PAPER • OPEN ACCESS

Robustness of location estimators under *t*-distributions: a literature review

View the article online for updates and enhancements.

Robustness of location estimators under *t*-distributions: a literature review

C Sumarni[1,2]**, K Sadik**[1,]**K A Notodiputro**[1] **and B Sartono**[1]

[1] Department of Statistics, Bogor Agriculture University, INDONESIA.
[2] Statistics Indonesia, Bali Province.

Email: cucu_s@bps.go.id,   kusmansadik@gmail.com,   khairilnotodiputro@gmail.com,
bagusco@gmail.com

Abstract. The assumption of normality is commonly used in estimation of parameters in statistical modelling, but this assumption is very sensitive to outliers. The *t*-distribution is more robust than the normal distribution since the *t*-distributions have longer tails. The robustness measures of location estimators under *t*-distributions are reviewed and discussed in this paper. For the purpose of illustration we use the onion yield data which includes outliers as a case study and showed that the *t* model produces better fit than the normal model.

1. Introduction

Estimation of parameters in a survey often assumes that the data is a random sample from a normally distributed population. Suppose that sample data y_i are recorded n units $(1 \leq i \leq n)$, and y_i are assumed as random samples from normal distribution,

$$y_i : \text{ iid } N\left(\mu, \sigma^2\right) \tag{1}$$

In statistical modeling it is also common to assume normality of the error terms. The model can be written as (2) and called the normal model.

$$y_i = \boldsymbol{x}_i^T \boldsymbol{\beta} + \varepsilon_i, \text{ where } \varepsilon_i : \ N\left(0, \sigma^2\right) \text{or}$$
$$y_i : \text{ iid } N\left(\mu(\boldsymbol{\beta}), \sigma^2\right) \tag{2}$$

Assuming normal distribution implies that the estimates become inaccurate when there are outliers in the data. We may assume robust distribution of the error terms to solve this problem. The *t*-distribution is useful for statistical modelling when data contains outliers. Lange *et al*. [1] has successfully demonstrated that the estimation results were robust to outliers in linear models if the assumptions of normality has been replaced by assuming a *t*-distribution, $\varepsilon_i : \ t(0, \sigma^2, v)$. The *t*-model can be written as

$$y_i : \text{ iid } t\left(\mu(\boldsymbol{\beta}), \sigma^2, v\right) \tag{3}$$

where $t(\mu(\boldsymbol{\beta}), \sigma^2, v)$ denotes the univariate t-distribution with location parameter $\mu(\boldsymbol{\beta})$ where $\boldsymbol{\beta}$ is a vector of coefficient's regression, scale parameter σ^2 and v degrees of freedom.

Lange *et al*. [1] has not shown explicit measures of the robustness of location estimators under t-distributions. In this paper, this robustness is reviewed by referring to the idea that the robustness of an estimator can be measured from the influence functions (IF), the asymptotic relative efficiency (ARE) and the breakdown point [2]. In this case, we focus on the influence functions and the asymptotic relative efficiency. We also study the influence of outliers toward modelling by illustrating the model of solid pesticides effects on the onion yield.

2. Location estimation under t-distributions

We assume in a robust estimation that y_i are random samples from t-distributed population which density function (pdf) is defined by (4).

$$f(y_i) = \frac{\Gamma((v+1)/2)}{\sqrt{\pi v \sigma^2} \Gamma(v/2)} \left(1 + \frac{1}{v} \left(\frac{y_i - \mu}{\sigma} \right)^2 \right)^{-(v+1)/2} \quad , -\infty \leq y \leq \infty \tag{4}$$

Note that, If $\mu = 0$ and $\sigma^2 = 1$, (4) is a density of univariate Student't distributions with v degrees of freedom (df), otherwise to be noncentral t-distributions with location parameter μ, scale parameter σ^2 and shape parameter v degrees of freedom [3]. Denote logarithm functions of (4) as $\ln f(y_i) = l_i$,

$$l_i = \ln \left(\Gamma((v+1)/2) \right) - \frac{1}{2} \ln(\pi v) - \frac{1}{2} \ln(\sigma^2) - \ln(\Gamma(v/2)) - \frac{(v+1)}{2} \ln \left(1 + \frac{1}{v} \left(\frac{y_i - \mu}{\sigma} \right)^2 \right) \tag{5}$$

or it can be written as

$$l_i = \text{constant} - \frac{(v+1)}{2} \ln \left(1 + \frac{1}{v} \left(\frac{y_i - \mu}{\sigma} \right)^2 \right) \tag{6}$$

The first derivative of l_i respect to μ is

$$\frac{\partial l_i}{\partial \mu} = \frac{v+1}{v + \left((y_i - \mu)/\sigma \right)^2} \left(\frac{y_i - \mu}{\sigma^2} \right) \tag{7}$$

If σ^2 is assumed known and v is fixed, we would get the maximum likelihood estimator (MLE) of μ by finding the solution of this equation,

$$\sum_{i=1}^{n} \frac{\partial l_i}{\partial \mu} = 0$$

$$\sum_{i=1}^{n} \frac{v+1}{v + \left((y_i - \mu)/\sigma \right)^2} \left(\frac{y_i - \mu}{\sigma^2} \right) = 0$$

If we denote $w_i = \dfrac{v+1}{v + \left((y_i - \mu)/\sigma \right)^2}$, then $\displaystyle\sum_{i=1}^{n} w_i \left(\frac{y_i - \mu}{\sigma^2} \right) = 0$

The MLE of μ under t-distributions is

$$\hat{\mu}_T = \left(\sum_{i=1}^{n} w_i y_i \right) \bigg/ \sum_{i=1}^{n} w_i \tag{8}$$

This is a weighted mean with depending on the sample, so the location estimator under t-distributions is one of an M-estimator [2]. Since w_i is a function of parameters (μ), equation (8) has no closed form solution. The EM algorithm can be used to get the solution (for more detailed see Lange *et al.* [1]).

3. The measure of robustness under t-distributions

Robustness of an estimator can be measured quantitatively and qualitatively. The quantitative robustness can be measured by a breakdown point and the qualitative ones can be seen from efficiency and stability.

3.1. The Influence Function (IF)

The stability of robust estimators can be seen from the influence function. The influence function can be computed according to the ψ function. The influence function of an M-estimate (maximum likelihood type estimates) is defined as [2].

$$\text{IF}(.) = \frac{\psi(y;\mu)}{-E\big[(\partial/\partial\mu)\psi(y;\mu)\big]} \tag{9}$$

If the ψ function of an M-estimator is computed from the first derivative of logarithm function of pdf, $\psi(y;\mu) = \frac{\partial}{\partial\mu}\big(-\ln f(y)\big)$, then the M-estimator is the usual MLE [4].

The location estimator under the t-distributions is a form of M-estimators with the ψ function defined as (10).

$$\psi(y;\mu) = \frac{v+1}{v+\big((y-\mu)/\sigma\big)^2}\left(\frac{y-\mu}{\sigma^2}\right) \tag{10}$$

The computing of the first derivative of the ψ function (10) can be seen in Appendix and the expectation is

$$E\big[(\partial/\partial\mu)\psi(y;\mu)\big] = E\left[\frac{\partial}{\partial\mu}\left[\frac{v+1}{v+\big((y-\mu)/\sigma\big)^2}\left(\frac{y-\mu}{\sigma^2}\right)\right]\right] = E\left[\frac{\frac{(v+1)}{\sigma^2}\left[\left(\frac{y-\mu}{\sigma}\right)^2 - v\right]}{\left(v+\big((y-\mu)/\sigma\big)^2\right)^2}\right]$$

$$= E\left[\frac{(v+1)}{v\sigma^2}\left[\frac{\frac{1}{v}\left(\left(\frac{y-\mu}{\sigma}\right)^2 - v\right)}{\left(\frac{v+\big((y-\mu)/\sigma\big)^2}{v}\right)^2}\right]\right]$$

Let $z = \dfrac{y-\mu}{\sigma}$, Lange *et al.* (1989) has computed that the integration yields,

$$E\left[\left(1+\frac{z^2}{v}\right)^{-m}\right] = \frac{\left(\frac{v}{2}+m-1\right)\text{L}\left(\frac{v}{2}\right)}{\left(\frac{v+1}{2}+m-1\right)\text{L}\left(\frac{v+1}{2}\right)} \quad \text{and}$$

$$E\left[\frac{z^2}{v}\left(1+\frac{z^2}{v}\right)^{-2}\right] = E\left[\frac{\left(\frac{z^2}{v}+1\right)-1}{\left(1+\frac{z^2}{v}\right)^2}\right] = E\left[\left(1+\frac{z^2}{v}\right)^{-1} - \left(1+\frac{z^2}{v}\right)^{-2}\right]. \text{ So we get}$$

$$E\left[(\partial/\partial\mu)\psi(y;\mu)\right] = E\left[\frac{(v+1)}{v\sigma^2}\left[\frac{\frac{z^2}{v}-1}{\left(1+\frac{z^2}{v}\right)^2}\right]\right] = E\left[\frac{(v+1)}{v\sigma^2}\left(\frac{z^2}{v}\right)\left(1+\frac{z^2}{v}\right)^{-2}\right] - E\left[\frac{(v+1)}{v\sigma^2}\left(1+\frac{z^2}{v}\right)^{-2}\right]$$

$$= \frac{(v+1)}{v\sigma^2}\left[\frac{v}{(v+1)(v+3)}\right] - \frac{(v+1)}{v\sigma^2}\left[\frac{v(v+2)}{(v+1)(v+3)}\right]$$

$$= \frac{1}{\sigma^2(v+3)} - \frac{(v+2)}{\sigma^2(v+3)}$$

$$= -\frac{(v+1)}{\sigma^2(v+3)}$$

and

$$-E\left[(\partial/\partial\mu)\psi(y;\mu)\right] = \frac{(v+1)}{\sigma^2(v+3)} \tag{11}$$

Thus the influence function of an M-estimator under t-distributions is

$$\text{IF}(\hat{\mu}_T) = \frac{v+3}{v+\left((y-\mu)/\sigma\right)^2}(y-\mu) \tag{12}$$

How to see that $\hat{\mu}_T$ (location estimator under t-distribution) is more robust than $\hat{\mu}_N$, the location estimator under normal distribution ($\hat{\mu}_N = \sum_{i=1}^n y_i /n$)? To answer this, we must know the ψ function and the influence function under normal distribution. The same way as before, we get under normal distribution the ψ function is

$$\psi*(y;\mu) = \frac{y-\mu}{\sigma^2} \tag{13}$$

and the influence function IF($\hat{\mu}_N$) is

$$\text{IF}(\hat{\mu}_N) = y - \mu \tag{14}$$

The influence function of location estimators under normal distribution is a trend line and unbounded. The increasingly an observed value ($y_i \to \infty$) will produce a higher value of the influence function (IF($\hat{\mu}_N$) $\to \infty$) and vice versa. It means that the data's outlier will influence highly on the estimates. Meanwhile, the influence function of location estimators under t-distributions is a bounded function (see figure 1). That is since any outlier data is down weighted by

$$w_i = \frac{v+1}{v+\left((y_i-\mu)/\sigma\right)^2} \tag{15}$$

where v is a shape parameter, then influence of such data is reduced. An outlier will not affect the estimate significantly. That is why the location estimator under t-distributions is more robust than under the normal distribution.
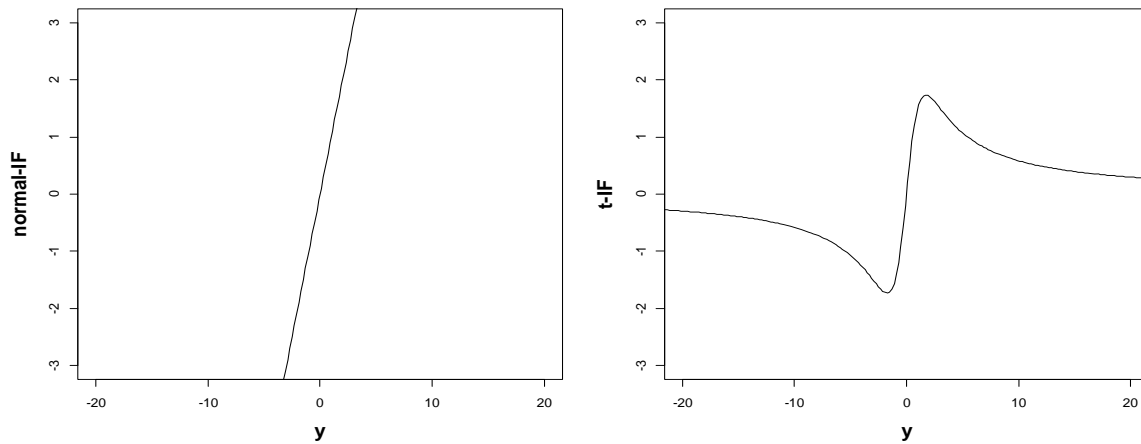

Figure 1. The influence function (IF) of location-estimators from normal distribution and t-distributions.

3.2. The Asymptotic distribution of location estimators under t-*distributions*

If the influence function (IF) of an M-estimates is proportional to the ψ function, then the asymptotic variance $(A(T))$ of $\sqrt{n}(T - \mu)$ satisfies: (Huber and Ronchetti 2009)

$$A(T) = E\left[\text{IF(.)}^2\right] \geq \frac{1}{\text{I}(\mu)} \tag{16}$$

where $\text{I}(\mu)$ is the Fisher Information. In other words, $\sqrt{n}(T - \mu)$ is asymptotically normal with mean zero and variance,

$$A(T) = E\left[\text{IF(.)}^2\right] \tag{17}$$

or we can write $\sqrt{n}\left(T - \mu\right) \to N\left(0, A(T)\right)$ (read: convergence in distribution), its mean,
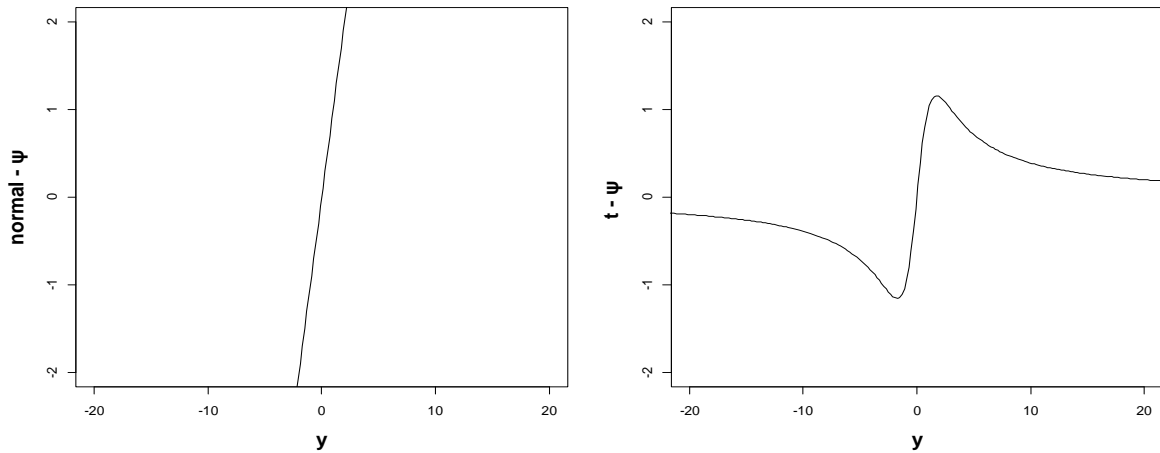
$$T \to N\left(\mu, A(T)/n\right) \tag{18}$$

Figure 2. The ψ function of M-estimates from normal distribution and t-distributions.

Comparing figure 1 by figure 2 and according to (12), we can understand that IF($\hat{\mu}_T$) is proportional to its ψ function. Now, we can find the asymptotic variance of $\hat{\mu}_T$ by (16) or (17).

$$A\left(\hat{\mu}_T\right) = E\left[IF\left(\hat{\mu}_T\right)^2\right] = \frac{E\left[\psi\left(y;\mu\right)\right]^2}{\left[-E\left[\left(\partial/\partial\mu\right)\psi\left(y;\mu\right)\right]\right]^2}$$

where $E\left[\psi\left(y;\mu\right)\right]^2 = E\left[\left(\frac{\partial}{\partial\mu}\left[\ln f\left(y\right)\right]\right)^2\right] = \mathrm{I}\left(\mu\right)$. Lange *et al.* (1989) showed that the expected

information of location estimators under *t*-distributions was $\mathrm{I}\left(\mu\right) = \frac{v+1}{\sigma^2\left(v+3\right)}$. According to (11),

$$\left[-E\left[\left(\partial/\partial\mu\right)\psi\left(y;\mu\right)\right]\right]^2 = \left[\frac{v+1}{\sigma^2\left(v+3\right)}\right]^2 = \left[\mathrm{I}\left(\mu\right)\right]^2. \quad \text{Thus,} \quad A\left(\hat{\mu}_T\right) = \frac{\mathrm{I}\left(\mu\right)}{\left[\mathrm{I}\left(\mu\right)\right]^2} = \frac{1}{\mathrm{I}\left(\mu\right)} = \frac{\left(v+3\right)}{\left(v+1\right)}\sigma^2 \text{ and}$$

the asymptotic variance of $\hat{\mu}_T$ is (19).

$$Var\left(\hat{\mu}_T\right) = \frac{\left(v+3\right)}{\left(v+1\right)}\frac{\sigma^2}{n} \tag{19}$$

The distribution of $\hat{\mu}_T$ converges to a normal distribution with mean μ and variance $Var(\hat{\mu}_T)$ as (19).

$$\hat{\mu}_T \rightarrow N\left(\mu, \frac{\left(v+3\right)}{\left(v+1\right)}\frac{\sigma^2}{n}\right) \tag{20}$$

Therefore the M-estimator under *t*-distributions is said to be asymptotically unbiased.

3.3. The Asymptotic relative efficiency
Definition (Casella and Berger 2002): if two estimator of μ, T and T' satisfy $\sqrt{n}\left(T-\mu\right) \rightarrow N\left(0, Var\left(T\right)\right)$ and $\sqrt{n}\left(T'-\mu\right) \rightarrow N\left(0, Var\left(T'\right)\right)$, the asymptotic relative efficiency (ARE) of T' with respect to T is defined as

$$ARE(T',T) = \frac{Var(T)}{Var(T')} \tag{21}$$

T is said to be asymptotically more efficient than T' if $ARE(T',T) \leq 1$ and T' is said to be asymptotically more efficient than T if $ARE(T',T) \geq 1$.

We know that the asymptotic variance of $\hat{\mu}_N$ is σ^2/n and according to (19) and (20), the asymptotic relative efficiency (ARE) of location estimators under t-distributions with v degrees of freedom ($\hat{\mu}_T$) respect to location estimators under normal distribution ($\hat{\mu}_N$) is

$$ARE(\hat{\mu}_T, \hat{\mu}_N) = \frac{Var(\hat{\mu}_N)}{Var(\hat{\mu}_T)} = \frac{\sigma^2/n}{(v+3)\sigma^2/(n(v+1))} \tag{22}$$

In normal data case (no outlier), an estimator under normal distribution of course will be better than t-distribution. The asymptotic relative efficiency of $\hat{\mu}_T$ with respect to $\hat{\mu}_N$ (22) will be

$$ARE(\hat{\mu}_T, \hat{\mu}_N, N(\mu, \sigma^2)) = \frac{v+1}{v+3} \tag{23}$$

We can see in table 1 that for normal data case, the efficiency of location estimators under t-distributions with 3 degrees of freedom ($v = 3$) is 0.67. If the degrees of freedom is raised to 30 ($v = 30$), the efficiency will increase to 0.94. Theoretically the ARE will tend to 1 as $v \to \infty$, and a location estimator under t-distributions as efficient as under the normal distribution.

Table 1. The Asymptotic Relative Efficiency of $\hat{\mu}_T$ respect to $\hat{\mu}_N$.

v	ARE
3	0.67
4	0.71
30	0.94
⋮	
∞	1

In the case our data comes from t-distributions, the location estimator under t-distributions is more efficient than under the normal distribution ($ARE(\hat{\mu}_T, \hat{\mu}_N, F = t(\mu, \sigma^2, v)) > 1$). Back to table 1, if the data came from a t-distribution with 30 degrees of freedom then we would have got an $ARE = 1.06$, whereas if the degrees of freedom were equal to 3, we would have got $ARE = 1.5$.

It is interesting, if v is fixed, which degrees of freedom would be chosen? We know that an estimator under normal distribution is very sensitive to outlier which can be seen from an unbounded IF. We can see from figure 3 that the greater the degrees of freedom closer to normal IF. In other words, we can say that the smaller the degrees of freedom is more robust.
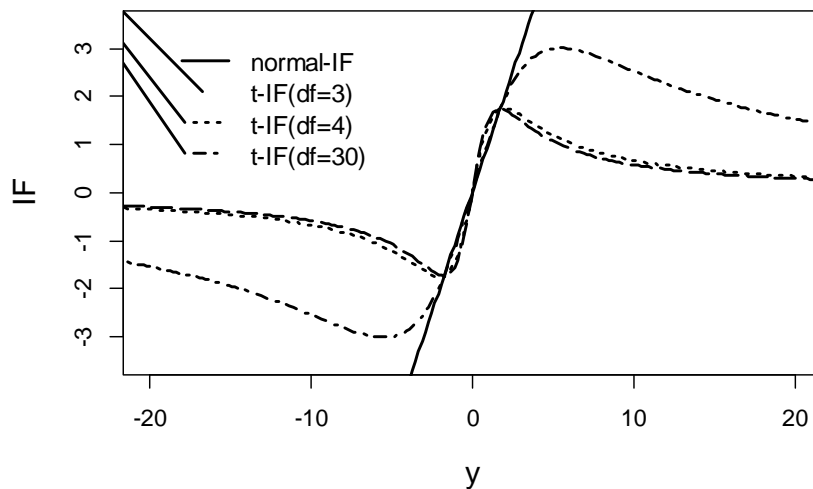
Figure 3. The Influence function of M-estimates under normal and *t*-distributions with some degrees of freedom (*df*).

4. Case study

One characteristics of Indonesian food is rich in flavors and a favorite seasoning is the onion. Brebes is the largest producer of onion in Indonesia with the best quality. The quality of onion is dependent on how treat the plants especially in dealing with pests. In this article, we studied how to fit the effects of solid pesticides on onion production when the data contains outliers. We used onion yield data presented in Listianawati [5] as a case study with little modification. The data is retrieved based on a survey of 67 onion farmers in the Kupu village, sub Regency Wanasari, Brebes Regency – the Middle Java Province at planting time May-August 2013. We used "heavy" package in R version 3.1.3 for the computing.

Figure 4(a) indicates that there is an outlier in the data. This happened on the 59th observation. He has produced 6.5 tons of onions when using solid pesticides as much as 2.6 kilograms, while the most other onion farmers harvested about 1 ton of onions and used about 1 kilogram of solid pesticides. We can also see in figure 4(a) that are likely to have a linear relationship between the use of solid pesticides and the onion productions. So we can use the linear models according to the normal model (2) or the *t* model (3). We set the degrees of freedom is equal to 4 ($v = 4$) for the *t* model.

Figure 4(b) shows the fitting model between the normal model and the *t* model. The regression line of the normal model is slightly above the regression line of the *t* model. We have just discussed in section before that the normal model is very sensitive to an outlier, and here we can see how an outlier influences the normal fit. If the outlier is removed, we will get two regression lines that overlap (figure 4(c)). It shows that the *t* model as good as the normal model when the data has no outlier. We can conclude from figure 4(b) and 4(c) that the *t* model was not influenced significantly by outliers or we can say that the *t* model is robust from outliers.
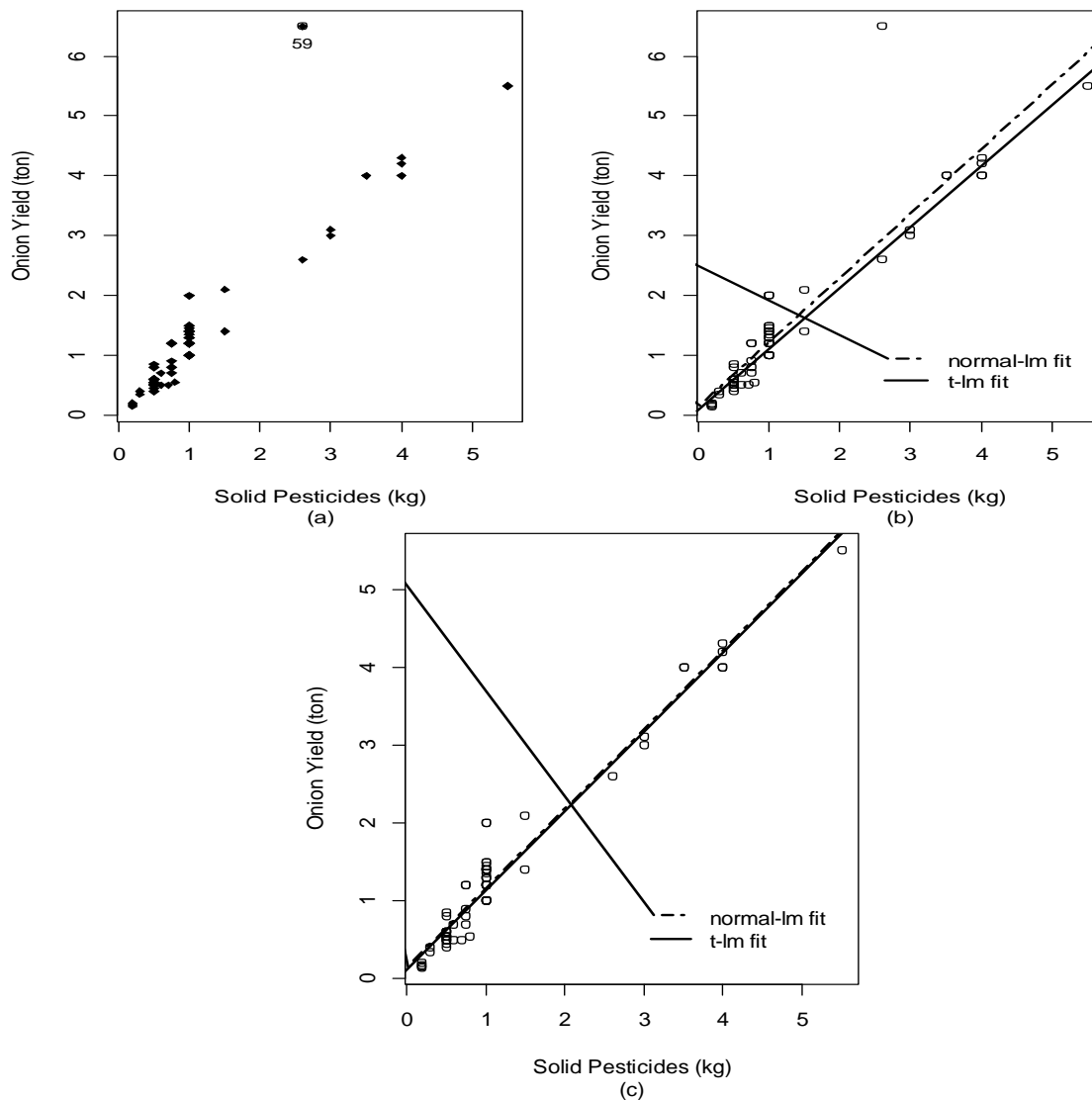
Figure 4. The plot of solid pesticides on the onion yield (a), the fitting model of normal and *t* models (b), the fitting model of normal and *t* models when the outlier is removed (c).

Table 2 presents that solid pesticides have significant effects against the onion yield both on the normal model and the *t* model. The addition of 1 kilogram of solid pesticides would increase production of the onion as much as 1.077 tons according to the normal model or 1.019 tons according to the *t* model. We see on a normal model would likely yield a higher estimation than the *t* model when there is an outlier to the right. The estimate of intercept in the normal model is also higher than the *t* model (0.131 for the normal model and 0.091 for the *t* model), but.it is not significant even at 90 percent of confidence level (*p-value*=0.146). While in the *t* model, it is significant at 95 percent of confidence level (*p-valu*e=0.018). This is because an outlier causing the estimated standard error of the regression coefficients in the normal model tend to be high so that it produces a high *p-value*.

The goodness of fit can be viewed from the value of log likelihood and/or the mean squared error (MSE). A good model would generate a maximum of log likelihood and/or minimum of MSE. Back to table 2, we see that a *t* model shows better fit than the normal model. The value of log-likelihood for the *t* model (-7.46) is greater than normal model (-49.34). The difference value is simply fantastic. It is about seven-times larger. It means that when the data includes an outlier, modelling which assumes *t-*

distributions will produce the log-likelihood maximum than normal model. The precision of the *t* model is also better than the normal model. The *t*-model produces smaller mean squared error (MSE=0.029) when compared to the normal model (MSE=0.255). The relative efficiency of the *t* model respect to the normal model is 8.8, or we may say the *t* model is about eight-times more efficient than the normal model in this cases. It is clear that the *t* model is more appropriate than the normal model.

Table 2. The estimate of the solid pesticides effect on the onion yield.

	normal-model			t-model ($v = 4$)		
	estimate	Std. error	*p-value*	estimate	Std. error	*p-value*
Parameters:						
Intercept	0.131	0.090	0.146	0.091	0.038	0.018
Solid Pesticides	1.077	0.054	0.000	1.019	0.023	0.000
Log Likelihood	-49.344			-7.460		
MSE	0.255			0.029		

5. Concluding Remarks

A location estimation which assumes *t*-distributions is one of robust estimations especially M-estimation because of the down weighted as discussed before. It has a bounded influence function and more efficient than under normal distribution when data contains outliers. Robustness of location estimators under *t*-distributions depends on its degrees of freedom. The smaller one is more robust. The case study on modeling of the onion yield has pointed out that the *t* model can overcome an outlier and preferably from normal models.

References

[1]   Lange K L, Little R J A and Taylor J M G 1989 *Journal of the American Statistical Association* **408** 881-96

[2]   Huber P J and Ronchetti E M 2009 *Robust Statistics Second Edition* (Hoboken, New Jersey: John Wiley & Son Inc.)

[3]   Kotz S and Nadarajah S 2004 *Multivariate t Distribution and Their Applications* (Cambridge: Cambridge University Press)

[4]   Casella G and Berger R L 2002 *Statistical Inference 2nd Edition* (Pacific Grove, USA: Duxbury)

[5]   Listianawati N N 2014 The analysis of factors affecting the production of onion in the Kupu village sub Regency Wanasari Brebes Regency [under graduate Thesis] (Jakarta: Syarif Hidayatullah State Islamic University Jakarta)

Appendix

The derivative of the ψ function under *t*-distribution

$$\psi = \frac{\partial}{\partial \mu}\left[\ln f(y)\right] = \frac{v+1}{v+\left((y-\mu)/\sigma\right)^2}\left(\frac{y-\mu}{\sigma^2}\right)$$

$$\psi' = \frac{\partial \psi}{\partial \mu} = \frac{\partial}{\partial \mu}\left[\frac{v+1}{v+\left((y-\mu)/\sigma\right)^2}\left(\frac{y-\mu}{\sigma^2}\right)\right]$$

Note, $w = \frac{u}{v}$, $w' = \frac{u'v - v'u}{v^2}$

$$\mathrm{u} = (v+1)\left(\frac{y-\mu}{\sigma^2}\right), \quad \mathrm{u}' = -\frac{v+1}{\sigma^2}, \quad \mathrm{v} = v + \left(\frac{y-\mu}{\sigma}\right)^2, \quad \mathrm{v}' = -2\left(\frac{y-\mu}{\sigma^2}\right)$$

$$\psi' = \frac{\left(-\frac{v+1}{\sigma^2}\right)\left(v+\left(\frac{y-\mu}{\sigma}\right)^2\right) - \left(-2\left(\frac{y-\mu}{\sigma^2}\right)\right)\left((v+1)\left(\frac{y-\mu}{\sigma^2}\right)\right)}{\left(v+\left(\frac{y-\mu}{\sigma}\right)^2\right)^2}$$

$$= \frac{\left(-\frac{(v+1)v}{\sigma^2}\right) - \left(\frac{(v+1)}{\sigma^2}\left(\frac{y-\mu}{\sigma}\right)^2\right) + \left(2\frac{(v+1)}{\sigma^2}\left(\frac{y-\mu}{\sigma}\right)^2\right)}{\left(v+\left(\frac{y-\mu}{\sigma}\right)^2\right)^2}$$

$$= \frac{\left(-\frac{(v+1)v}{\sigma^2}\right) + \left(\frac{(v+1)}{\sigma^2}\left(\frac{y-\mu}{\sigma}\right)^2\right)}{\left(v+\left(\frac{y-\mu}{\sigma}\right)^2\right)^2}$$

$$= \frac{\left(\frac{v+1}{\sigma^2}\right)\left(\left(\frac{y-\mu}{\sigma}\right)^2 - v\right)}{\left(v+\left(\frac{y-\mu}{\sigma}\right)^2\right)^2}$$