Model Averaging Method for Supersaturated Experimental Design

Deiby T. Salaki[1] , Anang Kurnia[2], Bagus Sartono[3]

[1]Department of Mathematics, Sam Ratulangi University, Manado, Indonesia
Graduate School, Bogor Agricultural University, Indonesia
[2,3]Department of Statistics, Bogor Agricultural University, Indonesia

E-mail: deibyts17@gmail.com

Abstract. In this paper, a new modified model averaging method was proposed. The candidate model construction was performed by distinguishing the covariates into focus variables and auxiliary variables whereas the weights selection was implemented using Mallows criterion. In addition, the illustration result shows that the applied model averaging method could be considered as a new alternative method for supersaturated experimental design as a typical form of high dimensional data. A supersaturated factorial design is an experimental series in which the number of factors exceeds the number of runs, so its size is not enough to estimate all the main effect. By using the model averaging method, the estimation or prediction power is significantly enhanced. In our illustration, the main factors are regarded as focus variables in order to give more attention to them whereas the lesser factors are regarded as auxiliary variables, which is along with the hierarchical ordering principle in experimental research. The limited empirical study shows that this method produces good prediction.

1. Introduction

Nowadays, the necessity of high dimensional data is increasing in all fields of study and has wide applications. Fortunately, the rapid growth in information technology supports the development of high dimensional data analysis. According to [1], high dimensional statistics refer to statistical inference when the number of unknown parameter is much larger order the sample size. Nevertheless, analysis of such data is impossible to be performed without additional assumption or penalization.

A typical form of high dimensional data stemmed from experimental data is supersaturated factorial design. A supersaturated factorial design is an experimental series in which the number of factors exceeds the number of runs. The supersaturated design (SSD) is recommended in high cost experiments and in situation where the number of effects is large, to screen few of a huge number of potentially active factors might be believed to be significant. There were some designs initiated to SSD. According to [2], Satterthwaite (1959), the initiator of the SSD, considered SSD to be useful in the initial investigation of large experiments and also proposed a random balance design for constructing SSD. Another construction method suggested by [3], namely $E(s^2)$ criterion to measure the efficiency of SSD. Hadamard matrix was a well-known design proposed by [2] in constructing SSD. Several authors have proposed various strategies in solving the challenges in SSD data analysis using some regression procedures, such as forward selection by [4], stepwise and all-subsets regression by [5] and partial least squares by [6]. There are also shrinkage regression approaches such as Smoothly Clipped Absolute Deviation procedure [7] and the Dantzig selector [8].

Many physical experiments can be high costly and time consuming to be studied using conventional method. In computer experiments or computer-based simulation, for example, the experiments can only be simulated by a computer modelling program in which the number of factor can be large as possible. Consequently, the running time is the mostly thing to be considered instead and the goal is turning to how to produce a best prediction model at any combination of values of the inputs that fitted the data [9]. Such concerns are in accordance with the motivation of model averaging as an alternative to variable selection or screening in SSD point of view.

Model averaging (MA) is another approach to improve prediction accuracy in analysing high dimensional data. Instead of relying on only one best model, the method refers both inference and prediction to the average over a set potential models in particular manner. The key issues of the methods are for example, weight selection criterion and the way of construction the candidate model [10]. Getting the best prediction model is preferred goal instead of estimation. Application of MA in SSD, have not received many attention yet. So far, the major focus of SSD analysis has been on screening or variable selection. MA approach in SSD, based on literature study, had been conducted by [11]. The paper proposed model averaging method by using BIC weight. Another weight used in model averaging namely mallows averaging which is refer to least square model averaging proposed by Hansen [12]. In his paper, Hansen performed different approach in preparing candidate models that is by constructing nested models while the weights is selected by minimizing Mallows criterion.

The current paper introduced a new typical MA approach which is tailored to candidate models preparation. Here modification, the main factors of matrix model are categorized as focus variables that must be included in the model and their interactions that are less certain as auxiliary variables. The approach went along with effect hierarchical principle to be considered to address the relative importance and relationship among effects. The principle set the lower order effects as more important factors than the higher level effects while the same order consider to be equally likely important [4]. Henceforth, the candidate model is constructed by augmenting the matrix of focus variable with the different combination of one or several auxiliary variables. Besides reducing the number of candidate model, the reason of this differentiation is to improve the prediction performance of the model [13]. In weights selection, Mallows criterion is applied which is proven asymptotically optimal in the sense of achieving the lowest possible squared error [12].

This paper is organized as follows. In Section 2, the supersaturated experimental design is described. Section 3 reviews the intended model averaging. In the Section 4, an illustration of MA approach in SSD is performed to show how well the prediction matched the actual model. Section 6 gives some concluding remarks

2. Supersaturated Design

Supersaturated design is fractional factorial design in which the number of actual effects exceeded the number of runs. The SSD is divided into two classes, multi-level designs with *s* level on each factor and mixed-level in which each factor has different level. The current paper was concerned with multi levels of linear main effects model. Supposed the model intended is

$$\boldsymbol{y} = \mathbf{X}\,\boldsymbol{\theta} + \boldsymbol{\varepsilon} \tag{2.1}$$

where $\boldsymbol{y}$ is the $n \times 1$ response vector, $\mathbf{X}$ is an $n \times (m+1)$ model matrix, vector of $\boldsymbol{\theta} = (\theta_0, \dots, \theta_m)^T$ constitutes the parameter and $\boldsymbol{\varepsilon}$ is a vector of independent normally distributed random errors with mean 0 and variance σ^2. The matrix $\mathbf{X}$ can be a union of main effects and their interactions. It is assumed that each of the m factors has two levels, contrast between high and low denoted by 0 and 1.

If the model matrix $\mathbf{X}$ is non-singular then the matrix $(\mathbf{X}'\mathbf{X})^{-1}$ exists and the least squares estimator of parameter vector $\boldsymbol{\theta}$ is denoted by $\widehat{\boldsymbol{\theta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{y}$. According to [14], if the matrix model is away from orthogonality and presence of multicollinearity then detection of the true active factors is hardly to be performed.

3. Model Averaging

Model averaging is considered the alternative to model selection, especially in estimation of model with high dimensional data. Model selection will choose one of a wide set of competed models based on certain selection criterion with possibly contradictory conclusion such as R^2, BIC and AIC. On the other hand, model averaging treats each potential model as an information contributor about the parameters of interest. The contribution is expressed in a weight assigned to the candidate model which is determined based on certain criterion. The goal of model averaging is not to find the best possible model but rather to get the best possible prediction model.

The first step in model averaging is to construct the model candidate. After getting estimates of each model, model averaging assigning weight to each of them and finally, takes the weighted averaging of all parameter vectors to obtain the model averaging estimator.

3.1. Framework of MA

Consider the n observations $\{(y_i, x_i); i = 1, \dots, n\}$, where y_i is response variable and $\boldsymbol{x}_i$ is a vector of m-dimensional explanatory variables where the number of predictor is allowed to increase with the sample size n. The relationship between the response variable and the explanatory variables will be considered from observed data. The relationship is explained in a multiple linear regression model, can be expressed as

$$\boldsymbol{y} = \sum_{j=1}^{m} \theta_j \boldsymbol{x}_j + \boldsymbol{\varepsilon} \tag{2.2}$$

where $E(\varepsilon) = 0$ and $Var(\varepsilon) = \sigma^2$. The intercept term is dropped by assuming that the means of the response variable and the input variable are already subtracted out.

Claimed M candidate models and the ith model can be written as follow

$$M_i: \ \boldsymbol{y}_i = \sum_{j \epsilon T_i} \theta_j \boldsymbol{x}_j + \boldsymbol{\varepsilon} \tag{2.3}$$

Where T_i was subset of explanatory variables contained in ith model. In matrix form, equation (2.2) can be written as equation (2.1) and equation (2.3) as below

$$M_i: \ \boldsymbol{y}_i = \mathbf{X}_i \boldsymbol{\theta}^i + \boldsymbol{\varepsilon} \tag{2.4}$$

Where $\boldsymbol{\theta}^i = [\theta_j]_{j \in T_i}$ is the vector of parameter $n \times s; s \leq m$

In model selection methods, according to certain criterion, it will be chosen a single best model:

$$\widehat{\boldsymbol{y}} = \widehat{\boldsymbol{\mu}} = \ \textstyle\sum_{j \in T_i} \hat{\theta}_j \boldsymbol{x}_j.$$

Recall the model M_i in equation (2.4), its prediction can be written as

$$\widehat{\boldsymbol{\mu}}_i = \ \mathbf{X}_i \widehat{\boldsymbol{\theta}}^i$$

Supposed w_i is the weight associated with ith model and let $\boldsymbol{w} = (w_1, \dots, w_M)'$ be the vector in unit simplex in $\mathbb{R}^M$

$$H_n = \{\boldsymbol{w} \in [0,1]^M \}$$

then *model averaging* estimator for equation (2.1) takes the form

$$\widehat{\boldsymbol{\mu}}_{MA} = \sum_{i=1}^{M} w_i \widehat{\boldsymbol{\mu}}_i \tag{2.5}$$

Based on the previously framework, the steps of MA analysis is itemized below:

1. Construct several candidate models M_i; $i = 1, \dots, M$, as described in equation (2.4).
2. Estimate each model M_i
3. Determine weights based on certain criterion
4. Average out weighted prediction of those models, as in equation (2.5)

3.2. Model and Estimation

Supposed the linear regression model of model in equation (2.1) is mentioned in the framework of [13] as below

$$\boldsymbol{y} = \mathbf{X}\boldsymbol{\theta} + \boldsymbol{\varepsilon} = \ \mathbf{X}_1 \boldsymbol{\beta}_1 + \mathbf{X}_2 \boldsymbol{\beta}_2 + \ \boldsymbol{\varepsilon} \tag{2.6}$$

The matrix of explanatory variables is $\mathbf{X} = (\mathbf{X}_1 | \mathbf{X}_2)$; $\mathbf{X}_1$ $(n \times q)$ is matrix of focus variables refer to intercept and main factor; $\mathbf{X}_2$ $(n \times r)$ is matrix of auxiliary variables refer to their interaction. Thus,

vector of parameter consists of $\boldsymbol{\beta_1} = (\theta_0, \theta_1, \theta_2, \dots, \theta_q)'$ and $\boldsymbol{\beta_2} = (\theta_{q+1}, \theta_{q+2}, \dots, \theta_r)'$. It is also assumed that $q \geq 1$; $r \geq 0$, $+r \leq n-1$; that is $\mathbf{X} = (\mathbf{X}_1 | \mathbf{X}_2)$ has full column-rank. According to Magnus [13] the model in equation (2.6) is more general than the conventional case where typically all variables are auxiliary except the constant term. Such form also allows some variables to be included in the model even when diagnostic tests suggest it to be removed or permitted ones as must contained in the model. Moreover, such classification could reduce the number of candidate model.

The least squares estimator of $\boldsymbol{\theta}$ for a full model where all the auxiliary variables are included in model is

$$\widehat{\boldsymbol{\theta}} = \begin{pmatrix} \widehat{\boldsymbol{\beta}}_1 \\ \widehat{\boldsymbol{\beta}}_2 \end{pmatrix} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{y}$$

The set of candidate models consist of different models arise when a different subset of the $\boldsymbol{\beta}_2$, is set to zero in other words each model matrix of candidate models consist of focus variables incorporate with combination of auxiliary variables so there are totally M=2^r candidate models to be consider.

Consider $S_i = (r_i \times r)$ a selection matrix to choose the included auxiliary variables, where r_i is the number of auxiliary variables contained in M_i. So, the estimators for the candidate model M_i, is

$$\widehat{\boldsymbol{\theta}}_i = \left(A_{M_i}{}' A_{M_i}\right)^{-1} A'_{M_i}\, \boldsymbol{y} \tag{2.7}$$

where $A_{M_i} = (\mathbf{X}_1 | \mathbf{X}_2 S_i')$. For a full model least square estimator, $\widehat{\boldsymbol{\theta}}_i = \widehat{\boldsymbol{\theta}}$, with $S_i = I_r$, an identity matrix with r dimension. On the contrary, $S_i = \mathbf{0}$ is for the least square estimator of the smallest model among all possible candidate models, that was the model with $\mathbf{X}_1$ as the design matrix. General form of the selection matrix is $S_i = (I_{r_i} | 0_{r-ri})$ and its column combination.

Suppose $\boldsymbol{w} = (w_1, \dots, w_M)'$ is a weight vector with $w_i \geq 0$ and satisfied $\sum_1^M w_i = 1$, that is the weight vector belongs to the unit simplex in $\mathbb{R}^M$:

$$\mathcal{H}_M = \left\{ \boldsymbol{w} \in [0,1]^M ; \sum_1^M w_i = 1 \right\}$$

then the averaging estimator of $\boldsymbol{\theta}$ is

$$\widehat{\boldsymbol{\theta}}_{MA} = \sum_1^M w_i \widehat{\boldsymbol{\theta}}_i$$

If the corresponding prediction candidate model is

$$\widehat{\boldsymbol{\mu}}_i = A_{M_i}\widehat{\boldsymbol{\theta}}_i = A_{M_i}\left(A_{M_i}{}' A_{M_i}\right)^{-1} A'_{M_i}\, \boldsymbol{y} = P_{M_i}\boldsymbol{y}$$

where

$$P_{M_i} = A_{M_i}\left(A_{M_i}{}' A_{M_i}\right)^{-1} A'_{M_i}$$

then model averaging prediction of equation (2.6) can be written as

$$\widehat{\boldsymbol{\mu}}_{\boldsymbol{MA}} = \sum_1^M w_i \widehat{\boldsymbol{\mu}}_i = \sum_1^M w_i P_{M_i}\boldsymbol{y} = P(\boldsymbol{w})\boldsymbol{y} \tag{2.8}$$

3.3. Mallow Weight Criterion

There are various methods in determining the weights of model averaging. One of them was Mallow averaging proposed by Hansen [12] in which the selection weights is performed by minimizing a Mallows criterion. Such criterion for the averaging estimator is

$$C(\boldsymbol{w}) = \hat{\varepsilon}(\boldsymbol{w})'\hat{\varepsilon}(\boldsymbol{w}) + 2\sigma^2 tr(P(\boldsymbol{w}))$$

where

$$\hat{\varepsilon}(\boldsymbol{w}) = \boldsymbol{y} - A\,\widehat{\boldsymbol{\theta}}_{MA},$$

If $tr(P(\boldsymbol{w})$ is rewritten as $\boldsymbol{K} = (k_1, \dots, k_M)'$; k_i is the number of parameters in ith model, then

$$\boldsymbol{C}(\boldsymbol{w}) = \boldsymbol{w}'\hat{\boldsymbol{\varepsilon}}'\boldsymbol{w} + 2\sigma^2\boldsymbol{w}'\boldsymbol{K} \tag{2.9}$$

which is a quadratic form in the vector $\boldsymbol{w}$.

By minimizing the criterion in (2.9), the weight vector is

$$\widehat{\boldsymbol{w}} = \underset{\boldsymbol{w} \in \mathcal{H}_M}{\operatorname{argmin}} \boldsymbol{C}(\boldsymbol{w}) \tag{2.10}$$

4. Illustration of MA Application in Supersaturated Designs

In order to apply model averaging method in SSD, an illustration will be presented based on [6] on pg. 443. Its goal was to develop an epoxide adhesive system for bonding a polyester cord. The experiment involved 23 factors with 14 runs or experiments. The response is strip adhesion, which needs to be maximized. After a screening analysis using stepwise regression, there are 6 significant main effects to be considered. Those effects are 4,10,12,15 and 19. A further analysis was planned by involving also two-factor interaction effects. As there are 5, hereafter denoted by A,B,C,D and E respectively, the analysis would consist of 5 main effects and 10 interaction effects with an additional parameter of intercept. Thus, there were as many as 16 parameters to be estimated with only 14 observations that was a 14 × 16 design matrix. It was a typical problem of high dimensional data, typically, supersaturated design experiment that reasonable to be solved using model averaging analysis.

Based on those resulted factors, the further analysis conducted by using model averaging with the follow steps.

1. Define the model by setting the main effects A, B, C, D and E as focus variables and the interaction effects, namely AB, AC,…, DE as auxiliary variables as was displayed in Table 1. Supposed the intercept and all factors are denoted by $\boldsymbol{x}_0, \boldsymbol{x}_1, \dots, \boldsymbol{x}_{15}$ respectively, the model could be written as

$$\boldsymbol{y} = \beta_0 + \sum_{i=1}^{5} \beta_i \boldsymbol{x}_i + \sum_{i=6}^{15} \beta_i \boldsymbol{x}_i + \boldsymbol{\varepsilon} \tag{2.11}$$

2. Prepare the candidate models based on the factors in step 1, each of which is composed of intercept and focus variables appended with combination of auxiliary variables sequentially. Thus, the design matrix of the first group of candidate model consists of model with focus variables only because $A_{M_0} = (\mathbf{X}_1 | \mathbf{X}_2 S_0') = X_1$ where $S_0' = \mathbf{0}$. The second group consists of model with focus variables and one of auxiliary variables. In this case, the form of selection matrix was $S_i = (I_1 | 0_{10-1}) = (I_1 | \mathbf{0_9})$ and its column combination, those are $(1\,0\,0\,0\,0\,0\,0\,0\,0\,0), (0\,1\,0\,0\,0\,0\,0\,0\,0\,0), \dots, (0\,0\,0\,0\,0\,0\,0\,0\,0\,1)$ respectively. The next group is contained model with focus variables and two auxiliary with the form of selection matrix was $S_i = (I_2 | \mathbf{0_8}) = \left(\begin{array}{cc|cccccccc} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right)$ and its column combination, for example $\left(\begin{array}{cccccccccc} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right)$ and $\left(\begin{array}{cccccccccc} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \end{array}\right)$. So do the next groups.

3. This step was performed using R program. The package `optimx` used to solve quadratic programming in order to get the vector weight proposed in equation (2.10). To estimate the candidate models, least squares method is applied. The predictive ability of model averaging prediction is validated using leave-one out-cross validation (LOOCV). According to [15], LOOCV is almost unbiased but has high variance, however, recommended to be used where the sample data is limited. It is a typical case of *k*-fold CV with *k* = 1 in which each of 14 iteration nearly all the data except for a single observation are used for training and the model is tested on the single observation. The MAPE used to measure mean of the absolute value of the percent errors, that is $\text{MAPE} = \frac{1}{n}\left(\sum_1^n \frac{actual - predicted}{actual}\right) 100\%$

Table 1. Supersaturated design matrix with 5 main factors and their interactions

Run	Adhesion	A	B	C	D	E	AB	AC	AD	AE	BC	BD	BE	CD	CE	DE
1	133	0	1	0	0	1	1	0	0	1	1	1	0	0	1	1
2	62	0	0	0	1	0	0	0	1	0	0	1	0	1	0	1
3	45	1	1	1	1	0	0	0	0	1	0	0	1	0	1	1
4	52	1	1	0	1	1	0	1	0	0	1	0	0	1	1	0
5	56	1	0	0	1	1	1	1	0	0	0	1	1	1	1	0
6	47	1	1	0	1	1	0	1	0	0	1	0	0	1	1	0
7	88	0	1	1	0	1	1	1	0	1	0	1	0	1	0	1
8	230	0	1	0	0	0	1	0	0	0	1	1	1	0	0	0
9	32	0	0	1	1	0	0	1	1	0	1	1	0	0	1	1
10	53	1	0	1	1	0	1	0	0	1	1	1	0	0	1	1
11	276	1	0	0	0	0	1	1	1	1	0	0	0	0	0	0
12	145	0	1	1	0	0	1	1	0	0	0	1	1	1	1	0
13	130	1	0	1	0	1	1	0	1	0	1	0	1	1	0	1
14	127	0	0	1	0	1	0	1	0	1	1	0	1	1	0	1

The value of actual and prediction are described in Figure 1 and Figure2

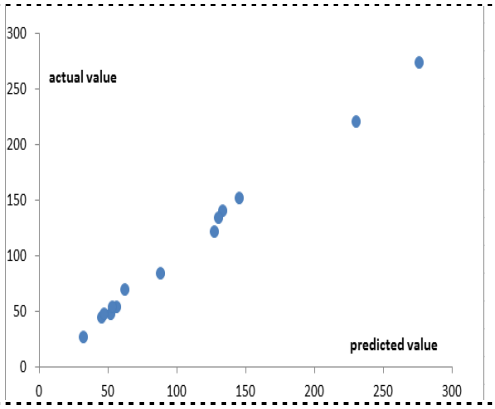

Figure 1. Plot of predicted against actual values

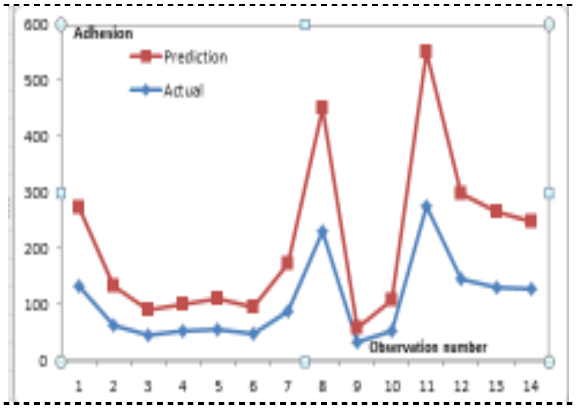

Figure 2. Plot of actual and predicted values

Refer to Figure 1, the value of actual against prediction is scattered closely to the absolute line. The value of both actual and prediction are shown in Figure 2. Their value is relatively having the same trend. As one of best prediction performance criterion, the mean absolute percent error (MAPE) is equal to 91%. The value of R^2 is equal to 0.99. It designates a high performance result of fitting because performed in collaboration with cross validation that is able to avoid overfitting. Thus, the value of R^2 could indicate a good prediction.

5. Conclusion

This study concludes that model averaging method could be applied in supersaturated experimental design as a typical of high dimensional data. That method was a modification of model averaging in which the candidate model construction is performed by distinguishing the matrix model into focus variables and auxiliary variables whereas the weighted selection is using Mallows criterion. In the case of experimental data set, main effects are regarded as focus variables, while the interactions as auxiliary variables. This differentiation follows the hierarchical ordering principle in experimental researches to address the relative importance among effects.

Based on our limited empirical study as was illustrated above, this method results a prediction model with good performance according to the value of MAPE and R^2. However, an advance study is needed to assess its performance in handling multilevel and mixed level experiments. Comparison with other type of model averaging is also highly recommended to be done in order to evaluate whether the modified model averaging method is better than others. Moreover, the further study in concern with criterions of estimator such as unbiasedness, efficiency and consistency would be a worth complement task.

References

[1] Buhlmann P and Van de Geer S 2011 *Statistics for High Dimensional Data* (Berlin: Spinger Verlag)
[2] Wu C F J and Hamada M S 2009 *Experiments Planning, Analysis, and Optimization* 2nd ed (NJ: John Wiley & Sons)
[3] Booth K H V and Cox D R 1962 Some Systematic Supersaturated Experimental Designs. *Technometrics* 4: 489-495
[4] Wu C F J 1993 Constructing of Supersaturated Designs Through Partially Aliased Interactions *Biometrika* 80:661-669
[5] Abraham B, Chipman H and Vijayan K 1999 Some Risks in the Construction and Analysis of Supersaturated Designs *Technometrics* 41(2):135-141
[6] Zhang Q Z, Zhang RC and Liu M Q 2007 A method for Screening Acrive Effect in Supersaturated Designs. *Journal of Statistical Planning and Inference* 137(9):235-245
[7] Li R and Lin D K J 2002 Data Analysis in Supersaturated Designs *Statistics & Probability*

Letters 59: 135-144
[8] Phoa F K H, Pan YH and Xu H. 2009 Analysis of Supersaturated Designs via Dantzig Selector. *Journal of Statistical Planning and Inference* 139(7):2362-2372
[9] Minqian L, Kaitai F 2006 A Case Study in The Application of Supersaturated Designs to Computer Experiment. *Acta Mathematica Scientia* 26B(4): 595-602
[10] Ando T and Li K C 2014 A Model Averaging Approach for High Dimensional Regression. *Journal of the American Statistical Association* 505(109): 254-265
[11] Marley, C J and Woods, D C 2010 A Comparison of Design and Model Selection Methods for Supersaturated Experiments *Journal of Computational Statistics & Data Analysis* 54(12): 3158-3167
[12] Hansen B E 2007 Least Squares Model Averaging. *Econometrica* 75:1175-1189
[13] Magnus J R, Powell O R and Schneck P 2010 A Comparison of Two Averaging Techniques with an Application to Growth Empirics. *Journal of Econometrics* 154: 139-153
[14] Georgiou S D 2014 Supersaturated designs: A review of their construction and analysis. *Journal of Statistical Planning and Inference* 144: 92-109
[15] Efron B 1983 Estimating the error rate of a prediction rule: improvement on cross-validation. *Journal American Statistics Association* 78:316-331