Thailand Statistician
July 2015; 13(2): 243-269
http://statassoc.or.th
Contributed paper

Elastic Net Regression with the Value of the L_2 Penalty Parameter Associated with Bayesian Analysis

Kanyalin Jiratchayut* and Chinnaphong Bumrungsup
Department of Mathematics and Statistics, Faculty of Science and Technology, Thammasat University, Pathum Thani 12120, Thailand.
*Corresponding author; e-mail: kanyalinn@gmail.com

Received: 6 February 2015
Accepted: 7 May 2015

Abstract

The aim of this article is to propose the method for choosing the value of the L_2 penalty parameter, λ_2, of elastic net linear regression model using Bayesian analysis. The value of λ_2 is specified through the behavior of Bayes factor. We study the performance of elastic net estimators where the value of λ_2 is based on Bayes factor and the value of λ_2 is chosen by 10-fold cross-validation method. Simulation studies and real data examples show that the elastic net estimator where the value of λ_2 is based on Bayes factor performs better in prediction accuracy.

____

Keywords: Bayes factor, Bayesian analysis, elastic net, L_2 penalty, shrinkage

1. Introduction

The elastic net proposed by Zou and Hastie [1] is a penalized regression method for variable selection and coefficient estimation. The elastic net simultaneously performs automatic variable selection and continuous shrinkage, it can select groups of correlated variables and overcomes the difficulty when the number of predictor variables (p) is greater than the number of observation (n). The elastic net is based on a combination of the ridge [2, 3] and the lasso [4] penalties. Consider a linear regression model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \,, \tag{1}$$

where $\mathbf{y}$ is an $n \times 1$ vector of response variable, $\mathbf{X}$ is an $n \times p$ matrix of predictor variables, $\boldsymbol{\beta}$ is an $p \times 1$ vector of parameter of regression coefficients, $\boldsymbol{\varepsilon}$ is an $n \times 1$ vector of random errors, p is the number of predictors, and n is the number of observations. The errors are assumed to be independent identically normally distributed random variable with mean 0 and finite variance σ^2. Without loss of generality, we assume the response is centered and the predictors are standardized, so the intercept is not included in the regression function. The elastic net is defined in two stages. Assuming that the response is centered and the predictors are standardized, naïve elastic net estimator is first found via

$$\widehat{\boldsymbol{\beta}}_{\text{Naïve elastic net}} = \arg\min_{\beta}[\|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2 + \lambda\, \mathcal{J}(\boldsymbol{\beta})] \tag{2}$$

with the elastic net penalty $\mathcal{J}(\boldsymbol{\beta}) = \alpha\|\boldsymbol{\beta}\|_2^2 + (1-\alpha)\|\boldsymbol{\beta}\|_1$, $\lambda = \lambda_1 + \lambda_2$,and let $\alpha = \lambda_2/(\lambda_1 + \lambda_2)$. The elastic net procedure can be viewed as a penalized least squares method. The elastic net penalty $\mathcal{J}(\boldsymbol{\beta}) = \alpha\|\boldsymbol{\beta}\|_2^2 + (1-\alpha)\|\boldsymbol{\beta}\|_1$ is a convex combination of the lasso and ridge penalties. When $\alpha = 1$, the naïve elastic net becomes simple ridge regression. If $\alpha = 0$ then $\lambda_2 = 0$, the naïve elastic net becomes the lasso. The elastic net has the Bayesian connection. The elastic net penalty corresponds to a new prior given by

$$p_{\lambda,\alpha}(\boldsymbol{\beta}) = c(\lambda,\alpha)exp\{-\lambda[\alpha\|\boldsymbol{\beta}\|_2^2 + (1-\alpha)\|\boldsymbol{\beta}\|_1]\}, \tag{3}$$

which is a compromise between the Gaussian and Laplacian priors. Zou and Hastie [1] pointed out that the elastic net estimator can be viewed as the Bayes posterior mode of $\boldsymbol{\beta}$ under the prior in (3). The final elastic net estimator is taken to be a rescaled version of the naïve estimator,

$$\widehat{\boldsymbol{\beta}}_{\text{elastic net}} = (1+\lambda_2)\widehat{\boldsymbol{\beta}}_{\text{Naïve elastic net}} \,. \tag{4}$$

The scaling was introduced to reduce perceived overshrinkage of the naïve estimator [1]. Hence, the *elastic net estimator* is defined as follows:

$$\widehat{\boldsymbol{\beta}}_{\text{elastic net}} = (1+\lambda_2)\big\{\arg\min_{\beta}[\|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2 + \lambda_2\|\boldsymbol{\beta}\|_2^2 + \lambda_1\|\boldsymbol{\beta}\|_1]\big\}, \tag{5}$$

where $\lambda_1 \geq 0$ and $\lambda_2 \geq 0$ are the penalty parameters, $\|\boldsymbol{\beta}\|_1 = \sum_{j=1}^{p} |\beta_j|$ is the L_1 norm of $\boldsymbol{\beta}$, and $\|\boldsymbol{\beta}\|_2^2 = \sum_{j=1}^{p} \beta_j^2$ is the L_2 norm of $\boldsymbol{\beta}$. The L_1 part of the elastic net performs automatic variable selection, while the L_2 part stabilizes the solution parts and, hence, improves the prediction.

The penalty parameters (λ_1 and λ_2) control the amount of shrinkage imposed on the coefficients, where some weak effects are forced to be exactly zero if the shrinkage level is large enough. Hence, the penalty parameters (λ_1 and λ_2) can be named the shrinkage parameters. If the value of λ_1 and λ_2 are too small, then no shrinkage will be performed. If the value of λ_1 and λ_2 are too high, then all coefficients will be shrunk to zero. Hence, the penalty parameters (λ_1 and λ_2) are important for the elastic net estimator. To solve the elastic net estimator, the researchers usually search for the optimal values of λ_1 and λ_2. K-fold cross-validation method is commonly used for choosing the values of λ_1 and λ_2 [1]. To avoid intensive computation, a grid of values for λ_2 is first specified. Zou and Hastie [1] suggested to pick a relatively small grid value of λ_2, They used (0, 0.01, 0.1, 1, 10, 100, 1000). For each λ_2, a 10-fold cross-validation is then used to choose λ_1. The chosen λ_2 is the one giving the smallest cross-validation error.

The elastic net has Bayesian interpretation as shown in (3), so many researchers adopted the elastic net in the Bayesian frameworks over the past years such as Bornn, Gottardo and Doucet [5], Kyung, Gill, Ghosh and Casella [6], Li and Lin [7], and Hans [8]. They did not propose the method for choosing the penalty parameters directly. Kyung, Gill, Ghosh and Casella [6] and Li and Lin [7] chose the penalty parameters λ_1 and λ_2 via Gibbs sampling approach. Bornn, Gottardo, and Doucet [5] and Hans [8] used cross-validation method to choose the values of penalty parameters λ_1 and λ_2.

For elastic net method, there are two penalty parameters (λ_1 and λ_2). In this article, we concentrate on λ_2. The penalty parameter λ_2 is the penalty parameter of the L_2 part which stabilizes the solution part of the elastic net estimator. The value of λ_2 is suggested to specify in the first step of the elastic net procedure [1]. Hence, the value of penalty parameter λ_2 plays an important role for elastic net method.

The objective of this article is to propose the method for choosing the value of the L_2 penalty parameter, λ_2, of elastic net linear regression model using Bayesian analysis. The value of λ_2 is chosen by using its effect on the posterior model probabilities and the behavior of Bayes factor [9]. Hence, the value of λ_2 is specified through Bayes factor. The Bayes factor proposed by Kass and Raftery [9] is a quantity for comparing models and for testing hypotheses in the Bayesian framework. Bayes factor has played a major role in assessing the goodness of fit of the competing models [10]. The advantages of the Bayes factor can see in Kass and Raftery [9], Raftery [11], and Ntzoufras [12]. The method proposed in this article is developed from Lykou and Ntzoufras [13] which proposed Bayesian lasso variable selection and the specification of the value of the shrinkage parameter λ_1 of the lasso [4] through Bayes factor. The prior distributions for $\boldsymbol{\beta}$ and σ^2 used in this article differ from the prior distribution used in [13]. In this research, we limit our attention to full rank model. This article is organized as follows. Section 2 describes the process for choosing the value of the penalty parameter λ_2 based on Bayes factor, λ_2BF. Section 3 presents some simulation data to study the performance of naïve elastic net estimators where the value of λ_2 is chosen by λ_2BF method and the value of λ_2 is chosen by 10-fold cross-validation method. Section 4 illustrates the performance of the λ_2BF on some real data examples. Conclusion and discussion are provided in Section 5.

2. Process for choosing the value of the penalty parameter λ_2 based on Bayes factor (λ_2BF)

In this process, the response and the predictor variables are transformed by the correlation transformation. The method of λ_2 BF is the following.

Step 1: Inclusion parameter

We begin by indexing each candidate model with one binary vector, $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_j, \dots, \gamma_p)^T$. An element γ_j takes value 0 or 1 depending on whether or not the *j*th predictor is excluded from the model. Hence, there are 2^p possible models $M_1, \dots, M_{2^p}$ where $M_{\boldsymbol{\gamma}}$ corresponds to the $\boldsymbol{\gamma}$th subset of $\mathbf{X}_1, \dots, \mathbf{X}_p$. Each submodel is of the form

$$M_{\boldsymbol{\gamma}}: \mathbf{y} = \mathbf{X}_{\boldsymbol{\gamma}}\boldsymbol{\beta}_{\boldsymbol{\gamma}} + \boldsymbol{\varepsilon} \, , \tag{6}$$

where $\mathbf{X}_{\boldsymbol{\gamma}}$ is an $n \times q_{\boldsymbol{\gamma}}$ design matrix whose columns correspond to the $\boldsymbol{\gamma}$th subset of $\mathbf{X}_1, \dots, \mathbf{X}_p$, $\boldsymbol{\beta}_{\boldsymbol{\gamma}}$ is a $q_{\boldsymbol{\gamma}} \times 1$ vector of regression coefficients for the $\boldsymbol{\gamma}$th subset, $q_{\boldsymbol{\gamma}} \equiv \boldsymbol{\gamma}^T\mathbf{1}$ denotes the size of the $\boldsymbol{\gamma}$th subset, and $\boldsymbol{\varepsilon} \sim N_n(\mathbf{0}, \sigma^2\mathbf{I})$. The linear regression model for each submodel of form (6) is

$$\mathbf{y}|\boldsymbol{\beta}_{\boldsymbol{\gamma}}, \sigma^2, \boldsymbol{\gamma} \sim N_n(\mathbf{X}_{\boldsymbol{\gamma}}\boldsymbol{\beta}_{\boldsymbol{\gamma}}, \sigma^2\mathbf{I}). \tag{7}$$

Step 2: Prior distribution for $\boldsymbol{\beta}$, σ^2, and $\boldsymbol{\gamma}$

In the process for choosing the value of λ_2, we use the hierarchical prior models in the form of submodel (7) as follows:

Prior distribution for $\boldsymbol{\beta}$

For elastic net estimator, the penalty parameter λ_2 is the L_2 part which is ridge penalty when $\lambda_1 = 0$. Suppose $\mathbf{y} \sim N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$ and each parameter β_j is distributed as $N(0, \tau^2)$, independently of one another, with τ^2 and σ^2 assumed known. Then the ridge estimator is the mode of the posterior distribution of $\boldsymbol{\beta}$ with $\lambda = \sigma^2/\tau^2$ [14]. Hence, we assume

$$\beta_j|\sigma^2 \sim N\left(0, \frac{\sigma^2}{\lambda_2}\right) \text{ for } j = 1, \ldots, p.$$

The prior distribution of $\boldsymbol{\beta}$ in matrix form is

$$\boldsymbol{\beta}|\sigma^2 \sim N_p(\mathbf{0}, \frac{\sigma^2}{\lambda_2}\mathbf{I}). \quad (8)$$

Prior distribution for σ^2

Assume σ^2 has inverse gamma prior distribution [8, 15-17]. The inverse gamma prior for σ^2 would maintain conjugacy which gives the posterior distribution in the closed form.

$$\sigma^2 \sim \text{inverse gamma}\left(\frac{\nu}{2}, \frac{\nu\xi}{2}\right), \quad (9)$$

(which is equivalent to $\nu\xi/\sigma^2 \sim \chi^2_\nu$). In this research, we choose $\nu = 3$ as suggested by Chipman, George and McCulloch [15] and choose $\xi = S^2_{FULL}$ (the traditional unbiased estimator of σ^2 based on saturated model) [16].

Prior distribution for $\boldsymbol{\gamma}$

For the specification of the model space prior, most Bayesian variable selection implementations have used independence priors of the form

$$p(\boldsymbol{\gamma}) = \prod_{j=1}^{p} w_j^{\gamma_j} (1 - w_j)^{1-\gamma_j}. \quad (10)$$

Under this prior, each $\mathbf{X}_j$ enters the model independently of the other coefficients, with probability $p(\gamma_j = 1) = 1 - p(\gamma_j = 0) = w_j$. In this research, we set $w_j = 1/2$ which yields the uniform prior.

Hence, the hierarchical prior models are summarized as follows:

$$\mathbf{y}|\boldsymbol{\beta}_{\boldsymbol{\gamma}},\sigma^2,\boldsymbol{\gamma} \sim N_n(\mathbf{X}_{\boldsymbol{\gamma}}\boldsymbol{\beta}_{\boldsymbol{\gamma}},\sigma^2\mathbf{I}),$$
$$\boldsymbol{\beta}_{\boldsymbol{\gamma}}|\sigma^2,\boldsymbol{\gamma} \sim N_{q_{\boldsymbol{\gamma}}}(\mathbf{0},\frac{\sigma^2}{\lambda_2}\mathbf{I}), \tag{11}$$
$$\sigma^2|\boldsymbol{\gamma} = \sigma^2 \sim \text{inverse gamma}\left(\frac{\nu}{2},\frac{\nu\xi}{2}\right),$$
$$p(\boldsymbol{\gamma}) = (1/2)^p.$$

Step 3: Posterior model probability of $\boldsymbol{\gamma}$ given $\mathbf{y}$

Using the hierarchical prior models described in Step 2, the posterior model probability of $\boldsymbol{\gamma}$ given $\mathbf{y}$, $f(\boldsymbol{\gamma}|\mathbf{y})$, is

$$f(\boldsymbol{\gamma}|\mathbf{y}) = f(\mathbf{y}|\boldsymbol{\gamma})p(\boldsymbol{\gamma}). \tag{12}$$

Step 4: Bayes factor

Bayes factor proposed by Kass and Raftery [9] is the posterior odds of one hypothesis when the prior probabilities of the two hypotheses are equal. For the Bayes factor used in this research, we define the hypotheses associated with the Bayes factor as

H_0: Reduced model (M_R) versus H_1: Full model (M_F),

where M_R (Reduced model) is the linear regression model with the predictors $\mathbf{X}_{\boldsymbol{\gamma}}$, and M_F (Full model) is the linear regression model with the predictors $\mathbf{X}_{\boldsymbol{\gamma}}$ of the reduced model and additional predictor $\mathbf{X}_j$.

Suppose $\mathbf{X}_{\boldsymbol{\gamma}}$ be the predictor variables of the reduced model, so

M_R: $\mathbf{y} = \mathbf{X}_{\boldsymbol{\gamma}}\boldsymbol{\beta}_{\boldsymbol{\gamma}} + \boldsymbol{\varepsilon}$, and M_F: $\mathbf{y} = \mathbf{X}_{\boldsymbol{\gamma}}\boldsymbol{\beta}_{\boldsymbol{\gamma}} + \mathbf{X}_j\boldsymbol{\beta}_j + \boldsymbol{\varepsilon}$.

The posterior odds of model M_F versus model M_R is given by

$$PO_{10} = \frac{f(M_F|\mathbf{y})}{f(M_R|\mathbf{y})} = \frac{f(\mathbf{y}|M_F)}{f(\mathbf{y}|M_R)} \times \frac{f(M_F)}{f(M_R)} \quad . \tag{13}$$

The Bayes factor for comparing the evidence of model M_F versus model M_R is

$$\mathrm{BF}_{10} = \frac{f(\mathbf{y}|M_F)}{f(\mathbf{y}|M_R)} \quad . \tag{14}$$

In this research, we use the uniform prior $p(\boldsymbol{\gamma}) = \left(\frac{1}{2}\right)^{p}$. Thus, the prior model probabilities $f(M_F)$ and $f(M_R)$ are equal for all competing models. Hence,

$$\mathrm{BF}_{10} = \frac{f(M_F|\mathbf{y})}{f(M_R|\mathbf{y})} \ . \tag{15}$$

If the BF_{10} is greater than one, we choose the hypothesis H_1. Otherwise, we choose H_0. In this research, we use Bayes factor interpretation [9] as follows:

Table 1. Bayes factor interpretation.

BF_{10}	Evidence against H_0
$1 < BF_{10} < 3$	Negligible
$3 < \mathrm{BF}_{10} < 20$	Positive
$20 < \mathrm{BF}_{10} < 150$	Strong
$\mathrm{BF}_{10} > 150$	Very strong

Step 5: Specification of the value of the penalty parameter λ_2 based on Bayes factor
Step 5.1: For all pairs of hypotheses, we compute the Bayes factor for multiple linear regression model

$$\mathrm{BF}_{10(Multiple\ model)} = \frac{f(\mathbf{y}|M_F)}{f(\mathbf{y}|M_R)} \tag{16}$$

where the posterior probabilities $f(\mathbf{y}|M_R)$ and $f(\mathbf{y}|M_F)$ are

$$f(\mathbf{y}|M_R) = (\lambda_2)^{\frac{q_{M_R}}{2}} \left|\mathbf{X}_{M_R}^T\mathbf{X}_{M_R} + \lambda_2\mathbf{I}_{q_{M_R}}\right|^{-\frac{1}{2}} \left(\nu\xi + S_{M_R}^2\right)^{-\frac{(n+\nu)}{2}}, \tag{17}$$

where

$$S_{M_R}^2 = \mathbf{y}^T\mathbf{y} - \mathbf{y}^T\mathbf{X}_{M_R}\left[\mathbf{X}_{M_R}^T\mathbf{X}_{M_R} + \lambda_2\mathbf{I}_{q_{M_R}}\right]^{-1}\mathbf{X}_{M_R}^T\mathbf{y}. \tag{18}$$

$$f(\mathbf{y}|M_F) = (\lambda_2)^{\frac{q_{M_F}}{2}} \left|\mathbf{X}_{M_F}^T\mathbf{X}_{M_F} + \lambda_2\mathbf{I}_{q_{M_F}}\right|^{-\frac{1}{2}} \left(\nu\xi + S_{M_F}^2\right)^{-\frac{(n+\nu)}{2}}, \tag{19}$$

where

$$S_{M_F}^2 = \mathbf{y}^T\mathbf{y} - \mathbf{y}^T\mathbf{X}_{M_F}\left[\mathbf{X}_{M_F}^T\mathbf{X}_{M_F} + \lambda_2\mathbf{I}_{q_{M_F}}\right]^{-1}\mathbf{X}_{M_F}^T\mathbf{y}. \quad (20)$$

To find the posterior probabilities $f(\mathbf{y}|M_F)$ and $f(\mathbf{y}|M_R)$, the shrinkage parameter λ_2 is replaced by θ_{M_F} and θ_{M_R}, respectively. Thus, the posterior probabilities $f(\mathbf{y}|M_R)$ and $f(\mathbf{y}|M_F)$ are

$$f(\mathbf{y}|M_R) = \left(\theta_{M_R}\right)^{\frac{q_{M_R}}{2}}\left|\mathbf{X}_{M_R}^T\mathbf{X}_{M_R} + \theta_{M_R}\mathbf{I}_{q_{M_R}}\right|^{-\frac{1}{2}}\left(\nu\xi + S_{M_R}^2\right)^{-\frac{(n+\nu)}{2}}, \quad (21)$$

where

$$S_{M_R}^2 = \mathbf{y}^T\mathbf{y} - \mathbf{y}^T\mathbf{X}_{M_R}\left[\mathbf{X}_{M_R}^T\mathbf{X}_{M_R} + \theta_{M_R}\mathbf{I}_{q_{M_R}}\right]^{-1}\mathbf{X}_{M_R}^T\mathbf{y}. \quad (22)$$

$$f(\mathbf{y}|M_F) = \left(\theta_{M_F}\right)^{\frac{q_{M_F}}{2}}\left|\mathbf{X}_{M_F}^T\mathbf{X}_{M_F} + \theta_{M_F}\mathbf{I}_{q_{M_F}}\right|^{-\frac{1}{2}}\left(\nu\xi + S_{M_F}^2\right)^{-\frac{(n+\nu)}{2}}, \quad (23)$$

where

$$S_{M_F}^2 = \mathbf{y}^T\mathbf{y} - \mathbf{y}^T\mathbf{X}_{M_F}\left[\mathbf{X}_{M_F}^T\mathbf{X}_{M_F} + \theta_{M_F}\mathbf{I}_{q_{M_F}}\right]^{-1}\mathbf{X}_{M_F}^T\mathbf{y}. \quad (24)$$

The shrinkage parameter $\lambda_2 \geq 0$. Hence, $\theta_{M_F} \geq 0$ and $\theta_{M_R} \geq 0$. In this research, we choose the value of θ_M by the method proposed by Hoerl, Kennard, and Baldwin [18] (cited by [19]). The value of θ_M is

$$\theta_M = \frac{q_M S_M^2}{\left[\widehat{\boldsymbol{\beta}}_{LS(\mathrm{M})}\right]^T\left[\widehat{\boldsymbol{\beta}}_{LS(\mathrm{M})}\right]} , \quad (25)$$

where q_M is the number of parameters in the model M (not counting the intercept term), S_M^2 is the residual mean square in the analysis of variance table obtained from the standard least squared fit of the model M, and $\widehat{\boldsymbol{\beta}}_{LS(M)}$ is the least squared estimator of the parameter in model M.

Step 5.2: Specification of the value of the penalty parameter λ_2 based on Bayes factor, λ_2BF, for elastic net regression model.

Step 5.2.1: If the Bayes factor $\mathrm{BF}_{10(Multiple\ model)} > 1$, the model M_F is a significance model as defined by Bayes factor and θ_{M_F} associated with this model is considered to be the choice of the value of λ_2. There are many models M_F in the set $\mathrm{BF}_{10(Multiple\ model)} > 1$, so there are many θ_{M_F} associated with these models. The value θ_{M_F} of the model M_F which has the highest posterior model probability $f(\mathbf{y}|M_F)$ (the posterior model probability using the prior in (11)) is selected to be the value of λ_2 associated with Bayes factor and this λ_2 is called λ_2BF.

Step 5.2.2: For checking the validity of λ_2BF to fit the elastic net model, the Bayes factor for elastic net linear regression model, $\mathrm{BF}_{elastic\ net}$, is computed. (The derivation of $\mathrm{BF}_{elastic\ net}$ is in Appendix.) The appropriate value of λ_2BF should give the Bayes factor $\mathrm{BF}_{elastic\ net} > 1$.

3. Simulation Study

In this section, we present simulations to study the performance of the naïve elastic net estimator. We considered two methods for choosing the value of the penalty parameter λ_2: the value of λ_2 is based on Bayes factor (λ_2BF), and the value of λ_2 is chosen by the 10-fold cross-validation method (λ_2CV). The elastic net method is implemented using lasso command of MATLAB2012a software. The 10-fold cross-validation (CV) method for tuning the penalty parameters (λ_1 and λ_2) is CV random partition using MATLAB2012a software. The value of λ estimated by 10-fold CV method is the λ with minimum mean prediction squared error as calculated by CV. For naïve elastic net estimator, the relationship between the shrinkage parameters is $\alpha = \lambda_2/(\lambda_1 + \lambda_2)$ where $\alpha \in (0,1)$. Thus, we study the performance of the elastic net estimator with variety values of α. The decision criterions are the following.

1. The prediction accuracy is measured by the prediction error (PE) defined as $E(\mathbf{y} - \hat{\mathbf{y}})^2$ where $\hat{\mathbf{y}} = \mathbf{X}\widehat{\boldsymbol{\beta}}$.

2. For each estimator $\widehat{\boldsymbol{\beta}}$, its estimation accuracy is measured by the mean square error ($MSE(\widehat{\boldsymbol{\beta}})$) defined as $E\left[(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}})^T(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}})\right]$.

3. The variable selection performance is gauged by (C, IC), where C is the number of zero coefficients that are correctly estimated by zero and IC is the number of nonzero coefficients that are incorrectly estimated by zero.

For each value of α, the average of PE, $MSE(\hat{\boldsymbol{\beta}})$, C and IC are computed based on 100 datasets of each simulation design. The standard errors of PE and $MSE(\hat{\boldsymbol{\beta}})$ are estimated using the bootstrap with B = 500 resampling from 100 PE's and 100 $MSE(\hat{\boldsymbol{\beta}})$'s, respectively.

3.1 Simulation study I

We generate 100 datasets using the simulation design proposed by Lykou and Ntzoufras [13], which consists of 15 predictor variables of 50 observations each. The first 10 predictors follow independent standard normal distribution and the last 5 predictors are generated as follows,

$$(x_{11},\ldots,x_{15}) = (x_1,\ldots,x_5) \times (0.3,0.5,0.7,0.9,1.1)^{\mathrm{T}} \times (1,1,1,1,1) + \mathbf{E},$$

where **E** consists of 5 independent $N(0,1)$ random variables. The response variable is generated as

$$y = 2x_1 - x_5 + 1.5x_7 + x_{11} + 0.5x_{13} + \varepsilon,$$

where $\varepsilon \sim N(0,2.5^2)$. This set of simulated data comprises of predictor variables that are correlated with each other. The simulation method is repeated 100 times. This dataset has different correlations between predictor variables. The last five predictors are highly correlated, whereas, there are small to moderate correlations between x_j, $j = 1,\ldots,5$ and $x_{11},\ldots,x_{15}$.

Using the process for choosing the value of the penalty parameter λ_2 based on Bayes factor (λ_2BF) described in Section 2, the value of λ_2BF is computed for 100 simulation datasets. We correct the validity of λ_2BF by using Bayes factor $\mathrm{BF}_{elastic\ net}$. The $\mathrm{BF}_{elastic\ net}$ is computed using variety value of α (Table 2). For all values of α, the simulation result reveals that the value λ_2BF gives $\mathrm{BF}_{elastic\ net} > 1$, whereas $\mathrm{BF}_{elastic\ net} > 3$ is derived using the small value of α. Hence, the appropriate value of λ_2BF is the value θ_{M_F} of the submodel M_F which has highest posterior model probability $f(\mathbf{y}|M_F)$, (the posterior model probability using the prior in (11)).

Table 3 shows result of naïve elastic net estimators for simulation study I where the penalty parameters λ_2 are chosen by λ_2CV and λ_2BF. Using the value of λ_2BF and α is close to one, the naïve elastic net estimator has the prediction performance better than the naïve elastic net estimator where the value of λ_2 is chosen by λ_2CV. Using λ_2BF, the prediction error of the naïve elastic net estimator tends to be large when α is close to zero.

At some value of α, the naïve elastic net estimator where the value of λ_2 is chosen by λ_2BF has $MSE(\widehat{\boldsymbol{\beta}})$ less than the naïve elastic net estimator where the value of λ_2 is chosen by λ_2CV. At some value of α, the naïve elastic net estimator where the value of λ_2 is chosen by λ_2BF performs both prediction performance and estimation accuracy better than the naïve elastic net estimator where the value of λ_2 is chosen by λ_2CV. The variable selection performance is gauged by (C, IC), the naïve elastic net estimator where the value of λ_2 is chosen by λ_2CV has the variable selection performance better than the naïve elastic net estimator where the value of λ_2 is chosen by λ_2BF. Nevertheless, the naïve elastic net estimator where the value of λ_2 is chosen by λ_2BF has the variable selection performance, C is close to true value of C, better than the naïve elastic net estimator where the value of λ_2 is chosen by λ_2CV when α is small.

Table 2. Summary $\mathrm{BF}_{elastic\ net}$ for simulation study I.

α	1 < $\mathrm{BF}_{elastic\ net}$< 3	3 < $\mathrm{BF}_{elastic\ net}$< 20	$\mathrm{BF}_{elastic\ net}$ > 20
0.9	100 datasets	-	-
0.8	100 datasets	-	-
0.7	100 datasets	-	-
0.6	100 datasets	-	-
0.5	100 datasets	-	-
0.4	100 datasets	-	-
0.3	100 datasets	-	-
0.2	100 datasets	-	-
0.1	100 datasets	-	-
0.05	100 datasets	16 datasets	-
0.04	100 datasets	51 datasets	1 dataset
0.03	100 datasets	94 datasets	4 datasets
0.02	100 datasets	98 datasets	38 datasets
0.01	100 datasets	98 datasets	92 datasets

Table 3. Model selection and fitting results of the naïve elastic net estimators for simulation study I where the value of penalty parameters λ_2 are chosen by λ_2CV and λ_2BF .

α	Method for choosing the value of the shrinkage parameter λ_2							
	λ_2CV				λ_2BF			
	PE	$MSE(\hat{\boldsymbol{\beta}})$	C	IC	PE	$MSE(\hat{\boldsymbol{\beta}})$	C	IC
0.9	4.7364 (0.1110)	0.1739 (0.0187)	0.53	0.07	4.2591 (0.0925)	0.1858 (0.0089)	0.13	0.04
0.8	4.7545 (0.1171)	0.1735 (0.0248)	1.22	0.13	4.2749 (0.0944)	0.1780 (0.0086)	0.32	0.05
0.7	4.7669 (0.1112)	0.1550 (0.0122)	1.81	0.23	4.2968 (0.0973)	0.1696 (0.0084)	0.55	0.07
0.6	4.8117 (0.1217)	0.1819 (0.0204)	2.47	0.29	4.3275 (0.0995)	0.1609 (0.0080)	0.84	0.09
0.5	4.8143 (0.1263)	0.1788 (0.0260)	2.92	0.37	4.3742 (0.1029)	0.1514 (0.0075)	1.25	0.11
0.4	4.8442 (0.1268)	0.1762 (0.0202)	3.49	0.37	4.4501 (0.1092)	0.1417 (0.0067)	1.81	0.19
0.3	4.8483 (0.1300)	0.1723 (0.0223)	4.04	0.50	4.5850 (0.1028)	0.1335 (0.0063)	2.69	0.28
0.2	4.9220 (0.1100)	0.1636 (0.0210)	4.70	0.58	4.8654 (0.1217)	0.1320 (0.0061)	4.02	0.43
0.1	4.8910 (0.1256)	0.1733 (0.0268)	5.00	0.64	5.7001 (0.1472)	0.1525 (0.0067)	6.14	0.92
0.09	4.9151 (0.1247)	0.1718 (0.0254)	5.12	0.68	5.8816 (0.1560)	0.1585 (0.0074)	6.55	1.02
0.08	4.8919 (0.1213)	0.1720 (0.0263)	5.02	0.71	6.1082 (0.1674)	0.1662 (0.0079)	6.94	1.13
0.07	4.8459 (0.1186)	0.1581 (0.0152)	4.93	0.69	6.4014 (0.1720)	0.1760 (0.0079)	7.30	1.21
0.06	4.8836 (0.1288)	0.1888 (0.0278)	4.95	0.69	6.7972 (0.1985)	0.1898 (0.0095)	7.75	1.38
0.05	4.8858 (0.1179)	0.1737 (0.0226)	5.08	0.69	7.3519 (0.2223)	0.2094 (0.0097)	8.15	1.52
0.04	4.8900 (0.1275)	0.1760 (0.0233)	5.16	0.78	8.2403 (0.2690)	0.2418 (0.0116)	8.71	1.69
0.03	4.9169 (0.1205)	0.1676 (0.0238)	5.36	0.74	9.7310 (0.2931)	0.2971 (0.0128)	9.19	2.12
0.02	4.8752 (0.1151)	0.1751 (0.0274)	5.11	0.73	12.5960(0.3705)	0.3943 (0.0121)	9.65	2.92
0.01	4.9237 (0.1168)	0.1761 (0.0242)	5.42	0.77	17.6766(0.3707)	0.5385 (0.0061)	9.97	4.63

The numbers in parenthesis are the corresponding standard errors of PE and $MSE(\hat{\boldsymbol{\beta}})$ estimated using the bootstrap with B = 500 resampling from the 100 PE's and 100 $MSE(\hat{\boldsymbol{\beta}})$'s, respectively. For simulation study I, the true value of C is 10.

3.2 Simulation study II

In this section, we study the performance of the λ_2BF with the simulation design where the number of parameters (p) depend on the sample size (n). The datasets are simulated by the simulation method proposed by Zou and Zhang [20].

Let $p = p_n = [4n^{1/2}] - 5$ for *n* = 100, 200, 400. The data is generated from the linear regression model

$$\mathbf{y} = \mathbf{X}^{\mathbf{T}}\boldsymbol{\beta} + \boldsymbol{\varepsilon} ,$$

where $\mathbf{y}$ is an $n \times 1$ vector of response variable, $\boldsymbol{\beta}$ is an $p \times 1$ vector of parameter of regression coefficients, and $\boldsymbol{\varepsilon}$ is an $n \times 1$ vector of random errors where $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I})$, σ = 6. Let $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_p]^T$; $\mathbf{X}_j$ is an $n \times 1$ vector of the *j* th predictor variables. $\mathbf{X}$ follows a p-dim multivariate normal distribution with zero mean and covariance $\mathbf{\Sigma}$, $\mathbf{X} \sim N_p(\mathbf{0}, \mathbf{\Sigma})$, where the covariance matrix $\mathbf{\Sigma}$ has the entry $\mathbf{\Sigma}_{j,k}$= *corr*$(j, k) = \rho^{|j-k|}$, $1 \le k$, $j \le p$. In this research, we set ρ = 0.5 and ρ = 0.75. Let $\mathbf{1}_q$ denotes a $q \times 1$ vector of 1's, and $\mathbf{0}_{p-3q}$ denotes a $(p - 3q) \times 1$ vector of 0's. The true coefficients $\boldsymbol{\beta} = \left(3 \cdot \mathbf{1}_q, 3 \cdot \mathbf{1}_q, 3 \cdot \mathbf{1}_q, \mathbf{0}_{p-3q}\right)^T$ where $q = [p_n/9]$. Let $\mathcal{A} = \{j : \beta_j \ne 0, j = 1, 2, \dots, p\}$. The size of $\mathcal{A}$ is the number of non-zero coefficients which are used to generate the response variable of the model. For this simulation method, the size of $\mathcal{A}$ is denoted by $|\mathcal{A}| = 3q$. There are six cases for combination of *n* = 100, 200, 400 and ρ = 0.5, 0.75. The simulation method is repeated 100 times.

Table 4 – Table 6 show the model selection and fitting results of the naïve elastic net estimators for simulation study II with different value of α. For every combination of (n, p, ρ) and α is not close to zero, the naïve elastic net estimator where the value of λ_2 is chosen by λ_2BF has the prediction performance better than the naïve elastic net estimator where the value of λ_2 is chosen by λ_2CV. For almost cases, the naïve elastic net estimator where the value of λ_2 is chosen by λ_2CV performs the estimation accuracy better than the naïve elastic net estimator where the value of λ_2 is chosen by λ_2BF. At some small value of α, the naïve elastic net estimator where the value of λ_2 is chosen by λ_2BF has the estimation accuracy better than the naïve elastic net estimator where the value of λ_2 is chosen by λ_2CV. For variable selection performance, the naïve elastic net estimator where the value of λ_2 is chosen by λ_2CV has the variable selection performance better than the naïve elastic net estimator where the value of λ_2 is chosen by λ_2BF. For small value of α (α is close to zero), the naïve elastic net estimator where the value of λ_2 is chosen by λ_2BF has the value C tends to the true value of C better than the naïve elastic net estimator where the value of λ_2 is chosen by λ_2CV.

Table 4. Model selection and fitting results of naïve elastic net estimators for simulation study II: n = 100 and p_n = 35.

n = 100, p_n = 35, ρ = 0.5											
ρ	Truth		α	λ_2CV				λ_2BF			
	C	IC		PE	$MSE(\widehat{\boldsymbol{\beta}})$	C	IC	PE	$MSE(\widehat{\boldsymbol{\beta}})$	C	IC
0.5	26	0	0.9	27.0197 (0.4712)	0.3476 (0.0116)	1.25	0	23.3527 0.4215)	0.7631 (0.0224)	0.26	0
			0.8	27.3919 (0.5048)	0.3190 (0.0098)	2.82	0	23.3792 (0.3803)	0.7469 (0.0215)	0.49	0
			0.7	27.6590 (0.5507)	0.3032 (0.0093)	4.50	0	23.4165 (0.3985)	0.7268 (0.0212)	0.77	0
			0.6	28.2900 (0.5340)	0.2769 (0.0097)	6.82	0	23.4717 (0.3931)	0.7014 (0.0200)	1.16	0
			0.5	28.6017 (0.5310)	0.2649 (0.0113)	8.82	0	23.5589 (0.4174)	0.6680 (0.0196)	1.65	0
			0.4	29.1849 (0.5599)	0.2465 (0.0110)	11.06	0	23.7110 (0.4057)	0.6218 (0.0179)	2.26	0
			0.3	29.7014 (0.5257)	0.2379 (0.0109)	13.57	0	24.0115 (0.4244)	0.5548 (0.0167)	3.51	0
			0.2	29.7958 (0.5952)	0.2433 (0.0119)	15.38	0	24.7206 (0.4334)	0.4537 (0.0139)	5.79	0
			0.1	30.2554 (0.5788)	0.2441 (0.0120)	17.60	0	27.1133 (0.4779)	0.2989 (0.0125)	11.76	0
			0.09	30.4487 (0.6386)	0.2339 (0.0099)	18.04	0	27.6419 (0.4904)	0.2807 (0.0120)	12.82	0
			0.08	30.2920 (0.6345)	0.2476 (0.0127)	17.92	0	28.2887 (0.5066)	0.2630 (0.0115)	14.06	0
			0.07	30.5599 (0.5508)	0.2417 (0.0116)	18.38	0	29.0935 (0.5255)	0.2466 (0.0111)	15.42	0
			0.06	30.3694 (0.6509)	0.2514 (0.0130)	18.20	0	30.1075 (0.5470)	0.2324 (0.0106)	16.95	0
			0.05	30.0570 (0.5936)	0.2626 (0.0135)	18.08	0	31.4413 (0.5579)	0.2220 (0.0098)	19.14	0
			0.04	30.4318 (0.6136)	0.2564 (0.0134)	18.71	0	33.3806 (0.6844)	0.2204 (0.0091)	20.96	0
			0.03	30.2753 (0.5783)	0.2610 (0.0135)	18.63	0	36.5138 (0.7253)	0.2393 (0.0088)	23.48	0.01
			0.02	30.3068 (0.6351)	0.2735 (0.0172)	18.98	0	43.1768 (0.8871)	0.3232 (0.0117)	25.29	0.03
			0.01	30.2835 (0.6037)	0.2658 (0.0140)	19.00	0	72.1491 (2.2532)	0.7680 (0.0285)	25.94	0.69

Table 4. (Continued).

n = 100, p_n = 35, ρ = 0.75											
	Truth			λ_2CV				λ_2BF			
ρ	C	IC	α	PE	$MSE(\hat{\beta})$	C	IC	PE	$MSE(\hat{\beta})$	C	IC
0.75	26	0	0.9	28.2136 (0.4770)	0.3010 (0.0129)	2.14	0	23.2912 (0.3677)	1.4239 (0.0483)	0.24	0
			0.8	28.4762 (0.5303)	0.2843 (0.0140)	4.30	0	23.3149 (0.3845)	1.3957 (0.0454)	0.44	0
			0.7	28.9119 (0.4746)	0.2666 (0.0173)	6.73	0	23.3477 (0.3959)	1.3608 (0.0489)	0.71	0.01
			0.6	29.4100 (0.5178)	0.2438 (0.0179)	9.28	0	23.3953 (0.3796)	1.3163 (0.0442)	1.04	0.03
			0.5	29.8884 (0.5297)	0.2239 (0.0158)	11.98	0	23.4690 (0.3887)	1.2578 (0.0444)	1.44	0.03
			0.4	30.2770 (0.4994)	0.2110 (0.0137)	14.56	0	23.5919 (0.4130)	1.1776 (0.0441)	2.31	0.04
			0.3	30.7705 (0.4875)	0.2146 (0.0155)	17.14	0	23.8238 (0.3991)	1.0618 (0.0414)	3.36	0.06
			0.2	30.6925 (0.4935)	0.2371 (0.0174)	18.69	0.01	24.3525 (0.3913)	0.8825 (0.0379)	5.55	0.05
			0.1	30.9058 (0.5215)	0.2809 (0.0189)	20.43	0.02	26.0741 (0.4032)	0.5825 (0.0305)	10.76	0.05
			0.09	31.1063 (0.5396)	0.2778 (0.0174)	20.98	0.01	26.4299 (0.4073)	0.5456 (0.0289)	11.78	0.04
			0.08	31.0708 (0.5248)	0.2966 (0.0204)	20.82	0.02	26.8578 (0.4324)	0.5077 (0.0283)	12.56	0.04
			0.07	30.9802 (0.4698)	0.3116 (0.0227)	21.05	0.01	27.3832 (0.4320)	0.4700 (0.0261)	13.93	0.05
			0.06	31.1741 (0.5264)	0.3140 (0.0221)	21.46	0.04	28.0007 (0.4376)	0.4348 (0.0274)	15.45	0.05
			0.05	31.0699 (0.4493)	0.3234 (0.0205)	21.47	0.03	28.7370 (0.4628)	0.4021 (0.0238)	17.21	0.05
			0.04	30.8949 (0.5077)	0.3441 (0.0205)	21.44	0.04	29.6760 (0.4708)	0.3727 (0.0240)	18.86	0.05
			0.03	30.8157 (0.4875)	0.3605 (0.0233)	21.36	0.06	31.0206 (0.4897)	0.3507 (0.0210)	21.01	0.06
			0.02	30.8951 (0.4950)	0.3686 (0.0210)	21.63	0.08	33.3505 (0.5347)	0.3462 (0.0209)	23.38	0.07
			0.01	30.8513 (0.4806)	0.4132 (0.0269)	21.66	0.07	40.9169 (0.9822)	0.4196 (0.0217)	25.54	0.19

The numbers in parenthesis are the corresponding standard errors of PE and $MSE(\hat{\boldsymbol{\beta}})$ estimated using the bootstrap with B = 500 resampling from the 100 PE's, and 100 $MSE(\hat{\boldsymbol{\beta}})$'s, respectively.

Table 5. Model selection and fitting results of naïve elastic net estimators for simulation study II: n = 200 and p_n = 51.

n = 200, p_n = 51, ρ = 0.5											
ρ	Truth		α	λ_2CV				λ_2BF			
	C	IC		PE	$MSE(\hat{\boldsymbol{\beta}})$	C	IC	PE	$MSE(\hat{\boldsymbol{\beta}})$	C	IC
0.5	36	0	0.9	28.3981 (0.3431)	0.1940 (0.0048)	1.76	0	26.3703 (0.3098)	0.3395 (0.0075)	0.13	0
			0.8	28.5592 (0.3387)	0.1814 (0.0046)	3.59	0	26.3796 (0.3078)	0.3340 (0.0076)	0.35	0
			0.7	28.8540 (0.3537)	0.1691 (0.0046)	5.97	0	26.3935 (0.3155)	0.3272 (0.0075)	0.65	0
			0.6	29.1381 (0.3517)	0.1567 (0.0046)	8.64	0	26.4151 (0.2952)	0.3185 (0.0075)	1.13	0
			0.5	29.4533 (0.3756)	0.1451 (0.0042)	11.68	0	26.4508 (0.3037)	0.3068 (0.0077)	1.91	0
			0.4	29.8231 (0.3635)	0.1360 (0.0043)	15.07	0	26.5157 (0.3017)	0.2906 (0.0073)	2.81	0
			0.3	30.1617 (0.3671)	0.1290 (0.0043)	18.04	0	26.6495 (0.3141)	0.2665 (0.0068)	4.34	0
			0.2	30.5984 (0.3711)	0.1259 (0.0048)	21.45	0	26.9924 (0.3119)	0.2276 (0.0059)	7.24	0
			0.1	30.8230 (0.3596)	0.1258 (0.0048)	24.35	0	28.3058 (0.3297)	0.1618 (0.0048)	15.22	0
			0.09	30.9338 (0.4049)	0.1254 (0.0049)	24.73	0	28.6078 (0.3613)	0.1536 (0.0049)	16.82	0
			0.08	30.8147 (0.3607)	0.1261 (0.0042)	24.56	0	28.9835 (0.3447)	0.1452 (0.0048)	18.49	0
			0.07	30.9431 (0.3822)	0.1261 (0.0048)	25.22	0	29.4668 (0.3560)	0.1370 (0.0047)	20.19	0
			0.06	30.9112 (0.3852)	0.1277 (0.0047)	25.20	0	30.1043 (0.3428)	0.1293 (0.0044)	22.39	0
			0.05	30.9377 (0.3633)	0.1282 (0.0043)	25.56	0	30.9635 (0.3758)	0.1233 (0.0044)	25.14	0
			0.04	31.0337 (0.3759)	0.1281 (0.0047)	25.98	0	32.1606 (0.3791)	0.1207 (0.0043)	28.11	0
			0.03	30.9982 (0.3943)	0.1302 (0.0049)	26.15	0	34.1056 (0.4076)	0.1255 (0.0044)	31.41	0
			0.02	30.9534 (0.3943)	0.1316 (0.0051)	26.09	0	38.1089 (0.4803)	0.1543 (0.0054)	34.54	0
			0.01	30.9319 (0.3871)	0.1327 (0.0050)	26.23	0	55.1290 (0.8797)	0.3336 (0.0109)	35.94	0.02

Table 5. (Continued).

n = 200, p_n = 51, ρ = 0.75											
ρ	Truth		α	λ_2CV				λ_2BF			
	C	IC		PE	$MSE(\hat{\beta})$	C	IC	PE	$MSE(\hat{\beta})$	C	IC
0.75	36	0	0.9	30.0407 (0.3708)	0.1865 (0.0063)	2.73	0	26.9033 (0.3273)	0.6741 (0.0177)	0.27	0
			0.8	30.3418 (0.3610)	0.1687 (0.0062)	6.03	0	26.9161 (0.3165)	0.6617 (0.0169)	0.62	0
			0.7	30.6939 (0.3814)	0.1528 (0.0062)	9.58	0	26.9346 (0.3096)	0.6464 (0.0174)	0.97	0
			0.6	31.0237 (0.3534)	0.1417 (0.0067)	13.23	0	26.9625 (0.3036)	0.6268 (0.0166)	1.38	0
			0.5	31.4804 (0.3823)	0.1310 (0.0058)	17.33	0	27.0076 (0.3198)	0.6007 (0.0175)	2.03	0
			0.4	31.8323 (0.3675)	0.1267 (0.0070)	21.04	0	27.0871 (0.3135)	0.5646 (0.0155)	3.14	0
			0.3	32.1215 (0.3709)	0.1246 (0.0066)	24.29	0	27.2466 (0.3398)	0.5116 (0.0166)	4.83	0
			0.2	32.3657 (0.3673)	0.1319 (0.0062)	26.83	0	27.6306 (0.3271)	0.4277 (0.0147)	8.07	0
			0.1	32.4528 (0.3603)	0.1586 (0.0078)	29.03	0	28.9601 (0.3265)	0.2878 (0.0126)	15.48	0
			0.09	32.4124 (0.3588)	0.1647 (0.0083)	28.95	0	29.2411 (0.3284)	0.2710 (0.0130)	17.05	0
			0.08	32.3310 (0.3779)	0.1687 (0.0083)	29.08	0	29.5719 (0.3070)	0.2543 (0.0119)	18.61	0
			0.07	32.3409 (0.3801)	0.1750 (0.0083)	29.18	0	29.9589 (0.3472)	0.2384 (0.0113)	20.52	0
			0.06	32.3340 (0.3683)	0.1801 (0.0081)	29.30	0	30.4266 (0.3410)	0.2233 (0.0112)	22.52	0
			0.05	32.2802 (0.3879)	0.1884 (0.0085)	29.33	0	30.9956 (0.3377)	0.2098 (0.0100)	24.73	0
			0.04	32.2772 (0.3787)	0.1941 (0.0103)	29.62	0	31.7217 (0.3360)	0.1990 (0.0099)	27.41	0
			0.03	32.3147 (0.3559)	0.1990 (0.0094)	29.82	0	32.7258 (0.3594)	0.1918 (0.0087)	30.10	0
			0.02	32.1887 (0.3809)	0.2106 (0.0096)	29.62	0	34.4767 (0.3771)	0.1928 (0.0093)	33.23	0
			0.01	32.2076 (0.3716)	0.2187 (0.0104)	29.80	0	40.4271 (0.5470)	0.2383 (0.0089)	35.64	0.01

The numbers in parenthesis are the corresponding standard errors of PE and $MSE(\hat{\boldsymbol{\beta}})$ estimated using the bootstrap with B = 500 resampling from the 100 PE's, and 100 $MSE(\hat{\boldsymbol{\beta}})$'s, respectively.

Table 6. Model selection and fitting results of naïve elastic net estimators for simulation study II: n = 400 and p_n = 75.

n = 400, p_n = 75, ρ = 0.5											
ρ	Truth		α	λ_2CV				λ_2BF			
	C	IC		PE	$MSE(\hat{\beta})$	C	IC	PE	$MSE(\hat{\beta})$	C	IC
0.5	51	0	0.9	29.7342 (0.2344)	0.1124 (0.0021)	1.98	0	28.6442 (0.2195)	0.1678 (0.0033)	0.22	0
			0.8	29.9357 (0.2417)	0.1041 (0.0023)	4.32	0	28.6470 (0.2291)	0.1660 (0.0031)	0.49	0
			0.7	30.0597 (0.2493)	0.0970 (0.0019)	7.40	0	28.6513 (0.2339)	0.1638 (0.0031)	0.98	0
			0.6	30.3179 (0.2544)	0.0892 (0.0020)	11.00	0	28.6582 (0.2288)	0.1609 (0.0031)	1.54	0
			0.5	30.6452 (0.2618)	0.0811 (0.0021)	15.57	0	28.6699 (0.2248)	0.1570 (0.0031)	2.22	0
			0.4	30.9591 (0.2519)	0.0747 (0.0020)	20.14	0	28.6921 (0.2411)	0.1514 (0.0030)	3.10	0
			0.3	31.2381 (0.2626)	0.0699 (0.0019)	24.60	0	28.7402 (0.2357)	0.1428 (0.0027)	4.64	0
			0.2	31.5680 (0.2551)	0.0660 (0.0017)	29.46	0	28.8723 (0.2247)	0.1277 (0.0026)	7.77	0
			0.1	31.6470 (0.2607)	0.0662 (0.0019)	32.67	0	29.4628 (0.2376)	0.0965 (0.0021)	15.87	0
			0.09	31.7481 (0.2753)	0.0653 (0.0020)	33.42	0	29.6174 (0.2378)	0.0917 (0.0022)	17.36	0
			0.08	31.7449 (0.2682)	0.0657 (0.0019)	33.95	0	29.8161 (0.2338)	0.0866 (0.0021)	19.45	0
			0.07	31.8506 (0.2893)	0.0649 (0.0018)	34.71	0	30.0764 (0.2399)	0.0811 (0.0019)	22.03	0
			0.06	31.8116 (0.2718)	0.0655 (0.0020)	34.73	0	30.4254 (0.2410)	0.0754 (0.0019)	25.30	0
			0.05	31.9557 (0.2607)	0.0647 (0.0018)	35.91	0	30.9026 (0.2380)	0.0698 (0.0018)	29.11	0
			0.04	31.8538 (0.2731)	0.0658 (0.0019)	35.60	0	31.5871 (0.2544)	0.0648 (0.0016)	33.83	0
			0.03	31.8563 (0.2614)	0.0659 (0.0018)	35.91	0	32.6799 (0.2632)	0.0618 (0.0017)	39.25	0
			0.02	31.7761 (0.2814)	0.0664 (0.0019)	35.69	0	34.8073 (0.2924)	0.0652 (0.0017)	46.18	0
			0.01	31.8163 (0.2749)	0.0667 (0.0019)	36.05	0	42.1288 (0.4126)	0.1075 (0.0028)	50.75	0

Table 6. (Continued).

n = 400, p_n = 75, ρ = 0.75											
ρ	Truth		α	λ_2CV				λ_2BF			
	C	IC		PE	$MSE(\hat{\beta})$	C	IC	PE	$MSE(\hat{\beta})$	C	IC
0.75	51	0	0.9	31.2760 (0.2286)	0.1174 (0.0027)	4.34	0	29.2677 (0.2259)	0.3445 (0.0068)	0.38	0
			0.8	31.4973 (0.2576)	0.1062 (0.0025)	8.87	0	29.2735 (0.2247)	0.3391 (0.0066)	0.73	0
			0.7	31.7710 (0.2440)	0.0951 (0.0025)	14.12	0	29.2823 (0.2152)	0.3323 (0.0069)	1.21	0
			0.6	32.0469 (0.2591)	0.0870 (0.0025)	19.43	0	29.2961 (0.2148)	0.3235 (0.0067)	1.85	0
			0.5	32.3491 (0.2344)	0.0807 (0.0022)	24.68	0	29.3196 (0.2215)	0.3118 (0.0065)	2.70	0
			0.4	32.6519 (0.2501)	0.0776 (0.0025)	29.57	0	29.3627 (0.2273)	0.2955 (0.0060)	4.20	0
			0.3	32.7587 (0.2390)	0.0789 (0.0023)	33.61	0	29.4533 (0.2189)	0.2711 (0.0054)	6.42	0
			0.2	33.0183 (0.2323)	0.0826 (0.0025)	37.82	0	29.6874 (0.2169)	0.2315 (0.0053)	10.71	0
			0.1	33.1332 (0.2419)	0.0960 (0.0033)	40.53	0	30.5516 (0.2175)	0.1631 (0.0041)	21.54	0
			0.09	33.0487 (0.2551)	0.0995 (0.0032)	40.52	0	30.7440 (0.2307)	0.1543 (0.0041)	23.57	0
			0.08	33.0701 (0.2497)	0.1007 (0.0035)	40.92	0	30.9765 (0.2196)	0.1454 (0.0041)	25.77	0
			0.07	33.0901 (0.2514)	0.1029 (0.0033)	41.19	0	31.2627 (0.2411)	0.1365 (0.0038)	28.04	0
			0.06	33.0349 (0.2526)	0.1057 (0.0033)	41.08	0	31.6087 (0.2271)	0.1281 (0.0037)	31.19	0
			0.05	33.0648 (0.2486)	0.1078 (0.0033)	41.49	0	32.0283 (0.2409)	0.1206 (0.0036)	34.66	0
			0.04	33.0284 (0.2537)	0.1113 (0.0036)	41.62	0	32.5531 (0.2322)	0.1142 (0.0034)	38.46	0
			0.03	33.0120 (0.2446)	0.1137 (0.0035)	41.69	0	33.2632 (0.2436)	0.1095 (0.0035)	42.62	0
			0.02	33.1257 (0.2544)	0.1159 (0.0036)	42.49	0	34.4910 (0.2413)	0.1078 (0.0035)	47.19	0
			0.01	32.8665 (0.2389)	0.1221 (0.0038)	41.34	0	38.2721 (0.2797)	0.1226 (0.0036)	50.62	0

The numbers in parenthesis are the corresponding standard errors of PE and $MSE(\hat{\boldsymbol{\beta}})$ estimated using the bootstrap with B = 500 resampling from the 100 PE's, and 100 $MSE(\hat{\boldsymbol{\beta}})$'s, respectively.

4. Real data examples

In this section, we apply two real datasets to illustrate the efficiency of the method for choosing the value of the penalty parameter λ_2 based on Bayes factor. The two datasets are the diabetes data and prostate cancer data which are used in elastic net literature and related methods.

4.1 Diabetes Data

The diabetes data is a data from Efron, Hastie, Johnstone and Tibshirani [21]. The response variable (y) is a quantitative measure of disease progression one year after baseline for 442 diabetes patients. The dataset contains 10 baseline predictor variables: AGE, SEX, body mass index (BMI), average blood pressure (BP), and six blood serum measurements: tc(S1), ldl(S2), hdl(S3), tch(S4), ltg(S5), glu(S6).

Using the method for choosing the value of the penalty parameter λ_2 based on Bayes factor described in Section 2, λ_2BF = 0.0128 is the value of penalty parameter λ_2 based on Bayes factor for the diabetes data. Table 7 shows summary of $\mathrm{BF}_{elastic\ net}$ for diabetes data, the value λ_2BF gives $\mathrm{BF}_{elastic\ net} > 1$ for all α whereas $\mathrm{BF}_{elastic\ net} > 3$ is derived with the small value of α. The result $\mathrm{BF}_{elastic\ net}$ for diabetes data is similar to the result $\mathrm{BF}_{elastic\ net}$ for simulation study I; i.e., $\mathrm{BF}_{elastic\ net} > 3$ is derived when α is close to zero.

Table 8 and Table 9 show the results of the naïve elastic net estimators for diabetes data where the shrinkage parameters λ_2 are chosen by λ_2CV and λ_2BF, respectively. The prediction error (PE) is computed for each value of α. In this research, the CV method is CV random partition. This causes the different value of (λ_1, λ_2) at each value of α. For some value of α, the naïve elastic net estimator where the value of λ_2 is chosen by λ_2BF has the prediction performance better than the naïve elastic net estimator where the value of λ_2 is chosen by λ_2CV. Using the λ_2BF, the prediction error of the naïve elastic net estimator tends to be large when α is small. Using the λ_2BF with α = 0.01, the predictors AGE, ldl, and tch are excluded. This variable selection result is the same as the result of Li and Lin [7] when the variable selection criterion of Li and Lin [7] is the scaled neighborhood criterion.

Table 7. Summary $\mathrm{BF}_{elastic\ net}$ for diabetes data.

α	1 < $\mathrm{BF}_{elastic\ net}$ < 3	3 < $\mathrm{BF}_{elastic\ net}$ < 20	$\mathrm{BF}_{elastic\ net}$ > 20
0.9	✓	-	-
0.8	✓	-	-
0.7	✓	-	-
0.6	✓	-	-
0.5	✓	-	-
0.4	✓	-	-
0.3	✓	-	-
0.2	✓	-	-
0.1	✓	-	-
0.05	✓	-	-
0.04	✓	✓	-
0.03	✓	✓	-
0.02	✓	✓	✓
0.01	✓	✓	✓

Table 8. Naïve elastic net estimators of diabetes data using λ_2CV.

α	λ_2CV	λ_1	Predictor variables										PE
			AGE	BMI	BP	S1	S2	S3	S4	S5	S6	SEX	
0.9	0.0537	0.0059	-0.0080	5.4561	1.0692	-0.1798	-0.0654	-0.6519	4.1757	43.6550	0.3303	-20.9670	2,880.171
0.8	0.0605	0.0151	-0.0050	5.4289	1.0637	-0.1642	-0.0766	-0.6649	4.1560	43.0250	0.3342	-20.7470	2,881.734
0.7	0.0677	0.0290	-0.0015	5.4001	1.0576	-0.1504	-0.0858	-0.6759	4.1302	42.4330	0.3379	-20.5040	2,883.446
0.6	0.0362	0.0241	-0.0121	5.5233	1.0800	-0.2357	-0.0172	-0.6030	4.1335	45.7490	0.3160	-21.4150	2,876.303
0.5	0.0736	0.0736	0	5.3747	1.0506	-0.1382	-0.0907	-0.6865	4.0241	41.9670	0.3391	-20.1910	2,885.256
0.4	0.0030	0.0045	-0.0313	5.6152	1.1110	-0.8008	0.4859	0.0295	5.5871	61.2020	0.2855	-22.6540	2,861.407
0.3	0.0315	0.0735	-0.0088	5.5383	1.0786	-0.2502	0	-0.5932	3.9812	46.3570	0.3087	-21.3800	2,875.600
0.2	0.0387	0.1549	0	5.5059	1.0662	-0.2104	-0.0251	-0.6380	3.6790	45.1630	0.3084	-20.8860	2,878.200
0.1	0.0228	0.2048	0	5.5714	1.0723	-0.2438	0	-0.6062	3.6280	46.7150	0.2919	-21.1500	2,875.676
0.01	0.0036	0.3579	-0	5.6538	1.0717	-0.2454	0	-0.6085	3.2822	47.8290	0.2651	-21.0770	2,876.946

Table 9. Naïve elastic net estimators of diabetes data using λ_2BF.

α	λ_2BF	λ_1	Predictor variables										PE
			AGE	BMI	BP	S1	S2	S3	S4	S5	S6	SEX	
0.9	0.0128	0.0014	-0.0243	5.6048	1.1007	-0.4581	0.1766	-0.3636	4.6361	52.2420	0.2975	-22.2460	2,868.025
0.8	0.0128	0.0032	-0.0241	5.6050	1.1005	-0.4558	0.1746	-0.3659	4.6295	52.1870	0.2974	-22.2390	2,868.085
0.7	0.0128	0.0055	-0.0238	5.6053	1.1002	-0.4529	0.1719	-0.3690	4.6210	52.1170	0.2972	-22.2290	2,868.163
0.6	0.0128	0.0085	-0.0234	5.6056	1.0999	-0.4490	0.1684	-0.3731	4.6096	52.0230	0.2969	-22.2160	2,868.267
0.5	0.0128	0.0128	-0.0229	5.6061	1.0994	-0.4436	0.1634	-0.3788	4.5937	51.8910	0.2967	-22.1980	2,868.415
0.4	0.0128	0.0192	-0.0221	5.6068	1.0987	-0.4355	0.1560	-0.3874	4.5699	51.6930	0.2962	-22.1710	2,868.638
0.3	0.0128	0.0299	-0.0207	5.6081	1.0975	-0.4219	0.1436	-0.4017	4.5301	51.3640	0.2955	-22.1260	2,869.017
0.2	0.0128	0.0512	-0.0180	5.6105	1.0952	-0.3949	0.1189	-0.4303	4.4507	50.7050	0.2939	-22.0370	2,869.803
0.1	0.0128	0.1152	-0.0100	5.6177	1.0881	-0.3138	0.0447	-0.5161	4.2123	48.7290	0.2895	-21.7670	2,872.381
0.01	0.0128	1.2672	-0	5.5551	0.9960	-0.1148	-0	-0.8154	0	45.329	0.2163	-17.4370	2,893.634

4.2 Prostate cancer data

The prostate cancer data is a data from a prostate cancer study of Stamey, Kabalin, Mcneal, et al. [22]. The response variable (y) is the logarithm of prostate specific antigen (lpsa) for 97 patients. The predictor variables are eight clinical measures: the logarithm of cancer volume (lcavol), the logarithm of prostate weight (lweight), age, the logarithm of the amount of benign prostatic hyperplasia (lbph), seminal vesicle invasion (svi), the logarithm of capsular penetration (lcp), the Gleason score (gleason), and the percentage Gleason score 4 or 5 (pgg45).

Using the method for choosing the value of the penalty parameter λ_2 based on Bayes factor described in Section 2, λ_2BF = 0.0286 is the value of penalty parameter λ_2 based on Bayes factor for the prostate cancer data. Table 10 shows summary of $\mathrm{BF}_{elastic\ net}$ for prostate cancer data, the value λ_2BF gives $\mathrm{BF}_{elastic\ net} > 1$ for all α whereas $\mathrm{BF}_{elastic\ net} > 3$ is derived with the small value of α. The result $\mathrm{BF}_{elastic\ net}$ for prostate cancer data is similar to the result $\mathrm{BF}_{elastic\ net}$ for simulation study I and diabetes data; i.e., $\mathrm{BF}_{elastic\ net} > 3$ is derived when α is close to zero.

Table 11 and Table 12 show the results of the naïve elastic net estimators for prostate cancer data where the shrinkage parameters λ_2 are chosen by λ_2CV and λ_2BF, respectively. The prediction error (PE) is computed for each value of α. For some value of α, the naïve elastic net estimator where the value of λ_2 is chosen by λ_2BF has the prediction performance better than the naïve elastic net estimator where the value of λ_2 is chosen by λ_2CV. Using the λ_2BF, the prediction error of the naïve elastic net estimator tends to be large when α is small. For α = 0.8, 0.9 where the λ_2BF has the prediction performance better than the λ_2CV, all predictors are included in the optimal model. This variable selection result is the same as the naïve elastic net of Zou and Hastie [1].

Table 10. Summary $\mathrm{BF}_{elastic\ net}$ for prostate cancer data.

α	1 < $\mathrm{BF}_{elastic\ net}$< 3	3 < $\mathrm{BF}_{elastic\ net}$ < 20	$\mathrm{BF}_{elastic\ net}$ > 20
0.9	✓	-	-
0.8	✓	-	-
0.7	✓	-	-
0.6	✓	-	-
0.5	✓	-	-
0.4	✓	-	-
0.3	✓	-	-
0.2	✓	-	-
0.1	✓	-	-
0.05	✓	✓	-
0.04	✓	✓	-
0.03	✓	✓	✓
0.02	✓	✓	✓
0.01	✓	✓	✓

Table 11. Naïve elastic net estimators of prostate cancer data using λ_2CV.

α	λ_2CV	λ_1	Predictor variables								PE
			lcavol	lweight	age	lbph	svi	lcp	gleason	pgg45	
0.9	0.0314	0.0035	0.5223	0.6079	-0.0178	0.0882	0.7031	-0.0579	0.0521	0.0036	0.4462
0.8	0.0816	0.0204	0.4654	0.5611	-0.0103	0.0679	0.6144	0	0.0404	0.0026	0.4596
0.7	0.0047	0.0020	0.554	0.6169	-0.0201	0.0939	0.7420	-0.0910	0.0478	0.0042	0.4441
0.6	0.0083	0.0055	0.5416	0.6086	-0.0185	0.0897	0.7139	-0.0704	0.0436	0.0038	0.4451
0.5	0.0224	0.0224	0.4973	0.5699	-0.0117	0.0717	0.6117	0	0.0238	0.0025	0.4555
0.4	0.0286	0.0429	0.4869	0.5263	-0.0054	0.0528	0.5828	0	0.0033	0.0022	0.4656
0.3	0.0127	0.0296	0.5009	0.5566	-0.0098	0.0662	0.6006	0	0.0128	0.0024	0.4576
0.2	0.0067	0.0269	0.5055	0.5631	-0.0108	0.0692	0.6038	0	0.0138	0.0024	0.4562
0.1	0.0007	0.0061	0.5489	0.6089	-0.0188	0.0904	0.7188	-0.0759	0.0406	0.0039	0.4447

Table 12. Naïve elastic net estimators of prostate cancer data using λ_2BF.

α	λ_2BF	λ_1	Predictor variables								PE
			lcavol	lweight	age	lbph	svi	lcp	gleason	pgg45	
0.9	0.0286	0.0032	0.5256	0.6092	-0.0181	0.0889	0.7078	-0.0616	0.0520	0.0037	0.4459
0.8	0.0286	0.0072	0.5172	0.6007	-0.0167	0.0852	0.6833	-0.0446	0.0461	0.0034	0.4475
0.7	0.0286	0.0123	0.5064	0.5899	-0.0148	0.0803	0.6517	-0.0228	0.0386	0.0030	0.4504
0.6	0.0286	0.0191	0.4946	0.5755	-0.0124	0.0740	0.6167	-0	0.0291	0.0025	0.4548
0.5	0.0286	0.0286	0.4915	0.5558	-0.0096	0.0655	0.6031	-0	0.0188	0.0024	0.4583
0.4	0.0286	0.0430	0.4869	0.5263	-0.0054	0.0528	0.5827	0	0.0033	0.0022	0.4656
0.3	0.0286	0.0668	0.4774	0.4852	-0	0.0329	0.5510	0	0	0.0016	0.4805
0.2	0.0286	0.1146	0.4638	0.4410	-0	0.0062	0.4815	0	0	0.0008	0.5042
0.1	0.0286	0.2578	0.4154	0.1950	0	0	0.2763	0	0	0	0.6089

5. Conclusion and discussion

The method for choosing the value of λ_2 based on Bayes factor, λ_2BF, improves the prediction accuracy of the elastic net method. When α is not close to zero, the elastic net estimator where the value of λ_2 is chosen by λ_2BF has the prediction performance better than the elastic net estimator where the value of λ_2 is chosen by λ_2CV. This is expected according to the L_2 part stabilizes the solution parts and improves the prediction. Using λ_2BF, the result reveals that the elastic net model is significance model as defined by Bayes factor ($\mathrm{BF}_{elastic\ net} > 1$) when $\alpha \in (0,1)$. Using λ_2BF to fit elastic net model, the penalty parameter λ_1 is derived from the relationship between the shrinkage parameters, i.e., $\alpha = \lambda_2/(\lambda_1 + \lambda_2)$. This may cause the value of λ_1 associated with λ_2BF becomes higher than the value of λ_1 derived from CV method when α is close to zero. It affects the prediction error of elastic net estimator derived from the λ_2BF becomes larger when α is small.

Elastic net does both parameter estimation and variable selection. The elastic net is based on a combination of the ridge (L_2) and the lasso (L_1) penalties. The L_1 part of the elastic net performs automatic variable selection, while the L_2 part stabilizes the solution parts and, hence, improves the prediction. In this article, we propose the λ_2BF which is the value of the penalty parameter of the L_2 part of the elastic net method; nevertheless, the elastic net estimator where the value of λ_2 is chosen by λ_2BF performs the variable selection performance better than the elastic net estimator where the value of λ_2 is chosen by λ_2CV when α is close to zero. For some small value of α, the elastic net estimator where the value of λ_2 is chosen by λ_2BF has the estimation accuracy better than elastic net estimator where the value of λ_2 is chosen by λ_2CV. Using the appropriate combination of λ_1 and λ_2, the elastic net estimator performs best in the prediction performance, the estimation accuracy and the variable selection performance.

The λ_2BF can be applied to different dataset where the number of parameters (p) less than the sample size (n), e.g. small p or the cases where the number of parameters diverges with the sample size. The method of λ_2BF can be used for adaptive elastic net estimator where the adaptive weight is included in the L_1 penalty e.g. the adaptive elastic net proposed by Zou and Zhang [20] and Ghosh [23]. In this research, the prior for σ^2 is inverse gamma distribution. The other choice is a noninformative prior $p(\sigma^2) \propto 1/\sigma^2$, and Gibbs sampling method can be used to search for the model having highest posterior probability rather than compute the entire posterior probability.

The extensions of the method proposed in this article to choose the value of the shrinkage parameter for penalized estimation in generalized linear models (e.g. regularized logistic regression [24], regularized multinomial regression [24]) are interesting for future research. It is also interesting to develop the other generalized linear model such as the generalized zero-altered Poisson regression model [25] into the penalized (regularized) regression framework, and apply the method proposed in this article to choose the value of the shrinkage parameter for the penalized version of the model in [25].

The model selection criterions AIC, BIC, and C_p can be applied for choosing the value of the penalty parameters of the elastic net. The research of Keerativibool [26] will be guidance for using these model selection criterions.

Acknowledgements

The authors gratefully acknowledge the financial support provided by Thammasat University Research Fund under the TU Research Scholar, Contract No.30/2557. The authors would like to thank the editor and the referees for helpful comments and suggestions.

References

[1] Zou, H. and Hastie, T., Regularization and Variable Selection via the Elastic Net. *Journal of the Royal Statistical Society, Series B*, 2005; 67: 301-320.

[2] Hoerl, A.E. and Kennard, R.W., Ridge Regression: Applications to Nonorthogonal Problems. *Technometrics*, 1970a; 12(1): 69-82.

[3] Hoerl, A.E. and Kennard, R.W., Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics*, 1970b; 12(1): 55-67.

[4] Tibshirani, R., Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society, Series B*, 1996; 58: 267-288.

[5] Bornn, L., Gottardo, R. and Doucet, A., *Grouping Priors and the Bayesian Elastic Net*. Technical report, University of British Columbia, Department of Statistics, 2010.

[6] Kyung, M., Gill, J., Ghosh, M. and Casella, G., Penalized Regression, Standard Errors, and Bayesian Lassos. *Bayesian Analysis*, 2010; 5(2): 369-412.

[7] Li, Q. and Lin, N., The Bayesian Elastic Net. *Bayesian Analysis*, 2010; 5(1): 151-170.

[8] Hans, C., Elastic Net Regression Modeling with the Orthant Normal Prior. *Journal of the American Statistical Association*, 2011; 106: 1383-1393.

[9] Kass, R. E. and Raftery, A. E., Bayes Factors. *Journal of the American Statistical Association*, 1995; 90: 773-795.

[10] Ando, T., *Bayesian Model Selection and Statistical Modeling*. CRC Press, New York, 2010.

[11] Raftery, A.E., Bayes Factors and BIC: Comment on “A Critique of the Bayesian Information Criterion for Model Selection”. *Sociological Methods & Research,* 1999; 27(3): 411-427.

[12] Ntzoufras, I., *Bayesian Modeling Using WinBUGS*. John Wiley & Sons, New Jersey, 2009.

[13] Lykou, A. and Ntzoufras, I., On Bayesian Lasso Variable Selection and the Specification of the Shrinkage Parameter. *Statistics and Computing*, 2012; (published online: 2012, February 14).

[14] Hastie, T., Tibshirani, R. and Friedman, J., *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (2nd ed.). Springer, New York, 2009.

[15] Chipman, H.A., George, E.I. and McCulloch, R. E., The Practical Implementation of Bayesian Model Selection (with Discussion). In *Model Selection*, ed. P. Lahiri, Beachwood, OH: IMS Lecture Notes-Monograph Series, 2001; 38: 65–134.

[16] George, E.I. and McCulloch, R.E., Approaches for Bayesian Variable Selection. *Statistica Sinica,* 1997; 7: 339-373.

[17] Park, T. and Casella, G., The Bayesian Lasso. *Journal of the American Statistical Association*, 2008; 103: 681-686.

[18] Hoerl, A.E., Kennard, R.W. and Baldwin, K.F., Ridge Regression, Some Simulations. *Communications in Statistics*, 1975; A4: 105-123.

[19] Draper, N.R. and Smith, H., *Applied Regression Analysis* (3rd ed.). John Wiley & Sons, New York, 1998.

[20] Zou, H. and Zhang, H.H., On The Adaptive Elastic-Net with a Diverging Number of Parameters. *The Annals of Statistics*, 2009; 37: 1733-1751.

[21] Efron, B., Hastie, T., Johnstone, I. and Tibshirani, R., Least Angle Regression. *The Annals of Statistics*, 2004; 32: 407-499.

[22] Stamey, T.A., Kabalin, J.N., Mcneal, J.E., Johnstone, I.M., Freiha, F., Redwine, E. A. and Yang, N., Prostate Specific Antigen in the Diagnosis and Treatment of Adenocarcinoma of the Prostate: II. Radical Prostatectomy Treated Patients. *Journal of Urology*, 1989; 141(5): 1076-1083.

[23] Ghosh, S., On The Grouped Selection and Model Complexity of the Adaptive Elastic Net. *Statistics and Computing*, 2011; 21: 451-462.

[24] Friedman, J., Hastie, T. and Tibshirani, R., Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software,* 2010; 33(1): 1-22.

[25] Kunthong, T., Ghosh, S.K. and Chaimongkol, S., Bayesian Analysis of Zero-Altered Poisson Regression Models. *Thailand Statistician: Journal of Thai Statistical Association*, 2013; 11(2): 111-131.

[26] Keerativibool, W., Study on the Penalty Functions of Model Selection Criteria. *Thailand Statistician: Journal of Thai Statistical Association*, 2014; 12(2): 161-178.

Appendix

Bayes factor for elastic net linear regression model

Zou and Hastie [1] pointed out that, solving the elastic net problem is equivalent to find the marginal posterior mode of $\boldsymbol{\beta}|\mathbf{y}$ when the prior distribution of $\boldsymbol{\beta}$ is given by a compromise between the Gaussian (normal) and Laplace (double exponential) priors. Kyung, Gill, Ghosh and Casella [6] proposed hierarchical model prior for $\boldsymbol{\beta}$ as

$$\boldsymbol{\beta}_{\boldsymbol{\gamma}}|\sigma^2,\mathbf{D}_\tau,\boldsymbol{\gamma} \sim N_{q_\gamma}(\mathbf{0},\sigma^2\mathbf{D}_\tau) \ , \tag{26}$$

where $\mathbf{D}_\tau$ is a diagonal matrix with diagonal elements $\left(\tau_j^{-2}+\lambda_2\right)^{-1}$,

$$\tau_j^2 \sim Exponential\left(\frac{\lambda_1^2}{2}\right), \ j=1,2,3,\dots,p,$$

using the prior in (26), the Bayes factor for elastic net linear regression model is

$$\mathrm{BF}_{10(elastic\ net)}=\frac{g\left(M_{F(elastic\ net)}|\mathbf{y}\right)}{g\left(M_{R(elastic\ net)}|\mathbf{y}\right)} \tag{27}$$

where the posterior model probabilities $g\left(M_{R(elastic\ net)}|\mathbf{y}\right)$ and $g\left(M_{F(elastic\ net)}|\mathbf{y}\right)$ are as follows.

$$g\left(M_{R(elastic\ net)}|\mathbf{y}\right) \equiv |\mathbf{D}_\tau|^{-\frac{1}{2}}\left|\mathbf{X}_{M_R}^T\mathbf{X}_{M_R}+\mathbf{D}_\tau^{-1}\right|^{-\frac{1}{2}}\left(\nu\xi+S_{M_R\mathbf{D}}^2\right)^{-\frac{(n+\nu)}{2}} \tag{28}$$

where

$$S_{M_R\mathbf{D}}^2=\mathbf{y}^T\mathbf{y}-\mathbf{y}^T\mathbf{X}_{M_R}\left[\mathbf{X}_{M_R}^T\mathbf{X}_{M_R}+\mathbf{D}_\tau^{-1}\right]^{-1}\mathbf{X}_{M_R}^T\mathbf{y}. \tag{29}$$

$$g\left(M_{F(elastic\ net)}|\mathbf{y}\right) \equiv |\mathbf{D}_\tau|^{-\frac{1}{2}}\left|\mathbf{X}_{M_F}^T\mathbf{X}_{M_F}+\mathbf{D}_\tau^{-1}\right|^{-\frac{1}{2}}\left(\nu\xi+S_{M_F\mathbf{D}}^2\right)^{-\frac{(n+\nu)}{2}} \tag{30}$$

where

$$S_{M_F\mathbf{D}}^2=\mathbf{y}^T\mathbf{y}-\mathbf{y}^T\mathbf{X}_{M_F}\left[\mathbf{X}_{M_F}^T\mathbf{X}_{M_F}+\mathbf{D}_\tau^{-1}\right]^{-1}\mathbf{X}_{M_F}^T\mathbf{y}. \tag{31}$$