Int. J. Agricult. Stat. Sci. Vol. 17, Supplement 1, pp. 2395-2401, 2021 www.connectjournals.com/ijass
DocID: https://connectjournals.com/03899.2021.17.2395 ISSN : 0973-1903, e-ISSN : 0976-3392

ORIGINAL ARTICLE

SPARSE MINIMUM AVERAGE VARIANCE ESTIMATION VIA QUANTILE REGRESSION GROUP VARIABLE SELECTION PENALTIES (GLQMAVE)

Rihab Ahmed* and Waleed Rodeen

Department of Statistics, College of Administration and Economics, University of Basrah, Al-Basrah, Iraq.
E-mail: rihabahmed998@gmail.com

Abstract: In this paper we will deal with the study of the statistical properties of variable selection method which is called the Group Lasso estimator in high-dimensional data with quantile regression model. The most characteristic of quantile regression is that it allows us to identify all the conditional distribution by estimating many different conditional quantities, with the use of an effective method to reduce sufficient dimension is the method of MAVE. Our proposed method is GLQMAVE, it is new method similar to methodology of many approaches that are interested in estimation and select the informative covariates simultaneously.

JEL classification: F1, F150, F140.

Key words: Group lasso (GL), Dimension reduction, Minimum average variance estimator (MAVE), Quantile regression(QR), Sufficient dimension reduction (SDR), Central mean subspaces (CMS).

Cite this article

Rihab Ahmed and Waleed Rodeen (2021). Sparse Minimum Average Variance Estimation via Quantile Regression Group Variable Selection Penalties (GLQMAVE). *International Journal of Agricultural and Statistical Sciences.* DocID: https://connectjournals.com/03899.2021.17.2395

1. Introduction

As a result of the scientific development that led to an increase in the volume of data, especially in the medical field, and as a result of technological development in the process of collecting high-dimensional data, this led to obtaining estimates of high contrast and biased, due to the presence of a number of problems, and therefore, we cannot use standard regression methods, but rather resorting to alternative methods [Saini and Kumar (2019)]. Quantile regression (*QR)* method is distinguished by its ability to provide a comprehensive and accurate description of the relationship between the explanatory variables X_i ' s and the response variable (y) by modeling the conditional distribution $y / x = \left[x_1, \dots x_p\right]$ in various quantities, studies that focus on theoretical properties indicate that the *QR* is insensitive to heteroskedasticity and outliers and thus is able to accommodate errors (residues) that are not normally distributed in many applications [Koenker and Bassett (1978)]. The *QR* has been applied in many different fields such as econometrics and finance (financial markets) and medical and agricultural studies [Hashem *et al*. (2016), Dikheel and Abdalriadha (2020)]. However, one of the disadvantages of Quantile regression is that it does not give us a unique solution, so we resort to the penalized methods. Penalized methods, researchers Donoho and Johnstone (1994) first developed the idea and then Tibshirani (1996) developed it, among its advantages compared to traditional methods is that it is more stable than traditional methods., among these methods, among these methods, we mention Lasso Adaptive Lasso, SCAD , Elastic net and Group Lasso, as these methods impose a penalty on the size of the parameters, which makes it possible to estimate the regression coefficients with a large number of variables and a relatively small number of observations (*i.e.* $P > n$), which can improve the predictive error of the model by reducing the variance in the estimations of the regression coefficients

Author for correspondence** ***Received *January 28, 2021* ***Revised*** *May 31, 2021* ***Accepted*** *July 14, 2021*

by reducing the estimations towards zero. The beginning of the use of penalty methods with quantile regression was to develop a quantile regression method with L1-regularization in order to reduce the individual effects of common values, and there are many who have combined Quantile regression with the penalized methods [Hashem *et al*. (2016), Al-kenani and Malik (2019)]. As mentioned above when the number of explanatory variables (P) is large, greater than the number of observations (n) *i.e.* ($P > n$), we will face a problem in regression analysis, and in order to get rid of this problem, we have to reduce the dimensions of (P) the vector of explanatory variables (X) without losing the regression information and without re-diagnosing the model or distribution error, we can achieve this through the Sufficient Dimension Reduction (SDR) method suggested by Cook (1998), assuming the response variable (Y), $x=\left(x_1,\ldots,x_P\right)^T$ is $1\times P$, the vector of predictive variables, and the reduction of sufficient dimension (SDR) transforms the matrix $B\left(d\times P\right)$ where $Y \coprod X / X^T B$ the symbol $(\coprod)$ indicates the independence, Dimension reduction subspace (DRS) is the column space with extension (B). The dimensional intersections of the dimension reduction subspaces (DRS) are denoted by the symbol $S_{y/x}$ where $S_{y/x}$ contains all the regression information for (y / x). A number of methods have been proposed to obtain $S_{(y/x)}$ including SIR, SAVE as well as PHD method, whereas if the mean function was of interest [Cook and Li (2002)], introduce the Central Mean Subspace (CMS) $S_{E(y/x)}$ and for the sake of estimating $S_{E(y/x)}$ (CMS), a number of methods have been proposed, including the iterative Hessian transformation(IHD) [Cook and Li (2002)], as well as the MAVE method [Xia *et al.* (2002)]. As previously mentioned, SDR provides a method for finding sufficient dimension without the need to re-diagnose the model or distribution error. These methods give us linear combinations of the original variables with less dimensions, and here we have a problem with interpreting the results, and to solve this problem we combine the qunatile regression method with the penalized method with an effective and efficient method of sufficient dimension reduction (SDR), which is a method of estimating the minimum average variance estimator (MAVE) to obtain accurate and dispersed solutions. The parts of this paper are as follows. The first part is the introduction, the second part is review of the MAVE method and our proposed method GLQMAVE. In the third part we will review the algorithm of the GLQMAVE method. The fourth part is the practical part, and simulation studies are carried out. In the fifth part the conclusion is given.

2. MAVE and GLQMAVE

In this section we will highlight the MAVE method and our proposed GLQMAVE method. Suppose we have the following model:

$$Y=f\left(X_1,X_2,..,X_P\right)+\varepsilon, \tag{1}$$

where, $f\left(X_1,X_2,..,X_P\right)=E\left(Y|X\right)$, $E\left(Y|X\right)=0$, Y is the response variable , X is $\left(P\times 1\right)$ Vector of predictive variables and ε is the error term. The sufficient dimension reduction (SDR) seeks to find a subspace S such that

$$Y \coprod E\left(Y\mid X\right)\mid P_s X, \tag{2}$$

$\coprod$ where this symbol gives an indication of independence and $P(.)$ it is mean of projection operator when subspaces achieve constraint (2) are called mean dimension reduction subspaces [Cook and Li (2002)] . So, if $d=dim\left(s\right)$ and $B=\left(\beta_1,\beta_2,\ldots,\beta_d\right)$ is a basis for S then we can replace the predictors X by linear combinations $\beta_1^T X,\beta_2^T X,\ldots,\beta_d^T X, d\le p$ and without losing information of the conditional mean function $E\left(Y\mid X\right)$ [Cook and Li (2002)] explain that the central mean subspace is the intersections of all subspaces that satisfy condition (2). Explanation of the details of the MAVE method for estimating $S_{E(Y|X)}$ (CMS) was proposed by Xia *et al.* (2009) as follows.

Let we have orthogonal matrix

$B_{E(P\times d)}$

$B=\left(\beta_1,\beta_2,\ldots,\beta_d\right)$ is a solution to

$$\min_B \left\{E\left[Y-E\left(Y\mid X^T B\right)\right]^2\right\} \tag{3}$$

Table 1: Comparison between GLQMAVE and LQMAVE based on Mean Squared Error (MSE) criterion of Example4.1.

Model 1		
	MSE (Lasso)	**MSE(Group Lasso)**
$\tau = 0.25$		
$\sigma = 3$	0.0119989	0.0118216
$\sigma = 6$	0.0120073	0.0119839
$\tau = 0.50$		
$\sigma = 3$	0.0116006	0.0106343
$\sigma = 6$	0.0120004	0.0116096
$\tau = 0.75$		
$\sigma = 3$	0.0116122	0.0109507
$\sigma = 6$	0.0120004	0.0116096

Table 2: Comparison between GLQMAVE and LQMAVE based on Mean Squared Error (MSE) criterion of Example4.1.

Model 2		
	MSE (Lasso)	**MSE(Group Lasso)**
$\tau = 0.25$		
$\sigma = 3$	0.0155101	0.0128913
$\sigma = 6$	0.0157434	0.0137147
$\tau = 0.50$		
$\sigma = 3$	0.0117469	0.0105042
$\sigma = 6$	0.0120004	0.0116096
$\tau = 0.75$		
$\sigma = 3$	0.0150804	0.0117484
$\sigma = 6$	0.0154182	0.0118903

where,

$B^T B = I_d$ is shorthand condition

and the conditional variance $X^T B$ is

$$\sigma_B^2\left(X^T B\right) = E\left[\left\{Y - E\left(Y \mid X^T B\right)\right\}\right]^2 / X^T B \quad (4)$$

So,

$$\underset{B}{min}\, E\left[Y - E\left(Y \mid X^T B\right)\right]^2 = \underset{B}{min}\, E\left\{\sigma^2{}_B\left(X^T B\right)\right\}, \quad (5)$$

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

$$\sigma_B^2\left(X_0 B\right) \approx \sum_{i=1}^{n}\left\{Y_i - E\left(Y_i \mid X_i^T B\right)\right\}^2 W_{i0}$$

$$\approx \sum_{i=1}^{n}\left[Y_i - \left\{a_0 + \left(X_i - X_0\right)^T Bb_0\right\}\right]^2 W_{i0}$$

and surely , W_{i0} is a function that measures the distance between X_i and X_0. W_{i0} are the Kernel weights centered at $X_0^T B$ with $\sum_{i=1}^{n} W_{i0} = 1$. So the problem of finding the matrix $B_{P\times d}$ is equivalent to that of solving the following optimization:

$$\underset{B = B^T B = I}{min}\left(\sum_{j=1}^{n}\sum_{i=1}^{n}\left[Y_i - \left\{a_j + \left(X_i - X_j\right)^T Bb_j\right\}\right]^2 W_{ij}\right) \quad (6)$$

Hashem *et al.* (2016) suggested QMAVE and was combined with Lasso penalty function, as below:

$$\sum_{j=1}^{n}\sum_{i=1}^{n}\rho_\tau\left[Y_i - \left\{a_j + \left(X_i - X_j\right)^T Bb_j\right\}\right]$$

$$w_{ij} + \lambda\sum_{k=1}^{p}\left|\beta_k\right| \quad (7)$$

$m = 1,\ldots,d$ where the dimension d is known and estimated by modified BIC.

Tibshirani (1996) presented the Lasso technique, which works simultaneously to estimate parameters and select variables. Often we observe categorical of explanatory variables in the high-dimensional data, and that Lasso imposes a penalty on each variable, which in doing so strengthens individual discrimination, while it prefers to structure the predictive variables collectively, that is, collective sparse is preferred over individual sparse, in this case Lasso fails to deal with this data.

Yuan and Lin (2006) introduced Group Lasso by generalizing the penalty function Lasso, when the predictive variables are grouped together, the selection of the variable on the group level becomes necessary, *i.e.* necessary, while Lasso ignores the group structure, so Lasso is not suitable for the group structure. Among those who paid attention to group structure and variable selection at the group level [Yuan and Lin (2006), Meier *et al.* (2008)]. One of the most important issues addressed by the Group Lasso method is in the medical field, for example, genetic data and genetic engineering, where these data are always in the form of groups according to the common characteristic and thus form many different genetic paths. Later on, several studies

Table 3: Comparison between GLQMAVE and LQMAVE based on Mean Squared Error (MSE) criterion of Example 4.1.

Model 3		
	MSE (Lasso)	**MSE(Group Lasso)**
$\tau = 0.25$		
$\sigma = 3$	0.0211262	0.0204661
$\sigma = 6$	0.0237273	0.0205179
$\tau = 0.50$		
$\sigma = 3$	0.0188405	0.0186187
$\sigma = 6$	0.0189902	0.0187383
$\tau = 0.75$		
$\sigma = 3$	0.0189316	0.0187486
$\sigma = 6$	0.0190208	0.0188349

Table 4: Comparison between GLQMAVE and LQMAVE based on the number of zero coefficient (Av0,s) of Example 4.1.

Model 1		
	(Av0,s) (Lasso)	**(Av0,s) (Group Lasso)**
$\tau = 0.25$		
$\sigma = 3$	8.33	12.67
$\sigma = 6$	7.00	12.50
$\tau = 0.50$		
$\sigma = 3$	10.00	17.00
$\sigma = 6$	7.00	13.00
$\tau = 0.75$		
$\sigma = 3$	13.00	16.23
$\sigma = 6$	9.00	11.00

were presented on Group Lasso. Subsequently, several studies were presented on Group Lasso so several dirctions appeared among them Sparse GLasso, hierarchical Lasso and standard GLasso.

In this paper, and for the reason which mentioned above our suggestion is to combine the group lasso penalty function with the QMAVE estimation method, so we get a new method

$$\sum_{j=1}^{n}\sum_{i=1}^{n}\rho_{\tau}\left[Y_i-\left\{a_j+\left(X_i-X_j\right)^{T}Bb_j\right\}\right]$$

$$w_{ij}+\lambda\sum_{g=1}^{G}\|\beta_g\|_{S_g} \tag{8}$$

3. Algorithm of GLQMAVE

In this section, the GLQMAVE method is proposed in order to obtain sufficient dimension reduction (SDR) under quantile regression settings to reach a simple

Table 5: Comparison between GLQMAVE and LQMAVE based on the number of zero coefficient (Av0,s) of Example 4.1.

Model 2		
	(Av0,s) (Lasso)	**(Av0,s) (Group Lasso)**
$\tau = 0.25$		
$\sigma = 3$	9.00	13.50
$\sigma = 6$	7.00	13.00
$\tau = 0.50$		
$\sigma = 3$	11.54	12.75
$\sigma = 6$	9.00	12.00
$\tau = 0.75$		
$\sigma = 3$	11.00	13.53
$\sigma = 6$	4.00	13.50

Table 6: Comparison between GLQMAVE and LQMAVE based on the number of zero coefficient (Av0,s) of Example 4.1.

Model 3		
	(Av0,s) (Lasso)	**(Av0,s) (Group Lasso)**
$\tau = 0.25$		
$\sigma = 3$	5.50	13.00
$\sigma = 6$	9.00	13.50
$\tau = 0.50$		
$\sigma = 3$	10.00	13.50
$\sigma = 6$	12.00	13.00
$\tau = 0.75$		
$\sigma = 3$	10.00	13.50
$\sigma = 6$	8.50	12.00

interpretation of the resulting estimators. The following algorithm is suggested for GLQMAVE:

1. Let $m=1$ and $B=\beta_0$ *i.e.* an arbitrary vector $P\times 1$.
2. B is a known vector, we find the solution vector $\left(a_j, b_j\right)$ where j = 1, ..., n in the following

$$\min_{a_j, b_{j=1,\dots,n}} \sum_{j=1}^{n}\sum_{i=1}^{n}\rho_{\tau}\left[Y_i-\left\{a_j+\left(X_i\right.\right.\right.$$

$$\left.\left.\left.-X_j\right)^{T}b^{T}{}_{j}B\right\}\right]W_{ij} \tag{9}$$

3. The estimated solution vector $(\hat{a}_j, \hat{b}_j)$, j = 1, ..., n, we find the solution of $\beta_{GL\tau m}$ from the following

Table 7: Comparison between GLQMAVE and LQMAVE based on Mean Squared Error (MSE) criterion for Example 4.2 when the sample size is n = 60.

Model (1)		
***n* = 60**	**MSE (Lasso)**	**MSE(Group Lasso)**
$\tau = 0.25$		
MSE	22.05157	22.05034
$\tau = 0.50$		
MSE	26.9286	26.64045
$\tau = 0.75$		
MSE	25.75343	25.65207

Table 8: Comparison between GLQMAVE and LQMAVE based on mean squared error (MSE) criterion for Example 4.2 when the sample size is n=120.

Model 2		
***n* = 120**	**MSE (Lasso)**	**MSE(Group Lasso)**
$\tau = 0.25$		
MSE	0.1858737	0.1857578
$\tau = 0.50$		
MSE	0.2121364	0.2120953
$\tau = 0.75$		
MSE	0.1932994	0.1931527

$$\min_{B:B^T B=I} \sum_{j=1}^{n}\sum_{i=1}^{n} \rho_\tau \left[Y_i - \left\{ \hat{a}_j + \left(X_i - X_j \right) \hat{b}_j^T \left(\hat{\beta}_{1,} \hat{\beta}_2, \ldots, \hat{\beta}_{m-1}, \hat{\beta}_m \right)^T \right\} \right] w_{ij} + \lambda_n \sum_{g=1}^{G} \beta_g S_g \quad (10)$$

4. Now we put $\hat{\beta}_{GL\tau m}$ in the mth of the column in B, and continue with step 2 to 3 until we reach convergence.
5. We update B with $\hat{\beta}_{1GL,}\ \hat{\beta}_{2GL}, \ldots, \hat{\beta}_{GL\tau m}, \hat{\beta}_0$ and let m be equal to $m + 1$.
6. If $m < d$, repeat steps 2 through 5 until $m = d$

4. Simulation

On the other hand, we will clarify and show the efficiency of our proposed method GLQMAVE by comparing with the LQMAVE method [Hashem *et al.* (2016)] to show the efficiency of GLQMAVE and its ability to produce accurate and sparse solutions. We will take the following example:

Example 4.1: R = 100 sets of data are created of size n =100 observations of the model

$$y = \frac{\left(X^T \beta_1 \right)}{\left\{ 0.5 + \left(X^T \beta_2 + 1.5 \right)^2 \right\}} + 0.2\varepsilon$$

where, $X = \left(X_1, \ldots, X_{10} \right)^T$ X_i, and ε are independent and have the same distributed form of $N(0,1)$, with $S_{E(y/x)} = span\left(B_2 \right)$. This means, the model is

$$y = \frac{\left(X_1 \right)}{\left\{ 0.5 + \left(X_2 + 1.5 \right)^2 \right\}} + 0.2\varepsilon$$

and we have the three models as follows.

Model 1: $\beta_{1=(1,0,0,1,0,0,0,0,0,0)}$, $\beta_{2=(0,0,0,0,1,0,0,1,0,0)}$

Model 2: $\beta_{1=(1,1,1,2,2,2,0,0,0,0)}$, $\beta_{2=(0,0,0,0,2,2,2,1,1,1)}$

Model 3: $\beta_{1=(1,1,1,-2,-2,-2,0,0,0,0)}$, $\beta_{2=(0,0,0,0,2,2,2,-1,-1,-1)}$

The simulation was done for the number of iterations (100) and considering three groups with variables related to each other and assuming $\sigma = 3, \sigma = 6$.

Example 4.2: R = 100 datasets were generated from linear model $y = X^T \beta + \varepsilon$ with sample size (60 , 120) , we have $X = \left(X_1, \ldots, X_{24} \right)^T$ where X_i are independent and identically distributed from normal distribution N(0,1) $\beta_{=(1,1,1,2,2,2,0,\ldots,0)^T}$ with $S_{E(y/x)}$ $= span\left(B_1 \right)$. So the model is

$$y = X_1 + X_2 + X_3 + 2X_4 + 2X_5 + 2X_6 + \varepsilon,$$

Correlation for the first group variables is 0.95 and for the variables of the second group, the correlation value is 0.90.

We have relied on two criteria to interpret the results, namely the mean squares error (MSE) and the number of zeroed coefficients (Ave0,s). The MSE Tables 1, 2, 3, 7, 8 have been organized for both methods GLQMAVE and LQMAVE in order to compare which two methods give us accurate predictive solutions. While the Tables 4, 5, 6, 9, 10 (Ave0,s), which represent the number of coefficients that are zeroed for the two methods in order to show which of the two methods gives us more sparse solutions. The process of obtaining a single result took 5 days, as more than one high-

Table 9: Comparison between GLQMAVE and LQMAVE based on the criterion of number of zeroed coeffcients (Av0,s) for Example 4.2 when the sample size in n = 60.

Model 1		
***n* = 60**	**(Av0,s) (Lasso)**	**(Av0,s)(Group Lasso)**
$\tau = 0.25$		
MSE	0.1858737	0.1857578
$\tau = 0.50$		
MSE	0.2121364	0.2120953
$\tau = 0.75$		
MSE	0.1932994	0.1931527

Table 10: Comparison between GLQMAVE and LQMAVE based on the criterion of number of zeroed coeffcients (Av0,s) for Example 4.2, when the sample size in n = 120.

Model 2		
***n* = 120**	**(Av0,s) (Lasso)**	**(Av0,s)(Group Lasso)**
$\tau = 0.25$		
MSE	20.01	23.25
$\tau = 0.50$		
MSE	19.02	21.17
$\tau = 0.75$		
MSE	19.11	21.32

efficiency computer was used to shorten the time to carry out more than one operation simultaneously.

Interpretation of the simulation results will be based on two criteria to assess the accuracy of the estimate. By Tables 1, 2 and 3 for the Example 4.1, and Tables 7 and 8 for Example 4.2 we can summarize the result. It is clear that the performance of the LQMAVE method is less efficient than the proposed method GLQMAVE. We can deduce this clearly for all Models 1, 2 and 3 and for all quantile regression levels and at different σ values for Example 4.1, and for models 1 and 2, for Example 4.2 at the number of observations (n = 60, 120) and for all levels of quantile regression, where we notice that our proposed method (GLQMAVE) has given (MSE) less than (LQMAVE) method, and this seems clear from Tables 1, 2 and 3, for the Example 4.1, and Tables 7 and 8 for Example 4.2.. As for the second comparison criterion (Av0,s), we notice that the (GLQMAVE) method has given more sparse coefficients than the (LQMAVE) method, and this seems clear from Tables 4,5 and 6, for the Example 4.1, and Tables 9 and 10 for Example 4.2.

5. Conclusion

Through the results obtained, it is clear that the proposed method GLQMAVE is the best in obtaining sparse and accurate solutions. The proposed method is a combination of Group Lasso (GL), Quantile regression and MAVE method. Since GL is one of the penalized methods that encourages the selection of variables collectively, where in some cases the predictive variables have a structure that encourages group selection, as is the case with categorical data. As for QR, quantile regression provides us with a clearer and more comprehensive picture of the conditional distribution (y/x), as for MAVE, it is one of the efficient ways to find (SDR) where it estimates (CMS) (Central mean subspace). GLQMAVE has been proven to be a good and efficient method for getting accurate and dispersed results.

References

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

Cook, R.D. and B. Li (2002). Dimension reduction for conditional mean in regression. *The Annals of Statistics*, **30(2)**, 455-474.

Dikheel, T.R. and R. Abdalriadha (2020). Singular spectrum analysis to analyze coronavirus confirmed cases data in some countries. *International Journal of Agricultural and Statistical Sciences*, **16(Supplement 1)**, 1401-1419.

Donoho, D.L. and J.M. Johnstone (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, **81(3)**, 425-455.

Hashem, H., V. Vinciotti, R. Alhamzawi and K. Yu (2016). Quantile regression with group lasso for classification. *Advances in Data Analysis and Classification*, **10(3)**, 375-390.

Koenker, R. and Jr. G. Bassett (1978). Regression quantiles. *Econometrica*, **46(1)**, 33-50.

Meier, L., S. Van De Geer and P. Bühlmann (2008). The group lasso for logistic regression. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **70(1)**, 53-71.

Saini, M. and A. Kumar (2019). Improvement over traditional methods of estimation in stratified random sampling. *International Journal of Agricultural and Statistical Sciences*, **15(2)**, 643-648.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, **58(1)**, 267-288.

Xia, Y., H. Tong, W.K. Li and L.X. Zhu (2009). An adaptive estimation of dimension reduction space. In Exploration of a Nonlinear World: An Appreciation of Howell Tong's Contributions to Statistics, World Scientific, 299-346.

Yao, W. and Q. Wang (2013). Robust variable selection through MAVE. *Computational Statistics and Data Analysis*, **63**, 42-49.

Yuan, M. and Y. Lin (2006). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B Statistical Methodology*, **68(1)**, 49-67.