Sigma J Eng Nat Sci, Vol. 42, No. 5, pp. 1344−1356, October, 2024

Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2023.00108

Research Article

Robust methods for detecting bad leverage point in logistic regression

Ebru GÜNDOĞAN AŞIK[1,*], Arzu ALTIN YAVUZ[2], Zafer KÜÇÜK[1]

[1]*Department of Statistics and Computer Science, Karadeniz Technical University, Trabzon, 61080, Türkiye*
[2]*Department of Statistics, Eskişehir Osmangazi University, Eskişehir, 26040, Türkiye*

ARTICLE INFO

Article history
Received: 07 March 2023
Accepted: 12 July 2023

Keywords:
Bad Leverage; High-Leverage; Logistic Regression; Robust Cut-Off Value; Robust Deviance Residuals

ABSTRACT

High-leverage points, known as good and bad leverage points, are also known as points away from center of x space. Bad leverage points are marginal values that show the incompatibility with misclassified observations and other observation values at x space. In the identification of bad leverage points, the problems of masking and swamping constitute a problem for the logistic regression model just as in the linear regression model. In this research, in addition to existing deviance components (DEVC), robust deviance components (RobDEVC) that are used to identify bad leverage points, different robust methods recommended to be used at the management of deviance components were examined. Also, for these methods, robust cut-off value combinations were examined as well. With the conducted simulation, robust methods recommended to be used in the deviance component method have shown better performance to identify bad leverage points by showing different cut-off values.

Cite this article as: Gündoğan Aşık E, Altın Yavuz A, Küçük Z. Robust methods for detecting bad leverage point in logistic regression. Sigma J Eng Nat Sci 2024;42(5):1344−1356.

INTRODUCTION

The logistic regression model is a model that is widely used in many areas. Unusual observations are examined together in the logistic regression model, just as in the linear regression model. Regression diagnostics methods are used to identify unusual observations (outlier, high-leverage point, influential observation) in the model. Observations with large residues are named as an outlier. The observations that largely change calculated various statistics once removed from the dataset are called influential observation [1]. High-leverage points are observations, remote from the average of the independent variable. High-leverage points are split into two groups known as good and bad leverage points. Good leverage points (GLP) are the remote observations in the independent variable that contribute to the parameter estimation. Bad leverage points (BLP) are defined as the influential observations of the independent variable that are incompatible with the majority of the data. The presence of high leverage points has a significant effect on the parameter estimates for the logistic regression model. The impact from high leverage points is more severe than other bad points. It has been stated that high leverage points are not only responsible for obtaining incorrect parameter estimates but may also cause other problems such as masking.

In the logistic regression model, to identify high-leverage points Distance from the Mean (DM), Generalized Weights (GW), Deviance Component (DEVC) and Robust Logistic

***Corresponding author.**
*E-mail address: ebrugundogan@ktu.edu.tr

This paper was recommended for publication in revised form by Editor-in-Chief Ahmet Selim Dalkilic

Diagnostic (RLGD) methods are used [2–5]. High-leverage points can also be identified with Robust Mahalanobis Distance (RMD) which is calculated by Minimum covariance determinant (MCD) or minimum volume ellipsoid (MVE). Another method used to identify high-leverage points is Deviance Components (DEVC) method. Deviation residues that exceed the cut-off value are determined as high-leverage points. In addition, the precise identification of high-leverage points of the cut-off value that is used in the DEVC method is important. The examined identification methods indiscriminately determine high-leverage points regardless of good or bad leverage points. In situations where a robust method is used for parameter estimation good leverage points will also have lower weights just as bad leverage points. There are many studies available in the literature for the optimization of the parameters [6–8]. Giving lower weight to good leverage points causes misinterpretations in parameter estimations. Parameter estimations done with Maximum Likelihood Estimations (MLE) would be more useful but they are not valid for bad leverage points [9–11]. In case when multiple bad leverage points were also affected by the masking and swamping problem has been proven by the conducted works [2,4,5]. To identify bad leverage points Robust Deviance Components (RobDEVC) method was developed in the logistic regression model [3]. In this developed method where deviation components are used, β parameters were obtained through the Mallows type leverage dependent weights estimator (Mallows). It has been observed that good predictions were made with the robust method used as a result of the simulation process. However, for the robust estimation of β parameters, usable estimators are not limited to Mallows.

In the studies available in the literature, the Median + 3MAD cut-off values were used in the diagnosis of bad leverage points. The authors suggested alternative robust values for the cut-off values [12]. In this study, a new performance indicator has been revealed by calculating the robust state of the residues. New RobDEVC diagnostic methods different from robust estimators were presented as an alternative to the Mallows estimators used for the estimation of β parameters in RobDEVC identification method. As an alternative to the Mallows estimator, the conditionally unbiased bounded influence function (CUBIF) estimator, Bianco and Yohai estimator (BY) and Weighted Bianco and Yohai estimator were covered. In the diagnosis methods, as the identifications of residues, accurate identification of cut-off values is important. Recommended as an alternative to the Median + 3MAD cut-off value in the existing literature by Gundogan et al., the performance of the methods used for robust cut-off value combinations' have been shown as an alternative in their works [12].

DEVC, ROBDEVC and New Robust Diagnostic Methods in Logistic Regression Model

The logistic regression model where the dependent variable obtains 0,1 value and has Bernoulli distribution is expressed as follows:

$$P(Y=1|X=x)=\pi_i=\frac{exp\ (x_i^T\beta)}{1+exp\ (x_i^T\beta)},\ \ i=1,2,\dots,n \quad (1)$$

where π_i represents fractile, i'th factor's probability and ranged in between $0 \le \pi_i \le 1$. Y, n × 1 dimension dependent variable vector; β, (p + 1) × 1 dimension unknown parameters vector; X, n × (p + 1) independent variable matrix and ε, n × 1 dimension error terms vector. For the estimation of unknown β parameters MLE is used. Probability and log-probability function is defined in order as follows:

$$L(\beta;y)=\prod_i^n \pi_i^{y_i}(1-\pi_i)^{1-y_i} \quad (2)$$

$$l(\beta;y)=\sum\nolimits_i^n [y_i \log(\pi_i)+(1-y_i)\log(1-\pi_i)]. \quad (3)$$

Once the derivative of β parameter is calculated in reference to the log-likelihood function and equalized to zero, MLE is obtained as a result of iterative solutions obtained from equations. $\hat{\pi}_i$ values are obtained by the usage of $\hat{\beta}$ parameter estimations.

The deviation (DEV) used in the logistic regression model is determined as:

$$DEV=\sum\nolimits_i^n d(y_i,\hat{\pi}_i)^2=\sum\nolimits_i^n 2\left[y_i\log\left(\frac{y_i}{\hat{\pi}_i}\right)+(1-y_i)\log\left(\frac{1-y_i}{1-\hat{\pi}_i}\right)\right]. \quad (4)$$

Here $d(y_i,\hat{\pi}_i)$ is known as the bias residuals and is an outlier detection method based on the differences of the deviations [13]. The deviation residues are obtained as follows:

$$d_i=sign(y_i-\hat{\pi}_i)\left\{2\left[y_i\log\left(\frac{y_i}{\hat{\pi}_i}\right)+(1-y_i)\log\left(\frac{1-y_i}{1-\hat{\pi}_i}\right)\right]\right\}^{1/2} \quad (5)$$

where $sign(y_i-\hat{\pi}_i)$ is the sign function that makes it positive or negative. *i'th* deviation residual component, and expressed as follows:

$$dc_i=\begin{cases}2\log\left(\frac{1}{1-\hat{\pi}_i}\right) & ,\quad y_i=0\\ 2\log\left(\frac{1}{\hat{\pi}_i}\right) & ,\quad y_i=1\end{cases}. \quad (6)$$

Whether the given cut-off value is greater than the obtained deviation residue component is checked. In the DEVC method, identification of the bad leverage points of β parameter estimations is done via MLE. Though, every observation of MLE sampling is given equal weight and affected by the outliers. This situation negatively affects parameter predictions and statistics calculated through these estimations do not reflect reality. For the determination of bad leverage points, identification of β parameters with robust estimators allows obtaining trustworthy results.

Therefore, through the effective estimation of β parameters, trustworthy estimated probabilities of $\hat{\pi}_i$ can be made. For the diagnosis of RobDEVC existing in the literature, estimation of β parameter was done via Mallows estimator. Mallows estimator that is developed by Künsch et al., and one of the generalized M estimators, reduces the weighted probability function dependent on common variables of weights to a bare minimum [14]. Mallows estimator that is based on conversion to an estimator with limited effect through the reduction of outliers in the X space of MLE, can be obtained with the following equation [15].

$$\sum_{i=1}^{n} w_i x_i [y_i \mathrm{F}(x_i\beta) - c(x_i\beta)] = 0 \tag{7}$$

Here w_i indicates weights dependent on observations and $c(x_i\beta)$ indicates rectification term. Counted as $w_i = w(x_i, x_i\beta)$ and $c(x_i, \beta) = 0$. Also, to weigh observations, RMD is used. Weight w_i, allows high-leverage observation to gain less weight from low leverage observation. Mallows estimators are sturdy against outliers, but they are not effective under the model.

Estimation probabilities obtained by using Mallows estimator in the diagnosis of RobDEVC method for the estimation of β parameter, are expressed as:

$$\hat{\pi}_{i(rob)} = \frac{exp\ (x_i^T \hat{\beta}_{rob})}{1 + exp\ (x_i^T \hat{\beta}_{rob})}. \tag{8}$$

Deviation residue obtained via the RobDEVC method is calculated as:

$$rdc_i = \begin{cases} 2\log\left(\dfrac{1}{1-\hat{\pi}_i^{rob}}\right), & y_i = 0 \\ 2\log\left(\dfrac{1}{\hat{\pi}_i^{rob}}\right), & y_i = 1 \end{cases} \tag{9}$$

and values exceeding the Median + 3MAD cut-off value are defined as bad leverage points [3].

In the RobDEVC diagnosis method, for β parameter estimation, Mallows estimator is being used. Although, in the analysis of robust logistic regression, for the estimation of β CUBIF, BY and WBY estimators are commonly used as well. In this research, at the RobDEVC diagnosis method, for the estimation of β parameter these estimators were examined and briefly explained below.

CUBIF Estimator: Is known as another generalized M estimator recommended for the logistic regression model. Estimator recommended by Künsch et al.,can be obtained by the following equation [14].

$$\sum_{i=1}^{n} w_i x_i [y_i \mathrm{F}(x_i\beta) - c(x_i\beta)] = 0 \tag{10}$$

Above, weights are defined as $w_i = w(x_i, x_i\beta, y_i)$ and created by the reflection of variables dependent on weights. CUBIF estimator has high model activity, but in cases where outliers are less, it is a sensitive estimator towards outliers.

Bianco and Yohai Estimatior (BY): In the obtaining of the sturdy estimator through changing of an objective function, first work is made by Pregibon [16]. This objective function is given in Equation (11).

$$\hat{\beta} = \arg\min_{\beta} \sum_{i=1}^{n} \lambda[d(x_i\beta, y_i)] \tag{11}$$

Here $\boldsymbol{\lambda}$ function is a Huber type function and $d(x_i\beta, y_i)$ is deviation components function. It has been observed that the results obtained via proposition of the objective function are not coherent with the influential observations in the data matrix. Following the objective function proposed by Pregibon, Bianco and Yohai have proposed the following Bianco and Yohai (BY) estimator in Equation (12) as follows [17].

$$\hat{\beta} = \arg\min_{\beta} \sum_{i=1}^{n} \rho\big[d(x_i\beta, y_i) + G\big(F(x_i\beta)\big) + G\big(1 - F(x_i\beta)\big)\big] \tag{12}$$

Here

$$G(t) = \int_0^t \rho'(-\ln u)\, du. \tag{13}$$

is expressed as and defined ρ function as:

$$\rho(t) = \begin{cases} t - \dfrac{t^2}{2c}, & t \le c \\ \dfrac{c}{2}, & otherwise \end{cases} \tag{14}$$

where c is the positive tuning parameter.

Weighted Bianco and Yohai Estimator (WBY): To make BY estimator sturdier Croux and Haesbroeck have weighted BY estimator again and proposed Weighted Bianco and Yohai estimator [18]. WBY estimator is defined as follows:

$$\hat{\beta} = \arg\min_{\beta} \sum_{i=1}^{n} w(x_i)\{\rho\big[D(x_i\beta, y_i) + G\big(F(x_i\beta)\big) + G\big(1 - F(x_i\beta)\big)\big]\}. \tag{15}$$

Here the weight $w(x_i)$, become RMD's decreasing function MCD estimator is used to calculate distances as follows [19].

$$w(x_i) = \begin{cases} 1, & RMD_i^2 \le \chi^2_{(p,0.975)} \\ 0, & otherwise \end{cases} \tag{16}$$

With the help of these estimators explained above and examined as an alternative to the Mallows estimator, for the robust estimation of β parameters, the estimation predictions are calculated in Equation (8). Via Equation (9) robust deviation residues used CUBIF, BY and WBY estimators are obtained. In cases where these residue values traditionally exceed Median + 3MAD cut-off value, they are named as bad leverage points.

In this research, on the analysis of logistic regression β if CUBIF estimator is used to estimate, obtained diagnosis method is RobDEVC$_1$, if BY estimator is used RobDEVC$_2$ and if WBY estimator is used named as RobDEVC$_3$ respectively.

Robust Estimators in Identifying Cut-Off Values

In DEVC and RobDEVC methods, to determine bad leverage points Median + 3MAD cut-off value is used. This cut-off value used consists of location and scale parameters. The location parameter expresses the common position of the data array. The scale parameter expresses the measurable values or the prevalence of the variable at a central point. Once the literature is examined, despite there being many studies where residue calculation methods were improved, there is no change in the calculation of cut-off values. However, the precise identification of a cut-off value for the diagnosis of bad leverage points is influential in the performance of diagnostic methods. Thus, new robust cut-off values that were utilized by the robust estimators in determining cut-off values were used [12]. As an alternative to Median + 3MAD cut-off value location parameter, from M estimators Huber, Andrews, Tukey Bisquare Hampel; from L estimators Median, Trimmed Winsorize; and from R estimators Hodges-Lehman estimators were used. As an alternative to the scale estimator MAD, Q_n and S_n estimators were used. Used estimators for location and scale parameters in the creation of cut-off values were given in Table 1.

Table 1. Location and scale estimator for cut-off values

Location estimator		
	Influence function	**Tuning constant**
Huber	$\psi(x)=\begin{cases} x & ,\lvert x\rvert \le c \\ c\, sgn(x) & ,\lvert x\rvert > c \end{cases}$	c = 1.345
Andrews	$\psi(x)=\begin{cases} \sin(x/c) & ,\lvert x\rvert \le c\pi \\ 0 & ,\lvert x\rvert > c\pi \end{cases}$	c = 1.339
Tukey bisquare	$\psi(x)=\begin{cases} x\left[1-\left(\frac{x}{c}\right)^2\right]^2 & ,\lvert x\rvert \le c \\ 0 & ,\lvert x\rvert > c \end{cases}$	c = 4.685
Hampel	$\psi(x)=\begin{cases} x & ,\lvert x\rvert \le a \\ a\, sgn(x) & ,a<\lvert x\rvert \le b \\ a\dfrac{c-\lvert x\rvert}{c-b} sgn(x) & ,b<\lvert x\rvert \le c \\ 0 & ,\lvert x\rvert > c \end{cases}$ $0<a\le b\le c$	c = 8.5
Median	$Medyan=\begin{cases} x_{\left(\frac{n+1}{2}\right)} & , \quad if\ n\ is\ odd\ number \\ \dfrac{x_{(n/2)}+x_{(n/2+1)}}{2} & , \quad if\ n\ is\ even\ number \end{cases}$	
Trimmed	$T_n=T_n(l_n,u_n)=\dfrac{1}{u_n-l_n}\sum_{i=l_n+1}^{u_n} x_{(i)}$	
Winsorize	$W_n=W_n(l_n,u_n)=\dfrac{1}{n}\left\{l_n x_{(l_n+1)}+\sum_{i=l_n+1}^{u_n} x_{(i)}+(n-u_n)x_{(u_n)}\right\}$	
Hodges-Lehmann	$\hat{\theta}_{HL}=median\left\{\dfrac{x_{(i)}+x_{(j)}}{2}\right\}, \quad 1\le i<j\le n$	
Scale Estimator		
MAD	$MAD=\dfrac{median(\lvert x_i-median(x_i)\rvert)}{0.6745}$	
S_n	$S_n=c_n 1.1926\, median_i\{median_j\lvert x_i-x_j\rvert\}, \quad i,j=1,2,\dots,n$	
Q_n	$Q_n=d_n 2.2219\{\lvert x_i-x_j\rvert; i<j\}_{(k)}, \quad i,j=1,2,\dots,n$	

Table 2. New cut-off values

CT:	$Medyan + 3MAD$	CT8:	$Hampel + 3Q_n$	CT16:	$Trimmed + 3S_n$
CT1:	$Medyan + 3S_n$	CT9:	$Andrew + 3MAD$	CT17:	$Trimmed + 3Q_n$
CT2:	$Medyan + 3Q_n$	CT10:	$Andrew + 3S_n$	CT18:	$Tukey + 3MAD$
CT3:	$Huber + 3MAD$	CT11:	$Andrew + 3Q_n$	CT19:	$Tukey + 3S_n$
CT4:	$Huber + 3S_n$	CT12:	$Winsorize + 3MAD$	CT20:	$Tukey + 3Q_n$
CT5:	$Huber + 3Q_n$	CT13:	$Winsorize + 3S_n$	CT21:	$HL + 3MAD$
CT6:	$Hampel + 3MAD$	CT14:	$Winsorize + 3Q_n$	CT22:	$HL + 3S_n$
CT7:	$Hampel + 3S_n$	CT15:	$Trimmed + 3MAD$	CT23:	$HL + 3Q_n$

CT: Cut-off value, CT1-23: New cut-off values

In Table 1 $x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}$ indicates order statistics, l_n=[nα] indicates observation numbers dropped from the low end in the ordered observations, u_n=[nβ] shows the observation numbers dropped from the higher end. [·] operator denotes the largest integer function. Is generally taken as $0 \leq \alpha \leq 0.25$, $1 - \beta \leq 0.25$ respectively [20–22]. When the last rows of Table 1 are examined, the denominator invariant in the calculation of the MAD estimator is used to make it consistent with the normal standard deviation of the scale's estimation. c_n and 1.1926 invariants S_n estimator, d_n and 2.2219 invariant Q_n allow the estimator to be an unbiased estimator. Also $k = \binom{h}{2} \approx \binom{n}{2}/4$ and h value is $\left(\frac{n}{2} + 1\right)$.

The proposed new cut-off values with different combinations of robust estimators of location and scale parameters are shown in Table 2.

Simulation Study

In this section with the Monte Carlo simulation, robust methods were emphasized to identify bad leverage points for the logistic regression model. As an alternative to DEVC and RobDEVC methods existing in the literature, $RobDEVC_1$, $RobDEVC_2$ and $RobDEVC_3$ are a simulation order to compare the identification of bad leverage points of robust methods. Also in the diagnosis of bad leverage points, a comparison of using robust cut-off values and different cut-off value combinations was examined as well. Correct Identification Ratio (CIR) was used to identify the suggested effectiveness of success at bad leverage points and these methods at which cut-off values were identified bad leverage points accurately. Also, the calculation of the Swamping ratio (SR) identifies how many good observations are determined as bad observations. CIR represents the ratio of accurately defined problematic observation to total actual problematic observation in the data and, SR indicates the ratio of problematic observations defined as good observation number to the total good observation.

In the work independent variable number is taken as p = 2,3. Sampling size is n = 80,120. It is created as dependent variable from different independent variable numbers and starting parameter values in the following equation:

$$y_i = \begin{cases} 0, & \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + \varepsilon_i < 0\,, \quad i = 1, 2, \ldots, n, \quad x_i \sim N(0,1) \\ 1, & \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + \varepsilon_i \geq 0\,, \quad \varepsilon_i \sim \Lambda(0,1), \quad \beta_i = (1, \ldots, 1) \end{cases} \tag{17}$$

where p is an independent variable number and ε is the error terms obtained from logistic distribution.

Different percentages of high-leverage points are identified as α = 4, 10 and the dataset is created as $(\alpha/2)$% good and $(\alpha/2)$% bad leverage points. y values equaled to high-leverage points produced from normal distribution [2,4]; is arranged for good leverage points as y = 1, for bad leverage points as y = 0.

For simulation work, R program was used. For all combinations of the independent variable number, sampling size and bad leverage point 10000 trials were made. Obtained results are given in Table 3-10.

· ***When independent variable number is two;***

If there are two bad leverage points in the dataset, for n = 80 in Table 3 is examined instead of classically recommended DEVC and RobDEVC diagnosis methods $RobDEVC_1$ $RobDEVC_2$ and $RobDEVC_3$ methods recommended in our work shows a higher rate of success in all criteria of diagnosis methods location and scale parameters. Among the recommended diagnosis methods, $RobDEVC_3$ method is more effective compared to other methods. Identifying bad leverage points accurately at which cut-off values in the following methods can be told by examining high CIR values. While in the RobDEVC method Andrew + 3Sn has a high CIR value of identifying cut- off value bad leverage points, other diagnosing methods have high CIR value at Andrew + $3Q_n$ cut-off value. Also, SR values vary from location to scaling parameters and generally obtained lower values in Huber location parameters. For n = 120, when Table 4 is examined recommended diagnosis methods show better performance in all criteria. The effectiveness of $RobDEVC_2$ and $RobDEVC_3$ diagnosis methods at Andrew + $3Q_n$ bad leverage points at cut-off value can be seen by the high CIR value.

If there are five bad leverage points in the dataset once Table 5 is examined, for n = 80 $RobDEVC_1$ $RobDEVC_2$ and $RobDEVC_3$ methods are more successful compared to the classical methods. DEVC and RobDEVC methods have

Table 3. Simulation results for $p = 2$, $n = 80$ and bad leverage point 2

		DEVC		RobDEVC		RobDEVC_1		RobDEVC_2		RobDEVC_3	
		CIR	SR	CIR	SR	CIR	SR	CIR	SR	CIR	SR
Mad	Tukey Biq.	0.9483	0.0269	0.9701	0.0327	0.9798	0.0389	0.9848	0.0577	0.9850	0.0577
	Huber	0.8470	0.0031	0.8944	0.0073	0.9211	0.0117	0.9496	0.0267	0.9498	0.0268
	Humpel	0.8976	0.0122	0.9375	0.0180	0.9551	0.0240	0.9706	0.0438	0.9705	0.0438
	Andrew	0.9567	0.0326	0.9768	0.0389	0.9841	0.0449	0.9877	0.0632	0.9878	0.0632
	Winsor	0.8603	0.0043	0.9047	0.0084	0.9302	0.0127	0.956	0.0266	0.9561	0.0266
	Trimmed	0.8546	0.0032	0.8999	0.0073	0.9266	0.0117	0.9533	0.0255	0.9533	0.0256
	Medyan	0.8997	0.0135	0.9363	0.0184	0.9522	0.0238	0.9692	0.0414	0.9693	0.0415
	HodgesLeh.	0.8570	0.0040	0.8998	0.0080	0.9268	0.0125	0.9534	0.0265	0.9535	0.0265
$\mathbf{S_n}$	Tukey Biq.	0.9488	0.0271	0.9850	0.0341	0.9808	0.0420	0.9851	0.0676	0.9852	0.0675
	Huber	0.8459	0.0032	0.9522	0.0083	0.9254	0.0141	0.9522	0.0349	0.9524	0.0350
	Humpel	0.9024	0.0127	0.9738	0.0196	0.9605	0.0273	0.9738	0.0530	0.9739	0.0530
	Andrew	0.9564	0.0327	0.9870	0.0402	0.9836	0.0480	0.9874	0.0734	0.9875	0.0734
	Winsor	0.8603	0.0045	0.9572	0.0095	0.9340	0.0154	0.9572	0.0349	0.9571	0.0349
	Trimmed	0.8541	0.0035	0.9541	0.0085	0.9291	0.0141	0.9541	0.0337	0.9544	0.0338
	Median	0.9056	0.0141	0.9719	0.0201	0.9570	0.0270	0.9719	0.0505	0.9718	0.0505
	HodgesLeh.	0.8550	0.0041	0.9549	0.0092	0.9298	0.0149	0.9549	0.0347	0.9548	0.0347
$\mathbf{Q_n}$	Tukey Biq.	0.9636	0.0357	0.9805	0.0433	0.9869	0.0509	0.9899	0.0720	0.9899	0.0719
	Huber	0.8827	0.0094	0.9254	0.0150	0.9475	0.0211	0.9656	0.0381	0.9655	0.0382
	Humpel	0.9268	0.0200	0.9564	0.0275	0.9710	0.0353	0.9791	0.0568	0.9791	0.0568
	Andrew	0.9691	0.0415	0.9837	0.0496	0.9901	0.0572	0.9912	0.0782	0.9913	0.0781
	Winsor	0.8921	0.0108	0.9327	0.0165	0.9523	0.0224	0.9678	0.0383	0.9677	0.0383
	Trimmed	0.8875	0.0096	0.9297	0.0153	0.9504	0.0211	0.9666	0.0371	0.9665	0.037
	Median	0.9291	0.0211	0.9574	0.0277	0.9687	0.0348	0.9800	0.0544	0.9800	0.0544
	HodgesLeh.	0.8880	0.0104	0.9298	0.0161	0.9505	0.0220	0.9673	0.0382	0.9671	0.0381

CIR: Correct identification rate; SR: Swamping rate

Table 4. Simulation results for $p = 2$, $n = 120$ and bad leverage point 2

		DEVC		RobDEVC		RobDEVC_1		RobDEVC_2		RobDEVC_3	
		CIR	SR	CIR	SR	CIR	SR	CIR	SR	CIR	SR
Mad	Tukey Biq.	0.9799	0.0444	0.9872	0.0491	0.9908	0.0543	0.9956	0.0705	0.9957	0.0703
	Huber	0.8883	0.0201	0.9157	0.0238	0.9372	0.0277	0.9633	0.0393	0.9632	0.0393
	Humpel	0.9465	0.0325	0.9642	0.0370	0.9761	0.0421	0.9872	0.0573	0.9872	0.0573
	Andrew	0.9863	0.0487	0.9912	0.0534	0.9951	0.0586	0.9979	0.0756	0.9979	0.0757
	Winsor	0.9051	0.0219	0.9313	0.0254	0.9488	0.0290	0.9708	0.0394	0.9707	0.0392
	Trimmed	0.9018	0.0214	0.9265	0.0249	0.9462	0.0285	0.9687	0.0389	0.9688	0.0387
	Median	0.9367	0.0319	0.9537	0.0360	0.9672	0.0405	0.9818	0.0544	0.9819	0.0545
	HodgesLeh.	0.8952	0.0203	0.9209	0.0237	0.9421	0.0273	0.9655	0.0378	0.9654	0.0378
$\mathbf{S_n}$	Tukey Biq.	0.9838	0.0501	0.9906	0.0566	0.9950	0.0642	0.9976	0.0875	0.9977	0.0875
	Huber	0.9184	0.0248	0.9409	0.0297	0.9600	0.0350	0.9782	0.0509	0.9780	0.0509
	Humpel	0.9615	0.0377	0.9782	0.0437	0.9863	0.0506	0.9925	0.0717	0.9926	0.0717
	Andrew	0.9878	0.0548	0.9934	0.0612	0.9961	0.0690	0.9985	0.0935	0.9985	0.0935
	Winsor	0.9326	0.0266	0.9536	0.0313	0.9687	0.0362	0.9828	0.0507	0.9828	0.0507
	Trimmed	0.9297	0.0261	0.9498	0.0307	0.9669	0.0357	0.9817	0.0502	0.9816	0.0502
	Median	0.9534	0.0369	0.9702	0.0425	0.9806	0.0489	0.9889	0.0683	0.9888	0.0683
	HodgesLeh.	0.9248	0.0250	0.9465	0.0297	0.9636	0.0346	0.9801	0.0489	0.9801	0.0489
$\mathbf{Q_n}$	Tukey Biq.	0.9982	0.0649	0.9994	0.0706	0.9998	0.0771	0.9999	0.0955	0.9999	0.0953
	Huber	0.9757	0.0377	0.9850	0.0413	0.9911	0.0452	0.9948	0.0567	0.9947	0.0567
	Humpel	0.9925	0.0513	0.9962	0.0562	0.9984	0.0619	0.9996	0.0783	0.9996	0.0780
	Andrew	0.9988	0.0701	0.9995	0.0759	0.9998	0.0826	1.0000	0.1019	1.0000	0.1017
	Winsor	0.9821	0.0395	0.9893	0.0429	0.9942	0.0465	0.9966	0.0567	0.9966	0.0567
	Trimmed	0.9800	0.0390	0.9874	0.0424	0.9934	0.0460	0.9965	0.0562	0.9966	0.0560
	Median	0.9882	0.0504	0.9943	0.0551	0.9968	0.0601	0.9986	0.0747	0.9987	0.0748
	HodgesLeh.	0.9787	0.0379	0.9866	0.0413	0.9926	0.0448	0.9957	0.0549	0.9958	0.0548

CIR: Correct identification rate; SR: Swamping rate

Table 5. Simulation results for $p = 2$, $n = 80$ and bad leverage point 5

		DEVC		RobDEVC		$RobDEVC_1$		$RobDEVC_2$		$RobDEVC_3$	
		CIR	SR	CIR	SR	CIR	SR	CIR	SR	CIR	SR
Mad	Tukey Biq.	0.4943	0.0791	0.5918	0.0697	0.6096	0.0676	0.7781	0.0469	0.7791	0.0470
	Huber	0.3516	0.0993	0.4412	0.0913	0.4580	0.0894	0.6730	0.0687	0.6718	0.0689
	Humpel	0.3901	0.0941	0.4864	0.0851	0.5060	0.0828	0.7130	0.0601	0.7115	0.0602
	Andrew	0.5122	0.0761	0.6128	0.0661	0.6308	0.0637	0.7942	0.0428	0.7929	0.0429
	Winsor	0.3644	0.0978	0.4548	0.0898	0.4725	0.0878	0.6838	0.0672	0.6826	0.0674
	Trimmed	0.3559	0.0988	0.4462	0.0910	0.4633	0.0890	0.6766	0.0687	0.6753	0.0689
	Median	0.4282	0.0887	0.5163	0.0805	0.5333	0.0786	0.7256	0.0581	0.7242	0.0581
	HodgesLeh.	0.3542	0.0991	0.4448	0.0911	0.4617	0.0892	0.6768	0.0683	0.6754	0.0684
$\mathbf{S_n}$	Tukey Biq.	0.4634	0.0827	0.5593	0.0736	0.5750	0.0719	0.7573	0.0501	0.7584	0.0502
	Huber	0.3264	0.1020	0.4122	0.0944	0.4274	0.0928	0.6465	0.0717	0.6456	0.0718
	Humpel	0.3604	0.0973	0.4540	0.0886	0.4719	0.0867	0.6880	0.0633	0.6867	0.0634
	Andrew	0.4826	0.0798	0.5818	0.0700	0.5986	0.0679	0.7728	0.0460	0.7741	0.0461
	Winsor	0.3374	0.1006	0.4244	0.0930	0.4410	0.0912	0.6570	0.0704	0.6556	0.0704
	Trimmed	0.3298	0.1016	0.4164	0.0940	0.4318	0.0924	0.6492	0.0718	0.6482	0.0719
	Median	0.3969	0.0915	0.4837	0.0838	0.4984	0.0822	0.6994	0.0613	0.6980	0.0614
	HodgesLeh.	0.3280	0.1019	0.4149	0.0942	0.4302	0.0926	0.6499	0.0714	0.6489	0.0714
$\mathbf{Q_n}$	Tukey Biq.	0.5852	0.0665	0.6730	0.0584	0.6883	0.0566	0.8276	0.0380	0.8291	0.0381
	Huber	0.4043	0.0910	0.5017	0.0829	0.5173	0.0813	0.7170	0.0617	0.7156	0.0618
	Humpel	0.4567	0.0841	0.5599	0.0753	0.5746	0.0735	0.7614	0.0523	0.7600	0.0524
	Andrew	0.6024	0.0633	0.6930	0.0545	0.7086	0.0526	0.8415	0.0337	0.8420	0.0338
	Winsor	0.4241	0.0888	0.5192	0.0808	0.5358	0.0791	0.7284	0.0600	0.7274	0.0601
	Trimmed	0.4132	0.0901	0.5096	0.0822	0.5262	0.0804	0.7218	0.0615	0.7205	0.0616
	Median	0.4889	0.0788	0.5812	0.0710	0.5961	0.0693	0.7713	0.0505	0.7691	0.0507
	HodgesLeh.	0.4093	0.0907	0.5070	0.0825	0.5225	0.0808	0.7205	0.0612	0.7193	0.0613

CIR: Correct identification rate; SR: Swamping rate

Table 6. Simulation results for $p = 2$, $n = 120$ and bad leverage point 5

		DEVC		RobDEVC		$RobDEVC_1$		$RobDEVC_2$		$RobDEVC_3$	
		CIR	SR	CIR	SR	CIR	SR	CIR	SR	CIR	SR
Mad	Tukey Biq.	0.5579	0.0475	0.6400	0.0413	0.7084	0.0353	0.8524	0.0160	0.8523	0.0161
	Huber	0.3914	0.0628	0.4734	0.0581	0.5463	0.0535	0.7206	0.0383	0.7200	0.0383
	Humpel	0.4546	0.0577	0.5442	0.0520	0.6200	0.0464	0.7914	0.0278	0.7911	0.0278
	Andrew	0.5752	0.0452	0.6597	0.0385	0.7286	0.0320	0.8668	0.0121	0.8667	0.0121
	Winsor	0.4091	0.0617	0.4931	0.0568	0.5648	0.0522	0.7365	0.0371	0.7360	0.0372
	Trimmed	0.4072	0.0618	0.4906	0.0570	0.5622	0.0524	0.7335	0.0375	0.7330	0.0375
	Median	0.4728	0.0553	0.5559	0.0499	0.6263	0.0445	0.7853	0.0275	0.7848	0.0275
	HodgesLeh.	0.3976	0.0624	0.4809	0.0577	0.5525	0.0532	0.7257	0.0383	0.7253	0.0383
$\mathbf{S_n}$	Tukey Biq.	0.5791	0.0461	0.6565	0.0399	0.7202	0.0338	0.8614	0.0104	0.8611	0.0105
	Huber	0.4032	0.0618	0.4894	0.0567	0.5628	0.0518	0.7417	0.0341	0.7414	0.0341
	Humpel	0.4711	0.0563	0.5602	0.0506	0.6365	0.0447	0.8051	0.0229	0.8047	0.0229
	Andrew	0.5952	0.0439	0.6735	0.0372	0.7401	0.0305	0.8737	0.0061	0.8734	0.0061
	Winsor	0.4227	0.0604	0.5083	0.0554	0.5806	0.0504	0.7571	0.0329	0.7567	0.0330
	Trimmed	0.4205	0.0606	0.5056	0.0556	0.5781	0.0507	0.7546	0.0333	0.7540	0.0333
	Median	0.4873	0.0541	0.5702	0.0486	0.6413	0.0428	0.8004	0.0225	0.8000	0.0225
	HodgesLeh.	0.4108	0.0613	0.4965	0.0563	0.5695	0.0514	0.7479	0.0341	0.7473	0.0342
$\mathbf{Q_n}$	Tukey Biq.	0.6740	0.0336	0.7441	0.0266	0.8113	0.0190	0.9204	0.0059	0.9202	0.0058
	Huber	0.5208	0.0526	0.6048	0.0470	0.6746	0.0414	0.8302	0.0232	0.8298	0.0232
	Humpel	0.5860	0.0458	0.6648	0.0396	0.7352	0.0327	0.8775	0.0098	0.8773	0.0098
	Andrew	0.6876	0.0309	0.7600	0.0232	0.8281	0.0148	0.9289	0.0110	0.9290	0.0109
	Winsor	0.5402	0.0509	0.6226	0.0453	0.6909	0.0396	0.8415	0.0218	0.8413	0.0218
	Trimmed	0.5383	0.0511	0.6206	0.0456	0.6882	0.0400	0.8395	0.0222	0.8393	0.0222
	Median	0.5992	0.0434	0.6741	0.0373	0.7413	0.0308	0.8753	0.0097	0.8752	0.0097
	HodgesLeh.	0.5292	0.0519	0.6120	0.0464	0.6813	0.0408	0.8340	0.0232	0.8338	0.0232

CIR: Correct identification rate; SR: Swamping rate

Table 7. Simulation results for p = 3, n = 80 and bad leverage point 2

		DEVC		RobDEVC		RobDEVC$_1$		RobDEVC$_2$		RobDEVC$_3$	
		CIR	SR	CIR	SR	CIR	SR	CIR	SR	CIR	SR
Mad	Tukey Biq.	0.8821	0.0063	0.9395	0.0140	0.9835	0.0328	0.9850	0.0825	0.9860	0.0828
	Huber	0.8040	0.0118	0.8880	0.0064	0.9615	0.0053	0.9800	0.0424	0.9810	0.0427
	Humpel	0.8445	0.0044	0.9185	0.0030	0.9755	0.0190	0.9830	0.0666	0.9840	0.0667
	Andrew	0.8950	0.0099	0.9485	0.0181	0.9865	0.0392	0.9865	0.0908	0.9875	0.0908
	Winsor	0.8135	0.0112	0.8925	0.0061	0.9635	0.0042	0.9810	0.0356	0.9820	0.0359
	Trimmed	0.8125	0.0112	0.8925	0.0063	0.9635	0.0043	0.9805	0.0363	0.9815	0.0366
	Median	0.8420	0.0031	0.9130	0.0044	0.9705	0.0194	0.9830	0.0623	0.9840	0.0624
	HodgesLeh.	0.8080	0.0121	0.8900	0.0069	0.9615	0.0034	0.9800	0.0347	0.9810	0.0350
S_n	Tukey Biq.	0.8995	0.0113	0.9520	0.0215	0.9875	0.0470	0.9885	0.1082	0.9895	0.1081
	Huber	0.8265	0.0086	0.9035	0.0018	0.9710	0.0141	0.9815	0.0590	0.9825	0.0590
	Humpel	0.8610	0.0005	0.9285	0.0085	0.9815	0.0300	0.9865	0.0892	0.9875	0.0894
	Andrew	0.9090	0.0154	0.9580	0.0263	0.9875	0.0541	0.9885	0.1173	0.9895	0.1173
	Winsor	0.8350	0.0080	0.9090	0.0015	0.9735	0.0132	0.9825	0.0519	0.9835	0.0519
	Trimmed	0.8320	0.0081	0.9075	0.0017	0.9735	0.0134	0.9825	0.0526	0.9835	0.0526
	Median	0.8630	0.0014	0.9215	0.0099	0.9795	0.0303	0.9865	0.0835	0.9875	0.0836
	HodgesLeh.	0.8275	0.0088	0.9060	0.0025	0.9715	0.0120	0.9820	0.0503	0.9830	0.0505
Q_n	Tukey Biq.	0.9345	0.0255	0.9695	0.0354	0.9840	0.0601	0.9885	0.1099	0.9895	0.1097
	Huber	0.8760	0.0004	0.9355	0.0060	0.9825	0.0199	0.9860	0.0561	0.9870	0.0561
	Humpel	0.9030	0.0104	0.9550	0.0187	0.9895	0.0394	0.9870	0.0891	0.9880	0.0890
	Andrew	0.9400	0.0299	0.9745	0.0410	0.9882	0.0689	0.9890	0.1205	0.9900	0.1203
	Winsor	0.8800	0.0015	0.9390	0.0068	0.9830	0.0192	0.9860	0.0492	0.9870	0.0493
	Trimmed	0.8790	0.0012	0.9395	0.0065	0.9830	0.0193	0.9860	0.0503	0.9870	0.0503
	Median	0.9055	0.0121	0.9530	0.0197	0.9875	0.0381	0.9870	0.0819	0.9880	0.0817
	HodgesLeh.	0.8770	0.0003	0.9360	0.0054	0.9825	0.0179	0.9860	0.0481	0.9870	0.0480

CIR: Correct identification rate; SR: Swamping rate

Table 8. Simulation results for p = 3, n = 120 and bad leverage point 2

		DEVC		RobDEVC		RobDEVC$_1$		RobDEVC$_2$		RobDEVC$_3$	
		CIR	SR	CIR	SR	CIR	SR	CIR	SR	CIR	SR
Mad	Tukey Biq.	0.9830	0.0483	0.9936	0.0580	0.9985	0.0786	0.9993	0.1174	0.9995	0.1174
	Huber	0.9550	0.0188	0.9766	0.0248	0.9927	0.0384	0.9967	0.0688	0.9967	0.0689
	Humpel	0.9747	0.0336	0.9889	0.0422	0.9970	0.0616	0.9990	0.1000	0.9991	0.1003
	Andrew	0.9862	0.0546	0.9946	0.0653	0.9993	0.0870	0.9997	0.1252	0.9997	0.1252
	Winsor	0.9588	0.0192	0.9786	0.0247	0.9931	0.0372	0.9970	0.0632	0.9971	0.0632
	Trimmed	0.9581	0.0189	0.9784	0.0245	0.9930	0.0370	0.9967	0.0631	0.9969	0.0633
	Median	0.9719	0.0326	0.9865	0.0403	0.9955	0.0576	0.9980	0.0925	0.9985	0.0922
	HodgesLeh.	0.9565	0.0179	0.9769	0.0233	0.9929	0.0353	0.9966	0.0607	0.9967	0.0606
S_n	Tukey Biq.	0.9879	0.0635	0.9961	0.0772	0.9994	0.1048	0.9996	0.1512	0.9997	0.1512
	Huber	0.9655	0.0283	0.9848	0.0370	0.9951	0.0562	0.9985	0.0931	0.9985	0.0933
	Humpel	0.9807	0.0462	0.9927	0.0585	0.9982	0.0847	0.9992	0.1309	0.9994	0.1310
	Andrew	0.9904	0.0709	0.9967	0.0856	0.9995	0.1147	0.9998	0.1602	0.9998	0.1602
	Winsor	0.9674	0.0288	0.9867	0.0373	0.9954	0.0548	0.9985	0.0871	0.9987	0.0871
	Trimmed	0.9672	0.0285	0.9866	0.0370	0.9953	0.0546	0.9983	0.0871	0.9986	0.0870
	Median	0.9775	0.0450	0.9913	0.0562	0.9976	0.0797	0.9991	0.1213	0.9993	0.1215
	HodgesLeh.	0.9657	0.0273	0.9853	0.0354	0.9951	0.0526	0.9984	0.0840	0.9985	0.0840
Q_n	Tukey Biq.	0.9937	0.0797	0.9981	0.0909	0.9998	0.1117	0.9999	0.1447	0.9999	0.1449
	Huber	0.9797	0.0380	0.9916	0.0453	0.9975	0.0600	0.9992	0.0872	0.9993	0.0871
	Humpel	0.9895	0.0592	0.9960	0.0697	0.9995	0.0901	0.9997	0.1248	0.9999	0.1247
	Andrew	0.9951	0.0881	0.9985	0.1002	0.9998	0.1221	1.0000	0.1541	1.0000	0.1541
	Winsor	0.9807	0.0392	0.9930	0.0460	0.9977	0.0591	0.9994	0.0816	0.9994	0.0814
	Trimmed	0.9808	0.0387	0.9931	0.0456	0.9977	0.0588	0.9993	0.0812	0.9993	0.0817
	Median	0.9875	0.0572	0.9954	0.0665	0.9994	0.0845	0.9998	0.1150	0.9997	0.1151
	HodgesLeh.	0.9796	0.0372	0.9922	0.0439	0.9975	0.0565	0.9991	0.0786	0.9993	0.0786

CIR: Correct identification rate; SR: Swamping rate

Table 9. Simulation results p = 3, n = 80 and bad leverage point 5

		DEVC		RobDEVC		RobDEVC$_1$		RobDEVC$_2$		RobDEVC$_3$	
		CIR	SR	CIR	SR	CIR	SR	CIR	SR	CIR	SR
Mad	Tukey Biq.	0.3134	0.0891	0.3874	0.0828	0.5192	0.0708	0.6810	0.0467	0.6800	0.0470
	Huber	0.2036	0.1097	0.2750	0.1043	0.4002	0.0946	0.5820	0.0731	0.5862	0.0731
	Humpel	0.2418	0.1025	0.3194	0.0965	0.4596	0.0843	0.6282	0.0607	0.6314	0.0605
	Andrew	0.3292	0.0858	0.4060	0.0786	0.5402	0.0662	0.6966	0.0421	0.6952	0.0425
	Winsor	0.2138	0.1081	0.2882	0.1028	0.4100	0.0931	0.5896	0.0729	0.5932	0.0728
	Trimmed	0.2134	0.1084	0.2866	0.1031	0.4080	0.0935	0.5880	0.0734	0.5914	0.0735
	Median	0.2606	0.0991	0.3332	0.0936	0.4644	0.0827	0.6274	0.0608	0.6300	0.0605
	HodgesLeh.	0.2084	0.1091	0.2816	0.1037	0.4032	0.0941	0.5824	0.0740	0.5872	0.0739
S$_n$	Tukey Biq.	0.3214	0.0868	0.3996	0.0799	0.5338	0.0668	0.6902	0.0401	0.6896	0.0402
	Huber	0.2140	0.1076	0.2862	0.1023	0.4164	0.0914	0.6004	0.0679	0.6002	0.0681
	Humpel	0.2536	0.0999	0.3336	0.0934	0.4690	0.0812	0.6440	0.0542	0.6452	0.0542
	Andrew	0.3366	0.0832	0.4208	0.0756	0.5560	0.0616	0.7050	0.0355	0.7050	0.0354
	Winsor	0.2244	0.1061	0.2994	0.1004	0.4274	0.0895	0.6070	0.0676	0.6072	0.0678
	Trimmed	0.2228	0.1064	0.2958	0.1008	0.4242	0.0903	0.6040	0.0683	0.6044	0.0684
	Median	0.2694	0.0967	0.3476	0.0908	0.4768	0.0792	0.6412	0.0545	0.6412	0.0546
	HodgesLeh.	0.2178	0.1071	0.2904	0.1016	0.4204	0.0908	0.6030	0.0686	0.6040	0.0688
Q$_n$	Tukey Biq.	0.3848	0.0736	0.4554	0.0666	0.5904	0.0533	0.7310	0.0300	0.7322	0.0299
	Huber	0.2674	0.0986	0.3448	0.0926	0.4700	0.0821	0.6396	0.0607	0.6402	0.0610
	Humpel	0.3102	0.0898	0.3934	0.0825	0.5238	0.0702	0.6852	0.0455	0.6856	0.0458
	Andrew	0.3968	0.0695	0.4682	0.0620	0.6132	0.0473	0.7440	0.0243	0.7468	0.0243
	Winsor	0.2798	0.0966	0.3556	0.0908	0.4804	0.0801	0.6474	0.0602	0.6490	0.0605
	Trimmed	0.2790	0.0968	0.3540	0.0912	0.4778	0.0807	0.6452	0.0608	0.6456	0.0611
	Median	0.3286	0.0862	0.4018	0.0800	0.5286	0.0688	0.6852	0.0461	0.6854	0.0464
	HodgesLeh.	0.2722	0.0979	0.3490	0.0920	0.4736	0.0815	0.6434	0.0613	0.6438	0.0616

CIR: Correct identification rate, SR: Swamping rate

Table 10. Simulation results for p = 3, n = 120 and bad leverage point 5

		DEVC		RobDEVC		RobDEVC$_1$		RobDEVC$_2$		RobDEVC$_3$	
		CIR	SR	CIR	SR	CIR	SR	CIR	SR	CIR	SR
Mad	Tukey Biq.	0.4998	0.0374	0.6974	0.0242	0.9369	0.0031	0.9686	0.0469	0.9682	0.0465
	Huber	0.3100	0.0572	0.4810	0.0472	0.8158	0.0232	0.9346	0.0117	0.9310	0.0114
	Humpel	0.3874	0.0496	0.5848	0.0373	0.8931	0.0096	0.9562	0.0332	0.9550	0.0328
	Andrew	0.5342	0.0334	0.7315	0.0192	0.9508	0.0078	0.9724	0.0524	0.9722	0.0519
	Winsor	0.3227	0.0561	0.4988	0.0460	0.8289	0.0224	0.9382	0.0092	0.9351	0.0087
	Trimmed	0.3227	0.0561	0.4973	0.0461	0.8278	0.0226	0.9375	0.0087	0.9343	0.0083
	Median	0.4025	0.0476	0.5831	0.0362	0.8749	0.0107	0.9516	0.0289	0.9500	0.0285
	HodgesLeh.	0.3118	0.0571	0.4837	0.0472	0.8175	0.0236	0.9348	0.0086	0.9313	0.0082
S$_n$	Tukey Biq.	0.5524	0.0323	0.7556	0.0184	0.9621	0.0114	0.9752	0.0668	0.9750	0.0664
	Huber	0.3309	0.0548	0.5151	0.0441	0.8543	0.0175	0.9461	0.0260	0.9436	0.0256
	Humpel	0.4212	0.0461	0.6304	0.0329	0.9248	0.0026	0.9638	0.0510	0.9630	0.0506
	Andrew	0.5906	0.0278	0.7914	0.0129	0.9708	0.0164	0.9784	0.0729	0.9785	0.0724
	Winsor	0.3447	0.0536	0.5344	0.0427	0.8677	0.0167	0.9497	0.0231	0.9476	0.0227
	Trimmed	0.3442	0.0537	0.5330	0.0428	0.8663	0.0170	0.9493	0.0227	0.9470	0.0223
	Median	0.4343	0.0442	0.6276	0.0320	0.9084	0.0038	0.9611	0.0460	0.9596	0.0455
	HodgesLeh.	0.3332	0.0546	0.5194	0.0439	0.8560	0.0181	0.9468	0.0224	0.9442	0.0219
Q$_n$	Tukey Biq.	0.7321	0.0123	0.8892	0.0007	0.9860	0.0257	0.9878	0.0727	0.9879	0.0717
	Huber	0.4710	0.0423	0.6747	0.0302	0.9321	0.0066	0.9668	0.0292	0.9665	0.0285
	Humpel	0.5869	0.0301	0.7910	0.0163	0.9717	0.0103	0.9782	0.0559	0.9781	0.0549
	Andrew	0.7631	0.0069	0.9110	0.0069	0.9880	0.0316	0.9882	0.0793	0.9901	0.0783
	Winsor	0.4952	0.0402	0.6970	0.0281	0.9412	0.0055	0.9692	0.0269	0.9690	0.0259
	Trimmed	0.4930	0.0404	0.6956	0.0283	0.9404	0.0057	0.9690	0.0264	0.9688	0.0256
	Median	0.5957	0.0282	0.7845	0.0157	0.9655	0.0085	0.9770	0.0500	0.9769	0.0491
	HodgesLeh.	0.4775	0.0418	0.6798	0.0299	0.9340	0.0070	0.9674	0.0261	0.9672	0.0251

CIR: Correct identification rate; SR: Swamping rate

Table 11. Calculated cut-off values

		Scale estimators		
	Location estimators	**Mad**	**S_n**	**Q_n**
DEVC	**Medyan**	2.9217	3.0768	2.9084
	Huber	3.0207	3.1757	3.0074
	Hampel	3.0559	3.2109	3.0426
	Andrew	2.9077	3.0628	2.8944
	Winsorize	3.0026	3.1576	2.9893
	Trimmed	3.0228	3.1778	3.0095
	Tukey	2.9150	3.0700	2.9017
	HodgesLeh.	3.0093	3.1643	2.9960
RobDEVC	**Medyan**	2.8664	3.0626	2.8926
	Huber	2.9598	3.1559	2.9860
	Hampel	2.9985	3.1947	3.0247
	Andrew	2.8515	3.0476	2.8777
	Winsorize	2.9426	3.1387	2.9688
	Trimmed	2.9655	3.1616	2.9917
	Tukey	2.8586	3.0548	2.8849
	HodgesLeh.	2.9487	3.1449	2.9750
$RobDEVC_1$	**Medyan**	2.7855	2.9633	2.8551
	Huber	2.8996	3.0774	2.9692
	Hampel	2.9427	3.1204	3.0123
	Andrew	2.8006	2.9783	2.8701
	Winsorize	2.8950	3.0727	2.9645
	Trimmed	2.9082	3.0859	2.9777
	Tukey	2.8760	2.8760	2.8760
	HodgesLeh.	2.8984	3.0762	2.9680
$RobDEVC_2$	**Medyan**	2.4526	2.6258	2.6119
	Huber	2.5951	2.7683	2.7544
	Hampel	2.6360	2.8092	2.7953
	Andrew	2.4313	2.6045	2.5906
	Winsorize	2.5818	2.7550	2.7411
	Trimmed	2.6385	2.8117	2.7977
	Tukey	2.4990	2.6722	2.6583
	HodgesLeh.	2.4990	2.6722	2.6583
$RobDEVC_3$	**Medyan**	2.4553	2.6370	2.6254
	Huber	2.5947	2.7764	2.7648
	Hampel	2.6361	2.8178	2.8062
	Andrew	2.4318	2.6135	2.6018
	Winsorize	2.5819	2.7636	2.7519
	Trimmed	2.6380	2.8197	2.8081
	Tukey	2.4991	2.6808	2.6692
	HodgesLeh.	2.5857	2.7674	2.7558

extremely low CIR values on all cut-off values to determine bad leverage points. In our work, all examined methods identify bad leverage points accurately at Andrew + $3Q_n$ cut-off values. Also, when SR values are used for Andrew + $3Q_n$ cut-off values in all methods have extremely low values. When Table 6 is examined, it is seen that the diagnostic methods suggested in the study for n=120 give good results. It can be said that $RobDEVC_3$ method gives more effective results in determining bad leverage points compared to all methods. Except for SR values $RobDEVC_2$ and $RobDEVC_3$ in all methods Andrew + $3Q_n$ have the lowest values at cut-off values.

· ***When the independent variable number is three;***

If there are two bad leverage points in the dataset when Table 7 is examined for n = 80, once diagnosis methods to identify bad leverage points are compared $RobDEVC_1$

$RobDEVC_2$ and $RobDEVC_3$ methods are more effective than the classical DEVC and RobDEVC methods. It can be seen that all methods give more accurate results on identifying bad leverage points at Andrew + $3Q_n$ cut-off value. When Table 8 for n = 120 is examined, all diagnostic methods show a high rate of success. Once Andrew + $3Q_n$ cut-off values are used, all diagnostic methods have a high identification ratio in identifying bad leverage points. $RobDEVC_2$ and $RobDEVC_3$ methods have the highest CIR value at this cut-off value. Also, when SR values in all diagnostic methods for the cut-off value combination have Hodges-Lehmann as the location parameter, it gets the lowest value.

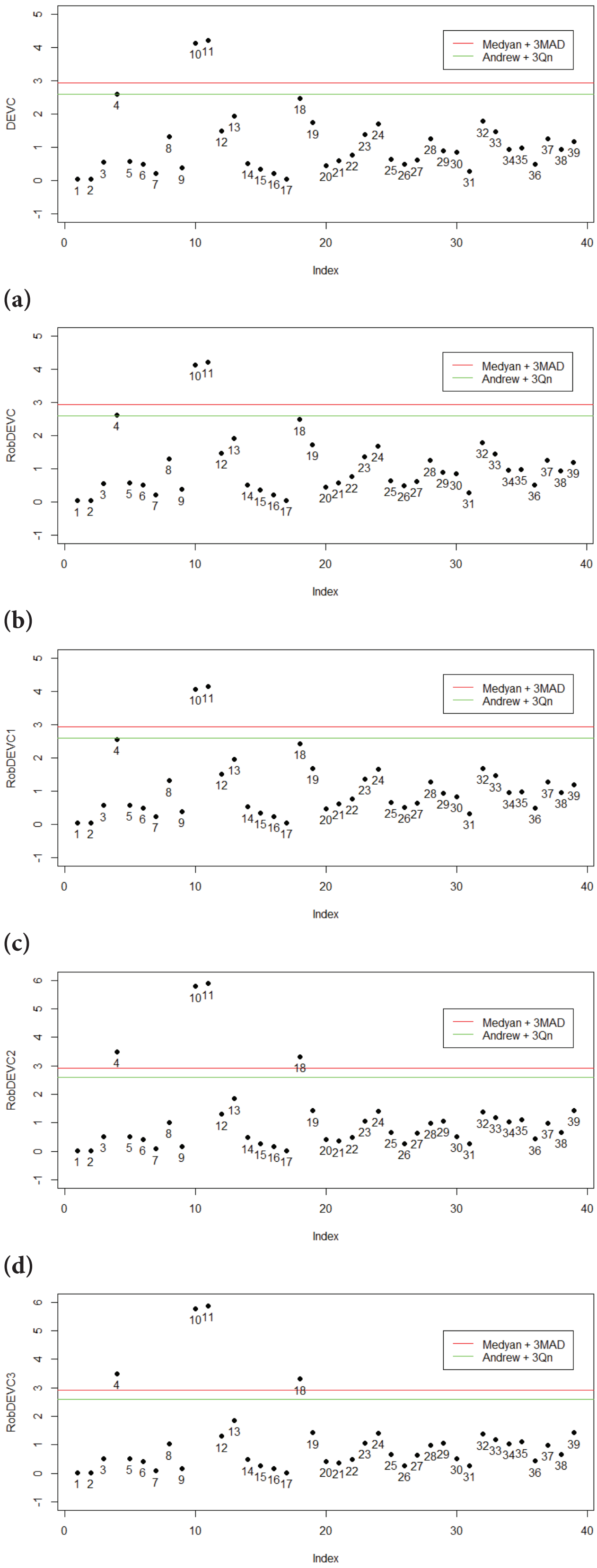
Finally, when there are five bad leverage points in the dataset when Table 9 is examined for n = 80 DEVC and RobDEVC methods have an extremely low rate of success at identifying bad leverage points once all criteria are considered. When the recommended diagnosis methods are examined, at high CIR value cut-off value combination $RobDEVC_3$ methods are more successful. Once all diagnostic methods are examined while Andrew + $3Q_n$ cut-off value has the highest CIR value on accurate identification of bad leverage points, it also has the lowest SR value. When Table 10 for n = 120 is examined $RobDEVC_1$ $RobDEVC_2$ and $RobDEVC_3$ diagnosis methods are more effective in identifying bad leverage points. In all diagnosis methods Andrew + $3Q_n$ cut-off values are the closest value to 1 in identifying bad leverage points. After Andrew + $3Q_n$ cut-off values, again the highest CIR value cut-off value combinations are in order Andrew + $3S_n$ and Andrew + 3MAD cut-off values can be seen in Table 15.

A Numerical Example

In this section of our work, our recommended robust diagnosis methods' real dataset application in the identification of bad leverage points for the logistic regression model was emphasized. Modified Finney dataset, has been obtained to examine respiration air's speed and the volume's effect on a temporary vasoconstriction in the skin of a finger [23]. The measuring process was constructed for reliable measurement of whether vasoconstriction alone will occur or not. In this dataset 4^{th}, 10^{th}, 11^{th} and 18^{th} observations are known as bad leverage points.

Classical diagnosis methods covered in our work and deviation residual obtained from the recommended robust diagnosis methods and robust deviation residual values were calculated. The cut-off values available in the literature and the cut-off values recommended in the study are given in Table 11 by using the deviation residuals of the methods discussed.

Once Table 11 is examined, the cut-off values to be used for the methods are given. The graphs of the diagnostic methods discussed in the study and the Median + 3MAD cut-off value available in the literature and the Andrew + $3Q_n$ cut-off value, which has a high correct classification rate in the simulation study, are given in Figure 1.



(a) (b) (c) (d) (e)

Figure 1. Index plots for (a) DEVC, (b) RobDEVC, (c) $RobDEVC_1$, (d) $RobDEVC_2$ and (e) $RobDEVC_3$ deviance residual valuesa

Additionally the values obtained by these cut-off values are calculated as 2,9217 and 2,6018 in order accordingly.

Once Figure 1 is examined when $RobDEVC_2$ and $RobDEVC_3$ methods are used, identified bad leverage points can be accurately identified for both cut-off values. For these methods, four bad leverage points in the data set were determined as bad leverage points in the 4th, 10th, 11th and 18th observations. Since these observations are known as bad leverage points in the dataset, Andrew + 3Qn cut-off value seems to correctly identify these observations. However, it is seen that the Medyan + 3MAD cut-off value available in the literature also determines more than one observation as a bad leverage point, apart from these bad leverages. In the DEVC and $RobDEVC_1$ diagnostic methods, observations 4 and 18 cannot be determined at both cut-off values, whereas when the RobDEVC diagnostic method is used, Andrew + $3Q_n$ cut-off value cannot accurately determine only the 18th observation.

CONCLUSION

Many diagnosis methods were developed to identify high-leverage points in the dataset. These developed methods identify high-leverage points, however, they cannot distinguish good or bad leverage points. RobDEVC method developed by Nurunabi et al. for a logistic regression model, when there are both good and bad leverage points in the dataset, it is successful in identifying bad leverage points [4]. In this work, new diagnostic methods were recommended as an alternative to the RobDEVC method and compared against DEVC and RobDEVC methods in the literature that are used to identify bad leverage points.

Once simulation results are examined, and all criteria are considered $RobDEVC_1$ $RobDEVC_2$ and $RobDEVC_3$ diagnosis methods show a higher rate of success compared to classic DEVC and RobDEVC methods for identifying bad leverage points. In order to accurately see the identification of bad leverage points at which cut-off value, CIR values are calculated. When CIR values are examined, in all methods, independent variables and sampling size Andrew + $3Q_n$ cut-off value has a higher CIR value compared to all other cut-off value combinations. Once diagnostic methods are examined $RobDEVC_3$ diagnosis method is effective. Also $RobDEVC_3$ and $RobDEVC_2$ methods have very close CIR values at cut-off value combinations.

In real data application dataset at $RobDEVC_2$ and $RobDEVC_3$ diagnosis methods when Andrew + $3Q_n$ cut-off value is used four bad leverage points are accurately identified. When other diagnosis methods' graphics are examined bad leverage points cannot be accurately identified at both used cut-off values.

As a result, in this study, it has been tried to determine bad leverage points by using robust deviation residuals. In the studies carried out so far, robust deviation residuals have been obtained by using CUBIF, BY and WBY robust estimators as an alternative to the robust deviation residual obtained by using deviation residuals and Mallows estimator. With these robust deviation residues, RobDEVC diagnostic methods have been proposed. With both real data application and simulation results have shown better performance at identifying bad leverage points than DEVC and RobDEVC diagnosis methods. When $RobDEVC_2$ and $RobDEVC_3$ diagnosis methods were used, bad leverage point identification ratio is high and $RobDEVC_3$ diagnosis methods are more effective. In addition to the Median + 3MAD cut-off value used in the literature for all diagnostic methods, the Andrew + $3Q_n$ cut-off value suggested by the authors gave good results in accurately identifying bad leverage points [12]. In the conducted research, during the identification of bad leverage points, usage of robust diagnosis methods and robust cut-off values will allow accurate identification of bad leverage points.

NOMENCLATURE

OLS	Ordinary Least Squares
MLE	Maximum Likelihood Estimator
MSE	Mean Square Error
$TPRE$	Two Parameter Ridge Estimator
$RMSE$	Residual Root Mean Square Error
$MMSE$	Matrix Mean Square Error
LRE	Logistic Rigde Estimator
BY	Bianco and Yohai
WBY	Weighted Bianco and Yohai
$CUBIF$	Conditionally Unbiased Bounded Influence Function
$Mallows$	Mallows Type Leverage Dependent Weights
$RLRE$	Robust Logistic Rigde Estimator
$IRLS$	Iterative Re-weighted Least Squares Method
LE	Logistic Estimator
$DEVC$	Deviance Components
$RobDEVC$	Robust Deviance Components
GM	Generalized M
RMD	Robust Mahalanobis Distance
MCD	Minimum Covariance Determinant
RLE	Robust Logistic Estimator
MAD	Median Absolute Deviation
CIR	Correct Identification Rate
SR	Swamping Rate
$TPLRE$	Two Parameter Logistic Rigde Estimator
$TPRLRE$	Two Parameter Robust Logistic Rigde Estimator

AUTHORSHIP CONTRIBUTIONS

Authors equally contributed to this work.

DATA AVAILABILITY STATEMENT

The authors confirm that the data that supports the findings of this study are available within the article. Raw data that support the finding of this study are available from the corresponding author, upon reasonable request.

CONFLICT OF INTEREST

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ETHICS

There are no ethical issues with the publication of this manuscript.

REFERENCES

[1] Belsley DA, Kuh E, Welsch RE. Regression Diagnostics. Hoboken, NJ, USA: John Wiley & Sons, Inc.; 1980. [CrossRef]

[2] Imon AHM. Identification of high leverage points in logistic regression. Pakistan J Stat 2006;22:147-156.

[3] Norazan MR, Sanizah A, Habshah M. Identifying bad leverage points in logistic regression model based on robust deviance components. Math Models Methods Modern Sci 2012:62-67.

[4] Nurunnabi AAM, Imon AHMR, Nasser M. Identification of multiple influential observations in logistic regression. J Appl Stat 2010;37:1605-1624. [CrossRef]

[5] Syaiba BA, Habshah M. Robust logistic diagnostic for the identification of high leverage points in logistic regression model. J Appl Sci 2010;10:3042-3050. [CrossRef]

[6] Arqub OA, Abo-Hammour Z. Numerical solution of systems of second-order boundary value problems using continuous genetic algorithm. Inform Sci 2014;279:396-415. [CrossRef]

[7] Abu Arqub O, Singh J, Alhodaly M. Adaptation of kernel functions-based approach with Atangana-Baleanu-Caputo distributed order derivative for solutions of fuzzy fractional Volterra and Fredholm integrodifferential equations. Math Methods Appl Sci 2021:7807-7834. [CrossRef]

[8] Abo-Hammour Z, Abu Arqub O, Momani S, Shawagfeh N. Optimization solution of Troesch's and Bratu's problems of ordinary type using novel continuous genetic algorithm. Discrete Dyn Nat Soc 2014;2014. [CrossRef]

[9] Ahmad S, Ramli NM, Midi H. Robust estimators in logistic regression: A comparative simulation study. J Modern Appl Stat Methods 2010;9:502-511. [CrossRef]

[10] Croux C, Haesbroeck G. Implementing the bianco and yohai estimator for logistic regression. Comput Stat Data Anal 2003;44:273-295. [CrossRef]

[11] Venter JH, De La Rey T. Detecting outliers using weights. South African Stat J 2007;41:127-160.

[12] Gündoğan Aşik E, Altin Yavuz A, Küçük Z. New robust cut-off values in determining bad leverage points in the logistic regression model. Mehmet Akif Ersoy Univ J Ec on Administrat Sci Fac 2021;8:630-650. [Turkish] [CrossRef]

[13] Sarkar K, Midi H, Rana S. Detection of outliers and influential observations in binary Logistic regression: An empirical study. J Appl Sci 2011;11:26-35. [CrossRef]

[14] Künsch HR, Stefanski LA, Carroll RJ. Conditionally unbiased bounded-influence estimation in general regression models, with applications to generalized linear models. J Am Stat Assoc 1989;84:460-466. [CrossRef]

[15] Carroll RJ, Pederson S. On robustness in the logistic regression model. J R Stat Soc Series B Stat Methodol 1993;55:693-706. [CrossRef]

[16] Pregibon D. Logistic regression diagnostics. Ann Stat 1981;9:705-724. [CrossRef]

[17] Bianco AM, Yohai VJ. Robust Estimation in the Logistic Regression Model. Springer V. New York, NY: 1996. [CrossRef]

[18] Croux C, Haesbroeck G. Implementing the Bianco and Yohai estimator for logistic regression. Comput Stat Data Anal 2003;44:273-295. [CrossRef]

[19] Rousseeuw PJ, Leroy AM. Robust Regression and Outlier Detection. Hoboken, NJ, USA: John Wiley & Sons, Inc.; 1987. [CrossRef]

[20] Hastings C, Mosteller F, Tukey JW, Winsor CP. Low moments for small samples: A comparative study of order statistics. Ann Math Stat 1947;18:413-426. [CrossRef]

[21] Hoaglin DC, Mosteller F. Understanding robust and explonatory data analysis,. New York, NY: John Wiley & Sons; 1983.

[22] Hodges JLJ, Lehmann EL. Estimates of location based on ranks tests. Ann Math Stat 1963;34:598-611. [CrossRef]

[23] Nurunnabi AAM, Imon AHMR, Nasser M. A new statistic for influence in linear regression. Technometrics 2008;47:1-11.