



## Variable Selection for Screening Experiments

Runze Li<sup>1</sup> and Dennis K. J. Lin<sup>2</sup>

<sup>1</sup>Department of Statistics and the Methodology Center, Pennsylvania State University, University Park, PA, USA

<sup>2</sup>Department of Supply Chain and Information Systems, Pennsylvania State University, University Park, PA, USA

(Received September 2006, accepted August 2007)

---

**Abstract:** The first step in many applications of response surface methodology is typically the screening process. Variable selection plays an important role in screening experiments when a large number of potential factors are introduced in a preliminary study. Traditional approaches, such as the best subset variable selection and stepwise deletion, may not be appropriate in this situation. In this paper we introduce a variable selection procedure via penalized least squares with the SCAD penalty. An algorithm to find the penalized least squares solution is suggested, and a standard error formula for the penalized least squares estimate is derived. With a proper choice of the regularization parameter, it is shown that the resulting estimate is root  $n$  consistent and possesses an oracle property; namely, it works as well as if the correct submodel were known. An automatic and data-driven approach was proposed to select the regularization parameter. Examples are used to illustrate the effectiveness of the newly proposed approach. The computer codes (written in MATLAB) to perform all calculation are available through the authors for an automatic data-driven variable selection procedure.

Keywords: Bayesian variable selection, penalized least squares, SCAD, supersaturated design.

---

### 1. Introduction

Many preliminary studies in industrial experimentation contain a large number of potentially relevant factors, but the number of actual active effects are believed to be sparse. The variable selection (screening) process is typically the first step in many applications of response surface methodology. It has received increasing attention recently. Traditional approaches, such as the best subset variable selection and stepwise deletion, may not be appropriate in this situation (Westfall *et al.* [18]). All-subset regression is typically preferable to stepwise regression, but all-subset regression is extremely time-consuming and in fact may be infeasible in practice.

Fan and Li [5] proposed a nonconcave penalized likelihood approach to selecting significant variables. Unlike stepwise variable selection and subset regression, their approaches delete insignificant variables by estimating their coefficients to be 0. Thus, their approaches estimate the regression coefficients and select significant variables simultaneously. This enables one to establish sampling properties for resulting estimates. From a theoretic point of view, they showed that their approach possesses an oracle property with proper choice of penalty function and regularization parameter; namely, the true coefficients that are zero are automatically estimated as zero, and the other coefficients are estimated as if the true submodel were known in advance.

Building on the work of Fan and Li [5], we introduce a variable selection procedure for analyzing experimental designs. Our approach can be directly implemented for cases in which the final (projective) design matrix is of full rank. The design matrix for the full

model, however, is allowed to be singular. Our settings are different from those in Fan and Li [5], which focus on likelihood-based linear models and in which covariates are assumed to be random. In our settings, we have shown that the resulting estimate is root  $n$  consistent and possesses an oracle property. We also propose an iterative ridge regression algorithm to find the penalized least squares solution. A standard error formula for the estimate is derived. This enables us to make statistical inference on the resulting model. An automatic data-driven approach to finding the tuning parameter is suggested.

This paper is organized as follows. In Section 2, the method for variable selection via nonconvex penalized least squares is introduced. Section 3 discusses the issues related to practical implementation and derive a standard error formula for the resulting estimate. Section 4 summarizes the proposed procedure by an algorithm. In Section 5, we demonstrate the nonconvex penalized least squares methods by two examples – one is based on Plackett and Burman design and the other one is based on uniform design (Fang and Lin, [7]). Section 6 presents conclusions and discussions.

## 2. Nonconcave Penalized Likelihood and Variable Selection

Suppose that the observations  $(x_i, y_i)$  are independent and identically distributed samples from a population with the density  $f(y, x^T \beta)$ . A form of penalized likelihood is

$$\frac{1}{n} \sum_{i=1}^n \log f(y_i, x_i^T \beta) - \sum_{j=1}^d p_\lambda(|\beta_j|),$$

where  $d$  is the dimension of  $x_i$ ,  $p_\lambda(\cdot)$  is a penalty function and  $\lambda$  is a tuning parameter, which controls the model complexity. The penalty function is not necessarily the same for all coefficients. For example, one may wish to keep important factors in a parametric model and hence not be willing to penalize their corresponding coefficients. For simplicity of presentation, we will assume that the penalty function is the same for all coefficients. Extension to the case with different penalty function is straightforward, however.

Some penalty functions have been used in the literature. The  $L_2$  penalty  $p_\lambda(|\beta|) = \frac{\lambda}{2} |\beta|^2$  yields a ridge regression, and the  $L_1$  penalty  $p_\lambda(|\beta|) = \lambda |\beta|$  results in LASSO (Tibshirani [17]). More generally, the  $L_q$  penalty leads to a bridge regression (see Frank and Friedman [10]). The hard thresholding penalty, defined by  $p_\lambda(|\beta|) = \lambda^2 - (|\beta| - \lambda)^2 I(|\beta| < \lambda)$ , where  $I(\cdot)$  is an indicator function, was studied by Antoniadis and Fan [1]. The penalized likelihood estimate is a hard thresholding rule (Donoho and Johnstone [4]) when the observations are from an ordinary linear regression model with an orthonormal design matrix (Fan and Li [5]).

Some conditions on  $p_\lambda(|\cdot|)$  are needed for the penalized likelihood approach to be effective. In particular,  $p_\lambda(|\cdot|)$  should be irregular at the origin, *i.e.*,  $p'_\lambda(0+) > 0$ . Antoniadis and Fan [1] provide more insights into how to choose a penalty function from theoretic points of view. A good penalty function should result in an estimator with the following three properties: *unbiasedness* for a large true coefficient to avoid excessive estimation bias, *sparsity* (estimating a small coefficient as zero) to reduce model complexity, and *continuity* to avoid unnecessary variation in model prediction. Necessary conditions for unbiasedness, sparsity and continuity have been derived by Antoniadis and Fan [1]. None of the  $L_q$ , including  $q=1$  and 2, and the hard thresholding penalty function simultaneously satisfy these three mathematical conditions.

A simple penalty function that satisfies all the three mathematical requirements is the smoothly clipped absolute deviation (SCAD) penalty proposed by Fan and Li [5], and its first order derivative is defined by

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

for some  $a > 2$  and  $\beta > 0$ . For simplicity of presentation, we will use SCAD for all procedures using the SCAD penalty. Denote by  $\beta_0$  the true value of  $\beta$ , and let  $\beta_0 = (\beta_{10}, \dots, \beta_{d0})^T = (\beta_{10}^T, \beta_{20}^T)^T$ . Without loss of generality, it is assumed that  $\beta_{20} = 0$ , and all components of  $\beta_{10}$  are not zero. Under some regularity conditions, the SCAD estimator  $\hat{\beta} = (\beta_1^T, \beta_2^T)^T$  possesses the following oracle property: With probability tending to 1, (i)  $\hat{\beta}_2 = 0$ ; and (ii)  $\sqrt{n}(\hat{\beta}_1 - \beta_{10}) \rightarrow N\{0, I_1^{-1}(\beta_{10}, 0)\}$  where  $I_1(\beta_{10}, 0)$  is the Fisher information matrix for  $\beta_1$  knowing  $\beta_2 = 0$  (see Fan and Li [5]).

From a Bayesian point of view and simulation comparisons, Fan and Li [5] suggested that  $a = 3.7$ . This will be used throughout this paper. Figure 1 depicts the  $L_2$ ,  $L_1$ ,  $L_{0.5}$  and SCAD penalties. From Figure 1, it can be seen that all penalties but the  $L_2$  penalty are singular at the origin. In fact, this is a necessary condition for the sparsity (Fan and Li [5]).

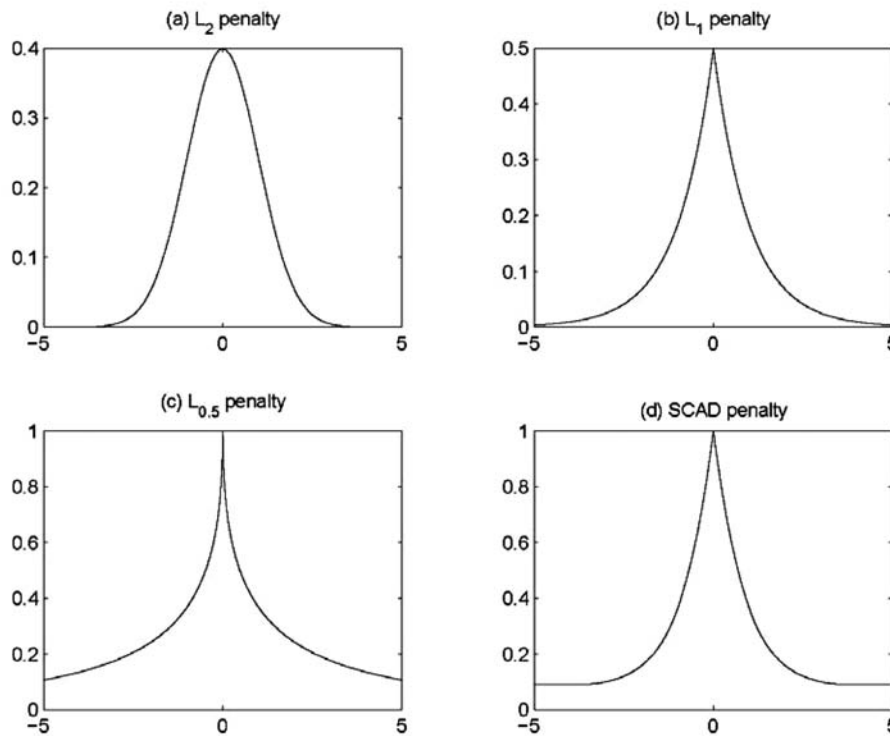


Figure 1. Plot of penalty functions: (a)  $L_2$ , (b)  $L_1$ , (c) hard thresholding, (d) SCAD penalties. The values of  $\lambda$  are 0.5, 1.05, 1.55 and 1.01 for (a), (b), (c) and (d), respectively. These values are the same as those in Fan and Li [5].

### 3. Variable Selection for Designed Experiments

Assume that observations  $y_1, \dots, y_n$  are an independent sample from the linear regression model

$$Y_i = \mathbf{x}_i^T \boldsymbol{\beta} + \varepsilon_i \quad (1)$$

with  $E(\varepsilon_i) = 0$  and  $\text{var}(\varepsilon_i) = \sigma^2$ , where  $\mathbf{x}_i$  corresponds to a fixed experiment design. As a natural extension of the penalized likelihood, we will propose a screening experiment approach via penalized least squares. Thus, the proposed approach does not require the error distribution to be known. Let  $\mathbf{y} = (y_1, \dots, y_n)^T$ , and  $\mathbf{X}$  is the corresponding design matrix. A form of penalized least squares is

$$Q(\boldsymbol{\beta}) = \frac{1}{2n} \sum_{i=1}^n (Y_i - \mathbf{x}_i^T \boldsymbol{\beta})^2 + \sum_{j=1}^d p_\lambda(|\beta_j|). \quad (2)$$

It has been shown by Li and Lin [13] that the oracle property holds for the penalized least squares estimate when the covariate  $\mathbf{x}$  is a fixed design under the following conditions. Let  $s$  be the dimension of  $\boldsymbol{\beta}_{10}$ , *i.e.*, the number of nonzero coefficients.

(C1)  $\max_{1 \leq k \leq n} \mathbf{x}_{1k}^T (\mathbf{X}_1^T \mathbf{X}_1)^{-1} \mathbf{x}_{1k} \rightarrow 0$ , as  $n \rightarrow \infty$ ; and

(C2)  $\mathbf{V}$  is finite and  $\mathbf{V}_{11} > 0$ ,

where,  $\mathbf{X}_1$  consists of the first  $s$  columns of  $\mathbf{X}$ ,  $\mathbf{V} = \lim_{n \rightarrow \infty} \mathbf{X}^T \mathbf{X} / n$  and  $\mathbf{V}_{11}$  consists of the first  $s$  columns and rows of  $\mathbf{V}$ . It is worth noting that these conditions are typical conditions for asymptotic normality of least squares estimator in linear regression model when the regressors are fixed designed. Furthermore,  $\mathbf{V}$  is finite for most experiment designs. Here we weaken these two conditions to Conditions (C1) and (C2), which guarantee that the asymptotic normality holds for the least squares estimator of  $\hat{\boldsymbol{\beta}}_1$ . Li and Lin [13] also show that the penalized least squares estimate is root  $n$  consistent and establish the oracle property for the penalized least squares estimator. This implies that the resulting penalized least squares is more efficient than the ordinary least squares estimate. Note that the proposed procedure possesses the oracle property under some mild conditions which do not require the design matrix to be of full rank. The conditions are frequently satisfied in practice.

### 4. An Algorithm for the Nonconvex Penalized Least Squares

Note that the SCAD penalty function is singular at the origin and may not have the second derivative at some points. In order to apply the Newton Raphson algorithm to the penalized least squares, Fan and Li [5] suggested to locally approximate the SCAD penalty function by quadratic functions as follows. Given an initial value  $\boldsymbol{\beta}^{(0)}$  that is close to the true value of  $\boldsymbol{\beta}$ , when  $\beta_j^{(0)}$  is not very close to 0, the penalty  $p_\lambda(|\beta_j|)$  can be locally approximated by the quadratic function as

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

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  $\beta_0$ :

$$\beta^{(1)} = \{\mathbf{X}^T \mathbf{X} + n\Sigma_\lambda(\beta^{(0)})\}^{-1} \mathbf{X}^T \mathbf{y}, \quad (3)$$

where

$$\Sigma_\lambda(\beta^{(0)}) = \text{diag}\{p'_\lambda(|\beta_1^{(0)}|)/|\beta_1^{(0)}|, \dots, p'_\lambda(|\beta_d^{(0)}|)/|\beta_d^{(0)}|\}.$$

This can be easily implemented in many statistical packages. Hunter and Li [12] studied the convergence behavior of this algorithm by using techniques related to the EM algorithm.

When the sample size is greater than the number of factors, and the design matrix is full rank, there exists a unique least squares estimate. The least squares estimate can serve as the initial value of  $\beta$ . When the sample size is relatively large, the least squares estimate possesses root  $n$  consistency, and hence it is very close to the true value of  $\beta$ . However, the number of experiments may be less than the number of potential candidates. It is obvious that the design matrix is not full rank, and thus the regression coefficients are not identifiable without further assumptions. In the context of analysis of screening experiments, it is assumed that only a few of factors are active. Here we suggest to use stepwise variable selection to construct an initial value of  $\beta$ . In other words, we first apply stepwise variable selection to the full model with small thresholding values (*i.e.* large value of significance level  $\alpha$ ) such that all active factors are included in the selected model. Of course, in this step, some of insignificant factors will remain in the selected model.

#### 4.1. Standard Error

The standard errors for estimated coefficients can be obtained directly because our procedure simultaneously performs parameter estimation and variable selection. From the iterative ridge regression, the covariance of the penalized least squares estimate can be estimated via

$$\widehat{\text{cov}}(\hat{\beta}) = \hat{\sigma}^2 \{\mathbf{X}^T \mathbf{X} + n\Sigma_\lambda(\beta_0)\}^{-1} \mathbf{X}^T \mathbf{X} \{\mathbf{X}^T \mathbf{X} + n\Sigma_\lambda(\beta_0)\}^{-1}, \quad (4)$$

where  $\hat{\sigma}^2$  is an estimator of  $\sigma^2$ , for example, the mean squared errors. Here the formula applies only to components which do not vanish.

#### 4.2. Selection of Regularization Parameter

To implement the SCAD, it is desirable to have an automatic data-driven method for selecting the tuning parameter  $\lambda$  in  $p_\lambda(\cdot)$ . Such a procedure is analogous to an automatic subset selection procedure, such as forward selection, backward elimination, but with good oracle properties. Cross validation and generalized cross-validation can be employed to select the tuning parameter  $\lambda$ . However, the cross validation may not be appropriate in the current setting when the number of experiments is small. Thus we estimate  $\lambda$  via minimizing an approximate generalized cross-validation (GCV) statistic (Craven and Wahba [3]).

For the penalized least squares, we update the solution by

$$\beta^{(1)}(\lambda) = \{\mathbf{X}^T \mathbf{X} + n\Sigma_\lambda(\beta^{(0)})\}^{-1} \mathbf{X}^T \mathbf{y},$$

with an initial value  $\beta_0$ . Thus the fitted value  $\hat{\mathbf{y}}$  of  $\mathbf{y}$  is  $\mathbf{X}\{\mathbf{X}^T \mathbf{X} + n\Sigma_\lambda(\beta^{(0)})\}^{-1} \mathbf{X}^T \mathbf{y}$  and

$$\mathbf{P}_x \hat{\beta}^{(1)}(\lambda) = \mathbf{X}\{\mathbf{X}^T \mathbf{X} + n\Sigma_\lambda(\beta)\}^{-1} \mathbf{X}^T,$$

can be regarded as a projection matrix. Define the number of effective parameters in the penalized least squares fit as  $e(\lambda) = \text{tr}[\mathbf{P}_x \{\hat{\beta}(\lambda)\}]$ . Therefore the generalized cross-validation statistics is

$$\text{GCV}(\lambda) = \frac{1}{n} \frac{\|\mathbf{y} - \mathbf{X}\hat{\beta}(\lambda)\|^2}{\{1 - e(\lambda)/n\}^2}, \quad (5)$$

and  $\hat{\lambda} = \text{argmin}_\lambda \{\text{GCV}(\lambda)\}$ .

#### 4.3. Implementation of SCAD Variable Selection Procedure

The SCAD variable selection procedure is summarized below, step by step.

- Step 1. If the number of regression coefficients is greater than the sample size, then first apply stepwise variable selection to the full model with small thresholding values.
- Step 2. Choose a grid point set for  $\lambda$ , say,  $\{\lambda_1, \dots, \lambda_L\}$ .
- Step 3. Based on the selected model by the stepwise regression, for each  $\lambda_l$ ,  $l=1, \dots, L$ , find the solution for the penalized least squares with the SCAD penalty by iterating (3) until it converges, and further compute the GCV statistic in (5).
- Step 4. The final estimate for  $\beta$  is the one with the lowest GCV score.

## 5. Examples

**Example 1.** A 12-run PB design and the corresponding outcomes are listed in Table 1, extracted from Hamada and Wu [11]. This data set was constructed by Hamada and Wu [11] based on the linear model  $Y = A + 2AB + 2AC + \varepsilon$  with  $\varepsilon \sim N(0, \sigma = 0.25)$ . We fit the data set by a linear regression model including all linear terms and interaction terms. The full model contains 66 (11 main effects plus 55 two-factor interactions) predictors.

To identify the active effects, stepwise variable selection was employed to the full model using SAS. The variables  $AB$ ,  $AC$ ,  $A$ ,  $AJ$ ,  $GJ$ ,  $FG$  and  $J$  were selected. The selected variables and their estimated coefficients, standard errors and  $p$ -values are depicted in Table 2. From Table 2, all selected variables are statistically significant at significance level  $\alpha = 0.05$ . In fact, compared with the true model, stepwise variable selection results in a very misleading model, though the truly significant variables are included the selected model. From this example, one should be cautious when using stepwise variable selection procedure.

Now we apply penalized least squares with the SCAD penalty for the selected model by stepwise procedure, the estimated tuning parameter  $\hat{\lambda} = 0.0121$ . The resulting model, including estimated coefficients and estimated standard errors, is depicted in Table 2. The square root of the mean square error is  $\hat{\sigma} = 0.2479$ . All estimated coefficients and their

standard errors are very close to their true values. In this case, the proposed method works very well. With small values of F-enter and F-remove, the stepwise variable selection procedure is able to retain active effects, though it also includes many inactive effects. On the other hand, this drawback of the stepwise variable selection can be alleviated by the SCAD, which may correctly identify active effects. Furthermore, the class of models considered here is much larger than the one used by Hamada and Wu [11].

Table 1. PB 12 run design and response in Example 1.

run	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>	<i>F</i>	<i>G</i>	<i>H</i>	<i>I</i>	<i>J</i>	<i>K</i>	response
1	+	+	-	+	+	+	-	-	-	+	-	1.058
2	+	-	+	+	+	-	-	-	+	-	+	1.004
3	-	+	+	+	-	-	-	+	-	+	+	-5.200
4	+	+	+	-	-	-	+	-	+	+	-	5.320
5	+	+	-	-	-	+	-	+	+	-	+	1.022
6	+	-	-	-	+	-	+	+	-	+	+	-2.471
7	-	-	-	+	-	+	+	-	+	+	+	2.809
8	-	-	+	-	+	+	-	+	+	+	-	-1.272
9	-	+	-	+	+	-	+	+	+	-	-	-0.955
10	+	-	+	+	-	+	+	+	-	-	-	0.644
11	-	+	+	-	+	+	+	-	-	-	+	-5.025
12	-	-	-	-	-	-	-	-	-	-	-	3.060

Table 2. The selected models in Example 1.

Factor	Stepwise Regression			SCAD		
	$\hat{\beta}$	$SE(\hat{\beta})$	<i>p</i> -value	$\hat{\beta}$	$SE(\hat{\beta})$	<i>p</i> -value
<i>AB</i>	1.9126	0.0096	<0.0001	2.0000	0.0679	<0.0001
<i>AC</i>	1.9620	0.0092	<0.0001	1.9808	0.0679	<0.0001
<i>A</i>	1.0209	0.0094	<0.0001	1.0967	0.0679	<0.0001
<i>AJ</i>	0.1424	0.0086	<0.0001			
<i>GJ</i>	0.1593	0.0097	<0.0001			
<i>FG</i>	-0.0679	0.0108	0.0015			
<i>J</i>	0.0350	0.0089	0.0112			

**Example 2.** In this example, we demonstrate the SCAD variable selection procedure by using analysis of a data set collected in the study of the effects of environmental pollutants on human health. The data set is taken from Fang [6]. A brief description is given below.

Details can be found in Fang [6]. It is believed that contents of some metal elements in water might directly have impact on human health. Of interest here is to study the association between contents of six selected metals and the mortality of some kind cell of mice. The six selected metals are Cadmium (Cd), Copper (Cu), Zinc (Zn), Nickel (Ni), Chromium (Cr), and Lead (Pb). The levels for the content of each metal element is set to

0.01, 0.05, 0.1, 0.2, 0.4, 0.8, 1, 2, 5, 8, 10, 12, 16, 18 and 20 (ppm). A uniform design (Fang and Wang [8]; and Fang *et al.* [9]) is used to construct a design with 17 experiments. For each run, three experiments were conducted. The average of these three outputs (mortality) is displayed in Table 3.

Table 3. Data for in Example 2.

Run	Cd	Cu	Zn	Ni	Cr	Pb	Response
1	0.01	0.20	0.80	5.00	14.00	16.00	18.5900
2	0.05	2.00	10.00	0.10	8.00	12.00	22.5200
3	0.10	10.00	0.01	12.00	2.00	8.00	32.4667
4	0.20	18.00	1.00	0.80	0.40	4.00	39.2967
5	0.40	0.10	12.00	18.00	0.05	1.00	32.2767
6	0.80	1.00	0.05	4.00	18.00	0.40	30.9933
7	1.00	8.00	2.00	0.05	12.00	0.10	40.0733
8	2.00	16.00	14.00	10.00	5.00	0.01	42.7100
9	4.00	0.05	0.10	0.40	1.00	18.00	24.8900
10	5.00	0.80	4.00	16.00	0.20	14.00	50.6833
11	8.00	5.00	16.00	2.00	0.01	10.00	60.2767
12	10.00	14.00	0.20	0.01	16.00	5.00	68.7067
13	12.00	0.01	5.00	8.00	10.00	2.00	32.4433
14	14.00	0.40	18.00	0.20	4.00	0.80	29.7600
15	16.00	4.00	0.40	14.00	0.80	0.20	68.0533
16	18.00	12.00	8.00	1.00	0.10	0.05	55.7867
17	20.00	20.00	20.00	20.00	20.00	20.00	79.1600

Table 4. The selected models in Example 2.

Factor	Stepwise Regression			SCAD		
	$\hat{\beta}$	$SE(\hat{\beta})$	$p$ -value	$\hat{\beta}$	$SE(\hat{\beta})$	$p$ -value
Intercept	36.8657	0.0761	<0.0001	38.7424	2.4792	<0.0001
Cd	14.7373	0.0406	<0.0001	13.8951	1.2747	<0.0001
Cu	12.8916	0.0301	<0.0001	11.8375	1.2473	<0.0001
Ni*Cr	-10.7835	0.0666	<0.0001	-8.9738	1.3672	0.0002
Zn*Pb	10.9933	0.0727	<0.0001	12.2283	1.3378	<0.0001
Cd*Cd	-5.7473	0.0569	0.0001	-7.7394	1.7256	0.0020
Cr	-7.4304	0.0353	<0.0001	-7.2145	1.3494	0.0007
Pb*Pb	8.4473	0.0723	0.0001	10.8729	1.8728	0.0004
Zn*Cr	-7.1860	0.0746	0.0001	-5.3017	1.6051	0.0108
Pb	4.0711	0.0348	0.0001			
Zn*Zn	2.4276	0.0518	0.0005			
Cu*Cr	-1.1144	0.0555	0.0025			
Ni	1.3451	0.0386	0.0008			
Cd*Ni	1.1082	0.0765	0.0047			
Ni*Pb	0.0790	0.0328	0.1376			



To avoid numerical instability, all  $x$  variables are standardized first. In our initial model, we include an intercept, all linear effects, quadratic effects and the first order interaction effects. This yields a linear regression models with 28 terms, including the intercept. Since there are only 17 design combination in this experiment. Thus, variable selection is necessary. Similar to the previous analysis, the stepwise regression analysis is first performed. The stepwise regression selects Cd, Cu, Ni\*Cr, Zn\*Pb, Cr, Pb\*Pb, Zn\*Cr, Pb, Zn\*Zn, Cu\*Cr, Ni, Cd\*Ni and Ni\*Pb, whose estimate, standard error and  $p$ -value are depicted in Table 4. All of them except Ni\*Pb are significant at level 0.05. Based on the selected model by stepwise regression, we apply the SCAD variable selection procedure to select significant variables. The selected  $\lambda$  is 0.0748. The resulting estimate of significant effects along with their standard errors and  $p$ -values are depicted in Table 4, from which it can be seen that all six metal elements have significant impact on mortality.

## 6. Conclusion and Discussion

In this paper, we have introduced the SCAD variable selection procedure for screening experiments. It has been shown that the proposed approach possesses an oracle property. Two data examples illustrated the effectiveness of the proposed approach. In our earlier work (Li and Lin [13, 14]), the proposed procedure has been successfully applied to analyzing data from supersaturated designs(Lin [15, 16]). The computer codes (written in MATLAB) to perform all calculations can be obtained from the authors.

The regularity conditions (C1) and (C2) in Section 3 ensure that the proposed procedure can be used to screen experiments. Although the design matrix for the full model can be singular, condition (C2) implies that the design sub-matrix responding to active factors is full rank, and therefore alias structure among the active factors is not allowed. One should be cautious in the situation that active factors or their interactions are aliased with another ones. It needs some further research on how to deal with such situations.

## Acknowledgement

The authors would like to thank referees for their valuable comments and suggestions. Li was supported by a National Institute on Drug Abuse (NIDA) grant P50 DA10075, and partially supported by NSF grants DMS 0348869 and CCF 0430349. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIDA or the National Institutes of Health.

## References

1. Antoniadis, A. and Fan, J. (2001). Regularization of wavelets approximations (with discussions). *Journal of American Statistical Association*, 96, 939-967.
2. Beattie, S. D., Fong, D. K. H. and Lin, D. K. J. (2002). A two-stage Bayesian model selection strategy for supersaturated designs. *Technometrics*, 44, 55-63.
3. 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. *Numerical Mathematics*, 31, 377-403.
4. Donoho, D. L. and Johnstone, I. M. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81, 425-455.
5. Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of American Statistical Association*, 96, 1348-1360.
6. Fang, K. T. (1994). *Uniform Design and Uniform Design Tables*. Science Press, Beijing.
7. Fang, K.T. and Lin, D. K. J. (2003). *Uniform experimental design and its applications*

- in industry. *Handbook of Statistics in Industry*, eds, C.R. Rao and R. Khattree. North Holland: New York, 22, 131-170.
8. Fang, K. T. and Wang, Y. (1994). *Applications of Number Theoretic Methods in Statistics*, Chapman and Hall, London.
  9. Fang, K. T., Lin, D. K. J., Winker, Peter and Zhang, Y. (2000). Uniform design: theory and application. *Technometrics*, 42, 237-248.
  10. Frank, I. E. and Friedman, J. H. (1993). A statistical view of some chemometrics regression tools. *Technometrics*, 35, 109-148.
  11. Hamada, M. and Wu, C. F. J. (1992). Analysis of designed experiments with complex aliasing. *Journal of Quality Technology*, 24, 130-137.
  12. Hunter, D. R. and Li, R. (2005). Variable selection using MM algorithms. *Annals of Statistics*, 33, 1617--1642.
  13. Li, R. and Lin, D. K. J. (2002). Data analysis in supersaturated designs, *Statistics and Probability Letters*, 59, 135-144.
  14. Li, R. and Lin, D. K. J. (2003). Comparisons of variable selection approaches for analysis of supersaturated design. *Journal of Data Science*, 1, 249-260.
  15. Lin, D. K. J. (1993). A new class of supersaturated designs. *Technometrics*, 35, 28-31.
  16. Lin, D. K. J. (1995). Generating system supersaturated designs. *Technometrics*, 37, 213-225
  17. Tibshirani, R. J. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B*, 58, 267-288.
  18. Westfall, P. H., Young, S. S. and Lin, D. K. J. (1998). Forward selection error control in analysis of supersaturated designs. *Statistica Sinica*, 8, 101-117.

*Authors' Biographies:*

**Runze Li** is a Professor of Statistics, The Pennsylvania State University. His research interests include variable selection, non- and semi-parametric regression modeling, nonparametric smoothing, functional and longitudinal data analysis, survival data analysis and modeling computer experiment data. He currently serves as associate editor for *Annals of Statistics*, *Journal of American Statistical Association* and *Statistica Sinica*. He is a life-time member of International Chinese Statistical Association, a permanent member of Institute of Mathematical Statistics, a member of American Statistical Association, and a member of Bernoulli society.-

**Dennis K. J. Lin** is a University Distinguished Professor of Supply Chain Management and Statistics at Penn State University. His research interests are quality assurance, industrial statistics, data mining and response surface. He has published over 100 papers in a wide variety of journals. He currently serves as associate editor for *Technometrics*, *Statistica Sinica*, *Journal of Data Science*, *Quality Technology & Quality Management*, *Journal of Quality Technology*; and *Taiwan Outlook*. He is editor of *Applied Stochastic Models for Business and Industry* from 2008. Dr. Lin is an elected fellow of the American Statistical Association (ASA), an elected member of International Statistical Institute (ISI), an elected fellow of American Society of Quality (ASQ), a lifetime member of International Chinese Statistical Association (ICSA), a fellow of the Royal Statistical Society (RSS), and has received the Most Outstanding Presentation Award from SPES, ASA. He is an honorary chair professor for various universities, including National Chengchi University (Taiwan), Renmin University of China, Fudan University and XiAn Statistical Institute. He is also the recipient of the 2004 Faculty Scholar Medal Award at Penn State University.