Journal of AL-Rafidain University College for Sciences (2023); Issue 54; 557 - 566

AL- Rafidain University College

PISSN: (1681-6870); EISSN: (2790-2293)

Journal of AL-Rafidain University College for Sciences

Available online at: https://www.jrucs.iq

JRUCS

Journal of AL-Rafidain University College for Sciences

Robust Sparse Sufficient Dimension Reduction via Adaptive Lasso Penalty

Prof Dr.Ali Alkenani	Prof Dr. Tahir R. Dikheel
ali.alkenani@qu.edu.iq	tahir_reisan@yahoo.com

Department of Statistics - College of Administration and Economics - University of Al-Qadisiyah, Al- Qadisiyah, Iraq.

Article Information

Article History:
Received: January, 6, 2023
Accepted: March, 3, 2023
Available Online: December, 31, 2023

Keywords:
Sufficient dimension reduction, Multi-index models, MAVE, Adaptive Lasso, Robust variables selection.

Correspondence:
Prof Dr.Ali Alkenani
ali.alkenani@qu.edu.iq

Abstract

In some multi-index models applications, there is an important role for dimension reduction and variable selection (VS) methods. The ALMAVE is a method for variables selection under sufficient dimension reduction theory settings. It combines the adaptive Lasso and MAVE (minimum average variance estimation) to produce sparse and accurate solutions. The ALMAVE is a very sensitive method to outliers in the response y due to use the least-squares criterion. In this article, we proposed robust ALMAVE. Also, an efficient estimation algorithm was proposed. The simulation studies and Logo design data analysis were employed to check the effectiveness of ALMAVE.

https://doi.org/10.55562/jrucs.v54i1.624

1. Introduction

In the last thirty years, the huge data have caused significant problems for the statistical methods. In multiple regression applications, only a number of covariates is thought to be truly related to the response. Thus, there is importance to do VS. In analysing the high dimensional data, the VS helps to achieve to goals. It plays important role to achieve better model interpretation and higher prediction precision. In the literature, a number of traditional VS methods has been proposed, such as AIC and BIC, etc. From another side, regularisation methods, such as Lasso (Tibshirani, 1996), SCAD (Fan and Li, 2001) and Elastic Net (Zou and Hastie, 2005) were proposed for simultaneous parameter estimation and VS. However, when the number of covariates is large it is not ease to formulate a parametric model. So, there is need to find model-free VS approaches.

A model-free alternative to VS was provided through introducing the idea of sufficient dimension reduction (SDR) by Cook (1998). The SDR focuses on replacing the original predictor vector with low dimensional projection without losing any information about the regression. The methods of SDR suffer from that the resulting directions are linear combinations of original predictors. Number of testing procedures was proposed by Cook (2004) and Li et al. (2005) to assess the effect of each covariate. Because of their inherent discreteness, these model free VS methods are not stable as is the case in the classical VS methods (Brieman, 1996). Ni et al. (2005), Li and Nachtsheim (2006), Li (2007), Bondell and Li (2008) and Li and Yin (2008) incorporated the regularisation methods into dimension reduction methods. Also, Wang and Yin (2008) combined Lasso with MAVE (Xia et al. 2002) to propose sparse MAVE (SMAVE). Alkenani and Yu (2013) incorporated adaptive Lasso (Zou, 2006), SCAD (Fan and Li, 2001) and MCP (Zhang, 2010) into MAVE to produce variables selection under SDR settings. Rahman and Alkenani (2020) and Alkenani and Rahman (2021) proposed sparse MAVE with adaptive elastic net and elastic net penalties, respectively.

Sparse MAVE versions, which are proposed in Wang and Yin (2008), Alkenani and Yu (2013), Rahman and Alkenani (2020) and Alkenani and Rahman (2021), need no strong assumptions about the covariates. However, these methods are very sensitive to outliers in y because of employing least-squares formulation. Cızek and Hardle (2006) showed that MAVE is very sensitive to outliers. Also, a robust version of MAVE(RMAVE) was proposed by the authors. Yao and Wang (2013) propose a robust sparse MAVE (RSMAVE) depending on the robust MAVE which is proposed in Cızek and Hardle (2006). Yao and Wang (2013) combined Lasso shrinkage with RMAVE to produce robust sparse dimension reduction. RSMAVE (Yao and Wang, 2013) inherits the advantages and disadvantages of their components.

Fan and Li (2001) showed that the Lasso produces biased estimates for the large coefficients. The authors explained that the Lasso does not have the oracle property. The adaptive Lasso was proposed by Zou (2006). The adaptive Lasso allows to penalise the different coefficients by using adaptive weights. The adaptive Lasso estimates are consistent and have the oracle property(Zou,2006).

The limitations of ALMAVE (Alkenani and Yu, 2013) and RSMAVE (Yao and Wang, 2013) motivate us to propose robust sparse dimension reduction method, which is called (RALMAVE). The RALMAVE has the robustness of RMAVE to the outliers in y and the ability of adaptive Lasso in oracle VS and consistent parameters estimation. The effectiveness of RALMAVE is assessed via analysis simulation examples and a real data.

The rest of the article is organised as follows. In Section 2, MAVE and ALMAVE were reviewed. RALMAVE is proposed in Section 3. The results of the simulation examples are reported in Section 4. In Section 5, RALMAVE was applied to a logo design data. The conclusions are summarized in Section 6.

2. A Summary of MAVE and ALMAVE

Suppose that

$$y = f(\mathrm{x}_1, \mathrm{x}_2, \dots, \mathrm{x}_p) + \varepsilon, \qquad (1)$$

where y, $\mathbf{x}$ and ε are the response, a $p \times 1$ predictor vector and the error variables, respectively. Assume $E(y|\mathbf{x}) = f(\mathrm{x}_1, \mathrm{x}_2, \dots, \mathrm{x}_p)$ and $E(\varepsilon\,|\mathbf{x}) = 0$. The $f(.)$ is an unknown smooth link function. For mean function, SDR investigates a subspace S such that

$$y \perp\!\!\!\perp E(y|\mathbf{x})|P_s\mathbf{x}, \qquad (2)$$

where $P_{(.)}$ is a projection operator. If d is the dimension of S and $\mathbf{B} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \dots, \boldsymbol{\beta}_d)$ is a basis for S, $\mathbf{x}$ can be replaced with $\mathbf{B}^{\boldsymbol{T}}\mathbf{x}$, $where\ d \leq p$. The central mean subspace $S_{E(y|\mathbf{x})}$ is the intersection of all subspaces satisfying (2) (Cook and Li, 2002). The MAVE was proposed to estimate $S_{E(y|\mathbf{x})}$. MAVE can estimate the effective dimension reduction (EDR) directions through

$$\min_{\mathbf{B}:\,\mathbf{B}^T\mathbf{B} = \mathbf{I}_d} \left(\sum_{j=1}^{n} \sum_{i=1}^{n} \left[y_i - \{a_j + \boldsymbol{b}_j{}^{\boldsymbol{T}}\mathbf{B}^{\boldsymbol{T}}(\mathbf{x}_i - \mathbf{x}_j)\} \right]^2 \omega_{ij} \right) \qquad (3)$$

where $\omega_{ij} \geq 0$ are the kernel weights. Iteratively, we can solve the minimization of (3) with respect to $\{(a_j, \boldsymbol{b}_j), j = 1, .., n\}$ and $\mathbf{B}$ separately. To improve the accuracy, a kernel weight $\tilde{\omega}_{ij}$ which is a function of $\tilde{\mathbf{B}}^{\boldsymbol{T}}(\mathbf{x}_i - \mathbf{x}_j)$ can be used.

Note that each estimated EDR directions is a linear combination of all original predictors. As result, the interpretation of resulting estimates is not ease. The interpretability of the model and the prediction precision can be improved through the selection of the important covariates. Alkenani and Yu (2013) proposed ALMAVE by combining the adaptive Lasso penalty with the least square formulation of MAVE in (3). The estimates of ALMAVE can be obtained by solving

$$\min_{\mathbf{B}} \left(\sum_{j=1}^{n} \sum_{i=1}^{n} \left[y_i - \{a_j + \boldsymbol{b}_j{}^{\boldsymbol{T}} \mathbf{B}^{\boldsymbol{T}} (\mathbf{x}_i - \mathbf{x}_j)\} \right]^2 \omega_{ij} + \lambda_k \sum_{k=1}^{d} \omega^*{}_k |\boldsymbol{\beta}_k| \right) \tag{4}$$

where $\{\lambda_k > 0, k = 1, \dots, d\}$ are the tuning parameters. The weights $\omega^*{}_k = 1/|\tilde{\beta}_k|^{\delta}$, $\tilde{\beta}$ is a MAVE estimate and $\delta > 0$. where $|.|$ represents the absolute value. We can solve the minimization of (4) by a standard adaptive Lasso algorithm. For details, see Alkenani and Yu (2013).

3. Robust SMAVE

3.1.Robust estimation

In the minimization problems (3) and (4), the least-squares criterion is used between y and $f(\mathrm{x}_1, \mathrm{x}_2, \dots, \mathrm{x}_p)$ to assess how well the model fits. The main drawback of the mentioned least-squares that it is not robust and sensitive to outliers in y. Cızek and Hardle (2006) employed the local L- and M- estimation instead of the local least squares to achieve the robustness. The robust MAVE estimates can be obtained by minimizing

$$\min_{\mathbf{B}: \mathbf{B}^T \mathbf{B} = \mathbf{I}_d} \sum_{j=1}^{n} \sum_{i=1}^{n} \rho(y_i - \{a_j + \boldsymbol{b}_j{}^{\boldsymbol{T}} \mathbf{B}^{\boldsymbol{T}} (\mathbf{x}_i - \mathbf{x}_j)\}) \, \omega_{ij}, \tag{5}$$

where $\rho(.)$ is a robust loss function. Let $\psi(.) = \acute{\rho}(.)$, where $\acute{\rho}(.)$ is the derivative of $\rho(.)$. The Huber's function (Huber, 1981) is a widely used, where $\psi(x) = max[-c, min(c, x)]$ and c controls the robustness amount. In practice, $c = 1.345\sigma$ is recommended by Huber (1981), where σ is the standard deviation of ε. Wilcox (1994) pointed out that the Huber's function is a monotonic and it gives a consistent estimator of location.

It is well known that M-estimators are sensitive to the high leverage outliers. However, the chance of appearing the high leverage outliers in a local window in the local linear approximation of MAVE is less likely. (Yao and Wang, 2013).

3.2. Robust ALMAVE (RALMAVE)

For robust VS, adaptive Lasso penalty can be incorporate into (5),

$$\min_{\mathbf{B}: \mathbf{B}^T \mathbf{B} = \mathbf{I}_d} \sum_{j=1}^{n} \sum_{i=1}^{n} \rho(y_i - \{a_j + \boldsymbol{b}_j{}^{\boldsymbol{T}} \mathbf{B}^{\boldsymbol{T}} (\mathbf{x}_i - \mathbf{x}_j)\}) \, \omega_{ij} + \lambda_k \sum_{k=1}^{d} \omega^*{}_k |\boldsymbol{\beta}_k|, \tag{6}$$

Noting that $\acute{\rho}(t) = t\acute{\rho}(t)/t$, the minimization of (6) can be done through (4) with the following updated weight

$$\omega^*_{ij} = \omega_{ij} \boldsymbol{W}(\hat{\varepsilon}_{ij}), \tag{7}$$

Where:

$$\boldsymbol{W}(\hat{\varepsilon}_{ij}) = \frac{\psi(\hat{\varepsilon}_{ij})}{\hat{\varepsilon}_{ij}}$$

$$\hat{\varepsilon}_{ij} = y_i - \{\hat{a}_j + \hat{\boldsymbol{b}}_j^T \hat{\mathbf{B}}^T (\mathbf{x}_i - \mathbf{x}_\mathrm{j})\}$$

$$\omega_{ij} = \frac{K_h\{\hat{\mathbf{B}}^T (\mathbf{x}_i - \mathbf{x}_\mathrm{j})\}}{\sum_{i=1}^{n} K_h\{\hat{\mathbf{B}}^T (\mathbf{x}_i - \mathbf{x}_\mathrm{j})\}},$$

and $K_h(v) = h^{-1}K(v/h)$, where $K(v)$ is a kernel function and h is the bandwidth. $\{\widehat{\mathbf{B}}, (\hat{a}_j, \widehat{\boldsymbol{b}}_j), j = 1, \ldots, n\}$ are initial estimators. With the weight ω^*_{ij}, the $\psi(.)$ controls the robustness. Also, as in Cızek and Hardle (2006), the algorithm of ALMAVE can be employed here to minimise (6) by replacing ω_{ij} in (4) with ω^*_{ij} in (7).

The following algorithm was proposed to minimize (6)

Algorithm 3.1. For $\{(y_i, \mathbf{x}_i), i = 1, .., n\}$,

1. Obtain $\{\widehat{\mathbf{B}}, (\hat{a}_j, \widehat{\boldsymbol{b}}_j), j = 1, \ldots, n\}$ in (5).
2. Calculate ω^*_{ij} in (7);
3. Replace ω_{ij} by ω^*_{ij} in (4), and update the estimator with ALMAVE algorithm as follows
   - **I.** For given $\widehat{\mathbf{B}}$, update $(a_j, \boldsymbol{b}_j)$ where $j = 1, \ldots, n$, from

$$\min_{\mathbf{B}:\mathbf{B}^T\mathbf{B}=\mathbf{I}_d} \left(\sum_{j=1}^{n}\sum_{i=1}^{n} \left[y_i - \{a_j + \boldsymbol{b}_j{}^{\boldsymbol{T}} \mathbf{B}^{\boldsymbol{T}} (\mathbf{x}_i - \mathbf{x}_j)\} \right]^2 \omega^*_{ij} \right) \tag{8}$$

   - **II.** For a given $(\hat{a}_j, \hat{b}_j)$, j = 1, … , n, solve B

$$\min_{\mathbf{B}} \left(\sum_{j=1}^{n}\sum_{i=1}^{n} \left[y_i - \{a_j + \boldsymbol{b}_j{}^{\boldsymbol{T}} \mathbf{B}^{\boldsymbol{T}} (\mathbf{x}_i - \mathbf{x}_j)\} \right]^2 \omega^*_{ij} + \lambda_k \sum_{k=1}^{d} \omega^*{}_k |\boldsymbol{\beta}_k| \right) \tag{9}$$

   - **III.** Iterate between (I) and (II) until convergence in B estimator.

4. Iterate between 2 and 3 until convergence.

According to our extensive simulations examples, the above Algorithm often converges within 5 to 10 iterations.

3.3. The selection of c value

The parameter c involves σ. This σ is unknown and we need to estimate it. A robust version of σ is the median absolute deviation (MAD) as

$$\widehat{\sigma} = \frac{Median(|\hat{\varepsilon}_i - Median(\hat{\varepsilon}_i)|)}{0.675} \tag{10}$$

The amounts 1.345 in c for Huber function can be modified. The more suitable value of c is the value which makes balance between the robustness to outliers and the efficiency of estimation (Yao and Wang, 2013).

3.4. Determination of d

In SDR, the issue of estimation d is a very crucial. In this article, a robust version of cross-validation (RCV) was employed to estimate d. The RCV based on Hampel's piecewise linear function (Hample et al., 1986) was used, where the Hampel's piecewise linear function is

$$\rho(t) = \begin{Bmatrix} t^2/2 & |t| \le a \\ a|t| - a^2/2 & a < |t| \le b \\ \dfrac{a(c|t| - t^2/2)}{c-b} - \dfrac{7a^2}{6} & b < |t| \le c \end{Bmatrix} \tag{11}$$

For a given dimension k, the CV value can be calculated depending on the estimated $\widehat{\mathbf{B}}$ as

$$CV_k = n^{-1} \sum_{i=1}^{n} \rho \left(y_i - \frac{\sum_{j\neq i} y_j K_h\{\widehat{\mathbf{B}}^T(\mathbf{x}_l - \mathbf{x}_i)\}}{\sum_{l\neq i} K_h\{\widehat{\mathbf{B}}^T(\mathbf{x}_j - \mathbf{x}_i)\}} \right) \tag{12}$$

After that, the estimated d can be obtained as follows

$$\hat{d} = arg \min_{0\le k\le p} CV_k \tag{13}$$

Other robust loss functions such as Huber or Tukey loss functions also can be used. From our simulation studies, Hampel's piecewise linear function showed slightly outperforms the others.

4. Simulation studies

The performance of RALMAVE was compared with the performance of SMAVE, ALMAVE, and RSMAVE through simulation studies. The trace correlation r^* which is used in Zhu and Zeng (2006) was adopted for measuring the estimation accuracy. Let $S(A)$ and $S(B)$ are columns spaces spanned by two $p \times d$ of full rank matrices. $P_A = A(A^TA)^{-1}A^T$ and $P_B = B(B^TB)^{-1}B^T$ are projection matrices on $S(A)$ and $S(B)$, respectively. $r^* = \sqrt{\frac{1}{d}tr(P_AP_B)}$, where, $0 \le r^* \le 1$. The true positive rate (TPR) and the false positive rate (FPR) were employed to measure the ability of the compared methods based on VS. TPR is the ratio of predictors number which is correctly identified as effective to actual effective predictors number. While, FPR is the ratio of predictors number which is falsely identified as effective to ineffective predictors number. The ideal situation is TPR near to to 1 and the FPR near to 0 at the same time.

An efficient adaptive Lasso algorithm was employed to solve the minimization in (9). A residual information criterion (RIC) (Shi and Tsai, 2002) was employed to choose λ for the adaptive Lasso,

$$RIC = \{n - p(\lambda)\}\log(RSS/\{n - p(\lambda)\}) + p(\lambda)\{\log(n) - 1\} + 4/\{n - p(\lambda) - 2\}, \tag{14}$$

where, the RSS is sum of squares of residual in the fit of adaptive Lasso, and $p(\lambda)$ is the non-zero coefficients number. RCV was employed for selection h.

4.1. Direction estimation and VS

The data were generated from:

$$y = \frac{\boldsymbol{\beta}_1^T\mathbf{x}}{0.5 + (1.5 + \boldsymbol{\beta}_2^T\mathbf{x})^2} + \varepsilon, \tag{15}$$

where,
$\beta_1 = (1,0,\dots,0)^T$, $\beta_2 = (0,1,0,\dots,0)^T$, and $\mathbf{x} \in \mathbb{R}^{10}$ with $d = 2$. The settings for $\mathbf{x}$ is as follows:
(a) $\mathbf{x} \sim N_{10}(\mathbf{0}_{10}, \boldsymbol{I}_{10})$ (b) $\mathbf{x} \sim N_{10}(\mathbf{0}_{10}, \boldsymbol{\Sigma})$, where $(i,j)^{th}$ element of $\boldsymbol{\Sigma}$ is $0.5^{|\boldsymbol{i}-\boldsymbol{j}|}$. The studied distributions of ε were as follows:

Dist.1. $N(0,1)$, the standard normal.

Dist.2. $t_3/\sqrt{3}$, t-distribution with 3 degree of freedom.

Dist.3. $0.95\,N(0,1) + 0.05\,N(0,10^2)$.

Dist.4. $0.95\,N(0,1) + 0.05\mathrm{U}(-50,50)$, 95% from standard normal and 5% uniform distribution.

The directions of EDR were obtained through SMAVE, ALMAVE, RSMAVE and RALMAVE methods. 200 datasets were generated for each sample size $n = 100,200, and\ 400$. The comparison among SMAVE, ALMAVE, RSMAVE and RALMAVE methods was carried out in Table 1 and 2. To assess the accuracy of estimation, the mean of r^* ($\mu(r^*)$) and standard error of r^* ($\mathrm{SE}(r^*)$) were summarized. Also, TPR and FPR were used to check the ability of RALMAVE in VS.

Table 1: $\mu(r^*)$, SE(r^*), TPR, and FPR for SMAVE, ALMAVE, RSMAVE and RALMAVE in case of uncorrelated predictors.

Dist.	n	Criterion	SMAVE	ALMAVE	RSMAVE	RALMAVE
Dist.1	100	$\mu(r^*)$ (SE(r^*))	0.876(0.144)	0.881(0.146)	0.850(0.160)	0.855(0.157)
		TPR	0.853	0.860	0.818	0.823
		FPR	0.126	0.123	0.142	0.140
	200	$\mu(r^*)$ (SE(r^*))	0.970(0.078)	0.973(0.080)	0.958(0.094)	0.963(0.090)
		TPR	0.968	0.972	0.958	0.961
		FPR	0.062	0.059	0.083	0.081
	400	$\mu(r^*)$ (SE(r^*))	0.999(0.007)	0.999(0.006)	0.998(0.005)	0.999(0.005)
		TPR	1.000	1.000	1.000	1.000
		FPR	0.037	0.036	0.052	0.051
Dist.2	100	$\mu(r^*)$ (SE(r^*))	0.877(0.151)	0.880(0.150)	0.907(0.129)	0.912(0.125)
		TPR	0.873	0.875	0.897	0.900
		FPR	0.178	0.177	0.159	0.157
	200	$\mu(r^*)$ (SE(r^*))	0.962(0.089)	0.965(0.088)	0.993(0.033)	0.996(0.032)
		TPR	0.960	0.961	0.998	0.1000
		FPR	0.092	0.090	0.093	0.090
	400	$\mu(r^*)$ (SE(r^*))	0.992(0.036)	0.995(0.034)	0.999(0.003)	1.000 (0.002)
		TPR	0.995	0.998	1.000	1.000
		FPR	0.068	0.066	0.083	0.080
Dist.3	100	$\mu(r^*)$ (SE(r^*))	0.659(0.231)	0.660(0.230)	0.862(0.143)	0.865(0.142)
		TPR	0.738	0.740	0.820	0.824
		FPR	0.407	0.406	0.153	0.150
	200	$\mu(r^*)$ (SE(r^*))	0.685(0.208)	0.687(0.207)	0.943(0.112)	0.947(0.110)
		TPR	0.708	0.709	0.933	0.935
		FPR	0.374	0.372	0.088	0.085
	400	$\mu(r^*)$ (SE(r^*))	0.737(0.209)	0.739(0.208)	0.998(0.007)	1.000 (0.004)
		TPR	0.755	0.757	1.000	1.000
		FPR	0.343	0.340	0.061	0.060
Dist.4	100	$\mu(r^*)$ (SE(r^*))	0.451(0.262)	0.454(0.260)	0.837(0.154)	0.841(0.152)
		TPR	0.683	0.686	0.800	0.805
		FPR	0.632	0.630	0.161	0.159
	200	$\mu(r^*)$ (SE(r^*))	0.389(0.278)	0.393(0.276)	0.958(0.097)	0.961(0.096)
		TPR	0.593	0.596	0.953	0.954
		FPR	0.540	0.538	0.075	0.074
	400	$\mu(r^*)$ (SE(r^*))	0.439(0.279)	0.440(0.279)	0.997(0.021)	0.999(0.017)
		TPR	0.573	0.575	0.998	0.999
		FPR	0.518	0.517	0.066	0.064

Table 2: $\mu(r^*)$, SE(r^*), TPR, and FPR for SMAVE, ALMAVE, RSMAVE and RALMAVE in case of correlated predictors.

Dist.	n	Criterion	SMAVE	ALMAVE	RSMAVE	RALMAVE
Dist.1	100	$\mu(r^*)$ (SE(r^*))	0.805(0.156)	0.812(0.150)	0.802(0.146)	0.811(0.140)
		TPR	0.797	0.803	0.792	0.797
		FPR	0.164	0.159	0.198	0.194
	200	$\mu(r^*)$ (SE(r^*))	0.917(0.123)	0.923(0.120)	0.887(0.138)	0.897(0.132)
		TPR	0.945	0.949	0.902	0.906
		FPR	0.101	0.098	0.126	0.123
	400	$\mu(r^*)$ (SE(r^*))	0.975(0.074)	0.980(0.071)	0.969(0.083)	0.977(0.078)
		TPR	0.998	0.999	0.993	0.997
		FPR	0.079	0.077	0.102	0.097
Dist.2	100	$\mu(r^*)$ (SE(r^*))	0.828(0.148)	0.835(0.143)	0.847(0.145)	0.854(0.140)
		TPR	0.863	0.868	0.900	0.904
		FPR	0.209	0.206	0.239	0.234
	200	$\mu(r^*)$ (SE(r^*))	0.906(0.126)	0.912(0.121)	0.949(0.100)	0.955(0.095)
		TPR	0.933	0.937	0.985	0.990
		FPR	0.139	0.135	0.153	0.150

	400	$\mu(r^*)$ (SE(r^*))	0.968(0.082)	0.973(0.078)	0.995(0.030)	0.999(0.027)
		TPR	0.990	0.993	1.000	1.000
		FPR	0.109	0.104	0.139	0.144
Dist.3	100	$\mu(r^*)$ (SE(r^*))	0.646(0.249)	0.650(0.246)	0.780(0.141)	0.785(0.138)
		TPR	0.705	0.709	0.767	0.771
		FPR	0.366	0.364	0.203	0.200
	200	$\mu(r^*)$ (SE(r^*))	0.710(0.169)	0.714(0.166)	0.879(0.135)	0.885(0.131)
		TPR	0.708	0.713	0.907	0.910
		FPR	0.326	0.323	0.153	0.151
	400	$\mu(r^*)$ (SE(r^*))	0.756(0.183)	0.760(0.186)	0.957(0.097)	0.963(0.095)
		TPR	0.797	0.800	0.993	0.996
		FPR	0.306	0.304	0.122	0.121
Dist.4	100	$\mu(r^*)$ (SE(r^*))	0.453(0.273)	0.458(0.270)	0.788(0.147)	0.793(0.143)
		TPR	0.693	0.697	0.787	0.790
		FPR	0.623	0.620	0.214	0.210
	200	$\mu(r^*)$ (SE(r^*))	0.418(0.292)	0.423(0.289)	0.887(0.131)	0.893(0.127)
		TPR	0.570	0.574	0.927	0.931
		FPR	0.509	0.506	0.149	0.146
	400	$\mu(r^*)$ (SE(r^*))	0.461(0.281)	0.465(0.278)	0.967(0.082)	0.972(0.078)
		TPR	0.630	0.634	0.995	0.998
		FPR	0.489	0.487	0.127	0.123

From the results in Table 1 and 2, the following observations were noticed.

1. For the errors which are follow Dist.1, the performance of RALMAVE is similar to the performance of ALMAVE and the performance of RSMAVE is similar to SMAVE. Also, the performance ALMAVE and RALMAVE is better than the performance of SMAVE and RSMAVE, respectively, based on estimation accuracy and VS.
2. The ALMAVE and SMAVE were showed some robustness when the errors were followed Dist.2. But their performance was negatively affected according to estimation accuracy and VS when the errors follow Dist.3 or Dist.4
3. For the errors which are follow Dist.2, Dist.3 and Dist.4, the RALMAVE and RSMAVE performed almost well as they did in the case of Dist.1. according to estimation accuracy and VS, RALMAVE outperformed the RSMAVE. In addition, RALMAVE also exceeded ALMAVE, especially when the errors followed Dist.3 and Dist.4. In summary, the proposed RALMAVE method gave very consistent estimates and it showed good performance in terms of estimation accuracy and VS for all error distributions considered. Also, the performance of RALMAVE was the best among all compared methods.

4.2. Estimation of d

The ability of robust CV in (13) was checked for the estimation of d in this section. We generated the data as in model (15) settings. The value of d was 2. The results in case of the independent predictors with $n = 100 \; and \; 200$ were reported. For each sample size, 200 datasets were generated. Table 3 reports the frequency of $\hat{d}$ out of 200 datasets. The results of L1-based CV (Cızek and Hardle, 2006) were also reported for the sake of comparison. It is clear that the robust CV based on Hample loss function gave very consistent estimation for all settings. It did well under Dist2, Dist3 and Dist4 settings, although a bit worse than those under Dist1. The performance of RCV based on Hample loss function a bit exceeds the performance of L1-based CV for Dist3 and Dist4.

Table 3: Frequency of $\hat{d}$ out of 200 datasets

Dist.	n	CV_{Hample}					CV_{L1}				
		d=1	d=2	d=3	d=4	d>=5	d=1	d=2	d=3	d=4	d>=5
Dist.1	100	10	156	33	1	0	7	153	40	0	0
	200	1	180	19	0	0	4	180	15	1	0
Dist.2	100	13	147	37	3	0	9	141	47	3	0
	200	1	174	23	2	0	5	176	19	0	0
Dist3.	100	30	100	46	16	8	46	91	43	17	3
	200	3	133	51	12	1	16	120	51	10	3

Dist4.	100	42	98	32	10	18	49	90	46	10	5
	200	17	132	27	9	15	21	122	32	19	6

5. Logo design (LD) data

The LD data were gathered by Henderson and Cote (1998) to know how LD may impact consumers' response to logos. The data contain 22 predictors with $n = 195$ observations. The response variable y refers to the logo effect.

1. ALMAVE (Alkenani and Yu, 2013) identified $d = 1$ direction with 8 important predictors.

To verify RALMAVE, LD were re-analysed by incorporate some outliers in y. 5% of contaminated observations and an outlier were inserted in the data. The value y_i is increased to $y_i + c$ and we report the results for $c = 10$ and 20.

W reported the number of selected variables (NSV) by ALMAVE and RALMAVE in Table 4. In addition, we reported $corr(\hat{\beta}, \hat{\beta}_{AL0})$ which is the correlation between $\hat{\beta}$ and $\hat{\beta}_{AL0}$ from ALMAVE without outliers to evaluate the estimation accuracy of RALMAVE.

Table 4: The NSV and $corr(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\beta}}_{AL0})$ for ALMAVE and RALMAVE methods.

Outliers	NSV		$corr(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\beta}}_{AL0})$	
	ALMAVE	RALMAVE	ALMAVE	RALMAVE
No outlier	8	8	1	0.9920
Single outlier ($c = 10$)	9	7	0.9242	0.9911
Single outlier ($c = 20$)	8	7	0.8432	0.9902
5% outliers ($c = 10$)	16	7	0.3638	0.9910
5% outliers ($c = 20$)	18	7	0.0653	0.9899

It can be seen that the performance of ALMAVE is very similar to the performance of RALMAVE for the data without outliers. In case of the data were contaminated with the outliers, the performance of ALMAVE is dramatically affected. But very consistent results were produced by RALMAVE, even with 5% outliers.

6. Conclusion

In this article, the RALMAVE method was proposed under SDR settings. The RALMAVE benefits from the merits of robust VS under SDR settings. The simulation studies indicate that RALMAVE was better than the SMAVE, ALMAVE, and RSMAVE under different settings. Also, RCV criterion based on Hample loss function was very efficient in estimating d. The idea of RALMAVE can be expanded to models with discrete response. For examples, logistic regression and Poisson regression.

References

[1] Alkenani, A. and Yu, K. (2013). "Sparse MAVE with oracle penalties". Advances and Applications in Statistics 34, 85–105.

[2] Alkenani, A. and Rahman, E. (2021). "Regularized MAVE through the elastic net with correlated predictors". Journal of Physics: Conference Series 1897 (1), 012018.

[3] Bellman, R. E. (1961). Adaptive Control Processes. Princeton University Press, Princeton, NJ.

[4] Bondell, H.D. and Li, L. (2008). "Shrinkage inverse regression estimation for model free variable selection". Journal of Royal Statistical Society B, 71, 287-299.

[5] Breiman, L. (1996). "Heuristics of instability and stabilization in model selection". Annals of Statistics, 24, 2350–2383.

[6] Cızek, P. and Hardle, W. (2006). "Robust estimation of dimension reduction space". Computational Statistics and Data Analysis, 51, 545-555.

[7] Cook, R. D. (1998). Regression Graphics: Ideas for studying regressions through graphics. Wiley, New York.

[8] Cook, R. D. (2004). "Testing predictor contributions in sufficient dimension reduction". Annals of Statistics, 32, 1062–1092.

[9] Cook, R.D., Li, B., 2002. “Dimension reduction for the conditional mean in regression”. The Annals of Statistics 30, 455–474.

[10] Fan, J. and Li, R. (2001). “Variable selection via nonconcave penalized likelihood and its oracle properties”. Journal of the American Statistical Association, 96, 1348–1360.

[11] Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J. and Stahel,W. A. (1986). Robust Statistics: The Approach Based on Influence Functions, Wiley, New York.

[12] Henderson, P. W. and Cote, J. A. (1998). “Guidelines for selecting or modifying logos”. Journal of Marketing, 62(April), 14–30.

[13] Huber, P. J. (1981). Robust Statistics. Wiley, New York.

[14] Li, L. (2007). “Sparse sufficient dimension reduction”. Biometrika, 94, 603–613.

[15] Li, L., Cook, R. D. and Nachtsheim, C. J. (2005). “Model-free variable selection”. Journal of Royal Statistical Society B, 67, 285–299.

[16] Li, L. and Nachtsheim, C. J. (2006). “Sparse sliced inverse regression”, Technometrics, 48, 503–510.

[17] Li, L. and Yin, X. (2008). “Sliced Inverse Regression with regularizations”. Biometrics 64, 124–131.

[18] Ni, L., Cook, R. D. and Tsai, C. L. (2005). “A note on shrinkage sliced inverse regression”. Biometrika, 92, 242–247.

[19] Rahman, E. and Alkenani, A. and (2020). “Sparse minimum average variance estimation via the adaptive elastic net when the predictors correlated”. Journal of Physics: Conference Series 1591 (1), 012041.

[20] Shi, P. and Tsai, C. L. (2002). “Regression model selection - a residual likelihood approach”. J. R. Statist. Soc. B 64, 237–52.

[21] Tibshirani, R. (1996). “Regression shrinkage and selection via the lasso”. Journal of Royal Statistical Society B, 58, 267–288.

[22] Wang, Q. and Yin, X. (2008). “A nonlinear multi-dimensional variable selection method for high dimensional data: Sparse MAVE”. Computational Statistics and Data Analysis, 52, 4512-4520.

[23] Wilcox, R. R. (1994). “The percentage bend correlation coefficient”. Psychometrika 59, 601–616.

[24] Xia, Y., Tong, H., Li, W. and Zhu, L. (2002). “An adaptive estimation of dimension reduction space”. Journal of Royal Statistical Society B, 64, 363-410.

[25] Yao, W. and Wang, Q. (2013). “Robust variable selection through MAVE”. Computational Statistics and Data Analysis, 63, 42-49.

[26] Zhang, C. H. (2010). “Nearly unbiased variable selection under minimax concave penalty”. Annals of Statistics 38, 894–942.

[27] Zhu, Y. and Zeng, P. (2006). “Fourier methods for estimating the central subspace and the central mean subspace in regression”. Journal of the American Statistical Association, 101, 1638–1651.

[28] Zou, H. (2006). “The adaptive Lasso and its oracle properties”. Journal of the American Statistical Association 101, 1418–142.

[29] Zou, H. and Hastie, T. (2005). “Regularization and variable selection via the elastic net”. Journal of Royal Statistical Society B, 67, 301–320.

مجلة كلية الرافدين الجامعة للعلوم (2023)؛ العدد 54؛ 557 - 566

AL- Rafidain University College

PISSN: (1681-6870); EISSN: (2790-2293)
مجلة كلية الرافدين الجامعة للعلوم

Available online at: https://www.jrucs.iq

JRUCS
Journal of AL-Rafidain University College for Sciences

تقليل قوي ومتفرق للأبعاد عن طريق عقوبة اللاسو التكيفية

أ.د. طاهر ريسان دخيل	أ.د. علي الكناني
tahir_reisan@yahoo.com	ali.alkenani@qu.edu.iq
قسم الاحصاء - كلية الادارة والاقتصاد - جامعة القادسية، القادسية، العراق	

المستخلص

في بعض تطبيقات النماذج متعددة الفهارس، هناك دور مهم لطرق تقليل الأبعاد واختيار المتغير (VS). ALMAVE هي طريقة لاختيار المتغيرات في ظل إعدادات نظرية تقليل الأبعاد الكافية. فهو يجمع بين Lasso التكيفي وMAVE (الحد الأدنى لتقدير متوسط التباين) لإنتاج حلول متفرقة ودقيقة. تعد ALMAVE طريقة حساسة جدًا للقيم المتطرفة في الاستجابة y نظرًا لاستخدام معيار المربعات الصغرى. في هذه المقالة، اقترحنا ALMAVE القوي. كما تم اقتراح خوارزمية تقدير فعالة. تم استخدام دراسات المحاكاة وتحليل بيانات تصميم الشعار للتحقق من فعالية ALMAVE.

معلومات البحث

تواريخ البحث:
تاريخ تقديم البحث: 2023/1/6
تاريخ قبول البحث: 2023/3/3
تاريخ رفع البحث على الموقع: 2023/12/31

الكلمات المفتاحية:
تقليل الأبعاد بشكل كافٍ، نماذج متعددة الفهارس، MAVE، Adaptive Lasso، اختيار قوي للمتغيرات.

للمراسلة:
أ.د. علي الكناني
ali.alkenani@qu.edu.iq

https://doi.org/10.55562/jrucs.v54i1.623