Robust Kurtosis Projection for Multivariate OutlierLabeling

Dyah E. Herwindiati, Rahmat Sagara, Janson Hendryli
Faculty of Information Technology, Tarumanagara University
Email: herwindiati@untar.ac.id, rahmat.sagara.01@gmail.com, jansonhendryli@gmail.com

Abstract—**Outlier labeling can be considered as an early procedure to get the information of 'suspects'. This paper introducesrobust kurtosis projection algorithm for multivariate outlier labeling of data set with moderate, high and very high percentage outlier. The algorithm works in two stages. In the first stage, we propose a projection approach to findthe orthonormal set of all vectors that maximize the kurtosis of the projected standardized data. In the second stage, we estimate robust covariance matrix minimizing vector variance to label high dimensional outliers. In this stage, we use the robust estimator on the lower-dimensional data space to identify the suspected anomolous observations. The simulation experiments reveal that theintroduced algorithm has a good performance to identify an anomalous observation hidden in a moderate, high, and very high percentage of contamination data and it seems to work well in data analysis.**

Keywords: kurtosis, orthonormal, outlier, robust, vectorvariance

I. Introduction

Outlier detectionis one of the basic problems of data mining. Outlier detection has the important role in modeling, statistical inference, and even data processing because outlier can lead to model misspecification, biased parameter estimation and poor forecasting. Outlier detection has also extensive use in wide variety of computer science applications, such as intrusion detection, image detection, content-based image retrieval, and classification of remote sensing data.

Awareness on outlier occurrence had emerged since early XVI century. It was when Francis Bacon on 1620 wrote about the importance to know phenomenon of nature deviations, cited by Werner [10].Studies on outlier detection have been developed for centuries. Thompson [16] proposed statistics test for univariate data in form of ratio between deviation of every observation to the mean and standard deviation of sample.Wilks [14] introduced a method of outlier testing for multivariate data based on the ratio of volume of aparallelotop. Rousseeuw [12] introduced the minimum volume ellipsoid(MVE) to estimate location parameters and covariance matrices as measure of outlierdetection. Rousseeuw and van Driessen [13] introduced the fast minimum covariance determinant(FMCD) for the same future goal. A criterion for robust estimation of location and covariance matrix for outlier labeling based on minimum vector variance (MVV) was also proposed byHerwindiati etal. [2].Various techniques on identification of outliers were proposed, one of them is outlier labeling.

The outlier labeling can be considered as the early procedure to get the information of "suspects". The aim of outlier labeling is to flag observations as possible outliers for further investigation, Iglewiczand Hoglin [1]. The preliminary information on the number and location of outlier is able to help in formally identifying the potential outliers.

Authors have proposed many definitions for an outlier with seemingly no universally accepted definition. This paper uses the famous definition given by Barnet and Lewis [15] that outlier to be one or more observations, which are not consistent among others.

Numerous outlier detection techniques have been proposed in the data analysis. Our paper presents a new technique to label outlier. The procedure combines two advantages of both kurtosis projection pursuit and robust covariance estimation. The goal of projection pursuit is to use the data to find low (one, two, or three) dimensional space providing the most revealing views of the full dimensional data. Robust statistics deals with deviations from the assumptions of normality, linearity, and independence stick on the classic estimation methods frequently are not satisfied, see Huber[11]. Robust statistics is a convenient modern way of summarizing results when we suspect that they include small proportion of outliers.

The interesting properties of kurtosis projection pursuit and the powerful of robust estimation tend us to introduce the new measure for multivariate outlier labeling of dataset with moderate, high, and very high percentage outlier.

Our algorithm works in two stages. In the first stage, we propose the projection approach finding the orthonormal set of all vectors that maximize the kurtosis of the projected standardized data. This approach improves on the slow convergence rate proposed byPena and Prieto[3]. In the second stage, we estimate robust covariance matrix minimizing vector variance (MVV) to label high dimensional outliers. In this stage we use the MVV estimator on the lower-dimensional data space from kurtosis projection to indentify the suspect anomalous observations.

978-1-5090-0363-1/15/$31.00 ©2015 IEEE

II. The Algorithm of Kurtosis Projection Approach

Projection pursuit; Friedman [5]; is a technique aiming at identifying low-dimensional projections of data that reveal interesting structures. The framework of projection pursuit is formulated as an optimization problem with the goal of finding projection axes that minimize or maximize a measure of interest called projection index.

Kurtosis can be formally defined as the standardized fourth population moment about the mean,

$$K = \frac{E(X-\mu)^4}{(E(X-\mu)^2)^2} = \frac{\mu^4}{\sigma^4} \tag{1}$$

where$E$ is the expectation,$\mu^4$is the fourth moment about the mean, and $\sigma$is standard deviation. The role of kurtosis is a measure of normality; in issues of robustness, outliers, and bimodality.

Let X $= (\vec{x}_1, \vec{x}_2, \cdots, \vec{x}_n)'$be a data matrix of size$n \times p$ as the observation result of$p$ variables to$n$ individual objects and$\vec{d}$a unit vector in$\mathbb{R}^p$. The orthogonal projection of each observation result on to one-dimensional space spanned by $\vec{d}$is $y_i = \vec{d}'\vec{x}_i$.

Write the projected data as Y $= (y_1, y_2, \cdots, y_n)$. The kurtosis of the projected data is formulated as

$$K = \frac{\frac{1}{n}\sum_{i\in\mathbb{N}_n}(y_i - t)^4}{s^4} \tag{2}$$

where$t = \frac{1}{n}\sum_{i\in\mathbb{N}_n} y_i$and s$^2 = \frac{1}{n}\sum_{i\in\mathbb{N}_n}(y_i - t)^2$are the sample mean and the sample variance of the data$Y$ respectively. It is noted that s^4 is the square of s^2 and $\mathbb{N}_n$is the set of all natural numbers less than or equal to$n$.

Centering and scaling transformation Y gives a new dataZ $= (z_1, z_2, \cdots, z_n)$, where for all$i \in \mathbb{N}_n$, $z_i = \frac{y_i - t}{s}$. Since the sample mean of$Z$is$0$ and the sample variance of$Z$is$1$, the kurtosis of$Z$is formulated as:

$$K = \frac{1}{n}\sum_{i\in\mathbb{N}_n} {z_i}^4 \tag{3}$$

Define the sample mean$\vec{t}$ and the covariance matrixS of the data matrix$X$ as:

$$\vec{t} = \frac{1}{n}\sum_{i\in\mathbb{N}_n} \vec{x}_i\, \vec{t} \tag{4}$$

$$S = \frac{1}{n}\sum_{i\in\mathbb{N}_n} (\vec{x}_i - \vec{t})(\vec{x}_i - \vec{t})' \tag{5}$$

then the kurtosis$K$can be written as

$$K = \frac{1}{n}\sum_{i\in\mathbb{N}_n}\left(\frac{\vec{d}'(\vec{x}_i - \vec{t})}{\sqrt{\vec{d}'S\vec{d}}}\right)^4 \tag{6}$$

Consider the objective function$f$defined as

$$f(\vec{d}) = \frac{1}{n}\sum_{i\in\mathbb{N}_n}(\vec{d}'\vec{y}_i)^4 - \lambda(\vec{d}'\vec{d} - 1) \tag{7}$$

whereλ is the Lagrange multiplier. The first derivative of$f$and setting it to be zero results:

$$\frac{df}{d\vec{d}} = \frac{1}{n}\sum_{i\in\mathbb{N}_n} 4(\vec{d}'\vec{y}_i)^3\vec{y}_i - 2\lambda\vec{d} = 0 \tag{8}$$

$$\left(\frac{1}{n}\sum_{i\in\mathbb{N}_n}(\vec{d}'\vec{y}_i)^2\vec{y}_i\vec{y}_i'\right)\vec{d} = \frac{\lambda}{2}\vec{d} \tag{9}$$

Write$\frac{\lambda}{2}$asκ_1. We see that$\kappa_1$is the eigenvalue of $\frac{1}{n}\sum_{i\in\mathbb{N}_n}(\vec{d}'\vec{y}_i)^2\vec{y}_i\vec{y}_i'and\vec{d}$is the corresponding eigenvector. Multiplying the last equation by$\vec{d}'$from the left results

$$\begin{aligned}\kappa_1 &= \vec{d}'\left(\frac{1}{n}\sum_{i\in\mathbb{N}_n}(\vec{d}'\vec{y}_i)^2\vec{y}_i\vec{y}_i'\right)\vec{d} \\ &= \frac{1}{n}\sum_{i\in\mathbb{N}_n}(\vec{d}'\vec{y}_i)^4 = K\end{aligned} \tag{10}$$

So the unit vector$\vec{d}$that maximizes$K$is the eigenvector of

$$M_1 = \frac{1}{n}\sum_{i\in\mathbb{N}_n}(\vec{d}'\vec{y}_i)^2\vec{y}_i\vec{y}_i' \tag{11}$$

corresponding to the maximum eigenvalue of that matrix. We will call such eigenvector by$\vec{d}_1$. In order to obtain the second unit vector$\vec{d}$ that maximizes$K$ and orthogonal to$\vec{d}_1$, we have to maximize the objective function$f$that is redefined as

$$f(\vec{d}) = \frac{1}{n}\sum_{i\in\mathbb{N}_n}(\vec{d}'\vec{y}_i)^4 - \lambda_1(\vec{d}'\vec{d} - 1) - \lambda_2\vec{d}'\vec{d}_1 \tag{12}$$

where$\lambda_1$and$\lambda_1$are the Lagrange multipliers. The first derivative of$f$ and setting it to be zero results,

$$\frac{df}{d\vec{d}} = \frac{1}{n}\sum_{i\in\mathbb{N}_n} 4(\vec{d}'\vec{y}_i)^3\vec{y}_i - 2\lambda_1\vec{d} - \lambda_2\vec{d}_1 = 0 \tag{13}$$

$$\frac{1}{n}\sum_{i\in\mathbb{N}_n}(\vec{d}'\vec{y}_i)^3\vec{y}_i - \frac{\lambda_2}{4}\vec{d}_1 = \frac{\lambda_1}{2}\vec{d} \tag{14}$$

Where$\frac{\lambda_1}{2}$is the kurtosis$K$. Write $\frac{\lambda_1}{2}$ as κ_2. Now, multiplying by$\vec{d}_1'$ from the left results

$$\vec{d}_1'\frac{1}{n}\sum_{i\in\mathbb{N}_n}\left(\vec{d}'\vec{y}_i\right)^3\vec{y}_i-\frac{\lambda_2}{4}\vec{d}_1'\vec{d}_1=\kappa_2\vec{d}_1'\vec{d} \tag{15}$$

$$\frac{1}{n}\sum_{i\in\mathbb{N}_n}\left(\vec{d}'\vec{y}_i\right)^3\vec{d}_1'\vec{y}_i-\frac{\lambda_2}{4}=0 \tag{16}$$

$$\frac{\lambda_2}{4}=\frac{1}{n}\sum_{i\in\mathbb{N}_n}\left(\vec{d}'\vec{y}_i\right)^3\vec{d}_1'\vec{y}_i \tag{17}$$

So we have

$$\frac{1}{n}\sum_{i\in\mathbb{N}_n}\left(\vec{d}'\vec{y}_i\right)^3\vec{y}_i-\frac{1}{n}\sum_{i\in\mathbb{N}_n}\left(\vec{d}'\vec{y}_i\right)^3\vec{d}_1'\vec{y}_i\,\vec{d}_1=\kappa_2\vec{d} \tag{18}$$

$$\left(\left(I-\vec{d}_1\vec{d}_1'\right)\frac{1}{n}\sum_{i\in\mathbb{N}_n}\left(\vec{d}'\vec{y}_i\right)^2\vec{y}_i\vec{y}_i'\right)\vec{d}=\kappa_2\vec{d} \tag{19}$$

The second unit vector$\vec{d}$ that maximizes$K$ and orthogonal to$\vec{d}_1$ is the eigenvector of

$$M_2=\left(I-\vec{d}_1\vec{d}_1'\right)\frac{1}{n}\sum_{i\in\mathbb{N}_n}\left(\vec{d}'\vec{y}_i\right)^2\vec{y}_i\vec{y}_i' \tag{20}$$

corresponding to the maximum eigenvalue of that matrix. We will call such eigenvector by$\vec{d}_2$. We can verify that for $2\le k\le p$, $\vec{d}_k$is the eigenvector of the matrix

$$M_k=\left(I-\sum_{j=1}^{k-1}\vec{d}_j\vec{d}_j'\right)\frac{1}{n}\sum_{i\in\mathbb{N}_n}\left(\vec{d}'\vec{y}_i\right)^2\vec{y}_i\vec{y}_i' \tag{21}$$

corresponding to the maximum eigenvalue of that matrix.

The outlier labeling can be considered as the early procedure to get the information of the "suspects". This paper use projection of maximize kurtosis coefficient to separate suspected data. Our proposed algorithm is inspired by Pena and Prieto's work [3]. Pena Prieto proposed the projection of data on p-orthogonal axis maximizing kurtosis. Their projection pursuit is not easy for large dataset and the computational difficulties are formidable.

We introduce the kurtosis projection approach for the initial step of our proposed method that is 'Robust Kurtosis Projection' to label outliers on multivariate data case. The projection algorithm is written as,

Algorithm 2.1: Finding p unit vector that maximizes the kurtosis

Input: matrix data$X=(\vec{x}_1,\vec{x}_2,\cdots,\vec{x}_n)'$ of size$n\times p$

Process:

1. Compute the sample mean$\vec{t}$ and the sample covariance matrix S of X.

$$\vec{t}=\frac{1}{n}\sum_{i\in\mathbb{N}_n}\vec{x}_i$$

$$S=\frac{1}{n}\sum_{i\in\mathbb{N}_n}(\vec{x}_i-\vec{t})(\vec{x}_i-\vec{t})'$$

2. Standardize the matrix data$X$such that the projected data has mean 0 and variance 1.

$$\vec{y}_i=S^{-\frac{1}{2}}(\vec{x}_i-\vec{t})$$

3. Find the first unit vector$\vec{d}_1$ as the eigenvector of $\frac{1}{n}\sum_{i\in\mathbb{N}_n}\left(\vec{d}'\vec{y}_i\right)^2\vec{y}_i\vec{y}_i'$ corresponding to the maximum eigenvalue of that matrix.
4. For $k=2,3,\cdots,p$, find the k-th unit vector $\vec{d}_k$ as the eigenvector of

$$\left(I-\sum_{j=1}^{k-1}\vec{d}_j\vec{d}_j'\right)\frac{1}{n}\sum_{i\in\mathbb{N}_n}\left(\vec{d}'\vec{y}_i\right)^2\vec{y}_i\vec{y}_i'$$

corresponding to the maximum eigenvalue of that matrix. The output is an orthonormal set of all vectors that maximize the kurtosis$\{\vec{d}_1,\vec{d}_2,\cdots,\vec{d}_p\}$.

5. Find the projection$z_i=\vec{d}'\vec{y}_i$. Determine the minimum amount of variation that we want, defined by the new variablez_k.
6. Compute the distance$z_k$to center point,$zero(p)$.

III. The Illustration of Identifying an Anomalous Observations using Kurtosis Projection Approach

To illustrate the identification of an anomalous observations using kurtosis projection in the manner described in Section II, we generate a small dataset of size$n$from a mixture model $(1-\varepsilon)N_p(\vec{\mu}_1,\Sigma)+\varepsilon N_p(\vec{\mu}_2,\Sigma)$. For this purpose, we set$n=200$and$p=8$. We have three experiments with small and moderate proportion of contaminations$\varepsilon=0.01$, $\varepsilon=0.05$, and $\varepsilon=0.15$.

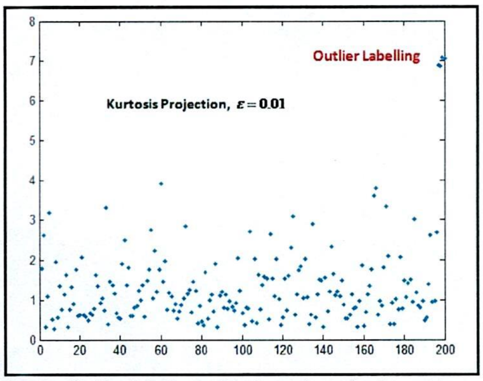

Fig. 1. Outlier labeling using kurtosis projection with **very small** contamination

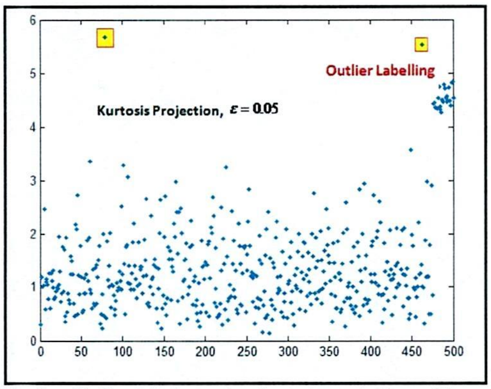

Fig. 2. Outlier labeling using kurtosis projection with **small** contamination

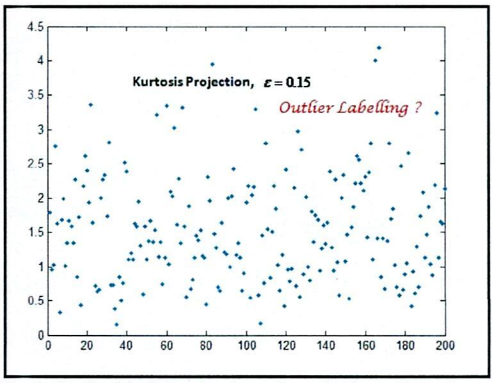

Fig. 3. Outlier labeling using kurtosis projection with **moderate** contamination

Figures 1, 2, and 3 show outlier labeling using kurtosis projection with very small, small, and moderate percentage of contamination in a dataset. Outlier labeling can be well identified by Kurtosis projection only for 1% contamination data. The process of identification failed on a moderate percentage of data contamination. The projection could be heavily distorted by the presence of outliers.

Our projection method is faster than ideas on Pena and Prieto's [3] but it has a good performance only for a small contamination data. It is different with Pena and Prieto's work. To improve the performance we introduce 'Robust Kurtosis Projection'.

IV. ROBUST KURTOSIS PROJECTION

Exploratory projection pursuit is a technique for finding interesting direction in low p-dimensional space of high dimensional, Jolliffe [4]. Projection pursuit is a tool for finding cluster, which can be labeled as 'unclean' cluster or an anomalous observation cluster. The technique gives a good clustering result in specific case, which is an anomalous observation hidden in a small proportion.

To improve the power against outliers we introduce a method combining two advantages of both kurtosis projection pursuit and robust covariance estimation, it is called as robust kurtosis projection method.

The robust kurtosis projection algorithm is composed in two steps. The first step is kurtosis projection step, which is explained in the previous section. In the second step, the robust step, we use robust minimum vector variance method.

Minimum Vector Variance (MVV) is the criteria to identify an outlier by using the minimization of vector variance (VV) proposed byHerwindiati et.al [2]. The estimator MVV is the pair(T_{MVV}, C_{MVV})giving minimum vector variance. The MVV is the good robust measure emerged sinceDjauhari [9] proposed the new multivariate dispersion that is vector variance (VV).

Suppose that$(\vec{X} - T)' C^{-1} (\vec{X} - T) = d^2$is an arbitrary ellipsoid. Let $\lambda_1, \lambda_2, \cdots, \lambda_p$be eigenvalues of C_{MVV}, vector variance (VV) is formulated as $Tr(C_{MVV}^2) = \lambda_1^2 + \lambda_2^2 + \cdots + \lambda_p^2$.

Consider a dataset$X = \{\vec{X}_1, \vec{X}_2, \cdots, \vec{X}_n\}$ of p-variate observations and let $H \subseteq X$.Suppose $T_{MVV}$and $C_{MVV}$are MVV estimator for the location parameter and covariance matrix. This two estimators are determined based on the set$H$ consists of$h = \left\lfloor \frac{n+p+1}{1} \right\rfloor$ data which give covariance matrix $C_{MVV}$of minimum $Tr(C_{MVV}^2)$among all possible sets of h data. Therefore,

$$T_{MVV} = \frac{1}{h} \sum_{\vec{X}_i \in H} \vec{X}_i \tag{22}$$

$$C_{MVV} = \frac{1}{h} \sum_{\vec{X}_i \in H} (\vec{X}_i - T_{MVV})(\vec{X}_i - T_{MVV})' \tag{23}$$

$T_{MVV}$and$C_{MVV}$are affine equivariant property.

The algorithm of Robust Kurtosis Projection is divided into two stages,

Algorithm 4.1: Robust Kurtosis Projection
Stage 1
Find the orthonormal set of all vectors that maximize the kurtosis $\{\vec{d}_1, \vec{d}_2, \cdots, \vec{d}_p\}$(see the algorithm in

Section II).

Stage 2

Identify the anomalous observations; labeled outliers; by using the algorithm:

1. Determine input data

2. Let H_{old}be an arbitrary subset containing $h = \left\lfloor \frac{n+p+1}{1} \right\rfloor$data points. Compute the mean vector$\vec{\bar{X}}_{H_{old}}$ and covariance matrix $\mathrm{S}_{H_{old}}$of all observations belonging toH_{old}. Then compute,

$$d^2_{H_{old}}(i) = \left(\vec{X}_i - \vec{\bar{X}}_{H_{old}}\right)' \mathrm{S}^{-1}_{H_{old}} \left(\vec{X}_i - \vec{\bar{X}}_{H_{old}}\right)$$

   for all $i = 1, 2, \cdots, n$

3. Sort these distances in increasing order,

$$d^2_{H_{old}}(\pi(1)) \leq d^2_{H_{old}}(\pi(2)) \leq \cdots \leq d^2_{H_{old}}(\pi(n))$$

4. Define$H_{new} = \{\vec{X}_{\pi(1)}, \vec{X}_{\pi(2)}, \cdots, \vec{X}_{\pi(h)}\}$

5. Calculate $\vec{\bar{X}}_{H_{new}}$, $\mathrm{S}_{H_{new}}$, and $d^2_{H_{new}}(i)$

6. If $Tr\left(\mathrm{S}^2_{H_{new}}\right) = 0$, repeat steps 1 to 5.
   If $Tr\left(\mathrm{S}^2_{H_{new}}\right) = Tr\left(\mathrm{S}^2_{H_{old}}\right)$, the process is stopped.
   Otherwise, continue until k-th iteration if$Tr(\mathrm{S}^2_1) \geq Tr(\mathrm{S}^2_2) \geq \cdots \geq Tr(\mathrm{S}^2_k) = Tr(\mathrm{S}^2_{k+1})$

7. Identify the labeled outlier by using robust MVV distance

V. The Illustration of Labeled Outliers Process using Robust Kurtosis Projection

In this section, we illustrate the good performance of robust kurtosis projection to separate the anomalous observations, which are hidden in moderate, high, and very high proportion or percentage in a dataset.

In the previous section, we see that kurtosis projection failed to identify the labeled outliers on a moderate percentage of data contamination. The projection is not robust against the outliers.

To indicate the power of robust kurtosis projection method, we use the same simulation data in Section III. We generate a mixture model $(1-\varepsilon)N_p(\vec{\mu}_1, \Sigma) + \varepsilon N_p(\vec{\mu}_2, \Sigma)$ of size $n = 200$, $p = 8$ with a moderate, high, and very high percentage contaminationε, which are 20%, 30%, and 55% of contaminated data.

It is surprising to find that the algorithm of robust kurtosis projection has a good performance to identify the anomalous observation hidden in a moderate, high, and very high percentage of contamination dataε. As seen on Figure 6, the contamination can still be well detected though the percentage is very high (55%).

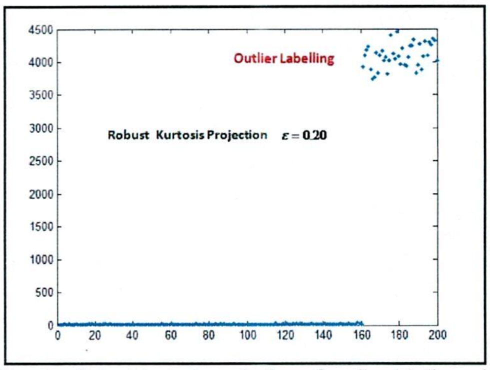

Fig. 4. Robust kurtosis projection of outlier labeling with **moderate** contamination

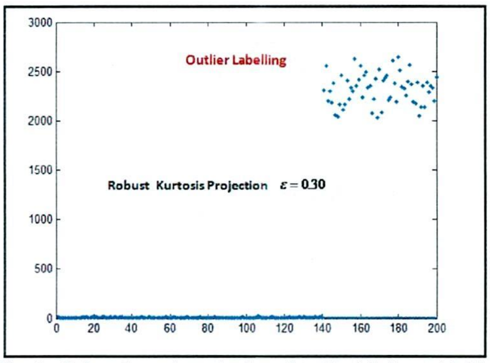

Fig. 5. Robust kurtosis projection of outlier labeling with **high** contamination

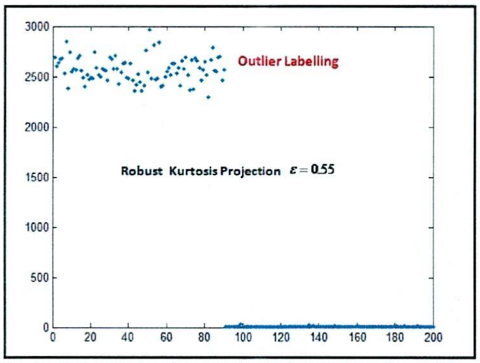

Fig. 6. Robust kurtosis projection of outlier labeling with **very high** contamination

VI. Remarks

Robust kurtosis projection has a good performance and robust to detect data contamination in a low, moderate, high even very high percentage. This method has been used and developed for multivariate outlier labeling and its application.

REFERENCES

[1] B. Iglewicz and D. C. Hoaglin, *How to Detect and Handle with Outliers*, American Society for Quality Statistics Division, 16, Milwaukee, 1964.

[2] D. E. Herwindiati, M. A. Djauhari, and M. Mashuri, "Robust multivariate outlier labeling", *Communication in Statistics (COMSTAT) B*, vol. 36, no. 6, 2007, pp. 1287-1294.

[3] D. Pena and J. F. Prieto. "Multivariate outlier detection and robust covariance matrix estimation", *Technometrics*, 43(3), 2001, pp. 286-300.

[4] I. T. Jolliffe, *Principle Component Analysis*, Springer Verlag, 1986.

[5] J. H. Friedman, "Exploratory projection pursuit", *Journal of the American Statistical Association*, vol. 82, no. 397, 1987, pp. 249-266.

[6] J. H. Friedman and J. W. Tukey, *A projection pursuit algorithm for exploratory data analysis*, C23(9), 1974, pp. 881-890.

[7] J. W. Tukey, *Exploratory Data Analysis*, Canada: Addison-Wesley, 1977.

[8] K. V. Mardia, J. T. Kent, and J. M. Bibby. *Multivariate Analysis*, Academic Press, 1979.

[9] M. A. Djauhari, "Improved monitoring of multivariate process variability", *Journal of Quality Technology*, 37(1), 2005, pp. 32-39.

[10] M. Werner, *Identification of Multivariate Outliers in Large Data Sets*, PhD Thesis, University of Colorado at Denver.

[11] P. J. Huber, "Robust estimation of location parameter", *Annals of Mathematical Statistics*, 35, 1964, pp. 73-101.

[12] P. J. Rousseeuw, "Multivariate estimation with high breakdown point," in *Mathematical Statistics and Applications*, B, W. Grossman, G. Pflug, I. Vincze, and W. Wertz, Ed, D. Reidel Publishing Company, 1985, pp. 283-297.

[13] P. J. Rousseeuw and K. van Driessen, "A fast algorithm for the minimum covariance determinant estimator", *Technometrics*, 41, 1999, pp. 212-223.

[14] S. S. Wilks, "Multivariate statistical outliers", *Sankya* A, 25, 1963, pp. 407-426.

[15] V. Barnett and T. Lewis, *Outliers in Statistical Data*, 2nd ed, New York: John Wiley, 1984.

[16] W. R. Thompson, "On a criterion for the rejection of observation and the distribution of the ratio of deviation to sample standard deviation", *The Annals of Mathematical Statistics*, 6(4), 1935, pp. 214-219.