Sparse dimension reduction via penalized quantile MAVE

Doaa Tahir Malik[1] and Ali Alkenani[2]

[1]M.sc. student in Statistics, Department of Statistics, College of Administration and Economics, University of Al-Qadisiyah, Iraq.

[2]Assistant. Prof Dr. in Statistics, Department of Statistics, College of Administration and Economics, University of Al-Qadisiyah, Iraq.

Correspondence: Ali Alkenani,

E-mail: ali.alkenani@ qu.edu.iq

http://orcid.org/0000-0001-5067-2321

Abstract

In this paper, the quantile minimum average variance estimator method (QMAVE) and the sparse quantile minimum average variance estimator with lasso penalty (LQMAVE) were proposed. In addition, this paper introduced an inclusive study of QMAVE and LQMAVE. Efficient algorithms proposed to solve QMAVE and LQMAVE minimization problems. The real data analysis and simulations were used to examine the performance of QMAVE and LQMAVE, respectively. From the numerical results, it is clear that the QMAVE and LQMAVE are useful methods in practice. Where its achieved the best and finest results of other methods.

Keywords:Dimension regression, Quantile regression, MAVE, Quantile MAVE, Lasso.

1.Introduction

In some applications of multiple regression, the number of the predictors P in the data matrix X became large and therefore the analysis of this data becomes difficult. In order to cope with this problem, we need to reduce the dimension of X while preserving full regression information and imposing few assumptions. Sufficient dimension reduction (SDR) theory (Cook, 1998) was developed to achieve this aim. Many methods were suggested to estimate the SDR space. Some of them focuses on finding the central subspace $S_{Y/X}$. Examples for these methods are graphical regression (Cook, 1994) and sliced inverse regression (SIR) (Li, 1991) among others.

For regression problems and when the mean function is of interest, Cook and Li (2002) introduced the concept of the Central Mean Subspace (CMS) for reducing the dimension. Many dimension reduction methods were suggested under this concept, for examples, Principal Hessian Direction (PHD) (Li, 1992) and the minimum average variance estimation (MAVE) (Xia et al., 2002) among others. MAVE is powerful dimension reduction method and it is effective in dealing with high-dimensional data. The MAVE was proven to be an efficient method to deal with the dimensionality problem in conditional mean regression.

Quantile regression (QR) is a good tool to explain the relationship of the outcome variable Y and the predictors X. It gives a whole picture of the relationships among variables. It used in many areas. For examples, economics, microarrays and other fields. It is robust to non–normal errors and outliers (Yu et al., 2003).

Let y_i be a response variable and X_i a $p \times 1$ vector of predictors for the ith observation, q_τ (X_i) is the inverse cumulative distribution function $(ICDF)$ of Y_i/X_i. Then, $q_\tau(X_i)$ can be modeled as $q_\tau(X_i)$ $=X_i^T\beta_\tau$. β_τ is a vector of ρ unknown parameters and τ is the quantile level. Koenker and Bassett (1978) suggest to obtain β_τ as minimizes of the following:

$$\min_{\beta_\tau} \quad \sum_{i=1}^{n} \rho_\tau \left(y_i - X_i^T \beta_\tau\right), \tag{1}$$

where $\rho_{\tau(.)}$ is the check loss function

$$\rho_\tau(u) = \tau u I_{[0,\infty)}(u) - (1-\tau) u I_{(-\infty,0)}(u) \tag{2}$$

The first contribution in this paper is quantile regression MAVE (QMAVE) has proposed. QMAVE combines the strength of QR with the effective method MAVE under the sufficient dimension reduction framework. The details of QMAVE have reported in Section 3. Later we will also compare the proposed method (QMAVE) with the two methods quantile regression (QR) and sliced inverse quantile regression(SIQR).

QMAVE method gives us a good tool to obtain sufficient dimensions reduction under quantile regression settings, however, this method suffers from that each dimension reduction component is a linear combination (L.C) of the predictors, which may be difficult to explain the resulting estimates.

Variable selection(V.S) is necessary for construct the model of multiple regression. It works on the improving the prediction of the models, providing model with low cost (Guyon and Elisseeff, 2003). V.S methods such as stepwise selection (Efroymson, 1960), Akaike information criteria (AIC) (Akaike, 1973) and Bayesian information criteria (BIC) (Schwarz, 1978) may suffer from instability, (Brieman,1996). In order to deal with the instability, regularization techniques can also implement V.S. Regularization methods can be formed by adding penalty to the loss functions. In regularization methods, the V.S is implemented through the parameter estimation process. Examples of regularization methods are the Least absolute shrinkage and selection operator (Lasso) (Tibshirani, 1996), the adaptive lasso (Zou, 2006) and Smoothly clipped absolute deviation (SCAD) (Fan and Li, 2001), among others.

Under the framework of the SDR, Ni et al. (2005) proposed a penalized SIR. Li and Nachtsheim (2006) suggested another version of the sparse SIR. Wang and Yin (2008) proposed the sparse MAVE (SMAVE) method. Alkenani and Yu (2013) proposed the SMAVE with adaptive lasso, SCAD and MCP penalties. Alkenani and Reisan (2016) proposed the sparse sliced inverse quantile regression.

The second contribution is sparse QMAVE with Lasso penalty which is proposed in order to solve the problem of that each dimension reduction component was produced through QMAVE is a L.C of all the predictors. Later we will also compare the proposed method (LQMAVE) with the two methods sparse quantile regression (LQR) and sparse sliced inverse quantile regression(LSIQR).

The rest of the paper is organized as follows: In Section 2, a short review of MAVE is given. QMAVE and Sparse QMAVE (LQMAVE) are proposed in Section3 and 4, respectively. Numerical experiments and real data were reported in Sections 5 and 6 respectively. The conclusions were reported in sections7.

2. Short Review of MAVE

Xia et al., (2002) proposed MAVE which was employed to estimate the CMS. MAVE such that B is the solution of:

$$\min_{B} \{ E[y - E(y \mid X^T B)]^2 \},$$

where $B^T B = I_d$.

The variance given $X^T B$ is

$$\sigma^2_B (X^T B) = E[\{y - E(y \mid X^T B)\}^2 \mid X^T B].$$

Thu

$$\min_{B} E[y-E(y\,|\,X^T B)]^2 = \min_{B} E\{\sigma^2_B(X^T B)\}.$$

For given X_0, $\sigma_B^2(X^T B)$ can be approximated as follows:

σ_B^2 $(X_0 B) \approx \sum_{i=1}^{n}\{y_i - E(y_i|X_i^T B)\}^2 \omega_{i0}$

$\approx \quad \sum_{i=1}^{n}[y_i - \{a_0 + (X_i - X_0)^T B b_0\}]^2 \; w_{i0}$, where $a_0 + (X_i - X_0)^T B b_0$ is the local linear expansion of $E(y_i\,|\,X_0^T B)$ at X_0, and $\omega_{i0} \geq 0$ are the kernel weights centered at $X_0^T B$ with $\sum_{i=1}^{n} w_{i0} = 1$

B can be obtained from solving (3):

$$\min_{\substack{B: B^T B = I \\ a_j, b_{j, j=1,\dots,n}}} \left(\sum_{j=1}^{n}\sum_{i=1}^{n}[y_i - \{a_j + (X_i - X_j)^T B b_j\}]^2 w_{ij}\right) \qquad (3) .$$

The algorithm of MAVE was explained as follows:

1. Let $m = 1$ and $B = \beta_0$, any arbitrary $p \times 1$ vector.

2. For known B, solve (a_j, b_j) where $j = 1, \dots, n$, from the minimization below:

$$\min_{a_j b_{j, j=1,\dots,n}} \left(\sum_{j=1}^{n}\sum_{i=1}^{n}[y_i - \{a_j + (X_i - X_j)^T B b_j\}]^2 w_{ij}\right)$$

3. For a given $(\hat{a}_j, \hat{b}_j), j = 1, \dots, n,$ solve β_m from the constrained quadratic minimization below:

$$\min_{B: B^T B = I} \left(\sum_{j=1}^{n}\sum_{i=1}^{n}[y_i - \left(\hat{a}_j \;\; + (X_i - X_j)^T \hat{b}_j^T (\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_{m-1}, \beta_m)\right)]^2 \, w_{ij}\right)$$

4. Now, put $\hat{\beta}_m$ in the mth column of β, and continue step 2 and 3 till convergence is attained.

5. Update B by $(\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_m, \beta_0)$, and let m to be $m + 1$.

6. If $m < d$, go to steps 2 to 5 till $m = d$.

Xia et. al. (2002) suggested to employ the refined multidimensional Gaussian kernel to compute the weights for MAVE as follows:

$W_{ij} = K_h \{\hat{B}^T(X_i - X_j)\} / \sum_{k=1}^{n} \{\hat{B}^T(X_i - X_j)\}$

3. Quantile MAVE (QMAVE)

In this section, we are propose QMAVE method. QMAVE gives us a good tool to obtain sufficient dimension reduction under quantile settings. QMAVE estimates can be obtained according to solve the following algorithm:

1. Let $m = 1$ and B= $\beta_{0,}$ any arbitrary $p \times 1$ vector.

2.For known B, solve (a_j, b_j) where $j = 1, \dots, n$, from the minimization below:

$$\min_{a_j b_{j=1,\dots,n}} \left(\sum_{j=1}^{n} \sum_{i=1}^{n} \rho_\tau [y_i - \{a_j + (X_i - X_j)^T b_j^T B\}] w_{ij} \right)$$

3. For a given$(\hat{a}_j, \hat{b}_j)$, $j = 1, \dots, n$, solve $\beta_{\tau m}$ from the following minimization:

$$\min_{B : B^T B = I} \left(\sum_{j=1}^{n} \sum_{i=1}^{n} \rho_\tau [y_i - \{\hat{a}_j + (X_i - X_j)^T \hat{b}_j^T (\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_{m-1}, \beta_m)\}] w_{ij} \right)$$

4. Now, put $\hat{\beta}_{tm}$ in the mth column of B, and continue step 2 and 3 till convergence is attained.

5. Update B by $(\hat{\beta}_{\tau 1}, \hat{\beta}_{\tau 2}, \dots, \hat{\beta}_{\tau m}, \beta_0)$, and assume m to be $m + 1$.

6. If $m < d$, repeat steps 2 to 5 till $m = d$.

The refined Gaussian kernel which was used for MAVE was employed for computing the weights in our proposed method.

4.Sparse QMAVE with Lasso (LQMAVE)

In this section, we are propose LQMAVE to obtain sparse sufficient dimension reduction under quantile settings. LQMAVE proposed according to the following algorithm:

The algorithm of LQMAVE was described as follows:

1. Let $m = 1$ and B= $\beta_{0,}$ any arbitrary $p \times 1$ vector.

2.For known B, solve (a_j, b_j) where $j = 1, \dots, n$, from the minimization below:

$$\min_{a_j b_{j=1,\dots,n}} \left(\sum_{j=1}^{n} \sum_{i=1}^{n} \rho_\tau [y_i - \{a_j + (X_i - X_j)^T b_j^T B\}] w_{ij}\right)$$

3.For given$(\hat{a}_j, \hat{b}_j)$, $j = 1, \dots, n$, solve $\beta_{L\tau m}$ from the constrained minimization below:

$$\min_{B: B^T B = I} \left(\sum_{j=1}^{n} \sum_{i=1}^{n} \rho_\tau \left[y_i - \left\{\hat{a}_j + \left(X_i - X_j\right) \hat{b}_j^T \left(\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_{m-1}, \hat{\beta}_m\right)^T\right\}\right] w_{ij} + \lambda_n \sum_{k=1}^{p} |\beta_{m,k}|\right)$$

4.Now, put $\hat{\beta}_{L\tau m}$ in the mth column of B, and continue step 2 and 3 till convergence is attained.

5.Update B by $\left(\hat{\beta}_{1L}, \hat{\beta}_{2L}, \dots, \hat{\beta}_{L\tau m}, \beta_0\right)$, and let m to be $m + 1$.

6. If $m < d$, repeat steps 2 to 5 till $m = d$

5. A simulation study

A- The effectiveness of QMAVE was examined through a numerical examples. QMAVE were compared with SIQR and QR for $\tau = (0.25, 0.50 \; and \; 0.75)$. The mean and standard deviation (SD) of the absolute correlation ($|r|$) between $\hat{\beta}^T X$ and the true index $\beta^T X$ and the median of mean squared errors (MMSE) for $\hat{\beta}^T X$ were reported for the sake of comparison.

Example 1:500 datasets (samples) with size $n = 400$ from $y = \beta^T X + \sigma\varepsilon$ were generated. The parameters vector $\beta = (1,1,1,1,1,1,1,1,1,0)^T$. $X_i (i = 1, \dots, 10)$ and ε were generated from standard normal. In this example, we assumed $\sigma = 1$ and $\sigma = 3$.

Example 2:we generated 200 data sets (samples) with size $n = 400$ from:

$$y = \sin\left\{\frac{\pi(u - A)}{C - A}\right\} + \varepsilon ,$$

where $u = \beta^T X, X = (X_1, \dots, X_8)$, $\beta = (1,1,1,1,1,1,1,1)^T/\sqrt{8}$, $A = \frac{\sqrt{3}}{2} - \frac{1.645}{\sqrt{12}}$, $C = \frac{\sqrt{3}}{2} + \frac{1.645}{\sqrt{12}}$, $X_i \sim$ Unif(0,1) and $\varepsilon \sim N(0,1)$ β is estimated with $\tau = (0.25, 0.50 \text{ and } 0.75)$.

B- According to V.S, the efficiency of LQMAVE was checked via a numerical examples. The LQMAVE was compared with LSIQR and LQR for $\tau = (0.25, 0.50\ and\ 0.75)$. The average number of zero coefficients (Ave 0's), the mean and SD of $|r|$ between $\hat{\beta}^T X$ and $\beta^T X$ and MMSE for $\hat{\beta}^T X$ were reported.

Example 3:we generated 500 data sets (samples) with $n = 400$ from $y = \beta^T X + \sigma\varepsilon$, where β as follows:

Model 1: $\beta = (1,1,0,1,0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0)^T$

Model 2: $\beta = (1,1,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0)^T$

Model 3: $\beta = (0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,0,1)^T$

The three models true unknown parameters vector, $X \in R^{20}$ and X= $(X_1, \dots, X_{20})$ were from $N(0, \Sigma)$ and the (i,j) element of Σ is $0.5^{|i-j|}$. The $\varepsilon \sim$ N(0,1). We assumed $\sigma = 1$ and $\sigma = 3$.

Table 1. The comparison of QMAVE, SIQR and QR was depended on example 1

		$\tau = 0.25$			τ=0.50			τ=0.75		
$\sigma = 1$		QMAVE	SIQR	QR	QMAVE	SIQR	QR	QMAVE	SIQR	QR
	Mean|r|	0.9687	0.9654	0.9621	0.9699	0.9677	0.9644	0.9690	0.9662	0.9632
	SD|r|	0.0004	0.0006	0.0007	0.0002	0.0004	0.0006	0.0003	0.0005	0.0007
	MMSE	0.0011	0.0019	0.0023	0.0007	0.0013	0.0017	0.0009	0.0016	0.0020
$\sigma = 3$	Mean|r|	0.9665	0.9633	0.9610	0.9674	0.9654	0.9620	0.9711	0.9642	0.9617
	SD|r|	0.0005	0.0007	0.0008	0.0003	0.0005	0.0007	0.0004	0.0006	0.0008
	MMSE	0.0017	0.0025	0.0030	0.0009	0.0019	0.0022	0.0012	0.0022	0.0025

Table 2. The comparison of QMAVE, SIQR and NQR was depended on example 2.

$\tau = 0.25$				τ=0.50			τ=0.75		
	QMAVE	SIQR	NQR	QMAVE	SIQR	NQR	QMAVE	SIQR	NQR
Mean|r|	0.8887	0.8854	0.8721	0.9139	0.9077	0.8954	0.9101	0.9060	0.8807
SD|r|	0.0977	0.0998	0.1137	0.0951	0.0972	0.1100	0.0966	0.0989	0.1120
MMSE	0.0071	0.0079	0.0093	0.0062	0.0069	0.0080	0.0067	0.0073	0.0087

Table 3. The comparison the LQMAVE, LSIQR and LQR was depended on example 3 model 1.

		LQMAVE	LSIQR	LQR
$\tau = 0.25$				
$\sigma = 1$	Ave0's	10.5900	10.5800	2.6636
	Mean|r|	0.9910	0.9899	0.9760
	SD|r|	0.0023	0.0027	0.0078
	MMSE	0.0017	0.0021	0.0032
$\sigma = 3$	Ave0's	9.9500	9.1000	9.5433
	Mean|r|	0.9667	0.9522	0.8644
	SD|r|	0.0177	0.0189	0.0250
	MMSE	0.0023	0.0028	0.0224
$\tau = 0.50$				
$\sigma = 1$	Ave0's	10.600	10.500	3.9800
	Mean|r|	0.9935	0.9900	0.9780
	SD|r|	0.0024	0.0028	0.0044
	MMSE	0.0006	0.0007	0.0013
$\sigma = 3$	Ave0's	9.9900	9.9800	8.8500
	Mean|r|	0.9708	0.9588	0.8990
	SD|r|	0.01911	0.0200	0.0233
	MMSE	0.0008	0.0009	0.0016
$\tau = 0.75$				
$\sigma = 1$	Ave0's	10.5800	10.5300	2.8400
	Mean|r|	0.9922	0.9900	0.9788
	SD|r|	0.0021	0.0024	0.0073
	MMSE	0.0015	0.0017	0.0029
$\sigma = 3$	Ave0's	9.9700	9.5500	9.5000
	Mean|r|	0.9699	0.9673	0.8654
	SD|r|	0.0179	0.0169	0.0241
	MMSE	0.0021	0.0024	0.0210

Table 4. The comparison of LQMAVE, LSIQR and LQR was depended on example 3 model 2.

		LQMAVE	LSIQR	LQR
$\tau = 0.25$				
$\sigma = 1$	Ave0's	10.200	9.5500	3.1000
	Mean|r|	0.9915	0.9833	0.9785
	SD|r|	0.0018	0.0022	0.0053
	MMSE	0.0014	0.0017	0.0047
$\sigma = 3$	Ave0's	11.000	11.600	7.7998
	Mean|r|	9.6600	0.9577	0.8900
	SD|r|	0.0144	0.0175	0.0449
	MMSE	0.0098	0.0111	0.0390
$\tau = 0.50$				
$\sigma = 1$	Ave0's	11.100	10.900	4.1000
	Mean|r|	0.9955	0.9898	0.9888
	SD|r|	0.0011	0.0017	0.0033
	MMSE	0.0001	0.0005	0.0020
$\sigma = 3$	Ave0's	9.9000	9.6000	7.7500
	Mean|r|	0.9700	0.9666	0.0942
	SD|r|	0.0092	0.0101	0.0170
	MMSE	0.0007	0.0009	0.0018
$\tau = 0.75$				
$\sigma = 1$	Ave0's	10.700	9.6000	3.3550
	Mean|r|	0.9945	0.9840	0.9844
	SD|r|	0.0014	0.0021	0.0043
	MMSE	0.0009	0.0013	0.0029
$\sigma = 3$	Ave0's	10.500	10.400	8.7889
	Mean|r|	0.9666	0.9611	0.9225
	SD|r|	0.0133	0.0141	0.0361
	MMSE	0.0028	0.0039	0.0067

Table 5. The comparison of the LQMAVE, LSIQR and LQR was depended on example 3 model 3.

		LQMAVE	LSIQR	LQR
$\tau = 0.25$				
$\sigma = 1$	Ave0's	10.410	10.400	3.3000
	Mean|r|	0.9889	0.9866	0.9799
	SD|r|	0.0029	0.0038	0.0052
	MMSE	0.0009	0.0013	0.0038
$\sigma = 3$	Ave0's	10.000	9.9000	8.6500
	Mean|r|	0.9410	0.9390	0.8712
	SD|r|	0.0245	0.0270	0.0511
	MMSE	0.0023	0.0026	0.0166
$\tau = 0.50$				
$\sigma = 1$	Ave0's	10.9900	10.9500	2.9000
	Mean|r|	0.9975	0.9911	0.9834
	SD|r|	0.0012	0.0018	0.0033
	MMSE	0.0001	0.0003	0.0025
$\sigma = 3$	Ave0's	10.500	10.400	8.7800
	Mean|r|	0.9660	0.9535	0.9005
	SD|r|	0.0173	0.0180	0.0280
	MMSE	0.0006	0.0006	0.0023
$\tau = 0.75$				
$\sigma = 1$	Ave0's	10.6000	9.8800	3.1500
	Mean|r|	0.9900	0.9924	0.9811
	SD|r|	0.0017	0.0020	0.0034
	MMSE	0.0006	0.0008	0.0025
$\sigma = 3$	Ave0's	9.4800	9.4600	8.7700
	Mean|r|	0.9355	0.9420	0.8999
	SD|r|	0.0195	0.0210	0.0301
	MMSE	0.0017	0.0022	0.0023

Depending on the mean and SD of $|r|$ between $\hat{\beta}_j^T X$ and $\beta_j^T X$ and MMSE of $\hat{\beta}^T X$ with different quantile levels and different values for σ .

From Tables 1 and 2, it can be see that QMAVE has a better performance than the SIQR and QR, also from Tables 3,4 and 5, it is clear that the performance of LQMAVE is better than the performance of LSIQR and LQR for all cases.

From Tables 3-5, it is obvious that LQMAVE gives values of MMSE and SD less than that for LSIQR and LQR, also the results show that the MMSE for the LQMAVE, LSIQR and LQR increase when σ increase for all quantiles values.

6. Real data

To check the performance of QMAVE and LQMAVE, we employed the Newborn Jaundice (NJ) data. We collect The data by the authors from the women's and children hospital in Al-Diwaniya to achieve this aim.

- Newborn Jaundice (NJ) data

In this section, the considered methods were applied in NJ data. Newborn Jaundice is one of the most popular diseases seen in new babies. It often develops in the second or third day of life and reaches its peak around the fourth day, but jaundice can occur within the first 24 hours after birth in rare cases. NJ data contains $n = 100$ observations. We collectv The dataset from the Women's and children's hospital in AL Diwaniya. response Y is TSB mg/dl (JAUNDICE). The eight predictors are the baby's age (number of days) (X1), baby weight kg (X2), PCV to baby g/dl Hematocrit or (Packed Cell Volume) (X3), Hb to baby g/dl (Hemoglobin)(X4), PCV to mother g/dl Hematocrit or (Packed Volume) (X5), Hb to mother g/dl (Hemoglobin) (X6), RBS to baby mg/dl (blood's sugar) (X7), number of brothers infected(X8).

Now we compered proposed methods with others by used of the adjusted R^2, where it gives the percentage of variation explained by only those independent variables than in reality affect the dependent variable, also it can be interpreted as an unbiased and it is more appropriate through the formula:

$$\text{adjusted } R^2 = 1 - \frac{SSE/df_e}{SST/df_t},$$

$df_e = n - p - 1$, df_t=n-1

later we compered between this methods by used prediction error through the formula:

prediction error= $y_i - \hat{y}_i$

6.1: Non sparse methods

Table 6: The adjusted R-square values for the model fit for NJ data with $\tau = 0.25, 0.50$ *and* 0.75.

Model fit	$\tau = 0.25$			$\tau = 0.50$			$\tau = 0.75$		
	SIQR	QMAVE	QR	SIQR	QMAVE	QR	SIQR	QMAVE	QR
Linear	0.798	0.817	0.729	0.822	0.851	0.745	0.801	0.821	0.733
Quadratic	0.828	0.879	0.807	0.843	0.895	0.813	0.834	0.885	0.811
Cubic	0.856	0.879	0.837	0.871	0.895	0.850	0.862	0.885	0.841
Quartic	0.856	0.879	0.837	0.871	0.895	0.850	0.862	0.885	0.841

Table 7: The prediction error (P.E) of the cubic fit for SIQR, QMAVE and QR for NJ data with $\tau = 0.25, 0.50$ *and* 0.75.

Methods	Prediction error		
	$\tau = 0.25$	τ =0.50	τ =0.75
SIQR	0.827	0.792	0.823
QMAVE	0.814	0.787	0.811
QR	0.961	0.848	0.955

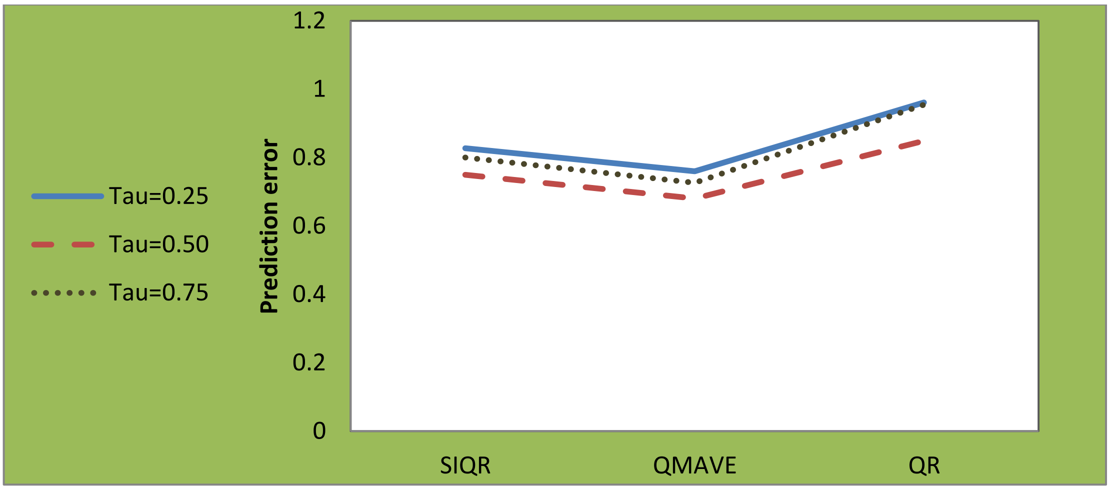

Figure 1: Plot and explanation of the estimated P.E for the studied methods based on NJ.

From table 6, it is clear that the adjusted R-squared for QMAVE is bigger than the adjusted R-squared for SIQR method and the adjusted R-squared for SIQR is bigger than its value for QR method. This means that QMAVE method is the best among the other for all quantile levels.

Table7 we presents the P.E for SIQR, QMAVE and QR based on the NJ data with different quantile levels. It is clear that the QMAVE method has a lower P.E value than the SIQR and the QR methods. This means the performance of QMAVE method is better than the performance of SIQR and QR under different quantile levels.

Figure1explains that the estimated prediction error with $\tau = 0.25, 0.50$ and 0.75, for the QMAVE is less than the estimated prediction error for SIQR and QR, where the three different lines in panel represent the P.E for the three methods in different quantile $\tau = 0.25, 0.50$ and 0.75. The blue line represents the PE at $\tau = 0.25$, the red line represents the PE at $\tau = 0.50$ and the green line represents the PE at $\tau = 0.75$.

6.2: Sparse methods

Table 8: The adjusted R-square values for the model fit for NJ data with $\tau = 0.25, 0.50$ *and* 0.75.

Model fit	$\tau = 0.25$			$\tau = 0.50$			$\tau = 0.75$		
	LSIQR	LQMAVE	LQR	LSIQR	LQMAVE	LQR	LSIQR	LQMAVE	LQR
Linear	0.744	0.884	0.741	0.787	0.895	0.777	0.752	0.886	0.744
Quadratic	0.873	0.902	0.850	0.884	0.918	0.871	0.879	0.907	0.861
Cubic	0.885	0.902	0.849	0.897	0.918	0.866	0.888	0.907	0.854
Quartic	0.885	0.902	0.849	0.897	0.918	0.866	0.888	0.907	0.854

Table 9: The P.E of the cubic fit for LSIQR, LQMAVE and LQR for BJ data with $\tau = 0.25, 0.50$ *and* 0.75.

Methods	Prediction error		
	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$
LSIQR	0.785	0.762	0.780
LQMAVE	0.761	0.738	0.757
LQR	0.817	0.791	0.808

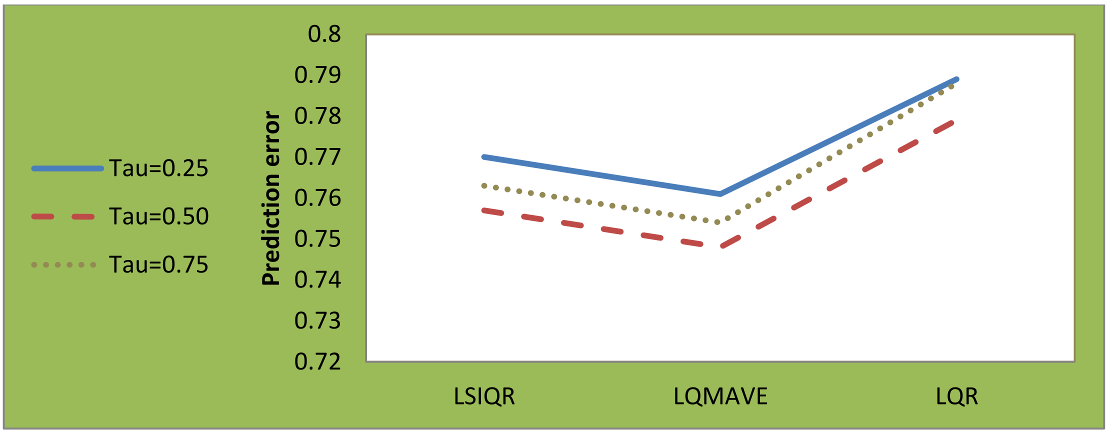

Figure 2: Plot and explanation of the estimated P.E for the studied methods based on NJ.

From table 8, it is clear that the adjusted R-squared values for LQMAVE is bigger than the adjusted R-squared for LSIQR and LQR. This means that LQMAVE method is the best among the other methods under different levels of quantile.

Table 9 we presents the P.E with $\tau = 0.25,\ 0.50$ and 0.75 of the LSIQR, LQMAVE and LQR on the NJ data. The results in the table show that LQMAVE method has a lower P.E than the LSIQR and the LQR methods. This means that LQMAVE has a better behavior than LSIQR and LQR for all the quantile levels.

From Figure 2, it is obvious that the values of estimated P.E with $\tau = 0.25, 0.50$ and 0.75, for the LQMAVE are less than its values for LSIQR and LQR, where the three different lines in panel represent the prediction errors for the three methods in different quantile $\tau = 0.25, 0.50$ and 0.75. The blue line represents the PE at $\tau = 0.25$, the red line represents the PE at $\tau = 0.50$ and the green line represents the PE at $\tau = 0.75$.

7.Conclusion and possible future work

In this paper, QMAVE and LQMAVE were proposed. The QMAVE and LQMAVE were compared with SIQR, QR, LSIQR and LQR. In order to check the behavior of the QMAVE and LQMAVE, simulations were employed. Based on the simulation studies and NJ data, it is clear that the QMAVE and LQMAVE have better behavior in comparison to SIQR, QR, LSIQR and LQR and thus the QMAVE and LQMAVE are useful practically. Future direction or extension of the current work is sparse quantile MAVE with group variable selection penalties.

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In: second International Symposium on Information Theory. Akademiai Kiado, Budapest, pp. 267-281.

Alkenani, A. and Reisan, T (2016). Sparse sliced inverse quantile regression. Journal of Mathematics and Statistics. Volume 12, Issue 3.

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

Cook, R. D. (1994). On the interpretation of regression plots. J. Amer. Statist. Assoc. 89, 177189.

Cook, R. (1998). Regression graphics: ideas for studying the regression through graphics. New York. Wily.

Cook, R. D. and Li, B. (2002). Dimension reduction for the conditional mean in regression. The Annals of Statistics 30, 455–474.

Efroymson, M. A. (1960), “Multiple Regression Analysis,” in Mathematical Methods for Digital Computers, eds. A. Ralston and H. S. Wilf, New York: John Wiley, pp. 191-203.

Fan, J. and Li, R. Z. (2001). Variable selection via non-concave penalized likelihood and its oracle properties. Journal of the American Statistical Association 96, 1348-1360.

Guyon, I. and Elisseeff, A. (2003). An introduction to variable and feature selection. Journal of Machine Learning Research 3, 1157–1182.

Koenker, R. and Bassett, G. (1978). Regression quantiles. Econometric 46,33-50.

Li, K. C. (1991). Sliced inverse regression for dimension reduction. Journal of the American Statistical Association 86, 316-342.

Li, K. (1992). On Principal Hessian Directions for Data Visualization and Dimension Reduction: Another Application of Stein’s Lemma. Journal of the American Statistical Association 87, 1025–1039.

Li, L. and Nachtsheim, C. J.(2006). Sparse sliced inverse regression. Technometrics 48, 503-510.

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

Schwarz, G. (1978). Estimating the dimension of a model. Annals of Statistics 6, 461-464.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society, Series B 58, 267-288.

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.

Xia, Y., Tong, H., Li, W. and Zhu, L. (2002). An adaptive estimation of dimension reduction space. Journal of the Royal Statistical Society Series B 64, 363–410.

Yu, K., Lu, Z. and Stander, J. (2003). Quantile regression: Applications and current research areas. The Statistician 52, 331–350.

Zou, H. (2006). The adaptive lasso and its oracle properties. Journal of the American Statistical Association 101, 1418-142.

Zou, H. Hastie, T. (2005). Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society Series B 67, 301-320.