A comparative study of different variable selection methods based on numerical simulation and empirical analysis

Dake Hou [1] , **Wenli Zhou** [2] , **Qiuxia Zhang** [3] , **Kun Zhang** [3] , **Jiaqi Fang** [Corresp. 2]

[1] School of Mathematics, Shandong University, Jinan, China

[2] School of Business, Wenzhou Univeristy, Wenzhou, China

[3] School of Economics and Management, Beijing Jiaotong University, Beijing, China

Corresponding Author: Jiaqi Fang
Email address: 20210634@wzu.edu.cn

In this paper, based on statistics and computer science, four common classical variable selection methods (Lasso, Elastic-Net, Adaptive Lasso and SCAD) are compared and analyzed through numerical simulation and empirical analysis, so as to realize the application of data science. To measure the performance of the built linear random effects model, we use four indicators: variable selection consistency, model prediction accuracy, stability, and efficiency. Regarding the consistency of variable selection, based on the geometric interpretation of the Pearson correlation coefficient formula, this paper proposes to analyze by calculating the angle between the unitized estimated coefficient vector β̂ and the unitized true coefficient vector β, and by using the boxplot tool the distribution of included angles is visually analyzed. The stability of the model is judged from one side according to the number of outliers. The research shows that Adaptive Lasso and SCAD have slightly better variable selection ability than Elastic-Net and Lasso. However, under the influence of random effects, the model coefficients fluctuate significantly due to random effects, while Elastic-Net and Lasso are relatively stable, and the mean square error and stability of each model are affected by random factors to a similar degree. Therefore, in the actual application process, the appropriate variable selection method can be selected to fit the model according to the needs.

A comparative study of different variable selection methods based on numerical simulation and empirical analysis

Different variable selection methods

Dake Hou[1], Wenli Zhou[2], Qiuxia Zhang[3], Kun Zhang[3], Jiaqi Fang[2*]

[1] School of Mathematics, Shandong University, Jinan, Shandong, China.
[2] School of Business, Wenzhou University, Wenzhou, Zhejiang, China.
[3] School of Economics and Management, Beijing Jiaotong University, Beijing, China.

Corresponding Author:
Jiaqi Fang
Wenzhou University, Wenzhou, Zhejiang, 325035, China.
Email address:20210634@wzu.edu.cn

Abstract

In this paper, based on statistics and computer science, four common classical variable selection methods (Lasso, Elastic-Net, Adaptive Lasso and SCAD) are compared and analyzed through numerical simulation and empirical analysis, so as to realize the application of data science. To measure the performance of the built linear random effects model, we use four indicators: variable selection consistency, model prediction accuracy, stability, and efficiency. Regarding the consistency of variable selection, based on the geometric interpretation of the Pearson correlation coefficient formula, this paper proposes to analyze by calculating the angle between the unitized estimated coefficient vector β and the unitized true coefficient vector β, and by using the boxplot tool the distribution of included angles is visually analyzed. The stability of the model is judged from one side according to the number of outliers. The research shows that Adaptive Lasso and SCAD have slightly better variable selection ability than Elastic-Net and Lasso. However, under the influence of random effects, the model coefficients fluctuate significantly due to random effects, while Elastic-Net and Lasso are relatively stable, and the mean square error and stability of each model are affected by random factors to a similar degree. Therefore, in the actual application process, the appropriate variable selection method can be selected to fit the model according to the needs.

Keywords: Linear random effect model; Variable selection; Coefficient consistency; Prediction accuracy; Boxplot; Stability

Introduction

Model selection is the key point of statistics. To achieve the goal of high prediction accuracy, strong inference and interpretation power of statistical modeling. Modern statistical theories and methods in the field of statistics are changing with each passing day. The classical linear statistical model is an effective and widely used statistical methods, in solving the problem of reality with inference ability that nonlinear model cannot surpass, and many new methods based on it, such as polynomial regression, spline regression, regression, generalized additive models and partial linear model can be regarded as is the extension and expansion. Variable selection is an important link in modeling, statistical disciplines for comparative analysis of various methods which are hot research topics, a suitable model should be from the early stages of modeling to introduce a large number of covariate parsing out the real response variables related elements, in order to achieve the model which has strong ability of inference, interpretation and prediction accuracy of the target, the model should be stable, the results of variable

selection should not be destabilized by random noise and contaminated data. Longitudinal data with random effects that appear in many fields such as biomedicine, clinical trials, meteorological observation, industrial engineering and e-commerce platforms are important data types. Therefore, we conduct a comparative study on Lasso variable selection methods based on the linear random effects model.
Scholars around the world have studied the theoretical research and application practice of Lasso regression methods, which also proposed many Lasso methods to solve variable selection problems of different models. Wang and Leng studied the *Adaptive group Lasso method* [1]. Yuan and Lin studied the *Graphical Lasso* method [2]. In terms of single-indicator model research, Wang and Yin proposed the *sMave* method with the corresponding iterative algorithm [3]. Zeng and He studied the estimation and variable selection of the single-index model which needs to solve a large number of sample data with derivatives are zero, selected the penalty partial derivative, and proposed *sim-Lasso* method [4], which could eliminate the effect of the point on the estimation where the partial derivative was zero with the coefficient of compression.
In the field of high and ultra-high dimensional models, there are a lot of academic achievements in the research and application of Lasso methods. For instance, Fan and Lv studied the application of ultra-high-dimensional variable selection method in the generalized linear model and proposed *Sure Independence Screening (SIS)* [5]. which selects the marginal Screening method, according to the Pearson correlation coefficient between the response variable and the single covariable, measured the importance of each covariable, and the dimension of the covariable is reduced to the extent that the traditional method of selecting penalty variables can be used effectively. Yuan and Lin established the systematic connection between Lasso and Bayes, and used *LARS* algorithm to complete the calculation of Bayesian posterior distribution in high-dimensional problems, they selected the method of maximizing marginal likelihood to provide a feasible method for Lasso's penalty parameter selection [6]. Since then, many papers were published to study the connection between various classical methods and Bayes. Yuan Jing used *Inverse Bayes Formulae (IBF)* to propose two new algorithms based on non-iterative sampling technology, which can solve Bayes Lasso problems quickly and effectively [7]. Li Hanfang improved Bayes Lasso estimation and proposed Bayesian Adaptive Lasso estimation equivalent to *Adaptive Lasso* [8][9]. In 2001, Statisticians Fan and Li proposed a new non-convex penalty function which called *Smoothly Clipped Absolute Deviation (SCAD)*, the punishment function is a continuous differentiable piecewise function. These kinds of literature show that the estimated value obtained by this method meets the properties of Oracle [10] .
In this study, the linear random effect models built by Lasso variable selection methods (*Lasso, Elastic-Net, Adaptive Lasso and SCAD*) are compared and analysed in terms of variable selection consistency, model prediction accuracy and stability.
The consistency respectively based on the formula of Pearson correlation coefficient of the geometric interpretation, we analysed the Angle between calculating unit $\hat{\beta}$ estimated coefficient vector and the real beta coefficient vector β, and using the boxplot tool intuitive analyze the distribution of the Angle and outliers, according to outliers from the side to evaluate the stability of the model. We use numerical simulation and empirical analysis to compare and analyze Lasso variable selection methods. The results show that the model analysis results what we obtained from the study are consistent with the theoretical analysis, further more, which is also consistent with the analysis results obtained by calculating the true positive rate (sensitivity), true negative rate (specificity) and calculating the probability of the predictor variable entering the models which were proposed in the relevant research.
In this paper, the penalty parameter λ of the model is obtained by the ten-fold cross-validation method, the α parameter in SCAD is 3 according to the recommendation of Fan and Li, and the weight coefficient γ in Adaptive Lasso is 1. But there are many options beyond that to determine these parameters in the penalty function. Secondly, since a large number of actual statistical data sets usually have data pollution problems, more comprehensive discussions are needed based on the work done in this paper.

Model and variable selection methods

A classical linear regression model is described as:

$$y_i = \beta_0 + \sum_{j=1}^{p} x_{ij}\beta_j + e_i \quad i = 1,2,...,n \tag{1}$$

There are n groups of observations, each of which consists of an output variable y_i and p associated predictive variables $x_i = (x_{i1},x_{i2},...,x_{ip})^T$. β_0 and $\beta = (\beta_1,\beta_2,...,\beta_p)^T$ are unknown parameters and e_i is the error term. The goal of linear regression is to predict the output from the predictors and to find out which predictors are important. Estimating the unknown parameter β is the core work. The traditional method is to minimize the least square method of the objective function.

$$\min_{\beta_0,\beta} mize \sum_{i=1}^{n} (y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j)^2 \tag{2}$$

Generally, in Equation 2, all the least squares estimates of β are not equal to zero. If p is large, the final model is difficult to explain. Moreover, when $n < p$, the results of the least squares estimates are not unique, and there are infinitely many solutions that can make the objective function equal to zero. Therefore, it is necessary to constrain (regularize) the estimation process, that is, to introduce a penalty function. According to the degree of concavity, the penalty function can be divided into convex punishment (*e.g., Lasso, Adaptive Lasso*) and non-convex punishment (*e.g., SCAD, MCP, etc.*). The convex function is defined on a convex set S which is meet the conditions: $\forall x_1,x_2 \in S, \forall a \in [0,1]$, there is

$$f(ax_1 + (1-a)x_2) \le af(x_1) + (1-a)f(x_2).$$

(*Definition of S: for* $\forall x_1,x_2 \in S, \forall a \in [0,1]$, *when* $x = ax_1 + (1-a)x_2,\ x \in S$)

The convex punishment can ensure the uniqueness of the solution, and there are efficient algorithms to get the estimator which has good stability and sparsity, but it is a partial estimator without Oracle properties.

The Non-convex penalty can realize the sparsity of coefficients which also meet the Oracle properties, but its non-convex properties cannot guarantee the uniqueness of solutions which will produce multiple local optima, so that ultimately leads to relatively poor stability of results. Moreover, a concave parameter is added to the penalty function, which increases the difficulty of calculation.

The Penalty function can be divided into（*Penalized Residual Sums of Squares*）and（*Penalty likelihood function*）according to different models applied. The sum of squares of the penalty residuals is used for a general linear regression model LM in which the dependent variable is normally distributed, the joining function is the identity, and the loss function is the sum of squares of residuals. The penalty likelihood function is applicable to Poisson distribution, Binomial distribution, Gamma distribution of dependent variables, the exponential cluster of connection function, loss function is a generalized linear model of likelihood function, such as Logistic regression model, Poisson regression model.

With parameter set $\beta = (\beta_1,\beta_2,...,\beta_p)^T$ as the study object, on the basis of *OLS*, increase the constraint conditions about the β, the penalty function $p_\lambda(|\beta|)$, thus, to establish the punishment least squares estimation (*Penalized Least Squares, PLS*)

$$\hat{\beta} = argmin\left\{\sum_{i=1}^{n}(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j)^2 + \sum_{j=1}^{p} p_\lambda(|\beta_j|)\right\} \tag{3}$$

Different variable selection methods can be constructed when we select different penalty functions.

Lasso method and basic characteristics

The statistician Tibshirani (1996) [13] proposed a variable selection method. The core idea of Lasso（*Least Absolute Shrinkage and Selection Operator*）is to introduce penalty factors to constrain the L_1 norm of the estimator β on the basis of ordinary Least square estimation.

As for the data set with p predictive variables and n predictive variable-response variable pairs $\{(x_i,y_i)\}_{i=1}^{n}$, we could use Lasso to find an estimate of $\hat{\beta}$ that can better fit the data through minimization of $RSS(\hat{\beta})$.

$$RSS(\hat{\beta}) = \underset{\hat{\beta}}{\text{argmin}} \sum_{i=1}^{n}\left(y_i - \sum_{j=1}^{p}\beta_j x_{ij}\right)^2 + \lambda\sum_{j=1}^{p}|\beta_j| \tag{4}$$

$\lambda \geq 0$ is the Penalized Parameter, $\lambda\sum_{j=1}^{p}|\beta_j|$ is compression constraints.

$$\hat{\beta} = argmin\left\{\|y - x\beta\|_2^2\right\} \quad s.t \ \ \|\beta\| \leq t. \tag{5}$$

$t \geq 0$ is a tuning Parameter, which controls the intensity of compression. If the Parameter obtained by the least squares is estimated as β^0, Lasso can be compression as long as $t < \sum_{j=1}^{p}\left|\widehat{\beta_j^0}\right|$. In addition, for some models with small absolute values, the coefficient is compressed to zero. Therefore, the inequality $\|\beta\| \leq t$ effectively limits the parameter space and makes the final model explicable.
It can be proved that the relation between t and λ [14] is

$$t = \sum_{j=1}^{p} sign(\widehat{\beta_j^0}) \cdot \widehat{\beta_j^0} - p\tfrac{\lambda}{2}. \tag{6}$$

The estimator $\hat{\beta}$ can be obtained that using numerical approximation algorithms (*commonly coordinate descent and minimum Angle regression*).
The basic characteristics of Lasso are as follows: in essence it is a process of seeking the sparse representation of the model. This process is accomplished by optimizing a "loss + penalty" function problem. However, when there is a group of highly correlated characteristic predictive variables, Lasso regression tends to choose one of them and ignore the others, resulting in unstable results, that is, collinearity cannot be accurately and effectively dealt with.

Adaptive Lasso method and basic characteristics

It can be seen from Lasso analysis that Lasso penalty function is the same for all the estimators in $\beta = (\beta_1, \beta_2, \cdots\cdots, \beta_p)$. However, if the penalty function can be made to give a smaller penalty to the larger estimators, thus effectively reducing the model's bias while ensuring the sparsity of the model. Therefore, Zou (2006) proposed Adaptive Lasso [15], in which the punishment parameter can be adjusted according to the size of the estimator, instead of λ being a fixed punishment in Lasso. Before the Adaptive Lasso in $\|\beta_j\|$ increase weight coefficient $\hat{w}_j = \frac{1}{|\hat{\beta}_j|^\gamma}$, γ constant is greater than zero, Zou advised to choose a $\sqrt{n}$ consistent estimator satisfy, which uses least squares estimator $\hat{\beta}$ as the initial estimator, $\hat{\beta}_j = \hat{\beta}^{OLS}$, Adaptive Lasso is in the form of

$$\hat{\beta} = arg\min_{\beta} \|y - X\beta\|_2^2 + \lambda\sum_{j=1}^{p}\hat{w}_j\|\beta_j\|_1. \tag{7}$$

By formula 7, when $\gamma \geq 1$, if some components of the initial estimate of $\hat{\beta}$ are large (such as $\hat{\beta}_j$), the weight of the corresponding punishment item weight coefficient $\hat{w}_j = \frac{1}{|\hat{\beta}_j|^\gamma}$ relative to other components, so the Adaptive - Lasso has the following features:
(1) The penalty function reduces the Lasso estimator bias while ensuring the estimator
(2) Zou (2006) [15] also proves that Adaptive Lasso can satisfy the following Oracle properties with an appropriate λ value:
(a) The consistency of variable selection:

$$\left\{j, \hat{\beta}_j \neq 0\right\} = \left\{j, \beta_j \neq 0\right\} \triangleq S_0.$$

(b) Asymptotic normality:

$$\sqrt{n}\left(\hat{\beta} - \beta_{S_0}\right)\xrightarrow{d}N(0,\sigma^2).$$

* S_0 is the active set of predictive variables

$$S_0 = \left\{j:\beta_j^0 \neq 0, j = 1,2,\cdots\cdots,p\right\},$$

$$\hat{S}(\lambda) = \left\{j:\hat{\beta}_j(\lambda) \neq 0, j = 1,2,\cdots\cdots,p\right\},$$

$\hat{S}(\lambda)$ Is a nonzero coefficient subscript set of parameters estimated by Lasso method. According to the CV method to get $\hat{\lambda}_{CV}$, eventually get $\hat{S}(\hat{\lambda}_{CV})$ which has a high probability to include S_0 and $\left|\hat{S}(\hat{\lambda}_{CV}) \leq min\ (n,p)\right|$.

(3) Adaptive Lasso also satisfies the three features of the penalty function proposed by Fan and Li, namely unbiasedness, sparsity and continuity, which are improved on the basis of Lasso, but it still cannot effectively deal with collinearity.

Elastic Net method and basic features

Ridge regression is characterized with evenly distributing weights to related characteristic variables, the Lasso method cannot accurately and effectively deal with collinearity, so that Zou.H and Hastic (2005) proposed Elastic Net by combining the advantages of Ridge regression and Lasso regression [16], which Is the convex combination of ridge regression and Lasso regression:

$$\hat{\beta} = arg\min_{\beta \in R^p}\left\{\|y - X\beta\|_2^2 + \lambda_1\sum_{j=1}^{p}|\beta_j| + \lambda_2\sum_{j=1}^{p}\beta_j^2\right\} \tag{8}$$

The penalty term is $P_\lambda(|\beta_j|) = \lambda_1|\beta_j| + \lambda_2|\beta_j|^2$, λ_1, λ_2 are two nonnegative punishment parameters, order $a = \frac{\lambda_2}{\lambda_1 + \lambda_2}$,

$$\hat{\beta} = arg\min_{\beta \in R^p}\left\{\|y - X\beta\|_2^2 + (1 - a)\sum_{j=1}^{p}|\beta_j| + a\sum_{j=1}^{p}\beta_j^2\right\} \tag{9}$$

Penalty term $P_\lambda(|\beta_j|) = (1 - a)\sum_{j=1}^{p}|\beta_j| + a\sum_{j=1}^{p}\beta_j^2$ at zero point is the guide, for $\forall a > 0$ is strictly convex.

$$\text{To: } X^*_{(n+p)\times p} = \frac{1}{\sqrt{1+\lambda_2}}\binom{X}{\sqrt{\lambda_2}I}, \qquad Y^*_{n+p} = \binom{Y}{0},$$

$$\gamma = \frac{\lambda_1}{\sqrt{1+\lambda_2}}, \qquad \beta^* = \sqrt{1+\lambda_2}\beta,$$

then

$$\hat{\beta}^* = arg\min_{\beta^* \in R^p}\left\{\left\|y^* - X^*\beta^*\right\|_2^2 + \gamma\sum_{j=1}^{p}\left|\beta_j^*\right|\right\}.$$

Therefore, the Elastic Net problem can be converted to the Lasso problem. The optimal Elastic Net solution $\hat{\beta}^*$ can be found on $\hat{\beta} = \frac{1}{\sqrt{1+\lambda_2}}\hat{\beta}^*$. Because $I_{p \times p}$ is full rank, considering the composition of $X^*_{(n+p)\times p}$ and it has p columns, the Lasso matrix $X^*_{(n+p)\times p}$ (matrix rank p), the applied matrix is $X_{(n+p)\times p}$. So Elastic Net can select up to p variables which solved Lasso can only select a maximum of n（$n < p$） variables. Zou and Hastie (2005) proved that Elastic Net can select group variables [16], that is, for several highly correlated independent variables, Elastic Net can select all these variables, which solves the problem that Lasso regression methods tend to select one and ignore the others. It is suitable for some application scenarios that need to analyse the relationship between dependent variables and predictive variables with group characteristics, such as variable screening and prediction of gene expression profile data.

SCAD method and basic features

Fan and Li (2001) proposed a non-convex penalty function [17] to achieve the unbiasedness of the estimator β (that is, the penalty on the coefficient is guaranteed to decrease with the increase of the coefficient estimator), thus guaranteeing the approximate unbiasedness of the large coefficient.

$$\hat{\beta} = \underset{\beta \epsilon \mathrm{R}^p}{argmin}\left\{\|y - X\beta\|_2^2 + \sum_{j=1}^{p} P_\lambda(|\beta_j|)\right\}, \quad (10)$$

$P_\lambda(|\beta_j|)$P is the penalty term. The proposed method is a continuously differentiable penalty function, called smooth cohesion absolute deviation penalty SCAD（*Smoothly Clipped Absolute Deviation Penalty*）.

$$\mathrm{P}(\beta_j|_{\lambda,a}) = \begin{cases} \lambda|\beta_j| & |\beta_j| \le \lambda, \\ \dfrac{2\mathrm{a}\lambda|\beta_j| - |\beta_j|^2 - \lambda^2}{2(a-1)} & \lambda < |\beta_j| < a\lambda, \\ \dfrac{(a+1)\lambda^2}{2} & |\beta_j| \ge a\lambda, \end{cases} \quad (11)$$

Take the derivative of the penalty function *SCAD*, get

$$\mathrm{P}'(\beta_j|_{\lambda,a}) = \begin{cases} sgn(\beta_{\mathrm{j}})\lambda & |\beta_j| \le \lambda \\ sgn(\beta_{\mathrm{j}})(\alpha\lambda - |\beta_{\mathrm{j}}|)/(\alpha - 1) & \lambda < |\beta_j| < a\lambda \\ 0 & |\beta_j| \ge a\lambda \end{cases} \quad (12)$$

Thus, its coefficient estimate speed of punished decreases with the coefficient estimators $|\beta_j|$ increases, namely when the β_j absolute value is larger, SCAD penalty is constant, when the β_j absolute value is small, the smaller the absolute value of the coefficient of the degree of compression than the LASSO method is bigger, therefore, better able to induce the sparse structure, t is convenient for screening and obtaining a sparse subset of variables. Therefore, SCAD guarantees the unbiasedness of large coefficients. Not differentiable at the origin, but continuously differentiable at $(-\infty,0) \cup (0,+\infty)$, a local quadratic approximation is used to obtain a local optimal solution while ensuring sparsity and continuity. Fan, Huang, and Kim respectively proved that N-SCAD (conventional), H-SCAD (high dimensional), and UH-SCAD (super high dimensional) have Oracle properties under certain hypothetical conditions, and have the advantage over Adaptive Lasso in that it does not require prior access to the consistent estimator of $\sqrt{n}$ to $\hat{\beta}^0$. The Local optimal solution is not necessarily the global optimal solution, numerical methods usually get some specific local solution, which makes the gap between theory and practical application, research on this aspect is also one of the hot topics.

Numerical simulation and comparative analysis

In this study, we compare and analyse the performance of the penalty function of Lasso, Adaptive Lasso, Elastic-Net and SCAD variable selection methods under different conditions by numerical simulation, and their applicable application scenarios.

Four indicators were used to evaluate the model performance: (1) consistency of model variable selection, (2) Err (square error), (3) efficiency of model algorithm, (4) stability of model. As for the consistency analysis of model variable selection, we do not adopt the false discovery rate and false exclusion rate methods [18]][19] [20][21] or calculate the probability of predictive variables entering the model [22][23]. However, we choose variable selection compatibility of the model to be found through calculating the Angle between uniformed estimated coefficient vector $\hat{\beta}$ and the uniformed true coefficient vector β Angle $= \frac{180}{\pi}\arccos\left(\beta^T\hat{\beta}\right)$ reflection coefficient estimation accuracy, namely consistency ($\{j:\hat{\beta}_j \neq 0\} = \{j:\beta_j \neq 0\} \triangleq S_0$; *$S_0$ is the active set of predictive variables*). The Angle is closer to zero, the higher the degree of compatibility between β and $\hat{\beta}$, the stronger the model interpretation ability (*geometric interpretation of Pearson correlation coefficient formula* [11][12]). $Err = \frac{1}{n}\sum(\hat{y}_0 - y_0)^2$, the mean square error of the predicted variable obtained from the training set which can be carried to the model, the mean square error of the response variable $\hat{y}_0$ and the actual response value y_0 in the test set can reflect the prediction accuracy of the model. If the predicted response value is close to the actual response value, ΔErr will be small; As for the stability of the model, the boxplot tool was used to analyse the outliers of Angle and Err under different variable selection methods to reflect the stability of the established model from one side. The efficiency of the model algorithm is evaluated the time required by the computer to complete a regression fitting.

Design of simulated data set

Our data are generated from the following linear random effects model

$$Y = X\beta + Z\delta + \varepsilon \tag{13}$$

$y_{n\times 1}$ is the response variable, $X_{n\times p}$ is the design matrix of the prediction variable, $Z_{n\times q}$ is the design matrix of the random effect variable, which is constructed in the same way as *X*, $q \leq p$. $\beta_{p\times 1}$ is a nonrandom parameter vector called deterministic effect, $\delta_{q\times 1}$ is a random parameter vector called random effect, $\varepsilon_{n\times 1}$ is a random noise vector. δ follows the normal distribution with zero mean and *G* covariance matrix. $\delta \sim N(0,G)$, ε follows normal distribution, $\varepsilon \sim N(0,R)$, $R = \sigma^2 I_n$, $cov(\delta,\varepsilon) = 0$. In order to design the simulation data set, the function mvrnorm() in MASS package is firstly used to design the $n \times p$ dimensional multivariate normal simulation matrix about the prediction variable *X*, $X_{n\times p}$ is the prediction variable, let $\beta_{p\times 1} = (1,1.2,\text{-}3.2,4.1,\text{-}5,\underbrace{0,\ldots 0}_{45})$ is the coefficient vector, and $y_{n\times 1}$ is the response variable. The model discussed is the sparse model, $\varepsilon_{n\times 1}$ is the random noise vector, which is independently identically distributed and follows the standard normal distribution, $\varepsilon_{n\times 1} \sim N(0,I_n)$, *X* and ε are independent. The numerical simulation experiment was repeated for 300 times, and the data in the X matrix was updated each run. In this study, the mean value of the prediction variables in each column of the X matrix was designed to be a random variable *mu*~ $N(0,I_n)$ subject to normal distribution instead of constant, to simulate the internal fluctuation of the prediction variables and the non-homogeneity of the random factors $Z\delta$ between individuals. The fitted model has linear random effects. The degree of collinearity between predictor variable X is closely related to the correlation coefficient between them, the design of experiments with correlation coefficient to measure the collinearity, completely collinearity corresponding correlation coefficient is 1, the collinearity does not exist if the correlation coefficient is 0. The correlation coefficient is approaching to 1, collinearity between variables will be more and more strong.

The data set was divided into training set and test set, and four variable selection methods, Lasso, Elastic-Net, SCAD and Adaptive Lasso, were used for regression modelling respectively. The numerical simulation experiment was conducted m=300 times, and the average value of the four indicators was taken as the basis of the evaluation model.

The numerical simulation is carried out in the following two scenarios.

(1) $n \gg p$, where the number of observed samples n is much larger than the conventional data set with the number of predictive variables P, n:p=2500:50; The correlation coefficient between the predictive variables was cor=0.2/0.4/0.6/0.8/0.98

(2) high dimensional data set which satisfying $log\,(p) = n^a$ $(0 < a < 1)$

n:p=40:50; Correlation coefficient cor=0.2/0.4/0.6/0.8/0.98

We use R to analyze and complete numerical simulation, mainly using the following software packages that MASS, car, psych, glmnet, ncvreg, msgps, ISLR, etc. We select parameters in software analysis:
For SCAD, we used ncvreg(), where the gamma parameter selection is 3.7 (*i.e., the punishment function where a is recommended by Fan*); alpha=1 (*MCP/SCAD penalty*); When n<p, lambda.min is set to 0.05 instead of 0.001;
For Adaptive Lasso, we used msgps() : where gamma parameter selection is 1 (*i.e., γ in the penalty function*); Initial estimators, as suggested by Zou when $n \gg p$, select the minimum square estimator that meets the requirements of $\sqrt{n}$ convergence $\hat{\beta}_j = \hat{\beta}^{OLS}$ as the initial estimator, corresponding to the parameter $Lambda = 0$; If n<p, $Lambda = 0.001$ as the initial estimate of $\hat{\beta}_j = \hat{\beta}^{ridge}$.
For Elastic-net, we used msgps() where alpha=0.5;
For Lasso, we used msgps() where alpha defaults to 1;
For SCAD, we used local quadratic approximation to obtain the locally optimal solution. We can obtain the Lambda of Lasso 、Adaptive Lasso and Elastic-net that used Generalized-Cross-Validationg（*GCV*）to solve the optimal value and obtain the corresponding coefficient value $\hat{\beta}$.

Comparative analysis

(1) First examine n:p=2500:50(*number of samples on the training set: number of predictors*); Number of trial repeats *m*=300, the correlation coefficient is *cor*=0.2/0. 4/0.6/0.8/0.98
Numerical simulation obtains Angle 、Err boxplot Figure 1, Time data Table 1 and Angle mean-correlation coefficient, Err mean~correlation coefficient relationship curve Figure 2.
(2)Comparative analysis. Under the condition of the large sample (n≫p), without considering the stability of the model, When cor<0.8, Investigate the model prediction accuracy and model coefficient coincidenceErrAngle, SCAD performed best, followed by Adaptive-Lasso and Lasso. This is because in the design of the penalty function, SCAD can achieve an approximate unbiased estimation of large coefficients. Theoretically, Adaptive-Lasso can improve variable selectivity by weighting corrections to the Lasso penalty terms, However, the results of this numerical simulation experiment show that the two perform basically the same in terms of variable selectivity（For Adaptive Lasso iterative solution of the minimum loss function, Zou (2006) proposed that to obtain satisfactory Oracle properties which required an initial estimator $\hat{\beta}^0$ with fully satisfies the $\sqrt{n}$ consistency. It is difficult to achieved whaterver using coordinate descent method or gradient descent method for iterative solution）.The models fitted in these four methods all perform well in terms of prediction accuracy (*Err≤1.005*), where the irreducible error ε~N(0,1) is an upper bound on the prediction accuracy of y, maximum value is 1(*i.e.*$\Delta Err \leq 0.005$).
However, when there is a set of highly correlated predictors *(i.e. when cor>0.8*), the variable selection compatibility Angle of the built model deteriorates rapidly, but the Err of the model is still small, which is basically not affected by the increase of the correlation coefficient impact. Therefore, the model built under the condition of severe collinearity of the predictor variables is suitable for prediction (*that is, predicting the outcome through the predictor variables*), but not suitable for inference (*exploring the relationship between each predictor variable and the response variable*).
The model stability and algorithm efficiency can be judged by Angle、Err boxplot and regression fitting timetable Time. The characteristics of the convex penalty function can bring good stability to the models built by Lasso, Elastic-Net and Adaptive-Lasso.
Lasso algorithm is simple and efficient, so the model fitting speed is the fastest. Therefore, Lasso is very suitable for modeling when *n>p* and the collinearity between prediction variables is small. While Elastic-Net performs best in terms of stability when the correlation coefficient becomes large, which is suitable for the application scenarios where the predictive variables have group characteristics. Due to the non-convex nature of SCAD's penalty functions, the model form is more complex than that of Lasso, Elastic-Net and Adaptive Lasso, the iterative algorithm runs significantly slower, and the instability of the local

quadratic approximation algorithm is manifested Angle 、Err in the boxplot with the largest number of outliers, and the model fitted with this method has the problem of poor stability.

(3) Next, consider the *n*:*p*=40:50 (*number of samples on the training set: number of predictors*) high-dimensional dataset; The number of test replicates *m*=300 and the correlation coefficient was *cor*=0.2/0.4/0.6/0.8/0.98

Numerical simulation obtains Angle、Err boxplot Figure 3, Time data Table 2, and Angle mean-correlation coefficient, Err mean-correlation coefficient relationship curve Figure 4.

(4)Comparative analysis. Examining the indicators of Angle and Err, SCAD still has the best performance, but compared with the condition of large sample data, the Angle becomes several to dozens of times larger. It can be seen that under the condition of high-dimensional data, the interpretability of the models built by the four classical methods is very weak. However, the prediction errors of the models built by Lasso, SCAD and Adaptive-Lasso variable selection methods are less than 3, which can be used in application scenarios that do not require high prediction accuracy. However, when the correlation coefficient between variables is greater than 0.8, the indicators of each model deteriorate rapidly, and it can be considered that the model has lost its guiding significance.

Through numerical simulation, it is proved that the Lasso-like classical variable selection method suitable for conventional data is not suitable for modelling high-dimensional data. Therefore, in the past ten years, the field of statistics at home and abroad has ushered in a research upsurge on high-dimensional models. The modelling theories and application practices of different types of high-dimensional data have emerged in large numbers [24], becoming a powerful tool for processing massive amounts of information in the era of big data.

In order to facilitate the comparison with other variable selection methods in the literature, this paper also shows that under the conditions of correlation coefficient cor=0.6 (weak collinearity), large sample and high dimensionality, the estimated value of each model coefficient can correctly select the selection rate of predictor variables and rejection rate indicator that correctly culls irrelevant variables.

True coefficient vector $\beta = (1,1.2,-3.2,4.1,-5,\underbrace{0,\ldots0}_{45})$,

$$\text{Selection rate} = VC/(\text{the number of nonzero values in } \beta)$$

$$\text{Rejection rate} = VD/(\text{the number of zeros in } \beta)$$

$$\text{Mean Squared Error of Coefficient Estimates} = \frac{1}{5}\text{sqrt}[\textstyle\sum_{j=1}^{5}\left(\beta_j - \hat{\beta}_j\right)^2]$$

where VC is the mean of correctly chosen predictors;

VD is the mean value of correctly excluding irrelevant variables;

Use four regression methods to repeat the numerical simulation for m=300 times, and the results are shown in Table 3 and Table 4.

Through the comparative analysis, it can be seen that under the condition of large samples, the four methods can accurately select the correct non-zero variables, but in terms of eliminating irrelevant variables, SCAD performs the best, the elimination rate reaches 100%, and the coefficient estimation error minimum. Similar to Lasso and Adaptive-Lasso, Elastic-Net basically has no ability to compress invalid variables. In high-dimensional conditions, although SCAD performed slightly worse than the other three methods in correctly selecting variables, it far surpassed them in the ability and accuracy of eliminating invalid variables.

In the literature[21][22][23], the method or model used to calculate the true positive rate (*sensitivity*) and true negative rate (*specificity*), and to calculate the probability of predicting variables entering the model compare the mean *cc* that can correctly select the predictor variables and the mean *cd* that can correctly eliminate irrelevant predictors. In addition to the conclusion that SCAD regression fitting is the best under conventional data conditions, this paper also analyzes the stability and efficiency of the model with the help of the boxplot tool and the time required for regression fitting, and points out that the problems of poor stability and low algorithm efficiency of SCAD.

Empirical Analysis and Discussion

Statistical analysis and modeling only give suggestions on the relative optimal model from the perspective of data. Ultimately, a reasonable and efficient regression model needs to be established in combination with professional practice. This paper uses the 462 × 9 dataset Heart provided by *R* on the baseline survey of coronary risk factors in rural South Africa for empirical analysis. The following variables are included in the dataset: systolic blood pressure (sbp), tobacco consumption (tobacco), low-density lipoprotein cholesterol (ldl), adipose tissue concentration (adiposity), family history (famhist), test score for type A personality (typea), obesity(obesity), alcohol consumption (alcohol) and age (age). Because the units of measure for each variable in the dataset are different, the dataset is first standardized. The dataset is then divided into training and test sets. On the training set, different variable selection methods (*Elastic-Net*, *Adaptive-Lasso*, *Lasso*, *SCAD*) and cross-validation methods are used to find out the factors that have a greater impact on systolic blood pressure and establish a regression model. On the test set, the model is validated, and the mean squared error is calculated to compare the quality of the models fitted by different methods.

First, the OLS linear regression model was used to model the data set, and the linear regression relationship between systolic blood pressure and other variables was established. Determine whether the response variable satisfies the statistical assumptions of the linear regression model, namely, normality, independence, homoscedasticity, and linear correlation with the predictor variables. The statistical properties obtained by performing OLS regression on systolic blood pressure (sbp) in the Heart dataset are shown in Figure 5.

From the QQ plot, Residuals vs Fitted plot and Scale-Location plot, it can be seen that the response variable sbp basically obeys normality, linearity, and homoscedasticity, and independence cannot be distinguished from the figure. But there is no a priori reason to think that one person's systolic blood pressure affects another, so it is reasonable to assume that sbp also satisfies independence.

First, perform OLS linear regression on Heart on the training set to establish a linear function relationship between systolic blood pressure and various influencing factors. The estimated values of the coefficients are shown in Table 5.

The model is verified on the test set, and the mean square error of the index used to measure the accuracy of the model is calculated, Err=0.92. The model established by OLS is that all predictors are related to the response variable, but usually the response variable is mainly related to a subset of the predictor variables. Therefore, in order to achieve variable selection and improve prediction accuracy, this paper uses the following Lasso regression method to build the model and conduct comparative analysis.

Elastic-Net

Figure 6 shows the parsing path of the Elastic-Net penalty function and the cross-validation MSE graph. Using the *R* language built-in cross-validation function (*ten-fold cross-validation*), the adjustment parameter $\lambda = 0.107$ when the training mean square error is the smallest is obtained, and then fitted to obtain Elastic net regression model, the corresponding regression coefficients are shown in Table 6.

Validate the model on the test set, and calculate the mean square error of the index reflecting the accuracy of the model Err=0.926, realize variable selection, screen out the factors that have a greater impact on systolic blood pressure as tobacco consumption, adipose tissue concentration, obesity, alcohol consumption and age, and give the quantitative relationship between them, where age and adipose tissue concentration had the greatest impact.

Lasso

Figure 7 shows the parsing path of the Lasso penalty function and the cross-validation MSE graph. Using the R language built-in cross-validation function (*ten-fold cross-validation*), the adjustment parameter λ= 0.0536 when the training mean square error is the smallest is obtained, and then fitted to obtain Lasso regression model, the corresponding regression coefficients are shown in Table 7.
Validate the model on the test set, and calculate the mean square error of the index reflecting the accuracy of the model Err=0.929, realize variable selection. Compared with the Elastic-Net method, eliminating the obesity term and increasing the influence factor of adipose tissue concentration on systolic blood pressure not only simplifies the model, but also has its rationality (*calling the corr.test() function in the psych package to analyze the obesity term and adipose tissue concentration with a correlation coefficient of 0.72*). The parsing path is also similar, and compresses faster as the coefficient increases with lambda.

SCAD

Figure 8 shows the parsing path of the SCAD penalty function. The shaded part indicates that SCAD achieves a local optimum in this area (*which will bring about the defect of unstable solution*). Use the built-in cross-validation function in R language (*ten-fold cross-validation*) to obtain the adjustment parameter λ=0.149 when the training mean square error is the smallest, and obtain the SCAD fitting regression model. The corresponding regression coefficients are shown in Table 8.

Validate the model on the test set, and calculate the mean square error of the index reflecting the accuracy of the model Err=0.946, realize variable selection. Five factors were screened out that had a greater impact on systolic blood pressure, and the quantitative relationship between them was given. However, the solution given by SCAD is unstable, and the coefficient of ldl given is negative, which is unreasonable.

Adaptive Lasso

Figure 9 shows the parsing path of the Adaptive Lasso penalty function. The package msgps gives the adjustment parameters under different variable selection criteria of AIC, BIC, GCV and Cp. In this paper, *GCV (ms.tuning= 2.86*) is selected to obtain the fitting Adaptive Lasso regression model. The corresponding regression coefficients are shown in Table 9.

We validate the model on the test set, and calculate the mean square error of the index reflecting the accuracy of the model Err=0.92, realizing variable selection. It is showed in Table 10. The factors that have a greater impact on systolic blood pressure are adipose tissue concentration, age, tobacco and alcohol, and the quantitative relationship between them is given more refined.

Result analysis:
(1) The Lasso method can well realize the variable selection, and screen out the factors that have a greater impact on the systolic blood pressure. Compared with the Elastic-Net method, the obesity item is eliminated and the influencing factors of the adipose tissue concentration on the systolic blood pressure are increased. The model is simple and reasonable (*the correlation coefficient between the obesity term and the adipose tissue concentration is 0.72*). The parsing path is also similar, and compresses faster as the coefficient increases with lambda.
(2) The solution given by SCAD is unstable, and the coefficient of ldl given is negative, which is unreasonable。

(3) Like Lasso, Adaptive Lasso screened out adipose tissue concentration, age, tobacco and alcohol as factors that have a greater impact on systolic blood pressure, and the mean square error was the smallest. To compare and analyse the different characteristics and results of four variable selection methods in fitting data sets with linear random effects. In this paper, a random effect data set of the same dimension and normal distribution with the mean zero and the variance covariance matrix is $\sigma^2 I_q$ is designed to be linearly superimposed on this data set (σ^2 *is taken as 0.1, and the number of randomly generated datasets is m=30*). Based on this design, Elastic-Net, Lasso, Adaptive Lasso and SCAD are used to fit the random effect of the coefficient estimate $\hat{\beta}$ in Figure 10 and the random effect of the mean square error in Figure 11. To observe and analyse the random effect of $\hat{\beta}$, you can see Although the variable selection ability of the model established by Adaptive Lasso and SCAD is slightly better than that of Elastic-Net and Lasso. However, under the influence of random effects, the model coefficients fluctuate significantly due to random effects, while Elastic-Net and Lasso are relatively stable, and the mean square error and stability of each model are affected by random factors to a similar degree. Therefore, in the actual application process, the appropriate variable selection method can be selected to fit the model according to the needs.

Conclusion

The current COVID-19 pandemic is still serious in the world, based on the data source from public platforms and Lasso class variable selection method what we proposed in the paper, to establish the linear random effects model to analysis the relations of interaction among infection rate of COVID-19, mortality, the countries and regions population density, urban population, the proportion of people 65 and older, vaccination rates, per capital GDP of the previous year, number of hospital beds per one thousand people, the human development index, the proportion of patients with underlying diseases and the overall government response index, etc. That is useful to help the government to grasp the situation of COVID-19 transmission, timely formulate and adjust the policies of epidemic prevention and control.

References

[1] Hansheng Wang,Chenlei Leng .A note on adaptive group Lasso [J]. Computational Statistics＆Data Analysis, 2008, 52(12):5277-5286.

[2] Yuan M, Lin Y. Model selection and estimation in the Gaussian graphical model [J]. Biometrika, 2007, 94(1):19-35.

[3] Qin Wang, Xiangrong Yin .A nonlinear multi-dimensional variable selection method for high dimensional data:Sparse MAVE [J]. Computational Statistics＆Data Analysis, 2008, 52(9):4512-4520.

[4] Peng Zeng, Tianhong He& Yu Zhu. A Lasso-type approach for estimation and variable selection in single index models[J]. Journa1 of Computational and Graphical Statisdcs, 2012, 21(1):92—109.

[5] Jianqing Fan, Jinchi Lv .Sure independence screening for ultrahigh dimensional feature space[J]. Journal of the Royal Statistical Society, 2008, 70(5):849—911.

[6] Yuan M&Lin Y. Efficient Empirical Bayes Variable Selection and Estimation in Linear Models[J]. Journal of America Statistical Association. (2012) 1215-1225.

[7] Yuan Jing. Bayesian Methods for Variable Selection Problems [D]. Shandong University doctoral dissertation 2013.

[8] Hanfang Li. The Research of Random Effects Panel Data Models Based on Bavesian Methed [D]. Central China Normal University doctoral dissertation 2019.

[9] Robert Richardson,Brian Hartman. Bayesian nonparametric regression models for modeling and predicting healthcare claims[J] . Insurance:Mathematics and Economics . 2018.

[10] J.Q. Fan& R. Li, Variable Selection via Nonconcave Penalized Likelihood and its Oracle Properties[J]. Journal of American Statistical Association, 2001, 96(456):1 348-1 360.

[11] Joseph Lee Rodgers &W.Alan Nicewander, Thirteen ways to look at the correlation Coefficient[J]. The American Statistician , 1988 42(1): 59-66.

[12] Liang Jiye, Feng Chenjiao, Song Peng. A Review of Big Data Correlation Analysis [J] . Chinese Journal of Computers, 2016, 39(1): 5-6.
[13] Tibshirani R. Regression Shrinkage and Selection via the Lasso[J]. Journal of the Royal Statistical Society. Series B (Statistical Methodology), 2011, 73(3): 273-282.
[14] Breiman ,Leo .Better Subset Regression Using The Nonnegative Garrote[J]. Technometrics, 1995，37(4)：373-384.
[15] Zou, Hui.The Adaptive Lasso And Its Oracle Properties[J].American Statistical Association，2006，101(476)：1418-1429.
[16] Zou H and Hastie T.Regularization and variable selection via the elastic net[J]. Journal of the Royal Statistical Society：Series B(Statistical Methodology)，(2005). 67(2)：301—320.
[17] Fan J.Q and Li R. Variable Selection Via Nonconvex Penalized Likelihood And Its Oracle Properties[J]. Journal of American Statistical Association，2001， 96(456)：1 348-1360.
[18] Chu Yi. Variable selection and feature selection under several models [D]. Shandong University doctoral dissertation 2018.
[19] Hastie T , Tibshirani R , Wainwright M . Statistical Learning with Sparsity: The Lasso and Generalizations[M]. 2015.
[20] Sim.N.Wood, translated by Shi Liwei. The core method of statistics and its application [M].People's Posts and Telecommunications Publishing House.
[21] Wu Xizhi. Statistical Methods for Complex Data (3rd Edition) [M]. Beijing: Renmin University of China Press, 2015.
[22] Bai Xuemei, Zhao Songshan. A deeper understanding of multicollinearity [J]. Journal of Dongbei University of Finance and Economics,2005(2):8-12.
[23] Xu Liqing, Tian Kai, Xiong Qin. Application of Elastics Net Method in Longitudinal Data Model [J]. Mathematical Theory.2016,36(2):61-66.
[24] Niu Yong, Li Huapeng. A review of feature screening methods for ultra-high-dimensional data [J]. Applied Probability and Statistics ,2021, 37(1): 69-104.

Table 1(on next page)

Regression fitting schedule Time(s)

Note：Computer processor（Intel(R) Core(TM) i7-10710U CPU @ 1.10GHz 1.61GHz）

Table 1 Regression fitting schedule Time(s)

Regression method	Time/cor= 0.2	Time/cor= 0.4	Time/cor= 0.6	Time/cor= 0.8	Time/cor= 0.98
Elastic-Net	1.9883	2.3013	2.2543	2.3370	2.3300
Lasso	1.8800	2.0140	1.9773	1.9900	2.1473
SCAD	2.2453	2.6367	4.1060	4.5927	12.4660
Adaptive-Lasso	2.0137	1.8079	1.9650	1.9113	1.8607

Note :Computer processor：Intel(R) Core(TM) i7-10710U CPU @ 1.10GHz 1.61GHz）

Table 2 (on next page)

Regression fitting schedule Time(s)

Table 2 Regression fitting schedule Time(s)

Regression method	Time/cor =0.2	Time/cor =0.4	Time/cor =0.6	Time/cor =0.8	Time/cor =0.98
Elastic-Net	0.334	0.328	0.358	0.396	0.502
Lasso	0.286	0.285	0.293	0.293	0.345
SCAD	0.135	0.141	0.154	0.156	0.201
Adaptive-Lasso	0.308	0.312	0.362	0.338	0.410

Table 3 (on next page)

Selection rate/Rejection rate (cor=0.6; m=300)

1 Table 3 Selection rate/Rejection rate (cor=0.6; m=300)

	n:p=2500:50				*n:p*=40:50			
Regression method	VC	Selection rate	VD	Rejection rate	VC	Selection rate	VD	Rejection rate
Elastic-Net	5	100%	0.093	≈ 0	5	100%	0.057	≈ 0
Lasso	5	100%	24.5	53.8%	4.98	99.6%	5.99	13.3%
SCAD	5	100%	45	100%	4.66	93.1%	43.6	96.7%
Adaptive-Lasso	5	100%	26.2	58.2%	4.91	98.2%	4.78	10.6%

2

Table 4 (on next page)

Coefficient estimation accuracy (cor=0.6;m=300)

Table 4 Coefficient estimation accuracy (cor=0.6;m=300)

	n:p=2500:50	*n:p*=40:50
Regression method	Coefficient estimate error	Coefficient estimate error
Elastic-Net	0.0135	0.474
Lasso	0.0162	0.210
SCAD	0.0121	0.147
Adaptive-Lasso	0.0134	0.288

Table 5 (on next page)

Heart : OLS Regression Model Coefficients

Table 5 Heart *:* OLS Regression Model Coefficients

(*Intercept*)	*tobacco*	*ldl*	*adiposity*	*famhist*	*typea*	*obesity*	*alcohol*	*age*
-0.0313	-0.0251	0.0317	0.1329	-0.0953	-0.052	0.0908	0.0846	0.3016

Table 6(on next page)

Heart*:* Elastic-Net-Coefficients

Table 6 Heart: Elastic-Net-Coefficients

(*Intercept*)	*tobacco*	*ldl*	*adiposity*	*famhist*	*typea*	*obesity*	*alcohol*	*age*
0.0333	0.0560		0.1815			0.0127	0.0335	0.1372

Table 7 (on next page)

(on next page)

Heart: Lasso- Coefficients

Table 7 Heart: Lasso- Coefficients

(*Intercept*)	*tobacco*	*ldl*	*adiposity*	*famhist*	*typea*	*obesity*	*alcohol*	*age*
0.0332	0.0561		0.2015				0.0347	0.1359

Table 8 (on next page)

Heart: SCAD- Coefficients

Table 8 Heart: SCAD- Coefficients

(*Intercept*)	*tobacco*	*ldl*	*adiposity*	*famhist*	*typea*	*obesity*	*alcohol*	*age*
0.0225		-.0053		-0.0351		0.2288	0.0081	0.3915

Table 9 (on next page)

Heart: Adaptive Lasso-Coefficients

Table 9 Heart: Adaptive Lasso-Coefficients

(*Intercept*)	*tobacco*	*ldl*	*adiposity*	*famhist*	*typea*	*obesity*	*alcohol*	*age*
0.0305	0.0506		0.2132				0.0450	0.1754

Table 10(on next page)

Comparison of Mean Square Errors by Models

Table 10 Comparison of Mean Square Errors by Models

Method	Err
Elastic-Net	0.926
Lasso	0.929
SCAD	0.946
Adaptive-Lasso	0.920

Figure 1

Angle、Err Boxplots

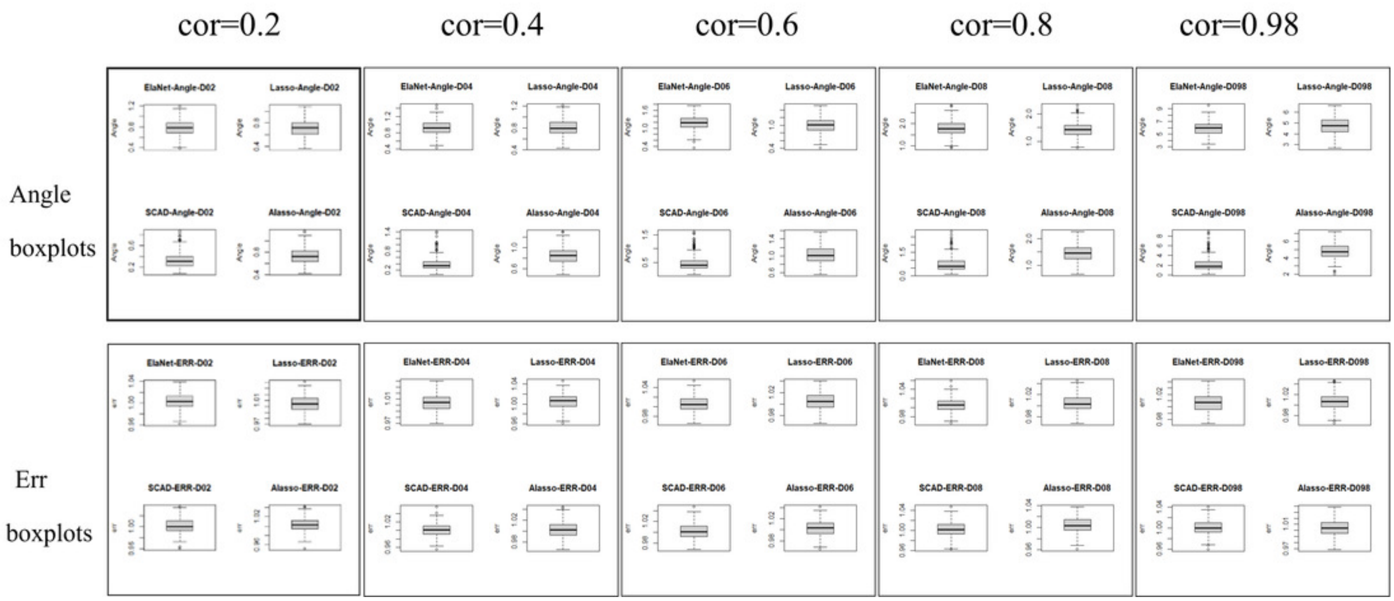

Figure 2

Relationship Graphics

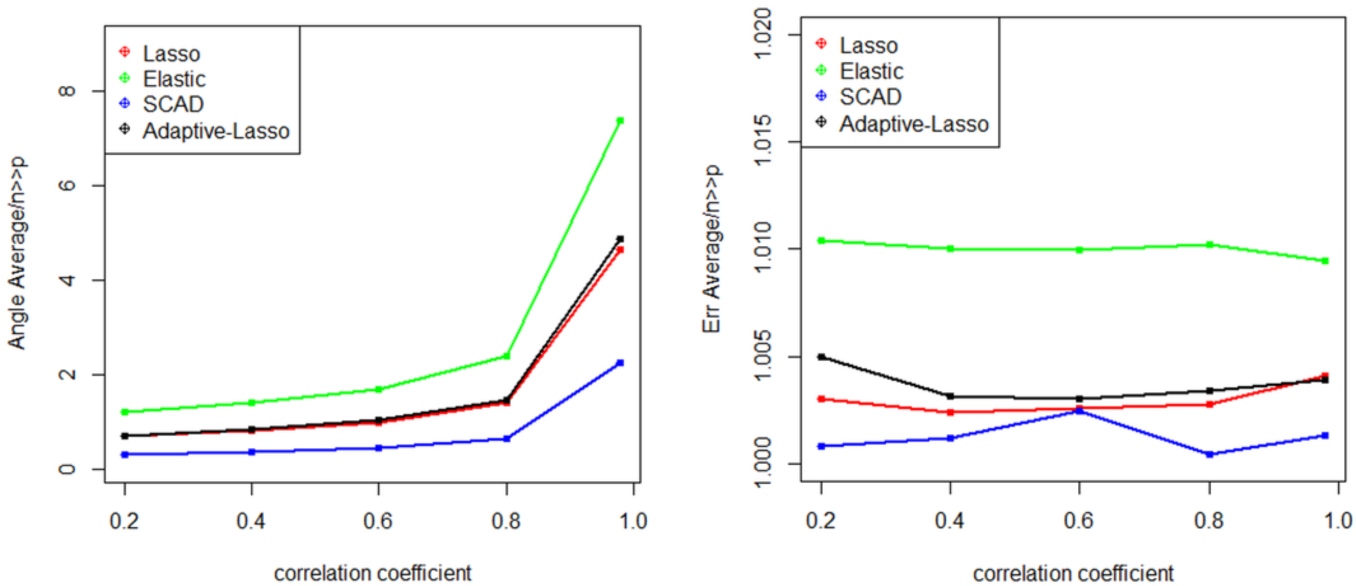

Figure 3

Angle□Err Boxplots

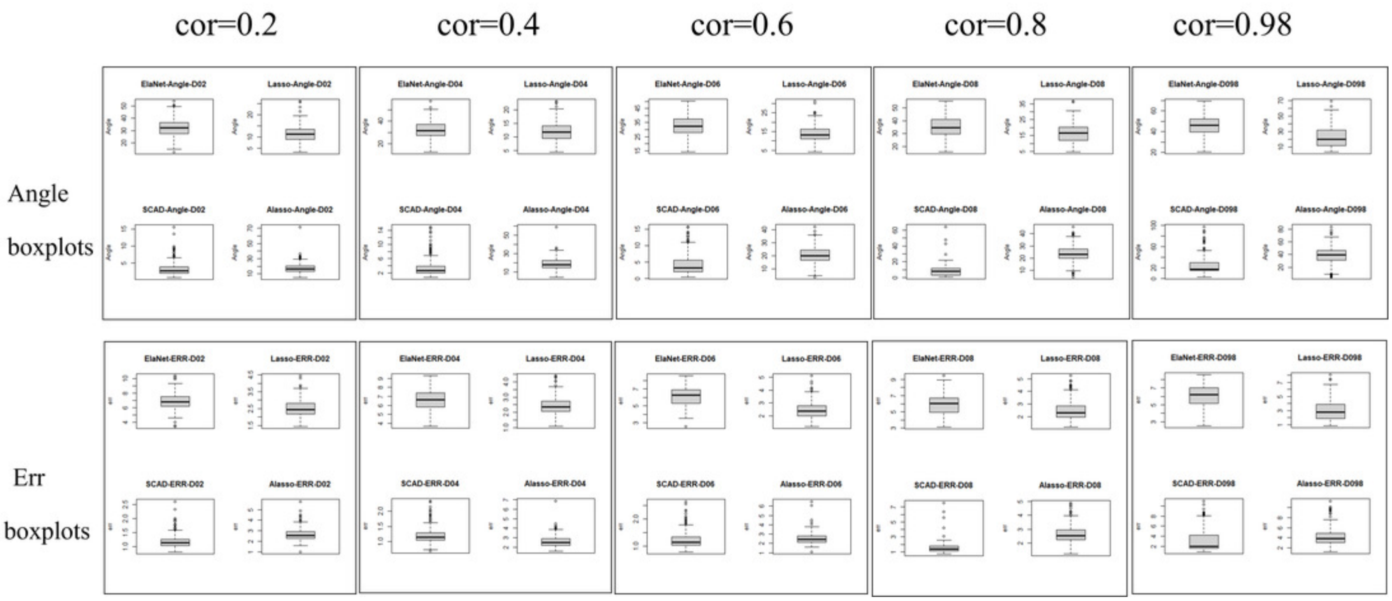

Figure 4

Relationship Graphics

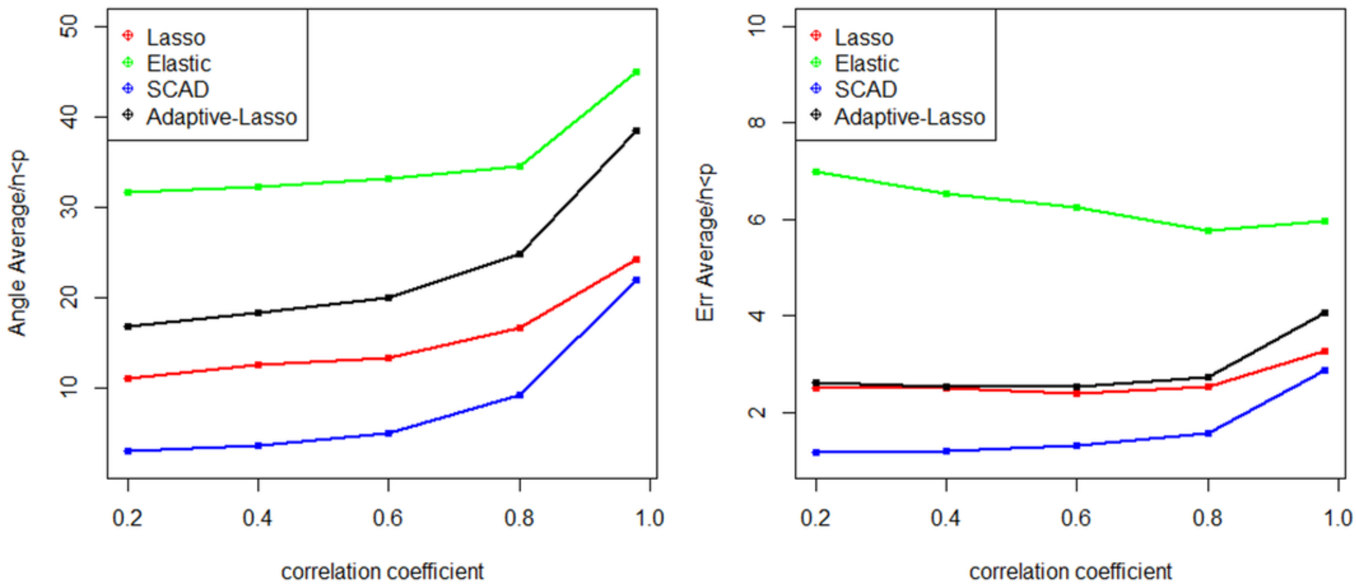

Figure 5

sbp regression diagnostic graph

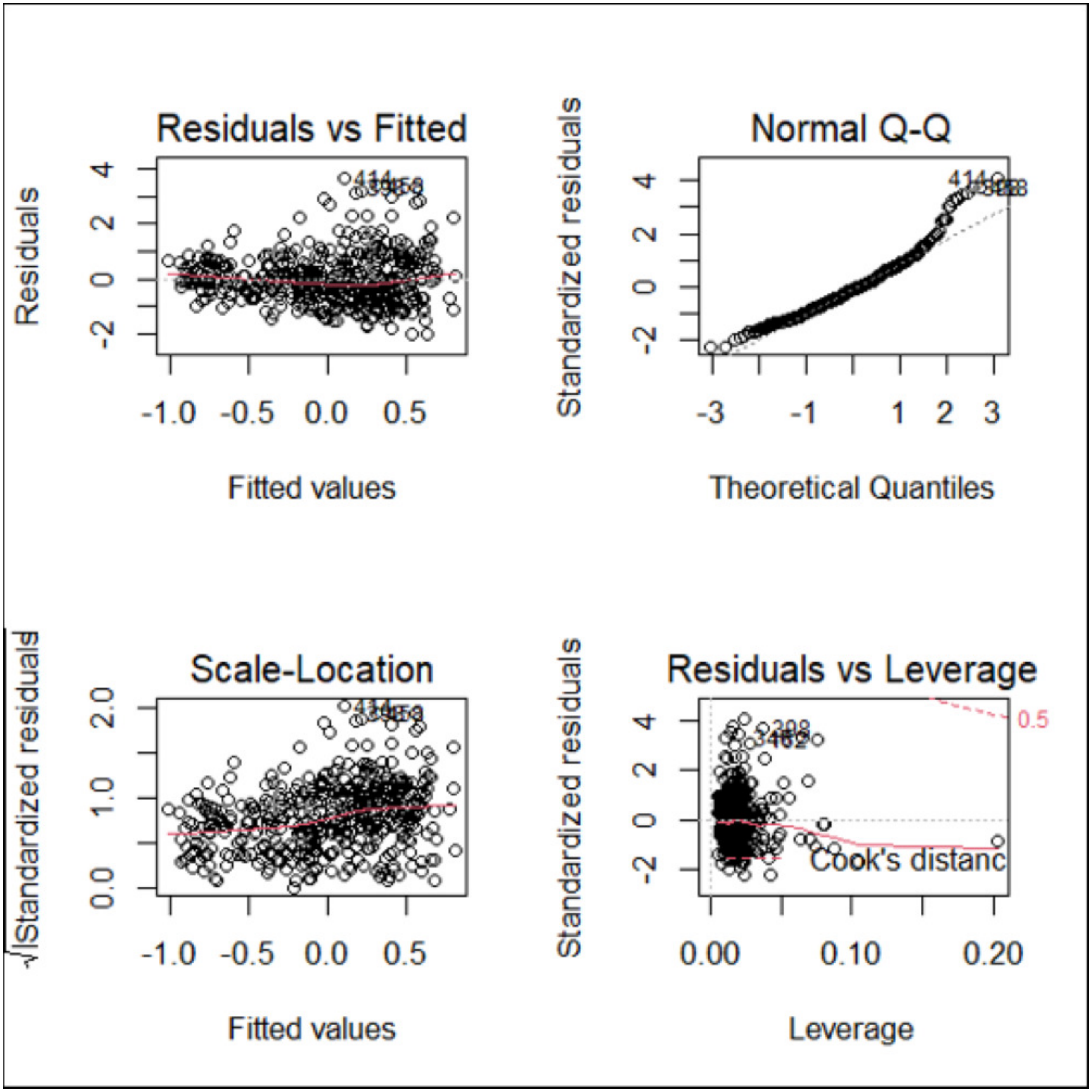

Figure 6

Elastic-Net parsing path and cross-validation MSE graph

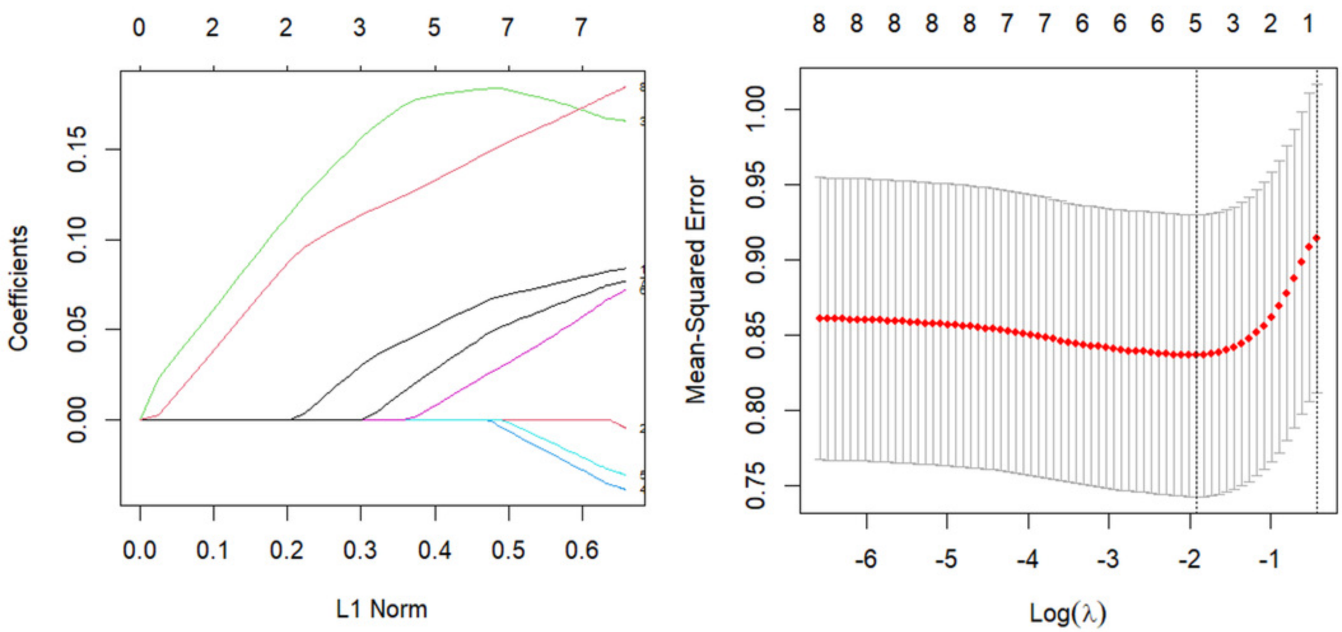

Figure 7

Lasso parsing path and cross-validation MSE graph

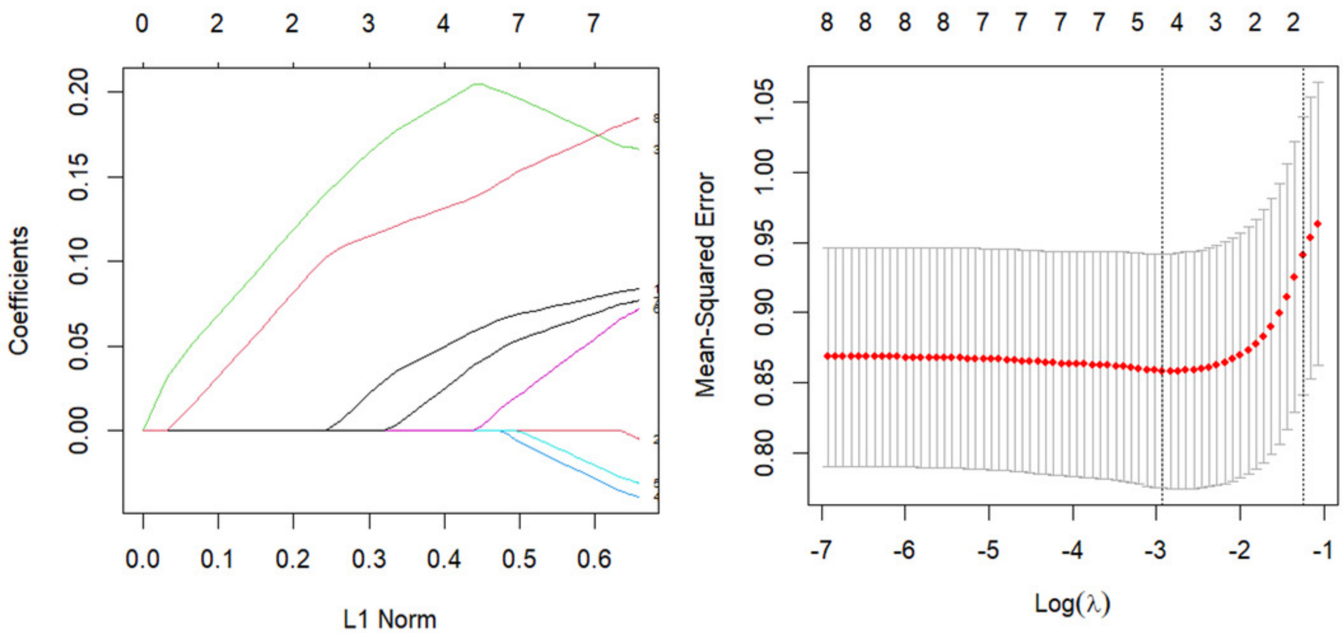

Figure 8

SCAD parsing path and cross-validation MSE graph

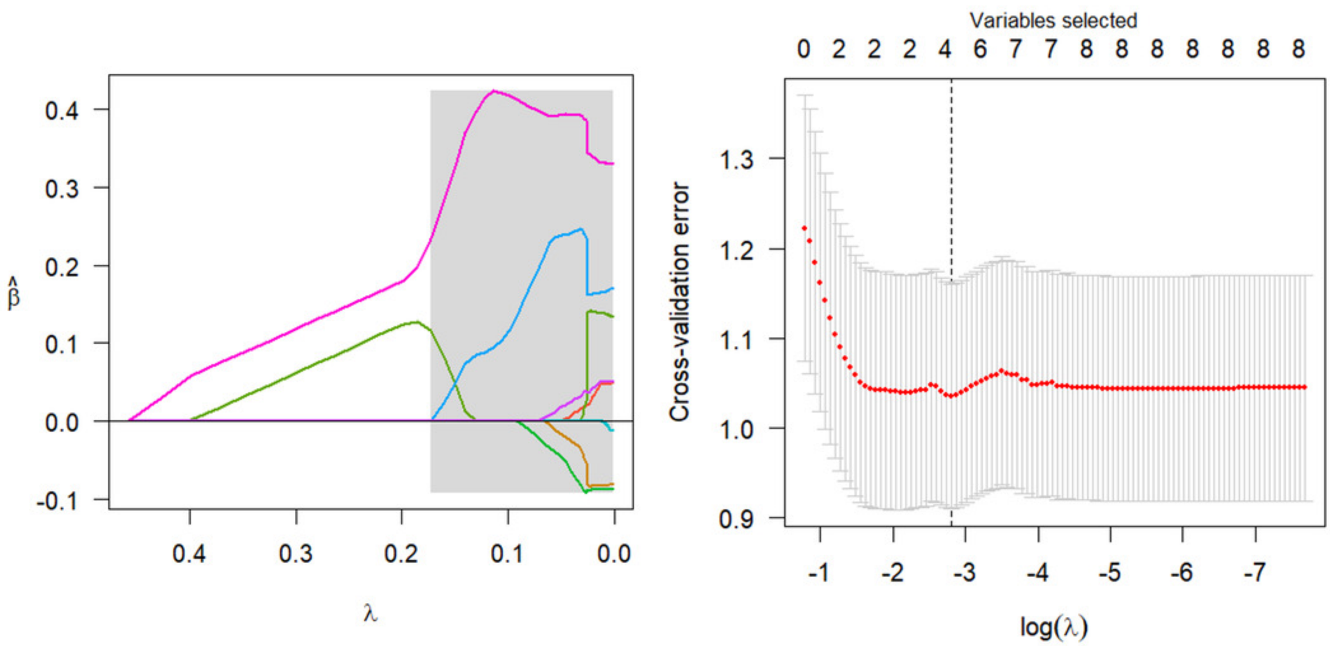

Figure 9

Adaptive Lasso parsing path

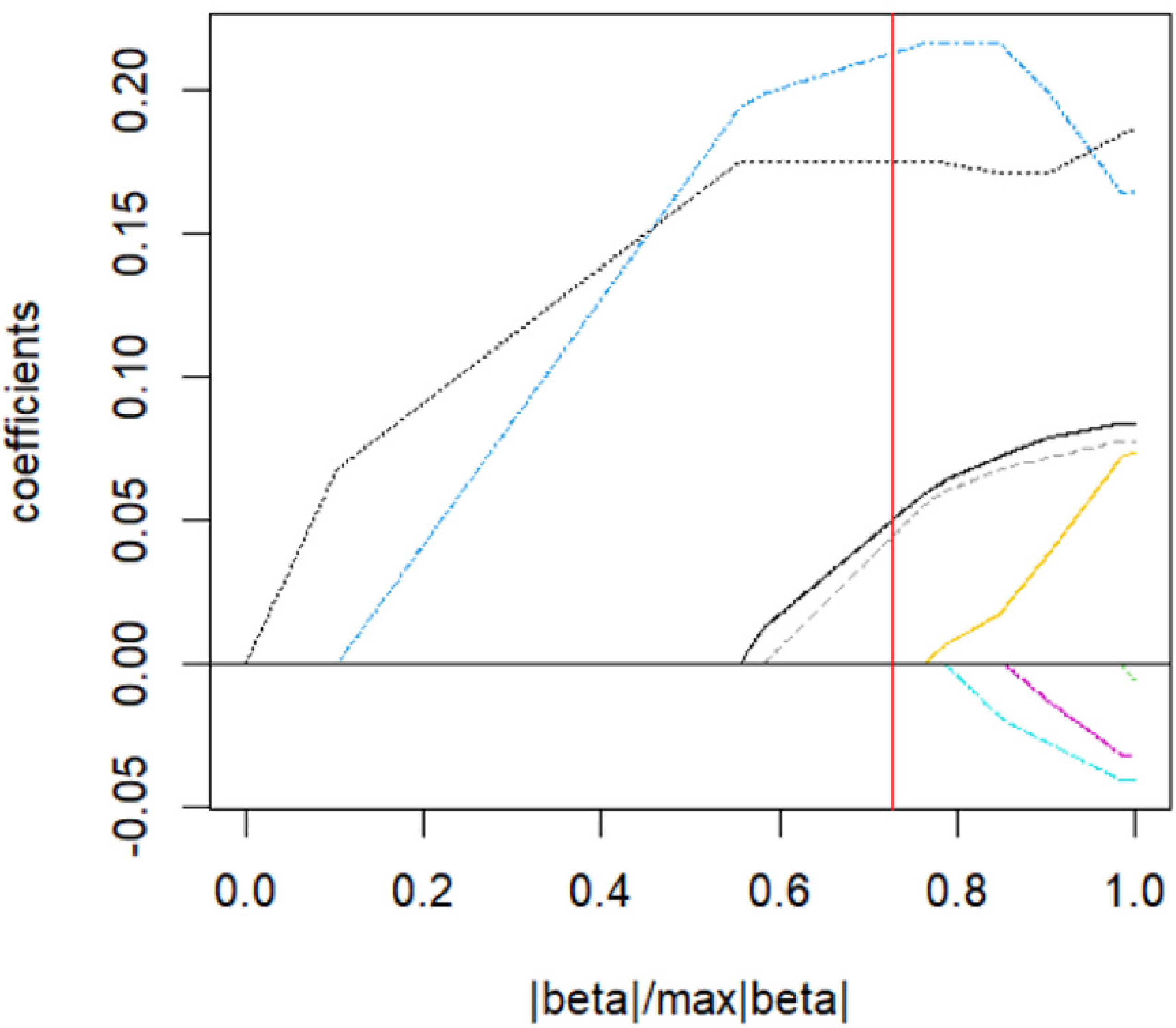

Figure 10

β̂ Random Effect Diagram

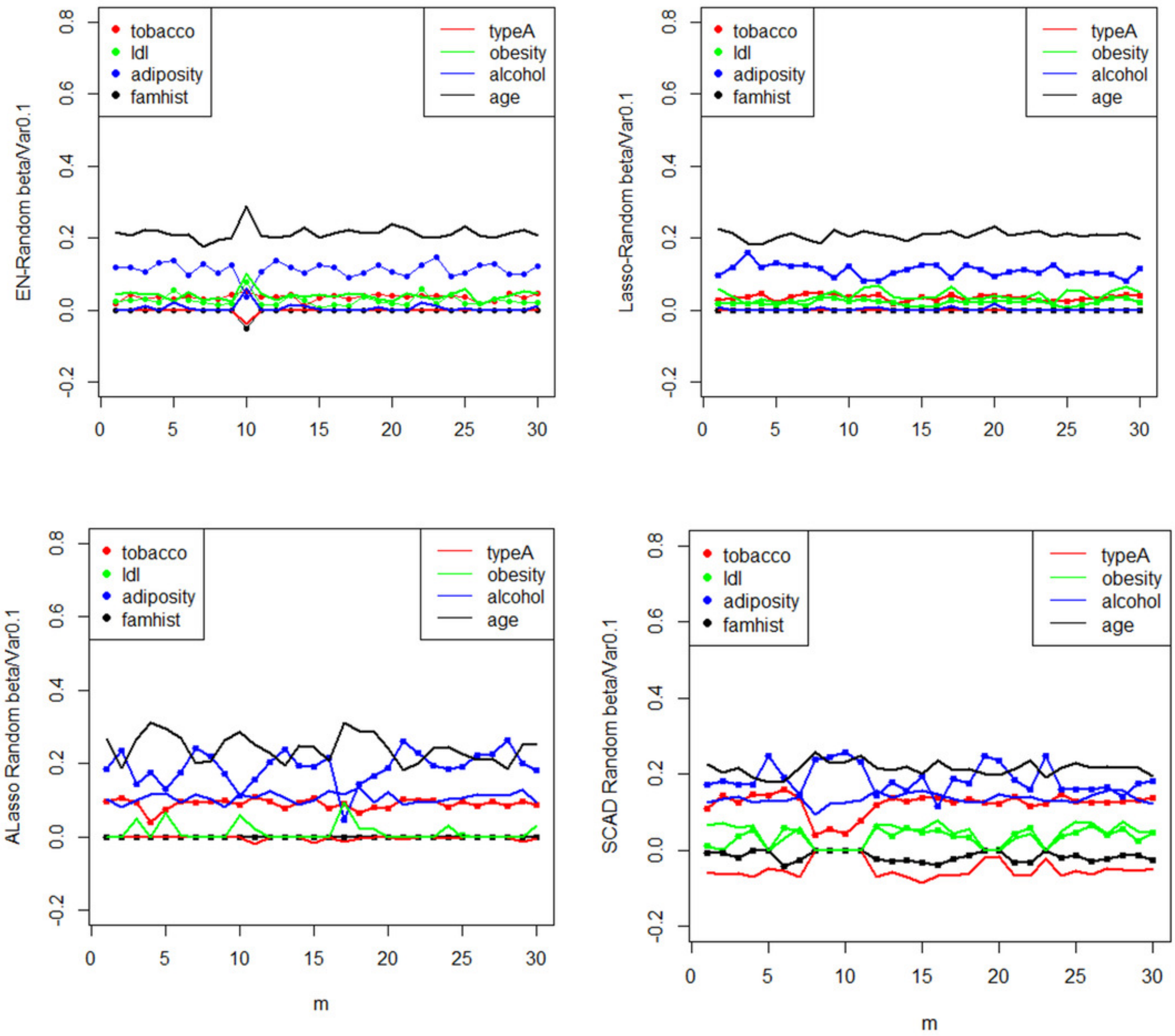

Figure 11

MSE Random Effect Diagram