_{j=0}^p x_{ij} \beta_j \right) \right)$$

and we set estimator for σ :

$$\hat{\sigma} = \frac{MAD}{0.6745} = \frac{\text{median} |e_i - \text{median}(e_i)|}{0.6745}.$$

Furthermore, we get for the first partial derivative $\hat{\beta}_j^M$ to β_j so that

$$\sum_{i=1}^n x_{ij} \psi \left(\frac{1}{\hat{\sigma}} \left(y_i - \sum_{j=0}^p x_{ij} \beta_j \right) \right) = 0, \quad (5)$$

where $\psi = \rho'$, x_{ij} is i -th observation on the j -th independent variable and $x_{i0}=1$.

The Huber function finds the Huber M-estimator of a location parameter with the scale parameter estimated with the MAD. In this case

$$\psi_k(x) = \max(-k, \min(k, x)), \quad (6)$$

where k is a constant the user specifies. Its limit as $k \rightarrow 0$ is the median, and as $k \rightarrow \infty$ is the mean. The value $k = 1.345$ gives 95% efficiency at the Gaussian model.

Remember that huber's function has the drawback that large outliers are not down weighted to zero. This can be achieved with Tukey's *biweight* function which is a redescending estimator such that $\psi(x) \rightarrow 0$ for $x \rightarrow \infty$. In particular, it has

$$\psi_k(x) = x(k - x^2)^2, \quad (7)$$

for $-k \geq x \geq k$ and 0 otherwise, so that it gives extreme observations zero weights. The usual value of k is 4.685. In general, the standard error of these estimates are slightly smaller than in Huber fit.

For equation (5), Draper and Smith [8] give a solution by defining weighted function

$$w(e_i) = \frac{\psi \left(\frac{1}{\hat{\sigma}} \left(y_i - \sum_{j=0}^p x_{ij} \beta_j \right) \right)}{\frac{1}{\hat{\sigma}} \left(y_i - \sum_{j=0}^p x_{ij} \beta_j \right)},$$

so equation (5) becomes

$$\sum_{i=1}^n x_{ij} \psi \left(y_i - \sum_{j=0}^p x_{ij} \beta_j \right) = 0. \quad (8)$$

In matrix notation, equation (8) can be written as

$$\mathbf{X}^T \mathbf{W}_i \mathbf{X} \boldsymbol{\beta} = \mathbf{X}^T \mathbf{W}_i \mathbf{Y}, \quad (9)$$

where \mathbf{W}_i is a $n \times n$ matrix with its diagonal elements are the weights. Equation (9), as known as weighted least squares (WLS) equation, can be solved by *iteratively reweighted least squares* (IRLS) method.

2.2. MM-estimation

M-estimators are not very resistant to leverage points, unless they have redescending ψ functions. The breakdown point of M-estimators is cannot exceed $1/p$ for other robust M-estimators (that is, it decreases with increasing dimension where there are more opportunities for outliers to occur). Several robust estimators of regression have been proposed in order to remedy this shortcoming, and are high breakdown point estimators of regression. Many of them are the *least median of squares* (LMS) regression, the *least trimmed squares* (LTS) regression and S-estimator. All of them have breakdown point of 50% (Rousseeuw [9]). The disadvantages of these high breakdown point estimators are the highly inefficiency of their parameter estimates and require heavy computational effort (Venables and Ripley [3]).

It is possible to combine the resistance of these high breakdown estimators with the efficiency of M-estimation. The MM-estimator proposed by Yohai, Stahel and Zamar [10] (see also Susanti *et al.* [7]) is an M-estimator starting at the coefficients and fixed scale given by the S-estimator. The MM-estimates have an asymptotic efficiency as close to one as desired, and simultaneously breakdown point 50%. Formally, the MM-estimate $\hat{\boldsymbol{\beta}}_j^{MM}$ consists in many procedures: First, compute an S-

estimate with breakdown point 50%, denoted by $\hat{\boldsymbol{\beta}}_j^*$. Second, compute the residuals $e_i^* = y_i - \sum_{j=0}^p x_{ij} \hat{\boldsymbol{\beta}}_j^*$ and find $\hat{\sigma}^*$ as solution of $\sum_{i=1}^n \varphi \left(\frac{e_i^*}{\sigma} \right) = (n-p)k_2$. Third, find the minimum $\hat{\boldsymbol{\beta}}_j^{MM}$ of $\sum_{i=1}^n \rho \left(\frac{1}{\hat{\sigma}^*} \left(y_i - \sum_{j=0}^p x_{ij} \beta_j \right) \right)$, where $\rho(\cdot)$ is a Tukey's biweight function.

3. Robust regression using student-t distribution

Statistical inference based on the normal distribution is known to be vulnerable to outliers. The robustness procedures which discussed in Section 2 are mainly directed at detecting outliers. After editing outliers, subsequent analysis is often still restricted to least squares based on the normal linear model. A serious problem with this approach is that resulting inferences fail to reflect uncertainty in the exclusion process; in particular, standard errors tend to be too small. Lange *et al.* [1] proposed a method to robust inference on regression models using the t distribution. Its approach is to replace the normal distribution by the t distribution in statistical models.

Suppose that sample data $y_i (i = 1, \dots, n)$ are recorded for n units. Typically, one assumes that the y_i , are independent normal random vectors, then

$$y_i \underset{ind}{\sim} N \left(\mu_i(\boldsymbol{\theta}), \boldsymbol{\Sigma}_i(\boldsymbol{\varphi}) \right). \quad (10)$$

The vector mean μ_i is a known form indexed by a set of unknown parameters θ , and the covariance matrix Σ_i is known form indexed by a set of unknown parameters φ . The method proposed by Lange *et al.* [1] is replacing (10) with model

$$y_i \underset{ind}{\sim} t(\mu_i(\theta), \Psi_i(\varphi), \nu), \quad (11)$$

where $t(\mu_i, \Psi_i, \nu)$ denotes the t distribution with location parameter μ_i , scale parameter Ψ_i , ν degrees of freedom, and density

$$P(y_i | \mu_i, \Psi_i, \nu) = \frac{|\Psi_i|^{-1/2} \Gamma[(\nu+1)/2]}{\Gamma(1/2)\Gamma(\nu/2)\nu^{1/2}} \times \left(1 + \frac{1}{\nu} \sum_{i=1}^n (y_i - \mu_i(\theta))^2 \Psi_i^{-1}(\varphi)\right)^{-(\nu+1)/2}. \quad (12)$$

Inferences about θ and φ in the univariate t distribution can be solved by likelihood methods. The log-likelihood for (11), ignoring constants, is $l(\theta, \varphi, \nu) = \sum_{i=1}^n l_i(\theta, \varphi, \nu)$, where

$$l_i(\theta, \varphi, \nu) = -\frac{1}{2} \ln |\Psi_i(\varphi)| - \frac{1}{2} (\nu+1) \ln \left(1 + \frac{\delta_i^2(\theta, \varphi)}{\nu}\right) - \frac{1}{2} \ln(\nu) + \ln \left[\Gamma\left(\frac{\nu+1}{2}\right)\right] - \ln \Gamma\left(\frac{\nu}{2}\right). \quad (13)$$

The letter Γ denotes the gamma function and

$$\delta_i^2(\theta, \varphi) = \sum_{i=1}^n (y_i - \mu_i(\theta))^2 \Psi_i^{-1}(\varphi). \quad (14)$$

Suppose $g(s, \nu) = \frac{\Gamma[(\nu+1)/2]}{\Gamma(1/2)\Gamma(\nu/2)^{1/2}} \times \left(1 + \frac{s}{\nu}\right)^{-(\nu+1)/2}$, so (12) becomes

$$P(y_i | \mu_i, \Psi_i, \nu) = |\Psi_i|^{-1/2} g(\delta_i^2(\theta, \varphi), \nu). \quad (15)$$

The first and second partial derivatives of (15) with respect to θ, φ and ν in order to obtain the score and expected information matrix can be found at Lange *et al.* [1].

Note that as $\nu \rightarrow \infty$, the univariate t distribution approaches the normal distribution. When $\nu < \infty$, maximum likelihood (ML) estimation of θ and certain functions of φ are robust in the sense that outlying cases with large Mahalanobis distances δ^2 are downweighted. In particular, ML estimates of θ (with q components, say) for the normal model (10) satisfy the likelihood equation

$$\frac{\partial l}{\partial \theta} = \sum_{i=1}^n A_i \Sigma_i^{-1} (y_i - \mu_i) = 0, \text{ where } l \text{ denotes the log-likelihood and } A_i \text{ denotes } q-1 \text{ matrix of partial}$$

derivatives of μ_i with respect to θ . ML estimate of θ under the t model (11) satisfy

$$\sum_{i=1}^n w_i A_i \Sigma_i^{-1} (y_i - \mu_i) = 0, \text{ where}$$

$$w_i = \frac{\nu+1}{\nu + \delta_i^2}. \quad (16)$$

is weight assigned to case i . Clearly, w_i decreases with increasing δ_i^2 .

Although the t modeling is not a solution for all robustness problems, this model has many advantages. In particular, data with shorter-than-normal tails or asymmetric error distributions, varying

degrees of long-tails among the variables, or extreme outliers are not well modeled by [11]. An advantage of the t modeling approach, however, is that a clear statement of assumptions that is incorporated in the model specification, and a critical assessment of them can yield modifications of the model that deal with some of its limitations by allowing different degrees of freedom parameters for different variables.

4. Simulation

This section presents two simulation studies. Simulation 1 is designed to analyze the behavior of parameter estimates from models using t distribution and M-estimation. Simulation 2 studies the robustness of parameter estimates from the models by using Sensitivity Curve.

4.1. Simulation 1

The target of simulation 1 is to investigate the comparativeness of the models using t distribution with M-estimation models. The procedures of a simulation study are denoted for the following steps.

1. Suppose the independently and identically distributed samples (x_i, y_i) , $i = 1, \dots, n$ are sampled from the linear model.
2. Set X as a sequence of 30 number with minimum 1 and maximum 1000 ($x_1 = 1$, $x_2 = x_1 + I$, $x_3 = x_2 + I$, ..., $x_{29} = x_{28} + I$, $x_{30} = 1000$), where $I = 999/29$.
3. Generate the error $e_i \sim N(0,1)$, $i = 1, \dots, 30$.
4. Consider the following eight cases for the contaminated error density of e_i :
 - *Case 1*: 1 contaminant from $N(0,100)$ was added to replace 1 observation from point 3.
 - *Case 2*: 1 contaminant from $N(50,100)$ was added to replace 1 observation from point 3.
 - *Case 3*: 3 contaminants from $N(0,100)$ were added to replace 3 observations from point 3.
 - *Case 4*: 3 contaminants from $N(50,100)$ were added to replace 3 observations from point 3.
 - *Case 5*: 6 contaminants from $N(0,100)$ were added to replace 6 observations from point 3.
 - *Case 6*: 6 contaminants from $N(50,100)$ were added to replace 6 observations from point 3.
 - *Case 7*: 30 observations from $t_{\nu=3}$ to replace all observations from point 3.
 - *Case 8*: 30 observations from $t_{\nu=1}$ to replace all observations from point 3.
5. In accordance with point 2 to 4, compute the value $y_i = a + bx_i + e_i$, where we set $a = 1$ and $b = 2$.
6. Estimate the parameter models with Huber M-estimation, Tukey's bisquare estimation, MM-estimation and t distribution with degrees of freedom $\nu = 1, 3, 5$ dan $\hat{\nu}$ (estimated degrees of freedom).
7. Replicate the procedures 4 until 6 with the number of replicates is 1000.
8. Calculate and compare the performance of the models, which are the Mean value of Absolute Biases and Mean Residual Standard Error (RSE) as the following:

$$\text{mean Absolute Bias } (\beta_j) = \frac{1}{R} \sum_{r=1}^R |\hat{\beta}_j^{(r)} - \beta_j|, \quad j = 1, 2 \quad (17)$$

$$\text{mean RSE} = \frac{1}{R} \sum_{r=1}^R \left[\left(\frac{1}{n-p} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \right)^{1/2} \right], \quad (18)$$

where $r = 1, \dots, R$ ($R = 1000$), $i = 1, \dots, n$ ($n = 30$) and $p = 2$.

Table 1 shows that the model with t distribution gives better results than the M-estimation. The mean value of absolute biases and mean residual standard error (RSE) of t distribution models (particularly with small degrees of freedom) is worth less than M-estimation. The interesting results can be obtained from the simulation. For the estimated degrees of freedom, $\hat{\nu}$, with the increasing

number of outliers, the model t distribution will respond with smaller value of $\hat{\nu}$. It occurs both in the case of an error with the contaminants that the variance increase (Case 1, 3 and 5) as well as by shifting the mean and variance increase (Case 2, 4 and 6). As for the case of error with student- t distribution (Case 7 and 8), the model t distribution with $\nu = 1$ and $\hat{\nu} = 0.965$ still has the model performed better than the M-estimation.

Table 1. Mean of absolute biases and mean of residual standard error of models

Case	Criterion	Huber	Tukey	MM	$t_{\nu=1}$	$t_{\nu=3}$	$t_{\nu=5}$	$t_{\nu=\hat{\nu}}$
Case 1	Bias a	0.540	0.526	0.498	0.080	0.392	0.474	0.145
	Bias b	0.004	0.004	0.003	0.005	0.002	0.003	0.003
	RSE	1.056	1.072	0.897	0.506	0.760	0.853	0.590
	$\hat{\nu}$							1.267
Case 2	Bias a	0.520	0.527	0.498	0.081	0.393	0.475	0.139
	Bias b	0.004	0.004	0.003	0.005	0.002	0.003	0.003
	RSE	1.042	1.072	0.898	0.506	0.760	0.854	0.584
	$\hat{\nu}$							1.245
Case 3	Bias a	0.562	0.540	0.529	0.156	0.464	0.536	0.184
	Bias b	0.005	0.004	0.004	0.004	0.003	0.004	0.006
	RSE	1.094	1.114	1.034	0.591	0.955	1.258	0.485
	$\hat{\nu}$							0.708
Case 4	Bias a	0.501	0.538	0.528	0.150	0.458	0.528	0.184
	Bias b	0.005	0.004	0.004	0.004	0.003	0.004	0.006
	RSE	1.086	1.113	1.033	0.590	0.955	1.262	0.476
	$\hat{\nu}$							0.686
Case 5	Bias a	0.604	0.543	0.546	0.302	0.532	1.397	0.336
	Bias b	0.007	0.004	0.005	0.004	0.004	0.023	0.008
	RSE	1.245	1.209	1.263	0.750	1.714	8.946	0.453
	$\hat{\nu}$							0.486
Case 6	Bias a	0.468	0.541	0.545	0.297	0.520	1.301	0.342
	Bias b	0.007	0.004	0.005	0.004	0.004	0.023	0.008
	RSE	1.301	1.209	1.264	0.750	1.732	10.078	0.444
	$\hat{\nu}$							0.470
Case 7	Bias a	0.371	0.374	0.370	0.391	0.360	0.364	0.365
	Bias b	0.006	0.006	0.006	0.007	0.006	0.006	0.006
	RSE	1.109	1.103	1.166	0.654	0.959	1.084	0.988
	$\hat{\nu}$							3.539
Case 8	Bias a	0.585	0.526	0.528	0.429	0.506	0.581	0.437
	Bias b	0.010	0.009	0.009	0.008	0.008	0.010	0.008
	RSE	1.570	1.518	1.699	0.954	1.772	2.357	0.959
	$\hat{\nu}$							0.964

4.2. Simulation 2

The target of simulation 2 is to demonstrate the Sensitivity Curve to assess the stability and robustness of estimators from the models. The Sensitivity Curve $SC_n(y_m)$ is a translated and rescaled version of the empirical influence function. In many situations, $SC_n(y_m)$ will converge to the influence function when $n \rightarrow \infty$. The procedures for making Sensitivity Curve are:

1. Set X as simulation 1.
2. Generate the error $e_i \sim N(0,1)$, $i = 1, \dots, 30$.
3. Compute the value $y_i = a + bx_i + e_i$, where we set $a = 1$, and $b = 2$.
4. Consider the arbitrary value of y_m , $m=1,2,\dots,1000$, where we set y_m as a sequence of 1000 numbers with minimum 100 and maximum 300.
5. Replace one observation of the sample by an arbitrary value y_m and count the value of $SC_n(y_m) = n[T_n(x_1, \dots, x_{n-1}, y_m) - T_{n-1}(x_1, \dots, x_{n-1}, y_m)]$, where $T_n(x)$ denotes the estimator of interest based on the sample x of size n .
6. Plot the value of $SC_n(y_m)$ as the y -axis dan y_m as the x -axis.

Figure 1 shows that the t distribution model with $\nu = 1$ looks more robust than the M-estimation. But with the increasingly large ν (i.e $\nu = 5$), models of t distribution is no more robust than the M-estimate models.

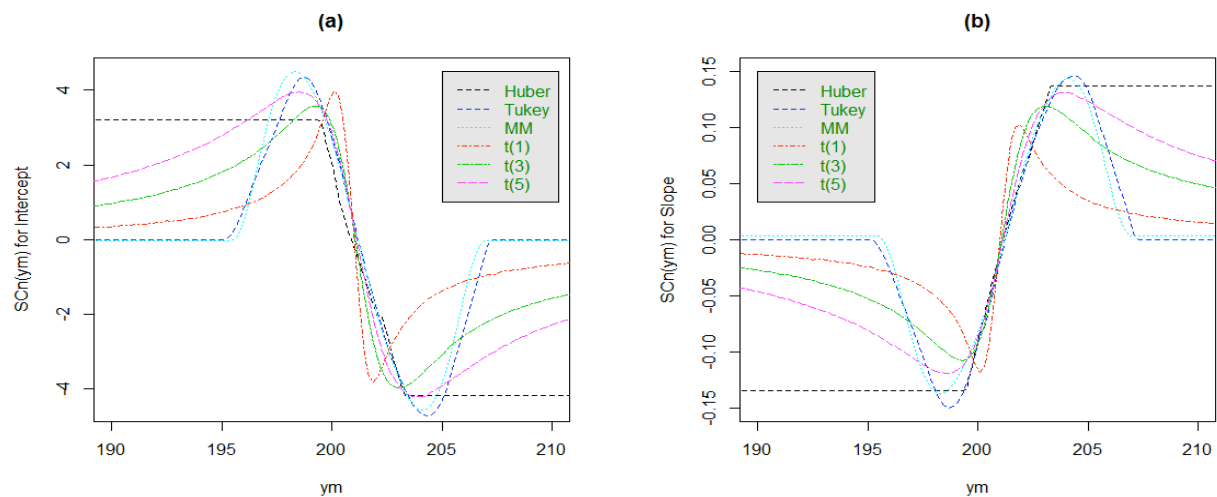


Figure 1. Sensitivity Curve for Intercept (a) and Slope (b) Parameter Estimates

5. Application to real data

In this paper we apply the models to the infant birth weight data in Indonesia at 2012. The infant birth weight data have been collected by the Indonesia Demographic and Health Survey (IDHS) which have been conducted regularly by Statistics Indonesia (BPS) [11] every five years. The sample size of data is 15124 infant which was born five years ago from the period of survey. The variables used in this study are the infant birth weight as dependent variable (Y) and the age of their mother as independent variable (X).

Table 2 shows the model t distribution with $\nu = 1$ has the smallest of standard error of estimates and residual standard error (RSE) value. The model t distribution with estimate of ν seems not more efficient compared to M-estimates, but has smaller RSE value than M-estimates. For the model t distribution with $\nu = 2, 3, 4$, and 5 have smaller of both the standard error and RSE value than the M-estimation.

6. Discussion

In general, the results of this study indicate that the model using t distribution with degrees of freedom $\nu=1$ (as named Cauchy distribution) has a performance and robustness are better than the models of M-

estimation. Almost equal to the research conducted by Lange *et al.* [1] which analyzed Stack-Loss data which the model t distribution with $\nu = 1$ also showed better results than the M-estimation.

The interesting thing from this article is the pattern of robustness shown by Sensitivity Curve where the model t distribution with $\nu = 1$ looks more effective of robustness in estimate change of its fluctuations as compared to the model M-estimate. Moreover, with the decreasing of ν , the performance and robustness of its models look better.

Table 2. Regression of infant birth weight data: estimated from 10 models

Model	β_0		β_1		RSE
	Mean	Std. Error	Mean	Std. Error	
LS	3018.154	21.192	5.320	0.750	567.938
Huber	3008.644	19.778	5.663	0.700	494.070
Tukey	3010.980	19.800	5.664	0.700	497.531
MM	3010.981	19.799	5.664	0.700	498.975
$t_{\nu=1}$	2981.949	16.286	6.016	0.576	308.620
$t_{\nu=2}$	3000.419	18.112	5.783	0.641	375.971
$t_{\nu=3}$	3005.356	18.845	5.712	0.667	412.357
$t_{\nu=4}$	3007.730	19.247	5.670	0.681	435.949
$t_{\nu=5}$	3009.202	19.507	5.639	0.690	452.734
$t_{\hat{\nu}=6.004}$	3010.226	19.917	5.614	0.706	465.411

7. Conclusion

This article shows the ability of models based on the t distribution to overcome outliers compared with M-estimates models. From the simulation and real data set application results we found that the best model fit to handle outliers is the models based on the t distribution with smallest degrees of freedom ν . The t distribution model with small ν also gives the good performance in robustness of its estimator.

Acknowledgments

The authors would like to express a sincere gratitude to Statistics Indonesia for providing funds for this research as a part of the PhD grant research program and for its human resource development projects in cooperation with Bogor Agricultural University, Indonesia.

References

- [1] Lange K L, Little R J A and Taylor J M G 1989 *J. Am. Statist. Ass.* **84** 881-96
- [2] Micceri T 1989 *Psyc. Bulletin* **105** 156-66
- [3] Venables W N and Ripley B D 2002 *Modern applied statistics with S* 4th Ed. (New York: Springer) chapter 5 pp 119-26
- [4] Huber P J 1981 *Robust statistics* (New York: John Wiley and Sons)
- [5] Hampel F R, Ronchetti E M, Rousseeuw P J and Stahel W A 1986 *Robust statistics. The approach based on influence functions* (New York: John Wiley and Sons)
- [6] Staudte R G and Sheather S J 1990 *Robust estimation and testing* (New York: John Wiley and Sons)
- [7] Susanti Y, Pratiwi H, Sulistijowati S H and Liana T 2014 *Int. J. Pure and Appl. Math.* **91** 349-60
- [8] Draper N R and Smith H 1998 *Applied Regression Analysis* 3th Ed (New York: Wiley-Interscience)

- [9] Rousseeuw P J 1984 *J. Am. Statist. Ass.* **79** 871-81
- [10] Yohai V, Stahel W A and Zamar R H 1991 A procedure for robust estimation and inference in linear regression *In Directions in Robust Statistics and Diagnostics* part 2. ed W A Stahel and S W Weisberg (New York: Springer-Verlag)
- [11] BPS 2012 *Indonesia Demographic and Health Survey* (Jakarta: BPS)