Pertanika J. Sci. & Technol. 20 (2): 313 - 325 (2012)

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

The Performance of Classical and Robust Logistic Regression Estimators in the Presence of Outliers

Habshah, M.[1] **and Syaiba, B. A.**[2*]

[1]*Laboratory of Applied and Computational Statistics, Institute for Mathematical Research, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia*

[2]*Department of Mathematics, Faculty of Science, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia*

ABSTRACT

It is now evident that the estimation of logistic regression parameters, using Maximum Likelihood Estimator (MLE), suffers a huge drawback in the presence of outliers. An alternative approach is to use robust logistic regression estimators, such as Mallows type leverage dependent weights estimator (MALLOWS), Conditionally Unbiased Bounded Influence Function estimator (CUBIF), Bianco and Yohai estimator (BY), and Weighted Bianco and Yohai estimator (WBY). This paper investigates the robustness of the preceding robust estimators by using real data sets and Monte Carlo simulations. The results indicate that the MLE behaves poorly in the presence of outliers. On the other hand, the WBY estimator is more efficient than the other existing robust estimators. Thus, it is suggested that the WBY estimator be employed when outliers are present in the data to obtain a reliable estimate.

Keywords: Maximum Likelihood Estimator, Robust Estimators, Outliers, Goodness of Fit, Monte Carlo Simulation

INTRODUCTION

Logistic regression model is used for prediction of the probability of an occurrence $Y = 0$ or a non occurrence $Y = 1$ of an event with predictor variables X(s) that may be either numerical, categorical or both. From its original acceptance in epidemiology, the application of this model is now widely used in many research fields. In practice, the Maximum Likelihood Estimator (MLE) is used to estimate the coefficients, standard errors and to compute the goodness of fit test. The MLE is known as the most efficient estimator with good optimality properties for estimating the parameters in the logistic

Article history:
Received: 17 January 2011
Accepted: 15 April 2011

Email addresses:
habshahmidi@gmail.com (Habshah, M.),
syaibabalqish@gmail.com (Syaiba, B. A.)
*Corresponding Author

ISSN: 0128-7680 © 2012 Universiti Putra Malaysia Press.

regression model. Unfortunately, the MLE is not robust toward outliers. It is now evident that the MLE estimates are known to be severely sensitive to outliers (Croux *et al*., 2002; Victoria-Feser, 2002; Croux & Haesbroeck, 2003; Imon & Hadi, 2008; Nurunnabi *et al*., 2009; Syaiba & Habshah, 2010; Sarkar *et al*., 2011). Even a single outlier is good enough to cause the estimates to suffer, and thus, resulting in a completely erroneous estimation. In a logistic regression problem, outlying observations which are corresponding to excessively large fitted values and highly influential to the model fit are treated as outliers (Hao, 1992; Croux & Haesbroeck, 2003). Nurunnabi *et al*. (2009) defined outliers as influential observations that need not to be outlined in the sense of having large fitted values. As an alternative, robust estimators which are much less affected by outliers are considered (Künsch *et al*., 1989; Carroll & Pederson, 1993; Bianco & Yohai, 1996; Croux & Haesbroeck, 2003).

In the next section, a brief background of the classical MLE, robust estimators and goodness of fit tests is reviewed. This is followed by an evaluation of the performance of MLE and robust estimators in the real examples and the Monte Carlo simulation study (see sub-section 3). Finally, the conclusion is given in sub-section 4.

MATERIALS AND METHODS

Maximum Likelihood Estimator

Consider a multiple logistic regression model:

$$Y = \pi(X) + \varepsilon \tag{1}$$

where, with $\eta = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p = X\beta$. Here, Y is an $n \times 1$ vector of response. Let $y_i = 0$ if the i^{th} unit does not have the characteristic and $y_i = 1$ if the i^{th} unit does possess that characteristic. X is an $n \times k$ matrix of explanatory variables with $k = p + 1$. $\beta^T = (\beta_0, \beta_1, \beta_2, \ldots \beta_p)$ is the vector of the regression parameters and ε is an n × 1 vector of the unobserved random errors. The quantity π_i is known as probability or fitted value for the i^{th} covariate. The model given in Eq.(2) satisfies $0 \le \pi_i \le 1$. The fitted values in logistic regression model are calculated for each covariate pattern which is dependent on the estimated probability for the covariate pattern, denoted as $y_i = \hat{\pi}_i$. Thus, the i^{th} residual is defined as:

$$\hat{\varepsilon} = y_i - \hat{\pi}_i \qquad i = 1, 2, ..., n \tag{3}$$

A logit transformation of the logistic regression model which is linear in its parameter is defined in terms of $\pi = (X)$ as follows:

$$g(X) = \log\left(\frac{\pi}{1-\pi}\right) = X\beta \tag{4}$$

Here, "log" shall designate the base *e* logarithm. The conditional distribution of response variable follows a Bernoulli distribution with a probability given by the conditional mean,

$\pi(X)$. Since $Y_i = 0$ for $i = 1, 2, ..., n$ are assumed to be independent with n corresponding to the random variables of $(Y_1, Y_2, ..., Y_n)$, the joint probability density function is written as:

$$\begin{aligned} g(Y_1, Y_2, \ldots, Y_n) &= \prod_{i=1}^{n} f_i(Y_i) \\ &= \prod_{i=1}^{n} \pi_i^{Y_i}(1-\pi_i)^{(1-Y_i)} \end{aligned} \tag{5}$$

Then, the MLE is obtained by maximizing the logarithm of the likelihood function produces, as follows:

$$\begin{aligned} \log(g(Y_1, Y_2, \ldots, Y_n)) &= \sum_{i=1}^{n} Y_i \log\left(\frac{\pi_i}{1-\pi_i}\right) + \sum_{i=1}^{n} \log(1-\pi_i) \\ &= \sum_{i=1}^{n} Y_i(X\beta) - \sum_{i=1}^{n} \log[1+\exp(X\beta)] \end{aligned} \tag{6}$$

By differenting Eq. (6) with respect to β_0, produces $\sum_{i=1}^{n}[Y_i - \pi(X_i)] = 0$ and $\sum_{i=1}^{n} X_i[Y_i - \pi(X_i)] = 0$ for $\beta_1, \beta_2, ..., \beta_p$. The iterative estimates of β (s) are then obtained as follows:

$$\begin{aligned} \hat{\beta}^{(k+1)} &= (X^T WX)^{-1} X^T W(X\hat{\beta}^k + W^{-1}e) \\ &= \hat{\beta}^{(k)} + (X^T WX)^{-1} X^T e \end{aligned} \tag{7}$$

where W is a diagonal matrix with an element of $\pi_i(1-\pi_i)$, $e = Y - \hat{\pi}$ and k is number of iterations.

It is important to point out that when a complete separation is found in the data, the parameters of the logistic regression model cannot be estimated by the MLE. The complete separation of data (means no overlapping cases) is when the X values, that correspond to $Y = 1$, exceed all of the X values that correspond to $Y = 0$ (Albert & Anderson, 1984; Santner & Duffy, 1986).

Recently, there are many robust estimators available in the literature due to the sensitivity of the MLE in the presence of outliers. In this section, several selected robust estimators are utilized to compare their performances with the classical MLE in the presence of outliers. These robust estimators are briefly discussed in the subsequent sections.

The Mallows Type Leverage Dependent Weights Estimator (MALLOWS)

Künsch *et al.* (1989) introduced the Mallow-type estimator by minimizing the weighted log-likelihood function where the weights are dependent on covariates. Carrol and Pederson (1993) investigated more on the Mallow-type estimator and proposed to turn the MLE into an estimate with bounded influence by down-weighting the outliers in the X-space. The MALLOWS estimator was obtained by minimizing the log-likelihood on a particular weight function.

$$\sum_{i=1}^{n} w_i[y_i \log(\pi_i(\beta)) + (1-y_i)\log(1-\pi_i(\beta))] \tag{8}$$

where $w_i = W(h_n(x_i))$. W is a non-increasing function such that $W(u)u$ is bounded which is dependent on a parameter $c > 0$, and $W(u) = \left(1 - \frac{u^2}{c^2}\right)\Big|(| u | \leq c)$. W is computed by Robust Mahalanobis Distance (RMD) based on the robust estimation of the centre and scatter matrix of the covariates.

The Conditionally Unbiased Bounded Influence Function estimator (CUBIF)

The CUBIF estimator minimizes a measure of efficiency based on the asymptotic covariance matrix under the model subject to a bound on a measure of infinitesimal sensitivity that is similar to the gross error sensitivity (Künsch *et al.*, 1989). It is a consistent M-estimator in the form of $\sum_{i=1}^{n} \psi(y_i, x_i, \beta) = 0$ such that $E\left(\psi(y,x,\beta) \mid x_i\right) = 0$. The optimal function of ψ is written as follows:

$$\psi(y,x,\beta,B) = W(\beta,y,x,b,B)\left\{y - g(\beta^T x) - c\left(\beta^T x, \frac{b}{h(x,B)}\right)\right\}x \tag{9}$$

where b is bounded on the measure of infinitesimal sensitivity, B is a dispersion matrix, and $h(x,B) = (x^T B^{-1} x)^{1/2}$ is a leverage measure. The function $c\left(\beta^T x, b \big/ h(x,B)\right)$ is a corrected bias with corrected residual as shown below:

$$r(y,x,\beta,b,B) = y_i - g(\beta^T x) - c\left(\beta^T x, \frac{b}{h(x,B)}\right) \tag{10}$$

The weights are in the form of $W(\beta,y,x,b,B) = W_b(r(y,x,\beta)h(x,B))$, where W_b is the Huber weight given by $W_b(x) = \min\left\{1, \frac{b}{|x|}\right\}$. The function W downweights the observation with a large corrected residual and a large leverage making the M-estimator to have a bounded influence.

The MALLOWS and the CUBIF estimators are available in the Robust Packages of SPLUS and R under the command of glmRob.

The Bianco and Yohai Estimator (BY)

Pregibon (1981) proposed robust M-estimates to replace the total deviance function based on minimizing the weighted total deviance.

$$M(\beta) = \sum_{i=1}^{n} \rho(d^2(\pi_i(\beta), y_i)) \tag{11}$$

where $\rho(u)$ is an increasing Huber loss function. Meanwhile, deviance residuals measure the discrepancies between the probabilities fitted using the regression coefficients β and the observed values. Later, Bianco and Yohai (1996) found that this estimator does not downweight the high leverage points and is not consistent as well. They improved this estimator by minimizing it, as follows:

$$M(\beta) = \sum_{i=1}^{n} [\rho(d^2(\pi_i(\beta), y_i)) + q(\pi_i(\beta))] \tag{12}$$

where $\rho(u)$ is a bounded, differentiable and non-decreasing function, which is defined by:

$$\rho(u) = \begin{cases} u - \left(x^2 / 2k\right), & x \le k \\ k / 2, & \text{otherwise} \end{cases} \tag{13}$$

with k is a positive number. The researchers defined $q(u) = v(u) + v(1-u)$ with $v(u) = 2\int_0^u \psi(-2\log t)dt$ and $\psi = \rho^T$.

The Weighted Bianco and Yohai Estimator (WBY)

Croux and Haesbroeck (2003) noticed that when working with Huber loss function, $\rho(u)$ which was suggested by Bianco and Yohai (1996) previously, occurred frequently that the BY estimator did not exist even for uncontaminated data. Thus, Croux and Haesbroeck (2003) accomplished the BY estimator and proposed an extra weights to downweight the high leverage points, $\psi_c^{CH}(u) = \exp\left(-\sqrt{\max(u,c)}\right)$. The WBY estimator minimizes:

$$\sum_{i=1}^{n} w(x_i)\left[\psi(d_i^2(\beta))(y_i - \pi_i(\beta)) - E_\beta\left(\psi(d_i^2(\beta))(y_i - \pi_i(\beta)) \,\middle|\, x_i\right)\right] \tag{14}$$

The weight is to be a decreasing function of RMD with distance is computed using the Minimum Covariance Determinant (MCD) (Rousseeuw & Leroy, 1987) that is taken as:

$$w(x_i) = \begin{cases} 1 & \text{if} \quad RMD_i^2 \le \chi^2_{(\mathrm{p},0.975)} \\ 0 & \text{else} \end{cases} \tag{15}$$

The WBY estimator consists a loss-function to guarantee the existence of the BY estimator and to provide a stable and fast algorithm to compute the BY estimator.

$\chi_{\downarrow}$arcsin$^{\uparrow}$2 *Goodness of Fit Test*

There are several measurements used to test the goodness of fit for logistic regression model. Nonetheless, Cox and Wermuth (1992) warned not to use R^2 when Y only has two possible values; this shows that frequently $R^2 = 0.1$ when good models are used. Meanwhile, Collett (2003) has shown that the deviance, which is dependent on the fitted success probabilities π_i, can only be used to summarise the goodness of fit test for a group binary data and unreliable for binary data or when data are sparse. The Pearson's χ^2 statistics is the most popular alternative instead the deviance. Both the deviance and this Pearson's χ^2 statistics have the same asymptotic χ^2 distribution when the fitted model is correct. Even if Pearson's χ^2 statistics can be computed to access the goodness of fit test for logistic regression model in the presence of outliers, one cannot solely rely on this statistics. Kordzakhia *et al.* (2001) suggested an alternative measure by using the chi-square statistics based on the arcsin transformation, χ^2_{arc}. Later, this statistic was applied to compute the goodness of fit test and to evaluate the performance of robust estimators (Künsch *et al.*, 1989; Croux & Haesbroeck, 2003). The χ^2_{arc} are defined as follows:

$$\chi^2_{arc} = \sum_{i=1}^{n} 4\left[\arcsin\sqrt{y_i} - \arcsin\sqrt{\pi_i}\right]^2; \quad i = 1,2,\ldots n \tag{16}$$

The arcsin transformation converts a Bernoulli random variable into one that is nearly normal and whose variance is slightly dependent on the parameter π_i. The arcsin is used to normalize the data in percentages or proportions whose distribution fits the Bernoulli distribution.

RESULTS AND DISCUSSIONS

In this study, the investigation was focused on the usefulness of the robust estimators on several well-known real data and simulation study.

The Prostate Cancer Data

First, the Prostate Cancer (PC) data given by Brown (1980) were taken into consideration. The data contain the values for two continuous variables, which are an elevated level of acid phosphates (AP) in the blood serum and the age of patients (AGE) that would be of value so as to predict whether or not PC patients also had lymph node involvement (LNI). The original data consisted of 53 patients and this was modified by adding two more outliers, namely, cases 54 $(y, x_1, x_2) = (0, 200, 67)$ and 55 $(y, x_1, x_2) = (0, 200, 68)$. The character plot of the PC data is presented in Figure 1 where AGE is plotted against AP and the character corresponding to occurrence $Y = 1$ and non-occurrence $Y = 0$ is denoted by symbols triangle and circle, respectively.

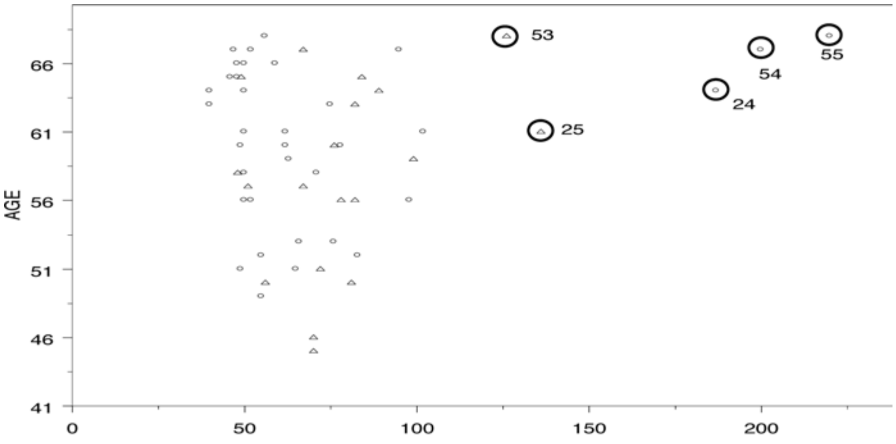

Fig.1: Scatter Plot of AGE vs AP with Outliers (Cases 24, 25, 53, 54, 55) for PC Data

It has been reported by Imon (2006) that the original data on the 53 patients may contain three outliers (cases 24, 25 and 53). Nonetheless, five outliers (*see* Fig. 1) were omitted from 55 observations to perform uncontaminated data.

The Neuralgia Data

Next, other data given by Piergorsch (1992) were considered. The data contain the values for two continuous variables, namely, the age of patients in completed years (AGE) and the pre-treatment duration of symptoms in month (DUR). There were 18 patients involved in this study and the outcome was whether the patients experienced any pain relief after the treatment. The character plot of the Neuralgia data is presented (*see* Fig.2) where DUR is plotted against AGE and the character corresponding to occurrence $Y = 1$ and non-occurrence $Y = 0$ is denoted by the triangle and circle, respectively.

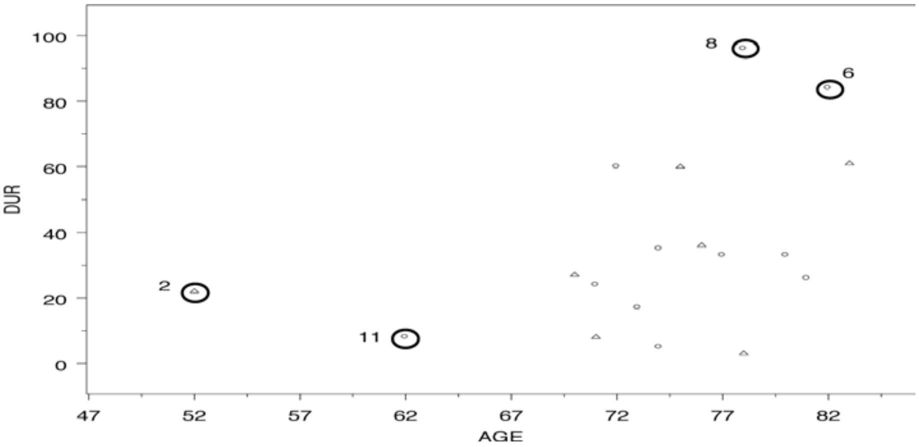

Fig. 2: Scatter Plot of DUR vs AGE with Outliers (Cases 2, 6, 8, 11) for Neuralgia Data

It is crucial to highlight that no identification of outliers for the Neuralgia data can be found in the literature. From the scatter plot (*see* Fig. 2), it pinpoints cases 2, 6, 8 and 11, as the outlying points. Therefore, the four suspected outliers should be removed to perform uncontaminated data.

The Erythrocyte Sedimentation Rate Data

The final data in the current study were the Erythrocyte Sedimentation Rate (ESR) data. In this case, the main objective was to see whether the levels of two plasma proteins (namely, Fibrinogen and γ.Globulin) in the blood plasma would be the factor increasing the ESR for healthy individuals. The study was carried out by the Institute of Medical Research, in Kuala Lumpur, Malaysia, involving 32 patients and the original data were collected by Collett and Jemain (1985). The responses of zero signify a healthy individual while the responses of unity refer to an unhealthy individual. Here, the continuous variables are (FIB and γ.GLO) versus the binary response of ESR. The character plot of the ESR data is presented in Fig.3 where γ.GLO is plotted against FIB and the character corresponding to occurrence $Y = 1$ and non-occurrence $Y = 0$ is denoted by the triangle and circle, respectively.

Syaiba and Habshah (2010) identified two outliers (namely, cases 13 and 29) in X-space for the original ESR data. As illustrated in Fig 3, it was observed that cases 14 and 15 are influential observations. Therefore, deleting cases 14 and 15 would create non-overlapping cases. In order to perform uncontaminated data, one more overlapping case was added by modifying case 13 with $(y, x_1, x_2) = (3.06, 37)$. From the uncontaminated data, the ESR data were contaminated where the occurrences ($Y = 1$) and non-occurrences ($Y = 0$) were replaced

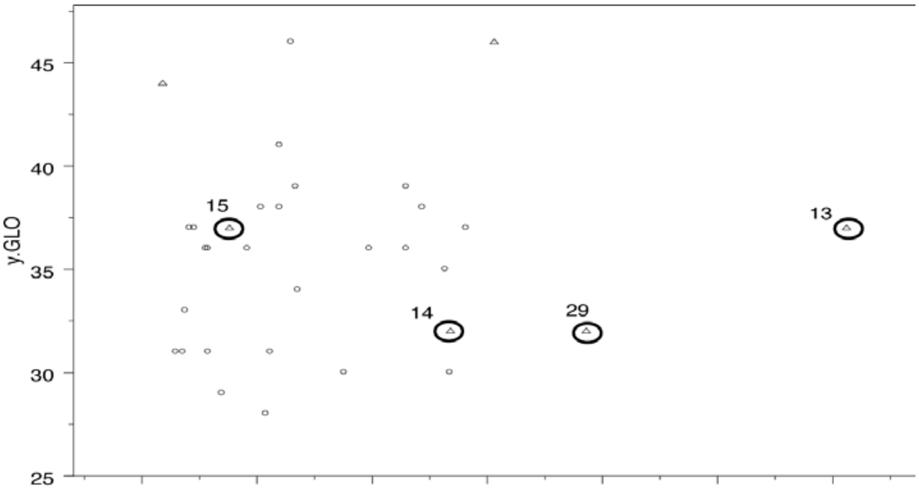

Fig.3: Scatter Plot of γ.GLO vs FIB with Outliers (Cases 13, 14, 15 and 29) for ESR Data

with each other for cases 14 and 15, and this might only leave one out of the three overlapping cases for the ESR data.

Monte Carlo Simulation Study

A simulation study was conducted to further assess the performance of the MLE and robust estimators. The evaluations focused on the severity of the outliers and also the number of observations by adding the outliers to the uncontaminated data. Following the work by Croux and Haesbroeck (2003), three different types of data were considered, and these are uncontaminated (Type 1), 5% moderate contaminated (Type 2), and 5% extreme contaminated (Type 3). The explanatory variables for the uncontaminated data (Type 1) were generated according to a standard normal distribution, $x_1 \sim N(0,1)$ and $x_2 \sim N(0,1)$, with four different numbers of observations, $n = (100, 200, 300, 400)$. The choice of a larger sample size was to ensure the existence of the overlapping cases in each replication. As pointed out by Victoria-Feser (2002), small data may lead to unidentifiable parameter estimates for no overlapping cases even without contamination. According to Victoria-Feser (2002), in practice, the number of observations with n = 50 is considered to be small. Thus, setting the true parameters as $\beta = (\beta_0, \beta_1, \beta_2)^T = (0.5, 1, -1)^T$ and the responses are defined as the following model equations:

$$y_i = \begin{cases} 0 & \text{if} \quad \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon_i < 0 \\ 1 & \text{if} \quad \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon_i \geq 0 \end{cases} \tag{17}$$

where the error terms were generated according to a logistic distribution, $\varepsilon_i \sim \Lambda(0,1)$. The explanatory variables for the contaminated data were generated according to the standard normal distributions, $z_1 \sim N(0,1)$ and $z_2 \sim N(0,1)$. In addition, the percentage of contamination denoted as s was also considered, as such that s = (5%) with magnitude of outlying shift distance in X-space for Type 2 and Type 3 taken as $\delta = 5$ and $\delta = 10$ respectively. The new x

values are defined as $x_1^* = z_1 + \delta$ and $x_2^* = z_2 - \delta$ and the responses are defined as they are in the following model equations:

$$y_i^* = \begin{cases} 0 & \text{if} \quad \beta_0 + \beta_1 x_1^* + \beta_2 x_2^* + \varepsilon_i \geq 0 \\ 1 & \text{if} \quad \beta_0 + \beta_1 x_1^* + \beta_2 x_2^* + \varepsilon_i < 0 \end{cases} \tag{18}$$

The performance of the estimators of MLE, CUBIF, MALLOWS, BY and WBY was evaluated based on the summary measures combining the individual estimated coefficients over $M = 1000$ replications. Therefore, BIAS and Root Mean Squared Error (RMSE) measures are computed as follows:

$$BIAS = \left\| \frac{1}{M} \sum_{i=1}^{M} \hat{\beta}_i^{(k)} - \beta_i^t \right\| \quad \text{and} \qquad RMSE = \sqrt{\frac{1}{M} \sum_{i=1}^{M} \left\| \hat{\beta}_i^{(k)} - \beta_i^t \right\|^2}$$

for $k = 1, 2, ..., M$ and $i = 1, 2, ..., p$, where $\| \cdot \|$ indicates the Euclidean norm.

RESULTS AND DISCUSSION

A “good” estimator is the one that has parameter estimates fairly close to the MLE estimates of the uncontaminated data. The second criterion is based on the goodness of fit test for the estimator which has the smallest value of χ^2_{arc}. Nevertheless, the complete tables of estimated coefficients, standard errors, and goodness of fit test could not be attached for the uncontaminated data of each real example due to space limitation in this paper. In general, the estimates and χ^2_{arc} values for the MLE, MALLOWS and CUBIF estimators are reasonably closer for the uncontaminated data. It was observed that for uncontaminated PC and Neuralgia data, the BY and WBY estimators gave different results for the parameter estimates when the outliers were omitted from the data. On deleting the outliers, the remaining data may have few overlapping cases, and thus, leaving the data in situation of quasi-complete separation. This is the reason why the BY and WBY estimators that downweight the outliers have larger estimated coefficients and standard errors compared to the MLE estimator. For the uncontaminated ESR data, the estimated coefficients of the BY and WBY estimators are slightly smaller compared to the MLE when the number of overlapping cases was increased.

Table 1: Estimated coefficients, standard errors, and the goodness of fit for PC (contaminated data)

		MLE^{uc}	MLE	MALLOWS	CUBIF	BY	WBY
Int.	β_0	-0.980	11.492	7.141	7.121	-1.912	-4.795
	$se(\beta_0)$	13.538	10.940	11.020	11.061	10.822	11.844
AP	β_1	3.031	1.141	1.663	1.749	1.112	2.913
	$se(\beta_1)$	1.376	0.781	0.817	0.826	0.770	1.030
	β_2	-3.004	-4.145	-3.607	-3.687	-0.817	-1.951
AGE	$se(\beta_2)$	2.936	2.735	2.744	2.751	2.653	2.805
	χ^2_{arc}	100.516	120.013	119.514	119.651	112.512	103.003

MLE^{uc} indicates the results for the uncontaminated PC data

Table 1 presents the comparison of the five estimators based on parameter estimates, standard errors and goodness-of-fit test for the contaminated PC data. Under the contaminated data, β_0 of all estimators is mostly affected by the outliers compared to other coefficients. The standard errors were found to be smaller but the χ^2_{arc} values increased compared to the uncontaminated data. The results presented in Table 1 indicate that the MLE is mostly influenced by the outliers. Among the robust estimators, the WBY is the most efficient estimator because it produces the lowest χ^2_{arc} value and its estimates are closer to the MLE of uncontaminated data.

Table 2: Estimated coefficients, standard errors, and the goodness of fit for Neuralgia (contaminated data)

		MLE^{uc}	MLE	MALLOWS	CUBIF	BY	WBY
Int.	β_0	-6.507	14.449	3.109	10.784	14.558	30.921
	$se(\beta_0)$	46.406	20.390	19.688	19.794	20.373	25.749
AGE	β_1	1.581	-3.312	-0.710	-2.492	-3.411	-7.480
	$se(\beta_1)$	10.798	4.836	4.674	4.698	4.836	6.133
	β_2	-0.196	-0.213	-0.200	-0.185	-0.123	0.209
DUR	$se(\beta_2)$	0.612	0.542	0.541	0.541	0.544	0.604
	χ^2_{arc}	33.687	41.680	40.371	40.816	40.750	38.338

MLE^{uc} indicates the results for the uncontaminated Neuralgia data

For the Neuralgia data with outliers (*see* Fig. 2), it is difficult to judge which estimator is the best by inspecting their parameter estimates. However, it is evident that the WBY is the best estimator as it has the smallest χ^2_{arc} value.

Table 3: Estimated coefficients, standard errors, and the goodness of fit for ESR (contaminated data)

		MLE^{uc}	MLE	MALLOWS	CUBIF	BY	WBY
Int.	β_0	12.263	19.882	20.449	20.579	14.231	20.441
	$se(\beta_0)$	5.839	9.417	9.809	10.031	7.099	10.381
FIB	β_1	1.830	2.597	2.648	3.053	1.791	2.572
	$se(\beta_1)$	1.062	1.543	1.611	1.681	1.308	1.877
	β_2	0.153	0.278	0.286	0.256	0.189	0.271
γ.GLO	$se(\beta_2)$	0.116	0.165	0.170	0.170	0.135	0.178
	χ^2_{arc}	42.237	27.050	26.628	26.047	27.805	25.724

MLE^{uc} indicates the results for the uncontaminated ESR data

Under the contaminated of the ESR data, β_0 and $se(\beta_0)$ of all the estimators are mostly affected by the outliers as compared to the other parameters (see Table 3). The results shown in Table 3 also indicate that the MLE is mostly influenced by the outliers. On modifying the contaminated data, there is only one overlapping observation, case 13 remains. This is the reason why the WBY that downweight this observation has large coefficients and standard errors. Even though the WBY has the smallest χ^2_{arc} value, the BY estimator should also be taken into consideration. The results illustrated in Tables 3 signify that the BY is a good estimator for the ESR data as its estimates are fairly closer to the MLE for the uncontaminated data.

Tables for the results of the summary measures consist BIAS and RMSE, whereby the first row indicates that the computation does not include the intercept term and second row indicates the computation including the intercept term. A "good" estimator is the one that has BIAS and RMSE, which are relatively small or closest to zero.

For the uncontaminated data shown in Table 4, the estimators of MLE, MALLOWS, CUBIF, BY and WBY behave not too differently. It can be seen that the BIAS and RMSE will reduce when the number of observations is increased. Under 5% of the intermediate contamination (*see* Table 5), the WBY estimator performs best in term of BIAS and RMSE. Meanwhile, the weighting step in the WBY estimator becomes more advantageous in the extreme contamination (*see* Table 6). However, the MLE estimator behaves very poorly in

Table 4: BIAS and RMSE of the estimators (Type 1)

	n = 100		*n* = 200		*n* = 300		*n* = 400	
	BIAS	RMSE	BIAS	RMSE	BIAS	RMSE	BIAS	RMSE
MLE	0.085	0.451	0.040	0.293	0.025	0.235	0.020	0.203
	0.086	0.522	0.040	0.340	0.026	0.273	0.021	0.238
MALLOWS	0.082	0.451	0.038	0.293	0.023	0.235	0.018	0.203
	0.082	0.522	0.038	0.341	0.024	0.274	0.018	0.238
CUBIF	0.083	0.452	0.040	0.294	0.025	0.235	0.020	0.203
	0.084	0.522	0.041	0.341	0.026	0.274	0.020	0.238
BY	0.094	0.472	0.045	0.304	0.026	0.240	0.022	0.207
	0.095	0.544	0.045	0.352	0.027	0.279	0.023	0.242
WBY	0.096	0.498	0.046	0.315	0.028	0.249	0.022	0.215
	0.097	0.569	0.047	0.363	0.289	0.287	0.023	0.249

Table 5: Bias and RMSE of the estimators (Type 2)

	n = 100		*n* = 200		*n* = 300		*n* = 400	
	BIAS	RMSE	BIAS	RMSE	BIAS	RMSE	BIAS	RMSE
MLE	0.684	0.747	0.698	0.723	0.706	0.724	0.704	0.719
	0.741	0.824	0.754	0.792	0.759	0.784	0.758	0.777
MALLOWS	0.615	0.684	0.636	0.668	0.645	0.664	0.645	0.661
	0.666	0.758	0.687	0.728	0.693	0.720	0.694	0.715
CUBIF	0.586	0.660	0.605	0.639	0.613	0.634	0.614	0.630
	0.639	0.736	0.658	0.702	0.663	0.691	0.664	0.686
BY	0.492	0.604	0.512	0.562	0.521	0.553	0.521	0.546
	0.537	0.675	0.556	0.619	0.563	0.603	0.563	0.595
WBY	0.281	0.504	0.319	0.417	0.336	0.401	0.342	0.389
	0.318	0.576	0.354	0.472	0.368	0.446	0.375	0.434

the contamination data. There are some losses in the precision (increased RMSE when BIAS is small) for the estimator based on weighting step**.** The intercept coefficient is more affected in the contaminated data, and consequently, the BIAS is larger compared to slope coefficients.

Table 6: BIAS and RMSE of the estimators (Type 3)

	n = 100		*n* = 200		*n* = 300		*n* = 400	
	BIAS	RMSE	BIAS	RMSE	BIAS	RMSE	BIAS	RMSE
MLE	1.352	1.369	1.352	1.361	1.354	1.359	1.353	1.357
	1.378	1.410	1.379	1.395	1.380	1.390	1.379	1.387
MALLOWS	0.386	0.548	0.475	0.537	0.501	0.541	0.516	0.544
	0.398	0.600	0.488	0.569	0.513	0.566	0.529	0.566
CUBIF	0.733	0.784	0.746	0.770	0.753	0.769	0.753	0.764
	0.752	0.826	0.765	0.800	0.771	0.794	0.771	0.788
BY	0.852	1.050	0.904	1.029	0.921	1.018	0.916	1.005
	0.872	1.090	0.926	1.061	0.943	1.046	0.939	1.032
WBY	0.096	0.492	0.046	0.313	0.028	0.247	0.021	0.213
	0.097	0.563	0.047	0.360	0.029	0.286	0.022	0.248

CONCLUSIONS

The purpose of this analysis was to compare the performance of the MLE and four robust estimators under contaminated and uncontaminated data. The results showed that the MLE estimator is severely affected by the presence of outliers. Among the robust estimators, the WBY estimator produced the smallest BIAS and RMSE in the contaminated data and their estimates are closer to the MLE for the uncontaminated data, followed by MALLOWS, BY and CUBIF. Therefore, it can be concluded that the WBY estimators perform better compared to the MLE estimator and the rest of the robust estimators in the presence of outliers. The results from the real data indicate that the WBY estimator produced the smallest χ^2_{arc} in the presence of outliers even though its estimates are slightly difference from the MLE estimator in the uncontaminated data due to quasi-complete separation. To protect against the outliers, weighting the covariates is effective. The weighting step can be seen as a way of the uncontaminated in the data before the estimation procedure.

REFERENCES

Albert, A., & Anderson, J. A. (1984). On the existence of maximum likelihood estimates in logistic regression. *Biometrika, 71*, 1-10.

Bianco, A. M., & Yohai, V. J. (1996). Robust estimation in the logistic regression model. Robust statistics, data analysis and computer intensive methods. *Proceedings of the Workshop in Honor of Huber, P.J. and Rieder, H. Lecturer notes in statistics.* Springer, New York, *109*, 17-34.

Brown, B. W. (1980). Prediction analysis for binary data. In R. G. Miller, B. Efron, B. W. Brown, and L. E. Moses (Eds.). *Biostatistics Casebook* (p. 3-18). New York: John Wiley and Sons, Inc.

Carroll, R. J., & Pederson, S. (1993). On robustness in the logistic regression model. *Journal of Royal Statistics Society B, 55*, 693-706.

Collett, D. (2003). *Modelling binary data* (2nd ed). London: Chapman & Hall.

Collett, D., & Jemain, A. A. (1985). Residuals, outliers and influential observations in regression analysis. *Sains Malaysiana, 14*, 493-511.

Cox, D. R., & Wermuth, N. (1992). A comment on the coefficient of determination for binary response. *American Statisticians, 46,* 1-4.

Croux, C., Flandre, C., & Haesbroeck, G. (2002). The breakdown behaviour of the maximum likelihood estimator in the logistic regression model. *Statistics and Probability Letters, 60,* 377-386.

Croux, C., & Haesbroeck, G. (2003). Implementing the bianco and yohai estimator for logistic regression. *Computational Statistics and Data Analysis, 44,* 273-295.

Hao, Y. (1992). *Maximum median likelihood and maximum trimmed likelihood estimations.* Published Doctoral Dissertation, University of Toronto, Canada.

Imon, A. H. M. R. (2006). Identification of high leverage points in logistic regression. *Pakistan Journal of Statistics, 22,* 147-156.

Imon, A. H. M. R., & Hadi, A. S. (2008). Identification of multiple outliers in logistic regression. *Communication in Statistics – Theory and Methods, 37,* 1697-1709.

Kordzakhia, N., Mishra, G. D., & Reiersølmoen, L. (2001). Robust estimation in logistic regression model. *Journal of Statistical Planning and Inference*, *98,* 211-223.

Künsch, H. R., Stefanski, L. A., & Carroll, R. J. (1989). Conditionally unbiased bounded influence estimation in general regression models with applications to generalized linear models. *Journal of the American Statistical Association, 84*, 460-466.

Nurunnabi, A. A. M., Imon, A. H. M. R., & Nasser, M. (2009). *Identification of multiple influential observations in logistic regression*. Philadelphia: Taylor and Francis, Inc.

Piegorsch, W. W. (1992). Complementary log regression for generalized linear models. *American Statisticians*, *46*, 94-99.

Pregibon, D. (1981). Logistic regression diagnostics. *Annals of Statistics, 9*, 705-724.

Rousseeuw P. J., & Leroy, M. (1987). *Robust regression and outlier detection* (p. 216-247). New York: John Wiley and Sons, Inc.

Santner, T. J., & Duffy, D. E. (1986). A note on A. Albert's and J.A. Anderson's conditions for the existence of maximum likelihood estimates in logistic regression models. *Biometrika, 73,* 755-758.

Sarkar, S. K., Midi, H., & Rana, S. (2011). Detection of outliers and influential observations in binary logistic regression: An empirical study. *Journal of Applied Sciences, 11,* 26-35.

Syaiba, B. A., & Habshah, H. (2010). Robust logistic diagnostic for the identification of high leverage points in logistic regression model. *Journal of Applied Sciences*, *10,* 3042-3050.

Victoria-Feser, M-P. (2002). Robust inference with binary data. *Psychometrika, 67*, 21-32.