Journal of Applied Statistics

J Appl Stat. 2021; 48(12): 2205–2230.
Published online 2020 Jun 30. doi: 10.1080/02664763.2020.1787355

PMCID: PMC9041692
PMID: 35706613

An elastic-net penalized expectile regression with applications

Q.F. Xu,[a,b] X.H. Ding,[a] C.X. Jiang,[a] K.M. Yu,[c] and L. Shi[d]

Abstract

To perform variable selection in expectile regression, we introduce the elastic-net penalty into expectile regression and propose an elastic-net penalized expectile regression (ER-EN) model. We then adopt the semismooth Newton coordinate descent (SNCD) algorithm to solve the proposed ER-EN model in high-dimensional settings. The advantages of ER-EN model are illustrated via extensive Monte Carlo simulations. The numerical results show that the ER-EN model outperforms the elastic-net penalized least squares regression (LSR-EN), the elastic-net penalized Huber regression (HR-EN), the elastic-net penalized quantile regression (QR-EN) and conventional expectile regression (ER) in terms of variable selection and predictive ability, especially for asymmetric distributions. We also apply the ER-EN model to two real-world applications: relative location of CT slices on the axial axis and metabolism of tacrolimus (Tac) drug. Empirical results also demonstrate the superiority of the ER-EN model.

Keywords: Expectile regression, elastic-net, SNCD, variable selection, high-dimensional data

Classification codes: 62J05

1. Introduction

In statistical modeling, regression analysis is an effective tool for exploring the relationship among variables. According to loss functions used in optimization, we distinguish three types of regressive methods including: (1) mean regression (or OLS) based on the quadratic loss $l(u) = u^2$, (2) quantile regression based on the asymmetric absolute loss $l(u) = u \cdot (\tau - I(u < 0))$, and (3) expectile regression based on the asymmetric quadratic loss $l(u) = u^2 \cdot |\theta - I(u < 0)|$. Both τ and θ determine the degree of asymmetry of the loss function. Expectile regression proposed in [1,51] is an extension of the classical linear regression model that allows for different weights on positive and negative residuals called asymmetric least-squares regression. This estimator shares good properties such as homogeneity, monotonicity with respect to weight level θ and translation invariance, see [51].

In mean regression problems, we are interested in finding important covariates to predict the response variable. With the study of the high-dimensional problems being enthusiastic, regression model with variable selection is very valuable in practice. For example, it is very important to consider high-dimensional regression in X-ray tomography and medical clinical trials field, etc. Several classical penalized methods, including the least absolute shrinkage and selection operator (LASSO, L_1) of [65], the smoothly clipped absolute deviation (SCAD) of [22], the elastic-net penalty of [93], the adaptive LASSO of [92], and the minimax concave penalty (MCP) of [88], are widely used in variable selection. In many practice, we need to select groups of variables rather than individual variables. To this end, there emerges variable selection in a group manner or group variable selection methods, such as the group LASSO of [87] and the adaptive group LASSO of [72]. [53] introduce a novel variant of such method called the adaptive simultaneous variable selection (SVS), which is closely linked with the adaptive group LASSO. In addition, two MCP-based penalization approaches are proposed by [49] for marker selection under the heterogeneity models. [4,5] present a Bayesian stochastic search variable selection method for subset selection based on a heteroscedastic linear regression model.

As an alternative to mean regression, quantile regression proposed by [39] can reveal the complete distribution of a response conditional on covariates information, which provides more useful information for decision making. Since then, quantile regression has drawn more and more attention from both academics and practitioners. There has been a large number of classical models, such as the local linear quantile regression of [86], the single-index quantile regression of [77], the composite quantile regression of [94], the CAViaR of [20], the exponentially weighted quantile regression of [64], the dynamic additive quantile (DAQ) model of [29], and the quantile regression on panel data in [42]. To solve high-dimensional problems, penalized methods have also been introduced into quantile regression. For example, L_1 penalized quantile regression model is developed in [8,40,42,46,79] via the LASSO approach. In [6,54,58,76], variable selection for quantile regression is implemented by the SCAD and adaptive LASSO penalties. In [10], the elastic-net method is used to identify important covariates for quantile regression. In [50], both the convex (e.g. LASSO and elastic-net) and nonconvex penalties (e.g. SCAD and MCP) are considered. Under the Bayesian framework, regularized quantile regression is conducted with the LASSO, group LASSO and the elastic-net methods in [44], with the adaptive LASSO in [2,3]. As for nonlinear quantile regressions, [43] propose a kernel quantile regression, where the kernel trick is used twice, one is for nonlinear fitting and the other is for smoothness penalty. [11] develops a quantile regression neural networks (QRNN) model, where large weights are penalized using a quadratic penalty term. In addition, additive quantile regression models are estimated by the LASSO and total variation roughness penalties in [38], the two-fold SCAD penalty in [47], and the boosting method in [24]. Recently, wild residual bootstrap inference is proposed by [71] for penalized quantile regression.

Similar to quantile regression, expectile regression is also able to investigate heterogeneous effects of covariates on the distribution of the response variable and provide its complete distribution information. According to [36,84], there exists an one-to-one mapping from expectiles to quantiles, i.e. for each θ-th expectile, there is a corresponding τ-th quantile. The attention devoted to expectile regression has increased substantially in recent years. In [70], they find that expectile regression tends to have less crossing and be more robust for heavy-tailed distribution. In [7,13,15,64], they relate the expectile to the expected shortfall (ES) and develop the conditional autoregressive expectile (CARE) model to measure ES risk. Moreover, [41] apply the CARE model to measure VaR risk, [78] develop a varying-coefficient expectile model for estimating VaR, and [32] propose a dynamic autoregressive expectile model for portfolio selection. Besides point prediction, forecast evaluations on expectile regression are comprehensively studied in [19,31]. For applications, a multivariate expectile regression model has been successfully applied by [12] to analyze the tail events of large cross-sectional and spatial data. In nonlinear situations, [37] consider a type of nonlinear expectile regression models and study the consistency and asymptotic normality of the estimates, a geoadditive expectile regression is considered by [60,61,67,68], an expectile regression neural network (ERNN) model is developed by [35] through adding a neural network structure to the expectile regression approach, a nonparametric CARE model is advanced by [80] via neural network, a flexible nonparametric expectile regression is considered by [81] using the kernel trick, and [23] propose an SVM-like expectile regression. Recently, expectile regression on high-dimensional data has received increasing attention and many penalized expectile regression methods have been developed. [74] introduce an L_0 penalty into an expectile regression with a Whittaker smoother to better model accelerometer data. [14,30,48] consider the sparse expectile regression under high dimensions with the LASSO (L_1 penalty), adaptive LASSO, and nonconvex penalties. In [89,91], they propose a penalized linear expectile regression with the SCAD penalty and show the oracle properties of the estimator. Moreover, [90] develop a partially linear additive expectile regression using nonconvex penalties such as SCAD and MCP. In [57,63], a semi-parametric expectile regression with penalized splines is developed to fit smooth expectile curves. Furthermore, [62] apply the semi-parametric expectile regression with penalized splines to model spatio-temporal effects beyond the mean, facilitating the interpretation of main effects and interactions between space and time.

We consider the elastic-net penalty of [93] as regularization to solve a major limitation of the expectile regression in that it cannot automatically select relevant variables. The elastic-net penalty is a convex combination of the LASSO penalty of [65] and the ridge penalty of [33], i.e. $L_1 + L_2$. As a penalization technique, it combines the strengths of the LASSO penalty and the ridge penalty, and effectively solves both continuous shrinkage and automatic variable selection simultaneously. The main difference between elastic-net penalty and SCAD penalty is that elastic-net penalty is convex, which has some advantages in calculation and it is particularly useful in the $p>n$ situation, or any situation where there are many correlated predictors. For example, in [16], the elastic-net penalty is added on the loading vectors of principal component analysis to identify the few stable regions in brain images. [45] propose an iterative algorithm to solve the generalized elastic-net regularization problem. [5] propose the Bayesian elastic-net for single index quantile regression for estimation and variables selection. [73] propose the weighted elastic-net to deal with the variable selection of generalized linear models on high-dimensional data and give the theoretical properties of the proposed method with a diverging number of parameters.

It is well known that the pathwise optimization is a primary scheme to solve penalized regression models. The classical algorithm is the Least Angle Regression (LARS) of [18], which is based on the residual error of the target. However, the LARS is extremely sensitive to the sample noise. As an alternative, the coordinate descent (CD) algorithm of [66] performs better in the pathwise optimization problem, for example, LASSO penalized least squares in [25], elastic-net penalized generalized linear models (GLM) in [26], and nonconvex penalized least squares and logistic regression in [9]. Unfortunately, the coordinate descent algorithm needs a longer computing time especially in high-dimensional data, see [82,83]. Furthermore, superlinear convergence of a Newton-type method is established by [55] for solving finite-dimensional nonsmooth equations, and hereinafter called semismooth Newton algorithm (SN). Moreover, [85] combine the strengths of the coordinate descent algorithm (CD) and the semismooth Newton algorithm (SN), and develop a semismooth Newton coordinate descent (SNCD) algorithm. The superiority of SNCD to CD lies in two folds, refer to [9,26,59], among others. First, the SNCD algorithm updates a regression coefficient and its corresponding subgradient simultaneously in each iteration, which makes it run much faster in practice. Second, the SNCD algorithm generalizes coordinate descent to work on a wider class of models where the loss functions, like the Huber loss regression of [34], only need to be first-order differentiable. In general, the SNCD algorithm always converges quickly and outperform the CD algorithm, especially in high-dimensional cases.

In this paper, we propose a novel elastic-net penalized expectile regression (ER-EN) model. The ER-EN model is similar to but different from that of [85] in terms of the used loss functions. Actually, in [85], they use the Huber loss and the asymmetric absolute loss to do mean regression and quantile regression. However, we consider an asymmetric quadratic loss, which yields expectile regression. To a certain extent, the asymmetric quadratic loss can be viewed as a combination of the asymmetric absolute loss and the (symmetric) quadratic loss. In this regard, we further introduce the elastic-net penalty into the asymmetric quadratic loss to construct the ER-EN model. This is a major innovation of the paper. In terms of statistical properties, the quantile regression only considers the position (or order) relationship in the data without considering the distance, while the expectile regression can take into account both of them. The quadratic loss function brings computational advantages and estimation effectiveness to expectile regression, where the estimation of the covariance matrix does not need to estimate a conditional density function. Thus expectile regression has more excellent properties than quantile regression, and we focus on expectile regression here. In addition, the ER-EN model also differs from [83] in two aspects. First, the multiple expectile regression of [83] is a nonparametric model, while our ER-EN is a parametric one. Second, the nonparametric multiple expectile regression model is solved by a regression tree-based gradient boosting estimator named as ER-Boost, while the ER-EN model is solved via the SNCD algorithm for its computational advantages in high-dimensional settings.

This ER-EN model is highly efficient and scalable in high-dimensional settings, and has at least three advantages. First, since it uses an asymmetric quadratic loss function, the ER-EN model can describe complete distribution of a response conditional on covariates information. Second, the ER-EN model is suitable for data in the presence of outliers or heterogeneity, has excellent stability and robustness in high-dimensional data applications particularly. Third, the ER-EN model is faster in computing, which is a vital merit of the ER-EN model when dealing with high-dimensional data. We first compare the ER-EN model with other models, including the elastic-net penalized least squares regression (LSR-EN), the elastic-net penalized Huber regression (HR-EN), the elastic-net penalized quantile regression (QR-EN) and conventional expectile regression (ER), and illustrate the advantage of the proposed ER-EN model through extensive Monte Carlo simulations. The numerical results show that the ER-EN model outperforms the others in terms of variable selection and predictive ability, especially for asymmetric distributions. We then apply the ER-EN model in two real-world applications to predict relative location of CT slices on the axial axis and metabolism of tacrolimus (Tac) drug. The empirical results also indicate that the ER-EN model is superior to several popular models for processing high-dimensional data in the medical field. The ER-EN model can not only select important variables to enhance the interpretability of the response but also improve the prediction accuracy of related diseases.

This paper proceeds as follows. Section 2 briefly reviews some elastic-net penalized linear regression. Section 3 presents a novel ER-EN model and discusses its modeling techniques. Simulation experiments are conducted in Section 4 to illustrate the efficacy of the proposed the ER-EN model. In Section 5, we consider real applications and report the empirical results of the ER-EN model. Section 6 summarizes and concludes the paper.

2. Penalized linear regression review

In this section, we briefly review the classical linear regression and its penalized versions in high-dimensional settings.

2.1. Linear regression

In regression analysis, the classical linear model relates y to p-dimensional $\boldsymbol{x}$ with

$$y = \beta_0 + \boldsymbol{x}_i^T \beta + \varepsilon, \tag{1}$$

where $y \in R$, $\boldsymbol{x} \in R^p$, $\beta \equiv \left(\beta_1, \beta_2, \ldots, \beta_p\right)^T \in R^p$, and ϵ is the random error.

The regression coefficients in β can be estimated through optimizing the following empirical loss or cost

$$\min_{\beta} \frac{1}{N} \sum_{i}^{N} l\left(y_i - \beta_0 - \boldsymbol{x}_i^T \beta\right), \tag{2}$$

where $l(\cdot)$ a generic loss function. In practice, commonly used loss functions include: (1) the quadratic loss $l\left(u\right) = u^2$ with $-\infty < u < \infty$, which yields mean regression or least squares regressions; (2) the asymmetric absolute loss $l(u) = u \cdot (\tau - I(u < 0))$ with $0 < \tau < 1$, which yields quantile regressions; (3) the asymmetric quadratic loss $l\left(u\right) = u^2 \cdot \left|\theta - I\left(u < 0\right)\right|$ with $0 < \theta < 1$, which yields expectile regressions; (4) the Huber loss $l\left(u\right) = \begin{cases} \frac{u^2}{2\gamma}, & \text{if} \quad |u| \le \gamma \\ \left|u\right| - \frac{\gamma}{2}, & \text{if} \quad |u| > \gamma \end{cases}$ with a threshold γ and $(-\infty < u < \infty)$, which yields Huber regression.

As we all know, mean regression is an efficient tool and estimates the conditional expectation of dependent variable given the values of independent variables. However, it is not suitable for data in the presence of outliers or heterogeneity in practice, see [13,17,56,75]. By contrast, quantile regression as well as expectile regression can provide comprehensive information about conditional distribution. In addition, the asymmetric quadratic loss function of expectile regression has the nature of continuity and smoothness and yields computational advantages. Huber regression is the same as mean regression for small residuals, but allows large residuals as the Huber loss provides a smooth transition between absolute and quadratic errors [11]. For this reason, Huber regression has attracted considerable attentions, see [52,80].

2.2. Penalized linear regression

In the high-dimensional case where $p>n$, we should consider variable selection of the model since it is often sparse with small proportion of nonzero coefficients. It is therefore important to identify and estimate the nonzero coefficients. As an alternative way to subset selection and information criteria, the penalized method is able to perform variable selection and model estimation simultaneously.

For a generic penalty function $P(\cdot)$, we consider a penalized regression

$$\min_{\beta} \frac{1}{N} \sum_{i}^{N} l\left(y_i - \beta_0 - \boldsymbol{x}_i^T \beta\right) + \lambda P\left(\beta\right), \quad (3)$$

where $\lambda \geq 0$ is the tuning parameter. The choice of penalty functions is important in the penalized regression. When using the ridge penalty of [33], regression coefficients are shrunken towards zero, but will never become exactly zero. Thus the ridge penalty will not provide a sparse model that is easy to interpret, refer to [26]. The LASSO penalty of [65] reduces the variability of the estimates by shrinking the coefficients and at the same time produces interpretable models by shrinking some coefficients to exactly zero. The SCAD penalty of [22] is symmetric and nonconvex, which can produce sparse solutions. However, its iterative algorithm runs slowly, which is not suitable for the situation of large amount of data. The MCP penalty of [88] has minimax optimality. Just like SCAD, the MCP penalty poses a challenge to the numerical calculation in fitting model because of its nonconvexity. The elastic-net penalty of [93] combines the advantages of LASSO penalty and ridge penalty, which can effectively solve the problems of continuous contraction and automatic variable selection. The elastic-net penalty is particularly useful when the number of predictors (p) is much bigger than the number of observations (n).

Given the above discussion, we use the elastic-net penalty function with the form

$$P\left(\beta\right) \equiv P_{\alpha}\left(\beta\right) = \alpha \parallel \beta \parallel_1 + \left(1-\alpha\right)\frac{1}{2} \parallel \beta \parallel_2^2, \quad 0 \leq \alpha \leq 1. \quad (4)$$

which is a convex combination of the LASSO penalty ($\alpha = 1$) and the ridge penalty ($\alpha = 0$). In Equation (3), when using the (symmetric) quadratic loss and the elastic-net penalty, we obtain the elastic-net penalized least squares regression (LSR-EN)

$$\min_{\beta} f_s\left(\beta\right) = \frac{1}{2N}\sum_{i}^{N}\left(y_i - \beta_0 - \boldsymbol{x}_i^T\beta\right)^2 + \lambda\left(\alpha \parallel \beta \parallel_1 + \left(1-\alpha\right)\frac{1}{2} \parallel \beta\Big\|_2^2\right), \quad 0 \le \alpha \le 1. \tag{5}$$

When using the Huber loss and the elastic-net penalty, we obtain the elastic-net penalized Huber regression (HR-EN)

$$\min_{\beta} f_H\left(\beta\right) = \frac{1}{N}\sum_{i}^{N} h_\gamma\left(y_i - \beta_0 - \boldsymbol{x}_i^T\beta\right) + \lambda\left(\alpha \parallel \beta \parallel_1 + \left(1-\alpha\right)\tfrac{1}{2} \parallel \beta\|_2^2\right), \quad 0 \le \alpha \le 1. \tag{6}$$

When using the quantile loss and the elastic-net penalty, we obtain the elastic-net penalized quantile regression (QR-EN)

$$\min_{\beta} f_Q\left(\beta\right) = \frac{1}{N}\sum_{i}^{N} \rho_\theta\left(y_i - \beta_0 - \boldsymbol{x}_i^T\beta\right) + \lambda\left(\alpha \parallel \beta \parallel_1 + \left(1-\alpha\right)\tfrac{1}{2} \parallel \beta\|_2^2\right), \quad 0 \le \alpha \le 1. \tag{7}$$

As for regression equations and mathematical properties, there are many similarities between quantile regression and expectile regression. In addition, the objective function of expectile regression is an asymmetric quadratic loss function, which is continuous and smooth, then the optimization and calculation are more convenient than quantile regression. In this regard, we intend to study expectile regression with the elastic-net penalty in the next section.

3. Elastic-net penalized linear expectile regression

In this section, we introduce a novel ER-EN model. We then provide its estimation method via the SNCD algorithm. In addition, we discuss model selection and model evaluation for the ER-EN model.

3.1. Model setup

Consider the asymmetric quadratic loss

$$l\left(u\right) \equiv \rho_\theta\left(u\right) = u^2 \cdot \left|\theta - I\left(u < 0\right)\right|, \tag{8}$$

where $I(\cdot)$ is the indicator function, $\theta \in (0,1)$ is the weight on positive and negative residuals, which determines the extent of asymmetry of the loss function, with $\theta = 0.5$ denoting median or expectation value. It is not hard to find that the quadratic loss is a special case of the asymmetric quadratic loss function when $\theta = 0.5$.

Now the asymmetric quadratic loss is used to expectile regression (ER) in [51]. Suppose that we have observations $(y_i, x_{1i}, x_{2i}, \ldots, x_{pi})$ for $i = 1, 2, \ldots, N$, the ER estimators are obtained by minimizing the empirical loss

$$L\left(\theta\right) = \frac{1}{N}\sum_{i}^{N} \rho_\theta\left(y_i - \beta_0 - \boldsymbol{x}_i^T \beta\left(\theta\right)\right)^2, \tag{9}$$

where $\beta(\theta) \equiv \left[\beta_1(\theta), \beta_2(\theta), \ldots, \beta_p(\theta)\right]^T$ is a conformable vector of parameters related to θ, for fixed values of θ in $(0,1)$.

[51] prove that the ER estimator has good properties and asymptotic normal distribution. ER models the conditional expectile rather than the conditional mean of the response given the values of covariates. For heterogeneous data, the functional relationship between the response and the covariates may vary in different segments of the conditional distribution. By choosing different θ, ER provides a powerful technique for exploring data heterogeneity in addition to outlier robustness.

In the regularization and variable selection framework, we add the elastic-net penalty into the empirical loss in Equation (9) to obtain the ER-EN model

$$\min_{\beta_0, \beta} f_E\left(\beta\right) = \frac{1}{N}\sum_i^N \rho_\theta\left(y_i - \beta_0 - \boldsymbol{x}_i^T \beta\right)^2 + \lambda\left(\alpha \parallel \beta \parallel_1 + (1-\alpha)\frac{1}{2} \parallel \beta \parallel_2^2\right), \quad 0 \le \alpha \le 1, \tag{10}$$

where λ is the tuning parameter. It is obvious that ER-LASSO and ER-ridge are two special cases of the ER-EN when $\alpha = 1$ and $\alpha = 0$, respectively.

3.2. Model estimation

In Equation (10), the objective function is composed of two parts: an expectile loss and an elastic-net penalty. The expectile loss is convex and differentiable, see [91]. In addition, the elastic-net penalty of [93] is a convex function since it is a convex combination of the LASSO and ridge penalty. Thus the objective function of Equation (10) is also a convex one.

In [85], they study the elastic-net penalized Huber loss regression. To solve the model, a semismooth Newton coordinate descent (SNCD) algorithm is developed. The SNCD is demonstrated to be a convergence and efficient algorithm to ultra-high dimensions.

It is well known that the Huber loss is the combination of a quadratic loss and an absolute one. In particular, when the γ approaches to infinity, the Huber loss degenerates into the quadratic loss. Additionally, the expectile loss is actually an asymmetric quadratic loss, which shares some properties (e.g. differential at the origin) of the (symmetric) quadratic loss. In this regard, it is nature to use the SNCD algorithm to solve our ER-EN model defined in Equation (10). We present its solution procedure as below.

Step 1, set $s = 0$ and initialize $\left(\beta_0^{(s)}, \beta_1^{(s)}, \beta_2^{(s)}, \ldots, \beta_p^{(s)}\right)$.

Step 2, implement SNCD to sequentially update β_0 and β_j for $j = 1, 2, \ldots, p$,

1. using $\beta_0^{(s+1)} = \arg\min_{\beta_0} f_E\left(\beta_0, \beta_1^{(s)}, \beta_2^{(s)}, \ldots, \beta_p^{(s)}\right)$;
2. using $\beta_j^{(s+1)} = \arg\min_{\beta_j} f_E\left(\beta_0^{(s+1)}, \ldots, \beta_{j-1}^{(s+1)}, \beta_j, \beta_{j+1}^{(s)}, \ldots, \beta_p^{(s)}\right)$.

Step 3, let $s + 1 \to s$ and repeat the step 2 until convergence.

3.3. Model selection

There are two tuning parameters, α and λ, in ER-EN model. For α, the status $\alpha = 1$ corresponds to the LASSO penalty and $\alpha = 0$ to the ridge penalty. The larger the α, the stronger the variable selection ability, and fewer variables to be selected. Conversely, the weaker the ability, more variables selected in the model. We consider three representatives $\alpha = 0.1, 0.5, 0.9$ for discussion. Due to space constraints, we only report the results of $\alpha = 0.9$ here, which is in consistent with [85].

For λ, in the ER-EN model, in order to ensure global convergence and actually implement the algorithm, selecting a good tuning parameter λ is essential in high-dimensional cases. Cross validation is a generally applicable way to predict the performance of a model on a validation set using computation in place of mathematical analysis. It combines averages of prediction error to derive a more accurate estimate of model prediction performance, see [27,28] for more details. In this paper, we adopt the widely used 10-fold cross validation to conduct model selection. Moreover, we use 1SE (Standard Error) criteria to determine the optimal λ. The 1SE criteria is a commonly used one in cross-validation and gives the most regularized model such that error is within one standard error of the minimum, see [69].

3.4. Model evaluation

3.4.1. Variable selection

Following [21], we use three measures, FP, FN and NZ, to evaluate the performance of variable selection.

1. FP: The number of false positive, i.e. $\widehat{\beta}_j \neq 0$ when the true component $\beta_j = 0$.
2. FN: The number of false negative, i.e. $\widehat{\beta}_j = 0$ when the true component $\beta_j \neq 0$.
3. NZ: The total number of variables that are not equal to zero, i.e. $\#\left\{j : \widehat{\beta}_j \neq 0\right\}$.

3.4.2. Coefficient estimation

Following [91], we use $\left\|\widehat{\beta}-\beta\right\|_2$ to appraise the accuracy of coefficient estimation. Here, $\widehat{\beta}$ is the estimator, β represents the true values of the regression coefficients, and $\|\cdot\|_2$ denotes the 2-norm.

3.4.3. Prediction ability

To evaluate the predictive ability of the models, we consider

1. The root mean square error (RMSE) defined as

$$RMSE\left(\theta\right) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(Expectile_{y_i}\left(\theta|\boldsymbol{x_i}\right) - \widehat{Expectile}_{y_i}\left(\theta|\boldsymbol{x_i}\right)\right)^2}, \tag{11}$$

   where $\widehat{Expectile}_{y_i}\left(\theta|\boldsymbol{x_i}\right)$ is the estimated value of true expectile $Expectile_{y_i}\left(\theta|\boldsymbol{x_i}\right)$ of y_i conditional on $\boldsymbol{x_i}$. In general, the smaller value of RMSE is, the better the model performs.
2. The mean absolute error (MAE) defined as

$$MAE\left(\theta\right) = \frac{1}{N}\sum_{i=1}^{N}\left|Expectile_{y_i}\left(\theta|\boldsymbol{x_i}\right) - \widehat{Expectile}_{y_i}\left(\theta|\boldsymbol{x_i}\right)\right|. \tag{12}$$

   Similarly, the smaller value of MAE is, the better the model performs.
3. The expectile-based prediction error (EPE) defined as

$$EPE\left(\theta\right) = \frac{1}{N}\sum_{i=1}^{N}\rho_\theta\left(y_i - \widehat{Expectile}_{y_i}\left(\theta|\boldsymbol{x_i}\right)\right). \tag{13}$$

   Thus the smaller value of EPE is, the better the model performs.

4. Simulation experiments

In this section, we demonstrate the superiority of the ER-EN model through Monte Carlo simulations. All numerical experiments are conducted on two Intel Core i5–2400 (3.10 GHz) processors and 8 GB RAM. All computations are implemented through R coding. Specifically, the R package ‘hqreg’ is used and modified to solve models.

4.1. *p<n*

4.1.1. Data generation

We generate the data from the following linear model:

$$y = \boldsymbol{x}_{\boldsymbol{i}}^T \beta + \sigma \varepsilon, \tag{14}$$

where $\boldsymbol{x}$ comes from a multivariate normal distribution, i.e. $\boldsymbol{x} \sim N_p(\mathbf{0}, \mathbf{\Sigma})$ with $\mathbf{\Sigma}_{i,j} = \rho^{|i-j|}$ and $\rho = 0.5$, $\beta = (3, 1.5, 0, 0, 2, 0, 0, 0)^T$; σ is a scale parameter to control signal-to-noise ratio. We consider two different types of random errors: $N(0,1)$ and $\chi^2(2)$, for ϵ. This data generation process (DGP) has been considered in many works, see for instance [22,65,76,92]. In addition, we set the scale parameter $\sigma = 3$, sample size *n* = 1200 and repetition *k* = 500.

4.1.2. Simulation process

We design the entire simulation process as follows.

Step 1: Generate data. We generate a data set with 1200 observations. The first 1000 observations are used as in-sample data, while the remaining 200 observations are used as out-of-sample data.

Step 2: Estimate models. We estimate the ER-EN model based on the optimal λ selected through the 10-fold cross validation. For comparison purposes, we also estimate other models including the elastic-net penalized least squares regression (LSR-EN), the elastic-net penalized Huber loss regression (HR-EN), the elastic-net penalized quantile regression (QR-EN) and the conventional expectile regression (ER).

Step 3: Evaluate performance. We evaluate the model from three aspects: variable selection, coefficient estimation, and prediction ability.

Step 4: Compare results. Repeat the above three steps 500 times and report the average and standard deviation.

4.1.3. Comparison results

In this section, fix $\alpha = 0.9$ when selecting the tuning parameter λ. We perform tuning parameter selection based on the 10-fold cross validation and use the 1SE (Standard Error) criteria to select the optimal λ in four models. Here we present in Table 1 the selection results of λ across different θ s.

Table 1.

Optimal values of λ in the ER-EN model via 10-fold cross validation with $p<n$.

θ	$\varepsilon \sim N(0,1)$		$\varepsilon \sim \chi^2(2)$	
	λ	1SE	λ	1SE
0.1	0.047	0.548	0.028	0.628
0.3	0.046	0.533	0.027	0.625
0.5	0.045	0.530	0.026	0.622
0.7	0.043	0.527	0.024	0.621
0.9	0.041	0.525	0.023	0.617

We then compare the performance of five models, namely LSR-EN, HR-EN, QR-ER, ER-EN and ER. Following [85], we consider three values of γ in the HR-EN model, i.e. IQR(y), IQR(y)/2, IQR(y)/10, where IQR(y) is the inter-quartile range of y. We report the simulation results for the two different random errors in Tables 2 and 3, respectively. Table 2 shows the results of the simulation experiment for the case of $\varepsilon \sim N(0,1)$ with $p<n$. We find that the ER model without variable selection functioning is inferior to the LSR-EN, HR-EN, HR-EN and ER-EN models in this case.

Table 2.

Simulation results for the case of $\varepsilon \sim N(0,1)$ with *p*<*n*.

Models	γ or θ	$\left\lVert\widehat{\beta} - \beta\right\rVert_2$	RMSE	MAE	EPE	NZ	FP	FN	Time
LSR-EN	–	0.168	0.220	0.276	0.184	3.173	1.123	0.000	0.239
		(0.042)	(0.026)	(0.012)	(0.010)	(0.009)	(0.352)	0.000	
HR-EN	0.91	0.188	0.185	0.214	0.195	3.406	1.154	0.000	0.387
		(0.073)	(0.056)	(0.046)	(0.054)	(0.139)	(0.966)	0.000	
	0.455	0.189	0.190	0.213	0.187	3.859	1.067	0.000	0.365
		(0.084)	(0.059)	(0.048)	(0.050)	(0.140)	(0.980)	0.000	
	0.091	0.180	0.196	0.220	0.186	3.742	1.044	0.000	0.390
		(0.087)	(0.057)	(0.049)	(0.057)	(0.147)	(0.952)	0.000	
QR-EN	0.1	0.267	0.285	0.305	0.191	3.220	1.175	0.000	0.357
		(0.051)	(0.059)	(0.055)	(0.063)	(0.060)	(0.065)	0.000	
	0.3	0.238	0.273	0.284	0.185	3.287	1.138	0.000	0.381
		(0.059)	(0.052)	(0.058)	(0.068)	(0.065)	(0.067)	0.000	
	0.5	0.251	0.255	0.280	0.186	3.275	1.162	0.000	0.355
		(0.050)	(0.051)	(0.053)	(0.069)	(0.060)	(0.069)	0.000	
	0.7	0.267	0.260	0.287	0.170	3.236	1.137	0.000	0.325
		(0.058)	(0.055)	(0.056)	(0.065)	(0.069)	(0.066)	0.000	
	0.9	0.242	0.264	0.296	0.181	3.507	1.149	0.000	0.370
		(0.056)	(0.057)	(0.055)	(0.066)	(0.067)	(0.067)	0.000	
ER-EN	0.1	0.178	0.223	0.303	0.187	3.230	1.107	0.000	0.226
		(0.061)	(0.050)	(0.020)	(0.010)	(0.005)	(0.809)	0.000	
	0.3	0.179	0.227	0.302	0.176	3.226	1.165	0.000	0.207
		(0.054)	(0.043)	(0.016)	(0.018)	(0.006)	(0.508)	0.000	
	0.5	0.168	0.220	0.276	0.184	3.173	1.123	0.000	0.239
		(0.042)	(0.026)	(0.012)	(0.010)	(0.009)	(0.352)	0.000	
	0.7	0.169	0.264	0.281	0.157	3.268	1.144	0.000	0.202
		(0.028)	(0.020)	(0.013)	(0.009)	(0.007)	(0.195)	0.000	
	0.9	0.170	0.265	0.290	0.163	3.295	1.108	0.000	0.218
		(0.037)	(0.025)	(0.014)	(0.008)	(0.008)	(0.208)	0.000	
ER	0.1	0.369	0.383	0.355	0.227	8.000	5.000	0.000	0.295

Note: (1) We present the standard deviation in parentheses below the average value across 500 repetitions; (2) Time shows CPU computing time (in seconds).

Table 3.

Simulation results for the case of $\varepsilon \sim \chi^2(2)$ with *p*<*n*.

Models	γ or θ	$\|\widehat{\beta} - \beta\|_2$	RMSE	MAE	EPE	NZ	FP	FN	Time
LSR-EN	–	0.164	0.620	0.481	0.144	3.029	1.080	0.000	0.230
		(0.052)	(0.036)	(0.018)	(0.014)	(0.017)	(0.355)	0.000	
HR-EN	0.88	0.215	0.558	0.445	0.119	3.979	1.264	0.000	0.462
		(0.085)	(0.063)	(0.054)	(0.055)	(0.143)	(0.965)	0.000	
	0.44	0.218	0.626	0.557	0.220	4.352	1.205	0.000	0.469
		(0.096)	(0.067)	(0.053)	(0.057)	(0.155)	(0.948)	0.000	
	0.088	0.222	0.651	0.572	0.223	4.365	1.059	0.000	0.440
		(0.097)	(0.068)	(0.057)	(0.058)	(0.141)	(0.946)	0.000	
QR-EN	0.1	0.148	0.632	0.580	0.191	3.220	1.257	0.000	0.380
		(0.065)	(0.060)	(0.063)	(0.063)	(0.062)	(0.063)	0.000	
	0.3	0.147	0.607	0.569	0.182	3.157	1.238	0.000	0.381
		(0.062)	(0.067)	(0.060)	(0.068)	(0.064)	(0.060)	0.000	
	0.5	0.140	0.595	0.567	0.180	3.205	1.206	0.000	0.387
		(0.060)	(0.064)	(0.064)	(0.066)	(0.068)	(0.062)	0.000	
	0.7	0.133	0.567	0.451	0.175	3.202	1.200	0.000	0.375
		(0.061)	(0.066)	(0.063)	(0.067)	(0.063)	(0.064)	0.000	
	0.9	0.146	0.585	0.555	0.189	3.148	1.219	0.000	0.370
		(0.069)	(0.063)	(0.068)	(0.068)	(0.067)	(0.062)	0.000	
ER-EN	0.1	0.165	0.615	0.563	0.198	3.088	1.117	0.000	0.219
		(0.077)	(0.041)	(0.027)	(0.015)	(0.015)	(0.807)	0.000	
	0.3	0.164	0.602	0.461	0.180	3.050	1.145	0.000	0.218
		(0.063)	(0.033)	(0.023)	(0.017)	(0.012)	(0.503)	0.000	
	0.5	0.164	0.620	0.481	0.144	3.029	1.080	0.000	0.230
		(0.052)	(0.036)	(0.018)	(0.014)	(0.017)	(0.355)	0.000	
	0.7	**0.155**	**0.487**	**0.357**	**0.082**	**3.011**	**0.790**	0.000	0.231
		(0.035)	**(0.025)**	**(0.017)**	**(0.010)**	**(0.010)**	**(0.190)**	0.000	
	0.9	0.158	0.568	0.473	0.178	3.027	0.957	0.000	0.229
		(0.040)	(0.023)	(0.013)	(0.012)	(0.011)	(0.207)	0.000	
ER	0.1	0.360	0.789	0.585	0.184	8.000	5.000	0.000	0.387

Note: (1) We present the standard deviation in parentheses below the average value across 500 repetitions. (2) Boldface denotes the best performance. (3) Time shows CPU computing time (in seconds).

In Table 3, we present the results for the case of $\varepsilon \sim \chi^2(2)$ with $p<n$. The advantage of the ER-EN model is obvious when the random error is asymmetric. It is superior to the other models in terms of variable selection and predictive ability. First, the ER-EN model selects the number of variables that are not equal to 0 more closely to the real number than the LSR-EN, the HR-EN and the QR-EN models in terms of NZ, while the ER model is invalid for variable selection. As for FP and FN, the ER-EN model also outperforms the other models especially when $\theta = 0.7$ in that all signal variables are selected, which means that our selected model contains the underlying true model. Second, the ER-EN model gets better fitting effects at different quantiles especially when $\theta = 0.7$ in terms of the averages of RMSE, MAE and EPE. In addition, the difference between $\theta = 0.5$ and the best $\theta = 0.7$ is relatively large. For example, $\left\|\widehat{\beta} - \beta\right\|_2$ changes from 0.164 to 0.155 by decreasing 5.49%, and RMSE changes from 0.620 to 0.487 by decreasing 21.45%.

The superiority of the ER-EN model at the best $\theta = 0.7$ can be verified more intuitively through the out-of-sample performance. The detailed comparisons with setting $\gamma = IQR(y)$, $\theta = 0.7$, $\alpha = 0.9$, and $\varepsilon \sim \chi^2(2)$ are shown in Figures 1–3, which present the boxplots of the RMSE, MAE and EPE for different models. The boxplots indicate that the ER-EN model is superior to the LSR-EN, HR-EN, QR-EN and ER models in terms of accuracy and robustness of prediction. We find that the prediction biases of ER model is significantly higher than those of the other four models.

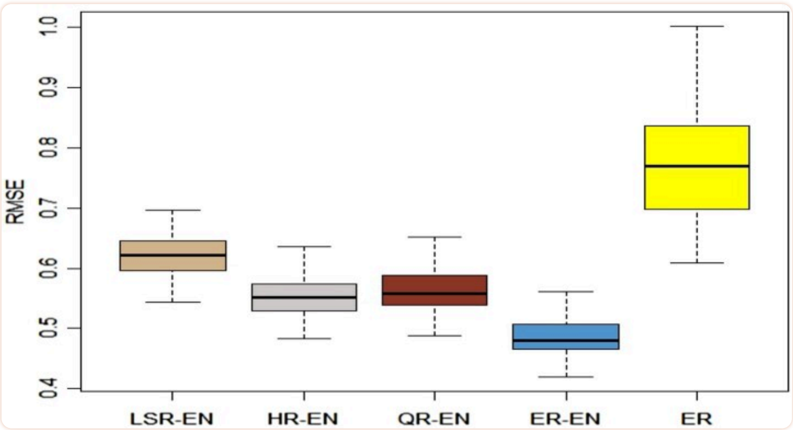

Figure 1.

Boxplots of out-of-sample RMSE for the case of $\varepsilon \sim \chi^2(2)$ with $p<n$.

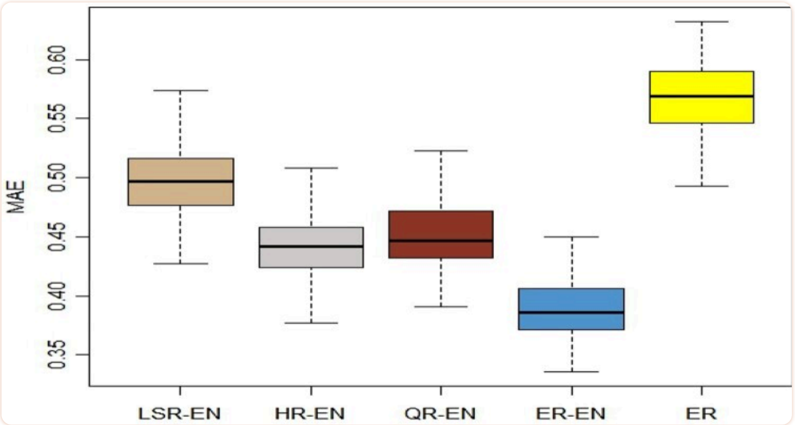

Figure 2.

Boxplots of out-of-sample MAE for the case of $\varepsilon \sim \chi^2(2)$ with $p<n$.

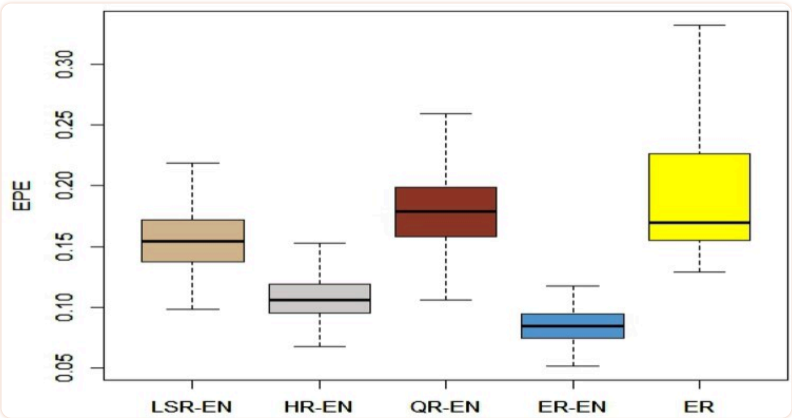

Figure 3.

Boxplots of out-of-sample EPE for the case of $\varepsilon \sim \chi^2(2)$ with $p<n$.

To sum up, the numerical results show that the ER-EN model, as well as the LSR-EN, HR-EN, QR-EN models, performs well when the random error is symmetrically distributed. However, its advantage is prominent when the random error follows an asymmetric distribution. It outperforms the LSR-EN, HR-EN, QR-EN and ER models in terms of variable selection, prediction accuracy, and computing time. In this case, the QR-EN model performs relatively better than the LSR-EN and HR-EN models because the former is also suitable for asymmetric error distribution. In addition, the ER-EN model is able to provide more precise and robust predictions for out-of-sample testing.

4.2. $p>n$

4.2.1. Data generation We still use model (14) to generate data with the same settings as the case of $p<n$. The differences are in the following two aspects.

First, we consider $p = 100$ and $n = 80$ for the case of $p>n$.

Second, we consider three different high-dimensional situations with 15%, 10% and 5% active variables. To this end, we set the coefficient vector β with 100 entries as follows:

Situation 1: 15% active variables. The first 5 entries in β are (3, 1.5, 2, 2.5, 1), the second and the third 5 entries are the same, while the left entries are zeros, i.e. $\beta_j = 0$ for $j = 16, 17, \ldots, 100$.

Situation 2: 10% active variables. The first 5 entries in β are (3, 1.5, 2, 2.5, 1), the second 5 entries are the same, while the left entries are zeros, i.e. $\beta_j = 0$ for $j = 11, 12, \ldots, 100$.

Situation 3: 5% active variables. The first 5 entries in β are (3, 1.5, 2, 2.5, 1), while the left entries are zeros, i.e. $\beta_j = 0$ for $j = 6, 7, \ldots, 100$.

4.2.2. Simulation process The simulation process is similar to the case of $p<n$, the main difference is that the first 60 observations are used as in-sample data, while the remaining 20 observations are used as out-of-sample data. In addition, the ER model is not applicable for the $p>n$ case and we do not consider its performance in this case.

4.2.3. Comparison results Table 4 presents the optimal values of the tuning parameter λ in the ER-EN model with $p>n$ via 10-fold cross validation in Situation 1. In addition, we report in Tables 5 and 6 the simulation results for the cases of $\varepsilon \sim N(0,1)$ and $\varepsilon \sim \chi^2(2)$, respectively. Overall, the numerical results also show that the ER-EN model outperforms the LSR-EN, HR-EN and QR-EN models in terms of variable selection, prediction accuracy, and computing time, especially for the asymmetric error distribution. In this regard, we only present in Figures 4–6 the boxplots of out-of-sample RMSE, MAE and EPE for the case of $\varepsilon \sim \chi^2(2)$. The superiority of the ER-EN model is also obvious in out-of-sample testing.

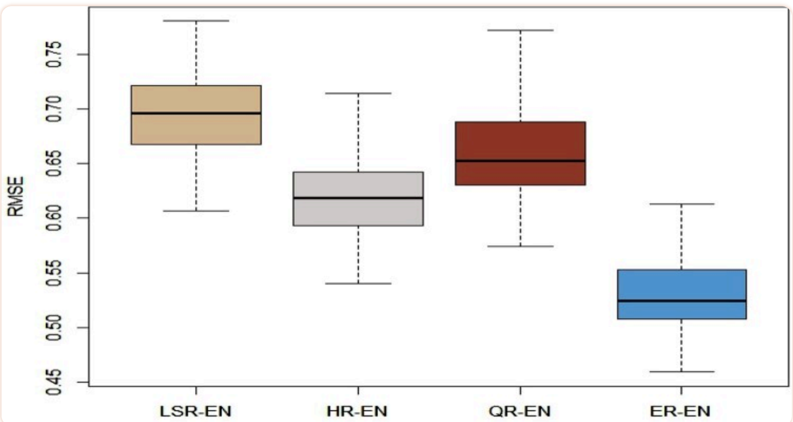

Figure 4.

Boxplots of out-of-sample RMSE for the case of $\varepsilon \sim \chi^2(2)$ with $p>n$ in Situation 1.

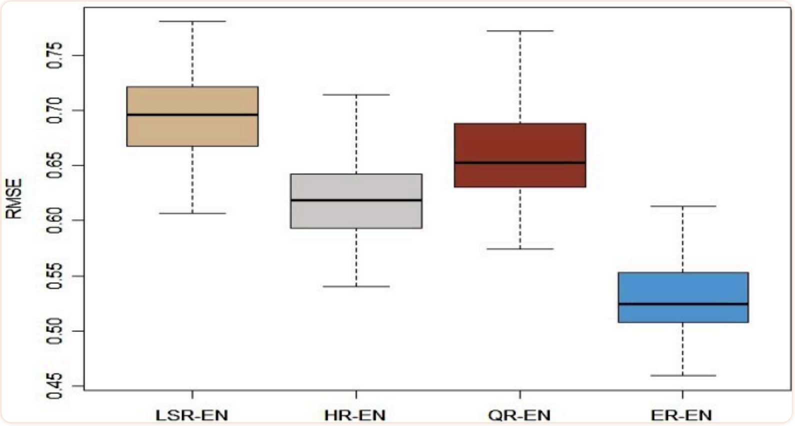

Figure 5.

Boxplots of out-of-sample MAE for the case of $\varepsilon \sim \chi^2(2)$ with $p>n$ in Situation 1.

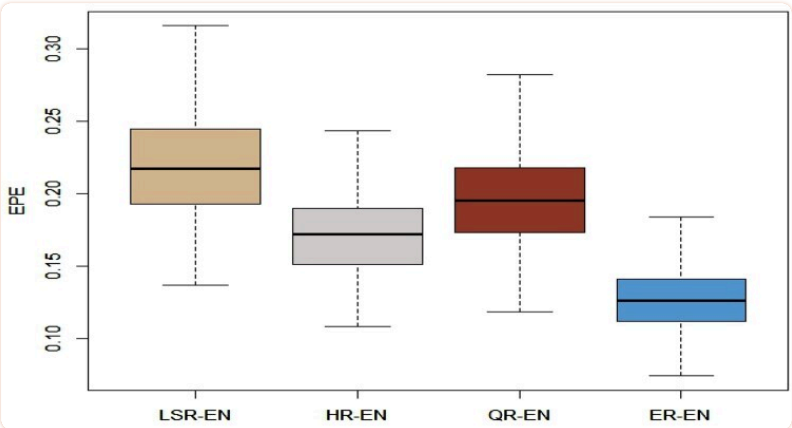

Figure 6.

Boxplots of out-of-sample EPE for the case of $\varepsilon \sim \chi^2(2)$ with $p>n$ in Situation 1.

Table 4.

Optimal values of λ in the ER-EN model with $p>n$ via 10-fold cross validation in Situation 1.

θ	$\varepsilon \sim N(0,1)$		$\varepsilon \sim \chi^2(2)$	
	λ	**1SE**	λ	**1SE**
0.1	0.053	0.659	0.036	0.759
0.3	0.054	0.658	0.037	0.747
0.5	0.056	0.657	0.038	0.742
0.7	0.057	0.660	0.039	0.741
0.9	0.058	0.661	0.040	0.740

Table 5.

Simulation results for the case of $\varepsilon \sim N(0,1)$ with *p*>*n* in Situation 1.

Models	γ or θ	$\lVert\widehat{\beta} - \beta\rVert_2$	RMSE	MAE	EPE	NZ	FP	FN	Time
LSR-EN	–	0.328	0.600	0.465	0.193	15.446	2.206	0.000	0.341
		(0.133)	(0.145)	(0.134)	(0.151)	(0.151)	(0.538)	0.000	
HR-EN	0.95	0.322	0.591	0.448	0.196	15.453	1.258	0.000	0.517
		(0.160)	(0.150)	(0.142)	(0.136)	(0.229)	(0.989)	0.000	
	0.475	0.339	0.587	0.456	0.252	15.375	1.581	0.000	0.533
		(0.121)	(0.148)	(0.136)	(0.141)	(0.241)	(0.977)	0.000	
	0.095	0.336	0.572	0.462	0.288	15.366	1.433	0.000	0.538
		(0.178)	(0.151)	(0.138)	(0.140)	(0.229)	(0.936)	0.000	
QR-EN	0.1	0.342	0.589	0.513	0.212	15.391	2.364	0.000	0.395
		(0.150)	(0.149)	(0.149)	(0.146)	(0.148)	(0.148)	0.000	
	0.3	0.345	0.597	0.474	0.204	15.275	2.999	0.000	0.419
		(0.147)	(0.150)	(0.147)	(0.152)	(0.159)	(0.152)	0.000	
	0.5	0.331	0.603	0.467	0.198	15.206	2.281	0.000	0.426
		(0.148)	(0.148)	(0.151)	(0.148)	(0.146)	(0.147)	0.000	
	0.7	0.346	0.606	0.465	0.191	15.220	2.262	0.000	0.432
		(0.151)	(0.151)	(0.148)	(0.144)	(0.151)	(0.150)	0.000	
	0.9	0.347	0.607	0.457	0.197	15.212	2.281	0.000	0.439
		(0.149)	(0.146)	(0.147)	(0.151)	(0.146)	(0.147)	0.000	
ER-EN	0.1	0.379	0.537	0.510	0.207	15.589	2.144	0.000	0.356
		(0.161)	(0.132)	(0.149)	(0.135)	(0.139)	(0.798)	0.000	
	0.3	0.347	0.534	0.468	0.188	15.433	2.162	0.000	0.334
		(0.150)	(0.141)	(0.146)	(0.154)	(0.153)	(0.499)	0.000	
	0.5	0.328	0.600	0.465	0.193	15.446	2.206	0.000	0.341
		(0.133)	(0.145)	(0.134)	(0.151)	(0.151)	(0.538)	0.000	
	0.7	0.335	0.601	0.463	0.154	15.225	2.164	0.000	0.338
		(0.155)	(0.150)	(0.154)	(0.150)	(0.160)	(0.379)	0.000	
	0.9	0.337	0.586	0.446	0.185	15.159	2.179	0.000	0.326
		(0.158)	(0.148)	(0.143)	(0.141)	(0.142)	(0.287)	0.000	

Note: (1) We present the standard deviation in parentheses below the average value across 500 repetitions. (2) Time shows CPU computing time (in seconds).

Table 6.

Simulation results for the case of $\varepsilon \sim \chi^2(2)$ with *p*>*n* in Situation 1.

Models	γ or θ	$\|\widehat{\beta} - \beta\|_2$	RMSE	MAE	EPE	NZ	FP	FN	Time
LSR-EN	–	0.467	0.701	0.611	0.229	15.069	1.134	0.000	0.483
		(0.109)	(0.110)	(0.130)	(0.136)	(0.130)	(0.408)	0.000	
HR-EN	0.94	0.502	0.663	0.524	0.228	16.004	2.325	0.000	0.696
		(0.124)	(0.114)	(0.156)	(0.155)	(0.192)	(0.959)	0.000	
	0.47	0.527	0.769	0.642	0.284	15.417	2.257	0.000	0.699
		(0.116)	(0.117)	(0.114)	(0.112)	(0.303)	(0.998)	0.000	
	0.094	0.275	0.804	0.647	0.320	15.390	2.083	0.000	0.712
		(0.147)	(0.115)	(0.109)	(0.106)	(0.195)	(0.990)	0.000	
QR-EN	0.1	0.547	0.722	0.609	0.244	15.299	1.301	0.000	0.564
		(0.116)	(0.114)	(0.114)	(0.180)	(0.119)	(0.118)	0.000	
	0.3	0.536	0.715	0.610	0.236	15.166	1.292	0.000	0.598
		(0.111)	(0.116)	(0.117)	(0.116)	(0.117)	(0.114)	0.000	
	0.5	0.519	0.707	0.611	0.230	15.263	1.294	0.000	0.585
		(0.120)	(0.119)	(0.115)	(0.114)	(0.120)	(0.119)	0.000	
	0.5	0.542	0.704	0.588	0.223	15.249	1.257	0.000	0.584
		(0.116)	(0.120)	(0.120)	(0.117)	(0.117)	(0.120)	0.000	
	0.9	0.545	0.709	0.603	0.229	15.246	1.264	0.000	0.590
		(0.115)	(0.111)	(0.119)	(0.116)	(0.120)	(0.119)	0.000	
ER-EN	0.1	0.475	0.704	0.602	0.225	15.101	1.167	0.000	0.487
		(0.126)	(0.102)	(0.121)	(0.133)	(0.133)	(0.858)	0.000	
	0.3	0.470	0.703	0.603	0.228	15.106	1.190	0.000	0.492
		(0.120)	(0.114)	(0.129)	(0.134)	(0.131)	(0.557)	0.000	
	0.5	0.467	0.701	0.611	0.229	15.069	1.134	0.000	0.483
		(0.109)	(0.110)	(0.130)	(0.136)	(0.130)	(0.408)	0.000	
	0.7	**0.451**	**0.618**	**0.486**	**0.186**	**15.063**	**0.842**	0.000	0.489
		(0.119)	**(0.121)**	**(0.136)**	**(0.131)**	**(0.133)**	**(0.149)**	0.000	
	0.9	0.464	0.702	0.582	0.217	15.075	0.999	0.000	0.491
		(0.101)	(0.123)	(0.134)	(0.132)	(0.132)	(0.257)	0.000	

Note: (1) We present the standard deviation in parentheses below the average value across 500 repetitions. (2) Boldface denotes the best performance. (3) Time shows CPU computing time (in seconds).

As for Situation 2 and Situation 3, we only report the simulation results in Tables 7–10. It is clear that the ER-EN model is especially suitable for the case of $\varepsilon \sim \chi^2(2)$. Comparing the simulation results of three situations in Tables 6, 8, and 10, we further find that as the proportion of active variables decreases, the error of the estimated coefficient decreases, and the accuracy of the ER-EN model increases. Actually, the situation with more active variables is a dense one, which means that variables tend to be more related to each other. This relatively stronger correlation may lead to a decline in the efficiency of variable selection and model estimation. On the contrary, the situation with less active variables is a sparse one. In this sparse situation, the relationship between variables is relatively simple for better performance of the model.

Table 9.

Simulation results for the case of $\varepsilon \sim N(0,1)$ with *p*>*n* in Situation 3.

Models	γ or θ	$\|\widehat{\beta} - \beta\|_2$	RMSE	MAE	EPE	NZ	FP	FN	Time
LSR-EN	–	0.252	0.524	0.389	0.107	5.270	0.330	0.000	0.282
		(0.109)	(0.091)	(0.072)	(0.077)	(0.108)	(0.044)	0.000	
HR-EN	0.89	0.247	0.515	0.372	0.120	5.324	0.382	0.000	0.458
		(0.136)	(0.126)	(0.118)	(0.112)	(0.106)	(0.065)	0.000	
	0.445	0.258	0.511	0.380	0.176	5.305	0.359	0.000	0.474
		(0.097)	(0.124)	(0.112)	(0.117)	(0.108)	(0.073)	0.000	
	0.089	0.260	0.496	0.386	0.212	5.317	0.357	0.000	0.479
		(0.154)	(0.127)	(0.114)	(0.116)	(0.114)	(0.070)	0.000	
QR-EN	0.1	0.266	0.513	0.437	0.136	5.328	0.312	0.000	0.336
		(0.126)	(0.125)	(0.125)	(0.122)	(0.113)	(0.064)	0.000	
	0.3	0.269	0.521	0.398	0.128	5.313	0.323	0.000	0.360
		(0.123)	(0.126)	(0.123)	(0.128)	(0.110)	(0.058)	0.000	
	0.5	0.255	0.527	0.391	0.122	5.285	0.302	0.000	0.367
		(0.124)	(0.124)	(0.127)	(0.124)	(0.112)	(0.053)	0.000	
	0.7	0.270	0.530	0.389	0.115	5.297	0.306	0.000	0.373
		(0.127)	(0.127)	(0.124)	(0.120)	(0.115)	(0.060)	0.000	
	0.9	0.271	0.531	0.381	0.121	5.316	0.325	0.000	0.380
		(0.125)	(0.122)	(0.123)	(0.127)	(0.130)	(0.057)	0.000	
ER-EN	0.1	0.303	0.501	0.440	0.121	5.325	0.368	0.000	0.297
		(0.137)	(0.108)	(0.085)	(0.071)	(0.117)	(0.064)	0.000	
	0.3	0.271	0.508	0.392	0.122	5.314	0.346	0.000	0.275
		(0.126)	(0.097)	(0.082)	(0.073)	(0.108)	(0.068)	0.000	
	0.5	0.252	0.524	0.389	0.107	5.270	0.330	0.000	0.282
		(0.109)	(0.091)	(0.072)	(0.077)	(0.108)	(0.044)	0.000	
	0.7	0.259	0.532	0.387	0.078	5.296	0.295	0.000	0.279
		(0.091)	(0.086)	(0.075)	(0.074)	(0.105)	(0.045)	0.000	
	0.9	0.261	0.530	0.380	0.109	5.312	0.301	0.000	0.267
		(0.104)	(0.084)	(0.073)	(0.075)	(0.108)	(0.043)	0.000	

Note: (1) We present the standard deviation in parentheses below the average value across 500 repetitions. (2) Time shows CPU computing time (in seconds).

Table 7.

Simulation results for the case of $\varepsilon \sim N(0,1)$ with $p>n$ in Situation 2.

Models	γ or θ	$\left\|\widehat{\beta} - \beta\right\|_2$	RMSE	MAE	EPE	NZ	FP	FN	Time
LSR-EN	–	0.311	0.593	0.448	0.176	10.261	0.389	0	0.367
		(0.050)	(0.032)	(0.013)	(0.018)	(0.109)	(0.055)	0	
HR-EN	0.92	0.306	0.574	0.431	0.179	10.357	0.541	0	0.543
		(0.077)	(0.067)	(0.059)	(0.053)	(0.121)	(0.066)	0	
	0.46	0.317	0.570	0.439	0.235	10.308	0.564	0	0.559
		(0.038)	(0.065)	(0.053)	(0.058)	(0.178)	(0.047)	0	
	0.092	0.319	0.555	0.445	0.271	10.370	0.612	0	0.564
		(0.095)	(0.068)	(0.055)	(0.057)	(0.148)	(0.048)	0	
QR-EN	0.1	0.325	0.572	0.496	0.195	10.325	0.347	0	0.421
		(0.067)	(0.066)	(0.066)	(0.063)	(0.129)	(0.065)	0	
	0.3	0.328	0.580	0.457	0.187	10.322	0.382	0	0.445
		(0.064)	(0.067)	(0.064)	(0.069)	(0.125)	(0.069)	0	
	0.5	0.314	0.586	0.450	0.181	10.296	0.364	0	0.452
		(0.065)	(0.065)	(0.068)	(0.065)	(0.113)	(0.064)	0	
	0.7	0.329	0.589	0.448	0.174	10.317	0.445	0	0.458
		(0.068)	(0.068)	(0.065)	(0.061)	(0.108)	(0.067)	0	
	0.9	0.330	0.590	0.440	0.180	10.315	0.367	0	0.465
		(0.066)	(0.063)	(0.064)	(0.068)	(0.124)	(0.063)	0	
ER-EN	0.1	0.362	0.520	0.490	0.180	10.318	0.427	0	0.382
		(0.078)	(0.049)	(0.026)	(0.012)	(0.123)	(0.045)	0	
	0.3	0.330	0.517	0.451	0.181	10.306	0.445	0	0.360
		(0.067)	(0.038)	(0.023)	(0.014)	(0.119)	(0.059)	0	
	0.5	0.311	0.593	0.448	0.176	10.261	0.389	0	0.367
		(0.050)	(0.032)	(0.013)	(0.018)	(0.109)	(0.055)	0	
	0.7	0.318	0.581	0.446	0.137	10.291	0.347	0	0.364
		(0.032)	(0.027)	(0.016)	(0.015)	(0.102)	(0.076)	0	
	0.9	0.320	0.579	0.429	0.168	10.298	0.362	0	0.352
		(0.045)	(0.025)	(0.014)	(0.016)	(0.127)	(0.064)	0	

Note: (1) We present the standard deviation in parentheses below the average value across 500 repetitions. (2) Boldface denotes the best performance. (3) Time shows CPU computing time (in seconds).

Table 8.

Simulation results for the case of $\varepsilon \sim \chi^2(2)$ with *p*>*n* in Situation 2.

Models	γ or θ	$\|\hat{\beta} - \beta\|_2$	RMSE	MAE	EPE	NZ	FP	FN	Time
LSR-EN	–	0.418	0.652	0.562	0.220	10.362	0.385	0.000	0.360
		(0.158)	(0.139)	(0.119)	(0.113)	(0.128)	(0.087)	0.000	
HR-EN	0.9	0.453	0.614	0.475	0.179	10.403	0.576	0.000	0.573
		(0.173)	(0.163)	(0.155)	(0.154)	(0.135)	(0.098)	0.000	
	0.45	0.478	0.720	0.593	0.235	10.308	0.318	0.000	0.576
		(0.133)	(0.166)	(0.153)	(0.151)	(0.152)	(0.057)	0.000	
	0.09	0.226	0.755	0.598	0.271	10.343	0.454	0.000	0.589
		(0.196)	(0.164)	(0.158)	(0.155)	(0.240)	(0.069)	0.000	
QR-EN	0.1	0.498	0.673	0.560	0.195	10.305	0.452	0.000	0.441
		(0.165)	(0.163)	(0.163)	(0.169)	(0.147)	(0.067)	0.000	
	0.3	0.487	0.666	0.561	0.187	10.303	0.443	0.000	0.475
		(0.160)	(0.165)	(0.166)	(0.165)	(0.142)	(0.063)	0.000	
	0.5	0.470	0.658	0.562	0.181	10.301	0.345	0.000	0.462
		(0.169)	(0.168)	(0.164)	(0.163)	(0.143)	(0.068)	0.000	
	0.7	0.493	0.655	0.539	0.174	10.273	0.278	0.000	0.461
		(0.165)	(0.169)	(0.169)	(0.166)	(0.129)	(0.069)	0.000	
	0.9	0.496	0.660	0.554	0.180	10.272	0.275	0.000	0.467
		(0.164)	(0.160)	(0.168)	(0.165)	(0.125)	(0.068)	0.000	
ER-EN	0.1	0.426	0.665	0.553	0.176	10.306	0.418	0.000	0.364
		(0.175)	(0.147)	(0.128)	(0.116)	(0.114)	(0.087)	0.000	
	0.3	0.421	0.664	0.574	0.179	10.302	0.414	0.000	0.369
		(0.169)	(0.135)	(0.120)	(0.115)	(0.105)	(0.086)	0.000	
	0.5	0.418	0.652	0.562	0.170	10.362	0.385	0.000	0.360
		(0.158)	(0.139)	(0.119)	(0.113)	(0.128)	(0.087)	0.000	
	0.7	**0.402**	**0.569**	**0.437**	**0.137**	**10.248**	**0.213**	0.000	0.366
		(0.130)	**(0.128)**	**(0.113)**	**(0.118)**	**(0.093)**	**(0.038)**	0.000	
	0.9	0.415	0.603	0.533	0.168	10.255	0.250	0.000	0.368
		(0.148)	(0.126)	(0.115)	(0.117)	(0.109)	(0.036)	0.000	

Note: (1) We present the standard deviation in parentheses below the average value across 500 repetitions. (2) Boldface denotes the best performance. (3) Time shows CPU computing time (in seconds).

Table 10.

Simulation results for the case of $\varepsilon \sim \chi^2(2)$ with *p*>*n* in Situation 3.

Models	γ or θ	$\|\widehat{\beta} - \beta\|_2$	RMSE	MAE	EPE	NZ	FP	FN	Time
LSR-EN	–	0.392	0.606	0.536	0.144	5.227	0.219	0.000	0.301
		(0.084)	(0.065)	(0.045)	(0.039)	(0.104)	(0.048)	0.000	
HR-EN	0.87	0.427	0.588	0.449	0.153	5.360	0.353	0.000	0.514
		(0.099)	(0.089)	(0.081)	(0.080)	(0.129)	(0.074)	0.000	
	0.435	0.452	0.694	0.567	0.209	5.346	0.348	0.000	0.517
		(0.059)	(0.092)	(0.079)	(0.077)	(0.158)	(0.073)	0.000	
	0.087	0.200	0.729	0.572	0.245	5.352	0.401	0.000	0.530
		(0.122)	(0.090)	(0.084)	(0.081)	(0.144)	(0.075)	0.000	
QR-EN	0.1	0.472	0.647	0.534	0.169	5.349	0.406	0.000	0.382
		(0.091)	(0.089)	(0.089)	(0.095)	(0.137)	(0.063)	0.000	
	0.3	0.461	0.640	0.535	0.161	5.346	0.347	0.000	0.416
		(0.086)	(0.091)	(0.092)	(0.091)	(0.124)	(0.069)	0.000	
	0.5	0.444	0.632	0.536	0.155	5.275	0.299	0.000	0.403
		(0.095)	(0.094)	(0.090)	(0.089)	(0.123)	(0.064)	0.000	
	0.7	0.467	0.629	0.513	0.148	5.228	0.282	0.000	0.402
		(0.091)	(0.095)	(0.095)	(0.092)	(0.120)	(0.065)	0.000	
	0.9	0.470	0.634	0.528	0.154	5.225	0.239	0.000	0.408
		(0.090)	(0.086)	(0.094)	(0.091)	(0.116)	(0.074)	0.000	
ER-EN	0.1	0.400	0.639	0.527	0.150	5.326	0.322	0.000	0.305
		(0.101)	(0.073)	(0.054)	(0.042)	(0.105)	(0.063)	0.000	
	0.3	0.395	0.628	0.528	0.153	5.240	0.245	0.000	0.310
		(0.095)	(0.061)	(0.046)	(0.041)	(0.102)	(0.062)	0.000	
	0.5	0.392	0.606	0.536	0.144	5.227	0.219	0.000	0.301
		(0.084)	(0.065)	(0.045)	(0.039)	(0.104)	(0.048)	0.000	
	0.7	**0.376**	**0.543**	**0.411**	**0.111**	**5.201**	**0.207**	0.000	0.307
		(0.056)	**(0.054)**	**(0.039)**	**(0.044)**	**(0.107)**	**(0.043)**	0.000	
	0.9	0.389	0.607	0.507	0.142	5.218	0.214	0.000	0.309
		(0.074)	(0.052)	(0.041)	(0.043)	(0.100)	(0.042)	0.000	

Note: (1) We present the standard deviation in parentheses below the average value across 500 repetitions. (2) Boldface denotes the best performance in each column. (3) Time shows CPU computing time (in seconds).

5. Practical applications

In this section, we apply our ER-EN model to two real-world applications. The main purpose is to investigate its interpretation and predictive ability.

5.1. Relative location of CT slices on axial axis data set

5.1.1. Data CT scans play an important role in medical imaging. It is still a challenging work to identify important one in tens of thousands of CT scans. Our ER-EN model can effectively solve the problem of screening important scans. We demonstrate the efficacy of ER-EN model with a dataset: relative location of CT slices on axial axis data set, from the public UCI database (http://archive.ics.uci.edu/ml/datasets/Relative+location+of+CT+slices+on+axial+axis).

The data are retrieved from a set of 53500 CT images from 74 different patients and consists of 384 features. The response variable is the relative location of the CT slice on the axial axis of the human body. We randomly partition the data by 3:2 into two sets, 32,100 observations used as in-sample data and the remaining 21,400 observations as out-of-sample data.

5.1.2. Variable selection and prediction In order to screen out the most important features from the 384 image features and reduce the prediction error, we use the elastic-net penalized function to screen the relative important features and predict the relative location of CT slices on the axial axis.

We compare the ER-EN model to the LSR-EN, HR-EN, QR-EN and ER models and report the comparison results in Table 11, where we use $\alpha = 0.9$ in the elastic-net penalty for all models. In Table 11, we present the number of nonzero regression coefficients, the expectile prediction error (EPE) of out-of-sample and the computing time across 500 repetitions.

Table 11.

Analysis of the relative location of CT slices on axial axis data set.

		Nonzero		EPE		
Models	γ **or** θ	**Average**	**Std.**	**Average**	**Std.**	**Time**
LSR-EN	–	153.412	1.674	43.248	0.949	194.126
HR-EN	26.744	108.347	2.472	44.127	0.849	308.159
	13.372	110.309	2.241	45.346	1.370	341.450
	2.674	110.246	3.745	47.317	1.043	340.052
QR-EN	0.1	121.415	1.529	48.951	0.979	201.694
	0.3	125.364	1.493	48.436	0.975	206.428
	0.5	132.058	1.428	47.521	0.973	217.459
	0.7	139.615	1.674	47.198	0.998	205.320
	0.9	145.692	1.958	46.549	0.986	204.219
ER-EN	0.1	140.437	1.549	44.765	0.942	195.703
	0.3	144.176	1.347	44.007	0.984	198.875
	0.5	153.412	1.674	43.248	0.949	194.126
	0.7	154.760	1.540	42.576	0.920	198.349
	0.9	165.713	1.671	42.207	0.956	195.756
ER	0.1	384.000	0.000	573.208	5.017	267.157
	0.3	384.000	0.000	547.670	4.452	268.383
	0.5	384.000	0.000	485.126	3.754	269.743
	0.7	384.000	0.000	482.462	4.451	273.429
	0.9	384.000	0.000	492.473	4.478	274.482

Note: Time shows the run time of the CPU (in seconds).

In this practical use, we find that the ER-EN model is still superior to the other models in terms of predictive ability and computing time. In addition, the ER-EN model is able to select important variables and tends to choose more nonzero variables than the QR-EN model. Taken together, the results imply that accurate prediction can be obtained by only using a part (about 40%) of variables, not all of them. This means that the complexity of matching can be reduced by simplifying the number of slice data through variable selection in the ER-EN model. Moreover, the EPE values and standard deviations at $\theta = 0.7$ become smaller, which confirms that the data are heterogeneous. It shows that the ER-EN model is able to predict relative location of CT slices on axial axis data accurately, especially at upper tails, which plays an important role in medical imaging.

5.2. Metabolism of tacrolimus (Tac) drug data set

5.2.1. Data In recent years, there has been an increasing demand for analysis of high-dimensional data, especially in the field of clinical medicine. The focus of the research is on the relationship between biochemical indexes and drug metabolism. The main difficulty is to solve the $p > n$ problem. We can solve this problem effectively by using the ER-EN model and consider using metabolism of tacrolimus (Tac) drug data set to compare model performance. Tac is an immunosuppressant drug that belongs to the class of calcineurin inhibitors and has an important role in the prevention of allograft rejection in liver transplantation (LT). This data is collected from the China Liver Transplant Registry (CLTR) database (http://www.cltr.org/pages/statistics/statisticslivercount.jsp).

The metabolism of tacrolimus (Tac) drug data set is a typical high-dimensional data with $p > n$. It has 390 biochemical indicators as variables, but only has 180 observations. We randomly partition the data by 3:2 into two sets: 108 observations are used as in-sample data and the remaining 72 observations as out-of-sample data.

5.2.2. Variable selection and prediction The selection of important biochemical indexes can be realized with using the elastic-net penalty. In this real-world application, it is necessary to consider the ER-EN model since $p > n$.

In this case, the ER model is invalid and Table 12 only reports the comparison results of the other four models: ER-EN, LSR-EN, HR-EN and QR-EN. The performance is evaluated via the number of nonzero regression coefficients, the expectile prediction error (EPE), and the computing time. Due to the limited space, we only report the result of $\alpha = 0.9$ for the elastic-net penalty in all models.

Table 12.

Analysis of the metabolism of tacrolimus (Tac) drug data set.

		Nonzero		EPE		Time
Models	**γ or θ**	**Average**	**Std.**	**Average**	**Std.**	
LSR-EN	–	221.429	7.023	149.358	35.741	115.486
HR-EN	56.862	247.896	6.140	185.048	198.064	137.398
	38.431	231.846	7.845	179.478	203.145	158.676
	5.686	228.364	7.830	186.054	206.756	159.063
QR-EN	0.1	226.147	7.659	201.659	137.529	126.756
	0.3	225.256	7.421	195.364	164.474	128.437
	0.5	226.159	6.593	184.250	145.270	126.472
	0.7	214.328	6.478	153.469	149.638	124.326
	0.9	202.365	6.326	180.290	156.986	125.210
ER-EN	0.1	200.124	6.059	141.452	47.230	113.912
	0.3	208.453	6.750	150.434	37.686	117.753
	0.5	221.429	7.023	149.358	35.741	115.486
	0.7	224.593	5.758	131.058	39.120	119.904
	0.9	225.937	5.560	148.367	45.159	115.245

Note: Time shows the run time of the CPU (in seconds).

By comparing the four models, we can see the ER-EN model has the minimum value of EPE, which illustrates its significant superiority in forecasting ability. In addition, we find that this accurate prediction result is achieved by only using a part (about 50%) of rather than all variables. As for the computing time, the ER-EN model is obviously a dominant one.

Taken together, the ER-EN model can provide accurate and robust prediction of tacrolimus (Tac) drug metabolism, which is of great significance to understand the metabolism of patients after liver transplantation.

6. Conclusion

In this paper, we introduce a novel elastic-net penalized expectile regression (ER-EN) model to perform variable selection in expectile regressions. We then provide its solution scheme via the SNCD algorithm. The proposed ER-EN model has three significant advantages. First, it is able to describe complete distribution of a response conditional on covariates information. Second, it has an excellent stability in solving high-dimensional data and is robust to heterogeneous data. Third, our ER-EN model runs faster than several competing models.

We discuss the techniques of ER-EN modeling and conduct Monte Carlo simulations to illustrate its strength. The numerical results show that the ER-EN model outperforms the LSR-EN, HR-EN, QR-EN and ER models in terms of variable selection and prediction ability. More specifically, the ER model without variable selection functioning is relatively inferior to the other four models for $p<n$ and become invalid for $p>n$. For the case of $\varepsilon \sim N(0,1)$, the LSR-EN, HR-EN, QR-EN, and ER-EN models have achieved good performance and none of them is a dominant one across all measures. But things are different for the case of $\varepsilon \sim \chi^2(2)$ where the ER-EN model is consistently preferred. In addition, the LSR-EN model is a special case of the ER-EN model when $\theta = 0.5$. This nature enables the ER-EN model to be a more general one that can be used in many fields.

We then consider two real-world applications and apply the ER-EN model to the predictions of relative location of CT slices on the axial axis and metabolism of tacrolimus (Tac) drug. The empirical results confirm our work and the findings are significant for practical use.

Acknowledgments

The authors are grateful to the Editor-in-Chief, the associated Editor, and three anonymous referees for their helpful comments and constructive guidance. This work was supported by the National Natural Science Foundation of China (71671056), the Humanity and Social Science Foundation of the Ministry of Education of China (19YJA790035), the Nature Science Foundation in the Universities of Anhui Province (XJ2019000103, KJ2017A391), and the National Statistical Science Research Projects of China (2019LD05).

Funding Statement

This work was supported by the National Natural Science Foundation of China (71671056), the Humanity and Social Science Foundation of the Ministry of Education of China (19YJA790035), the Nature Science Foundation in the Universities of Anhui Province (XJ2019000103, KJ2017A391), and the National Statistical Science Research Projects of China (2019LD05).

Disclosure statement

No potential conflict of interest was reported by the authors.

References

1. Aigner D.J., Amemiya T., and Poirier D.J., On the estimation of production frontiers: maximum likelihood estimation of the parameters of a discontinuous density function, *Int. Econ. Rev. (Philadelphia)* 17 (1976), pp. 377–396. doi: 10.2307/2525708 [CrossRef] [Google Scholar]

2. Alhamzawi R., Conjugate priors and variable selection for bayesian quantile regression, *Comput. Stat. Data Anal.* 64 (2013), pp. 209–219. doi: 10.1016/j.csda.2012.01.014 [CrossRef] [Google Scholar]

3. Alhamzawi R., Yu K., and Benoit D.F., Bayesian adaptive lasso quantile regression, *Stat. Model.* 12 (2012), pp. 279–297. doi: 10.1177/1471082X1101200304 [CrossRef] [Google Scholar]

4. Alshaybawee T., Alhamzawi R., Midi H., and Allyas I.I., Bayesian variable selection and coefficient estimation in heteroscedastic linear regression model, *J. Appl. Stat.* 45 (2018), pp. 2643–2657. doi: 10.1080/02664763.2018.1432576 [CrossRef] [Google Scholar]

5. Alshaybawee T., Midi H., and Alhamzawi R., Bayesian elastic net single index quantile regression, *J. Appl. Stat.* 44 (2017), pp. 853–871. doi: 10.1080/02664763.2016.1189515 [CrossRef] [Google Scholar]

6. Amin M., Song L., Thorlie M.A., and Wang X., SCAD-penalized quantile regression for high-dimensional data analysis and variable selection, *Stat. Neerl.* 69 (2015), pp. 212–235. doi: 10.1111/stan.12056 [CrossRef] [Google Scholar]

7. Bellini F., Mercuri L., and Rroji E., Implicit expectiles and measures of implied volatility, *Quant. Finance*. 18 (2018), pp. 1851–1864. doi: 10.1080/14697688.2018.1447680 [CrossRef] [Google Scholar]

8. Belloni A. and Chernozhukov V., L1-penalized quantile regression in high-dimensional sparse models, *Ann. Stat.* 39 (2011), pp. 82–130. doi: 10.1214/10-AOS827 [CrossRef] [Google Scholar]

9. Breheny P. and Huang J., Coordinate descent algorithms for nonconvex penalized regression, with applications to biological feature selection, *Ann. Appl. Stat.* 5 (2011), pp. 232–253. doi: 10.1214/10-AOAS388 [PMC free article] [PubMed] [CrossRef] [Google Scholar]

10. Burgette L.F., Exploratory quantile regression with many covariates an application to adverse birth outcomes, *Epidemiology* 22 (2011), pp. 859–866. doi: 10.1097/EDE.0b013e31822908b3 [PubMed] [CrossRef] [Google Scholar]

11. Cannon A., Quantile regression neural networks: Implementation in R and application to precipitation downscaling, *Comput. Geosci.* 37 (2011), pp. 1277–1284. doi: 10.1016/j.cageo.2010.07.005 [CrossRef] [Google Scholar]

12. Chao S.K., Härdle W.K., and Huang C., Multivariate factorizable expectile regression with application to fMRI data, *Comput. Stat. Data. Anal.* 121 (2018), pp. 1–19. doi: 10.1016/j.csda.2017.12.001 [CrossRef] [Google Scholar]

13. Chen J.M., On exactitude in financial regulation: Value-at-Risk, expected shortfall, and expectiles, *Risks* 6 (2018), pp. 1–29. doi: 10.3390/risks6020061 [CrossRef] [Google Scholar]

14. Ciuperca G., *Variable selection in high-dimensional linear model with possibly asymmetric or heavy-tailed errors*, Tech. rep., Universit Lyon 1 (2018).

15. Daouia A., Girard S., and Stupfler G., Estimation of tail risk based on extreme expectiles, *J. R. Stat. Soc. Ser. B* 80 (2018), pp. 263–292. doi: 10.1111/rssb.12254 [CrossRef] [Google Scholar]

16. de Pierrefeu A., Löfstedt T., Hadj-Selem F., Dubois M., Jardri R., Fovet T., Ciuciu P., Frouin V., and Duchesnay E., Structured sparse principal components analysis with the TV-elastic net penalty, *IEEE. Trans. Med. Imaging.* 37 (2018), pp. 396–407. doi: 10.1109/TMI.2017.2749140 [PubMed] [CrossRef] [Google Scholar]

17. Durbin J., Testing for serial correlation in least-squares regression when some of the regressors are lagged dependent variables, *Econometrica* 38 (1970), pp. 410–421. doi: 10.2307/1909547 [CrossRef] [Google Scholar]

18. Efron B., Hastie T., Johnstone I., and Tibshirani R., Least angle regression, *Ann. Stat.* 32 (2004), pp. 407–499. doi: 10.1214/009053604000000067 [CrossRef] [Google Scholar]

19. Ehm W., Gneiting T., Jordan A., and Krüger F., Of quantiles and expectiles: Consistent scoring functions, choquet representations and forecast rankings, *J. Royal Statist. Soc. B* 78 (2016), pp. 505–562. doi: 10.1111/rssb.12154 [CrossRef] [Google Scholar]

20. Engle R. and Manganelli S., CAViaR: conditional autoregressive value at risk by regression quantiles, *J. Business Econom. Stat.* 22 (2004), pp. 367–381. doi: 10.1198/073500104000000370 [CrossRef] [Google Scholar]

21. Fan J., Fan Y., and Barut E., Adaptive robust variable selection, *Ann. Statist.* 42 (2012), pp. 324–351. doi: 10.1214/13-AOS1191 [PMC free article] [PubMed] [CrossRef] [Google Scholar]

22. Fan J. and Li R., Variable selection via nonconcave penalized likelihood and its oracle properties, *J. Am. Stat. Assoc.* 96 (2001), pp. 1348–1360. doi: 10.1198/016214501753382273 [CrossRef] [Google Scholar]

23. Farooq M. and Steinwart I., An SVM-like approach for expectile regression, *Computat. Statist. Data Analysis* 109 (2017), pp. 159–181. doi: 10.1016/j.csda.2016.11.010 [CrossRef] [Google Scholar]

24. Fenske N., Kneib T., and Hothorn T., Identifying risk factors for severe childhood malnutrition by boosting additive quantile regression, *J. Am. Stat. Assoc.* 106 (2011), pp. 494–510. doi: 10.1198/jasa.2011.ap09272 [CrossRef] [Google Scholar]

25. Friedman J., Hastie T., and Tibshirani R., Pathwise coordinate optimization, *Ann. Appl. Stat.* 1 (2007), pp. 302–332. doi: 10.1214/07-AOAS131 [CrossRef] [Google Scholar]

26. Friedman J., Hastie T., and Tibshirani R., Pathwise coordinate optimization, *J. Stat. Softw.* 33 (2010), pp. 1–22. doi: 10.18637/jss.v033.i01 [PMC free article] [PubMed] [CrossRef] [Google Scholar]

27. Goldstein D.R., Analyzing microarray gene expression data, *J. Am. Stat. Assoc.* 100 (2005), pp. 1464–1465. doi: 10.1198/jasa.2005.s60 [CrossRef] [Google Scholar]

28. González M., Dominguez D., and Sánchez Á., Increase attractor capacity using an ensembled neural network, *Expert. Syst. Appl.* 71 (2017), pp. 206–215. doi: 10.1016/j.eswa.2016.11.035 [CrossRef] [Google Scholar]

29. Gouriéroux C. and Jasiak J., Dynamic quantile models, *J. Econom.* 147 (2008), pp. 198–205. doi: 10.1016/j.jeconom.2008.09.028 [CrossRef] [Google Scholar]

30. Gu Y. and Zou H., High-dimensional generalizations of asymmetric least squares regression and their applications, *Ann. Statist.* 44 (2016), pp. 2661–2694. doi: 10.1214/15-AOS1431 [CrossRef] [Google Scholar]

31. Guler K., Ng P.T., and Xiao Z., Mincerczarnowitz quantile and expectile regressions for forecast evaluations under aysmmetric loss functions, *J. Forecast.* 36 (2017), pp. 651–679. doi: 10.1002/for.2462 [CrossRef] [Google Scholar]

32. Hamidi B., Maillet B., and Prigent J.L., A dynamic autoregressive expectile for time-invariant portfolio protection strategies, *J. Econom. Dyn. Control* 46 (2014), pp. 1–29. doi: 10.1016/j.jedc.2014.05.005 [CrossRef] [Google Scholar]

33. Hoerl A.E., Kennard R.W., and Hoerl R.W., Practical use of ridge regression: A challenge met, *J. R. Stat. Soc. Ser. C* 34 (1985), pp. 114–120. [Google Scholar]

34. Huber P.J., Robust regression: Asymptotics, conjectures and monte carlo, *Ann. Stat.* 1 (1973), pp. 799–821. doi: 10.1214/aos/1176342503 [CrossRef] [Google Scholar]

35. Jiang C., Jiang M., Xu Q., and Huang X., Expectile regression neural network model with applications, *Neurocomputing* 247 (2017), pp. 73–86. doi: 10.1016/j.neucom.2017.03.040 [CrossRef] [Google Scholar]

36. Jones M.C., Expectiles and M-quantiles are quantiles, *Stat. Probab. Lett.* 20 (1994), pp. 149–153. doi: 10.1016/0167-7152(94)90031-0 [CrossRef] [Google Scholar]

37. Kim M. and Lee S., Nonlinear expectile regression with application to Value-at-Risk and expected shortfall estimation, *Comput. Stat. Data Anal.* 94 (2016), pp. 1–19. doi: 10.1016/j.csda.2015.07.011 [CrossRef] [Google Scholar]

38. Koenker R., Additive models for quantile regression: Model selection and confidence bandaids, *Brazilian J. Probab. Stat.* 25 (2011), pp. 239–262. doi: 10.1214/10-BJPS131 [CrossRef] [Google Scholar]

39. Koenker R. and Bassett Jr G.W., Regression quantiles, *Econometrica* 46 (1978), pp. 33–50. doi: 10.2307/1913643 [CrossRef] [Google Scholar]

40. Koenker R., Ng P., and Portnoy S., Quantile smoothing splines, *Biometrika* 81 (1994), pp. 673–680. doi: 10.1093/biomet/81.4.673 [CrossRef] [Google Scholar]

41. Kuan C.M., Yeh J.H., and Hsu Y.C., Assessing value at risk with CARE, the conditional autoregressive expectile models, *J. Econom.* 150 (2009), pp. 261–270. doi: 10.1016/j.jeconom.2008.12.002 [CrossRef] [Google Scholar]

42. Lamarche C., Robust penalized quantile regression estimation for panel data, *J. Econom.* 157 (2010), pp. 396–408. doi: 10.1016/j.jeconom.2010.03.042 [CrossRef] [Google Scholar]

43. Li Y., Liu Y., and Zhu J., Quantile regression in reproducing kernel hilbert spaces, *J. Am. Stat. Assoc.* 102 (2007), pp. 255–268. doi: 10.1198/016214506000000979 [CrossRef] [Google Scholar]

44. Li Q., Xi R., and Lin N., Bayesian regularized quantile regression, *Bayesian Anal.* 5 (2010), pp. 533–556. doi: 10.1214/10-BA521 [CrossRef] [Google Scholar]

45. Li S. and Ye W., A generalized elastic net regularization with smoothed l0 penalty, *Adv. Pure Math.* 7 (2017), pp. 66–74. doi: 10.4236/apm.2017.71006 [CrossRef] [Google Scholar]

46. Li Y. and Zhu J., L1-norm quantile regression, *J. Comput. Graph. Stat.* 17 (2008), pp. 163–185. doi: 10.1198/106186008X289155 [CrossRef] [Google Scholar]

47. Lian H., Semiparametric estimation of additive quantile regression models by two-fold penalty, *J. Bus. Econ. Stat.* 30 (2012), pp. 337–350. doi: 10.1080/07350015.2012.693851 [CrossRef] [Google Scholar]

48. Liao L., Park C., and Choi H., Penalized expectile regression: An alternative to penalized quantile regression, *Ann. Inst. Stat. Math.* 71 (2019), pp. 409–438. doi: 10.1007/s10463-018-0645-1 [CrossRef] [Google Scholar]

49. Liu J., Huang J., and Ma S., Integrative analysis of multiple cancer genomic datasets under the heterogeneity model, *Stat. Med.* 32 (2013), pp. 3509–3521. doi: 10.1002/sim.5780 [PMC free article] [PubMed] [CrossRef] [Google Scholar]

50. Mkhadri A., Ouhourane M., and Oualkacha K., A coordinate descent algorithm for computing penalized smooth quantile regression, *Stat. Comput.* 27 (2017), pp. 865–883. doi: 10.1007/s11222-016-9659-9 [CrossRef] [Google Scholar]

51. Newey W.K. and Powell J.L., Asymmetric least squares estimation and testing, *Econometrica* 55 (1987), pp. 819–847. doi: 10.2307/1911031 [CrossRef] [Google Scholar]

52. Niu J., Chen J., and Xu Y., Twin support vector regression with huber loss, *J. Intel. Fuzzy Syst.* 32 (2017), pp. 4247–4258. doi: 10.3233/JIFS-16629 [CrossRef] [Google Scholar]

53. Pecanka J., van der Vaart A.W., and Jonker M.A., Modeling association between multivariate correlated outcomes and high-dimensional sparse covariates: The adaptive SVS method, *J. Appl. Stat.* 46 (2019), pp. 893–913. doi: 10.1080/02664763.2018.1523377 [CrossRef] [Google Scholar]

54. Peng B. and Wang L., An iterative coordinate descent algorithm for high-dimensional nonconvex penalized quantile regression, *J. Comput. Graph. Stat.* 24 (2015), pp. 676–694. doi: 10.1080/10618600.2014.913516 [CrossRef] [Google Scholar]

55. Qi L. and Sun J., A nonsmooth version of newton's method, *Math. Program.* 58 (1993), pp. 353–367. doi: 10.1007/BF01581275 [CrossRef] [Google Scholar]

56. Sargan J.D. and Bhargava A., Testing residuals from least squares regression for being generated by the Gaussian random walk, *Econometrica* 51 (1983), pp. 153–174. doi: 10.2307/1912252 [CrossRef] [Google Scholar]

57. Schulze Waltrup L. and Kauermann G., Smooth expectiles for panel data using penalized splines, *Stat. Comput.* 27 (2017), pp. 271–282. doi: 10.1007/s11222-015-9621-2 [CrossRef] [Google Scholar]

58. Shen Y., Liang H.Y., and Fan G.L., Penalized empirical likelihood for quantile regression with missing covariates and auxiliary information, *Commun. Stat. Theory Methods* 47 (2018), pp. 2001–2021. doi: 10.1080/03610926.2017.1335413 [CrossRef] [Google Scholar]

59. Simon N., Friedman J., Hastie T., and Tibshirani R., Regularization paths for Cox's proportional hazards model via coordinate descent, *J. Stat. Softw.* 39 (2011), pp. 1–13. doi: 10.18637/jss.v039.i05 [PMC free article] [PubMed] [CrossRef] [Google Scholar]

60. Sobotka F., Kauermann G., Waltrup L.S., and Kneib T., On confidence intervals for semiparametric expectile regression, *Stat. Comput.* 23 (2013), pp. 135–148. doi: 10.1007/s11222-011-9297-1 [CrossRef] [Google Scholar]

61. Sobotka F. and Kneib T., Geoadditive expectile regression, *Comput. Stat. Data Anal.* 56 (2012), pp. 755–767. doi: 10.1016/j.csda.2010.11.015 [CrossRef] [Google Scholar]

62. Spiegel E., Kneib T., and Otto-Sobotka F., *Spatio-temporal expectile regression models*, Statist. Model. (2019).

63. Spiegel E., Sobotka F., and Kneib T., Model selection in semiparametric expectile regression, *Electron. J. Stat.* 11 (2017), pp. 3008–3038. doi: 10.1214/17-EJS1307 [CrossRef] [Google Scholar]

64. Taylor J.W., Estimating value at risk and expected shortfall using expectiles, *J. Financial Econom.* 6 (2008), pp. 231–252. doi: 10.1093/jjfinec/nbn001 [CrossRef] [Google Scholar]

65. Tibshirani R., Regression shrinkage and selection via the lasso, *J. R. Stat. Soc. Ser. B* 58 (1996), pp. 267–288. [Google Scholar]

66. Tseng P., Convergence of a block coordinate descent method for nondifferentiable minimization, *J. Optim. Theory Appl.* 109 (2001), pp. 475–494. doi: 10.1023/A:1017501703105 [CrossRef] [Google Scholar]

67. Waldmann E., Sobotka F., and Kneib T., Bayesian geoadditive expectile regression, *Comput. Stat. Data Anal.* 56 (2013), pp. 755–767. [Google Scholar]

68. Waldmann E., Sobotka F., and Kneib T., Bayesian regularisation in geoadditive expectile regression, *Stat. Comput.* 27 (2017), pp. 1539–1553. doi: 10.1007/s11222-016-9703-9 [CrossRef] [Google Scholar]

69. Walker M.A., Langkilde-Geary I., Hastie H.W., Wright J., and Gorin A., Automatically training a problematic dialogue predictor for a spoken dialogue system, *J. Artificial Intel. Res.* 16 (2002), pp. 293–319. doi: 10.1613/jair.971 [CrossRef] [Google Scholar]

70. Waltrup L.S., Sobotka F., Kneib T., and Kauermann G., Expectile and quantile regressiondavid and goliath? *Stat. Model.* 15 (2015), pp. 433–456. doi: 10.1177/1471082X14561155 [CrossRef] [Google Scholar]

71. Wang L., Keilegom I.V., and Maidman A., Wild residual bootstrap inference for penalized quantile regression with heteroscedastic errors, *Biometrika* 105 (2018), pp. 859–872. doi: 10.1093/biomet/asy037 [CrossRef] [Google Scholar]

72. Wang H. and Leng C., A note on adaptive group lasso, *Comput. Stat. Data Anal.* 52 (2008), pp. 5277–5286. doi: 10.1016/j.csda.2008.05.006 [CrossRef] [Google Scholar]

73. Wang X. and Wang M., Variable selection for high-dimensional generalized linear models with the weighted elastic-net procedure, *J. Appl. Stat.* 43 (2016), pp. 796–809. doi: 10.1080/02664763.2015.1078300 [CrossRef] [Google Scholar]

74. Wirsik N., Otto-Sobotka F., and Pigeot I., Modeling physical activity data using L0-penalized expectile regression, *Biometrical J.* 61 (2019), pp. 1371–1384. doi: 10.1002/bimj.201800007 [PubMed] [CrossRef] [Google Scholar]

75. Wold S., Ruhe A., Wold H., and Dunn III W.J., The collinearity problem in linear regression. The partial least squares (PLS) approach to generalized inverses, *SIAM J. Sci. Stat. Comput.* 5 (1984), pp. 735–743. doi: 10.1137/0905052 [CrossRef] [Google Scholar]

76. Wu Y. and Liu Y., Variable selection in quantile regression, *Stat. Sin.* 19 (2009), pp. 801–817. [Google Scholar]

77. Wu T., Yu K., and Yu Y., Single-index quantile regression, *J. Multivar. Anal.* 101 (2010), pp. 1607–1621. doi: 10.1016/j.jmva.2010.02.003 [CrossRef] [Google Scholar]

78. Xie S., Zhou Y., and Wan A.T.K., A varying-coefficient expectile model for estimating value at risk, *J. Bus. Econ. Stat.* 32 (2014), pp. 576–592. doi: 10.1080/07350015.2014.917979 [CrossRef] [Google Scholar]

79. Xu Q., Cai C., Jiang C., Sun F., and Huang X., Sampling lasso quantile regression for large-scale data, *Commun. Stat. Simulat. Comput.* 47 (2018), pp. 92–114. doi: 10.1080/03610918.2016.1277750 [CrossRef] [Google Scholar]

80. Xu Q., Liu X., Jiang C., and Yu K., Nonparametric conditional autoregressive expectile model via neural network with applications to estimating financial risk, *Appl. Stoch. Models. Bus. Ind.* 32 (2016), pp. 882–908. doi: 10.1002/asmb.2212 [CrossRef] [Google Scholar]

81. Yang Y., Zhang T., and Zou H., Flexible expectile regression in reproducing kernel Hilbert spaces, *Technometrics* 60 (2018), pp. 26–35. doi: 10.1080/00401706.2017.1291450 [CrossRef] [Google Scholar]

82. Yang Y. and Zou H., An efficient algorithm for computing the HHSVM and its generalizations, *J. Comput. Graphical Stat.* 22 (2013), pp. 396–415. doi: 10.1080/10618600.2012.680324 [CrossRef] [Google Scholar]

83. Yang Y. and Zou H., Nonparametric multiple expectile regression via ER-Boost, *J. Stat. Comput. Simul.* 85 (2015), pp. 1442–1458. doi: 10.1080/00949655.2013.876024 [CrossRef] [Google Scholar]

84. Yao Q. and Tong H., Asymmetric least squares regression estimation: A nonparametric approach, *J. Nonparametr. Stat.* 6 (1996), pp. 273–292. doi: 10.1080/10485259608832675 [CrossRef] [Google Scholar]

85. Yi C. and Huang J., Semismooth newton coordinate descent algorithm for elastic-net penalized huber loss regression and quantile regression, *J. Comput. Graph. Stat.* 26 (2017), pp. 547–557. doi: 10.1080/10618600.2016.1256816 [CrossRef] [Google Scholar]

86. Yu K. and Jones M., Local linear quantile regression, *J. Am. Stat. Assoc.* 93 (1998), pp. 228–237. doi: 10.1080/01621459.1998.10474104 [CrossRef] [Google Scholar]

87. Yuan M. and Lin Y., Model selection and estimation in regression with grouped variables, *J. R. Stat. Soc. Ser. B* 68 (2006), pp. 49–67. doi: 10.1111/j.1467-9868.2005.00532.x [CrossRef] [Google Scholar]

88. Zhang C.H., Nearly unbiased variable selection under minimax concave penalty, *Ann. Stat.* 38 (2010), pp. 894–942. doi: 10.1214/09-AOS729 [CrossRef] [Google Scholar]

89. Zhao J., Chen Y., and Zhang Y., Expectile regression for analyzing heteroscedasticity in high dimension, *Stat. Probab. Lett.* 137 (2018), pp. 304–311. doi: 10.1016/j.spl.2018.02.006 [CrossRef] [Google Scholar]

90. Zhao J., Yan G., and Zhang Y., *Semiparametric expectile regression for high-dimensional heavy-tailed and heterogeneous data*, Tech. rep., Zhejiang University City College (2019)

91. Zhao J. and Zhang Y., Variable selection in expectile regression, *Commun. Stat. Theory Methods* 47 (2018), pp. 1731–1746. doi: 10.1080/03610926.2017.1324989 [CrossRef] [Google Scholar]

92. Zou H., The adaptive lasso and its oracle properties, *J. Am. Stat. Assoc.* 101 (2006), pp. 1418–1429. doi: 10.1198/016214506000000735 [CrossRef] [Google Scholar]

93. Zou H. and Hastie T., Regularization and variable selection via the elastic net, *J. R. Stat. Soc. Ser. B* 67 (2005), pp. 301–320. doi: 10.1111/j.1467-9868.2005.00503.x [CrossRef] [Google Scholar]

94. Zou H. and Yuan M., Composite quantile regression and the oracle model selection theory, *Ann. Stat.* 36 (2008), pp. 1108–1126. doi: 10.1214/07-AOS507 [CrossRef] [Google Scholar]