Thailand Statistician
July 2007; 5: 41-55
http://statassoc.or.th
Contributed paper

A Partial Robustifying Weighted Least Squares Estimator

Renumas Gulasirima*[a], and Pachitjanut Siripanich [b]

[a] Program of Applied Statistics, Faculty of Science and Technology, Rajabhat Suan Dusit University, Bangkok 10300, Thailand.

[b] School of Applied Statistics, National Institute of Development Administration, Bangkok 10240, Thailand.

***Author for correspondence**; e-mail: renumas99@yahoo.com

Received: 15 May 2006

Accepted: 15 January 2007

Abstract

The objective of this study is to introduce an alternative weighted least squares estimator called a partial robustifying weighted least squares (RWLS2) estimator. The weight is coincided with the weight of robustifying weighted least squares (RWLS1) estimator proposed by Gulasirima and Siripanich [3] but is partially applied on residuals. Based on ideas of Windham [7] and Gervini and Yohai [2], the proposed weight function is assigned to be one for good observations and less than one for outliers or influential observations. In particular the weight is a proportion of a power of density of errors. RWLS2 is an alternative robust regression estimator which accommodates outliers whilst all assumptions are retained.

The weighted normal random variable has an invariance property with zero mean and decreasing variance. By the results of real data study, it is found that the proposed weight can reduce the effect of influential outliers. RWLS2 performs as well as RWLS1 by mean of R^2 but slightly better by means of relative efficiency based on the MSE of least squares (LS) estimator. Both estimators work as well as the LS in situation of no outlier but obviously better when outliers exist.

Keyword: outliers, Robustifying Weighted Least Squares Estimator.

1. Introduction

In regression analysis, the most well-known estimator is the least squares (LS) estimator. However, in practice, only a single outlier may distort the LS dramatically. One basic way to overcome the problem of outliers in regression analysis is the robust regression in which the effects of outliers are reduced. Many robust regression estimators have been developed. In this paper, we focus on weighting-type estimators which are obtained by $\min_{\hat{\beta}} \sum wr^2$. Rousseeuw and Leroy [6] suggested that outliers should be deleted by means of the following weight.

$$w(r) = \begin{cases} 1 & \text{if } \left|r/\hat{\sigma}\right| \le 2.5, \\ 0 & \text{if } \left|r/\hat{\sigma}\right| > 2.5. \end{cases} \tag{1}$$

Daniel Gervini and Yohai [2] introduced the robust and fully efficient regression estimator (REWLS) based on the idea of Rousseeuw and Leroy [6]. Their proposed weight function called a down-weight and was defined as

$$w(u) = \begin{cases} 1 & \text{if } u = 0, \\ g(u) & \text{if } 0 < |u| \le 1, \\ 0 & \text{if } |u| > 1, \end{cases} \tag{2}$$

where $g(u) > 0$, w is a non-increasing function and u is defined to be proportional to $|r|$.

Gulasirima and Siripanich, [3] and Gulasirima, [4] introduced the robustifying weighted least squares (RWLS1) estimator on which the weight is based on the idea of Windham [7] and was applied on residual distribution. That is, the weight function is of the form

$$w_R(r_j) = \frac{nf^c(r_j)}{\sum_{j=1}^{n} f^c(r_j)}, \tag{3}$$

where $r_j = y_j - \hat{y}_j$, $f\left(r_j\right)$ is a normal density function and c is a positive constant.

2. A Partial Robustifying Weighted Least Squares Estimator (RWLS2)

In this paper we concentrate on a linear regression model $y = X\beta + \varepsilon$, where outliers occur one way or another. Applying equations (1) – (3), the partial weighted least squares estimator (RWLS2) is introduced. For instant, let k be a positive constant depending on fraction α of influential outliers called a cut-off point, i.e.,

$P[|R_j - \mu| > \sigma k] = \alpha$ where $R_j = Y_j - \hat{Y}_j$ is the j^{th} estimator of residual, $\mu = E(R_j)$ and $\sigma^2 = Var(R_j)$ for $j = 1, 2, ..., n$. Let $r_j = y_j - \hat{y}_j$ be an j^{th} observed residual where j = 1, 2, …, n. Thus, the partial robustifying weight is defined as

$$w_P(r_j) = \begin{cases} 1 & ; \quad |r_j - \hat{\mu}| \le k\hat{\sigma}, \\ \min\left(1\,,\ \dfrac{nf^c(r_j)}{\sum_{j=1}^{n} f^c(r_j)}\right); & \quad |r_j - \hat{\mu}| > k\hat{\sigma}. \end{cases} \tag{4}$$

where $\hat{\mu} = \dfrac{\sum w_P r}{\sum w_P} = \dfrac{\sum w_P (y - \hat{y})}{\sum w_P}$, $\hat{\sigma}^2 = \dfrac{\sum w_P (r - \hat{\mu})^2}{\sum w_P}$ and $w_P = w_P(r)$. Hence the partial robustifying weighted least square estimator or RWLS2 can be obtained as follow.

$$\hat{\beta}_P = (X'W_P X)^{-1} X'W_P Y \text{ where } W_P = diag\left(w_P(r_1), w_P(r_2), ..., w_P(r_n)\right). \tag{5}$$

Note that, the equation (4) actually is a part of robustifying weight function using for RWLS1. Steps of computing (3) were presented by Gulasirima and Siripanich [3]. Hence substituting (3) in (4), a partial robustifying weight for each observation (j^{th}) and W_P can be obtained

3. Theoretical result

Some important properties of random variable X weighted by the partial robustifying weight function (X_{PW}) are investigated as follows. The main result here needs the usages of robustifying weight function of RWLS1 (see Gulasirima and Siripanich, [3] and Gulasirima, [4]):

Let $W_R = w_R(x) = \dfrac{f^c(x)}{E\left[f^c(x)\right]}$ be the robustifying weight function satisfied C1: c is given a positive constant and C2: a random variable $X \sim N(\mu, \sigma^2)$ with a density function f(x). The random variable weighted by W_R, $X_{RW} \sim N\left(\mu, \dfrac{\sigma^2}{(c+1)}\right)$. (6)

Consider the partial robustifying weight function

$$W_P = w_P(x) = \begin{cases} 1 & ; \ |x - \mu| \ge k_\alpha \sigma, \\ w_R(x) & ; \ |x - \mu| < k_\alpha \sigma. \end{cases} \tag{7}$$

Lemma 1. Let X satisfy C1 and C2, and define the weight function $w_R(x)$ = W_R. In addition, let W_P as in (7) be $W_{P0} = w_{P0}(x)$ using a given $\sigma = \sigma_0$, then

$$w_{P0}(x) = \begin{cases} 1 & \text{if } |x-\mu| \le k_\alpha \sigma_0, \\ w_R(x) & \text{if } |x-\mu| > k_\alpha \sigma_0, \end{cases} \quad \text{given that} \tag{8}$$

$k_\alpha > 0$, $0 < \alpha_w \le \alpha_0 < 0.5$, $P\left[|X-\mu| > k_\alpha \sigma_0\right] = \alpha_0$ and $P\left[|X_{RW}-\mu| > k_\alpha \sigma_0\right] = \alpha_w$. Moreover, X_{RW} is a random variable where X is weighted by W_R . Subsequently,

$$E\left[w_{P0}(X)\right] = 1 - \alpha_0 + \alpha_w > 0 . \tag{9}$$

Proof. By applying robustifying weight properties, $W_R = w_R(x)$ is as in (6), then $X_{RW} \sim N\left(\mu, \sigma^2/(c+1)\right)$ is yielded. Let a pdf. of X_{RW} be denoted by $f_{RW}(x)$, then

$$E\left[w_{P0}(X)\right] = \int_{-\infty}^{\infty} w_{P0}(x) f(x) dx$$

$$= \int_{-\infty}^{\mu - k_\alpha\sigma_0} w_{P0}(x) f(x) dx + \int_{\mu - k_\alpha\sigma_0}^{\mu + k_\alpha\sigma_0} w_{P0}(x) f(x) dx + \int_{\mu + k_\alpha\sigma_0}^{\infty} w_{P0}(x) f(x) dx$$

$$= \int_{-\infty}^{\mu - k_\alpha\sigma_0} w_R(x) f(x) dx + \int_{\mu - k_\alpha\sigma_0}^{\mu + k_\alpha\sigma_0} f(x) dx + \int_{\mu + k_\alpha\sigma_0}^{\infty} w_R(x) f(x) dx$$

$$= 1 - P\left[|x-\mu| > k_\alpha \sigma_0\right] + \int_{-\infty}^{\mu - k_\alpha\sigma_0} f_{RW}(x) dx + \int_{\mu + k_\alpha\sigma_0}^{\infty} f_{RW}(x) dx$$

$$= 1 - \alpha_0 + \alpha_w .$$

Since $0 < \alpha_w \le \alpha_0 < 0.5$, then $E\left[w_{P0}(X)\right] > 0$.

Theorem 1. Let X satisfy C1 and C2, and define the weight functions W_R and W_{P0} as in (8). In addition, let $k_\alpha > 0$, $0 < \alpha_w \le \alpha_0 < 0.5$, $P\left[|X-\mu| > k_\alpha \sigma_0\right] = \alpha_0$ and $P\left[|X_{RW}-\mu| > k_\alpha \sigma_0\right] = \alpha_w$, where X_{RW} is a random variable obtained by weighting X by W_R .

Define $W_{P1} = w_{P1}(x) = \dfrac{w_{P0}(x)}{E\left[w_{P0}(X)\right]}$, i.e.

$$w_{P1}(x) = \begin{cases} 1/E[w_{P0}(X)] & \text{if } |x-\mu| \le k_\alpha\sigma_0, \\ w_R(x)/E[w_{P0}(X)] & \text{if } |x-\mu| > k_\alpha\sigma_0. \end{cases} \tag{10}$$

Let X_{PW} be the random variable weighted by W_{P1}, then the following statements hold:

1). The pdf. of X_{PW} is defined in this dissertation as α-contaminated normal and was found to be

$$f_{PW}(x) = \begin{cases} \dfrac{f(x)}{E[w_{P0}(X)]} & \text{if } |x-\mu| \le k_\alpha\sigma_0, \\ \dfrac{f_{RW}(x)}{E[w_{P0}(X)]} & \text{if } |x-\mu| > k_\alpha\sigma_0, \end{cases} \tag{11}$$

where $f_{RW}(x) = w_R(x)f(x)$ is the pdf. of X_{RW}

2). $E[X_{PW}] = \mu$ and

$$V(X_{PW}) = \sigma_0^2\left[1 - \frac{c\alpha}{c+1} + \frac{2\alpha k_\alpha f_Z(-k_\alpha)}{\alpha_w}\left\{\frac{(2\pi)^{c/2} f_Z^c(-k_\alpha)}{\sqrt{c+1}} - 1\right\}\right], \tag{12}$$

where $\alpha = \dfrac{\alpha_w}{1-\alpha_0+\alpha_w}$ and $f_Z(-k_\alpha) = \dfrac{e^{-k_\alpha^2/2}}{\sqrt{2\pi}}$

Proof. Let $F(x_{PW}) = F_{PW}(x)$ be the distribution of the weighted random variable X_{PW}. Let the indicator function, $\underset{(-\infty,x_{PW}]}{I(x)}$, similar to theorem 3.1, be

$$\underset{(-\infty,x_{PW}]}{I(x)} = \begin{cases} 1 & \text{if } x \le x_{PW}, \\ 0 & \text{if } x > x_{PW}. \end{cases}$$

1). In order to show that the density of X_{PW} is normally contaminated, consider

$$F_{PW}(x) = F(x_{PW}) = P[X_{PW} \le x_{PW}] = E\left[\underset{(-\infty,x_{PW}]}{I(X_{PW})}\right]$$

$$= \int_{-\infty}^{\infty} w_{P1}(x) \underset{(-\infty,x_{PW}]}{I(x)} f(x)dx$$

$$= \int_{-\infty}^{x_{PW}} w_{P1}(x) \underset{(-\infty, x_{PW}]}{I(x)} f(x)dx + \int_{x_{PW}}^{\infty} w_{P1}(x) \underset{(-\infty, x_{PW}]}{I(x)} f(x)dx$$

$$= \int_{-\infty}^{x_{PW}} \frac{w_{P0}(x)}{E\left[w_{P0}(X)\right]} f(x)dx \, .$$

Hence, $F(x_{PW}) = \frac{1}{E\left[w_{P0}(X)\right]} \int_{-\infty}^{x_{PW}} w_{P0}(x) f(x)dx \, .$ (13)

When given normal probabilities, α_0 and α_w , normal distributions

$F(\mu - k_\alpha \sigma_0) = F_Z(-k_\alpha) = \frac{\alpha_0}{2}$ and $F_{RW}(\mu - k_\alpha \sigma_0) = F_Z\left(-k_\alpha \sqrt{c+1}\right) = \frac{\alpha_w}{2}$ can be respectively defined, where $F_Z(z)$ is a standard normal distribution.

With respect to W_{P1}, $F(x_{PW})$ will be considered in the following three cases:

Case 1. If $x_{PW} < \mu - k_\alpha \sigma_0$ and $W_{P0} = w_R(x)$,

then as the result of (13) it can be shown that

$$F(x_{PW}) = \frac{1}{E\left[w_{P0}(X)\right]} \int_{-\infty}^{x_{PW}} w_R(x) f(x)dx$$

$$= \frac{1}{E\left[w_{P0}(X)\right]} \int_{-\infty}^{x_{PW}} f_{RW}(x)dx \, .$$

By applying theorem 3.1 and corollary 3.1 of Gulasirima and Siripanich, [3] (see also Gulasirima, [4])

$f_{RW}(x) = w_R(x) f(x)$ is $N\left(\mu,\ \sigma^2/(c+1)\right)$ yielded

$$F_{PW}(x) = \frac{1}{E\left[w_{P0}(X)\right]} F_{RW}(x). \quad (14)$$

Case 2. If $\mu - k_\alpha \sigma_0 \le x_{PW} \le \mu + k_\alpha \sigma_0$ and $W_{P0} = 1$, so $F(x_{PW})$ is separated into two parts;

$$F(x_{PW}) = \frac{1}{E\left[w_{P0}(X)\right]} \left\{ \int_{-\infty}^{\mu - k_\alpha \sigma_0} w_R(x) f(x)dx + \int_{\mu - k_\alpha \sigma_0}^{x_{PW}} f(x)dx \right\}$$

$$= \frac{1}{E\left[w_{P0}(X)\right]} \left\{ \frac{\alpha_w}{2} + F(x_{PW}) - \frac{\alpha_0}{2} \right\}$$

$$F_{PW}(x) = \frac{1}{E\left[w_{P0}(X)\right]}\left\{\frac{\alpha_w}{2} + F(x) - \frac{\alpha_0}{2}\right\}. \qquad (15)$$

Case 3. If $x_{PW} > \mu - k_\alpha \sigma_0$ and $W_{P0} = w_R(x)$, there are three parts to $F(x_{PW})$;

$$F(x_{PW}) = \frac{1}{E\left[w_{P0}(X)\right]}\left\{\int_{-\infty}^{\mu-k_\alpha\sigma_0} w_R(x)f(x)dx + \int_{\mu-k_\alpha\sigma_0}^{\mu+k_\alpha\sigma_0} f(x)dx + \int_{\mu+k_\alpha\sigma_0}^{x_{PW}} w_R(x)f(x)dx\right\}$$

$$= \frac{1}{E\left[w_{P0}(X)\right]}\left[\frac{\alpha_w}{2} + (1-\alpha_0) + \left\{F_{RW}(x_{PW}) - F_{RW}(\mu + k_\alpha\sigma_0)\right\}\right]$$

$$= \frac{1}{E\left[w_{P0}(X)\right]}\left[F_{RW}(x_{PW}) - \alpha_0 + \alpha_w\right]$$

$$F_{PW}(x) = \frac{1}{E\left[w_{P0}(X)\right]}\left[F_{RW}(x) - \alpha_0 + \alpha_w\right]. \qquad (16)$$

From the three cases, (14) - (16), perform a distribution of X_{PW} as

$$F_{PW}(x) = \begin{cases} \frac{1}{E\left[w_{P0}(X)\right]} F_{RW}(x) & \text{if } x < \mu - k_\alpha\sigma_0, \\ \frac{1}{E\left[w_{P0}(X)\right]}\left\{F(x) + \frac{\alpha_w}{2} - \frac{\alpha_0}{2}\right\} & \text{if } \mu - k_\alpha\sigma_0 \le x \le \mu + k_\alpha\sigma_0, \\ \frac{1}{E\left[w_{P0}(X)\right]}\left\{F_{RW}(x) + \alpha_w - \alpha_0\right\} & \text{if } x > \mu + k_\alpha\sigma_0. \end{cases} \qquad (17)$$

Therefore the density of X_{PW}, $f_{PW}(x)$, is obtained by differentiating $F_{PW}(x)$ with respect to the property W_1. If $|x-\mu| > k_\alpha\sigma_0$, it follows that

$$\frac{dF_{PW}(x)}{dx} = \frac{1}{E\left[w_{P0}(X)\right]}\frac{dF_{RW}(x)}{dx} = \frac{1}{E\left[w_{P0}(X)\right]} w_R(x) f(x)$$

If $|x-\mu| \le k_\alpha\sigma_0$, then $\frac{dF_{PW}(x)}{dx} = \frac{1}{E\left[w_{P0}(X)\right]}\frac{dF(x)}{dx} = \frac{f(x)}{E\left[w_{P0}(X)\right]}$. Hence the first statement of theorem, (11) holds.

2). Derivation of the mean and variance of X_{PW}.

We can obtain

$$E[X_{PW}] = \int_{-\infty}^{\infty} xf_{PW}(x)dx$$

$$= \frac{1}{E[w_{P0}(X)]}\left\{\int_{-\infty}^{\mu-k_\alpha\sigma_0} xf_{RW}(x)dx + \int_{\mu-k_\alpha\sigma_0}^{\mu+k_\alpha\sigma_0} xf(x)dx + \int_{\mu+k_\alpha\sigma_0}^{\infty} xf_{RW}(x)dx\right\}$$

$$= \frac{1}{1-\alpha_0+\alpha_w}\{A+B+C\}, \qquad (18)$$

by using Lemma 1., $E[w_{P0}(X)] = 1-\alpha_0+\alpha_w$, and denoting $A = \int_{-\infty}^{\mu-k_\alpha\sigma_0} xf_{RW}(x)dx$,

$B = \int_{\mu-k_\alpha\sigma_0}^{\mu+k_\alpha\sigma_0} xf(x)dx$ and $C = \int_{\mu+k_\alpha\sigma_0}^{\infty} xf_{RW}(x)dx$.

Consider the integrals in (18);

$$A = \int_{-\infty}^{\mu-k_\alpha\sigma_0} xf_{RW}(x)dx$$

$$= \int_{-\infty}^{\mu-k_\alpha\sigma_0} x\left(\frac{c+1}{2\pi\sigma_0^2}\right)^{1/2} e^{-\frac{(c+1)}{2}\left(\frac{x-\mu}{\sigma_0}\right)^2} dx$$

$$= \int_{-\infty}^{-k_\alpha\sqrt{c+1}} \left(\mu+\frac{\sigma_0}{\sqrt{c+1}}z\right)\frac{e^{-z^2/2}}{\sqrt{2\pi}}dz, \quad \text{using the transformation } z = \frac{x-\mu}{\sigma_0/\sqrt{c+1}}$$

$$= \mu\frac{\alpha_w}{2} - \frac{\sigma_0}{\sqrt{2\pi(c+1)}}e^{-k_\alpha^2(c+1)/2}. \qquad (19)$$

$$B = \int_{\mu-k_\alpha\sigma_0}^{\mu+k_\alpha\sigma_0} xf(x)dx = \int_{-\infty}^{\infty} xf(x)dx - \int_{-\infty}^{\mu-k_\alpha\sigma_0} xf(x)dx - \int_{\mu+k_\alpha\sigma_0}^{\infty} xf(x)dx$$

$$= \mu - \int_{-\infty}^{\mu-k_\alpha\sigma_0} \frac{x}{\sigma_0\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma_0^2}\right)^2}dx + \int_{\mu+k_\alpha\sigma_0}^{\infty} \frac{x}{\sigma_0\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma_0^2}\right)^2}dx$$

$$= \mu - \int_{-\infty}^{-k_\alpha} (\mu+\sigma_0 z)\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma_0^2}\right)^2}dz + \int_{k_\alpha}^{\infty} (\mu+\sigma_0 z)\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma_0^2}\right)^2}dz$$

$$= \mu - \left\{ \mu \frac{\alpha_0}{2} - \frac{k_\alpha \sigma_0}{\sqrt{2\pi}} e^{-k_\alpha^2/2} \right\} - \left\{ \mu \frac{\alpha_0}{2} + \frac{k_\alpha \sigma_0}{\sqrt{2\pi}} e^{-k_\alpha^2/2} \right\}$$

$$= \mu - \alpha_0 \mu . \qquad (20)$$

$$C = \int_{\mu + k_\alpha \sigma_0}^{\infty} x f_{RW}(x) dx = \int_{\mu + k_\alpha \sigma_0}^{\infty} x \left(\frac{c+1}{2\pi\sigma_0^2} \right)^{1/2} e^{-\frac{(c+1)}{2}\left(\frac{x-\mu}{\sigma_0}\right)^2} dx$$

$$= \mu \frac{\alpha_w}{2} + \frac{\sigma_0}{\sqrt{2\pi(c+1)}} e^{-k_\alpha^2 (c+1)/2} . \qquad (21)$$

From (19)-(21), $A + B + C = \mu(1 - \alpha_0 + \alpha_w)$ and, by applying lemma 1., $E[X_{PW}] = \mu$.

To find the variance of $X_{PW}, V(X_{PW}) = E[X_{PW}^2] - \mu^2$, it can be defined without loss in a generalized form $V(X_{PW}) = E[X_{PW}^2]$ with $\mu = 0$ given. Now let us examine $E[X_{PW}^2]$.

$$E[X_{PW}^2] = \int_{-\infty}^{\infty} x^2 f_{PW}(x) dx$$

$$= \frac{1}{E[w_{P0}(X)]} \left\{ \int_{-\infty}^{\mu - k_\alpha \sigma_0} x^2 f_{RW}(x) dx + \int_{\mu - k_\alpha \sigma_0}^{\mu + k_\alpha \sigma_0} x^2 f(x) dx + \int_{\mu + k_\alpha \sigma_0}^{\infty} x^2 f_{RW}(x) dx \right\}$$

$$= \frac{1}{1 - \alpha_0 + \alpha_w} \{ D + E + F \} , \qquad (22)$$

by applying lemma 1., $E[w_{P0}(X)] = 1 - \alpha_0 + \alpha_w$, and denoting $D = \int_{-\infty}^{-k_\alpha \sigma_0} x^2 f_{RW}(x) dx$, $E = \int_{-k_\alpha \sigma_0}^{k_\alpha \sigma_0} x^2 f(x) dx$ and $F = \int_{k_\alpha \sigma_0}^{\infty} x^2 f_{RW}(x) dx$.

Consider the integrals in (22);

$$D = \int_{-\infty}^{-k_\alpha \sigma_0} x^2 f_{RW}(x) dx = \int_{-\infty}^{-k_\alpha \sqrt{c+1}} \frac{z^2 \sigma_0^2}{(c+1)} \frac{e^{-z^2/2}}{\sqrt{2\pi}} dz , \text{ for } z = \frac{x}{\sigma_0 / \sqrt{c+1}}$$

$$= \frac{\sigma_0^2}{(c+1)\sqrt{2\pi}} \int_{-\infty}^{-k_\alpha \sqrt{c+1}} z^2 e^{-z^2/2} dz$$

$$= \frac{\sigma_0^2}{(c+1)}\left\{\frac{k_\alpha\sqrt{c+1}}{\sqrt{2\pi}}e^{-k_\alpha^2(c+1)/2}+\frac{\alpha_w}{2}\right\}. \quad (23)$$

$$E = \int_{-k_\alpha\sigma_0}^{k_\alpha\sigma_0} x^2 f(x)dx = \int_{-\infty}^{\infty} x^2 f(x)dx - \int_{-\infty}^{-k_\alpha\sigma_0} x^2 f(x)dx - \int_{k_\alpha\sigma_0}^{\infty} x^2 f(x)dx$$

$$= \sigma_0^2 - 2\int_{-\infty}^{-k_\alpha\sigma_0} x^2 \frac{1}{\sigma_0\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x}{\sigma_0}\right)^2} dx$$

$$= \sigma_0^2 - 2\int_{-\infty}^{-k_\alpha} z^2\sigma_0^2 \frac{e^{-z^2/2}}{\sqrt{2\pi}} dz, \quad \text{for } z = \frac{x}{\sigma_0}$$

$$= \sigma_0^2 - 2\sigma_0^2\left\{\frac{k_\alpha}{\sqrt{2\pi}}e^{-k_\alpha^2/2}+\frac{\alpha_0}{2}\right\}. \quad (24)$$

$$F = \int_{k_\alpha\sigma_0}^{\infty} x^2 f_{RW}(x)dx = \int_{k_\alpha\sigma_0}^{\infty} x^2\left(\frac{c+1}{2\pi\sigma_0^2}\right)^{1/2} e^{-\frac{(c+1)}{2}\left(\frac{x-\mu}{\sigma_0}\right)^2} dx = D$$, for integral of an even function.

Therefore, by combining (22)-(24), we get

$$D+E+F = \sigma_0^2 - 2\sigma_0^2\left\{\frac{k_\alpha}{\sqrt{2\pi}}e^{-k_\alpha^2/2}+\frac{\alpha_0}{2}\right\} + \frac{2\sigma_0^2}{c+1}\left\{\frac{k_\alpha\sqrt{c+1}}{\sqrt{2\pi}}e^{-k_\alpha^2(c+1)/2}+\frac{\alpha_w}{2}\right\}$$

$$= \sigma_0^2\left[(1-\alpha_0+\alpha_w) - \frac{c\alpha_w}{c+1} + 2k_\alpha f_Z(-k)\left\{\frac{(2\pi)^{c/2} f_Z^c(-k)}{\sqrt{c+1}}-1\right\}\right] \quad (25)$$

where $f_z(z)$ is a standard normal density.

Hence, the variance of X_{PW} can be derived by applying (25) as follows;

$$V(X_{PW}) = \sigma_0^2\left[1-\frac{c\alpha}{c+1}+\frac{2\alpha k_\alpha f_Z(-k_\alpha)}{\alpha_w}\left\{\frac{(2\pi)^{c/2} f_Z^c(-k_\alpha)}{\sqrt{c+1}}-1\right\}\right].$$

4. Numerical Examples

Three examples are selected from 'Robust Regression and Outlier Detection' by Rousseeuw and Leroy [6]. The first data set has no outlier and other data sets consist of outliers in y – direction and xy – direction, respectively. In each data set, the LS and the LMS estimators are already given. The rest is to compute alternative estimators (RWLS2) and then compare with the LS and the LMS estimators.

Example 1: The data is obtained from Afifi and Azen [1] quoted in Rousseeuw and Leroy, [6] and it concern the calibration of an instrument that measures the lactic acid concentration in the blood where the explanatory variable is the true concentration.

In this case, there is no outlier or leverage point. After fitting the data, It is found that all 4 methods namely LS, least median square (LMS) suggested by Rousseeuw and Leroy, [6] RWLS1 and RWLS2 yield almost the same results. The estimates of the intercept term ($\hat{\beta}_0$) and the regression coefficient (β_1) are in narrow ranges: $\hat{\beta}_0$'s are between 0.15 – 0.16, except that of RWLS1 ($\hat{\beta}_0$ = 0.1207), and $\hat{\beta}_1$'s are between 0.120 – 0.125. Consequently, their regression lines are very close together (see figure 1). In addition, their coefficient of determinations (R^2) are slightly different: R^2 of RWLS1 and RWLS2 are respectively 98.26% and 98.11% and R^2 of both LS and LMS are 97.43%. Some differences can be seen in MSE: the MSE obtained from LS and LMS methods are respectively 1.1637 and 1.1639 but the MSE for RWLS1 and RWLS2 are 0.7726 and 0.7378, respectively.

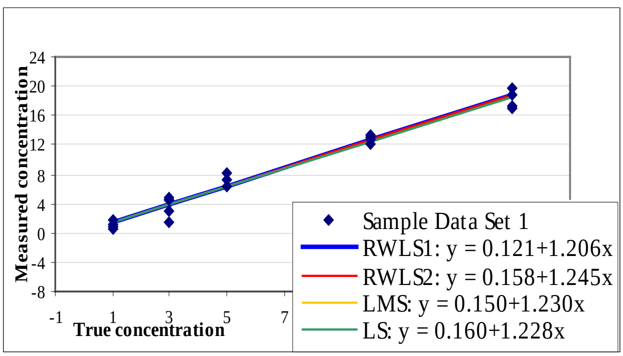

Figure 1. Observations and regression lines for Example 1: Data on the Calibration of an Instrument that Measures Lactic Acid Concentration in Blood.

Example 2: In this case, the data is the total number of international phone calls from Belgium in the years 1950-1973, provided by the Belgian Statistical Survey. Unusual data points occurred in the year 1964-1969 because the transitions did not happen exactly on New Year's Day. This caused a large fraction of outliers in the y-direction (Rousseeuw and Leroy, [6]).

Consequently, the regression line obtained from LS method is influenced by outliers. The regression lines obtained from the other three methods are almost the same but obviously different from LS regression line as can be seen in Figure 2. In the same manner, R^2 of LMS, RWLS1 and RWLS2 methods are slightly different but they are much difference from LS method. Also the MSE of RWLS1 and RWLS2 methods are almost negligible (0.0036 and 0.0028 respectively), MSE of LMS is 0.0160 and the MSE of LS method is 31.6107, very much larger than those obtained from the first 3 methods.

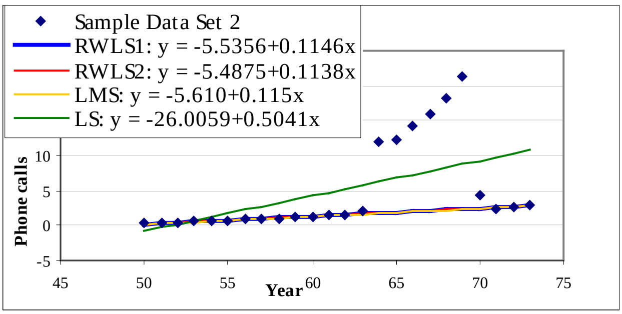

Figure 2. Observations and regression lines for Example 2: Number of International Calls from Belgium (in tens of millions).

Example 3: This data is taken from Mickey et al. [5] quoted in Rousseeuw and Leroy [6]. The response is the Gesell adaptive score corresponding to the explanatory, age (in month) of 21 children when they uttered their first word. This is a contaminated data sample which outliers appear in both of x and y directions.

Considering the regression lines obtained from 4 methods, they seem to be the same. The regression lines of RWLS2 and LMS are closed together whereas that of RWLS2 is rather different than the others (see figure 3). However, R^2 and MSE of those 4 methods are somewhat different. For R^2, the RWLS1 yields the highest value (R^2 = 60.86%). The next highest are obtained from LMS (R^2 = 57.16%) and RWLS2 (R^2 = 55.16%) method, respectively. The least R^2 is obtained from LS method and it is equal to 41.00%. For the MSE, they can be ordered ascending by their values as 66.4371 for RWLS1, 74.4458 for LMS, 75.6834 for RWLS2, and 121.5045 for LS method. It is seen that, though the regression lines are located closely, the LS method seems to have very low performance by means of R^2 and MSE.

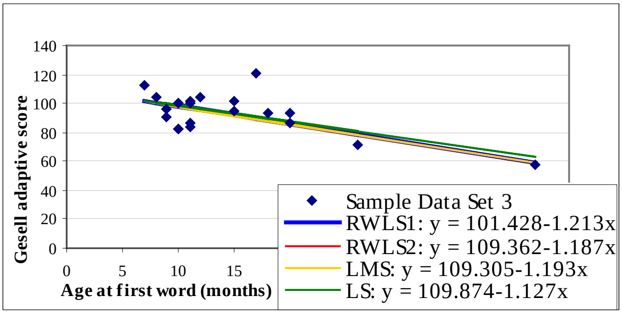

Figure 3. Observations and regression lines for Example 3: First Word-Gesell Adaptive Score Data.

The numerical examples of real data yield obviously that the proposed weight function not only reduce effect of the outliers, e.g. Example 2 and 3 but also performed as well as the LS in situation of no outlier, e.g. Example 1 (see Gulasirima [4]). The R^2 of RWLS2 is slightly less than that of RWLS1 but evidently differ from R^2 of LS in cases of outliers. Comparing between MSE of the four methods, it is found that RWLS2 has minimum values of MSE in two cases. To make comparison clearer, relative efficient (RE) of MSE for RWLS1, RWLS2 and LMS to the MSE of LS are computed. For the case of no outlier, the REs exhibit that RWLS1 and RWLS2 estimators are 1.5 – 1.6 times more efficient than LS estimators, while the LMS estimator is as efficient as the LS estimator. For the case with outliers in y – direction (Example 2), based on the values of RE, all 3 estimators are at least 1,900 times more efficient than the LS estimator. In the last case (Example 3), RWLS1, LMS and RWLS2 estimators are respectively 1.83, 1.63 and 1.61 time more efficient than the LS estimator. Details are shown in Table 1 and Table 2.

Table 1. The coefficient of determination (R^2) and mean squared error (MSE) obtained from the LS, LMS, RWLS1 and RWLS2 methods.

Data set	R^2				MSE			
	LS	LMS	RWLS1	RWLS2	LS	LMS	RWLS1	RWLS2
1.Normal	0.9743	0.9743	0.9826	0.9811	1.1637	1.1639	0.7726	0.7378
2.Outliers in Y	0.2959	0.9768	0.9940	0.9919	31.6107	0.0160	0.0036	0.0028
3.Outliers in X & Y	0.4100	0.5716	0.6086	0.5516	121.5045	74.4458	66.4371	75.6834

Table 2. The relative efficiency (RE) of the mean squared error (MSE) obtained from the LMS, RWLS1 and RWLS2 methods to the MSE obtained from the LS method.

Data set	RE		
	LMS	RWLS1	RWLS2
1. No outlier	1.00	1.51	1.58
2. Outliers in Y	1975.67	8780.75	11289.54
3. Outliers in X and Y	1.63	1.83	1.61

5. Conclusion

Results from three examples given above show that the partial robustifying weight least squared, RWLS2 method perform as good as RWLS1, LMS and LS methods in the case of no outlier but RWLS estimators are much more efficient than LS estimator.

Reconsider the weight. It can be seen that effect of outliers is ignored by small value of weight and the R^2, MSE and RE are computed from observations that are closed to a regression line. This makes the robustifying weight least squared (RWLS1 and RWLS2) methods yield "good" result, that is, high value of R^2 and small value of MSE (see Gulasirima [4]). It seems that we bestow an advantage on RWLS method rather than the LS method. Feasibly, RWLS2 may reduce this disadvantage of LS since only few observations (outliers) are transformed into weighted observations, the rest which is the most are unchanged. However, it would be interesting to find more appropriate statistic(s), besides R^2, MSE and RE, for comparison between these methods.

References

[1] Afifi. A. A., and Azen, S. P., *Statistical Analysis, a Computer Oriented Approach*, New York: Academic Press, 1979.

[2] Gervini, D., and Yohai, V.J., A Class of Robust and Fully Efficient Regression Estimators, *Annals of Statistics,* **30**; 2002: 583-616.

[3] Gulasirima, R., and Siripanich, S., Robustifying Weighted Least Squares Estimator, *Thailand Statistician,* **3**; 2005:13-27.

[4] Gulasirima, R., Robustifying Regression Models, *A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy (Statistics),*

National Institute of Development Administration, Bangkok: Kasemsri C.P. (Thewate), 2006.

[5] Mickey, M. R., Dunn, O. J., and Clark, V., Quoted in Rousseeuw, P. J. and Leroy, A. M. 1987, *Robust Regression and Outlier Detection,* New York: John Wiley & Sons, 1967.

[6] Rousseeuw, P.J., and Leroy, A.M., *Robust Regression and Outlier Detection,* New York: John Wiley & Sons, 1987.

[7] Windham, M.P., Robustifying Model Fitting**,** *Journal of the Royal Statistical Society.* **57**; 1995: 599-609.