Regularization Methods Based on the \(L_q\)-Likelihood for Linear Models with Heavy-Tailed Errors

Yoshihiro Hirose†
Hokkaido University, Japan

Abstract
We propose regularization methods for linear models based on the \(L_q\)-likelihood, which is a generalization of the log-likelihood using a power function. Some heavy-tailed distributions are known as \(q\)-normal distributions. We find that the proposed methods for linear models with \(q\)-normal errors coincide with the regularization methods that are applied to the normal linear model. The proposed methods work well and efficiently, and can be computed using existing packages. We examine the proposed methods using numerical experiments, showing that the methods perform well, even when the error is heavy-tailed.

Keywords: LASSO, MCP, \(q\)-normal distribution, SCAD, Sparse estimation

1 Introduction
We propose regularization methods based on the \(L_q\)-likelihood for linear models with heavy-tailed errors. These methods turn out to coincide with the ordinary regularization methods that are used for the normal linear model. The proposed methods work efficiently, and can be computed using existing packages.

Linear models are widely applied, and many methods have been proposed for estimation, prediction, and other purposes. For example, for estimation and variable selection in the normal linear model, the literature on sparse estimation includes the least absolute shrinkage and selection operator (LASSO) [18], smoothly clipped absolute deviation (SCAD) [8], Dantzig selector [4], and minimax concave penalty (MCP) [20]. The LASSO has been studied extensively and generalized to many models, including the generalized linear models [15].

Because the regularization methods for the normal linear model are useful, they are sometimes used in linear models with non-normal errors. Here,
popular errors include the Cauchy error and the $t$-distribution error, both of which are heavy-tailed errors. For example, [1] partly consider the Cauchy and $t$-distribution errors in their extensive experiments. These heavy-tailed distributions are known to be $q$-normal distributions, which are studied in the literature on statistical mechanics [10] [16] [19]. The $q$-normal model is also studied in the literature on the generalized Cauchy distribution. For example, see [2] [3] [5] [6].

In this study, we consider the problem of a linear regression with a $q$-normal error. We propose sparse estimation methods based on the $L_q$-likelihood, which is a generalization of the log-likelihood using a power function. The maximizer of the $L_q$-likelihood, the maximum $L_q$-likelihood estimator (MLqE), is investigated by [9] as an extension of the ordinary maximum likelihood estimator (MLE). [9] studies the asymptotic properties of the MLqE.

The rest of the paper is organized as follows. In Section 2, we introduce several tools, including the normal linear model, regularization methods, $L_q$-likelihood, and $q$-normal models. In Section 2 we describe the problem under consideration, that is, estimations in linear models with $q$-normal errors. Moreover, we propose several regularization methods based on the $L_q$-likelihood. In Section 4, we evaluate the proposed methods using numerical experiments. Section 5 concludes the paper.

2 Preliminaries

2.1 Normal Linear Model and Sparse Estimation

First, we introduce the normal linear model, the estimation of which is a basic problem in statistics and machine learning [11]. Furthermore, we briefly describe some well-known regularization methods.

The normal linear model is defined as follows. A response is represented by a linear combination of explanatory variables $x_1, x_2, \ldots, x_d$ as

$$y^a = \theta^0 + \sum_{i=1}^{d} x_i^a \theta^i + \varepsilon^a \quad (a = 1, 2, \ldots, n),$$

where $y^a$ is the response of the $a$-th sample, $n$ is the sample size, $d$ is the number of explanatory variables, $x_i^a$ is the i-th explanatory variable of the a-th sample, $\varepsilon^a$ is a normal error with mean zero and known variance, and the regression coefficient $\theta = (\theta^0, \theta^1, \ldots, \theta^d)^\top$ is the parameter to be estimated. The normal linear model is equivalently given by

$$\mu = X \theta,$$

where $\mu^a = E[y^a]$ is the expectation of the response $y^a$, $\mu = (\mu^a)$, and $X = (x_i^a)$ is a design matrix of size $n \times (d + 1)$, with $x_0^a = 1 (a = 1, 2, \ldots, n)$. Moreover, we define a row vector $x^a (a = 1, 2, \ldots, n)$ as $x^a = (1, x_1^a, x_2^a, \ldots, x_d^a)$, and a column vector $x_i (i = 0, 1, 2, \ldots, d)$ as $x_i = (x_1^i, x_2^i, \ldots, x_d^i)^\top$, which results in $X = (x_1^\top, x_2^\top, \ldots, x_n^\top)^\top = (x_0, x_1, x_2, \ldots, x_d)$. Let $y = (y^a)$ be the response
vector of length \( n \). We assume that each column vector \( x_i (i = 1, 2, \ldots, d) \) is
standardized, as follows: \( \sum_{a=1}^{n} x_i^a = 0 \) and \( \|x_i\| = 1 \), for \( i = 1, 2, \ldots, d \).

As is well known, some regularization methods for the normal linear model
are formulated as an optimization problem in the form of

\[
\min_{\theta \in \mathbb{R}^{d+1}} \left\{ \frac{1}{2n} \| y - X\theta \|^2 + \rho_\lambda(\theta) \right\},
\]

where \( \rho_\lambda(\theta) \) is a penalty term, and \( \lambda \geq 0 \) is a regularization parameter. The
LASSO \[18\] uses \( \rho_\lambda(\theta) = \lambda \|\theta\|_1 = \lambda \sum_{i=1}^{d} |\theta^i| \). The path of the LASSO estimator when \( \lambda \) varies can be made by the least angle regression (LARS) algorithm \[7\]. The SCAD \[8\] uses

\[
\rho_\lambda(\theta) = \begin{cases} 
\sum_{i=1}^{d} \lambda |\theta^i| & (|\theta^i| \leq \lambda), \\
-\sum_{i=1}^{d} |\theta^i|^2 - 2a\lambda |\theta^i| + \lambda^2 & (\lambda < |\theta^i| \leq a\lambda), \\
\sum_{i=1}^{d} (a + 1)\lambda^2 / 2 & (a\lambda < |\theta^i|),
\end{cases}
\]

and the MCP \[20\] uses

\[
\rho_\lambda(\theta) = \lambda \sum_{i=1}^{d} \int_{0}^{\theta^i} \left( 1 - \frac{u}{\gamma\lambda} \right)_+ du,
\]

where \( a(> 2) \) and \( \gamma(> 0) \) are tuning parameters.

The regularization problem given in \[2\] can be represented by

\[
\min_{\theta \in \mathbb{R}^{d+1}} \left\{ -\frac{1}{n} \log f(y|\theta) + \rho_\lambda(\theta) \right\},
\]

where \( f(y|\theta) \) is the probability density function of the statistical model. Note that \( \log f(y|\theta) \) is the log-likelihood.

### 2.2 \( L_q \)-Likelihood

The \( L_q \)-likelihood is a generalization of the log-likelihood that uses a power function instead of the logarithmic function. Let \( y = (y^1, y^2, \ldots, y^n)^T \) be a vector of independent and identically distributed (i.i.d.) observations, and let \( \theta \) be a parameter of a statistical model. For \( q > 0 (q \neq 1) \), the \( L_q \)-likelihood function is defined as

\[
L_q(\theta|y) = \sum_{a=1}^{n} \log_q f(y^a|\theta),
\]
where \( f(\cdot | \theta) \) is a probability density function of the statistical model, and
\[
\log_q(u) = \frac{1}{1 - q} (u^{1-q} - 1) \quad (u > 0)
\]
is the \( q \)-logarithmic function \[19\]. For \( q = 1 \), we define
\[
\log_1(u) = \log u \quad (u > 0),
\]
which is the ordinary logarithmic function. When \( q = 1 \), the \( L_q \)-likelihood is the log-likelihood.

The ML\( q \)E is defined as the estimator that maximizes the \( L_q \)-likelihood. \[9\] studied the asymptotic performance of the ML\( q \)E, showing that it enjoys good asymptotic properties (e.g., asymptotic normality).

### 2.3 \( q \)-Normal Model

Before defining the \( q \)-normal distribution \[10, 16, 19\], we introduce the \( q \)-exponential function. For \( q > 0 \) \((q \neq 1)\), the \( q \)-exponential function is the inverse function of the \( q \)-logarithmic function, and is given by
\[
\exp_q(u) = \begin{cases} 
1 + \left(1 - \frac{1}{q}\right)u & (u > -1/(1-q)), \\
0 & \text{(otherwise)}. 
\end{cases}
\]
For \( q = 1 \), the 1-exponential function is the ordinary exponential function
\[
\exp_1(u) = \exp u \quad (u \in \mathbb{R}).
\]

Using the \( q \)-exponential function, the \( q \)-normal model is given by
\[
S_q = \{ f_q(y|\xi, \sigma) | \xi \in \Xi, \sigma > 0 \},
\]
\[
f_q(y|\xi) = \frac{1}{Z_q} \exp_q \left\{ -\frac{1}{3 - q} \left( \frac{y - \xi}{\sigma} \right)^2 \right\}
\]
\[
= \frac{1}{Z_q} \left\{ 1 - \frac{1 - q}{3 - q} \left( \frac{y - \xi}{\sigma} \right)^2 \right\}^{\frac{1}{1-q}},
\]
where \( \xi \) is a location parameter, \( \Xi \subset \mathbb{R} \) is the parameter space, and \( \sigma \) is a dispersion parameter. The constant \( Z_q \) is a normalizing constant.

We assume that \( 1 \leq q < 3 \), which ensures that the sample space is the real line itself, not just part of it. Moreover, the parameter space is \( \Xi = \mathbb{R} \) when \( 1 \leq q < 3 \).

For example, the 1-normal model is the ordinary normal model. Another example is the Cauchy distribution for \( q = 2 \):
\[
f_2(y|\mu, \sigma) = \frac{1}{\sigma B(\frac{1}{2}, \frac{1}{2})} \left( 1 + \frac{(y - \xi)^2}{\sigma^2} \right)^{-1},
\]
where \( B(\cdot, \cdot) \) is the beta function. Furthermore, the \( t \)-distribution of the degree of freedom \( \nu \) is obtained for \( q = 1 + 2/(\nu + 1) \):

\[
f_{1+2/(\nu+1)}(y|\mu, \sigma) = \frac{1}{\sqrt{\nu}\sigma B(\frac{\nu}{2}, \frac{1}{2})} \left(1 + \frac{(y - \mu)^2}{\nu\sigma^2}\right)^{-\frac{\nu+1}{2}}.
\]

### 3 Problem and Estimation Method

#### 3.1 Linear Model with \( q \)-Normal Error

In this subsection, we formulate our problem, that is, a linear regression with a heavy-tailed error. The errors of the Cauchy and \( t \)-distributions in linear models have been studied by researchers in the context of heavy-tailed errors [12, 13, 14, 17]. However, they focused mainly on the MLE, whereas we are interested in sparse estimators. Moreover, our approach is based on the \( L_q \)-likelihood, not the ordinary log-likelihood.

We examine the problem of estimating the linear model given in (1) with i.i.d. errors from a \( q \)-normal distribution; henceforth, we refer to this as the \( q \)-normal linear model. In terms of probability distributions, we wish to estimate the parameter \( \theta \) of the \( q \)-normal linear model \( M_q \):

\[
M_q = \{ f(\cdot|\theta) | \theta \in \mathbb{R}^{d+1} \},
\]

\[
f(y|\theta) = \frac{1}{Z_q^n} \prod_{a=1}^{n} \exp_q \left\{ - \frac{(y^a - x^a\theta)^2}{3-q} \right\} = \frac{1}{Z_q^n} \prod_{a=1}^{n} \left\{ 1 - \frac{1-q}{3-q} (y^a - x^a\theta)^2 \right\}^{\frac{1}{1-q}},
\]

where the dispersion parameter is assumed to be known (\( \sigma = 1 \)). The 1-normal linear model is identical to the normal linear model, as described in subsection 2.1.

#### 3.2 \( L_q \)-likelihood-based Regularization Methods

We propose regularization methods based on the \( L_q \)-likelihood. For \( q \)-normal linear models, the proposed methods coincide with the original regularization methods for the normal linear model. In other words, we apply the ordinary regularization methods as if the error distribution were a normal distribution. The literature describes how to compute the proposed methods efficiently. Moreover, our method calculates the MLqE.

We define the \( L_q \)-likelihood for the \( q \)-normal linear model in (7) as (6), where \( \theta \) is the regression coefficient. Note that the components of \( y \) are not assumed to be identically distributed because their distributions are dependent on the explanatory variables.
The $L_q$-likelihood for the $q$-normal linear model is

$$L_q(\theta | y) = \sum_{a=1}^{n} \log_q f(y^a | \theta)$$

$$= \sum_{a=1}^{n} \log_q \left[ \frac{1}{Z_q} \exp_q \left\{ - \frac{(y^a - x^a \theta)^2}{3 - q} \right\} \right]$$

$$= - \frac{Z_q^{-1}}{3 - q} \|y - X \theta\|^2 - n \log_q(Z_q),$$  \hfill (8)

where the second term is a constant. The ML$q$E of the parameter $\theta$ is defined as the maximizer of the $L_q$-likelihood. In the $q$-normal linear model, the ML$q$E is equal to the ordinary least square, the MLE for the normal linear model.

We propose a LASSO, SCAD, and MCP based on the $L_q$-likelihood by replacing the log-likelihood with the $L_q$-likelihood in the optimization problem in (5). That is, the $L_q$-likelihood-based regularization methods are given in the form of

$$\min_{\theta \in \mathbb{R}^{d+1}} \left\{ - \frac{1}{n} L_q(\theta | y) + \rho_\lambda(\theta) \right\}. \hfill (9)$$

The penalty $\rho_\lambda$ is $\lambda \|\theta\|_1$ for the LASSO, (3) for the SCAD, and (4) for the MCP. Note that the estimator for $\lambda = 0$ is the ML$q$E. As a special case, the proposed methods are the ordinary regularization methods when $q = 1$.

Because of (5) and (9), for the $q$-normal linear models, the $L_q$-likelihood-based regularization methods are essentially the same as the penalized least square (2). In other words, we implicitly use the $L_q$-likelihood-based regularization methods when we apply the ordinary LASSO, SCAD, and MCP to data with heavy-tailed errors.

4 Numerical Experiments

In this section, we describe the results of our numerical experiments and compare the proposed methods. Here, we focus on model selection and generalization.

Our methods do not require additional implementations because the LASSO, SCAD, and MCP are already implemented in software packages. In the experiments, we use the ncvreg package of the software R.

4.1 Setting

The procedure for the experiments is as follows. We fix the value $q$ of the $q$-normal linear model and the $L_q$-likelihood, the dimension $d$ of the parameter $\theta$, the ratio of nonzero components $r_{nz}$ of $\theta$, the true value $\theta_0$ of the nonzero components of $\theta$, and the sample size $n$. The value of $q$ is selected from 1, 13/11, 3/2, 5/3, 2, 2.01, 2.1, and 2.5, where $q = 13/11$ is the $t$-distribution with $\nu = 10$ degrees of freedom, $q = 3/2$ is the $t$-distribution with $\nu = 3$ degrees of freedom, and $q = 5/3$ is the $t$-distribution with $\nu = 2$ degrees of freedom. The
Table 1: All cases in the experiments. Each case is studied for the values of $q$ and $n$.

| $\theta_0$ | $r_{nz} = 0.2$ | $r_{nz} = 0.4$ | $r_{nz} = 0.6$ | $r_{nz} = 0.8$ |
|------------|---------------|---------------|---------------|---------------|
| $d = 10$ | $d = 100$  | $d = 10$ | $d = 100$ | $d = 10$ | $d = 100$ | $d = 10$ | $d = 100$ |
| 10$^0$ | 1     | 5     | 9     | 13  | 17  | 21  | 25  | 29  |
| 10$^1$ | 2     | 6     | 10    | 14  | 18  | 22  | 26  | 30  |
| 10$^2$ | 3     | 7     | 11    | 15  | 19  | 23  | 27  | 31  |
| 10$^3$ | 4     | 8     | 12    | 16  | 20  | 24  | 28  | 32  |

sample size is $n = 100$ or $n = 1,000$. The true parameter consists of $d \times r_{nz}$ $\theta_0$s and $d \times (1 - r_{nz})$ zeros. All cases are illustrated in Table 1.

For each of $m = 1,000$ trials, we create the design matrix $X$ using the `rnorm()` function in R. The response $y$ is generated as $q$-normal random variables using the qGaussian package. For the estimation, we apply the `ncvreg()` function to ($y$, $X$) with the default options; for example, the values of the tuning parameters are $a = 3.7$ and $\gamma = 3$.

To select one model and one estimate from a sequence of parameter estimates generated by a method, we use the AIC and BIC:

$$\text{AIC} = -2 \log p(y|\hat{\theta}) + 2d', \quad (10)$$
$$\text{BIC} = -2 \log p(y|\hat{\theta}) + d' \log n, \quad (11)$$

where $d'$ is the dimension of parameters of the model under consideration. Moreover, we use other criteria based on the $L_q$-likelihood:

$$L_q\text{-AIC} = -2L_q p(y|\hat{\theta}) + 2d', \quad (12)$$
$$L_q\text{-BIC} = -2L_q p(y|\hat{\theta}) + d' \log n. \quad (13)$$

For a sequence ($\hat{\theta}(k)$) made by each of the methods, let $I(k) = \{i | \hat{\theta}_i^{(k)} \neq 0\}$ and $\hat{\theta}_I^{(k)}$, the MLE of the model $\mathcal{M}(k) = \{p(\cdot | \theta) | \theta^j = 0 (j \notin I(k))\}$. We call \(^{(1)}\) with $\hat{\theta} = \hat{\theta}_\text{MLE}$ AIC1, and \(^{(1)}\) with $\hat{\theta} = \hat{\theta}(k)$ AIC2. Similarly, \(^{(1)}\) with $\hat{\theta} = \hat{\theta}_I^{(k)}$ is BIC1, and \(^{(1)}\) with $\hat{\theta} = \hat{\theta}(k)$ is BIC2. The $L_q$-AIC and $L_q$-BIC are referred to in the same manner; for example, \(^{(1)}\) with $\hat{\theta} = \hat{\theta}_\text{MLE}$ is $L_q$-AIC1. Note that AIC1, BIC1, $L_q$-AIC1, and $L_q$-BIC1 are available only when the MLE exists; AIC2, BIC2, $L_q$-AIC2, and $L_q$-BIC2 are always applicable. Finally, we used cross-validation (CV) in addition to these information criteria.

### 4.2 Result

The results are presented in Tables 2–23, which report the best result for each method based on the various information criteria.
The model selection results are reported in Tables 2–15. Each number shows the number of trials (among \(m = 1,000\) trials) where a method selects the true model. Here, a larger value is better.

The generalization results are reported in Tables 16–23. To evaluate the generalization error of the proposed methods, we newly make \(m = 1,000\) independent copies \(\{(y'_1, X'_1), \ldots, (y'_m, X'_m)\}\) in each trial. We computed the difference between \((y'_1, \ldots, y'_m)\) and the \(m\) predictions using each of the methods. Each value shows the average prediction error over \(m\) trials. In this case, a smaller value is better.

Our first concern is whether the proposed methods work well. The results for \(q = 1\) can be regarded as a reference for the other values of \(q\). The tables show that the proposed methods work well in both model selection and generalization, especially for \(q < 2\). The methods also perform well in terms of model selection for \(q = 2, 2.01,\) and \(2.1\). However, they perform poorly for \(q = 2.5\) in terms of model selection and for \(q \geq 2\) in terms of generalization. As anticipated, a large \(q\) makes the problem difficult.

Second, we evaluate the performance of the proposed methods, finding that the MCP performs best in most cases. In a few cases, the MCP performed similarly to or slightly worse than the other methods. For model selection, the cases with \(q = 1, n = 1,000\) and large \(\theta_0\) are exceptions. Furthermore, the LASSO performed worse than the SCAD and MCP.

Third, we consider the effect of \(r_{nx}, \theta_0, d,\) and \(n,\) in addition to \(q\). The cases with large \(r_{nx}\) and/or small \(\theta_0\) are difficult. Moreover, a large \(d\) makes the problems difficult. However, if we have a small \(q (1 \leq q < 2),\) large \(\theta_0 (\theta_0 = 10^2, 10^3)\) and small \(r_{nx},\) the problems with large \(d\) can be easier than those with small \(d.\) Furthermore, a small \(n\) makes the problems difficult in a similar manner to a large \(d.\) These observations imply that, for \(1 \leq q < 2,\) small-sample problems can be easier than large-sample problems if \(r_{nx}\) is small and \(\theta_0\) is large.

Fourth, the choice of information criterion changes the methods’ performance. In terms of model selection, BIC2 was mostly the best for many values of \(q\). For \(3/2 \leq q \leq 2.1,\) BIC1 was a little better than BIC2 if BIC1 was available. For \(q = 1\) and \(13/11,\) BIC2 was better than BIC1. AIC1 and AIC2 were as good as BICs for \(2 \leq q \leq 2.1.\) Moreover, the \(L_q\)-BIC1 and -BIC2 were best only for \(q = 3/2,\) when BIC1 and BIC2 performed just as well. Overall, the \(L_q\)-information criteria performed poorly.

Furthermore, in terms of generalization, BIC2 was mostly the best. AIC2 was as good as BIC2, whereas AIC2 was sometimes a little worse than BIC2. The information criteria using the \(L_q\)-likelihood were poor for \(q = 13/11.\) For \(q = 1, 3/2,\) and \(5/3,\) the \(L_q\)-information criteria worked as well as the ordinary criteria and CV, except for some cases. The performance of CV was mostly good, but was occasionally very poor.

In summary, using an appropriate criterion, the proposed methods perform well for linear models with slightly heavy-tailed errors \((1 \leq q < 2).\) Moreover, the proposed methods work in terms of model selection, even if the error is heavy-tailed \((2 \leq q < 2.5).\) Overall, we recommend using the MCP and BIC2.
Table 2: The result on model selection for $q = 1$ and $n = 100$. Each number indicates how many trials a method correctly selects the true model using an appropriate information criterion.

| $\theta_0$ | Method | $r_{\text{nz}} = 0.2$ | $r_{\text{nz}} = 0.4$ | $r_{\text{nz}} = 0.6$ | $r_{\text{nz}} = 0.8$ |
|------------|--------|-----------------------|-----------------------|-----------------------|-----------------------|
|            |        | $d = 10$ | $d = 100$ | $d = 10$ | $d = 100$ | $d = 10$ | $d = 100$ | $d = 10$ | $d = 100$ |
| 10$^0$     | LASSO  | 32        | 0         | 9         | 0         | 5         | 0         | 8         | 0         |
|            | SCAD   | 32        | 0         | 8         | 0         | 5         | 0         | 8         | 0         |
|            | MCP    | 32        | 0         | 8         | 0         | 8         | 0         | 5         | 0         |
| 10$^1$     | LASSO  | 366       | 0         | 486       | 0         | 625       | 0         | 784       | 0         |
|            | SCAD   | 710       | 415       | 719       | 114       | 768       | 0         | 812       | 0         |
|            | MCP    | 778       | 519       | 793       | 84        | 797       | 0         | 813       | 0         |
| 10$^2$     | LASSO  | 366       | 35        | 487       | 0         | 627       | 0         | 788       | 0         |
|            | SCAD   | 781       | 1,000     | 782       | 822       | 808       | 2         | 821       | 0         |
|            | MCP    | 802       | 1,000     | 812       | 890       | 814       | 8         | 833       | 0         |
| 10$^3$     | LASSO  | 365       | 46        | 496       | 0         | 655       | 0         | 823       | 0         |
|            | SCAD   | 781       | 1,000     | 809       | 817       | 871       | 2         | 945       | 0         |
|            | MCP    | 808       | 1,000     | 832       | 889       | 880       | 4         | 938       | 0         |

Table 3: The result on model selection for $q = 1$ and $n = 1,000$.

| $\theta_0$ | Method | $r_{\text{nz}} = 0.2$ | $r_{\text{nz}} = 0.4$ | $r_{\text{nz}} = 0.6$ | $r_{\text{nz}} = 0.8$ |
|------------|--------|-----------------------|-----------------------|-----------------------|-----------------------|
|            |        | $d = 10$ | $d = 100$ | $d = 10$ | $d = 100$ | $d = 10$ | $d = 100$ | $d = 10$ | $d = 100$ |
| 10$^0$     | LASSO  | 13        | 0         | 7         | 0         | 5         | 0         | 8         | 0         |
|            | SCAD   | 13        | 0         | 7         | 0         | 5         | 0         | 10        | 0         |
|            | MCP    | 13        | 0         | 8         | 0         | 7         | 0         | 12        | 0         |
| 10$^1$     | LASSO  | 579       | 28        | 670       | 81        | 774       | 216       | 875       | 417       |
|            | SCAD   | 823       | 485       | 840       | 478       | 851       | 549       | 892       | 675       |
|            | MCP    | 790       | 611       | 781       | 630       | 820       | 655       | 886       | 665       |
| 10$^2$     | LASSO  | 577       | 28        | 670       | 85        | 774       | 224       | 875       | 449       |
|            | SCAD   | 828       | 833       | 842       | 847       | 853       | 856       | 891       | 842       |
|            | MCP    | 768       | 753       | 788       | 751       | 823       | 781       | 886       | 790       |
| 10$^3$     | LASSO  | 577       | 28        | 670       | 85        | 774       | 226       | 875       | 466       |
|            | SCAD   | 851       | 834       | 878       | 848       | 902       | 873       | 935       | 892       |
|            | MCP    | 778       | 754       | 812       | 760       | 848       | 791       | 906       | 827       |
Table 4: The result on model selection for $q = \frac{13}{11}$ and $n = 100$.  

| $\theta_0$ | Method | $r_{\text{nz}} = 0.2$ | $r_{\text{nz}} = 0.4$ | $r_{\text{nz}} = 0.6$ | $r_{\text{nz}} = 0.8$ |
|------------|--------|-----------------|-----------------|-----------------|-----------------|
|            |        | $d = 10$        | $d = 100$       | $d = 10$        | $d = 100$       |
| $10^0$     | LASSO  | 29              | 0               | 4               | 0               | 3               | 0               |
|            | SCAD   | 29              | 0               | 4               | 0               | 2               | 0               |
|            | MCP    | 29              | 0               | 4               | 0               | 2               | 0               |
| $10^1$     | LASSO  | 796             | 0               | 821             | 0               | 877             | 0               | 939             | 0               |
|            | SCAD   | 896             | 382             | 890             | 55              | 901             | 0               | 944             | 0               |
|            | MCP    | 869             | 588             | 884             | 135             | 901             | 0               | 939             | 0               |
| $10^2$     | LASSO  | 797             | 32              | 821             | 0               | 878             | 0               | 939             | 0               |
|            | SCAD   | 924             | 1,000           | 925             | 822             | 921             | 1               | 939             | 0               |
|            | MCP    | 872             | 1,000           | 884             | 897             | 902             | 7               | 944             | 0               |
| $10^3$     | LASSO  | 797             | 59              | 821             | 0               | 879             | 0               | 940             | 0               |
|            | SCAD   | 929             | 1,000           | 946             | 823             | 953             | 1               | 975             | 0               |
|            | MCP    | 875             | 1,000           | 897             | 886             | 921             | 8               | 959             | 0               |

Table 5: The result on model selection for $q = \frac{13}{11}$ and $n = 1,000$.  

| $\theta_0$ | Method | $r_{\text{nz}} = 0.2$ | $r_{\text{nz}} = 0.4$ | $r_{\text{nz}} = 0.6$ | $r_{\text{nz}} = 0.8$ |
|------------|--------|-----------------|-----------------|-----------------|-----------------|
|            |        | $d = 10$        | $d = 100$       | $d = 10$        | $d = 100$       |
| $10^0$     | LASSO  | 21              | 0               | 3               | 0               | 2               | 0               |
|            | SCAD   | 21              | 0               | 2               | 0               | 1               | 0               |
|            | MCP    | 22              | 0               | 2               | 0               | 1               | 0               |
| $10^1$     | LASSO  | 936             | 658             | 951             | 774             | 965             | 768             | 982             | 785             |
|            | SCAD   | 941             | 588             | 951             | 651             | 965             | 711             | 982             | 838             |
|            | MCP    | 955             | 792             | 958             | 769             | 969             | 792             | 982             | 873             |
| $10^2$     | LASSO  | 936             | 657             | 952             | 786             | 965             | 837             | 982             | 893             |
|            | SCAD   | 978             | 980             | 981             | 981             | 981             | 980             | 985             | 984             |
|            | MCP    | 961             | 939             | 967             | 940             | 975             | 947             | 984             | 956             |
| $10^3$     | LASSO  | 936             | 660             | 952             | 784             | 965             | 838             | 982             | 894             |
|            | SCAD   | 977             | 977             | 985             | 981             | 984             | 982             | 988             | 984             |
|            | MCP    | 961             | 940             | 968             | 944             | 975             | 945             | 984             | 961             |

Table 6: The result on model selection for $q = \frac{3}{2}$ and $n = 100$.  

| $\theta_0$ | Method | $r_{\text{nz}} = 0.2$ | $r_{\text{nz}} = 0.4$ | $r_{\text{nz}} = 0.6$ | $r_{\text{nz}} = 0.8$ |
|------------|--------|-----------------|-----------------|-----------------|-----------------|
|            |        | $d = 10$        | $d = 100$       | $d = 10$        | $d = 100$       |
| $10^0$     | LASSO  | 13              | 0               | 1               | 0               | 0               | 2               | 0               |
|            | SCAD   | 13              | 0               | 1               | 0               | 1               | 0               |
|            | MCP    | 16              | 0               | 1               | 0               | 1               | 0               |
| $10^1$     | LASSO  | 945             | 0               | 936             | 0               | 924             | 0               | 933             | 0               |
|            | SCAD   | 947             | 0               | 944             | 0               | 944             | 0               | 933             | 0               |
|            | MCP    | 945             | 264             | 947             | 19              | 947             | 0               | 933             | 0               |
| $10^2$     | LASSO  | 952             | 16              | 958             | 0               | 973             | 0               | 977             | 0               |
|            | SCAD   | 972             | 997             | 971             | 816             | 977             | 0               | 979             | 0               |
|            | MCP    | 965             | 999             | 969             | 887             | 976             | 7               | 978             | 0               |
| $10^3$     | LASSO  | 951             | 41              | 958             | 0               | 973             | 0               | 977             | 0               |
|            | SCAD   | 971             | 1,000           | 971             | 822             | 977             | 1               | 979             | 0               |
|            | MCP    | 965             | 1,000           | 969             | 889             | 976             | 4               | 978             | 0               |
Table 7: The result on model selection for $q = \frac{3}{2}$ and $n = 1,000$. 

| $\theta_0$ | Method | $r_{nz} = 0.2$ | $r_{nz} = 0.4$ | $r_{nz} = 0.6$ | $r_{nz} = 0.8$ |
|-------------|--------|----------------|----------------|----------------|----------------|
|             | $d = 10$ | $d = 100$ | $d = 10$ | $d = 100$ | $d = 10$ | $d = 100$ | $d = 10$ | $d = 100$ |
| $10^0$      | LASSO  | 7              | 0              | 1              | 0              | 1              | 0              | 0              |
|             | SCAD   | 8              | 0              | 1              | 0              | 2              | 0              | 0              |
|             | MCP    | 8              | 0              | 1              | 0              | 2              | 0              | 0              |
| $10^1$      | LASSO  | 971            | 735            | 967            | 539            | 978            | 376            | 990            | 419            |
|             | SCAD   | 971            | 726            | 966            | 543            | 978            | 421            | 990            | 501            |
|             | MCP    | 971            | 762            | 966            | 654            | 977            | 605            | 989            | 683            |
| $10^2$      | LASSO  | 994            | 963            | 996            | 968            | 996            | 975            | 999            | 973            |
|             | SCAD   | 994            | 994            | 998            | 994            | 998            | 994            | 999            | 996            |
|             | MCP    | 996            | 989            | 997            | 989            | 997            | 984            | 999            | 992            |
| $10^3$      | LASSO  | 994            | 963            | 996            | 967            | 996            | 975            | 999            | 974            |
|             | SCAD   | 998            | 995            | 998            | 993            | 998            | 994            | 999            | 996            |
|             | MCP    | 996            | 989            | 997            | 988            | 997            | 988            | 999            | 993            |

Table 8: The result on model selection for $q = \frac{5}{3}$ and $n = 100$. 

| $\theta_0$ | Method | $r_{nz} = 0.2$ | $r_{nz} = 0.4$ | $r_{nz} = 0.6$ | $r_{nz} = 0.8$ |
|-------------|--------|----------------|----------------|----------------|----------------|
|             | $d = 10$ | $d = 100$ | $d = 10$ | $d = 100$ | $d = 10$ | $d = 100$ | $d = 10$ | $d = 100$ |
| $10^0$      | LASSO  | 4              | 0              | 1              | 0              | 0              | 11             | 0              |
|             | SCAD   | 5              | 0              | 1              | 0              | 3              | 0              | 9              |
|             | MCP    | 4              | 0              | 1              | 0              | 3              | 0              | 5              |
| $10^1$      | SCAD   | 779            | 0              | 727            | 0              | 722            | 0              | 762            | 0              |
|             | MCP    | 791            | 12             | 739            | 0              | 745            | 0              | 779            | 0              |
| $10^2$      | LASSO  | 929            | 11             | 935            | 0              | 932            | 0              | 947            | 0              |
|             | SCAD   | 953            | 955            | 950            | 770            | 945            | 1              | 952            | 0              |
|             | MCP    | 953            | 956            | 951            | 857            | 953            | 9              | 949            | 0              |
| $10^3$      | LASSO  | 950            | 90             | 935            | 0              | 935            | 0              | 948            | 0              |
|             | SCAD   | 954            | 1,000          | 951            | 815            | 949            | 0              | 961            | 0              |
|             | MCP    | 954            | 1,000          | 954            | 893            | 937            | 3              | 959            | 0              |

Table 9: The result on model selection for $q = \frac{5}{3}$ and $n = 1,000$. 

| $\theta_0$ | Method | $r_{nz} = 0.2$ | $r_{nz} = 0.4$ | $r_{nz} = 0.6$ | $r_{nz} = 0.8$ |
|-------------|--------|----------------|----------------|----------------|----------------|
|             | $d = 10$ | $d = 100$ | $d = 10$ | $d = 100$ | $d = 10$ | $d = 100$ | $d = 10$ | $d = 100$ |
| $10^0$      | LASSO  | 2              | 0              | 2              | 0              | 2              | 0              | 5              |
|             | SCAD   | 2              | 0              | 2              | 0              | 2              | 0              | 0              |
|             | MCP    | 2              | 0              | 1              | 0              | 2              | 0              | 4              |
| $10^1$      | SCAD   | 810            | 81             | 790            | 22             | 729            | 18             | 794            | 35             |
|             | MCP    | 810            | 106            | 749            | 48             | 731            | 41             | 799            | 58             |
| $10^2$      | LASSO  | 981            | 984            | 984            | 984            | 986            | 975            | 994            | 956            |
|             | SCAD   | 981            | 987            | 984            | 983            | 986            | 977            | 994            | 971            |
|             | MCP    | 983            | 987            | 984            | 982            | 986            | 979            | 994            | 972            |
| $10^3$      | LASSO  | 983            | 988            | 986            | 987            | 986            | 981            | 995            | 969            |
|             | SCAD   | 986            | 989            | 989            | 983            | 990            | 981            | 995            | 974            |
|             | MCP    | 983            | 987            | 986            | 983            | 987            | 981            | 995            | 974            |
Table 10: The result on model selection for $q = 2$ and $n = 100$.

| $\theta_0$ | Method | $r_{nz} = 0.2$ | $r_{nz} = 0.4$ | $r_{nz} = 0.6$ | $r_{nz} = 0.8$ |
|------------|--------|----------------|----------------|----------------|----------------|
|            | $d = 10$ | $d = 100$ | $d = 10$ | $d = 100$ | $d = 10$ | $d = 100$ | $d = 10$ | $d = 100$ |
| $10^0$     | LASSO   | 3     | 0     | 2     | 0     | 0     | 0     | 2     | 0     |
|           | SCAD    | 3     | 0     | 2     | 0     | 0     | 0     | 2     | 0     |
|           | MCP     | 3     | 0     | 1     | 0     | 0     | 0     | 2     | 0     |
| $10^1$     | LASSO   | 81    | 0     | 57    | 0     | 53    | 0     | 58    | 0     |
|           | SCAD    | 83    | 0     | 58    | 0     | 50    | 0     | 54    | 0     |
|           | MCP     | 83    | 0     | 57    | 0     | 50    | 0     | 51    | 0     |
| $10^2$     | LASSO   | 821   | 0     | 777   | 0     | 745   | 0     | 732   | 0     |
|           | SCAD    | 824   | 314   | 782   | 219   | 765   | 2     | 735   | 0     |
|           | MCP     | 824   | 314   | 782   | 219   | 765   | 2     | 735   | 0     |
| $10^3$     | LASSO   | 961   | 29    | 935   | 0     | 895   | 0     | 879   | 0     |
|           | SCAD    | 966   | 915   | 944   | 746   | 913   | 1     | 879   | 0     |
|           | MCP     | 967   | 935   | 945   | 820   | 913   | 7     | 879   | 0     |

Table 11: The result on model selection for $q = 2$ and $n = 1,000$.

| $\theta_0$ | Method | $r_{nz} = 0.2$ | $r_{nz} = 0.4$ | $r_{nz} = 0.6$ | $r_{nz} = 0.8$ |
|------------|--------|----------------|----------------|----------------|----------------|
|            | $d = 10$ | $d = 100$ | $d = 10$ | $d = 100$ | $d = 10$ | $d = 100$ | $d = 10$ | $d = 100$ |
| $10^0$     | LASSO   | 1     | 0     | 1     | 0     | 0     | 0     | 1     | 0     |
|           | SCAD    | 3     | 0     | 1     | 0     | 0     | 0     | 1     | 0     |
|           | MCP     | 2     | 0     | 0     | 0     | 0     | 0     | 3     | 0     |
| $10^1$     | LASSO   | 5     | 0     | 1     | 0     | 1     | 0     | 1     | 0     |
|           | SCAD    | 5     | 0     | 1     | 0     | 1     | 0     | 1     | 0     |
|           | MCP     | 4     | 0     | 0     | 0     | 0     | 0     | 1     | 0     |
| $10^2$     | LASSO   | 518   | 226   | 476   | 162   | 451   | 132   | 494   | 123   |
|           | SCAD    | 519   | 226   | 476   | 162   | 453   | 142   | 494   | 145   |
|           | MCP     | 518   | 235   | 475   | 200   | 455   | 180   | 495   | 178   |
| $10^3$     | LASSO   | 933   | 910   | 936   | 883   | 941   | 845   | 926   | 779   |
|           | SCAD    | 933   | 910   | 936   | 883   | 942   | 841   | 925   | 840   |
|           | MCP     | 933   | 911   | 938   | 899   | 943   | 884   | 926   | 855   |

Table 12: The result on model selection for $q = 2.01$ and $n = 100$.

| $\theta_0$ | Method | $r_{nz} = 0.2$ | $r_{nz} = 0.4$ | $r_{nz} = 0.6$ | $r_{nz} = 0.8$ |
|------------|--------|----------------|----------------|----------------|----------------|
|            | $d = 10$ | $d = 100$ | $d = 10$ | $d = 100$ | $d = 10$ | $d = 100$ | $d = 10$ | $d = 100$ |
| $10^1$     | LASSO   | 1     | 0     | 1     | 0     | 0     | 0     | 1     | 0     |
|           | SCAD    | 2     | 0     | 0     | 0     | 0     | 0     | 1     | 0     |
|           | MCP     | 1     | 0     | 0     | 0     | 0     | 0     | 1     | 0     |
| $10^1$     | LASSO   | 84    | 0     | 44    | 0     | 40    | 0     | 48    | 0     |
|           | SCAD    | 86    | 0     | 45    | 0     | 40    | 0     | 49    | 0     |
|           | MCP     | 87    | 0     | 46    | 0     | 43    | 0     | 52    | 0     |
| $10^2$     | LASSO   | 834   | 1     | 793   | 0     | 775   | 0     | 773   | 0     |
|           | SCAD    | 835   | 314   | 792   | 210   | 774   | 0     | 772   | 0     |
|           | MCP     | 834   | 314   | 792   | 210   | 774   | 0     | 772   | 0     |
| $10^3$     | LASSO   | 971   | 22    | 961   | 0     | 937   | 0     | 921   | 0     |
|           | SCAD    | 972   | 894   | 961   | 737   | 936   | 1     | 927   | 0     |
|           | MCP     | 973   | 946   | 960   | 810   | 939   | 3     | 930   | 0     |
Table 13: The result on model selection for $q = 2.01$ and $n = 1,000$.

| $\theta_0$ | Method | $r_{\text{nz}} = 0.2$ | $r_{\text{nz}} = 0.4$ | $r_{\text{nz}} = 0.6$ | $r_{\text{nz}} = 0.8$ |
|------------|--------|------------------------|------------------------|------------------------|------------------------|
| $10^0$     | LASSO  | 2                      | 0                      | 0                      | 0                      |
|            | SCAD   | 3                      | 0                      | 0                      | 1                      |
|            | MCP    | 1                      | 0                      | 0                      | 1                      |
| $10^1$     | LASSO  | 6                      | 0                      | 1                      | 0                      |
|            | SCAD   | 6                      | 0                      | 1                      | 3                      |
|            | MCP    | 2                      | 0                      | 0                      | 3                      |
| $10^2$     | LASSO  | 493                    | 161                    | 447                    | 124                    |
|            | SCAD   | 493                    | 165                    | 445                    | 123                    |
|            | MCP    | 495                    | 174                    | 444                    | 144                    |
| $10^3$     | LASSO  | 931                    | 871                    | 928                    | 839                    |
|            | SCAD   | 931                    | 872                    | 929                    | 846                    |
|            | MCP    | 931                    | 875                    | 929                    | 861                    |

Table 14: The result on model selection for $q = 2.1$ and $n = 100$.

| $\theta_0$ | Method | $r_{\text{nz}} = 0.2$ | $r_{\text{nz}} = 0.4$ | $r_{\text{nz}} = 0.6$ | $r_{\text{nz}} = 0.8$ |
|------------|--------|------------------------|------------------------|------------------------|------------------------|
| $10^0$     | LASSO  | 2                      | 0                      | 0                      | 0                      |
|            | SCAD   | 2                      | 0                      | 0                      | 0                      |
|            | MCP    | 3                      | 0                      | 0                      | 3                      |
| $10^1$     | LASSO  | 6                      | 0                      | 0                      | 0                      |
|            | SCAD   | 6                      | 0                      | 0                      | 3                      |
|            | MCP    | 4                      | 0                      | 0                      | 3                      |
| $10^2$     | LASSO  | 619                    | 953                    | 512                    | 545                    |
|            | SCAD   | 618                    | 147                    | 554                    | 525                    |
|            | MCP    | 621                    | 139                    | 564                    | 537                    |
| $10^3$     | LASSO  | 931                    | 881                    | 864                    | 854                    |
|            | SCAD   | 931                    | 783                    | 890                    | 612                    |
|            | MCP    | 904                    | 778                    | 892                    | 652                    |

Table 15: The result on model selection for $q = 2.1$ and $n = 1,000$.

| $\theta_0$ | Method | $r_{\text{nz}} = 0.2$ | $r_{\text{nz}} = 0.4$ | $r_{\text{nz}} = 0.6$ | $r_{\text{nz}} = 0.8$ |
|------------|--------|------------------------|------------------------|------------------------|------------------------|
| $10^0$     | LASSO  | 6                      | 0                      | 0                      | 0                      |
|            | SCAD   | 6                      | 0                      | 0                      | 0                      |
|            | MCP    | 4                      | 0                      | 0                      | 0                      |
| $10^1$     | LASSO  | 118                    | 3                      | 82                     | 1                      |
|            | SCAD   | 118                    | 3                      | 84                     | 2                      |
|            | MCP    | 119                    | 4                      | 81                     | 3                      |
| $10^2$     | LASSO  | 759                    | 556                    | 744                    | 514                    |
|            | SCAD   | 758                    | 558                    | 741                    | 526                    |
|            | MCP    | 758                    | 560                    | 743                    | 544                    |

13
Table 16: The result on generalization for $q = 1$ and $n = 100$. Each value indicates the average generalization error among $m = 1000$ trials.

| $\theta$ | Method | $r_{\text{val}} = 0.2$ | $r_{\text{val}} = 0.4$ | $r_{\text{val}} = 0.6$ | $r_{\text{val}} = 0.8$ |
|----------|--------|---------------------|---------------------|---------------------|---------------------|
| 10^1     | LASSO  | 1.03 1.10 1.05 1.34 | 1.07 1.49 1.08 1.64 | 1.10 1.53 1.09 1.76 | 1.12 2.06 1.14 2.40 |
|          | SCAD   | 1.04 1.19 1.06 1.34 | 1.07 1.49 1.08 1.64 | 1.10 1.53 1.09 1.76 | 1.12 2.06 1.14 2.40 |
|          | MCP    | 1.05 1.20 1.06 1.34 | 1.07 1.49 1.08 1.64 | 1.10 1.53 1.09 1.76 | 1.12 2.06 1.14 2.40 |
| 10^3     | LASSO  | 1.07 1.55 1.09 2.53 | 1.11 3.78 1.11 4.76 | 1.15 4.92 1.16 5.92 | 1.18 6.01 1.18 7.01 |
|          | SCAD   | 1.04 1.27 1.06 1.86 | 1.08 3.05 1.10 3.54 | 1.15 4.92 1.16 5.92 | 1.18 6.01 1.18 7.01 |
|          | MCP    | 1.04 1.27 1.06 1.86 | 1.08 3.05 1.10 3.54 | 1.15 4.92 1.16 5.92 | 1.18 6.01 1.18 7.01 |
| 10^4     | LASSO  | 1.07 1.98 1.11 9.62 | 1.17 22.65 1.23 32.03 | 1.30 35.32 1.34 44.64 | 1.37 47.92 1.40 57.21 |
|          | SCAD   | 1.04 1.20 1.06 1.64 | 1.07 1.86 1.09 2.08 | 1.12 2.27 1.14 2.47 | 1.18 2.68 1.20 2.88 |
|          | MCP    | 1.04 1.20 1.06 1.64 | 1.07 1.86 1.09 2.08 | 1.12 2.27 1.14 2.47 | 1.18 2.68 1.20 2.88 |

Table 17: The result on generalization for $q = 1$ and $n = 1000$.

| $\theta$ | Method | $r_{\text{val}} = 0.2$ | $r_{\text{val}} = 0.4$ | $r_{\text{val}} = 0.6$ | $r_{\text{val}} = 0.8$ |
|----------|--------|---------------------|---------------------|---------------------|---------------------|
| 10^1     | LASSO  | 1.00 1.02 1.00 1.04 | 1.01 1.05 1.01 1.05 | 1.01 1.05 1.01 1.05 | 1.01 1.05 1.01 1.05 |
|          | SCAD   | 1.00 1.02 1.00 1.04 | 1.01 1.05 1.01 1.05 | 1.01 1.05 1.01 1.05 | 1.01 1.05 1.01 1.05 |
|          | MCP    | 1.00 1.02 1.00 1.04 | 1.01 1.05 1.01 1.05 | 1.01 1.05 1.01 1.05 | 1.01 1.05 1.01 1.05 |
| 10^2     | LASSO  | 1.00 1.02 1.00 1.04 | 1.01 1.05 1.01 1.05 | 1.01 1.05 1.01 1.05 | 1.01 1.05 1.01 1.05 |
|          | SCAD   | 1.00 1.02 1.00 1.04 | 1.01 1.05 1.01 1.05 | 1.01 1.05 1.01 1.05 | 1.01 1.05 1.01 1.05 |
|          | MCP    | 1.00 1.02 1.00 1.04 | 1.01 1.05 1.01 1.05 | 1.01 1.05 1.01 1.05 | 1.01 1.05 1.01 1.05 |
| 10^3     | LASSO  | 1.00 1.02 1.00 1.04 | 1.01 1.05 1.01 1.05 | 1.01 1.05 1.01 1.05 | 1.01 1.05 1.01 1.05 |
|          | SCAD   | 1.00 1.02 1.00 1.04 | 1.01 1.05 1.01 1.05 | 1.01 1.05 1.01 1.05 | 1.01 1.05 1.01 1.05 |
|          | MCP    | 1.00 1.02 1.00 1.04 | 1.01 1.05 1.01 1.05 | 1.01 1.05 1.01 1.05 | 1.01 1.05 1.01 1.05 |

Table 18: The result on generalization for $q = 13/11$ and $n = 100$.

| $\theta$ | Method | $r_{\text{val}} = 0.2$ | $r_{\text{val}} = 0.4$ | $r_{\text{val}} = 0.6$ | $r_{\text{val}} = 0.8$ |
|----------|--------|---------------------|---------------------|---------------------|---------------------|
| 10^1     | LASSO  | 1.27 1.46 1.29 1.63 | 1.31 1.73 1.32 1.81 | 1.33 1.92 1.34 2.01 | 1.35 2.10 1.36 2.19 |
|          | SCAD   | 1.27 1.47 1.29 1.63 | 1.31 1.73 1.32 1.81 | 1.33 1.92 1.34 2.01 | 1.35 2.10 1.36 2.19 |
|          | MCP    | 1.27 1.47 1.29 1.63 | 1.31 1.73 1.32 1.81 | 1.33 1.92 1.34 2.01 | 1.35 2.10 1.36 2.19 |
| 10^2     | LASSO  | 1.33 2.46 1.36 2.63 | 1.37 2.85 1.38 3.01 | 1.39 3.19 1.40 3.34 | 1.41 3.49 1.42 3.63 |
|          | SCAD   | 1.33 2.47 1.36 2.63 | 1.37 2.85 1.38 3.01 | 1.39 3.19 1.40 3.34 | 1.41 3.49 1.42 3.63 |
|          | MCP    | 1.33 2.47 1.36 2.63 | 1.37 2.85 1.38 3.01 | 1.39 3.19 1.40 3.34 | 1.41 3.49 1.42 3.63 |
| 10^3     | LASSO  | 1.33 2.46 1.36 2.63 | 1.37 2.85 1.38 3.01 | 1.39 3.19 1.40 3.34 | 1.41 3.49 1.42 3.63 |
|          | SCAD   | 1.33 2.47 1.36 2.63 | 1.37 2.85 1.38 3.01 | 1.39 3.19 1.40 3.34 | 1.41 3.49 1.42 3.63 |
|          | MCP    | 1.33 2.47 1.36 2.63 | 1.37 2.85 1.38 3.01 | 1.39 3.19 1.40 3.34 | 1.41 3.49 1.42 3.63 |
Table 19: The result on generalization for $q = 13/11$ and $n = 1,000$.

| $\theta_0$ | Method | $r_{nx} = 0.2$ | $r_{nx} = 0.4$ | $r_{nx} = 0.6$ | $r_{nx} = 0.8$ |
|------------|--------|----------------|----------------|----------------|----------------|
| $10^0$     | LASSO  | 1.25           | 1.27           | 1.26           | 1.29           |
|            | SCAD   | 1.25           | 1.27           | 1.26           | 1.29           |
|            | MCP    | 1.25           | 1.27           | 1.26           | 1.29           |
| $10^4$     | LASSO  | 1.26           | 1.32           | 1.26           | 1.35           |
|            | SCAD   | 1.25           | 1.28           | 1.26           | 1.30           |
|            | MCP    | 1.25           | 1.28           | 1.26           | 1.30           |
| $10^2$     | LASSO  | 1.25           | 1.32           | 1.26           | 1.35           |
|            | SCAD   | 1.25           | 1.32           | 1.26           | 1.35           |
|            | MCP    | 1.25           | 1.32           | 1.26           | 1.35           |
| $10^3$     | LASSO  | 1.25           | 1.28           | 1.26           | 1.30           |
|            | SCAD   | 1.25           | 1.28           | 1.26           | 1.30           |
|            | MCP    | 1.25           | 1.28           | 1.26           | 1.30           |

Table 20: The result on generalization for $q = 3/2$ and $n = 100$.

| $\theta_0$ | Method | $r_{nx} = 0.2$ | $r_{nx} = 0.4$ | $r_{nx} = 0.6$ | $r_{nx} = 0.8$ |
|------------|--------|----------------|----------------|----------------|----------------|
| $10^0$     | LASSO  | 2.87           | 3.29           | 2.89           | 3.50           |
|            | SCAD   | 2.87           | 3.29           | 2.89           | 3.50           |
|            | MCP    | 2.88           | 3.32           | 2.89           | 3.51           |
| $10^4$     | LASSO  | 3.00           | 4.05           | 3.07           | 5.82           |
|            | SCAD   | 2.94           | 4.47           | 3.01           | 6.94           |
|            | MCP    | 2.93           | 5.50           | 3.01           | 5.93           |
| $10^2$     | LASSO  | 3.01           | 23.33          | 3.08           | 103.52         |
|            | SCAD   | 2.91           | 3.72           | 2.96           | 46.11          |
|            | MCP    | 2.91           | 3.72           | 2.97           | 11.23          |
| $10^3$     | LASSO  | 5.00           | 19.73          | 5.98           | 9.900          |
|            | SCAD   | 2.91           | 5.72           | 2.96           | 10.23          |
|            | MCP    | 2.91           | 5.72           | 2.96           | 6.97           |

Table 21: The result on generalization for $q = 3/2$ and $n = 1,000$.

| $\theta_0$ | Method | $r_{nx} = 0.2$ | $r_{nx} = 0.4$ | $r_{nx} = 0.6$ | $r_{nx} = 0.8$ |
|------------|--------|----------------|----------------|----------------|----------------|
| $10^0$     | LASSO  | 2.92           | 2.92           | 2.92           | 2.94           |
|            | SCAD   | 2.92           | 2.92           | 2.92           | 2.94           |
|            | MCP    | 2.92           | 2.92           | 2.92           | 2.94           |
| $10^4$     | LASSO  | 2.93           | 3.05           | 2.94           | 3.12           |
|            | SCAD   | 2.93           | 3.05           | 2.94           | 3.12           |
|            | MCP    | 2.93           | 3.05           | 2.94           | 3.05           |
| $10^2$     | LASSO  | 2.93           | 3.05           | 2.94           | 3.12           |
|            | SCAD   | 2.93           | 3.05           | 2.94           | 3.12           |
|            | MCP    | 2.93           | 3.05           | 2.94           | 3.05           |
| $10^3$     | LASSO  | 2.93           | 3.05           | 2.94           | 3.12           |
|            | SCAD   | 2.93           | 3.05           | 2.94           | 3.12           |
|            | MCP    | 2.93           | 3.05           | 2.94           | 3.05           |
Table 22: The result on generalization for \( q = 5/3 \) and \( n = 100 \).

| \( \theta_0 \) | Method | \( r_{nk} = 0.2 \) | \( d = 10 \) | \( d = 100 \) | \( r_{nk} = 0.4 \) | \( d = 10 \) | \( d = 100 \) | \( r_{nk} = 0.6 \) | \( d = 10 \) | \( d = 100 \) | \( r_{nk} = 0.8 \) | \( d = 10 \) | \( d = 100 \) |
|--------------|-------|-----------------|-----------|-----------|-----------------|-----------|-----------|-----------------|-----------|-----------|-----------------|-----------|-----------|
| \( 10^0 \)   | LASSO | 10.68           | 14.07     | 10.71     | 14.26          | 10.72     | 14.46     | 10.74           | 14.65     |         |               |           |           |
|              | SCAD  | 10.68           | 14.07     | 10.71     | 14.26          | 10.72     | 14.47     | 10.74           | 14.66     |         |               |           |           |
|              | MCP   | 10.69           | 14.09     | 10.71     | 14.29          | 10.72     | 14.49     | 10.74           | 14.69     |         |               |           |           |
| \( 10^1 \)   | LASSO | 11.04           | 17.72     | 11.23     | 19.45          | 11.32     | 21.00     | 11.40           | 22.28     |         |               |           |           |
|              | SCAD  | 10.97           | 18.85     | 11.18     | 21.89          | 11.31     | 22.62     | 11.41           | 25.29     |         |               |           |           |
|              | MCP   | 10.97           | 19.12     | 11.19     | 22.11          | 11.33     | 23.14     | 11.42           | 24.98     |         |               |           |           |
| \( 10^2 \)   | LASSO | 11.32           | 37.31     | 11.64     | 120.93         | 11.74     | 245.48    | 11.78           | 346.97    |         |               |           |           |
|              | SCAD  | 11.00           | 18.13     | 11.31     | 38.66          | 11.51     | 149.72    | 11.65           | 199.90    |         |               |           |           |
|              | MCP   | 11.00           | 18.05     | 11.30     | 27.55          | 11.51     | 107.95    | 11.66           | 144.35    |         |               |           |           |
| \( 10^3 \)   | LASSO | 11.31           | 197.49    | 11.65     | 100.79         | 11.77     | 224.66    | 11.89           | 320.98    |         |               |           |           |
|              | SCAD  | 10.89           | 16.41     | 11.18     | 16.27          | 11.39     | 125.26    | 11.66           | 174.80    |         |               |           |           |
|              | MCP   | 10.90           | 16.41     | 11.18     | 26.64          | 11.89     | 88.13     | 11.61           | 120.66    |         |               |           |           |

Table 23: The result on generalization for \( q = 5/3 \) and \( n = 1,000 \).

| \( \theta_0 \) | Method | \( r_{nk} = 0.2 \) | \( d = 10 \) | \( d = 100 \) | \( r_{nk} = 0.4 \) | \( d = 10 \) | \( d = 100 \) | \( r_{nk} = 0.6 \) | \( d = 10 \) | \( d = 100 \) | \( r_{nk} = 0.8 \) | \( d = 10 \) | \( d = 100 \) |
|--------------|-------|-----------------|-----------|-----------|-----------------|-----------|-----------|-----------------|-----------|-----------|-----------------|-----------|-----------|
| \( 10^0 \)   | LASSO | 25.40           | 12.64     | 25.41     | 12.65          | 25.41     | 12.67     | 25.41           | 12.69     |         |               |           |           |
|              | SCAD  | 25.40           | 12.64     | 25.41     | 12.65          | 25.41     | 12.67     | 25.41           | 12.69     |         |               |           |           |
|              | MCP   | 25.40           | 12.64     | 25.41     | 12.65          | 25.41     | 12.67     | 25.41           | 12.69     |         |               |           |           |
| \( 10^1 \)   | LASSO | 25.45           | 13.08     | 25.48     | 13.32          | 25.49     | 13.49     | 25.50           | 13.59     |         |               |           |           |
|              | SCAD  | 25.45           | 13.07     | 25.48     | 13.34          | 25.50     | 13.50     | 25.50           | 13.63     |         |               |           |           |
|              | MCP   | 25.45           | 13.08     | 25.47     | 13.35          | 25.50     | 13.53     | 25.51           | 13.64     |         |               |           |           |
| \( 10^2 \)   | LASSO | 25.48           | 13.25     | 25.53     | 13.57          | 25.57     | 13.75     | 25.58           | 13.84     |         |               |           |           |
|              | SCAD  | 25.45           | 12.90     | 25.49     | 13.17          | 25.54     | 13.42     | 25.57           | 13.66     |         |               |           |           |
|              | MCP   | 25.45           | 12.90     | 25.50     | 13.18          | 25.54     | 13.41     | 25.57           | 13.66     |         |               |           |           |
| \( 10^3 \)   | LASSO | 25.52           | 13.25     | 25.56     | 13.57          | 25.58     | 13.79     | 25.65           | 14.00     |         |               |           |           |
|              | SCAD  | 25.44           | 12.86     | 25.47     | 13.11          | 25.50     | 13.37     | 25.59           | 13.63     |         |               |           |           |
|              | MCP   | 25.44           | 12.86     | 25.47     | 13.11          | 25.50     | 13.37     | 25.59           | 13.63     |         |               |           |           |
5 Conclusion

We proposed regularization methods for \(q\)-normal linear models based on the \(L_q\)-likelihood. The proposed methods coincide with the ordinary regularization methods. Our methods perform well for slightly heavy-tailed errors \((1 \leq q < 2)\) in terms of model selection and generalization. Moreover, they work well in terms of model selection for heavy-tailed errors \((2 \leq q < 2.5)\). A theoretical analysis of the proposed methods is left to future work.

References

[1] S. Ejaz Ahmed, Hwanwoo Kim, Gökhan Yildirim, and Bahadir Yüzbasi. High-dimensional regression under correlated design: An extensive simulation study. In *Matrices, Statistics and Big Data -Selected Contributions from IWMS 2016*, pages 145–175, 2019.

[2] Ayman Alzaatreh, Carl Lee, Felix Famoye, and Indranil Ghosh. The generalized cauchy family of distributions with applications. *Journal of Statistical Distributions and Applications*, 3:12, 2016.

[3] Nikoletta Bassiou, Constantine Kotropoulos, and Evangelia Koliopoulou. Symmetric \(\alpha\)-stable sparse linear regression for musical audio denoising. In *8th International Symposium on Image and Signal Processing and Analysis (ISPA 2013)*, pages 382–387, 2013.

[4] Emmanuel Candes and Terence Tao. The Dantzig selector: Statistical estimation when \(p\) is much larger than \(n\). *Annals of Statistics*, 36:2313–2351, 2007.

[5] Rafael E. Carrillo, Tuncer C. Aysal, and Kenneth E. Barner. Generalized cauchy distribution based robust estimation. In *International Conference on Acoustics, Speech, and Signal Processing (ICASSP) 2008*, pages 3389–3392, 2008.

[6] Rafael E. Carrillo, Tuncer C. Aysal, and Kenneth E. Barner. A generalized cauchy distribution framework for problems requiring robust behavior. *EURASIP Journal on Advances in Signal Processing*, 2010:1–19, 2010.

[7] Bradley Efron, Trevor Hastie, Ian Johnstone, and Robert Tibshirani. Least angle regression. *Annals of Statistics*, 32:407–499, 2004.

[8] Jianqing Fan and Runze Li. Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, 96:1348–1360, 2001.

[9] Davide Ferrari and Yuhong Yang. Maximum \(l_q\)-likelihood estimation. *Annals of Statistics*, 38:753–783, 2010.
[10] Shigeru Furuichi. On the maximum entropy principle and the minimization of the Fisher information in Tsallis statistics. *Journal of Mathematical Physics*, 50:013303, 2009.

[11] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning (2nd Edition)*. Springer, 2009.

[12] Melvin J. Hinich and Prem P. Talwar. A simple method for robust regression. *Journal of the American Statistical Association*, 70:113–119, 1975.

[13] Paul W. Holland and Roy E. Welsch. Robust regression using iteratively reweighted least-squares. *Communications in Statistics - Theory and Methods*, 6:813–827, 1977.

[14] K. R. Kadiyala and K. S. R. Murthy. Estimation of regression equation with Cauchy disturbances. *Canadian Journal of Statistics*, 5:111–120, 1977.

[15] Mee Young Park and Trevor Hastie. $l_1$-regularization path algorithm for generalized linear models. *Journal of the Royal Statistical Society, Series B*, 69:659–677, 2007.

[16] Domingo Prato and Constantino Tsallis. Nonextensive foundation of lévy distributions. *Physical Review E*, 60:2398–2401, 2000.

[17] V. Kerry Smith. Least squares regression with Cauchy errors. *Oxford Bulletin of Economics and Statistics*, 35:223–231, 1973.

[18] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B*, 58:267–288, 1996.

[19] Constantino Tsallis. *Introduction to Nonextensive Statistical Mechanics*. Springer, 2009.

[20] Cun-Hui Zhang. Nearly unbiased variable selection under minimax concave penalty. *Annals of Statistics*, 38:894–942, 2010.