Research Article GU J Sci 33(4): 872-890 (2020) DOI: 10.35378/gujs.642935

Gazi University

Journal of Science

http://dergipark.gov.tr/gujs

Adaptive Reweighted Minimum Vector Variance Estimator of Covariance Used for as a New Robust Approach to Partial Least Squares Regression

Esra POLAT[1,*], Hazlina ALI[2]

[1]*Department of Statistics, Faculty of Science, Hacettepe University, 06800, Ankara, Turkey*

[2]*Department of Mathematics & Statistics, School of Quantitative Sciences, College of Arts & Sciences, 06010 Universiti Utara Malaysia, Sintok Kedah, Malaysia*

Highlights

- A new robust PLSR method: PLS-ARWMVV is introduced.
- PLS-ARWMVV is compared with ordinary PLSR and four popular robust PLSR methods.
- The methods are compared in terms of efficiency, fitting to data and prediction capability.
- The proposed robust PLS-ARWMVV is robust, efficient and fitting to data set well.

Article Info

Received: 05/11/2019
Accepted: 10/06/2020

Keywords

Minimum vector Variance (MVV), Partial least squares regression, Robust covariance

Abstract

Partial Least Squares Regression (PLSR), which is developed as partial type of the least squares estimator of regression in case of multicollinearity existence among independent variables, is a linear regression method. If there are outliers in data set, robust methods can be applied for diminishing or getting rid of the negative impacts of them. Past studies have shown that if the covariance matrix is appropriately robustified, PLS1 algorithm (PLSR for one dependent variable) becomes robust against outliers. In this study, an adaptive reweighted estimator of covariance based on Minimum Vector Variance as the first estimator is used and a new robust PLSR method: "PLS-ARWMVV" is introduced. PLS-ARWMVV is compared with ordinary PLSR and four popular robust PLSR methods. The simulation and real data application are revealed that if there are contaminated observations, proposed robust PLS-ARWMVV is robust and efficient. It generally performs better than robust PRM and good alternative for other robust PLS-KurSD, RSIMPLS and PLS-SD methods.

1. INTRODUCTION

PLSR is a well-known method for multivariate analysis. The linear relationship between a group of independent variables and a group of dependent variables could be modeled by using it. The target of PLS is constituting latent variables (components) explaining most of the information (variability) in the explanatory variables that is beneficial for predicting dependent variables as diminishing the dimension by using less components than the number of explanatory variables [1]. These are constructed through the use of latent variables which maximize the covariance between the explanatory variables and dependent variables. This structure following an iterative process to satisfy the components' orthogonality [2]. Because knowing that the existence of outliers or deviations from normal data seriously affects ordinary PLSR, many PLSR methods have been proposed that demonstrate robustness against data contamination [3, 4]. The commonly used NIPALS and SIMPLS algorithms of PLSR are affected by existence of outliers. Various robust types of these algorithms have introduced for one or more dependent variable [5].

Usually robust versions of PLSR have been derived in literature by two strategies: one of them the reduction of the weights of outliers and the other one robustly estimating the covariance matrix. The first method is seen as semi-robust: for example, because of they got non-robust beginning weights or weights' nonresistance towards to leverage points [3]. Wakeling and Macfie [6] studied on PLS with multivariate

*Corresponding author, e-mail: espolat@hacettepe.edu.tr

dependent variables (PLS2) and they changed the regression steps in PLS2 algorithm with M estimations having basis on weighted regressions. An iterative reweighted least squares, Siegel's repeated median and least median of squares for one dependent variable PLS (PLS1) were compared by Griep et al. [7], however, they were unresisting to high leverage outliers [5]. Robust PLSR methods obtained by estimation of covariance robustly, however, overcome to all versions of outliers containing leverage points [3]. For example, by using Stahel-Donoho estimator (SDE) in order to robustly estimate covariance matrix in PLS1 algorithm, Gil and Romera [8] introduced a robust PLSR method [8]. The results of SIMPLS algorithm are influenced by outliers' existence, as the algorithm established on linear least squares regression and on empirical covariance matrix between y and x variables. A robust version of SIMPLS, RSIMPLS, was proposed by Hubert and Vanden Branden [3] used for both one or several dependent variables. In robust RSIMPLS, firstly ROBPCA was applied to x and y variables to obtain robust estimations of Sxy and Sx later that continues in a similar manner to the SIMPLS algorithm. In the second step, a robust regression method (ROBPCA regression) is applied. ROBPCA, a robust PCA technique, associates projection following opinions with Minimum Covariance Determinant (MCD) covariance estimates for low dimensions [3, 9]. A technique named as Partial Robust M (PRM) regression was suggested by Serneels et al. [10] that differed from other robust PLSR methods: robust regression's partial estimator was suggested in place of robust PLS. PRM generated for only univariate dependent variable from SIMPLS algorithm. PRM has provided that giving low weights for both leverage points and vertical outliers using properly selected weighing scheme [10]. Since the name implies, PRM is a partial type of robust M-regression. Weights ranging from zero to one are computed in an iterative system for diminishing outliers' effects in both x and y space. PRM is significantly effective in terms of calculation cost and statistical characteristics [4]. González et al. [5] dealt with PLS with one dependent variable (PLS1) and indicated that in case of appropriately sample covariance matrix robustification, also PLS1 algorithm can becomes robust. Hence, additional robustification of the linear regression stages of PLS1 is unneeded [5].

Here, we are interested on PLS1 and by integrating robust covariance estimators into the classical PLS1 algorithm using the similar approaches as in previous studies of González et al. [5] and Gil and Romera [8]. For our robust approach, the covariance matrix is robustly estimated by using the Minimum Vector Variance (MVV) estimators as first robust estimators of mean and covariance for an adaptive reweighted estimator of covariance.

2. PLS1 ALGORITHM

$z=(y,X)'$ is assumed to be vector of a p+1 dimension with sample size of n that it is separated as a group of p dimensional explanatory variables, x and one dependent variable y. $S_z=\begin{pmatrix} s_y^2 & s'_{y,X} \\ s_{y,X} & S_X \end{pmatrix}$, the sample covariance matrix of z, here $s_{y,X}$ denotes the $p\times 1$ dimensional vector of covariances. Estimation of linear regression model $\hat{y}=\hat{\beta}'x$ and the linearly explanation of the dependent variable could be done by using a set of a components $(t_1,\cdots,t_k)$ (k<<p). These components are linear functions of explanatory variables. Therefore, linear models can be obtained as Equations (1) and (2). Here, X denotes the $n\times p$ dimensional matrix of explanatory variables and x_i' shows its *i*th row [5]

$$x_i = Pt_i + \varepsilon_i \tag{1}$$

$$y_i = q't_i + \eta_i \ . \tag{2}$$

P shows the $p \times k$ dimensional loadings matrix of the vector $t_i = (t_{i1}, \cdots, t_{ik})'$; q denotes vector of the y-loadings with k-dimension. Error vectors ε_i and η_i show normal distributions with zero means and they are not correlated. $T = (t_1, \cdots, t_k)'$ component matrix can not be directly observed and could be estimated by using the maximum likelihood estimation as denoted in Equation (3) [5]

$$T = XW_k \,. \tag{3}$$

$W_k = [w_1, w_2, \cdots, w_k]$ is the $p \times k$ matrix of loadings and $w_i, 1 \le i < k$ vectors are the solutions of Equation (4) considering the constraint given in Equation (5) with $w_1 \alpha s_{y,X}$. As a result it is observed that components $(t_1, \cdots, t_k)$ are orthogonal. w_i vectors are obtained as the eigenvectors related to the greatest eigenvalues of the matrix demonstrated by Equation (6) [5]

$$w_i = \arg\max_{w} \operatorname{cov}^2 (Xw, y) \tag{4}$$

$$w_i' S_x w_j = 0 \text{ and } w'w = 1 \text{ for } 1 \le j < i \tag{5}$$

$$(I - P_X(i)) s_{y,X} s_{y,X}' \,. \tag{6}$$

The space spanned by $S_x W_i$ is showed by the projection matrix: $P_X(i) = (S_X W_i)\left[(S_X W_i)' (S_X W_i) \right]^{-1} (S_X W_i)'$. These results are showed that w_i vectors could be calculated iteratively as in following

$$w_1 \alpha s_{y,X} \tag{7}$$

$$w_{i+1} \alpha s_{y,X} - S_X W_i (W_i' S_X W_i)^{-1} W_i' s_{y,X}, \; 1 \le i < k \,. \tag{8}$$

Since Equation (7) and Equation (8) are obtained, the calculation of PLS components t_i is unneeded. w_{i+1} merely is based on the value of the i former vectors $w_1, w_2, \cdots, w_i$, on S_X and $s_{y,X}$ for each stage of algorithm. Furthermore, as w_1 is merely based on $s_{y,X}$, the computation of W is exactly stabilized by $s_{y,X}$ and S_X values. At last, since the regression coefficients in Equation (2) are not correlated, because of the uncorrelation of the t components, $\hat{\beta}_k^{PLS}$ PLSR coefficients are computed as shown by Equation (9) [5]

$$\hat{\beta}_k^{PLS} = W_k (W_k' S_X W_k)^{-1} W_k' s_{y,X} \,. \tag{9}$$

The practice of algorithm could be considered as a two stage process: (1) w_i weights, identifying the new orthogonal t_i vectors, are calculated using Equations (7) and (8) by the usage of the covariance matrix; (2)

the q_i ,the y-loadings, are calculated by regressing y on single component t_i [5]. Equation (9) clears that these two stages are based merely on the covariance matrix of the samples and replacing these covariance matrices with their robust counterparts will also cause the process becomes robust. Hence, following this idea we have suggested a robust PLSR technique. In the next section, robust covariance estimation method that we used for our approach is given.

3. THE NEW INTRODUCED ROBUST PLSR: PLS-ARWMVV

MVV estimators are used for robustly estimating the covariance used in ordinary PLSR. Our suggested robust PLSR technique will be introduced in this section. The covariance matrix of S_z in PLS1 algorithm will be robustly estimated by using an adaptive reweighted estimator that uses MVV estimators in the initial stage as robust beginning estimators of μ and Σ. Therefore, here equations are applied on $z_i = (y_i, X_i), i = 1, \cdots, n \in \Re^{p'}$, in where $p' = p + 1$. First of all, MVV estimator and functioning of the MVV algorithm is expressed.

3.1. Minimum Vector Variance (MVV) Estimator

One of the latest contributions in the study of robust estimators of μ and Σ is the MVV proposed by Herwindiati [11]. In this study, it was proven that MVV estimators show three main features of well robust estimators: high breakdown point (BP=0.5), affine equivariance and efficiency in terms of computation. Although this estimator appears similar with the popular MCD for its robustness, it has the advantage over MCD in terms of computational efficiency [12, 13]. In addition, MVV estimators are more effective in detecting outliers and in controlling Type I error compared with MCD [14].

The fundamental technique used for estimation of MVV is the Mahalanobis Squared Distances (MSD) that is given as in Equation (10) [14]:

$$d_i^2 = (z_i - \mu)' \Sigma^{-1} (z_i - \mu), \qquad i = 1, 2, \cdots, n \tag{10}$$

where n represents number of observations. Since $z_i' = (y_i, X_i)', i = 1, \cdots, n$ consider a data set $Z_n = \{z_1, \cdots, z_n\}$ of $p' = p + 1$ random variable and be $H \subseteq X$. The ideal value of the number of observations concerned in the calculation of MVV estimators ($\hat{\mu}_{MVV}$, $\hat{\Sigma}_{MVV}$) is $h = \left[\frac{n + p' + 1}{2}\right]$ that creates a covariance matrix $\hat{\Sigma}_{MVV}$ possessing minimal $\mathrm{Tr}\left(\hat{\Sigma}_{MVV}^2\right)$ from whole probable groups of h data. However, it is no guarantee that the iteration process for sets of h data from only one initial h-set can generate the final value of $\mathrm{Tr}\left(\hat{\Sigma}_{MVV}^2\right)$ as an extensive minimum of the MVV objective function. The approximation of MVV estimators could be held by lots of beginning selections of h-subsets. We calculated MVV estimators by MVV algorithm suggested in Yahaya et al. [15] which applies concentration step (C-step) for each initial subset, and then selects a particular number of subsets which generates the lowest vector variance. From the algorithm, the location and covariance estimators are given in Equation (11) [14]

$$\hat{\mu}_{MVV} = \frac{1}{h}\sum_{i=1}^{h} z_i \quad \text{and} \quad \hat{\Sigma}_{MVV} = \frac{1}{h}\sum_{i=1}^{h}\left(z_i - \mu_{MVV}\right)\left(z_i - \mu_{MVV}\right)'. \tag{11}$$

3.2. The MVV Algorithm

To compute the MVV estimators, Yahaya et al. [15] proposed the new MVV algorithm by combination of C-step used in Herwindiati et al. [13]. The C-step is identical to the one in Fast MCD algorithm (used for calculating MCD estimator), with the only difference is the calculation of covariance determinant is changed with the vector variance [15].

Stage 1: Generating Beginning Subsets.
This stage has to be repeated 500 times

1. Pull a random subset (H_0) with number of sample points, $h = p' + 1$. Calculate the mean vector $\bar{z}_{H_0} = \text{average}(H_0)$ and covariance matrix $S_{H_0} = \text{cov}(H_0)$.
2. Compute the MSDs $d_0^2(i) = (z_i - \bar{z}_{H_0})' S_{H_0}^{-1} (z_i - \bar{z}_{H_0})$ for $i = 1, \cdots, n$.
3. Order these MSDs in ascending order, $d_0^2(\pi(1)) \le d_0^2(\pi(2)) \le \cdots \le d_0^2(\pi(n))$. This sorting determines a permutation π on the index set.
4. Draw a new subset $H_1 = \{\pi(1), \cdots, \pi(h)\}$ here $h = \left[\frac{n + p' + 1}{2}\right]$, then calculate $\bar{z}_{H_1}$, S_{H_1}, $\text{Tr}(S_{H_1}^2)$ and compute MSD, where $d_1^2(i) = (z_i - \bar{z}_{H_1})' S_{H_1}^{-1} (z_i - \bar{z}_{H_1})$ for $i = 1, \cdots, n$.
5. Repeat step 3 and 4 for H_2.
6. Put in order the 500 values of $\text{Tr}(S_{H_2}^2)$ in increasing rank, later choose 10 subsets of H_2 having the lowest $\text{Tr}(S_{H_2}^2)$. These subsets are handled as beginning subsets and their mean vectors, $\bar{z}_{H_2}$ and covariance matrices, S_{H_2} are used in Stage 2.

Stage 2: Concentration Steps (C-Step)
This procedure will be duplicated until the convergence for each of the 10 subsets is executed. The convergence is obtained when $\text{Tr}(S_{k-1}^2) = \text{Tr}(S_k^2)$, here k is the number of iterations.

1. Calculate the MSDs by using $\bar{z}_{H_2}$ and S_{H_2} : $d_2^2(i) = (z_i - \bar{z}_{H_2})' S_{H_2}^{-1} (z_i - \bar{z}_{H_2})$ for $i = 1, \cdots, n$
2. Repeat step 3 and 4 in Stage 1 till $\text{Tr}(S_{k-1}^2) = \text{Tr}(S_k^2)$. If $\text{Tr}(S_{k-1}^2) > \text{Tr}(S_k^2)$ the procedure is proceeded. This procedure will be repeated until convergence is executed.
3. When convergence is executed for all the 10 subsets, select the subset (H^*) that creates the lowest $\text{Tr}(S_{H_k}^2)$. From H^*, compute $\bar{z}_{H^*} = \hat{\mu}_{MVV}$ and $S_{H^*} = \hat{\Sigma}_{MVV}$ as the location and covariance estimators for MVV respectively.

3.3. An Adaptive Reweighted Minimum Vector Variance Estimator of Covariance

In linear regression, lots of estimators were suggested for the purpose of providing high efficiency and robustness. To sum up, in case of both robustness and efficiency are important, best selection appears to be a two-step process. Different from Rousseeuw and Van Zomeren [16], Gervini [17] suggested a reweighted one-step estimator using adaptive threshold values. This adaptive reweighting technique can pursue the outlier resistance of the beginning estimator in bias and breakdown and as well obtain 100 % efficiency for normal distribution. Firstly, in Gervini and Yohai [18] this type of adaptive reweighting has been introduced for linear regression. Gervini [17], has widened this opinion and he suggested an adaptive technique for multivariate location and covariance estimation.

For $z_1, \cdots, z_n$ observations in $\Re^{p'}$ with $p' = p+1$ and beginning robust estimators of location and covariance $\left(\hat{\mu}_{0n}, \hat{\Sigma}_{0n}\right)$, the Mahalanobis distances are obtained as in Equation (12) [17]

$$d_i := d\left(z_i, \hat{\mu}_{0n}, \hat{\Sigma}_{0n}\right) = \left\{\left(z_i - \hat{\mu}_{0n}\right)' \hat{\Sigma}_{0n}^{-1} \left(z_i - \hat{\mu}_{0n}\right)\right\}^{1/2}. \tag{12}$$

As it is expected an outlier has a greater Mahalanobis distance than a 'good' sample. In case of normality assumption d_i^2 almost has the distribution of $\chi_{p'}^2$ and the observations with $d_i^2 \geq \chi_{p',0.975}^2$ can be suspected as an outlier. Rousseeuw and Van Zomeren [16] ignore these outliers and calculate the new estimators $\left(\hat{\mu}_{1n}, \hat{\Sigma}_{1n}\right)$ using remaining observations [17].

Herwindiati [11] proposed MVV as an alternative robust estimator of mean and covariance. Herwindiati et al. [13] showed that MVV was used as an objective function in Fast MCD algorithm [19] for substituting the MCD criteria. The results indicated that MVV has better performance in terms of efficiency than Fast MCD and has the similar efficiency with Fast MCD for labelling outliers. Since the MVV method, calculated by algorithm proposed by Yahaya et al. [15], is a good option to MCD and MVV estimators can be operated as the beginning robust estimators of μ and Σ in the 'adaptive reweighted' technique. By following this idea, in this study, putting the MVV estimators $\left(\hat{\mu}_{MVV}, \hat{\Sigma}_{MVV}\right)$ as beginning robust estimators $\left(\hat{\mu}_{0n}, \hat{\Sigma}_{0n}\right)$ in 'adaptive reweighted' technique, the robust $\left(\hat{\mu}_{1n}, \hat{\Sigma}_{1n}\right)$ estimators are obtained and they named as 'Adaptive Reweighted Minimum Vector Variance /ARWMVV' estimators $\left(\hat{\mu}_{ARWMVV}, \hat{\Sigma}_{ARWMVV}\right)$ [15, 17].

Gervini [17] mentioned that this reweighting stage is for efficiency improvement of the beginning estimator while preserving majority of its robustness. Nevertheless $\chi_{p',0.975}^2$ is a subjective threshold value. Although they show the normal distribution, a significant amount of observations is had to be omitted from the analysis in case of large data sets. An alternative method for avoiding from this issue is raising the threshold value to other arbitrary fix value, but in this situation the bias of the reweighted estimator is influenced. Consequently, as a better option 'an adaptive threshold value' is used which shows an increment with n in case of the data is 'clean', however, stays bounded in case of outliers' existence. Gervini [17] has suggested a technique of building up such adaptive threshold values. Let Equation (13) shows squared Mahalanobis distances' empirical distribution [17]

$$G_n(u) = \frac{1}{n}\sum_{i=1}^{n} I\left(d^2\left(z_i, \hat{\mu}_{MVV}, \hat{\Sigma}_{MVV}\right) \leq u\right). \tag{13}$$

$G_{p'}(u)$ is the distribution function of $\chi_{p'}^2$. In case of the data set has a normal distribution the expectation is G_n converging to $G_{p'}$. So that a method of detecting outliers is comparing the tails of G_n with the tails of $G_{p'}$. If $\eta = \chi_{p',1-\alpha}^2$ for a particular low α, for example α=0.025, Equation (14) is defined [17]

$$\alpha_n = \sup_{u \geq \eta} \left\{ G_{p'}(u) - G_n(u) \right\}^+. \tag{14}$$

Here, $\{\cdot\}^+$ shows the positive section. This α_n could be considered as a measure of outliers in the data set. As a negative difference does not show outliers' existence, merely positive differences in Equation (14) are considered. In case of $d^2_{(i)}$ shows the ith sort statistic of the squared Mahalanobis distances and $i_0 = \max\left\{ i : d^2_{(i)} < \eta \right\}$, later Equation (14) becomes as in Equation (15) [17]

$$\alpha_n = \max_{i>i_0} \left\{ G_{p'}\left(d^2_{(i)}\right) - \frac{i-1}{n} \right\}^+ . \tag{15}$$

The observations having the largest $\lfloor \alpha_n n \rfloor$ distances are taken under consideration as outliers and omitted in the reweighting stage. $\lfloor a \rfloor$ shows the largest integer which is less than or equal to a. The cut-off value is given as in Equation (16) in which $G_n^{-1}(u) = \min\left\{ s : G_n(s) \geq u \right\}$. Here $c_n = d^2_{(i_n)}$ with $i_n = n - \lfloor \alpha_n n \rfloor$ and that $i_n > i_0$ as a result of the description of α_n. Therefore, $c_n > \eta$. For defining the reweighted estimator, weights of the styles in Equation (17) are used [17]

$$c_n = G_n^{-1}(1-\alpha_n) \tag{16}$$

$$w_{in} = w\left(\frac{d^2\left(z_i, \hat{\mu}_{MVV}, \hat{\Sigma}_{MVV}\right)}{c_n} \right) . \tag{17}$$

The weight function that provides ***(W)*** $w : [0,\infty) \to [0,1]$ is non-increasing, $w(0) = 1$, $w(u) > 0$ for $u \in [0,1)$ and $w(u) = 0$ for $u \in [1,\infty)$. The easiest selection between those functions fulfilling ***(W)*** is the hard-rejection function $w(u) = I(u < 1)$ that is the most popular used one in application. When the weights in Equation (17) are calculated, the one-step reweighted estimators $\left(\hat{\mu}_{ARWMVV}, \hat{\Sigma}_{ARWMVV}\right)$ are obtained as in Equation (18). Under convenient circumstances, the threshold values in Equation (16) will be in tendency to infinity in case of multivariate normal distribution and later Equation (18) becomes asymptotically equal to general sample mean and covariance, and therefore reach exact asymptotic efficiency [17]

$$\hat{\mu}_{ARWMVV} = \sum_{i=1}^{n} w_{in} z_i \Big/ \sum_{i=1}^{n} w_{in} \qquad \hat{\Sigma}_{ARWMVV} = \sum_{i=1}^{n} w_{in} \left(z_i - \hat{\mu}_{ARWMVV}\right)\left(z_i - \hat{\mu}_{ARWMVV}\right)' \Big/ \sum_{i=1}^{n} w_{in} . \tag{18}$$

As a result, $\hat{\Sigma}_{ARWMVV}$ robust covariance estimator in Equation (18) is used for computing the robust covariance estimator: $\hat{S}_z = \begin{pmatrix} \hat{s}_y^2 & \hat{s}'_{y,X} \\ \hat{s}_{y,X} & \hat{S}_X \end{pmatrix}$. Then, by using $\hat{S}_z$ in the alternative definition of PLS1 algorithm given between Equations (7)-(9), robust "PLS-ARWMV" is suggested by us. Our new robust PLS-ARWMVV algorithm can be summarized as shown in Equation (19). $\hat{s}_{y,X}$ and $\hat{S}_X$ are calculated by

decomposing the robust covariance estimation of $z_i^{'} = \left(y_i, X_i\right)^{'}, i = 1, \cdots, n$ that is computed by ARWMVV method.

$$
\begin{aligned}
& w_1 \alpha \hat{s}_{y,X} \\
& w_{i+1} \alpha \hat{s}_{y,X} - \hat{S}_X W_i \left(W_i' \hat{S}_X W_i\right)^{-1} W_i' \hat{s}_{y,X}, \; 1 \le i < k \\
& \hat{\beta}_k^{PLS-ARWMVV} = W_k \left(W_k' \hat{S}_X W_k\right)^{-1} W_k' \hat{s}_{y,X}
\end{aligned} \tag{19}
$$

4. SIMULATION STUDY

Here, the comparison of robust methods; PLS-SD [8], RSIMPLS [3], PRM [10], PLS-KurSD [5] and the ordinary PLSR with our new method PLS-ARWMVV will be done about efficiency, fitting to data and prediction capability by doing simulation on clean and contaminated data sets. In regards to the first models denoted by Equations (1) and (2), and using a simulation design close as the one given in [5] the data sets are created as shown in Equation (20)

$$
T \sim N_2\left(0_2, \sum_t\right) \quad X = TI_{2,p} + N_p\left(0_p, 0.1\, I_p\right) \quad y = TA_{2,1} + N(0,1). \tag{20}
$$

I_p shows $p \times p$ dimensional unit matrix, $\left(I_{k,p}\right)_{i,j} = 0$ and $\left(I_{k,p}\right)_{i,j} = 1$ for $i = j$. $0_2 = (0,0)^{'}$ is a vector of zeros, $T_{n \times 2}$ is the component matrix, $A_{2,1} = (1,1)^{'}$ denotes a vector of ones. k=2 and $\sum_t = \begin{pmatrix} 4 & 0 \\ 0 & 2 \end{pmatrix}$. The comparison is made between our PLS-ARWMVV technique and four well-known robust PLSR techniques and ordinary PLSR method for five kinds of outliers.

1. Bad leverage points are sample points that are with large distance from the regression hyperplane though they projects on the regression hyperplane falling outside the vast bulk of the projected sample points (clean ones): $T_\in \sim N_2\left(10_2, \sum_t\right) \quad X_\in = T_\in I_{2,p} + N_p\left(0_p, 0.1\, I_p\right)$.

2. Vertical outliers are sample points far away from the hyperplane, however, they are projected within the vast bulk of the projected sample points: $y_\in = TA_{2,1} + N(10.0.1)$.

3. Good leverage points are in the neighborhood of the hyperplane, however, they are distant from the group of the vast majority of the samples: $T_\in \sim N_2\left(10_2, \sum_t\right) \quad X_\in = T_\in I_{2,p} + N_p\left(\left(0_2, 10_{p-2}\right), 0.1\, I_p\right)$.

4. Concentrated Outliers are groups of bad leverage points: $T_\in \sim N_2\left(10_2, \sum_t\right) \quad X_\in = T_\in I_{2,p} + N_p\left(10_p, 0.001\, I_p\right)$.

5. Orthogonal outliers lie distant from the t-space, however, they become good sample points after projecting in t-space. Therefore, it does not severely affect the calculation of the regression coefficients, however, they may affect the loadings: $X_\in = TI_{2,p} + N_p\left(\left(0_2, 10_{p-2}\right), 0.1\, I_p\right)$

For each of the cases, m=1000 data sets were created. The efficiency of the techniques is determined using Equation (21). The real coefficient vector is computed as $\beta_{p,1} = I_{p,2}' A_{2,1}$. $\hat{\beta}_k^{(l)}$ shows parameter estimation

for k components in the lth simulation. MSE denotes to what extent the parameter is truly estimated. MSE value close to zero is preferred [9]

$$\mathrm{MSE}_k\left(\hat{\beta}\right)=\frac{1}{m}\sum_{l=1}^{m}\left\|\hat{\beta}_k^{(l)}-\beta\right\|^2 \quad . \tag{21}$$

The methods' performances in terms of fitting to the regular observations (G_r) is considered. This measure is given as in Equation (22). $r_{i,k}$ shows residual of the ith sample point for model with k components. Aim is obtaining a GOF value near to 1 [9]

$$\mathrm{GOF}_k=1-\frac{\underset{i\in G_r}{\mathrm{var}}\left(r_i,k\right)}{\underset{i\in G_r}{\mathrm{var}}\left(y_i\right)} \quad . \tag{22}$$

Root Mean Squared Error (RMSE) can be used for measuring the predictive capability of the techniques. Firstly, a test set G_t of clean observations (size of n_t equals n/2) is created and later Equation (23) is calculated. $\hat{y}_{i,k}$ denotes the prediction for y-value of sample point i from the test set in case of the estimations of regression coefficient are obtained from training set (in which sample size n and number of components is k) [9]

$$\mathrm{RMSE}_k=\sqrt{\frac{1}{n_t}\sum_{i=1}^{n_t}\left(y_i-\hat{y}_{i,k}\right)^2} \quad . \tag{23}$$

m=1000 repetitions done, the mean angle among the estimated parameter $\hat{\beta}_{[y_\varepsilon,X_\varepsilon],k}$ and the true parameter β are also calculated [3, 5].

Table 1. *Clean data*

n		**p**	PLSR	RSIMPLS	PRM	PLS-SD	PLS-KurSD	***PLS-ARWMVV***
100	MSE	**6**	***0.0167***	0.0221	0.0183	0.0203	0.0220	0.0805
		12	***0.0251***	0.0365	0.0263	0.0289	0.0504	0.1136
	GOF	**6**	***0.8301***	0.8292	0.8293	0.8293	0.8291	0.8168
		12	***0.8321***	0.8315	0.8310	0.8315	0.8295	0.8225
	RMSE	**6**	***1.0993***	1.1020	1.1004	1.1015	1.1018	1.1393
		12	***1.1037***	1.1083	1.1053	1.1059	1.1134	1.1344
	Mean(angle)	**6**	***0.0673***	0.0801	0.0690	0.0758	0.0769	0.1480
		12	***0.0944***	0.1133	0.0948	0.1021	0.1256	0.1841
200	MSE	**6**	***0.0100***	0.0127	0.0111	0.0115	0.0112	0.0332
		12	***0.0140***	0.0174	0.0147	0.0147	0.0154	0.0381
	GOF	**6**	***0.8295***	0.8291	0.8291	0.8292	0.8292	0.8244
		12	***0.8317***	0.8314	0.8311	0.8315	0.8314	0.8283
	RMSE	**6**	***1.0962***	1.0978	1.0968	1.0971	1.0969	1.1123
		12	***1.0999***	1.1015	1.1008	1.1004	1.1007	1.1108
	Mean(angle)	**6**	***0.0495***	0.0583	0.0509	0.0538	0.0529	0.0938
		12	***0.0673***	0.0778	0.0674	0.0699	0.0716	0.1139

The results for clean data is given in Table 1. Table 1 shows that in case of no contamination, when the sample size increases all of methods' performances getting better in terms of efficiency, however, the increase in n value has no significant impact on fitting to data and prediction capability performances of methods. Moreover, it could be mentioned that proposed PLS-ARWMVV method's both of efficiency and prediction performance is affected positively from sample size increment. When p increases, GOFs and RMSE values are not significantly affected, while, MSE and mean (angle) values are affected badly, especially for smaller sample size. If the data set is clean classical method come to forefront regardless of the n, ε and p, which is the expected case.

The simulation settings in Table 2, by changing first 10 % and 20 % of the observations with below mentioned kinds of outliers, are used.

Table 2. *The simulation settings*

Variables	**Values**
Number of sample sizes (n)	100 and 200
Number of independent variables (p)	6 and 12
Proportion of contamination (ε)	0.1 and 0.2
Kinds of outliers	bad leverage points, vertical outliers, good leverage points, concentrated outliers, orthogonal outliers

The results for each types of outliers are given in separate tables across Tables 3-7. Hence, the performances of the existing robust methods and new proposed one is observed for each outliers' types.

Table 3. *Bad leverage points*

n		**p**	**ε**	PLSR	RSIMPLS	PRM	PLS-SD	PLS-KurSD	***PLS-ARWMVV***
100	MSE	**6**	**0.1**	1.7414	***0.0238***	0.0879	0.1290	***0.0243***	***0.0669***
			0.2	1.9109	***0.0258***	1.7915	0.5601	***0.0287***	***0.0578***
		12	**0.1**	1.7533	***0.0380***	0.1073	0.2709	***0.0546***	***0.0952***
			0.2	1.9805	***0.0398***	1.8153	1.1518	***0.0849***	***0.0772***
	GOF	**6**	**0.1**	0.2651	***0.8285***	0.8136	0.8029	***0.8284***	***0.8197***
			0.2	0.1879	***0.8296***	0.2413	0.6919	***0.8285***	***0.8236***
		12	**0.1**	0.2663	***0.8339***	0.8143	0.7755	***0.8313***	***0.8271***
			0.2	0.1911	***0.8322***	0.2446	0.5647	***0.8239***	***0.8274***
	RMSE	**6**	**0.1**	2.2825	***1.1008***	1.1497	1.1836	***1.1010***	***1.1279***
			0.2	2.4006	***1.1052***	2.3216	1.4930	***1.1076***	***1.1254***
		12	**0.1**	2.2788	***1.0991***	1.1554	1.2727	***1.1048***	***1.1180***
			0.2	2.3958	***1.0991***	2.3134	1.7810	***1.1188***	***1.1137***
	Mean(angle)	**6**	**0.1**	1.1403	0.0839	0.1076	0.1291	0.0824	0.1334
			0.2	1.3049	0.0852	1.1784	0.3422	0.0883	0.1251
		12	**0.1**	1.1700	0.1159	0.1345	0.2119	0.1354	0.1755
			0.2	1.3511	0.1212	1.2208	0.6931	0.1519	0.1572
200	MSE	**6**	**0.1**	1.7093	***0.0130***	0.0697	0.1150	***0.0121***	***0.0270***
			0.2	1.8979	***0.0135***	1.7712	0.5041	***0.0133***	***0.0238***
		12	**0.1**	1.7193	***0.0174***	0.0708	0.2370	***0.0161***	***0.0337***
			0.2	1.8998	***0.0191***	1.7802	0.9312	0.0479	***0.0308***
	GOF	**6**	**0.1**	0.2613	***0.8290***	0.8159	0.8029	***0.8291***	***0.8257***
			0.2	0.1836	***0.8296***	0.2391	0.6918	***0.8298***	***0.8274***
		12	**0.1**	0.2625	***0.8319***	0.8186	0.7685	***0.8318***	***0.8293***
			0.2	0.1850	***0.8320***	0.2380	0.5579	***0.8239***	***0.8302***
	RMSE	**6**	**0.1**	2.2833	***1.0994***	1.1422	1.1818	***1.0987***	***1.1088***
			0.2	2.3959	***1.0981***	2.3145	1.4789	***1.0977***	***1.1053***
		12	**0.1**	2.2800	***1.0946***	1.1367	1.2838	***1.0945***	***1.1030***
			0.2	2.3930	***1.0957***	2.3148	1.7692	***1.1153***	***1.1012***
	Mean(angle)	**6**	**0.1**	1.1289	0.0587	0.0837	0.1068	0.0562	0.0851
			0.2	1.3084	0.0585	1.1826	0.3044	0.0593	0.0811
		12	**0.1**	1.1465	0.0779	0.0941	0.1676	0.0740	0.1079
			0.2	1.3178	0.0813	1.1995	0.5710	0.0968	0.1048

The results when the data set is contaminated by bad leverage points is presented in Table 3. In this case, as it is expected for all of conditions robust methods (included the new proposed robust PLS-ARWMVV) outperform the traditional PLSR particularly in terms of predictive capability and efficiency. If the proportion of bad leverage points is 0.1, when p increases, PLS-ARWMVV outperforms classical PLSR, robust PLS-SD and PRM in terms of fitting, efficiency and prediction abilities. Moreover, if n increases, for both of dimensions PLS-ARWMVV outperforms classical PLSR, robust PLS-SD and PRM in terms of efficiency and prediction abilities that in higher dimension performance of PLS-ARWMVV is getting better against these two robust methods. If the proportion of bad leverage points, getting higher as 0.2, regardless

of dimension and sample size, the new proposed robust PLS-ARWMVV, PLS-KurSD and RSIMPLS techniques become prominent techniques with better fitting, efficiency and prediction properties, additionally, they have lower mean angle values compared with robust PLS-SD and PRM. In case of high proportion of bad leverage points existence, the new robust PLS-ARWMVV come to forefront as a second one following RSIMPLS with its better efficiency and prediction performances for high dimension of p=12 regardless of sample size. Overall, in case of bad leverage points existence RSIMPLS, PLS-KurSD and PLS-ARWMVV come to forefronts methods.

Table 4. *Vertical outliers*

n		**p**	**ε**	PLSR	RSIMPLS	PRM	PLS-SD	PLS-KurSD	***PLS-ARWMVV***
100	MSE	**6**	**0.1**	0.1107	0.0228	0.0232	0.0266	0.0241	0.0630
			0.2	0.1987	0.0221	0.0340	0.0397	0.0282	0.0567
		12	**0.1**	0.1655	0.0321	0.0321	0.0399	0.0482	0.0740
			0.2	0.3826	***0.0299***	***0.0446***	0.1237	0.1908	0.2543
	GOF	**6**	**0.1**	0.8040	0.8287	0.8283	0.8288	0.8289	0.8216
			0.2	0.7859	0.8293	0.8268	0.8262	0.8295	0.8242
		12	**0.1**	0.8062	0.8322	0.8315	0.8309	0.8306	0.8279
			0.2	0.7821	***0.8309***	***0.8288***	0.8153	0.8143	0.8145
	RMSE	**6**	**0.1**	1.1768	1.1042	1.1066	1.1058	1.1047	1.1292
			0.2	1.2287	1.1031	1.1142	1.1189	1.1054	1.1214
		12	**0.1**	1.1796	1.1028	1.1042	1.1090	1.1078	1.1178
			0.2	1.2425	***1.1046***	***1.1138***	1.1592	1.1557	1.1532
	Mean(angle)	**6**	**0.1**	0.1727	0.0764	0.0789	0.0843	0.0824	0.1329
			0.2	0.2257	0.0731	0.0947	0.1054	0.0893	0.1253
		12	**0.1**	0.2332	0.1058	0.1049	0.1198	0.1314	0.1612
			0.2	0.3120	***0.0982***	***0.1240***	0.2006	0.1923	0.1968
200	MSE	**6**	**0.1**	0.0522	0.0106	0.0124	0.0123	0.0112	0.0250
			0.2	0.0835	0.0117	0.0182	0.0201	0.0134	0.0257
		12	**0.1**	0.0678	0.0153	0.0174	0.0200	0.0170	0.0324
			0.2	0.1149	***0.0143***	***0.0234***	0.0510	0.0369	0.0502
	GOF	**6**	**0.1**	0.8171	0.8294	0.8294	0.8296	0.8299	0.8266
			0.2	0.8096	0.8311	0.8300	0.8298	0.8317	0.8294
		12	**0.1**	0.8172	0.8291	0.8293	0.8291	0.8299	0.8275
			0.2	0.8072	***0.8298***	***0.8291***	0.8230	0.8271	0.8268
	RMSE	**6**	**0.1**	1.1359	1.0982	1.0987	1.0985	1.0974	1.1077
			0.2	1.1628	1.0992	1.1041	1.1055	1.0989	1.1066
		12	**0.1**	1.1362	1.1005	1.1010	1.1022	1.0996	1.1071
			0.2	1.1653	***1.0971***	***1.1025***	1.1232	1.1095	1.1092
	Mean(angle)	**6**	**0.1**	0.1196	0.0482	0.0541	0.0560	0.0532	0.0831
			0.2	0.1520	0.0495	0.0671	0.0723	0.0584	0.0808
		12	**0.1**	0.1559	0.0672	0.0736	0.0823	0.0757	0.1069
			0.2	0.2010	***0.0636***	***0.0870***	0.1319	0.1021	0.1156

The results when the data set is contaminated by vertical outliers is presented in Table 4. In this case, as it is expected for all situations robust methods (included the new proposed robust PLS-ARWMVV) are

superior to the ordinary PLSR particularly in terms of predictive capability and efficiency. In case of vertical outliers' presence, generally robust RSIMPLS is the leading one especially in terms of efficiency and prediction. Especially, in case of n=100, p=12 and ε=0.2, the RSIMPLS method's efficiency and prediction performance compared to other robust methods is attractive. Also it is clear that PRM shows good performance in case of vertical outliers that regardless of sample size in higher dimension, existence of higher percentage of outliers in the data, it is the second robust method especially in terms of efficiency and prediction. However, it must be mentioned that, when sample size increases robust methods performances get closer to RSIMPLS.

Table 5. *Good leverage points*

n		**p**	**ε**	PLSR	RSIMPLS	PRM	PLS-SD	PLS-KurSD	***PLS-ARWMVV***
100	MSE	**6**	**0.1**	0.8877	***0.0243***	0.8529	***0.0274***	***0.0242***	0.0617
			0.2	0.8767	***0.0273***	0.9132	0.0652	***0.0277***	***0.0560***
		12	**0.1**	0.5447	***0.0383***	0.4965	***0.0352***	***0.0574***	0.0797
			0.2	0.5975	***0.0401***	0.5757	0.0969	***0.0589***	***0.0685***
	GOF	**6**	**0.1**	0.7128	*0.8286*	0.7718	0.8284	0.8284	0.8204
			0.2	0.6886	0.8297	0.7047	0.8270	0.8287	0.8239
		12	**0.1**	0.7484	0.8339	0.7818	0.8336	0.8312	0.8284
			0.2	0.7198	0.8322	0.7409	0.8300	0.8293	0.8282
	RMSE	**6**	**0.1**	1.4317	***1.1008***	1.2768	***1.1008***	***1.1009***	1.1258
			0.2	1.4885	***1.1052***	1.4516	***1.1139***	***1.1071***	1.1236
		12	**0.1**	1.3345	***1.0989***	1.2478	***1.0980***	***1.1059***	1.1144
			0.2	1.4007	***1.0991***	1.3469	***1.1065***	***1.1062***	1.1102
	Mean(angle)	**6**	**0.1**	0.7091	0.0850	0.6501	0.0860	0.0823	0.1309
			0.2	0.7135	0.0877	0.7177	0.1462	0.0882	0.1214
		12	**0.1**	0.5395	0.1162	0.5079	0.1128	0.1365	0.1645
			0.2	0.5522	0.1219	0.5586	0.2005	0.1412	0.1531
200	MSE	**6**	**0.1**	0.8646	***0.0132***	0.8424	***0.0166***	***0.0121***	0.0261
			0.2	0.8616	***0.0143***	0.8977	0.0433	***0.0133***	***0.0234***
		12	**0.1**	0.5345	***0.0176***	0.4767	***0.0195***	***0.0161***	0.0306
			0.2	0.5726	***0.0198***	0.5525	0.0682	***0.0188***	***0.0297***
	GOF	**6**	**0.1**	0.7144	0.8291	0.7785	0.8289	0.8291	0.8260
			0.2	0.6879	0.8298	0.7040	0.8275	0.8298	0.8275
		12	**0.1**	0.7462	0.8319	0.7818	0.8318	0.8318	0.8296
			0.2	0.7220	0.8322	0.7433	0.8304	0.8319	0.8303
	RMSE	**6**	**0.1**	1.4218	***1.0993***	1.2525	***1.0992***	***1.0987***	***1.1083***
			0.2	1.4831	***1.0982***	1.4462	***1.1053***	***1.0977***	***1.1052***
		12	**0.1**	1.3408	***1.0947***	1.2442	***1.0946***	***1.0945***	***1.1017***
			0.2	1.3963	***1.0957***	1.3432	***1.1011***	***1.0955***	***1.1009***
	Mean(angle)	**6**	**0.1**	0.7028	0.0599	0.6438	0.0630	0.0562	0.0840
			0.2	0.7111	0.0619	0.7147	0.1208	0.0591	0.0810
		12	**0.1**	0.5371	0.0784	0.5014	0.0801	0.0740	0.1056
			0.2	0.5400	0.0838	0.5491	0.1726	0.0805	0.1036

The results when the data set is contaminated by good leverage points is presented in Table 5. If the proportion of good leverage points is 0.1 or 0.2, new PLS-ARWMVV and robust RSIMPLS, PLS-SD, PLS-KurSD techniques perform better than classical one and robust PRM in terms of goodness of fit, efficiency and prediction capability for all cases. In Table 5, for each condition the best three methods in terms of efficiency and prediction performances are showed. In case of good leverage points presence, RSIMPLS and PLS-KurSD methods seem generally leading ones. It is obvious that sometimes PLS-ARWMVV performs even better than PLS-SD. Moreover, when n increases especially prediction performances get closer for these four methods.

Table 6. *Concentrated outliers*

n		**p**	**ε**	PLSR	RSIMPLS	PRM	PLS-SD	PLS-KurSD	***PLS-ARWMVV***
100	MSE	**6**	**0.1**	1.8488	***0.0240***	1.6460	***0.0446***	***0.0243***	0.0626
			0.2	1.7327	***0.0272***	*1.8448	0.3145	***0.0272***	***0.0514***
		12	**0.1**	1.1890	***0.0369***	1.1751	***0.0519***	***0.0481***	0.0789
			0.2	1.1432	***0.0400***	*1.2337	0.9219	***0.0532***	***0.0747***
	GOF	**6**	**0.1**	0.5653	0.8286	0.6999	0.8262	0.8284	0.8205
			0.2	0.5481	0.8297	0.5428	0.8090	0.8290	0.8240
		12	**0.1**	0.6824	0.8339	0.7206	0.8317	0.8316	0.8286
			0.2	0.6573	0.8322	0.6660	0.8134	0.8304	0.8270
	RMSE	**6**	**0.1**	1.7647	***1.1008***	1.4527	***1.1083***	***1.1010***	1.1245
			0.2	1.7943	***1.1053***	*1.8072	1.1723	***1.1062***	***1.1235***
		12	**0.1**	1.5017	***1.0988***	1.4121	***1.1029***	***1.1045***	1.1133
			0.2	1.5540	***1.0991***	1.5349	1.1697	***1.1035***	***1.1145***
	Mean(angle)	**6**	**0.1**	1.0411	0.0845	0.8267	0.0997	0.0823	0.1296
			0.2	1.0450	0.0878	*1.0501	0.3215	0.0877	0.1210
		12	**0.1**	0.8216	0.1151	0.7749	0.1334	0.1292	0.1620
			0.2	0.8309	0.1217	*0.8377	0.6048	0.1304	0.1577
200	MSE	**6**	**0.1**	1.7954	***0.0132***	1.6880	0.0300	***0.0121***	***0.0248***
			0.2	1.6823	***0.0142***	*1.7812	0.1999	***0.0132***	***0.0228***
		12	**0.1**	1.1611	***0.0176***	1.1351	0.0346	***0.0161***	***0.0303***
			0.2	1.1076	***0.0198***	*1.2030	0.7573	***0.0188***	***0.0315***
	GOF	**6**	**0.1**	0.5600	0.8291	0.7101	0.8267	0.8291	0.8260
			0.2	0.5391	0.8298	*0.5332	0.8108	0.8298	0.8275
		12	**0.1**	0.6773	0.8319	0.7159	0.8301	0.8318	0.8296
			0.2	0.6571	0.8322	0.6657	0.8149	0.8319	0.8301
	RMSE	**6**	**0.1**	1.7649	***1.0993***	1.4200	***1.1067***	***1.0987***	***1.1076***
			0.2	1.8029	***1.0982***	*1.8166	1.1582	***1.0977***	***1.1050***
		12	**0.1**	1.5132	***1.0946***	1.4212	***1.1006***	***1.0945***	***1.1017***
			0.2	1.5520	***1.0957***	1.5341	1.1531	***1.0955***	***1.1016***
	Mean(angle)	**6**	**0.1**	1.0367	0.0599	0.8332	0.0761	0.0562	0.0829
			0.2	1.0437	0.0619	1.0485	0.2728	0.0591	0.0806
		12	**0.1**	0.8199	0.0783	0.7718	0.1030	0.0740	0.1048
			0.2	0.8219	0.0838	0.8305	0.5578	0.0804	0.1045

It is very hard to handle with the concentrated outliers. The results when the data set is contaminated by concentrated outliers is presented in Table 6. If the proportion of concentrated outliers is 0.1, new PLS-ARWMVV and robust RSIMPLS, PLS-SD, PLS-KurSD techniques perform better than classical one and robust PRM in terms of goodness of fit, efficiency and prediction capability for both of sample sizes and dimensions. In Table 6, for each condition the best three methods in terms of efficiency and prediction performances are especially showed. If the proportion of concentrated outliers is 0.2, PLS-ARWMVV is more efficient, fits data better and has a better prediction capability than both PRM and PLS-SD, additionally, PLS-ARWMVV has a smaller mean angle values than these two techniques. Overall, in the case of concentrated outliers' presence, RSIMPLS and PLS-KurSD methods seem generally leading ones. Moreover, when the proportion of concentrated outliers increases as ε=0.2, PRM performs worse even than classic PLSR method.

Table 7. *Orthogonal outliers*

n		**p**	**ε**	PLSR	RSIMPLS	PRM	PLS-SD	PLS-KurSD	***PLS-ARWMVV***
100	MSE	**6**	**0.1**	0.2047	***0.0297***	0.1638	***0.0226***	***0.0243***	0.0616
			0.2	0.2228	***0.0436***	*0.2552	***0.0242***	***0.0309***	0.0586
		12	**0.1**	0.2269	***0.0430***	0.2052	***0.0304***	***0.0529***	0.0759
			0.2	0.2362	***0.0689***	*0.2701	***0.0329***	***0.0505***	0.0713
	GOF	**6**	**0.1**	0.7808	0.8295	0.7930	0.8301	0.8298	0.8221
			0.2	0.7731	0.8273	*0.7652	0.8289	0.8283	0.8234
		12	**0.1**	0.7739	0.8304	0.7812	0.8309	0.8282	0.8253
			0.2	0.7730	0.8308	*0.7647	0.8320	0.8299	0.8285
	RMSE	**6**	**0.1**	1.2453	***1.1049***	1.2110	***1.1025***	***1.1038***	1.1278
			0.2	1.2657	***1.1151***	*1.2883	***1.1074***	***1.1094***	1.1298
		12	**0.1**	1.2620	***1.1020***	1.2416	***1.0972***	***1.1058***	1.1161
			0.2	1.2605	***1.1035***	*1.2837	***1.0933***	***1.1006***	***1.1077***
	Mean(angle)	**6**	**0.1**	0.2956	***0.0925***	0.2548	***0.0799***	***0.0832***	0.1306
			0.2	0.3108	***0.1126***	*0.3332	***0.0813***	***0.0901***	0.1257
		12	**0.1**	0.3150	***0.1230***	0.2930	***0.1045***	***0.1332***	0.1632
			0.2	0.3221	0.1534	*0.3461	***0.1094***	***0.1352***	***0.1530***
200	MSE	**6**	**0.1**	0.1884	***0.0147***	0.1542	***0.0115***	***0.0116***	0.0280
			0.2	0.2108	***0.0224***	*0.2440	***0.0122***	***0.0128***	0.0259
		12	**0.1**	0.2177	***0.0229***	0.2007	***0.0154***	***0.0168***	0.0298
			0.2	0.2194	0.0361	*0.2540	***0.0163***	***0.0183***	***0.0289***
	GOF	**6**	**0.1**	0.7817	0.8292	0.7922	0.8296	0.8295	0.8263
			0.2	0.7784	0.8305	*0.7701	0.8318	0.8318	0.8293
		12	**0.1**	0.7752	0.8300	0.7806	0.8305	0.8303	0.8281
			0.2	0.7752	0.8297	*0.7664	0.8314	0.8313	0.8295
	RMSE	**6**	**0.1**	1.2408	***1.0997***	1.2115	***1.0988***	***1.0990***	1.1089
			0.2	1.2595	***1.1071***	*1.2841	***1.1030***	***1.1031***	1.1114
		12	**0.1**	1.2557	***1.0989***	1.2415	***1.0963***	***1.0973***	1.1032
			0.2	1.2610	1.1069	*1.2856	***1.0991***	***1.1000***	***1.1055***
	Mean(angle)	**6**	**0.1**	0.2884	0.0615	0.2529	0.0533	0.0538	0.0860
			0.2	0.3087	0.0780	*0.3335	0.0555	0.0578	0.0807

		12	**0.1**	0.3146	0.0869	0.2976	0.0713	0.0747	0.1032
			0.2	0.3158	0.1098	*0.3416	0.0745	0.0796	0.1025

The results when the data set is contaminated by orthogonal outliers is presented in Table 7. If the proportion of orthogonal outliers is 0.1, RSIMPLS, PLS-SD and PLS-KurSD methods come to forefront especially in terms of efficiency and prediction ability. If the proportion of orthogonal outliers is 0.2, new proposed PLS-ARWMVV and robust RSIMPLS, PLS-SD, PLS-KurSD techniques outperform both ordinary PLSR and robust PRM in terms of efficiency, fitting to data and prediction ability regardless of sample and dimension sizes. Overall, in case of orthogonal outliers' presence all robust methods except PRM perform good. Moreover, when the proportion of orthogonal outliers increases as ε=0.2, PRM performs worse even than classic PLSR method.

Overall, in situation of moderate proportion of different kinds of outliers' existence (ε=0.1), regardless of sample sizes and dimensions, the three robust PLSR techniques of literature (RSIMPLS, PLS-SD, PLS-KurSD) and the new robust PLS-ARWMVV outperform the ordinary PLSR particularly in terms of prediction capability and efficiency. Although PRM also performs better than ordinary PLSR, for especially good leverage points, concentrated outliers and orthogonal outliers it performs badly in comparison with the other robust techniques. For the whole kinds of outliers (vertical outliers not included) if the percentage of outliers increase as ε=0.2, robust PRM loose its efficiency and prediction capability and fitting worse, additionally, the mean angle values get higher compared to the other robust PLSR techniques (containing new PLS-ARWMVV). Regardless of sample sizes, dimensions and proportions of outliers, the mean angle values of PLS-ARWMVV is lower than the ordinary technique, however, is lower than the PRM for only good leverage points, concentrated outliers and orthogonal outliers. The remarkable point that if ratio of orthogonal outliers is reached a high proportion as 0.2, PRM has a bad performance similar as ordinary PLSR that it is less efficient and has a worse predictive capability than ordinary PLSR, moreover, for concentrated outliers PRM has worse efficiency and sometimes worse prediction ability than it. It can be referred that if the percentage of outliers reaches a high ratio 0.2, proposed PLS-ARWMVV technique still has better efficiency, prediction capability and it fits to data better compared to ordinary PLSR for whole kinds of outliers.

5. REAL-LIFE FISH DATA SET APPLICATION

PLS-ARWMVV and four well-known robust PLSR techniques are compared on fish data of Naes [20]. The performances of methods are evaluated using Equation (22) and Equation (23) with regards to fitting and prediction capability. The final 7 of 45 sample points of this data are outlying observations. Fat concentration (ratio, %) of 45 fish observations (rainbow trout) and explanatory variables of the absorbance at 9 Near Infrared Reflectance (NIR) wavelengths measurements were obtained after sample homogenization. The purpose is modelling the associations among the fat concentration (dependent variable) and these 9 spectrums (explanatory variables). Data set is partitioned in two pieces as the first 10 sample points are the test data while the rest of 35 observations are the training data [7, 20-22].

RMSE is computed from training set containing 7 outlying observations (the ratio of outliers' is 20%). Later, using regression parameters computed from six different models, the predictions are obtained for uncontaminated test data with 10 observations. The ideal numbers of components are selected as k_{opt}=3 as given in [22]. From Table 8 it is clear that the new robust PLS-ARWMVV has a higher performance than ordinary PLSR and robust PRM both in terms of prediction and fitting abilities. It has also better prediction performance than robust PLS-SD, while it shows a nearly similar performance in terms of fitting. Moreover, PLS-ARWMVV has a close performance to the both robust RSIMPLS and PLS-KurSD.

Table 8. *The results for fish data*

	PLSR	RSIMPLS	PRM	PLS-SD	PLS-KurSD	***PLS-ARWMVV***
GOF	0.9306	0.9741	0.7045	0.9720	0.9750	0.9742
RMSE	1.7827	1.4642	1.6713	1.5598	1.4495	1.4746

6. CONCLUSION

In this paper, a new robust PLSR technique is proposed, named as "PLS-ARWMVV", for the linear regression model with one dependent variable for obtaining robust and efficient results in case of outliers' existence. Doing a simulation study, proposed robust PLSR technique is compared with ordinary PLSR and four robust well-known PLSR techniques of literature in terms of fitting to data, efficiency and predictive ability on clean and contaminated data sets. 10 % and 20 % of this data set is changed with outlying observations. Hence, the increase in the ratio of outliers effects on performances of techniques is examined.

The real data application indicates that PLS-ARWMVV outperforms both ordinary PLSR and robust PRM in terms of prediction and fitting abilities. Also, it has a better predictive capability in comparison with robust PLS-SD but they show similar fitting performance. Moreover, PLS-ARWMVV shows nearly a similar performance with robust RSIMPLS and PLS-KurSD.

The results of simulation conclude that in situation of 10 % or 20 % of bad leverage points' existence, PLS-ARWMVV performs better compared to both PLS-SD and PRM in terms of fitting, efficiency and prediction abilities. In case of 10 % or 20 % of vertical outliers' existence, PLS-ARWMV is not leading one but still robust as compared to classical one and follows to four robust PLSR techniques in terms of efficiency, fitting and prediction capabilities. The situation of 10 % or 20 % of good leverage points' presence, PLS-ARWMVV performs much better compared to PRM in terms of fitting, prediction and efficiency. If data contains 20 % of good leverage points, PLS-ARWMVV has a better efficiency than PLS-SD. If the ratio of concentrated outliers is 10 %, PLS-ARWMVV has a better efficiency, fitting to data better and a higher predictive ability than PRM. Presence of 20 % of concentrated outliers indicates that PLS-ARWMVV is more efficient and better predictive than both PRM and PLS-SD. Presence 10 % or 20 % of orthogonal outliers shows that PLS-ARWMVV fits better, predicts better and is more efficient than PRM.

Overall, new robust PLS-ARWMVV could cope with various outliers efficiently and gives robust predictions, especially, it performs better than robust PRM (with the exception of vertical outliers that in this case PRM performs better than PLS-ARWMVV, however, when sample size increases their performances getting closer). PLS-ARWMVV also gives better results than robust PLS-SD in some cases. These results are also supported by real data application. Hence, PLS-ARWMVV is a good option to robust PLS-SD, RSIMPLS, PLS-KurSD and PRM techniques of robust PLSR literature that in some cases it outperforms some of them or shows a close performance with these four.

CONFLICTS OF INTEREST

No conflict of interest was declared by the authors.

REFERENCES

[1] Gurunlu Alma, O., Bulut, E., "Genetic algorithm based variable selection for partial least squares Regression using ICOMP Criterion" Asian Journal of Mathematics and Statistics, 5(3), 82-92, (2012).

[2] Bulut, E., Egrioglu, E., "A new partial least square method based on elman neural network" American Journal of Intelligent Systems, 4(4), 154-158, (2014).

[3] Hubert, M., Vanden Branden, K., "Robust methods for partial least squares regression", Journal of Chemometrics, 17: 537-549, (2003).

[4] Liebmann, B., Filzmoser, P., Varmuza, K., "Robust and classical PLS regression compared", Journal of Chemometrics, 24(3-4): 111-120, (2010).

[5] González, J., Peña, D., Romera, R, "A robust partial least squares regression method with applications", Journal of Chemometrics, 23, 78–90, (2009).

[6] Wakeling, I.N., Macfie, H.J.H., " A robust PLS procedure", Journal of Chemometrics, 6: 189–198, (1992).

[7] Griep, M.I., Wakeling, I.N., Vankeerberghen, P., Massart, D.L., "Comparison of semirobust and robust partial least squares procedures", Chemometrics and Intelligent Laboratory Systems, 29: 37-50, (1995).

[8] Gil, J.A., Romera, R., "On robust partial least squares (PLS) methods", Journal of Chemometrics, 12: 365-378, (1998).

[9] Engelen, S., Hubert, M., Vanden Branden, K., Verboven, S., "Robust PCR and robust PLSR: a comparative study", In: Hubert, M., Pison, G., Struyf, A., Aelst, S.V. (Eds.), Theory and Applications of Recent Robust Methods, Birkhäuser, Basel, 105–117, (2004).

[10] Serneels, S., Croux, C., Filzmoser, P., Van Espen, P.J., "Partial robust M-regression", Chemometrics and Intelligent Laboratory Systems, 79: 55-64, (2005).

[11] Herwindiati, D.E., "A new criterion in robust estimation for location and covariance matrix, and its application for outlier labeling". Phd. Thesis, Institut Teknologi Bandung, Bandung, (2006).

[12] Djauhari, M.A., Mashuri, M., Herwindiati, D.E., "Multivariate process variability monitoring", Communications in Statistics - Theory and Methods, 37(11): 1742-1754, (2008).

[13] Herwindiati, D.E., Djauhari, M.A., Mashuri, M., "Robust multivariate outlier labeling", Communication in Statistics–Computation and Simulation, 36: 1287-1294, (2007).

[14] Ali, H., Syed-Yahaya, S.S., "On robust mahalonobis distance issued from minimum vector variance", Far East Journal of Mathematical Sciences (FJMS), 74(2): 249-268, (2013).

[15] Yahaya, S.S., Ali, H., Omar, Z., "An alternative hotelling T^2 control chart based on minimum vector variance (MVV)", Modern Applied Science, 5(4): 132-151, (2011).

[16] Rousseeuw, P.J., Van Zomeren, B.C., "Unmasking multivariate outliers and leverage points", Journal of the American Statistical Association, 85: 633–639, (1990).

[17] Gervini, D., "A robust and efficient adaptive reweighted estimator of multivariate location and scatter", Journal of Multivariate Analysis, 84: 116–144, (2003).

[18] Gervini, D., Yohai, V.J., "A class of robust and fully efficient regression estimators", The Annals of Statistics, 30 (2): 583–616, (2002).

[19] Rousseeuw, P.J., Van Driessen, K., "A fast algorithm for the minimum covariance determinant estimator", Technometrics, 41: 212–224, (1999).

[20] Naes, T., "Multivariate calibration when the error covariance matrix is structured", Technometrics, 27(3): 301-311, (1985).

[21] Hardy, A.J., MacLaurin, P., Haswell, S.J., De Jong, S., Vandeginste, B.G.M., "Double-case diagnostic for outliers identification", Chemometrics and Intelligent Laboratory Systems, 34: 117-129, (1996).

[22] Polat, E., Gunay, S., "A new robust partial least squares regression method based on multivariate MM-estimators", International Journal of Mathematics and Statistics, 18(3): 82-99, (2017).