A numerical study on group quantile regression models

Doyoen Kim[a], Yoonsuh Jung[1,a]

[a]Department of Statistics, Korea University, Korea

Abstract

Grouping structures in covariates are often ignored in regression models. Recent statistical developments considering grouping structure shows clear advantages; however, reflecting the grouping structure on the quantile regression model has been relatively rare in the literature. Treating the grouping structure is usually conducted by employing a group penalty. In this work, we explore the idea of group penalty to the quantile regression models. The grouping structure is assumed to be known, which is commonly true for some cases. For example, group of dummy variables transformed from one categorical variable can be regarded as one group of covariates. We examine the group quantile regression models via two real data analyses and simulation studies that reveal the beneficial performance of group quantile regression models to the non-group version methods if there exists grouping structures among variables.

Keywords: group penalty, penalized quantile regression, variable selection

1. Introduction

In some regression problems we are interested in various percentage points of the distribution rather than the overall average of the data. For these cases, Koenker and Bassett (1978) proposed a quantile regression model that estimates the conditional quantile function of the response variable and identifies the effect for each quantile separately. With the advantages of quantile regression, its usage is clearly increasing in various fields such as economics (Hendricks and Koenker, 1992; Koenker and Hallock, 2001), microarray studies (Wang and He, 2007), demand analysis and empirical finance area (Wei and He, 2006; Wei *et al.*, 2006).

We focus on variable selection aspect of penalized quantile regression because variable selection is an important procedure in the modeling as many datasets include a large number of candidate predictors. Several methods were introduced under the regularization framework. Tibshirani R (1996) proposed the least absolute shrinkage and selection operator (LASSO) using l_1-norm and adaptive LASSO and its oracle properties were shown by Zou (2006). Fan (1997) then proposed a smoothly clipped absolute deviation (SCAD) penalty function and Fan and Li (2001) demonstrated its oracle properties under a penalized likelihood. Later, Wu and Liu (2009) extended the oracle properties of the SCAD penalty to the penalized quantile regression. After that Zhang (2010) proposed the minimax concave penalty (MCP) and showed its oracle properties. However, Hoerl and Kennard (1970) proposed ridge regression that employs l_2 type penalty. Zou and Hastie (2005) introduced an elastic net that combines l_1 penalty of LASSO and l_2 penalty of ridge.

For the case of variables showing grouping structure, it seems more appropriate to reflect a grouping structure for an accurate estimation and prediction. The most common example is the case when a

categorical variable is converted to several dummy variables. A group of dummy variables from one original categorical variable can be thought as the same group. Another example of the grouped variables is an additive model with polynomial variables. Terms forming a polynomial such as x_1, x_1^2, x_1^3 can be regarded as composing one group. Grouping can also be applied to the model which imposes prior scientific information. For example, genes having same biological pathway can be classified into the same group in genetic data analysis. When we analyze these type of data with group structure among the variables, we can apply certain penalization methods with group penalties that reflect the grouping information.

Group penalty can be applied when the grouping structure is unknown (thus must be estimated). In biological studies, genetic data usually have background scientific information. For example, genes with the same biological pathway are often located in a neighborhood forming a group. We need to estimate grouping structure of predictors by certain clustering methods for variables such as Chavent and Kuentz-Simonet (2012).

Several penalties considering grouping structure have been proposed. Group LASSO which uses the l_2-norm of the coefficients within a group was proposed by Bakin (1999) and extended by Yuan and Lin (2006). Huang *et al.* (2012) then showed group SCAD and group MCP for covariates possessing a grouping structure to select important groups. In the context of the quantile regression models, Ciuperca (2019) proposed adaptive group LASSO that adaptive LASSO penalty and established the sparsity and asymptotic normality of their methods. Kato (2011) considered high dimensional sparse quantile regression models with group LASSO penalty and attained non-asymptotic error bound of the estimation error. However, Group LASSO penalty was investigated for the classification problem by Hashem *et al.* (2016).

The later part of this paper is organized as follows. Penalized quantile regression models with a group penalty are introduced in Section 2. A real data set is analyzed in Section 3. In Section 4, we compare the prediction performance of introduced methods under various simulation settings. The conclusion and discussions are given in Section 5.

2. Group quantile regression models

We consider the linear model with p predictors:

$$\mathbf{Y} = \mathbf{X}\beta_\tau + \epsilon,$$

where $\mathbf{Y} = (y_1, \ldots, y_n)^T$ is an $n \times 1$ vector of response variable, $\mathbf{X} = (x_{ij})$, $i = 1, \ldots, n$, $j = 1, \ldots, p$ is the design matrix of predictors and ϵ is a vector of independent random errors with mean 0. Then the τth conditional quantile function can be estimated by solving

$$\underset{\beta_\tau \in \mathbb{R}^p}{\operatorname{argmin}} \sum_{i=1}^{n} \rho_\tau \left(y_i - x_i^T \beta_\tau \right), \tag{2.1}$$

where x_i^T is the ith row of $\mathbf{X}$, $\beta_\tau = (\beta_{\tau,1}, \ldots, \beta_{\tau,p})^T$ and $\rho(u) = \tau u - I(u < 0)u$ is the check loss function. To produce sparse solution, Koenker *et al.* (1994) and Koenker (2004) suggested penalized version of (2.1)

$$\underset{\beta_\tau \in \mathbb{R}^p}{\operatorname{argmin}} \sum_{i=1}^{n} \rho_\tau \left(y_i - x_i^T \beta_\tau \right) + \lambda J \left(\beta_\tau \right), \tag{2.2}$$

where $\lambda \geq 0$ is the tuning parameter and $J(\beta_\tau)$ denotes the penalty.
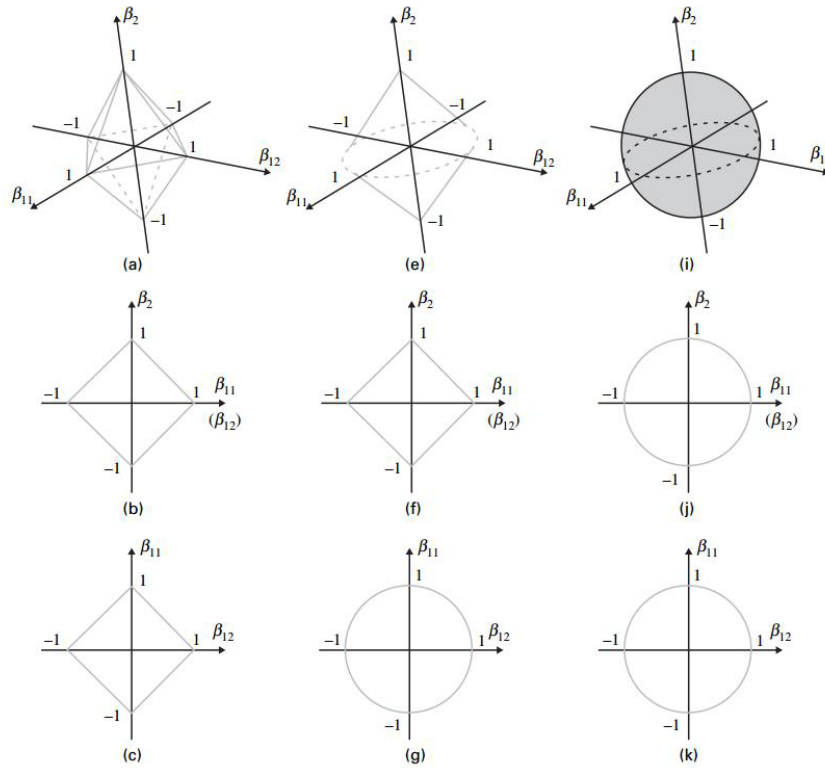
Figure 1: l_1 penalty, Group LASSO penalty, l_2 penalty from Yuan and Lin (2006).

Many penalties under mean regression exist for reflecting grouping information. In this study, we propose group penalties under the quantile regression models. Group LASSO penalty, group SCAD penalty, and group MCP under the quantile regression models will be newly defined.

First, group LASSO proposed by Yuan and Lin (2006) uses l_2-norm of the coefficients within a group. So it is regarded as a mixture of l_1 penalty and l_2 penalty, but a bit different from the elastic net penalty by Zou and Hastie (2005). Figure 1 illustrates the concept that the coefficients composing the group only by itself are affected by the l_1 penalty. Suppose there are two groups with three variables : $\boldsymbol{\beta}_1 = (\beta_{11}, \beta_{12})^T$ and a scalar β_2. Figure 1(a), (e), (i) show the contours of the penalties. Figure 1(a) refers to $|\beta_{11}| + |\beta_{12}| + |\beta_2| = 1$, Figure 1(g) refers to $\beta_{11}^2 + \beta_{12}^2 = 1$. In Figure 1(g), group LASSO selects both β_{11} and β_{12} at the same time. Consequently, in Figure 1(e), group LASSO then shows sparsity at the group level. From this process, the group LASSO selects only groups of variables, while not selecting variables within groups individually. Suppose that the predictor variables were divided into G different groups, then group LASSO penalty under quantile regression is defined as

$$\hat{\beta}(\text{gLASSO}) = \underset{\beta_\tau \in \mathbb{R}^p}{\text{argmin}} \sum_{i=1}^{n} \rho_\tau \left(y_i - x_{i,g}^T \beta_{\tau,g} \right) + \lambda \sum_{g=1}^{G} \sqrt{K_g} \|\beta_{\tau,g}\|_2, \qquad (2.3)$$

where $x_{i,g}$ is the design matrix for the ith sample where its columns correspond to the predictors in group g and $\beta_{\tau,g}$ is coefficient vector of group g and K_g is the number of covariates in group g. Here, K_g is used to adjust the size of groups.

Besides LASSO, other penalty functions can be applied. By applying SCAD penalty and MCP under quantile regression, they are defined as

$$\hat{\beta}(\text{gSCAD}) = \underset{\beta_\tau \in \mathbb{R}^p}{\text{argmin}} \sum_{i=1}^{n} \rho_\tau \left(y_i - x_{i,g}^T \beta_{\tau,g} \right) + \sum_{g=1}^{G} p_{1\lambda,\gamma}(\|\beta_{\tau,g}\|_2), \tag{2.4}$$

$$p_{1\lambda,\gamma}(\beta) = \begin{cases} \lambda|\beta|, & |\beta| \le \lambda, \\ \dfrac{2\gamma\lambda|\beta| - \beta^2 - \lambda^2}{2(\gamma - 1)}, & \lambda < |\beta| \le \gamma\lambda, \\ \dfrac{\lambda^2(\gamma + 1)}{2}, & |\beta| > \gamma\lambda. \end{cases}$$

$$\hat{\beta}(\text{gMCP}) = \underset{\beta_\tau \in \mathbb{R}^p}{\text{argmin}} \sum_{i=1}^{n} \rho_\tau \left(y_i - x_{i,g}^T \beta_{\tau,g} \right) + \sum_{g=1}^{G} p_{2\lambda,\gamma}(\|\beta_{\tau,g}\|_2), \tag{2.5}$$

$$p_{2\lambda,\gamma}(\beta) = \begin{cases} \lambda|\beta| - \dfrac{\beta^2}{2\gamma}, & |\beta| \le \gamma\lambda, \\ \dfrac{1}{2}\gamma\lambda^2, & |\beta| > \gamma\lambda. \end{cases}$$

From the (2.4) and (2.5), penalty is preferentially applied to each variables in g groups by the group unit. The summation over g is then substituted so that we can decide which groups of variables to be selected. The form of SCAD and MCP penalties are similar; however, the group SCAD is reported to show less grouping than group MCP (Ogutu and Piepho, 2014). The implementation of finding the solutions for the group LASSO, SCAD, MCP is done via the R package `rqPen`. `rqPen` produces estimates using the `QICD` command.

The theoretical properties of the group SCAD and group MCP are rarely known under the conditional mean regression despite the well-known oracle properties of the SCAD and MCP under the linear models. Additional complexity such as the true number of group and whether the number of group is fixed or increasing seems to hinder the exploration of the theoretical grounding. Unfortunately, we do not provide some theoretical properties under the quantile regression with these penalties.

3. Application to real data sets

3.1. Birth weight data

The birth weight data set from Hosmer and Lemeshow (1989) is used for the real data analysis. The data were collected at Baystate Medical Center in Springfield, Massachusetts. The data has the *birthweight*: birth weights of $n = 188$ babies and eight predictors about the mother, three of them are continuous variables and the rest are categorical variables. The three continuous variables are

- age: age of mother (years)

- weight: weight of mother at last menstrual period (pounds)

- visit: number of physician visits during the first trimester (0, 1, 2, or 3 or more).

The five categorical variables are

- race: race of mother (white, black, or the others)

Table 1: Birth weight data: Coefficients of group LASSO, LASSO, group SCAD, and SCAD penalized median regression from arbitrary one fit

Variable	LASSO(G)	LASSO(NG)	SCAD(G)	SCAD(NG)
intercept	3798.4927	2207.0331	2243.9668	2591.5665
age	−145.8383	0.0000	−3.2061	0.0000
age^2	3.0818	0.2200	0.2841	0.1924
weight	6.0441	4.9070	4.9070	2.1216
smoking	0.0000	0.0000	0.0000	0.0000
hyper	0.0000	0.0000	0.0000	0.0000
uterine	0.0000	0.0000	0.0000	0.0000
visit	0.0000	0.0000	0.0000	0.0000
$race_1$	0.0000	0.0000	0.0000	0.0000
$race_2$	0.0000	0.0000	0.0000	0.0000
pre_1	0.0000	0.0000	0.0000	0.0000
pre_2	0.0000	0.0000	0.0000	0.0000

LASSO = least absolute shrinkage and selection operator; SCAD = smoothly clipped absolute deviation.

- smoking: smoking status during pregnancy (yes or no)

- pre: history of premature labor (0, 1, 2, or 3)

- hyper: history of hypertension (yes or no)

- uterine: presence of uterine irritability (yes or no).

The dummy variables from each categorical variable are treated as one group for the group quantile regression model. $Race_1$ and $race_2$ are created for white and others respectively so that black is baseline category for race. For pre, pre_1 stands for history of premature labor once and pre_2 stands for two or more than two times of premature labors. After the preliminary analysis, age^2 is added to the model. Taking birthweight as a response variable, we fit quantile regression models at $\tau = 0.5$ to find significant variables. Our model is,

$$E(\text{birthweight}) = \beta_0 + \beta_1 \text{age} + \beta_2 \text{age}^2 + \beta_3 \text{weight} + \beta_4 \text{smoking} + \beta_5 \text{hyper}$$
$$+ \beta_6 \text{uterine} + \beta_7 \text{visit} + \beta_8 \text{race}_1 + \beta_9 \text{race}_2 + \beta_{10} \text{pre}_1 + \beta_{11} \text{pre}_2.$$

When a full model is fitted with LASSO, SCAD, group LASSO, group SCAD, we have the results for Table 1. From the results, group penalties select both age and age^2 simultaneously whereas LASSO and SCAD only choose age.

Simulation studies are conducted to see which methods yield a better estimate. First, we randomly divide the data into two parts. The first part is training data with 100 samples, and the other 88 samples are used for test data. The penalty parameter λ is chosen from the training data by employing a 10-fold cross validation. After selecting λ value, we fit LASSO penalized quantile regression, SCAD penalized quantile regression, group LASSO penalized quantile regression, and group SCAD penalized quantile regression to the training data. We then obtain the estimates for 12 coefficients and predict 88 values of response variable in the test data. Finally, predicted mean check error (PMCE) from 88 predicted values is calculated to gauge the performance. PMCE is therefore obtained by calculating the mean check error only with the test data. PMCE is defined as follow with the observed y_i and predicted $\hat{y}_i$ over the test data with $n_{\text{test}} = 88$:

$$\text{PMCE} = \frac{1}{n_{\text{test}}} \sum_{i=1}^{n_{\text{test}}} \{\tau |y_i - \hat{y}_i| I(y_i > \hat{y}_i) + (1 - \tau)|y_i - \hat{y}_i| I(y_i \le \hat{y}_i)\}. \tag{3.1}$$

Table 2: Birthweight data: Mean of PMCE, pure PMCE (and its standard error in the parentheses) for group LASSO, group SCAD, LASSO, and SCAD penalized median regression and mean regression from 500 different splits of data

	Median LASSO(G)	Median LASSO(NG)	Median SCAD(G)	Median SCAD(NG)
PMCE	292.74	297.71	296.16	298.39
Pure PMCE	12.40	17.37	15.82	18.05
	(0.86)	(0.78)	(0.81)	(0.78)
	Mean LASSO(G)	Mean LASSO(NG)	Mean SCAD(G)	Mean SCAD(NG)
PMCE	277.07	277.22	280.60	279.83
	(0.79)	(0.79)	(0.82)	(0.82)

PMCE = predicted mean check error; LASSO = least absolute shrinkage and selection operator; SCAD = smoothly clipped absolute deviation.

We repeat this procedure 500 times to reduce the variation from splitting the data. Table 2 then summarizes the mean of PMCE and its standard error.

From the results, the group version of LASSO and SCAD show better performance that the non-group version of penalties for median regression.

Noting that there is a certain lower bound for PMCE, we observe that PMCE cannot be smaller than the mean check error using the whole data in theory. This is because PMCE is calculated from the test data only. For this reason, the mean check error using all samples is obtained and subtracted from PMCE. Subtracting the mean check error is similar to removing the variance of the error σ^2 in the model as the mean check error is often used as $\hat{\sigma}^2$. It seems reasonable to use pure PMCE as we often remove σ^2 in simulation studies for calculating PMCE. Now, this pure PMCE shows a 12.4% to 28.6% reduction by the group penalty methods. Therefore, we can clearly see some benefits of the proposed methods.

3.2. Parkinson speech data

This data set contains measurement of voice recordings from equal number of healthy people and people with Parkinsonism who suffer from speech impairments. The covariates are $n = 1040$ measurements from $p = 26$ types of sound recordings. The time-frequency based 26 covariates can be categorized into 6 groups where the groups are known. See Sakar *et al.* (2013) for the details of the data set. The response variable is the united Parkinson's disease rating scale (UPDRS) having values from 1 (normal) to 55 (most severe). As the half of the measurements are from healthy people, about half of the response variable is equal to 1. For this reason, we do not fit the quantile regression with quantiles less than 0.5, but $\tau = 0.5, \ldots, 0.9$ are examined.

We compare the LASSO and SCAD penalized quantile regression models and their counterparts with group penalties. For this purpose, the 1,040 samples are randomly split into 2/3 of training data and other test data. Using the training data, we fit the considered models where the penalty parameters were sought by a 10-fold cross-validation. After choosing an 'optimal' penalty parameter, we fit the training data and then predict the values of the response variable in the test data. The prediction error is measured with the check error defined in (3.1). We repeat this procedure 100 times to reduce the variation that arises from the random split. Table 3 provides the mean of 100 PMCE value at each value of τ and shows a smaller PMCE from the group penalties. The methods with group penalties show the most reduction in the PMCE at $\tau = 0.7$ and very slight increase when $\tau = 0.9$. In general, we see the improvement by incorporating the group penalties.

Table 3: Parkinson speech data: Mean of PMCE (and its standard error in the parentheses) for LASSO, group LASSO, SCAD, and group SCAD penalized regression at various quantiles from 100 different splits of data. Percentage of PMCE reduction by incorporating the group penalties are shown as % reduction

τ	0.5	0.6	0.7	0.8	0.9
SCAD (NG)	6.00 (0.04)	6.71 (0.04)	6.45 (0.04)	5.46 (0.02)	3.36 (0.01)
SCAD(G)	5.85 (0.04)	6.57 (0.04)	5.79 (0.03)	5.03 (0.03)	3.39 (0.01)
% reduction	2.53	1.88	10.20	7.97	−1.00
LASSO (NG)	5.93 (0.07)	6.63 (0.08)	6.39 (0.08)	5.36 (0.08)	3.30 (0.05)
LASSO(G)	5.78 (0.06)	6.50 (0.08)	5.74 (0.07)	5.00 (0.08)	3.34 (0.05)
% reduction	2.42	1.95	10.25	7.56	−1.01

PMCE = predicted mean check error; LASSO = least absolute shrinkage and selection operator; SCAD = smoothly clipped absolute deviation.

Table 4: Simulation results for Model I: Mean of MSE, TP rate, TN rate for LASSO, SCAD, MCP are from 200 simulated data sets with $n = 100$

		LASSO(G)	LASSO(NG)	SCAD(G)	SCAD(NG)	MCP(G)	MCP(NG)
$\tau = 0.1$	MSE	873.18	983.28	1664.51	1423.33	1667.30	1430.00
		(37.2)	(30.2)	(62.2)	(39.1)	(59.3)	(37.9)
	TP rate	78.25	68.58	94.58	60.08	95.67	60.67
		(1.0)	(1.5)	(1.2)	(1.6)	(1.0)	(1.6)
	TN rate	42.00	39.82	64.71	48.93	64.39	48.68
		(1.8)	(1.5)	(2.4)	(1.6)	(2.5)	(1.5)
$\tau = 0.2$	MSE	403.28	899.76	549.90	1170.84	562.90	1211.79
		(15.4)	(26.1)	(19.3)	(26.6)	(20.2)	(28.3)
	TP rate	74.75	62.92	94.83	45.08	95.67	45.67
		(1.1)	(1.5)	(1.0)	(1.5)	(0.9)	(1.6)
	TN rate	54.43	46.75	72.96	67.86	71.68	67.79
		(1.8)	(1.6)	(2.2)	(1.6)	(2.2)	(1.7)
$\tau = 0.3$	MSE	299.36	862.30	305.80	1087.98	300.55	1106.97
		(8.5)	(24.1)	(12.2)	(23.7)	(11.7)	(25.9)
	TP rate	77.00	57.25	94.83	39.25	95.00	38.33
		(1.1)	(1.7)	(0.9)	(1.5)	(1.1)	(1.4)
	TN rate	55.64	53.71	76.32	75.36	76.79	76.64
		(2.0)	(1.7)	(2.1)	(1.6)	(2.2)	(1.5)
$\tau = 0.4$	MSE	254.80	833.40	217.07	1092.55	214.99	1082.13
		(7.7)	(25.0)	(8.6)	(27.3)	(8.4)	(25.4)
	TP rate	77.00	55.25	92.92	35.58	93.67	36.00
		(1.4)	(1.8)	(1.2)	(1.5)	(1.1)	(1.5)
	TN rate	58.86	57.82	81.00	81.21	80.96	81.68
		(2.0)	(1.9)	(1.9)	(1.5)	(2.0)	(1.5)
$\tau = 0.5$	MSE	243.28	821.00	209.41	1086.88	205.59	1084.43
		(7.6)	(26.7)	(8.4)	(26.7)	(8.6)	(26.1)
	TP rate	76.83	52.58	91.42	33.92	91.83	33.50
		(1.5)	(1.7)	(1.2)	(1.4)	(1.2)	(1.5)
	TN rate	60.29	60.96	77.11	82.07	77.79	82.39
		(2.1)	(1.9)	(2.1)	(1.3)	(2.1)	(1.4)

Standard errors are in parenthesis. All values are multiplied by 10^2.

LASSO = least absolute shrinkage and selection operator; SCAD = smoothly clipped absolute deviation; MCP = minimax concave penalty; MSE = mean squared error; TP = true positive; TN = true negative.

4. Simulation studies

In this section, we compare the prediction performance of LASSO, SCAD, MCP and grouped version of LASSO, SCAD, MCP under the four simulation scenarios. In the first scenario, we fit an additive

Table 5: Simulation results for Model II: Mean of MSE, TP rate, TN rate for LASSO, SCAD, MCP are from 200 simulated data sets with $n = 100$

		LASSO(G)	LASSO(NG)	SCAD(G)	SCAD(NG)	MCP(G)	MCP(NG)
	MSE	892.65	1361.60	991.70	1698.67	982.53	1706.56
		(11.1)	(29.5)	(19.9)	(38.8)	(19.2)	(39.8)
$\tau = 0.1$	TP rate	59.06	46.63	62.19	32.75	61.94	33.81
		(1.6)	(1.7)	(1.7)	(1.5)	(1.6)	(1.6)
	TN rate	47.35	48.06	84.18	65.12	85.59	64.53
		(1.7)	(1.4)	(1.3)	(1.5)	(1.2)	(1.5)
	MSE	818.28	1394.68	844.13	1645.89	836.67	1641.21
		(11.6)	(30.0)	(19.8)	(31.8)	(14.6)	(30.8)
$\tau = 0.2$	TP rate	62.06	53.44	73.63	35.38	74.19	33.94
		(1.4)	(1.8)	(1.2)	(1.7)	(1.2)	(1.6)
	TN rate	52.21	49.94	65.32	68.26	64.12	69.15
		(1.8)	(1.7)	(2.1)	(1.7)	(2.2)	(1.7)
	MSE	757.97	1364.15	720.11	1606.66	714.26	1614.34
		(14.2)	(27.2)	(14.6)	(28.3)	(14.1)	(27.4)
$\tau = 0.3$	TP rate	67.63	51.25	76.94	32.00	77.56	31.69
		(1.5)	(1.6)	(1.2)	(1.5)	(1.2)	(1.5)
	TN rate	50.44	51.71	65.06	70.68	64.06	71.18
		(1.8)	(1.6)	(2.1)	(1.6)	(2.1)	(1.6)
	MSE	672.45	1406.66	671.74	1627.18	678.25	1640.61
		(15.8)	(27.9)	(17.0)	(29.6)	(16.5)	(29.8)
$\tau = 0.4$	TP rate	71.56	52.31	79.19	32.81	79.19	31.56
		(1.4)	(1.7)	(1.2)	(1.5)	(1.1)	(1.5)
	TN rate	49.06	46.65	60.94	67.56	61.26	69.79
		(1.7)	(1.6)	(3.1)	(1.6)	(2.1)	(1.5)
	MSE	621.12	1414.99	590.80	1632.37	587.34	1633.92
		(14.2)	(29.6)	(17.1)	(28.9)	(17.3)	(27.9)
$\tau = 0.5$	TP rate	72.50	50.56	81.50	31.44	82.00	31.06
		(1.4)	(1.8)	(1.3)	(1.6)	(1.2)	(1.6)
	TN rate	48.50	47.24	61.03	67.68	60.97	67.65
		(1.6)	(1.5)	(2.0)	(1.6)	(2.0)	(1.6)

Standard errors are in parenthesis. All values are multiplied by 10^2.
LASSO = least absolute shrinkage and selection operator; SCAD = smoothly clipped absolute deviation; MCP = minimax concave penalty; MSE = mean squared error; TP = true positive; TN = true negative.

model of continuous variables through third-order polynomial. In the second scenario, we fit an additive model of both continuous and categorical variables also through a third-order polynomial. In the third we consider an additive model of only categorical variables. Finally, we fit an ANOVA model with all the two-way interactions. In all simulations, the errors ϵ are from the standard normal distribution. The penalty parameters are selected by 10-fold cross-validation where the check loss is used for both model construction and validation.

- Model I: Random variables $Z_1, \ldots, Z_{16}$ and W are independently generated from a standard normal distribution. The covariates are then defined as $X_i = (Z_i + W)/\sqrt{2}$. The model is:

$$Y = \frac{1}{2}X_3^3 + \frac{1}{2}X_3^2 + \frac{1}{2}X_3 + \frac{1}{6}X_6^3 - \frac{1}{2}X_6^2 + \frac{1}{3}X_6 + \epsilon.$$

- Model II: $X_1, \ldots, X_{20}$ are generated as in model I. Then $X_{11}, \ldots, X_{20}$ are first generated according to a centered multivariate normal distribution with a covariance between X_i and X_j being $0.5^{|i-j|}$. Then X_i is trichotomized as 0, 1, or 2 if it is smaller than $\Phi^{-1}(1/3)$, larger than $\Phi^{-1}(2/3)$ or in between.

Table 6: Simulation results for Model III. Mean of MSE, TP rate, TN rate for LASSO, SCAD, MCP are from 200 simulated data sets with $n = 100$

		LASSO(G)	LASSO(NG)	SCAD(G)	SCAD(NG)	MCP(G)	MCP(NG)
$\tau = 0.1$	MSE	282.81	818.73	394.63	825.53	398.62	825.67
		(8.9)	(6.2)	(11.5)	(6.6)	(11.6)	(6.7)
	TP rate	73.58	55.25	46.17	42.67	46.58	43.67
		(1.6)	(1.8)	(2.1)	(1.6)	(2.2)	(1.6)
	TN rate	36.83	52.42	59.54	65.71	58.54	64.42
		(1.5)	(1.8)	(2.4)	(1.6)	(2.4)	(1.6)
$\tau = 0.2$	MSE	202.82	772.30	202.72	768.74	198.41	770.85
		(6.8)	(5.9)	(8.2)	(5.3)	(7.9)	(5.5)
	TP rate	86.17	52.67	81.58	42.75	81.25	40.67
		(1.3)	(1.9)	(1.8)	(1.7)	(1.8)	(1.6)
	TN rate	31.04	57.75	43.71	69.13	45.08	70.42
		(1.7)	(1.9)	(2.2)	(1.8)	(2.2)	(1.7)
$\tau = 0.3$	MSE	160.78	747.47	168.09	740.63	168.95	739.28
		(7.0)	(7.2)	(9.4)	(7.5)	(9.6)	(7.3)
	TP rate	91.67	51.17	85.58	41.00	85.25	41.58
		(1.0)	(1.8)	(1.6)	(1.7)	(1.7)	(1.7)
	TN rate	28.92	57.17	44.67	68.25	43.75	68.42
		(1.6)	(1.9)	(2.1)	(1.8)	(2.0)	(1.8)
$\tau = 0.4$	MSE	139.23	726.39	158.78	714.23	151.52	711.59
		(5.9)	(6.8)	(9.7)	(6.6)	(9.1)	(6.2)
	TP rate	93.67	49.75	87.75	38.92	88.75	39.00
		(0.8)	(1.7)	(1.6)	(1.5)	(1.5)	(1.5)
	TN rate	27.96	54.88	41.33	68.17	41.33	69.00
		(1.6)	(1.7)	(2.2)	(1.6)	(2.1)	(1.5)
$\tau = 0.5$	MSE	138.13	719.82	153.13	715.73	157.20	714.26
		(6.3)	(8.1)	(9.6)	(8.5)	(10.0)	(9.0)
	TP rate	94.17	51.33	89.75	40.83	89.33	40.42
		(0.8)	(1.7)	(1.4)	(1.8)	(1.5)	(1.8)
	TN rate	28.75	52.71	41.50	65.38	40.83	64.71
		(1.5)	(1.6)	(2.1)	(1.6)	(2.2)	(1.7)

Standard errors are in parenthesis. All values are multiplied by 10^2.
LASSO = least absolute shrinkage and selection operator; SCAD = smoothly clipped absolute deviation; MCP = minimax concave penalty; MSE = mean squared error; TP = true positive; TN = true negative.

$\epsilon \sim N(0, 2^2)$. The model is:

$$Y = \frac{1}{2}X_3^3 + \frac{1}{2}X_3^2 + \frac{1}{2}X_3 + \frac{1}{6}X_6^3 - \frac{1}{2}X_6^2 + \frac{1}{3}X_6 + I(X_{11} = 0) + \frac{1}{2}I(X_{11} = 1) + \epsilon.$$

- Model III: Random variables $X_1, \ldots, X_{15}$ are first simulated according to a centered multivariate normal distribution with covariance between X_i and X_j being $0.5^{|i-j|}$. Then X_i is trichotomized as 0, 1, or 2 if it is smaller than $\Phi^{-1}(1/3)$, larger than $\Phi^{-1}(2/3)$ or in between. The model is:

$$Y = 0.9I(X_1 = 1) - 0.6I(X_1 = 0) + 0.5I(X_3 = 1) + 0.25I(X_3 = 0) + 0.5I(X_5 = 1) + 0.5I(X_5 = 0) + \epsilon.$$

- Model IV: 4 categorical factors X_1, X_2, X_3 and X_4 are generated as in model III. The model is:

$$Y = 1.5I(X_1 = 1) + I(X_1 = 0) + 1.5I(X_2 = 1) + 1I(X_2 = 0) + 0.5I(X_1 = 1, X_2 = 1)$$
$$+ 0.75I(X_1 = 1, X_2 = 0) + I(X_1 = 0, X_2 = 1) + 1.25I(X_1 = 0, X_2 = 0) + \epsilon.$$

Table 7: Simulation results for Model IV. Mean of MSE, TP rate, TN rate for LASSO, SCAD, MCP are from 200 simulated data sets with $n = 100$

		LASSO(G)	LASSO(NG)	SCAD(G)	SCAD(NG)	MCP(G)	MCP(NG)
$\tau = 0.1$	MSE	974.87	4382.14	1089.52	4378.63	1087.99	4348.43
		(35.7)	(48.4)	(61.4)	(46.9)	(61.1)	(43.8)
	TP rate	90.44	64.75	97.81	51.88	97.81	51.00
		(0.9)	(2.2)	(0.6)	(2.4)	(0.6)	(2.4)
	TN rate	23.25	35.25	8.38	48.13	8.00	49.00
		(1.0)	(2.2)	(1.1)	(2.4)	(1.1)	(2.4)
$\tau = 0.2$	MSE	761.39	4016.89	977.69	4143.90	994.03	4151.36
		(33.7)	(42.1)	(44.2)	(47.3)	(43.3)	(46.2)
	TP rate	93.38	71.75	83.44	54.38	82.69	55.13
		(0.8)	(1.9)	(1.7)	(2.5)	(1.7)	(2.5)
	TN rate	25.63	28.25	29.00	45.63	28.63	44.88
		(1.9)	(1.9)	(2.2)	(2.5)	(2.2)	(2.5)
$\tau = 0.3$	MSE	642.29	3853.86	946.14	3963.01	929.65	4026.50
		(29.1)	(51.1)	(42.8)	(54.7)	(42.5)	(58.1)
	TP rate	93.94	75.25	80.75	65.13	81.75	63.50
		(0.7)	(1.7)	(1.7)	(2.0)	(1.7)	(2.0)
	TN rate	25.00	24.75	38.50	34.88	37.38	36.50
		(1.8)	(1.7)	(2.6)	(2.0)	(2.6)	(2.0)
$\tau = 0.4$	MSE	641.29	3601.66	941.30	3751.45	975.85	3759.88
		(30.8)	(51.0)	(44.1)	(56.8)	(44.2)	(57.1)
	TP rate	92.00	74.25	81.56	62.88	80.44	62.13
		(0.9)	(1.7)	(1.7)	(1.9)	(1.7)	(1.9)
	TN rate	29.00	25.75	41.25	37.13	40.75	37.88
		(2.0)	(1.7)	(2.7)	(1.9)	(2.6)	(1.9)
$\tau = 0.5$	MSE	702.93	3445.67	1043.48	3571.69	1056.95	3599.84
		(34.0)	(49.4)	(47.6)	(52.0)	(47.4)	(51.9)
	TP rate	90.06	78.63	78.31	57.13	78.19	57.50
		(0.9)	(1.5)	(1.7)	(2.0)	(1.7)	(2.0)
	TN rate	28.63	21.38	45.63	42.88	43.63	42.50
		(1.9)	(1.9)	(2.9)	(2.0)	(2.8)	(2.0)

Standard errors are in parenthesis. All values are multiplied by 10^2.
LASSO = least absolute shrinkage and selection operator; SCAD = smoothly clipped absolute deviation; MCP = minimax concave penalty; MSE = mean squared error; TP = true positive; TN = true negative.

We generate $R = 200$ datasets for each example and penalty parameter is chosen through a 10-fold cross validation under 100 penalty parameter values. To measure the performance of the methods, we use mean squared error (MSE) for estimation accuracy calculated as MSE $= (\hat{\beta} - \beta)^{\text{TE}}(X^T X)(\hat{\beta} - \beta)$ along with the true positive (TP) rate and the true negative (TN) rate for identification accuracy over R simulated data sets. From above 4 models, we compare group penalized quantile regression of LASSO, SCAD, MCP (LASSO(G), SCAD(G), MCP(G)) and penalized quantile regression not considering grouping structure (LASSO(NG), SCAD(NG), MCP(NG)). Sample size $n = 100$ and 200 are considered, but the results from $n = 100$ are reported only.

For implementation, we use `rqPen` for both the grouped version and non-grouped version of LASSO, SCAD, MCP. The results of the simulation for Model I are summarized in Table 4, Model II are in Table 5, Model III are in Table 6, and Model IV are in Table 7.

From the results, we can notice that the group version of LASSO, SCAD, MCP perform better than the models not considering grouping structure in most cases.

Finally, we only investigate heavy tail distributions under the Model I. Now, the error distributions ϵ are generated from the t-distribution with 5 degrees of freedom. 200 Monte Carlo data sets with

Table 8: Mean of MSE, TP rate, and TN rate for LASSO and Group LASSO (and its standard error in the parentheses).

		MSE	TP rate	TN rate
$\tau = 0.1$	LASSO(NG)	30.78 (1.35)	99.58 (0.18)	55.36 (1.61)
	LASSO(G)	26.02 (1.04)	100.00 (0.00)	57.29 (1.42)
$\tau = 0.2$	LASSO(NG)	30.28 (1.21)	99.50 (0.20)	56.36 (1.53)
	LASSO(G)	26.39 (1.06)	100.00 (0.00)	57.82 (1.53)
$\tau = 0.3$	LASSO(NG)	30.26 (1.28)	6.71 (0.20)	55.71 (1.58)
	LASSO(G)	26.46 (1.04)	100.00 (0.00)	56.39 (1.50)
$\tau = 0.4$	LASSO(NG)	30.44 (1.27)	6.71 (0.18)	55.11 (1.56)
	LASSO(G)	26.27 (1.04)	100.00 (0.00)	58.21 (1.52)
$\tau = 0.5$	LASSO(NG)	30.52 (1.27)	6.71 (0.18)	56.46 (1.61)
	LASSO(G)	26.57 (1.04)	100.00 (0.00)	57.11 (1.50)

Errors follows t-distribution with 5 degrees of freedom. All values are multiplied by 10^2. (NG) and (G) respectively stands for non-group and group version of LASSO.
LASSO = least absolute shrinkage and selection operator; MSE = mean squared error; TP = true positive; TN = true negative.

n=200 samples are generated and we summarize the results by the mean of the MSE, mean true positive rate (TP rate), and mean true negative rate (TN rate).

The results in Table 8 show that the group version of the regression models yield the lower MSE with higher true positive and true negative rates under heavy tail error distribution.

5. Concluding remarks

This work presents numerical studies for group penalized quantile regression with the group LASSO, group SCAD, and group MCP penalties for analyzing the data with a grouping structure. The application of group penalty to quantile regression seems reasonable and improves the fit. We also applied group penalties to high dimensional data; however, these penalties did not work well in R package `rqPen`. It also failed to capture the grouping structure and provided the same results as the non-group version despite being designed for high dimensional data. High dimension data is common in genetic areas; therefore, future studies on its implementation look very interesting.

Acknowledgements

References

Bakin S (1999). *Adaptive regression and model selection in data mining problems* (PhD thesis), The Australian National University.

Chavent M and Kuentz-Simonet V (2012). ClustOfVar: an R package for the clustering of variables, *Journal of Statistical Software*, **50**, 1–16.

Ciuperca G (2019). Adaptive group LASSO selection in quantile models, *Statistical Papers*, **60**, 173–197.

Fan J (1997). Comments on wavelets in statistics: a review by A. Antoniadis, *Journal of the Italian Statistical Society*, **6**, 131–138.

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

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

Hendricks W and Koenker R (1992). Hierarchical spline models for conditional quantiles and the demand for electricity, *Journal of the American Statistical Association*, **87**, 58–68.

Hoerl AE and Kennard RW (1970). Ridge regression: biased estimation for nonorthogonal problems, *Technometrics*, **12**, 55–67.

Hosmer DW and Lemeshow S (1989). *Applied logistic regression*, Wiley, Vol. 2.

Huang J, Breheny P, and Ma S (2012). A Selective Review of Group Selection in High-Dimensional Models, *Statistical Science*, **27**, 481–499.

Kato K (2011). Group Lasso for high dimensional sparse quantile regression models, *arXiv:1103.1458 v2 [stat.ME]*.

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

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

Koenker R and Hallock KF (2001). Quantile regression, *Journal of Economic Perspectives*, **15**, 143–156.

Koenker R, NG P, and Portnoy S (1994). Quantile smoothing splines, *Biometrika*, **81**, 673–680.

Ogutu JO and Piepho HP (2014). Regularized group regression methods for genomic prediction: Bridge, MCP, SCAD, group bridge, group lasso, sparse group lasso, group MCP and group SCAD, *BMC Proceedings*, **8**.

Sakar BE, Isenkul ME, Sakar CO, Sertbas A, Gurgen F, Delil S, Apaydin H, and Kursun O (2013). Collection and analysis of a Parkinson Speech Dataset with multiple types of sound recordings, *IEEE Journal of Biomedical and Health Informatics*, **17**, 828–834.

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

Wang H and He X (2007). Detecting differential expressions in GeneChip microarray studies: a quantile approach, *Journal of the American Statistical Association*, **102**, 104–112.

Wei Y and He X (2006). Conditional growth charts, *The Annals of Statistics*, **34**, 2069–2097.

Wei Y, Pere A, Koenker R, and He X (2006). Quantile regression methods for reference growth charts, *Statistics in Medicine*, **25**, 1369–1382.

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

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

Zhang CH (2010). Nearly unbiased variable selection under minimax concave penalty, *The Annals of Statistics*, **38**, 894–942.

Zou H (2006). The adaptive lasso and its oracle properties, *Journal of the American Statistical Association*, **101**, 1418–1429,

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