



**HAL**  
open science

# Tuning parameter estimation in penalized least squares methodology

E. Androulakis, Christos Koukouvinos, Kalliopi Mylona

► **To cite this version:**

E. Androulakis, Christos Koukouvinos, Kalliopi Mylona. Tuning parameter estimation in penalized least squares methodology. *Communications in Statistics - Simulation and Computation*, 2011, 40 (09), pp.1444-1457. 10.1080/03610918.2011.575507 . hal-00703144

**HAL Id: hal-00703144**

**<https://hal.science/hal-00703144>**

Submitted on 1 Jun 2012

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



**Tuning parameter estimation in penalized least squares methodology**

Journal:	<i>Communications in Statistics - Simulation and Computation</i>
Manuscript ID:	LSSP-2011-0018.R1
Manuscript Type:	Original Paper
Date Submitted by the Author:	16-Mar-2011
Complete List of Authors:	Androulakis, E.; National Technical University of Athens Koukouvinos, Christos; National Technical University of Athens Mylona, Kalliopi; National Technical University of Athens
Keywords:	Penalized least squares, tuning parameter, error estimation, generalized cross validation
Abstract:	In general, tuning parameters are chosen by data-driven approaches, such as the commonly used generalized cross validation. In this article, we propose an alternative method for the derivation of the tuning parameter selector in penalized least squares framework, which can lead to an ameliorated estimate. Simulation studies are presented to support theoretical findings and a comparison of the Type I and Type II error rates, considering the $L_1$ , the hard thresholding and the Smoothly Clipped Absolute Deviation penalty functions, is performed. The results are given in tables and discussion follows.
Note: The following files were submitted by the author for peer review, but cannot be converted to PDF. You must view these files (e.g. movies) online.	
LSSP-2011-0018.tex	

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

SCHOLARONE™  
Manuscripts

For Peer Review Only

# Tuning parameter estimation in penalized least squares methodology

E. Androulakis<sup>1</sup>, C. Koukouvinos<sup>1</sup> and K. Mylona<sup>2</sup>

<sup>1</sup>Department of Mathematics

National Technical University of Athens

Zografou 15773, Athens, Greece

<sup>2</sup>Faculty of Applied Economics

Universiteit Antwerpen

2000 Prinsstraat 13, Antwerpen, Belgium

## Abstract

The efficiency of the penalization methods ([6]) depends strongly on a tuning parameter due to the fact that it controls the extent of penalization. Therefore, it is important to select it appropriately. In general, tuning parameters are chosen by data-driven approaches, such as the commonly used generalized cross validation. In this article, we propose an alternative method for the derivation of the tuning parameter selector in penalized least squares framework, which can lead to an ameliorated estimate. Simulation studies are presented to support theoretical findings and a comparison of the Type I and Type II error rates, considering the  $L_1$ , the hard thresholding and the Smoothly Clipped Absolute Deviation penalty functions, is performed. The results are given in tables and discussion follows.

*Key words and phrases:* Penalized least squares, tuning parameter, error estimation, generalized cross validation.

*AMS Subject Classification:* Primary 62J05, 62-07, Secondary 62K15

## 1 Introduction

The penalized least squares approach [6] has been consistently demonstrated to be an attractive regression shrinkage and selection method. This procedure is different from traditional approaches of variable selection in that it selects the significant variables and estimate regression coefficients simultaneously. The produced estimators are as efficient as the oracle estimator. In addition, insignificant variables are deleted by estimating their coefficients as 0. Recent related studies include [5], [7], [9] and [10].

1  
2  
3  
4 However, the efficiency of this methodology depends on appropriately choosing the tuning  
5 parameter which is involved in the penalty functions. Several methods are useful to select  
6 the tuning parameter. They are established through the use of an appropriate criterion.  
7 The minimization of this criterion with respect to the tuning parameter, can lead to the  
8 desirable selector. Data-driven approaches, such as cross validation (CV) and generalized  
9 cross validation (GCV) [4] are the most known existing methods. In this work, we are based in  
10 the estimates of the norm of the error in the solution of systems of linear equations, proposed  
11 by C. Brezinski et al. [2], [3]. The authors showed that it is possible to use these estimates  
12 as a regularization parameter selection method, by suitably adapting its expression to the  
13 chosen regularization technique. For this reason, they developed particular representations of  
14 it, specially suited to some classical regularization methods, such as Tikhonov regularization.  
15 This prompted us to derive a formula, which by minimization, can produce a tuning parameter  
16 selector in penalized least squares framework.  
17  
18  
19  
20  
21  
22

23 This paper is organized as follows: The error estimates introduced in [2] and [3] will be  
24 briefly discussed in Section 2. In Section 3 we modulate these estimates in order to obtain  
25 a tuning parameter selection method. In Section 4, we use some simulation experiments  
26 and we give the averages of the Type I and Type II errors derived from penalized least  
27 squares methodology, using our method and generalized cross validation to estimate the  
28 tuning parameter. The presented results are discussed and some comments, derived from the  
29 methods' comparison, are given. In Section 5, some general conclusions are presented.  
30  
31  
32  
33  
34

## 35 2 Error estimates for linear systems

36  
37  
38 C. Brezinski et al. [2] studied several estimates for the norm of the error in the solution of  
39 systems of linear equations. They also extended these estimates in [3] to the least squares  
40 problem  
41

$$42 \min_{x \in \mathbb{R}^n} \|\mathbf{A}\mathbf{x} - \mathbf{b}\| \quad (1)$$

43 where  $\mathbf{A} \in \mathbb{R}^{m \times n}$  is ill-conditioned and  $\mathbf{b} \in \mathbb{R}^m$  is a data vector which is considered to be  
44 contaminated by random noise with zero mean, so that the use of a regularization method  
45 is necessary in order to obtain feasible results. In contrast with their previous work [2], they  
46 imposed no restrictions on  $m$  and  $n$ , thus allowing either  $m \geq n$  or  $m < n$ .  
47  
48  
49

50 Consider the least squares problem (1), with  $\mathbf{A} \in \mathbb{R}^{m \times n}$ ,  $\text{rank}(\mathbf{A}) = r \leq \min(m, n)$  and  
51 let  $\mathbf{x}^*$  be the normal solution of (1). Denote also by  $\mathbf{x}$  any approximate solution of (1) and  
52 consider the error vector  $\mathbf{e} = \mathbf{x}^* - \mathbf{x}$ , the corresponding residual  $\mathbf{r} = \mathbf{b} - \mathbf{A}\mathbf{x}$  and the normal  
53 residual  $\boldsymbol{\rho} = \mathbf{A}^T \mathbf{r}$ . Let also  $\mathbf{A} = U\Sigma V^T$  be the singular value decomposition (SVD) of  $\mathbf{A}$  [8],  
54 with  
55  
56  
57

$$58 U = [\mathbf{u}_1, \dots, \mathbf{u}_m], \mathbf{u}_i \in \mathbb{R}^m, U^T U = I_m,$$

$$V = [\mathbf{v}_1, \dots, \mathbf{v}_n], \mathbf{v}_i \in \mathbb{R}^n, V^T V = I_n,$$

$$\Sigma = \begin{pmatrix} \Sigma_r & 0 \\ 0 & 0 \end{pmatrix} \in \mathbb{R}^{m \times n}, \Sigma_r = \text{diag}(\sigma_1, \dots, \sigma_r),$$

and  $\sigma_1 \geq \dots \geq \sigma_r > 0$ . C. Brezinski et al. proposed the following estimate of the norm of the error

$$\|\mathbf{e}\|^2 \simeq \eta_\nu^2 = d_0^{\nu-1} d_1^{5-2\nu} d_2^{\nu-3}, \nu \in \mathbb{R} \quad (2)$$

where

$$d_0 = \|\mathbf{r}\|^2 = \|U^T \mathbf{r}\|^2 = \sum_{i=1}^m \alpha_i^2$$

$$d_1 = \|\boldsymbol{\rho}\|^2 = \|\mathbf{A}^T \mathbf{r}\|^2 = \sum_{i=1}^r \sigma_i^2 \alpha_i^2$$

$$d_2 = \|\mathbf{A}\boldsymbol{\rho}\|^2 = \|\mathbf{A}\mathbf{A}^T \mathbf{r}\|^2 = \sum_{i=1}^r \sigma_i^4 \alpha_i^2 \quad (3)$$

and  $\alpha_i = \mathbf{u}_i^T \mathbf{r}$ . For more details regarding the derivation of the aforementioned estimate, as well as its application to various regularization methods such as Tikhonov regularization, the interested reader is referred to [3].

### 3 Tuning parameter selection

We now consider the case of the linear regression model  $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ , where  $\mathbf{y}$  is an  $n \times 1$  vector,  $\mathbf{X}$  is an  $n \times d$  matrix and  $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)'$  is the  $n \times 1$  vector of iid random errors, where  $\varepsilon_i \sim N(0, \sigma^2)$  for all  $i = 1, 2, \dots, n$ . The least squares estimate  $\hat{\boldsymbol{\beta}}$  is obtained via minimizing  $\|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2$ .

Fan and Li [6] proposed a family of variable selection procedures via penalized least squares and nonconcave penalized likelihood. These new methods select the significant variables and estimate regression coefficients simultaneously. In their paper, a form of the penalized least squares is

$$\frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) + n \sum_{j=1}^d p_\lambda(|\beta_j|), \quad (4)$$

where  $p_\lambda(\cdot)$  is a penalty function and  $\lambda$  is an unknown thresholding parameter, which can be chosen by data-driven approaches, such as cross-validation (CV) and generalized cross-validation (GCV, see [4]). Minimizing (4) with respect to  $\boldsymbol{\beta}$  leads to a penalized least squares estimator of  $\boldsymbol{\beta}$ . The most common penalties are the  $L_1$  penalty  $p_\lambda(|\beta|) = \lambda|\beta|$  which results in the Least Absolute Shrinkage and Selection Operator method (LASSO, see [11]), and the hard

thresholding penalty  $p_\lambda(|\beta|) = \lambda^2 - (|\beta| - \lambda)^2 I(|\beta| < \lambda)$  (see [1]), where  $I(\cdot)$  is an indicator function. However, these penalties do not simultaneously satisfy the necessary mathematical conditions for unbiasedness, sparsity and continuity. Therefore, Fan and Li [6] proposed the continuous differentiable penalty function, known as Smoothly Clipped Absolute Deviation penalty (SCAD), the first derivative of which is defined by

$$p'_\lambda(\beta) = \lambda \left\{ I(\beta \leq \lambda) + \frac{(\alpha\lambda - \beta)_+}{(\alpha - 1)\lambda} I(\beta > \lambda) \right\}, \quad (5)$$

for some  $\beta > 0$  and  $\alpha > 2$ , with  $p_\lambda(0) = 0$ . For the choice of  $\alpha$ , according to the relevant literature, the value  $\alpha \approx 3.7$  appears to perform quite satisfactorily in numerous variable selection problems. Compactly, the penalized least squares estimator is obtained as follows.

Suppose that we are given an initial value  $\beta^{(0)}$  that is close to the true value of  $\beta$ . When  $\beta_j^{(0)}$  is not very close to 0, the penalty  $p_\lambda(|\beta_j|)$  can be locally approximated by a quadratic function as

$$[p_\lambda(|\beta_j|)]' = p'_\lambda(|\beta_j|) \text{sgn}(\beta_j) \approx \{p'_\lambda(|\beta_j^{(0)}|)/|\beta_j^{(0)}|\} \beta_j, \quad (6)$$

otherwise, set  $\hat{\beta}_j = 0$ . With the local quadratic approximation, the solution for the penalized least squares can be found by iteratively computing the following ridge regression with an initial value  $\beta^{(0)}$ :

$$\beta^{(1)} = [\mathbf{X}^T \mathbf{X} + n \sum_\lambda (\beta^{(0)})]^{-1} \mathbf{X}^T \mathbf{y}, \quad (7)$$

where  $\sum_\lambda (\beta^{(0)}) = \text{diag}\{p'_\lambda(|\beta_1^{(0)}|)/|\beta_1^{(0)}|, \dots, p'_\lambda(|\beta_{d^*}^{(0)}|)/|\beta_{d^*}^{(0)}|\}$ . We note that  $d^*$  is the number of statistical significant factors, for the other factors the corresponding penalty is 0. The estimators from the full models can be used as initial estimators, as long as they are not overly parameterized.

Motivated by C. Brezinski et al. work, we now propose a new method for estimating the parameter  $\lambda$ . The solution of (4) is determined by

$$[\mathbf{X}^T \mathbf{X} + n \sum_\lambda (\beta^{(0)})] \beta^{(1)} = \mathbf{X}^T \mathbf{y}, \quad (8)$$

which implies the relation

$$\mathbf{X}^T \mathbf{r}_\lambda = n \sum_\lambda (\beta^{(0)}) \beta^{(1)}, \quad (9)$$

where  $\mathbf{r}_\lambda = \mathbf{y} - \mathbf{X} \beta^{(1)}$ . Moreover, by using our matrix notation and equations (3), the estimate (2) becomes

$$\eta_\nu(\lambda) = \|\mathbf{r}_\lambda\|^{\nu-1} \cdot \|\mathbf{X}^T \mathbf{r}_\lambda\|^{5-2\nu} \cdot \|\mathbf{X} \mathbf{X}^T \mathbf{r}_\lambda\|^{\nu-3}, \nu \in \mathbb{R}. \quad (10)$$

Hence, substituting  $\mathbf{X}^T \mathbf{r}_\lambda$  by  $n \sum_\lambda (\beta^{(0)}) \beta^{(1)}$  in (10), we have the following formula

$$\eta_\nu(\lambda) = \|\mathbf{r}_\lambda\|^{\nu-1} \cdot \|n \sum_\lambda (\beta^{(0)}) \beta^{(1)}\|^{5-2\nu} \cdot \|\mathbf{X} n \sum_\lambda (\beta^{(0)}) \beta^{(1)}\|^{\nu-3}, \nu \in \mathbb{R}. \quad (11)$$

1  
2  
3  
4 A proper minimization of (11) can lead to an ameliorated estimate of  $\lambda$ . Specifically, we start  
5 with a chosen interval containing the initial values of  $\lambda$ . For each  $\lambda$ , we use the Fan and Li  
6 algorithm [6] so as to compute a solution  $\beta^{(1)}$  and then we evaluate the norm of the error by  
7 formula (11). We note that in our formula,  $\beta^{(0)}$  is the before-last value when implementing  
8 the Fan and Li algorithm. The parameter which minimizes (11) is the finally chosen tuning  
9 parameter. We also note that, in the case when the produced penalties are all zeros, instead  
10 of (11), we minimize the initial expression (10). Otherwise, the diagonal matrix  $\sum_{\lambda}(\beta^{(0)})$   
11 would contain only zero elements and as a result, equation (11) would not be applicable.  
12  
13  
14  
15  
16

## 17 4 Simulation results

18  
19  
20 To compare the performance of the proposed method we applied simulations for a wide range  
21 of underlying models and for five statistical designs and two simulated datasets. We used three  
22 different penalty functions, the hard penalty, the  $L_1$  penalty (LASSO) and the SCAD penalty,  
23 with two different approaches for estimating the thresholding parameter  $\lambda$ : The generalized  
24 cross validation (gcv) and our method ( $\eta_{\nu}(\lambda)$ ) as an alternative. A  $2^5$  factorial design (FD),  
25 a  $2^{8-2}$  and a  $2^{10-3}$  fractional factorial of resolution V design (FFD), an OA(18,7,3,2) and  
26 an OA(27,13,3,2) are included in our simulations. In addition, the first simulated dataset  
27 (denoted as dataset 1) has 100 rows and 10 columns, the first six components (columns)  
28 of which are standard normal. The last four are independently identically distributed as a  
29 Bernoulli distribution with probability of success 0.5. The second dataset (denoted as dataset  
30 2) has also 100 rows and 10 columns and all its components are independently identically  
31 distributed as a  $U(0,20)$ .  
32  
33  
34  
35  
36  
37

38 For our simulation experiments, we developed linear models with coefficients taking values  
39 randomly selected from  $-5$  to  $5$ . When a generated coefficient was "almost zero" then it  
40 was replaced by 50% of the maximum coefficient. The distribution of the contrasts is not  
41 effected by the coefficients magnitude but it depends on the relative size of the coefficients.  
42 An  $\varepsilon_i \sim N(0,1)$  for all  $i = 1, 2, \dots, n$  random error has been added to each corresponding  
43 observation  $y_i$ . Only main effects models were considered in the simulations. The true active  
44 variables were also selected randomly, according to the uniform distribution, from the set  
45 of  $\{1, \dots, d\}$  potentially active factors, with respect to the number of active variables of the  
46 design matrix. The coefficients of the non-active variables, in the true model, were set equal  
47 to zero. The necessary parameter  $\alpha$  for the used methods was selected according to the  
48 suggestion in the original paper it appears. Specifically, we chose  $\alpha = 3.7$ . We initially set  
49 the level of significance to be 0.1 for both F-enter and F-remove in the stepwise variable  
50 selection procedure for choosing an initial value for the penalized least squares procedures  
51 with SCAD, LASSO and hard penalties. However, this choice could affect the penalized  
52  
53  
54  
55  
56  
57  
58  
59  
60



methods' performance, as well as our results. Therefore we also investigated the case in which the level of significance is set to be 0.05 for both F-enter and F-remove and also the case when it is set to be 0.05 for F-enter and 0.1 for F-remove.

The parameter  $\nu$  can take every real value. C. Brezinski et al. [3] found out that, for solving discrete ill-posed problems by Tikhonov regularization, it is better to choose  $\nu \geq 2$ . This is probably due to the need of giving a sufficient weight to the residual  $\|\mathbf{r}_\lambda\|$  in the expression of  $\eta_\nu(\lambda)$ . Our simulations prompted us to choose  $\nu \geq 3$  in order to gain the best results. Therefore, we used the value  $\nu = 4$  in all our simulations, which were conducted using MATLAB codes. We must stress that, due to the fact that gcv also uses some initial values of  $\lambda$  in its routine, the same values were given in both methods, in order to make a strict and fair comparison of them.

In Tables 1-21, the results obtained by our simulations are presented. Specifically, in the first column of these tables the specific type of the used design or dataset is presented. In the second column, the number of true active factors  $q$  used in the simulated models is shown. In the next columns we present the Type I and Type II error rates for the compared methods. These are the penalized least square approach with the SCAD penalty (SCAD),  $L_1$  penalty (LASSO) and hard thresholding penalty (Hard), using the gcv and the proposed, in this paper approach,  $\eta_\nu(\lambda)$ , as a thresholding parameter selection method. For each design type and simulated dataset and for each number of the active factors, we randomly (concerning the active variables included in the model and their coefficients) developed 1000 linear models and used them to evaluate the methods.

### Simulation results for the case when the level of significance is 0.1 for both F-enter and F-remove

**Table 1.** Methods performance for random model coefficients using 1000 simulations in a two level factorial design

Design	$q$ Factors	SCAD(gcv)		SCAD( $\eta_\nu(\lambda)$ )		LASSO(gcv)		LASSO( $\eta_\nu(\lambda)$ )		Hard(gcv)		Hard( $\eta_\nu(\lambda)$ )	
		Type I	Type II	Type I	Type II	Type I	Type II	Type I	Type II	Type I	Type II	Type I	Type II
$2^5$ FD	1	0.12	0.00	0.00	0.00	0.11	0.00	0.03	0.00	0.11	0.00	0.01	0.00
	2	0.11	0.00	0.01	0.00	0.10	0.00	0.02	0.00	0.10	0.00	0.01	0.00
	3	0.10	0.00	0.00	0.00	0.11	0.00	0.02	0.00	0.11	0.00	0.01	0.00
	4	0.11	0.00	0.00	0.00	0.10	0.00	0.01	0.00	0.10	0.00	0.01	0.00

**Table 2.** Methods performance for random model coefficients using 1000 simulations in a two level fractional factorial design

Design	$q$ Factors	SCAD(gcv)		SCAD( $\eta_\nu(\lambda)$ )		LASSO(gcv)		LASSO( $\eta_\nu(\lambda)$ )		Hard(gcv)		Hard( $\eta_\nu(\lambda)$ )	
		Type I	Type II	Type I	Type II	Type I	Type II	Type I	Type II	Type I	Type II	Type I	Type II
$2^{8-2}$ FFD	1	0.11	0.00	0.00	0.00	0.10	0.00	0.02	0.00	0.11	0.00	0.01	0.00
	2	0.10	0.00	0.00	0.00	0.11	0.00	0.01	0.00	0.11	0.00	0.01	0.00
	3	0.11	0.00	0.00	0.00	0.11	0.00	0.01	0.00	0.10	0.00	0.01	0.00
	4	0.11	0.00	0.00	0.00	0.10	0.00	0.01	0.00	0.10	0.00	0.01	0.00
	5	0.10	0.00	0.00	0.00	0.11	0.00	0.01	0.00	0.10	0.00	0.00	0.00
	6	0.10	0.00	0.00	0.00	0.10	0.00	0.01	0.00	0.10	0.00	0.01	0.00
	7	0.11	0.00	0.00	0.00	0.11	0.00	0.00	0.00	0.12	0.00	0.01	0.00

**Table 3.** Methods performance for random model coefficients using 1000 simulations in a two level fractional factorial design

Design	$q$ Factors	SCAD(gcv)		SCAD( $\eta_\nu(\lambda)$ )		LASSO(gcv)		LASSO( $\eta_\nu(\lambda)$ )		Hard(gcv)		Hard( $\eta_\nu(\lambda)$ )	
		Type I	Type II	Type I	Type II	Type I	Type II	Type I	Type II	Type I	Type II	Type I	Type II
$2^{10-3}$ FFD	1	0.10	0.00	0.00	0.00	0.11	0.00	0.01	0.00	0.10	0.00	0.00	0.00
	2	0.11	0.00	0.00	0.00	0.10	0.00	0.01	0.00	0.11	0.00	0.00	0.00
	3	0.10	0.00	0.00	0.00	0.10	0.00	0.00	0.00	0.10	0.00	0.01	0.00
	4	0.10	0.00	0.00	0.00	0.11	0.00	0.00	0.00	0.11	0.00	0.00	0.00
	5	0.10	0.00	0.00	0.00	0.10	0.00	0.00	0.00	0.10	0.00	0.00	0.00
	6	0.10	0.00	0.00	0.00	0.10	0.00	0.00	0.00	0.10	0.00	0.00	0.00
	7	0.11	0.00	0.00	0.00	0.11	0.00	0.00	0.00	0.10	0.00	0.01	0.00
	8	0.10	0.00	0.00	0.00	0.10	0.00	0.00	0.00	0.10	0.00	0.00	0.00
	9	0.11	0.00	0.00	0.00	0.00	0.11	0.00	0.00	0.00	0.09	0.00	0.01

**Table 4.** Methods performance for random model coefficients using 1000 simulations in a three level orthogonal array of strength 2

Design	$q$ Factors	SCAD(gcv)		SCAD( $\eta_\nu(\lambda)$ )		LASSO(gcv)		LASSO( $\eta_\nu(\lambda)$ )		Hard(gcv)		Hard( $\eta_\nu(\lambda)$ )	
		Type I	Type II	Type I	Type II	Type I	Type II	Type I	Type II	Type I	Type II	Type I	Type II
OA(18,7,3,2)	1	0.13	0.00	0.00	0.00	0.13	0.00	0.07	0.00	0.13	0.00	0.04	0.00
	2	0.12	0.00	0.01	0.03	0.13	0.00	0.06	0.01	0.13	0.00	0.04	0.01
	3	0.12	0.00	0.01	0.03	0.13	0.00	0.07	0.01	0.12	0.00	0.04	0.00
	4	0.12	0.00	0.01	0.02	0.12	0.00	0.07	0.00	0.12	0.00	0.03	0.01
	5	0.10	0.00	0.01	0.02	0.11	0.00	0.09	0.00	0.12	0.00	0.04	0.01
	6	0.11	0.00	0.01	0.02	0.11	0.00	0.09	0.00	0.12	0.00	0.05	0.00

**Table 5.** Methods performance for random model coefficients using 1000 simulations in a three level orthogonal array of strength 2

Design	$q$ Factors	SCAD(gcv)		SCAD( $\eta_\nu(\lambda)$ )		LASSO(gcv)		LASSO( $\eta_\nu(\lambda)$ )		Hard(gcv)		Hard( $\eta_\nu(\lambda)$ )	
		Type I	Type II	Type I	Type II	Type I	Type II	Type I	Type II	Type I	Type II	Type I	Type II
OA(27,13,3,2)	1	0.14	0.00	0.00	0.00	0.14	0.00	0.06	0.00	0.14	0.00	0.03	0.00
	2	0.14	0.00	0.01	0.01	0.14	0.00	0.05	0.00	0.14	0.00	0.03	0.00
	3	0.13	0.00	0.00	0.01	0.13	0.00	0.04	0.00	0.14	0.00	0.03	0.00
	4	0.13	0.00	0.00	0.01	0.14	0.00	0.05	0.00	0.14	0.00	0.04	0.00
	5	0.13	0.00	0.01	0.01	0.13	0.00	0.06	0.00	0.12	0.00	0.03	0.00
	6	0.13	0.00	0.01	0.01	0.12	0.00	0.06	0.00	0.13	0.00	0.03	0.00
	7	0.13	0.00	0.01	0.01	0.12	0.00	0.07	0.00	0.13	0.00	0.03	0.00
	8	0.11	0.00	0.01	0.01	0.11	0.00	0.08	0.00	0.12	0.00	0.03	0.00
	9	0.11	0.00	0.01	0.01	0.12	0.00	0.08	0.00	0.11	0.00	0.04	0.00
	10	0.11	0.00	0.01	0.01	0.12	0.00	0.09	0.00	0.12	0.00	0.04	0.00
	11	0.11	0.00	0.01	0.00	0.09	0.00	0.08	0.00	0.10	0.00	0.03	0.00
	12	0.11	0.00	0.00	0.00	0.00	0.11	0.00	0.10	0.00	0.11	0.00	0.04

**Table 6.** Methods performance for random model coefficients using 1000 simulations in dataset 1

Dataset	$q$ Factors	SCAD(gcv)		SCAD( $\eta_\nu(\lambda)$ )		LASSO(gcv)		LASSO( $\eta_\nu(\lambda)$ )		Hard(gcv)		Hard( $\eta_\nu(\lambda)$ )	
		Type I	Type II	Type I	Type II	Type I	Type II	Type I	Type II	Type I	Type II	Type I	Type II
Dataset 1	1	0.10	0.00	0.00	0.00	0.10	0.00	0.02	0.00	0.09	0.00	0.02	0.00
	2	0.10	0.00	0.00	0.00	0.10	0.00	0.01	0.00	0.10	0.00	0.02	0.00
	3	0.10	0.00	0.00	0.00	0.10	0.00	0.01	0.00	0.10	0.00	0.02	0.00
	4	0.10	0.00	0.00	0.00	0.10	0.00	0.01	0.00	0.10	0.00	0.02	0.00
	5	0.11	0.00	0.00	0.00	0.09	0.00	0.02	0.00	0.10	0.00	0.02	0.00
	6	0.10	0.00	0.00	0.00	0.10	0.00	0.02	0.00	0.10	0.00	0.02	0.00
	7	0.10	0.00	0.00	0.00	0.09	0.00	0.02	0.00	0.09	0.00	0.02	0.00
	8	0.11	0.00	0.00	0.00	0.10	0.00	0.01	0.00	0.10	0.00	0.02	0.00
	9	0.11	0.00	0.00	0.00	0.00	0.10	0.00	0.02	0.00	0.11	0.00	0.02

**Table 7.** Methods performance for random model coefficients using 1000 simulations in dataset 2

Dataset	$q$ Factors	SCAD(gcv)		SCAD( $\eta_\nu(\lambda)$ )		LASSO(gcv)		LASSO( $\eta_\nu(\lambda)$ )		Hard(gcv)		Hard( $\eta_\nu(\lambda)$ )	
		Type I	Type II	Type I	Type II	Type I	Type II	Type I	Type II	Type I	Type II	Type I	Type II
Dataset 2	1	0.10	0.00	0.00	0.00	0.10	0.00	0.01	0.00	0.11	0.00	0.02	0.00
	2	0.10	0.00	0.00	0.00	0.10	0.00	0.01	0.00	0.11	0.00	0.02	0.00
	3	0.10	0.00	0.00	0.00	0.10	0.00	0.01	0.00	0.11	0.00	0.02	0.00
	4	0.11	0.00	0.00	0.00	0.10	0.00	0.01	0.00	0.10	0.00	0.02	0.00
	5	0.11	0.00	0.00	0.00	0.10	0.00	0.01	0.00	0.09	0.00	0.02	0.00
	6	0.10	0.00	0.00	0.00	0.10	0.00	0.01	0.00	0.09	0.00	0.01	0.00
	7	0.11	0.00	0.00	0.00	0.10	0.00	0.01	0.00	0.09	0.00	0.01	0.00
	8	0.09	0.00	0.00	0.00	0.10	0.00	0.01	0.00	0.10	0.00	0.01	0.00
	9	0.10	0.00	0.00	0.00	0.00	0.00	0.10	0.00	0.01	0.00	0.01	0.00

Simulation results for the case when the level of significance is 0.05 for both F-enter and F-remove

**Table 8.** Methods performance for random model coefficients using 1000 simulations in a two level factorial design

Design	$q$ Factors	SCAD(gcv)		SCAD( $\eta_\nu(\lambda)$ )		LASSO(gcv)		LASSO( $\eta_\nu(\lambda)$ )		Hard(gcv)		Hard( $\eta_\nu(\lambda)$ )	
		Type I	Type II	Type I	Type II	Type I	Type II	Type I	Type II	Type I	Type II	Type I	Type II
$2^5$ FD	1	0.06	0.00	0.00	0.00	0.05	0.00	0.02	0.00	0.06	0.00	0.01	0.00
	2	0.06	0.00	0.00	0.00	0.05	0.00	0.01	0.00	0.06	0.00	0.01	0.00
	3	0.05	0.00	0.01	0.00	0.06	0.00	0.01	0.00	0.05	0.00	0.01	0.00
	4	0.06	0.00	0.01	0.00	0.05	0.00	0.01	0.00	0.05	0.00	0.00	0.00

**Table 9.** Methods performance for random model coefficients using 1000 simulations in a two level fractional factorial design

Design	$q$ Factors	SCAD(gcv)		SCAD( $\eta_\nu(\lambda)$ )		LASSO(gcv)		LASSO( $\eta_\nu(\lambda)$ )		Hard(gcv)		Hard( $\eta_\nu(\lambda)$ )		
		Type I	Type II	Type I	Type II	Type I	Type II	Type I	Type II	Type I	Type II	Type I	Type II	
$2^{8-2}$ FFD	1	0.05	0.00	0.00	0.00	0.05	0.00	0.02	0.00	0.05	0.00	0.00	0.00	
	2	0.05	0.00	0.00	0.00	0.05	0.00	0.01	0.00	0.05	0.00	0.00	0.00	
	3	0.05	0.00	0.00	0.00	0.06	0.00	0.00	0.00	0.06	0.00	0.01	0.00	
	4	0.05	0.00	0.00	0.00	0.05	0.00	0.00	0.00	0.05	0.00	0.01	0.00	
	5	0.05	0.00	0.00	0.00	0.00	0.00	0.05	0.00	0.06	0.00	0.01	0.00	
	6	0.05	0.00	0.00	0.00	0.00	0.04	0.00	0.00	0.00	0.05	0.00	0.01	0.00
	7	0.05	0.00	0.00	0.00	0.00	0.05	0.00	0.00	0.00	0.06	0.00	0.01	0.00

**Table 10.** Methods performance for random model coefficients using 1000 simulations in a two level fractional factorial design

Design	$q$ Factors	SCAD(gcv)		SCAD( $\eta_\nu(\lambda)$ )		LASSO(gcv)		LASSO( $\eta_\nu(\lambda)$ )		Hard(gcv)		Hard( $\eta_\nu(\lambda)$ )	
		Type I	Type II	Type I	Type II	Type I	Type II	Type I	Type II	Type I	Type II	Type I	Type II
$2^{10-3}$ FFD	1	0.05	0.00	0.00	0.00	0.05	0.00	0.01	0.00	0.05	0.00	0.00	0.00
	2	0.05	0.00	0.00	0.00	0.05	0.00	0.01	0.00	0.05	0.00	0.00	0.00
	3	0.04	0.00	0.00	0.00	0.06	0.00	0.01	0.00	0.05	0.00	0.01	0.00
	4	0.06	0.00	0.00	0.00	0.06	0.00	0.00	0.00	0.05	0.00	0.00	0.00
	5	0.06	0.00	0.00	0.00	0.06	0.00	0.00	0.00	0.05	0.00	0.00	0.00
	6	0.06	0.00	0.00	0.00	0.05	0.00	0.00	0.00	0.05	0.00	0.00	0.00
	7	0.05	0.00	0.00	0.00	0.05	0.00	0.00	0.00	0.06	0.00	0.01	0.00
	8	0.05	0.00	0.00	0.00	0.00	0.00	0.05	0.00	0.05	0.00	0.00	0.00
	9	0.05	0.00	0.00	0.00	0.00	0.06	0.00	0.00	0.00	0.06	0.00	0.01

Table 11. Methods performance for random model coefficients using 1000 simulations in a three level orthogonal array of strength 2

Design	q	SCAD(gcv)		SCAD( $\eta_\nu(\lambda)$ )		LASSO(gcv)		LASSO( $\eta_\nu(\lambda)$ )		Hard(gcv)		Hard( $\eta_\nu(\lambda)$ )	
		Type I	Type II	Type I	Type II	Type I	Type II	Type I	Type II	Type I	Type II	Type I	Type II
OA(18,7,3,2)	1	0.07	0.00	0.01	0.03	0.07	0.00	0.04	0.00	0.06	0.01	0.03	0.01
	2	0.06	0.00	0.00	0.03	0.06	0.01	0.03	0.01	0.06	0.00	0.03	0.00
	3	0.07	0.00	0.01	0.02	0.06	0.00	0.04	0.00	0.06	0.00	0.04	0.00
	4	0.05	0.00	0.01	0.02	0.06	0.00	0.06	0.00	0.05	0.01	0.03	0.01
	5	0.06	0.01	0.01	0.02	0.06	0.00	0.06	0.00	0.06	0.01	0.04	0.01
	6	0.05	0.01	0.01	0.02	0.06	0.01	0.06	0.01	0.05	0.01	0.04	0.01

Table 12. Methods performance for random model coefficients using 1000 simulations in a three level orthogonal array of strength 2

Design	q	SCAD(gcv)		SCAD( $\eta_\nu(\lambda)$ )		LASSO(gcv)		LASSO( $\eta_\nu(\lambda)$ )		Hard(gcv)		Hard( $\eta_\nu(\lambda)$ )	
		Type I	Type II	Type I	Type II	Type I	Type II	Type I	Type II	Type I	Type II	Type I	Type II
OA(27,13,3,2)	1	0.07	0.00	0.00	0.02	0.07	0.00	0.04	0.00	0.07	0.00	0.03	0.00
	2	0.07	0.00	0.00	0.01	0.07	0.00	0.03	0.00	0.06	0.00	0.03	0.00
	3	0.07	0.00	0.01	0.01	0.07	0.00	0.03	0.00	0.06	0.00	0.03	0.00
	4	0.06	0.00	0.01	0.01	0.07	0.00	0.03	0.00	0.07	0.00	0.03	0.00
	5	0.06	0.00	0.00	0.01	0.06	0.00	0.04	0.00	0.06	0.00	0.03	0.00
	6	0.06	0.00	0.01	0.01	0.06	0.00	0.04	0.00	0.06	0.00	0.03	0.00
	7	0.06	0.00	0.01	0.01	0.06	0.00	0.05	0.00	0.05	0.00	0.03	0.00
	8	0.06	0.00	0.01	0.00	0.06	0.00	0.05	0.00	0.05	0.00	0.03	0.00
	9	0.06	0.00	0.01	0.00	0.06	0.00	0.06	0.00	0.05	0.00	0.03	0.00
	10	0.06	0.01	0.00	0.01	0.07	0.01	0.06	0.01	0.06	0.00	0.03	0.00
	11	0.05	0.03	0.01	0.03	0.05	0.03	0.05	0.02	0.05	0.03	0.03	0.03
	12	0.05	0.08	0.01	0.08	0.05	0.08	0.04	0.07	0.05	0.07	0.03	0.07

Table 13. Methods performance for random model coefficients using 1000 simulations in dataset 1

Dataset	q	SCAD(gcv)		SCAD( $\eta_\nu(\lambda)$ )		LASSO(gcv)		LASSO( $\eta_\nu(\lambda)$ )		Hard(gcv)		Hard( $\eta_\nu(\lambda)$ )	
		Type I	Type II	Type I	Type II	Type I	Type II	Type I	Type II	Type I	Type II	Type I	Type II
Dataset 1	1	0.05	0.00	0.00	0.00	0.05	0.00	0.01	0.00	0.05	0.00	0.02	0.00
	2	0.04	0.00	0.00	0.00	0.06	0.00	0.01	0.00	0.05	0.00	0.02	0.00
	3	0.05	0.00	0.00	0.00	0.05	0.00	0.01	0.00	0.05	0.00	0.02	0.00
	4	0.05	0.00	0.00	0.00	0.05	0.00	0.01	0.00	0.05	0.00	0.02	0.00
	5	0.05	0.00	0.00	0.00	0.05	0.00	0.01	0.00	0.05	0.00	0.02	0.00
	6	0.05	0.00	0.00	0.00	0.05	0.00	0.01	0.00	0.05	0.00	0.02	0.00
	7	0.05	0.00	0.00	0.00	0.05	0.00	0.01	0.00	0.05	0.00	0.02	0.00
	8	0.05	0.00	0.00	0.00	0.05	0.00	0.01	0.00	0.05	0.00	0.02	0.00
	9	0.05	0.00	0.00	0.00	0.05	0.00	0.01	0.00	0.04	0.00	0.01	0.00

Table 14. Methods performance for random model coefficients using 1000 simulations in dataset 2

Dataset	q	SCAD(gcv)		SCAD( $\eta_\nu(\lambda)$ )		LASSO(gcv)		LASSO( $\eta_\nu(\lambda)$ )		Hard(gcv)		Hard( $\eta_\nu(\lambda)$ )	
		Type I	Type II	Type I	Type II	Type I	Type II	Type I	Type II	Type I	Type II	Type I	Type II
Dataset 2	1	0.05	0.00	0.00	0.00	0.05	0.00	0.02	0.00	0.05	0.00	0.01	0.00
	2	0.05	0.00	0.00	0.00	0.05	0.00	0.01	0.00	0.05	0.00	0.01	0.00
	3	0.04	0.00	0.00	0.00	0.05	0.00	0.01	0.00	0.05	0.00	0.01	0.00
	4	0.05	0.00	0.00	0.00	0.05	0.00	0.01	0.00	0.05	0.00	0.01	0.00
	5	0.06	0.00	0.00	0.00	0.05	0.00	0.01	0.00	0.05	0.00	0.01	0.00
	6	0.05	0.00	0.00	0.00	0.05	0.00	0.01	0.00	0.05	0.00	0.01	0.00
	7	0.05	0.00	0.00	0.00	0.05	0.00	0.01	0.00	0.05	0.00	0.01	0.00
	8	0.06	0.00	0.00	0.00	0.06	0.00	0.01	0.00	0.05	0.00	0.01	0.00
	9	0.05	0.00	0.00	0.00	0.04	0.00	0.01	0.00	0.06	0.00	0.01	0.00

Simulation results for the case when the level of significance is 0.05 for F-enter  
and 0.1 for F-remove

Table 15. Methods performance for random model coefficients using 1000  
simulations in a two level factorial design

Design	$q$ Factors	SCAD(gcv)		SCAD( $\eta_\nu(\lambda)$ )		LASSO(gcv)		LASSO( $\eta_\nu(\lambda)$ )		Hard(gcv)		Hard( $\eta_\nu(\lambda)$ )	
		Type I	Type II	Type I	Type II	Type I	Type II	Type I	Type II	Type I	Type II	Type I	Type II
$2^5$ FD	1	0.06	0.00	0.00	0.00	0.06	0.00	0.03	0.00	0.06	0.00	0.01	0.00
	2	0.10	0.00	0.00	0.00	0.10	0.00	0.01	0.00	0.12	0.00	0.01	0.00
	3	0.11	0.00	0.01	0.00	0.10	0.00	0.01	0.00	0.10	0.00	0.01	0.00
	4	0.10	0.00	0.01	0.00	0.10	0.00	0.01	0.00	0.12	0.00	0.01	0.00

Table 16. Methods performance for random model coefficients using 1000  
simulations in a two level fractional factorial design

Design	$q$ Factors	SCAD(gcv)		SCAD( $\eta_\nu(\lambda)$ )		LASSO(gcv)		LASSO( $\eta_\nu(\lambda)$ )		Hard(gcv)		Hard( $\eta_\nu(\lambda)$ )	
		Type I	Type II	Type I	Type II	Type I	Type II	Type I	Type II	Type I	Type II	Type I	Type II
$2^{8-2}$ FFD	1	0.07	0.00	0.00	0.00	0.07	0.00	0.02	0.00	0.07	0.00	0.01	0.00
	2	0.10	0.00	0.00	0.00	0.11	0.00	0.01	0.00	0.11	0.00	0.00	0.00
	3	0.11	0.00	0.00	0.00	0.10	0.00	0.01	0.00	0.11	0.00	0.00	0.00
	4	0.11	0.00	0.00	0.00	0.11	0.00	0.00	0.00	0.10	0.00	0.00	0.00
	5	0.10	0.00	0.00	0.00	0.10	0.00	0.00	0.00	0.11	0.00	0.01	0.00
	6	0.11	0.00	0.00	0.00	0.10	0.00	0.01	0.00	0.11	0.00	0.01	0.00
	7	0.10	0.00	0.00	0.00	0.10	0.00	0.00	0.00	0.10	0.00	0.00	0.00

Table 17. Methods performance for random model coefficients using 1000  
simulations in a two level fractional factorial design

Design	$q$ Factors	SCAD(gcv)		SCAD( $\eta_\nu(\lambda)$ )		LASSO(gcv)		LASSO( $\eta_\nu(\lambda)$ )		Hard(gcv)		Hard( $\eta_\nu(\lambda)$ )	
		Type I	Type II	Type I	Type II	Type I	Type II	Type I	Type II	Type I	Type II	Type I	Type II
$2^{10-3}$ FFD	1	0.08	0.00	0.00	0.00	0.08	0.00	0.02	0.00	0.07	0.00	0.00	0.00
	2	0.10	0.00	0.00	0.00	0.10	0.00	0.01	0.00	0.10	0.00	0.00	0.00
	3	0.12	0.00	0.00	0.00	0.12	0.00	0.01	0.00	0.11	0.00	0.00	0.00
	4	0.10	0.00	0.00	0.00	0.10	0.00	0.00	0.00	0.10	0.00	0.00	0.00
	5	0.10	0.00	0.00	0.00	0.10	0.00	0.00	0.00	0.10	0.00	0.00	0.00
	6	0.10	0.00	0.00	0.00	0.10	0.00	0.00	0.00	0.09	0.00	0.00	0.00
	7	0.10	0.00	0.00	0.00	0.11	0.00	0.00	0.00	0.10	0.00	0.00	0.00
	8	0.11	0.00	0.00	0.00	0.10	0.00	0.00	0.00	0.10	0.00	0.00	0.00
	9	0.10	0.00	0.00	0.00	0.10	0.00	0.00	0.00	0.11	0.00	0.00	0.00

Table 18. Methods performance for random model coefficients using 1000  
simulations in a three level orthogonal array of strength 2

Design	$q$ Factors	SCAD(gcv)		SCAD( $\eta_\nu(\lambda)$ )		LASSO(gcv)		LASSO( $\eta_\nu(\lambda)$ )		Hard(gcv)		Hard( $\eta_\nu(\lambda)$ )	
		Type I	Type II	Type I	Type II	Type I	Type II	Type I	Type II	Type I	Type II	Type I	Type II
OA(18,7,3,2)	1	0.08	0.00	0.01	0.04	0.08	0.00	0.04	0.00	0.08	0.00	0.04	0.00
	2	0.13	0.01	0.01	0.02	0.12	0.01	0.05	0.01	0.12	0.00	0.03	0.00
	3	0.13	0.00	0.01	0.03	0.12	0.00	0.06	0.00	0.12	0.00	0.04	0.00
	4	0.13	0.00	0.01	0.02	0.12	0.00	0.08	0.00	0.11	0.01	0.04	0.01
	5	0.10	0.00	0.01	0.02	0.10	0.00	0.08	0.00	0.11	0.00	0.04	0.00
	6	0.09	0.00	0.01	0.02	0.11	0.00	0.09	0.00	0.10	0.01	0.03	0.01

**Table 19.** Methods performance for random model coefficients using 1000 simulations in a three level orthogonal array of strength 2

Design	$q$ Factors	SCAD(gcv)		SCAD( $\eta_\nu(\lambda)$ )		LASSO(gcv)		LASSO( $\eta_\nu(\lambda)$ )		Hard(gcv)		Hard( $\eta_\nu(\lambda)$ )		
		Type I	Type II	Type I	Type II	Type I	Type II	Type I	Type II	Type I	Type II	Type I	Type II	
OA(27,13,3,2)	1	0.10	0.00	0.00	0.02	0.10	0.00	0.05	0.00	0.10	0.00	0.03	0.00	
	2	0.14	0.00	0.00	0.01	0.14	0.00	0.05	0.00	0.14	0.00	0.04	0.00	
	3	0.14	0.00	0.01	0.01	0.13	0.00	0.04	0.00	0.13	0.00	0.03	0.00	
	4	0.13	0.00	0.01	0.01	0.14	0.00	0.05	0.00	0.13	0.00	0.03	0.00	
	5	0.14	0.00	0.01	0.01	0.14	0.00	0.06	0.00	0.14	0.00	0.03	0.00	
	6	0.12	0.00	0.01	0.01	0.13	0.00	0.07	0.00	0.13	0.00	0.03	0.00	
	7	0.13	0.00	0.01	0.01	0.12	0.00	0.08	0.00	0.12	0.00	0.04	0.00	
	8	0.12	0.00	0.01	0.01	0.12	0.00	0.09	0.00	0.12	0.00	0.03	0.00	
	9	0.12	0.00	0.01	0.00	0.11	0.00	0.08	0.00	0.11	0.00	0.03	0.00	
	10	0.11	0.01	0.01	0.01	0.12	0.01	0.10	0.01	0.11	0.00	0.04	0.00	
	11	0.11	0.01	0.01	0.01	0.01	0.11	0.02	0.10	0.02	0.10	0.02	0.04	0.02
	12	0.10	0.04	0.00	0.04	0.09	0.03	0.08	0.03	0.10	0.03	0.04	0.03	

**Table 20.** Methods performance for random model coefficients using 1000 simulations in dataset 1

Dataset	$q$ Factors	SCAD(gcv)		SCAD( $\eta_\nu(\lambda)$ )		LASSO(gcv)		LASSO( $\eta_\nu(\lambda)$ )		Hard(gcv)		Hard( $\eta_\nu(\lambda)$ )	
		Type I	Type II	Type I	Type II	Type I	Type II	Type I	Type II	Type I	Type II	Type I	Type II
Dataset 1	1	0.07	0.00	0.00	0.00	0.06	0.00	0.01	0.00	0.06	0.00	0.01	0.00
	2	0.10	0.00	0.00	0.00	0.10	0.00	0.01	0.00	0.10	0.00	0.02	0.00
	3	0.11	0.00	0.00	0.00	0.10	0.00	0.01	0.00	0.09	0.00	0.02	0.00
	4	0.10	0.00	0.00	0.00	0.09	0.00	0.01	0.00	0.10	0.00	0.02	0.00
	5	0.09	0.00	0.00	0.00	0.10	0.00	0.01	0.00	0.10	0.00	0.02	0.00
	6	0.10	0.00	0.00	0.00	0.10	0.00	0.02	0.00	0.10	0.00	0.03	0.00
	7	0.09	0.00	0.00	0.00	0.10	0.00	0.01	0.00	0.09	0.00	0.02	0.00
	8	0.10	0.00	0.00	0.00	0.11	0.00	0.02	0.00	0.10	0.00	0.02	0.00
	9	0.10	0.00	0.00	0.00	0.11	0.00	0.02	0.00	0.11	0.00	0.02	0.00

**Table 21.** Methods performance for random model coefficients using 1000 simulations in dataset 2

Dataset	$q$ Factors	SCAD(gcv)		SCAD( $\eta_\nu(\lambda)$ )		LASSO(gcv)		LASSO( $\eta_\nu(\lambda)$ )		Hard(gcv)		Hard( $\eta_\nu(\lambda)$ )	
		Type I	Type II	Type I	Type II	Type I	Type II	Type I	Type II	Type I	Type II	Type I	Type II
Dataset 2	1	0.07	0.00	0.00	0.00	0.07	0.00	0.02	0.00	0.07	0.00	0.01	0.00
	2	0.10	0.00	0.00	0.00	0.11	0.00	0.01	0.00	0.10	0.00	0.02	0.00
	3	0.11	0.00	0.00	0.00	0.11	0.00	0.01	0.00	0.10	0.00	0.02	0.00
	4	0.10	0.00	0.00	0.00	0.10	0.00	0.01	0.00	0.10	0.00	0.01	0.00
	5	0.10	0.00	0.00	0.00	0.09	0.00	0.01	0.00	0.11	0.00	0.02	0.00
	6	0.10	0.00	0.00	0.00	0.09	0.00	0.01	0.00	0.10	0.00	0.01	0.00
	7	0.10	0.00	0.00	0.00	0.09	0.00	0.01	0.00	0.10	0.00	0.02	0.00
	8	0.09	0.00	0.00	0.00	0.09	0.00	0.01	0.00	0.10	0.00	0.01	0.00
	9	0.09	0.00	0.00	0.00	0.10	0.00	0.01	0.00	0.09	0.00	0.01	0.00

In the following remark, we quote some general conclusions, which can be obtained from the Type I and Type II error rates, appeared in the above tables.

**Remark.** From the Tables 1-21, we extract the following conclusions:

1. The improvement of Type I error rates is obvious and extremely satisfactory when implementing our method for selecting the thresholding parameter.
2. Concerning the rates of Type II errors, they generally remain at the zero level, regardless of the thresholding parameter selection method we use. There is only a negligible difference when we use the two OAs.

3. Between the penalized methods with the use of  $\eta_\nu(\lambda)$ , we notice that when the design matrix is one of the five statistical designs, the rate of the Type I error is lower in the SCAD( $\eta_\nu(\lambda)$ ) method, followed by the Hard( $\eta_\nu(\lambda)$ ) method and then the LASSO( $\eta_\nu(\lambda)$ ) method. On the other hand, the rate of the Type II error (at the expense of the Type I error) is lower, when using the LASSO( $\eta_\nu(\lambda)$ ) method, followed by an almost equal performance of the Hard( $\eta_\nu(\lambda)$ ) method. The SCAD( $\eta_\nu(\lambda)$ ) method has a slightly different Type II error rate in the case of the two OAs. However, when we use the two simulated datasets, the LASSO( $\eta_\nu(\lambda)$ ) and the Hard( $\eta_\nu(\lambda)$ ) method have an almost equal performance and the SCAD( $\eta_\nu(\lambda)$ ) method has both error rates equal to zero.
4. The important advantage when using our approach to select the parameter  $\lambda$ , is that the penalized methods (apart from the LASSO( $\eta_\nu(\lambda)$ ) method in the case of the two OAs) have a stable performance with both error rates generally at an almost zero level, regardless of the growth of the important factors.
5. Moreover, the alteration in the value of the significance level for F-enter and F-remove in the stepwise variable selection procedure does not substantially affect the performance of the penalized methods when using  $\eta_\nu(\lambda)$ . We only observe a slight increase in the Type II error rates as the number of the true active factors grows, but only for the case when we use the OA(27,13,3,2), which however, are almost equal with the Type II error rates of gcv. Therefore, the selection of the significance level is left to the decision of the experimenters.

## 5 Concluding Remarks

In this paper we extend the estimates given in [2] and [3] of the norm of the error in the solution of systems of linear equations, to the case of penalized least squares. We derive a new expression which by minimization can lead to an estimate of parameter  $\lambda$ . The choice of this parameter is very crucial because it controls the extent of penalization. In addition, it influences the magnitude of Type I and Type II error rates of the penalized methods. There is always a cost of declaring an inactive factor to be active (Type I error), and a cost of declaring an active effect to be inactive (Type II error). As a result, one must keep these error rates at a low level. To this end, generalized cross validation is essentially used. Despite its popularity and effectiveness, our simulations highlight that our method is a quite trustworthy procedure, as it produces low error rates, which in general are at a zero level. Moreover, it is easy to implement. Therefore, it could be considered an effective alternative.

## 6 Acknowledgments

We would like to thank Professor Claude Brezinski for his insightful comments and suggestions, and Professors Dennis K.J. Lin, Richard Runze Li and Giuseppe Rodriguez for sending the Matlab codes for the procedures proposed in their papers. We also thank the referees for their useful remarks which led to a substantial improvement in the content and the presentation of the paper.

## References

- [1] Antoniadis, A. (1997), Wavelets in statistics: a review (with discussion), *J. Italian Statist. Assos.*, 6, 97-144.
- [2] Brezinski, C., Rodriguez, G. and Seatzu, S. (2008), Error estimates for linear systems with applications to regularization, *Numer. Algor.*, 49, 85-104.
- [3] Brezinski, C., Rodriguez, G. and Seatzu, S. (2009), Error estimates for the regularization of least squares problems, *Numer. Algor.*, 51, 61-76.
- [4] Craven, P. and Wahba, G. (1979), Smoothing noisy data with spline functions: estimating the correct degree of smoothing by the method of generalized cross-validation, *Numer. Math.*, 31, 377-403.
- [5] Efron, B., Hastie, T., Johnstone, I. and Tibshirani, R. (2004), Least angle regression, *Ann. Statist.*, 32, 407-499.
- [6] Fan, J. and Li, R. (2001), Variable selection via nonconcave penalized likelihood and its oracle properties, *J. Amer. Statist. Assos.*, 96, 1348-1360.
- [7] Fan, J. and Li, R. (2006), *Statistical challenges with high dimensionality: Feature selection in knowledge discovery*, Proceedings of the International Congress of Mathematicians, M. Sanz-Sole, J. Soria, J.L. Varona, and J. Verdera, eds., Vol. III, 595-622.
- [8] Golub, G.H., Van Loan, C.F. (1989), *Matrix Computations*, The John Hopkins University Press, Baltimore.
- [9] James, G.M., Radchenko, P. and Lv, J. (2009), DASSO: connections between the Dantzig selector and lasso, *J. Roy. Statist. Soc. Ser. B*, 71, 127-142.
- [10] Leng, C., Lin, Y. and Wahba, G. (2006), A note on the LASSO and related procedures in model selection, *Stat. Sin.*, 16, 1273-1284.



1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

[11] Tibshirani, R.J. (1996), Regression shrinkage and selection via the LASSO, *J. Roy. Statist. Soc. Ser. B*, 58, 267-288.

For Peer Review Only

**Communications in Statistics - Simulation and Computation**

Letter to the Referees concerning our revision of the submission LSSP-2011-0018

Tuning parameter estimation in penalized least squares methodology

by E. Androulakis, C. Koukouvinos and K. Mylona.

This letter describes the revision of our LSSP-2011-0018 submission entitled "Tuning parameter estimation in penalized least squares methodology" according to the requests of the referee's report. Taking into consideration the points you kindly mentioned at your report we give a response as follows:

We left the reference [1], as it is used in order to cite the hard thresholding penalty. Please see page 4, line 1.