Republic of Iraq
Ministry of Higher Education
And Scientific Research
University of AL-Qadisiyah
College of Administration and Economics

Sparse dimension reduction through penalized quantile MAVE with application

A thesis

Submitted to the college of Administration and Economics, University of AL-Qadisiyah as a partial fulfillment of the requirements for the Degree of Master of science in statistics

By
Doaa Tahir Malik
Supervised by
Assist. Prof. Dr. Ali Alkenani

2019

Abstract

Minimum average variance estimation (MAVE) has become an important and widely used method, it is powerful dimension reduction method and it is effective in dealing with high dimensional data. It was proven to be an efficient to deal with the dimensionality problem in conditional mean regression. In this thesis, 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 thesis 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. In Chapter one, review of dimension reduction, quantile regression and variable selection concepts were presented. In Chapter two, minimum average variance estimator methods is reviewed, quantile minimum average variance estimator (QMAVE) and sparse quantile minimum average variance estimator with lasso (LQMAVE) penalty method were proposed. In Chapter three, simulations study. In Chapter four, data analysis were employed to check the performance of QMAVE and LQMAVE. In Chapter five, conclusions and future works were reported.

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

و قل رب زدني علما

صدق الله العظيم

سورة طه ,الآية114

The decision if the defense committee

We are the head and members of the defense committee certify that we have been looked at the thesis entitled (Sparse dimension reduction through penalized quantile MAVE with application) and we have debated the student (Doaa Tahir Malik). As a result, the student has defended her thesis and all its content. So that we have found the thesis is worthy to be accepted to award the degree of Master in statistics sciences.

Prof. Dr. Rahim Jabbar Alhamzawi

(The head of the committee)

Assist. Prof. Dr. Waleed Mia Rodeen

(Member)

Dr. Taha hussein Ali

(Member)

Assist. Prof. Dr. Ali Jawad Kadhim Alkenani

(Member and supervisor)

Approval of the college committee

The council of the college of administration and economics at the University of Al_qadissiah have been approved on the decision of the defense committee.

Assist. Prof. Dr Ali Jawad Kadhim

The Dean if the college of administration and economics/Al_qadissiah university

Acknowledge

The Most Compassionate all praise be to Allah the Lord of the worlds.

I would like to express my gratitude to my supervisor Assist. Prof. Dr Ali Alkenani for his supervision of this work, for the useful comments, remarks and engagement through the learning process of this master thesis. Furthermore, I would like to thank him for introducing me to this topic and for the support.

My thanks go also to Dr Tahir Reisan and Dr Taha Hussein Ali, for their contribution and their support.

My thanks go also to Dr Muhammad Dahsh and Pr. Dr Rahim Rahim Jabbar Alhamzawi.

Finally my thanks go also to my family members my father, my mother, my sisters and my brothers for the unlimited support

Dedication

To the first teacher, Prophet Muhammad, who taught us the purpose of life.

There are a number of people without whom this thesis might not have been written.

To my supervisor Assist. Prof. Dr Ali Alkenani.

To my family members my father the wonderful gift of God, my wonderful mother, my sisters and my brothers source of encouragement and inspiration to me throughout my life.

Table of Contents

Chapter five

List of abbreviations

AIC	Akaike information criterion
Adaptive Lasso	Adaptive least absolute shrinkage and selection operator
Ave 0′s	Average number of zero coefficients
BIC	Bayesian information criterion
BEP	Backward elimination procedure
CD	Curse of dimensionality
CMS	Central Mean Subspace
CV	Cross validation
DR	Dimension reduction
FSP	Forward selection procedure
GR	Graphical regression
HD	High dimensional
ICDF	Inverse cumulative distribution function
Lasso	Least absolute shrinkage and selection operator
L.C	Linear combination
LQR	Quantile regression with lasso penalty

LQMAVE	Sparse Quantile MAVE with Lasso penalty
LSIQR	Sparse sliced quantile regression with lasso penalty
MAVE	Minimum average variance estimator
MCP	Minimax concave penalty
MMSE	Median of mean squared errors
O.L.S	Ordinary least squares
OP's	Oracle properties
PHD	Principal hessian directions
QR	Quantile regression
QMAVE	Quantile MAVE
RSS	Residual sum of squares
SD	Standard deviation
SDR	Sufficient dimension reduction
SMAVE	Sparse MAVE
SIR	Sliced inverse regression

SIQR	Sliced inverse quantile regression
SWSE	Step wise selection procedure
SAVE	Sliced average variance estimation
SQMAVE	Sparse quantile MAVE
SCAD	Smoothly clipped absolute deviation
V.S	Variable selection

Chapter one

1.1. Introduction

In some applications of multiple regression, the number of the predictors p became large and therefore the analysis of this data becomes difficult. In order to deal with this problem, we need to shrinkage dimensionality of X with few assumptions. When we talk about reducing of dimensions this means that there are high dimensions, these dimensions are known as variables or features. The greater the number of these variables, the more difficult it will be to deal with them. Therefore, we will encounter a problem known as curse of dimensions, which occurs when the variables are increasing to the multivariate model, where the greater the number of dimensions, the more difficult it is to predict a specific quantity. These variables may not be all effective or influential or may be interconnected and therefore redundant and this requires reduction. Therefore, the process of reducing the dimension is the process of converting high-dimensional data to a less space. It plays an important role to address this problem and by reducing the number of random variables, in other word, simplifying the understanding of data only numerically or visually and thus maintaining the integrity of data.

In addition, there are other benefits to reduce the dimension where it works on data compression and reduces the time of calculations and there are some algorithms do not work well when the dimensions are so high. Therefore, we should work on reducing the dimension and make the algorithms useful.

Also we need to deal with the problem of links and that by removing the duplicated features and thus reduce the number of features in the data set without the need to lose a lot of information and maintain or improve the performance of the model. It

is an effective way to deal with huge sets of data and in general, the dimension can be reduced in two ways:

1-Variable selections:- A technique that is used extensively in machine learning to select partial sets of variables, where we specify subgroups of the original set of variables to get the smallest subset that can be used to model the problem. The process of selecting the variables helps to give a clearer understanding of the variables and data by informing us about the important variables and their relationship with each other, in addition to that is reduces the cost. In general it is divided into two types of methods, traditional and regularization methods.

Traditional methods include such as stepwise selection Efroymson (1960), AIC (Akaike, 1973) and BIC (Schwarz, 1978), and when it is compared with regularization methods, one can notice the instability high variance. whereas, the regularization methods which are first used (to determine the variable) by Donho and Johnston in (1994) such as lasso (least absolute shrinkage and selection operator) the Lasso (Tibshirani, 1996), adaptive lasso (Zou, 2006), (SCAD) (Fan and Li, 2001) and elastic-net (Zou and Hastie, 2005) among others. These methods are more stable than traditional method because the process of estimation and selection of variables is continuously achieved through continuous reduction

2 - Variable extraction:- is the process that leads to the reduction of dimensions when the input of algorithms is very large to be processed easily. we reduce the data with high dimensions to the area of less dimensions and there are many new

techniques to extract the variables and to reduce the dimensions without losing significant number of the information.

Sufficient dimension reduction SDR (Cook, 1998) was introduced to achieve this aim. Many methods were suggested to estimate the SDR space. Some of them focusing on finding the central subspace $S_{Y/X}$. Examples for these methods are Graphical regression GR (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 CMS for DR. Many DR methods were suggested under this concept, for examples, PHD (Li, 1992) and MAVE (Xia et al., 2002) among others. MAVE is a powerful dimension reduction method and it is effective in dealing with high-dimensional data. The MAVE was shown to be an efficient to deal with DR in conditional mean regression.

1.2. Literature Review

Many authors studied sufficient dimension reduction, which depend feature selection and feature extraction. Now outlining some rummages for this subject.

Examples of regularization methods are the Lasso (Tibshirani, 1996), adaptive lasso (Zou, 2006), (SCAD) (Fan and Li, 2001) and elastic-net (Zou and Hastie, 2005) among others.

Under the framework of the SDR, Ni et al. (2005) suggested a penalized SIR. Li and Nachtsheim (2006) suggested another version of the sparse SIR. Li (2007) proposed Sparse Sliced Inverse Regression. 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.

QR has attracted much significant interest in the literature. For examples, Yu et al. (2003), Koenker (2004), Melly (2006), Wang et al.(2007), Li and Zhu (2008), Zou and Yuan (2008) , Owen (2008), Wu and Liu (2009), Yuan and Yin (2010), Li et al. (2010), Alhamzawi and Yu (2012), Alkenani et al. (2012), and Alkenani and Yu (2013) Alkenani and Dikheel (2016), Ciuperca (2016), Ou et.,al. (2017).

1.3. Aim

The first aim in this thesis is QMAVE was proposed. QMAVE combines the strength of QR with the effective method MAVE under the sufficient dimension reduction framework. The details of QMAVE was reported in chapter 2.

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

The second aim is sparse QMAVE with Lasso penalty (LQMAVE) 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.

The rest of the thesis is organized as follows: In chapter 2,varaibles selection methods, a short review of MAVE was given, QMAVE and LQMAVE were proposed, respectively. Numerical experiments in chapter 3. real data was reported in chapter 4. The conclusions and possible future work were reported in 5.

Chapter two

2.1. Variable selection(V.S)

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) .

2.1.1. Traditional V.S

V.S techniques, such as stepwise selection Efroymson (1960), AIC (Akaike, 1973) and BIC (Schwarz, 1978) are highly time consuming, discrete procedures with high variance and suffer from instability (Brieman,1996).

2.1.1.1. Step wise selection procedure

This method is a development of the forward selection method establish its basis (Efroymson, 1960) to make it more efficient and the point of distinction between the two methods is that all independent variables at the end of each step are ascertained by relying on the choice ($F_{partail}$)and re-evaluated again because there are strong relationships between the independent variables that were introduced in the previouse steps and so the way stepwise selection is the best way to choose the best regression equation (Zahra Hasan AlTameemi et.,al,2014).

2.1.1.1.1. Forward selection procedure

This method depends on starting without any independent variable and the independent variables are chosen to be included in the equation one after the other depending on the comparison ($F_{partial}$) for each variable with a value ($F_{tabular}$) . the highest value is chosen ($F_{tabular}$) for each step and after making sure that value is greater than (F_{in}) the variable in question is entered into the eqution and the steps continue to show the independent variables one after the other untile we get to the top ($F_{partial}$) less than (F_{in}) according to the following formula.

$$F^* = \frac{SSR(x1)}{\frac{SSE(X1)}{n-2}} \quad (1)$$

where SSR:-represents the deviations showen

SSE:-represents the unelarified deviations

n:- sample size

2.1.1.1.2. Backward elimination procedure

This method begins with the adoption of all the independent variables in the equation and then deletes the variables from the equation one after the other depending on the value ($F_{tabular}$) ,which is called (F_{IN})the steps are described as follows:

Steps one:-work to include all independent variables in the regression equation and calculate the values ($F_{partial}$)for each variable according to the following formulas:-

$$F_{i\ partial}= \frac{SSR\left[\frac{xi}{all\ other\ explanatory\ variables}\right]}{\frac{SSE(x1,\dots,xk)}{n-k-1}} \qquad (2)$$

And chooses the variable that has the lowest value ($F_{partial}$) , and compare with (F_{IN}). If it proves that (Fi)<(F_{IN}) the relevant variable is deleted from the equation and the second step is moved to the degree of freedom of the numerator (1) and the denominator (n-k-1).

Step two :- all independent variables except those deleted in step (1) are included ($F_{partial}$) for each of the remaining variables of the first step the smallest is chosen and compared with (F_{IN}) to the degree of freedom of the numerator (1) and the denominator (n-k-2) . if ($F_{partial}$) <(F_{IN}) ,deletes the variable in question and moves to step (3) and so the steps continue untill the smallest value is obtained (F_{IN})< ($F_{partial}$) the solution stops .

2.1.1.2. Akaike Information Criteria(AIC)

Akaike (1973) developed AIC, which is a measure of the relative quality of statistical modeling. Its main objective is to make a distinction between the estimated models by relying on its lowest value (Sugiura, 1978), where the model with the lowest value is considered the best model and is expressed as follows:

$$AIC(k)=-2Ln(L)+2k, \quad (3)$$

where k: is the number of parameters.

L: is the value of MLE.

2.1.1.3. Bayesian information criteria(BIC)

Schwarz (1978) proposed BIC which is one of the traditional V.S methods. It is similar to AIC, but it differs than it by including the sample size where it was taken into consideration which makes it better than the Akaiki information criteria (Carlos and Sergioc, 2012). The model with lowest value is considered the best and it is expressed by the following formula:

$$BIC(k)=-2Ln(L)+k\ Ln(n), \quad (4)$$

where n: represents the number of observations .

2.1.2. Regularization methods

Regularization techniques can also carry out variable selection. Therefore Regularization methods is the technique which is used to solve complexity models problem. In Generalization performance is closely related to the complexity model. where the model with high complexity tends to have low bias and high variance. while the model low complexity tends to have high bias and low variance, therefore the Regularization methods are frequently used to control the model complexity by penalizing to more complex models. First use of regularization techniques for V.S is made by Donoho and Johnstone (1994). Regularization methods can be formed by adding penalty terms to the standard loss functions, such as O.L.S loss function . In regularization methods the V.S is implemented through the parameter estimation process (Wang and Yin, 2008). Examples of regularization methods are the Lasso (Tibshirani, 1996), adaptive lasso (Zou, 2006), (SCAD) (Fan and Li, 2001) and elastic-net (Zou and Hastie, 2005) among others.

2.1.2.1. Lasso

Lasso was proposed by Tibshirani (1996) for parameters estimation and V.S. It is considered a powerful and effective to tackle HD data. In lasso, the RSS was minimized subject to $\sum_{k=1}^{p} |\beta_k|$ being less than a constant. According to this condition, Lasso shrinks some coefficients and sets other to 0. In this method, (Tabshirani, 1996) added a penalty function to the least squares loss function as follows:

$$\underset{\beta}{\text{argmin}} \quad ||y - x\beta||_2^2 + \lambda \, ||\beta_k||_1 \tag{5}$$

where $||y - x\beta||_2^2 = \sum_{i=1}^{n}(y_i - X_i^{\mathrm{T}}\beta)^2$

$||\beta_k||_1 = \sum_{k=1}^{p}|\beta_k|$ is the l_1 norm of β

$\lambda \geq 0$ controls the strength of penalty, the bigger value of λ gives the greater amount of shrinkage

2.1.2.2. Adaptive Lasso

Zou (2006) proposed Adaptive Lasso, where Zou criticized Lasso and he explained that Lasso estimates are biased because lasso imposes the same restriction on all coefficients and this produces estimates will not be consistent. Zou (2006) explained that it is possible to impose different weights for different coefficients in the penalty function, thus we will reduce the bias and improve the accuracy of V.S and as a result, we get unbiased and consistent estimates. Adaptive Lasso estimates can be obtained by the following formula:

$$\underset{\beta}{min} \quad ||y - x\beta||_2^2 \quad + \lambda \overset{*}{w_k} \quad ||\beta_k||_1 \tag{6}$$

$\lambda > 0$ is the tuning parameter.

$\overset{*}{w} = \frac{1}{\hat{\beta}_{ols}}$ is a known weights vector.

2.1.2.3. Elastic net

Zou and Hastie (2005) proposed Elastic net. They studied lasso and they pointed a number of limitations of lasso as follow:

1. In case if p> n the Lasso selects almost n variables.

2. If there is a set of variables strongly related, then Lasso will select only one from the group and ignores the rest of the variables.

They also proved that the Elastic net method has the Oracle properties. Elastic net estimates can be obtained by the following formula:

$$\min_{\beta} \quad ||y - X\beta||_2^2 + \lambda_1||\beta_k||_1 + \lambda_2||\beta_k||_2^2, \tag{7}$$

where $\lambda_1 \; and \; \lambda_2$ are the tuning parameters

The elastic net penalty is a combination of the lasso and ridge penalties.

2.2. MAVE

Cook (1998) introduces the theory of sufficient dimension reduction to minimize the number of predictors and maintain the whole information of regression and use small number of hypothesis. Therefore, there are several proposed methods to estimate sufficient dimension reduction.

Some of these methods focus on finding the central subspace that referred to as $S_{Y/X}$,and examples include graphical regression(Cook, 1994), sliced average variance estimation (SAVE) (Cook and Weisberg,1991) and sliced inverse regression (SIR) (Li, 1991).

It is noted that in many cases of regression analysis, the focus is on estimating the central mean subspace. Cook and Li (2002) introduced the notion of central mean subspace to reduce dimensions when the mean function is interest. There are some dimension reduction methods included in this classify, for examples, Principal Hessian direction (PHD) (Li and Cook, 1992) and the most popular method to estimate the CMS which is the MAVE (Xia et al. 2002). The MAVE method is used to a wide range of models, with fewer constraints on the distribution of the covariates. The MAVE is especially effective in dealing with high-dimensional covariates and it was shown to be an powerful dimension reduction method in high dimensional regression problems (Zhu et al. 2006). MAVE was attracted much significant interest in the literature.

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

$$\min_{\mathrm{B}} \{ \mathrm{E}[y- \mathrm{E}(y|X^T\mathrm{B})]^2\} , \quad (8)$$

where $B^TB=I_d$.

The variance given X^TB is

$\sigma^2{}_{\mathrm{B}}(X^TB)= \mathrm{E}[\{y-\mathrm{E}(y|X^T\mathrm{B})\}^2|X^T\mathrm{B}]$.

Thus,

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

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

$$\sigma_B^2(\mathrm{X}_0\mathrm{B}) \approx \sum_{i=1}^n \{y_i - E(y_i|X_i^TB)\}^2 w_{i0}$$

$$\approx \sum_{i=1}^n [y_i - \{\mathrm{a}_0 + (X_i - X_0)^T B b_0\}]^2 \, w_{i0} \quad (9)$$

where a_0+$(X_i$ -$X_0)^{\mathrm{T}}B\mathrm{b}_0$ is the local linear expansion of $\mathrm{E}(y_i|X_0^T\mathrm{B})$ at X_0, and w_{i0} ≥ 0 are the kernel weights centered at $X_0^T\mathrm{B}$ with $\sum_{i=1}^n w_{i0=}1$

w_{i0}=$K_h\{B(X_i - X_0)^T\}/\sum_{i=1}^n K_h\{B(X_i - X_0)^T\}$,

where $K_h(.) = h^d K(\frac{.}{h})$, where $K(.)$ is the Gaussian Kernel as follows (Brillinger, 1983).

$$K(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \quad (10)$$

where h : represents the Smoothing parameter, Sometimes it is called the bandwidth, which is the parameter represents the size of the distribution or contrast window. Smooth parameter greatly affects the similarity of the estimated curve and the real curve. For proper approximation, it is necessary to find the best way to balance between the bias and the variance in order to obtain the least possible errors. Therefore, the selection of smooth parameter value must be done with high accuracy, (Altman, 1990). By increasing the value of this parameter the bias increases and the variance value decreases. But when we decrease this parameter the opposite occurs. There are many methods for estimating this the bandwidth, including the method of cross validation, which is one of the common methods used to find the value of the bandwidth where it plays an important role in both bias and the shape of the curve.

The basic idea of the method is that each time one value of the explanatory variable is excluded and it can be represented by the following formula:

$$cv(h) = n^{-1} \sum_{i=1}^{n} [y_i - \hat{m}_{h.j}(x_j)]^2 \, w(x_j) \quad (11)$$

$$\widehat{m}_{h.j}(x_j) = n^{-1} \sum_{i \neq j}^{n} w_{hi}(x_j) y_i$$

Can be obtained from solving (12)

$$\min_{\substack{\text{B: } B^T \text{B} = \text{I} \\ a_j, b_{j, j=1,\dots,n}}} \left(\sum_{j=1}^{n} \sum_{i=1}^{n} [y_i - \{a_j + (X_i - X_j)^T B b_j\}]^2 w_{ij} \right) \qquad (12)$$

2.3. Algorithm of MAVE

The algorithm of MAVE was explained as follows:

1. Let $m = 1$ and B=β_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:

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

4. Now, put $\hat{\beta}_m$ in the mth column of B, 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:

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

2.4.Quantile regression

Koenker and Bassett (1978) proposed QR is approach with great potential and is a very fertile in terms of a possible applications and it is analysis differs from more the traditional regression models in its focus on the distributions And the QR estimates actually includes far more information than can be presented in simple tables. QR is a powerful estimation method for regression models which offers a robust approach to observe how covariates influence the response distribution in different levels. In addition the QR estimates actually includes far more information than can be presented in simple tables. It has become well-known technique to describe a distribution of the response variable given the set of predictors. Therefore the QR gives a complete analysis of a stochastic relationships

between random variables and is much better suited to analyzing questions include changes in the distribution of the dependent variable.

Mean regression analysis focus on the relationship among the conditional mean of the Y_i given X_i. QR was attracted a lot of theoretical practical interest (see Koenker (2005), Yu et al. (2003) for details).

One of the attractiveness of QR is its ability to give a clear and whole picture of the relationship between Y_i and X_i. It is robust to non–normal errors and outliers (Yu et al., 2003). This can be illustrated in the following diagram:-

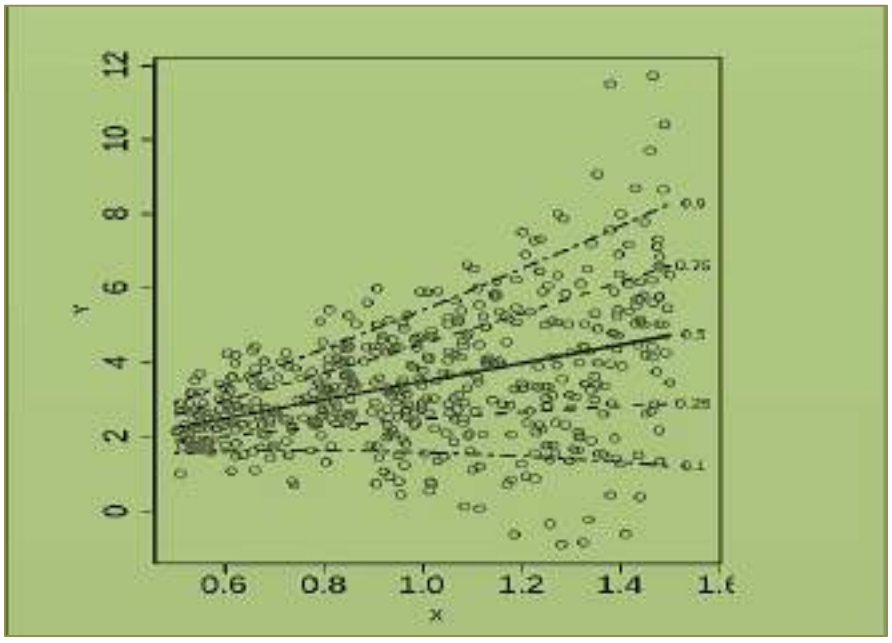

Figure 1: shows quantile regression levels. (Koenker et al.,2001)

To explain the importance of QR let y_i be a response variable and X_i a $p \times 1$ vector of predictors for the ith observation, $q_\tau\ (X_i)$ is $ICDF$ 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 minimizer of the following:

$$\min_{\beta_\tau} \sum_{i=1}^{n} \rho_\tau (y_i - X_i^T \beta_\tau), \tag{13}$$

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

$$\rho_\tau(u) = \tau u I_{[0,\infty)}(u) - (1-\tau) u I_{(-\infty,0)}(u) \tag{14}$$

$$\rho_\tau = \begin{cases} \tau u & u \geq 0 \\ -(1-\tau)u & u < 0, \end{cases}$$

where u: represents $(y_i - X_i^T \beta_\tau)$. The check loss function can be illustrated by the following diagram:

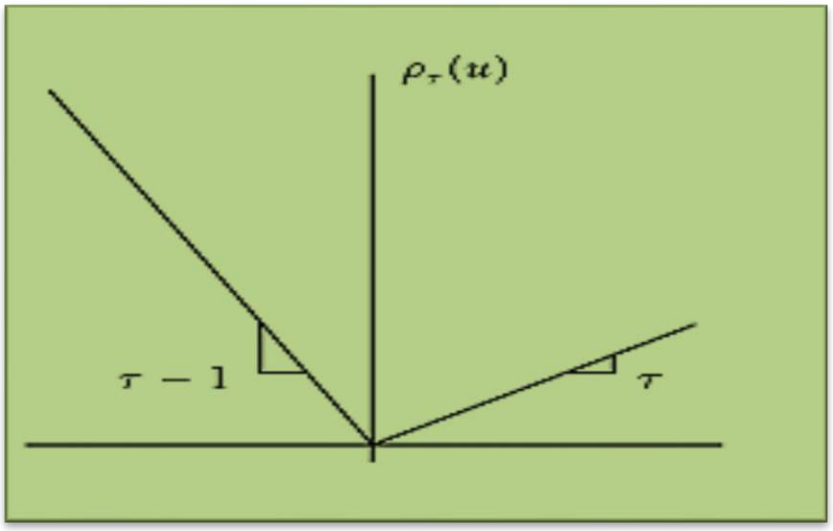

Figure 2: shows the check loss function.(Koenker et al.,2001)

2.5. QMAVE

QMAVE combines the strength of QR with the effective method MAVE under the sufficient dimension reduction framework, where the QR gives a complete analysis of a stochastic relationships between random variables and is much better suited to analyzing questions include changes in the distribution of the dependent variable and MAVE method it was shown to be an powerful dimension reduction method in high dimensional regression problems. Therefore QMAVE method give us a good tool to obtain sufficient dimensions reduction under quantile regression settings

The QMAVE was proposed as a minimizer of the following formula:

$$\left(\sum_{j=1}^{n}\sum_{i=1}^{n}\ \rho_\tau[y_i-(a_j+(X_i-X_j)^TBb_j)]w_{ij}\ \right) \quad (15)$$

2.6.Algorithm of QMAVE

In this chapter, QMAVE method has been proposed. 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:

$$\underset{a_j b_{j=1,\dots,n}}{min}\ \left(\sum_{j=1}^{n}\sum_{i=1}^{n}\rho_\tau[y_i-\{a_j+(X_i-X_j)^Tb_j^TB\}]w_{ij}\right)$$

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

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

4. Now, put $\hat{\beta}_{\tau m}$ 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.

2.7. LQMAVE

MAVE (Xia et al. 2002) suffers from the fact that each dimension reduction component is a linear combination of all the original predictors so that it is difficult to interpret the resulting estimates.

QMAVE method give 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 of all of the predictors, which may not be simple to explain the resulting estimates.

In this chapter, Sparse QMAVE with Lasso penalty (LQMAVE) was proposed.

LQMAVE can be solved as a minimizer of (16).

$$\left(\sum_{j=1}^{n}\sum_{i=1}^{n} \rho_\tau[y_i - (a_j + (X_i - X_j)^T B b_j)]w_{ij}\right) + \lambda\sum_{k=1}^{p}|\beta_k| \qquad (16)$$

where a_0+(X_i -X_0)$^T B$b$_0$ is the local linear expansion, $\sum_{k=1}^{p}|\beta_k|$ is the l_1 norm of β.

where $\lambda > 0$ is the tuning parameter. The tuning parameter plays an important role in the process of selection of significant variables, where it controls the amount of shrinkage or penalty. If the value of the tuning parameter a tend to zero, the estimation of the least - squares that is not penalized will be obtained. A wide range of methods have been developed for the estimation of tuning parameter such as the Cross - Validation method, which is one of the commonly used methods.

$$CV(\lambda) = \frac{1}{n}\sum_{i=1}^{n} e_i(\lambda) \qquad (17)$$

where $e_i(\lambda) = \sum_{i=1}^{k}(y_i - \hat{y}_i(\lambda))^2$,

2.8. Algorithm of LQMAVE

In this chapter, LQMAVE was proposed 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 $m = d$

Chapter three

A simulation study

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

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

Example 2: the same example was used as the first example, but the size of the sample was resized n=100

Example 3: We generated 200 datasets with $n = 400$ 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 $\tau = (0.25, 0.50 \text{ and } 0.75)$.

Example 4: : the same example was used as the third example, but the size of the sample was resized n=100

B- According to V.S, the efficiency of LQMAVE was checked via numerical examples. The LQMAVE was compared with LSIQR and LQR for $\tau = (0.25, 0.50\ and\ 0.75)$. The 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 5:We generated 500 datasets with $n = 400$ from $y = \beta^T X + \sigma\varepsilon$, where βas 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: $\beta = (0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,0,1)^T$

The three models true unknown parameters vector, $X \in R^{20}$ and X= $(X_1, \dots, X_{20})$ were from $N(0, \sum)$ and the (i, j) element of $\sum$ 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

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

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

		QMAVE	SIQR	QR
$\tau = 0.25$				
$\sigma = 1$	Mean|r|	0.9481	0.9466	0.9432
	SD|r|	0.0006	0.0009	0.0010
	MMSE	0.0015	0.0023	0.0030
$\sigma = 3$	Mean|r|	0.9332	0.9327	0.9311
	SD|r|	0.0007	0.0010	0.0010
	MMSE	0.0026	0.0039	0.0047
$\tau = 0.50$				
$\sigma = 1$	Mean|r|	0.9544	0.9519	0.9501
	SD|r|	0.0004	0.0007	0.0008
	MMSE	0.0011	0.0018	0.0023
$\sigma = 3$	Mean|r|	0.9522	0.9507	0.9489
	SD|r|	0.0005	0.0008	0.0009
	MMSE	0.0021	0.0029	0.0033
$\tau = 0.75$				
$\sigma = 3$	Mean|r|	0.9690	0.9662	0.9632
	SD|r|	0.0005	0.0008	0.0009
	MMSE	0.0014	0.0024	0.0027
$\sigma = 3$	Mean|r|	0.9711	0.9642	0.9617
	SD|r|	0.0006	0.0009	0.00010
	MMSE	0.0023	0.0034	0.0040

Table 3. The comparison of QMAVE, SIQR and NQR was depended on example 3.

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

Table 4. The comparison of QMAVE, SIQR and NQR was depended on example 4

		QMAVE	SIQR	NQR
$\tau = 0.25$				
	Mean|r|	0.8386	0.8366	0.8320
	SD|r|	0.1105	0.1221	0.1240
	MMSE	0.0121	0.0136	0.0142
$\tau = 0.50$				
	Mean|r|	0.8668	0.8406	0.8398
	SD|r|	0.1001	0.1122	0.1199
	MMSE	0.0098	0.0120	0.0131
$\tau = 0.75$				
	Mean|r|	0.8445	0.8381	0.8334
	SD|r|	0.0988	0.1197	0.1222
	MMSE	0.0101	0.0120	0.0140

Table 5. The comparison of LQMAVE, LSIQR and LQR was depended on example 5 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 6. The comparison of LQMAVE, LSIQR and LQR was depended on example 5 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 7. The comparison of LQMAVE, LSIQR and LQR was depended on example 5 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 1,2, contains information about results for the QMAVE, SIQR and QR based on the linear model, it can be seen that QMAVE has a better performance than the SIQR and QR.

From Tables 3,4, contains information about results for the QMAVE, SIQR and QR based on the nonlinear models, it can be seen that QMAVE has a better performance than the SIQR and QR.

From Tables 5-7, contains information about results for the LQMAVE, LSIQR and LQR based on the linear models, 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.

Chapter four

Real data

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

- Newborn Jaundice (NJ) data

In this chapter, the considered methods were applied on 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, but this is rare. Many babies become jaundiced because new born babies have a bigger number of red blood cells, which are also breaking down more quickly, therefore more bilirubin is released into the blood. Jaundice disappears by the time your baby is 10 days old , however a small number of babies do need some treatment depending on how much bilirubin there is in the blood.

NJ data contains $n = 100$ observations. The dataset was collected from the Women's and children's hospital in AL Diwaniya. The 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 Cell 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= $y_i - \hat{y}_i$

4.1. Non 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	$\tau = 0.25$			$\tau = 0.50$			$\tau = 0.75$		
	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 9: The prediction error (PE) 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

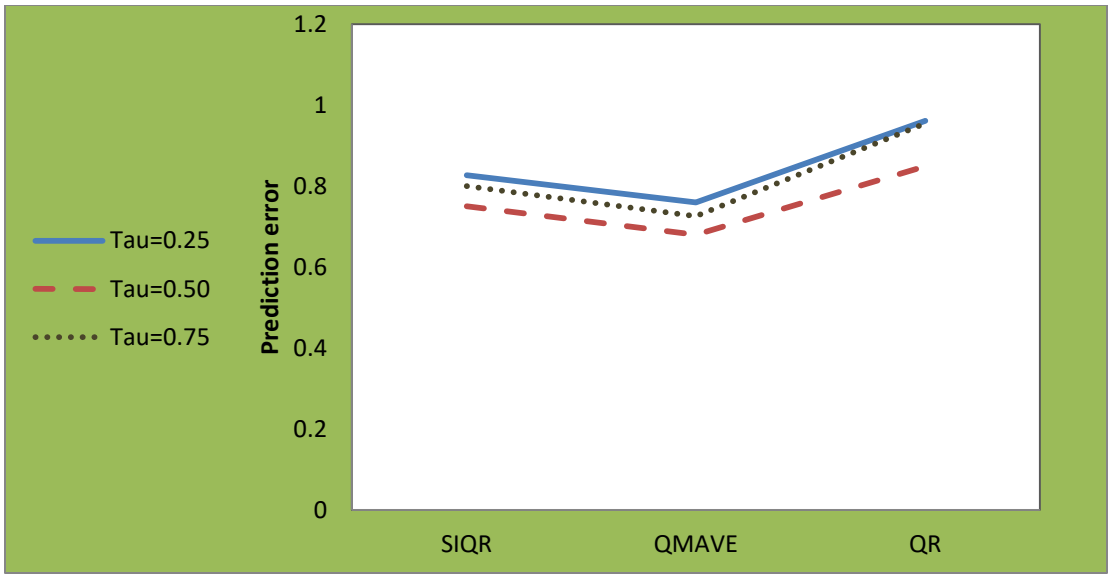

Figure 3: Plot and explanation of the estimated PE for the studied methods based on NJ.

Table 8 reports the values of adjusted R-squared with $\tau = 0.25, 0.50$ and 0.75 for the four model fit based on the NJ data for all the studied methods. It can be seen 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.

Table 9 presents the prediction error for all the considered methods which are studied based on the NJ data with different quantile levels. It is clear that the QMAVE method has a lower prediction error than the SIQR and the QR methods. This means that the QMAVE method show a better performance than SIQR and QR under different quantile levels.

Figure3 explains 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 is the P.E at $\tau = 0.50$ and the green line is the P.E at $\tau = 0.75$.

4.2. Sparse methods

Table 10: 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 11: 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

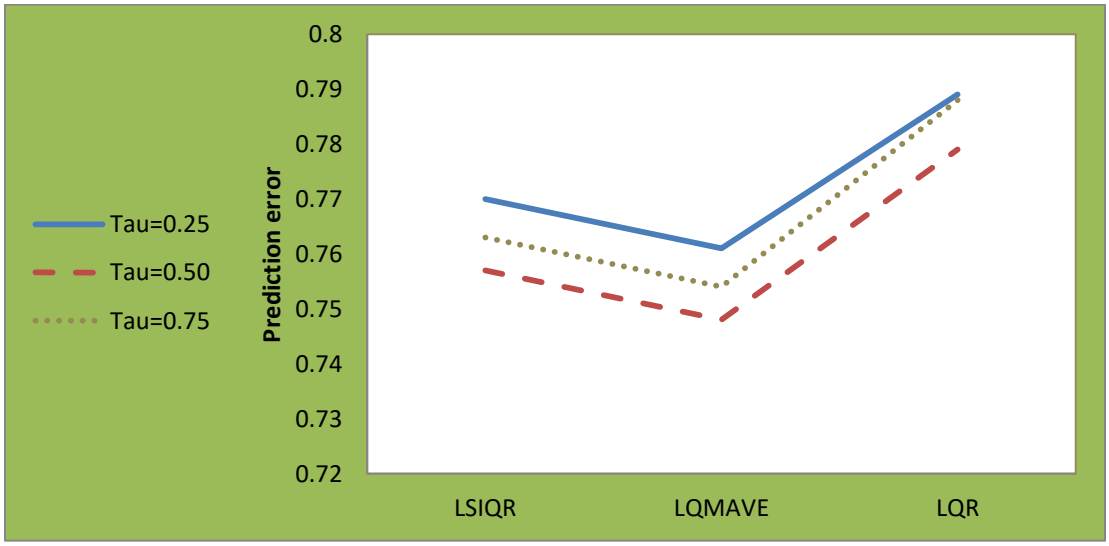

Figure 4: Plot and explanation of the estimated P.E for the studied methods based on NJ.

Table 10 reports the values of adjusted R-squared with $\tau = 0.25, 0.50$ and 0.75 for the model fit based on the NJ data for all the studied methods. It can be seen that the adjusted R-squared 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 11 presents the prediction error with $\tau = 0.25,\ 0.50$ and 0.75 of the methods which are studied based on the NJ data. The results in the table show that LQMAVE method has a lower prediction error than the LSIQR and the LQR methods. This means that the LQMAVE method has a better performance than LSIQR and LQR for all the quantile levels.

From Figure 4, 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 $\tau = 0.75$.

Chapter five

5.1. Conclusions

In this thesis, 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.

5.2. Possible Future work

Future direction or extension of the current work is sparse quantile MAVE with group variable selection penalties, sparse quantile MAVE with adaptive lasso and sparse quantile MAVE with elastic net.

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. Alhamzawi, R. and Yu, K. (2012). Penalized Flexible Bayesian Quantile Regression. Applied Mathematics 3, 2155-2168.

Alhamzawi, R. Yu, K. and Benoit, D. (2012). Bayesian adaptive lasso quantile regression. Statistical Modelling 12, 279-297.

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).

Altman, N. S. (1990). Kernel smoothing of data with correlated errors. J. Am. Statist. Ass., 85,749-759.

Brillinger, D. R. (1983). A generalized linear model with (Gaussian) regression variables. In A Festschrift for Erich L. Lehmann (eds P.J. Bickel, K. A. Doksum and J. L. Hodges, Jr), pp. 97-114. Belmont: Wadsworth.

Breiman, L. (1996). Heuristics of instability and stabilization in model selection. The Annals of Statistics 24, 2350-2383.

Bassett, G. W. and Chen, H. (2001). Portfolio style: return- based attribution using quantile regression. Emp. Econ., 26, 293-305.

- Cook, R. D. and Weisberg, S. (1991). Discussion of Li (1991). Journal of the American Statistical Association 86, 328-332.

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.

Carlos, A. M. and Sergioc, C. S.(2012). Does BIC Estimate and Forecast Better than AIC?. Available at (https://mpra.ub.uni-muenchen. de/42235/).

Donoho, D. and Johnstone, I. (1994). Ideal spatial adaptation by wavelet shrinkage. Biometrika 81, 425-455.

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.

Koenker et al.,(2001). Quantile Regression for Duration Data: A Re-appraisal of the Pennsylvania Reemployment Bo-nusExperiments. Empirical Economics. March,26:1, pp.199-220

Koenker, R. (2004). Quantile regression for longitudinal data. Journal of Multivariate Analysis 91, 74-89.

Koenker, R. (2005). Quantile Regression, Cambridge, U.K.: Cambridge University Press.

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.

Li, L.(2007). Sparse sufficient dimension reduction. Biometrika 94, 603- 613.

Li, Y. and Zhu, J. (2008). l_1-norm quantile regressions. Journal of Computational and Graphical Statistics 17, 163- 185.

Li, Q. Xi, R. and Lin, N. (2010). Bayesian Regularized Quantile Regression. Bayesian Analysis 5, 1- 24.

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

Naiman, D. Q. (2007). Quantile regression- series: quantitative applications in the social sciences. The United States of America, SAGE publications.

Sugiura, N. (1978). Further analysis of the data by Akanke's information criterion and the finite corrections, Communications in statistics- Theory and Methods, A7:13-26.

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.

Wu, Y. and Liu, Y. (2009). Variable selection in quantile regression. Statistica Sinica 19, 801-817.

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.

Yuan, Y. and Yin, G. (2010). Bayesian quantile regression for longitudinal studies with non- ignorable missing data. Biometrics 66, 105- 114.

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

Zhu, L, X. et., al. (2006). On sliced inverse regression with high- dimensional covariates. Journal of the American Statistical Association 101, 630-643.

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

Zou, H. and Yuan, M. (2008). Composite quantile regression and the oracle model selection theory. Annals of Statistics 36, 1108- 1126.