



## Comment: Model Selection With Strong and Weak Heredity Constraints

William Li & Ji Zhu

To cite this article: William Li & Ji Zhu (2014) Comment: Model Selection With Strong and Weak Heredity Constraints, *Technometrics*, 56:1, 21-22, DOI: [10.1080/00401706.2013.860921](https://doi.org/10.1080/00401706.2013.860921)

To link to this article: <http://dx.doi.org/10.1080/00401706.2013.860921>



Published online: 20 Feb 2014.



Submit your article to this journal [↗](#)



Article views: 221



View related articles [↗](#)



View Crossmark data [↗](#)

# Comment: Model Selection With Strong and Weak Heredity Constraints

William Li

Supply Chain and Operations Department  
University of Minnesota  
Minneapolis, MN 55455  
(wli@umn.edu)

Ji Zhu

Department of Statistics  
University of Michigan  
Ann Arbor, MI 55313  
(jizhu@umich.edu)

## 1. INTRODUCTION

We congratulate the authors for an excellent article on an important topic. Screening designs have been an active topic in the literature in recent years. However, the work in this field has been heavy on designs and light on analysis methodologies. This is a timely article that provides a nice summary on design analysis.

The article compared several important analysis methodologies including LASSO, Dantzig selector, and SCAD, which are shown to perform well on analyzing designed experiments. A fundamental assumption behind most of the methods is *effect sparsity*, which holds in many experimental design data. Indeed, effect sparsity is considered to be one of the three principles in design of experiments (DOE; Wu and Hamada 2009). The other two principles are *hierarchy* and *heredity*. According to effect hierarchy, lower-order effects are more likely to be significant than higher-order effects, and effects of the same order are equally important. For models following the (strong) heredity principle, if an interaction term  $x_1x_2$  is in the model, then both  $x_1$  and  $x_2$  should also be in the model. Additional methods have been proposed to address such models, which are shown to be more effective than, say, LASSO. Choi, Li, and Zhu (2010) proposed a variable selection method that is particularly tailored to models following heredity, called the strong heredity interaction model (SHIM). They used simulated data including a DOE example to demonstrate that SHIM generally outperforms other competing methods including LASSO and SCAD.

The idea of SHIM can be easily extended to models following weak heredity, in which if an interaction term  $x_1x_2$  is in the model, then either  $x_1$  or  $x_2$  should also be in the model. In the following, we propose a method for the weak heredity interaction model (WHIM). We then discuss the performance of both SHIM and WHIM for analyzing DOE data.

## 2. WHIM METHOD AND NUMERICAL RESULTS

Consider the usual linear model with  $p$  main effects and their two-factor interactions:

$$y = \beta_0 + \beta_1x_1 + \cdots + \beta_px_p + \alpha_{12}x_1x_2 + \cdots + \alpha_{p-1,p}x_{p-1}x_p + \epsilon. \quad (1)$$

Following Choi, Li, and Zhu (2010), we reparameterize the coefficients for the interaction terms  $\alpha_{jj'}$ ,  $j < j'$ ,  $j, j' =$

$1, \dots, p$  as  $\alpha_{jj'} = \gamma_{jj'}(|\beta_j| + |\beta_{j'}|)$ . Then for the purpose of variable selection, we consider the following penalized least-squares criterion:

$$\min \sum_{i=1}^n (y_i - \beta_0 - \sum_j \beta_j x_{ij} - \sum_{j < j'} \alpha_{jj'} x_{ij} x_{ij'})^2 + \lambda_\beta (|\beta_1| + \cdots + |\beta_p|) + \lambda_\gamma (|\gamma_{12}| + \cdots + |\gamma_{p-1,p}|). \quad (2)$$

The first tuning parameter  $\lambda_\beta$  controls the estimates at the main effect level, and the second tuning parameter  $\lambda_\gamma$  controls the estimates at the interaction level. Note that if the coefficient for  $x_jx_{j'}$  is not equal to zero, Equation (2) implies that at least one of  $\beta_j$  and  $\beta_{j'}$  is not equal to zero. Thus, the WHIM formulation (Equation (2)) is particularly suitable for models following weak heredity. The main difference between SHIM of Choi, Li, and Zhu (2010) and WHIM is that in the former method, the coefficients for the interaction terms are reparameterized as  $\alpha_{jj'} = \gamma_{jj'}\beta_j\beta_{j'}$ , such that if the coefficient for  $x_jx_{j'}$  is not equal to zero, it implies that both  $\beta_j$  and  $\beta_{j'}$  are not equal to zero.

The WHIM constrained optimization of Equation (2) can be analyzed in a similar way as the method described in Choi, Li, and Zhu (2010). In what follows, we present several simulation studies. We first evaluate the accuracy of predictions of the three methods: SHIM, WHIM, and LASSO, using a similar simulation study used in Choi, Li, and Zhu (2010). We consider a model with five variables  $x_1, \dots, x_5$ . The true model follows weak heredity:  $y = 2x_2 + x_3 + 2x_1x_3 + \epsilon$ . Each of the five predictors is normally distributed with mean of zero and variance of one. For each simulation, 40 runs are used, and the coefficients were obtained by the three methods. We then compute the MSE that is defined as the sum of squared differences between the fitted coefficients and true coefficients for all effects. Figure 1 compares three boxplots of the MSE values resulting from 1,000 simulations. It can be seen that WHIM clearly outperforms LASSO. The performance of SHIM is also significantly better than LASSO, even if the former was developed for models following strong heredity.

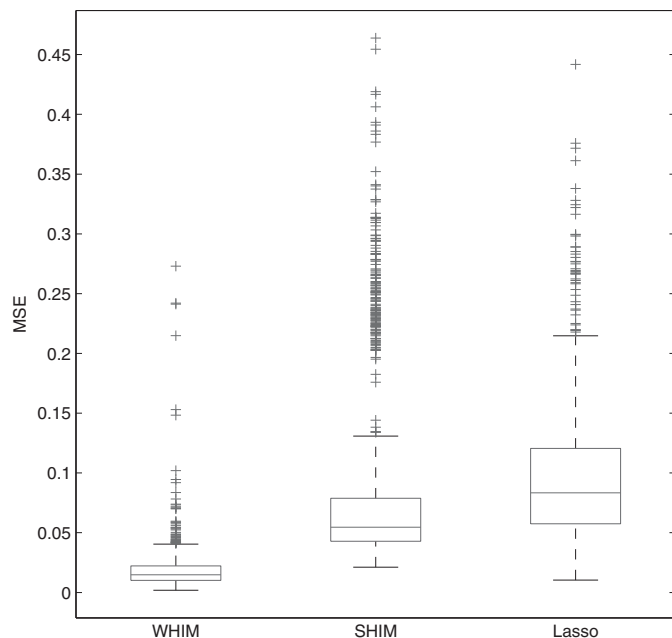


Figure 1. Simulation results: the boxplots show the distribution of mean squared differences between fitted beta and true beta values for all effects of the model.

We now compare the variable selection performance of WHIM with both LASSO and SHIM for DOE data. We consider a simple  $2^{5-1}$  design that was defined by  $x_5 = x_1x_2x_3x_4$ . For each of the three cases considered in Table 1, we generate 100 simulations. We select the tuning parameter values using BIC, and then compute both sensitivity and specificity values, which are defined in the usual way:

$$\text{sensitivity} = \frac{\text{number of correctly selected terms}}{\text{total number of relevant terms}},$$

$$\text{specificity} = \frac{\text{number of correctly unselected terms}}{\text{total number of irrelevant terms}}.$$

For the three cases considered in Table 1, models for Cases 1 and 2 follow weak heredity but not strong heredity. The last model for case 3 includes all two-factor interactions of active factors  $x_1$ ,  $x_2$ , and  $x_3$ . An ideal method should have large values for both sensitivity and specificity. For the simulation cases considered in Table 1, the three methods result in almost iden-

Table 1. Coefficients of true models for the  $2^{5-1}$  design case

	$x_1$	$x_2$	$x_3$	$x_1x_2$	$x_1x_3$	$x_2x_3$
Case 1	7	2	0	0	7	0
Case 2	7	2	0	0	1	0
Case 3	7	2	1	7	7	7

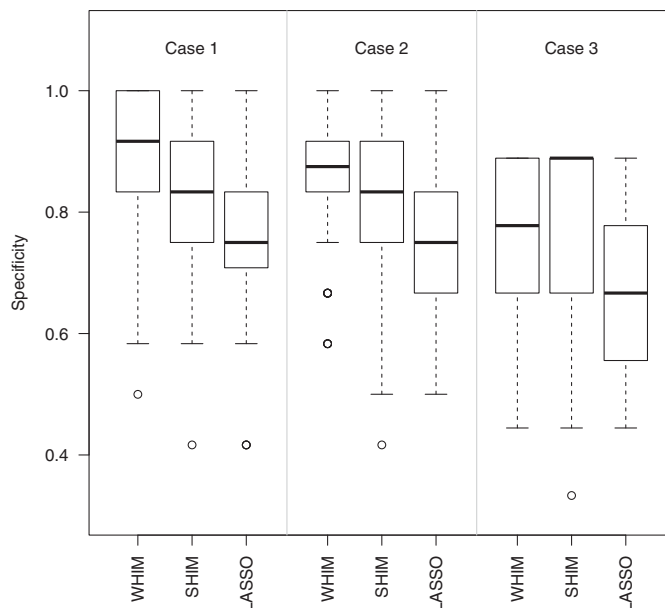


Figure 2. Boxplots of specificity values for Cases 1–3.

tical sensitivity values. For Cases 1 and 2, they all showed nearly 100% sensitivity, which is not surprising given the relative small number of active effects. For Case 3, their average sensitivity values are all equal to 0.833. Consequently, we focus on comparisons of specificity values of the models chosen by the three methods. Figure 2(a) shows a boxplot of the distribution of specificity values of the three methods for Case 1. With a model that follows weak heredity and a large coefficient of the interaction term of  $x_1x_3$ , WHIM clearly outperforms SHIM and LASSO. The mean specificity of WHIM is 0.87, compared with specificity values of 0.82 and 0.77 for SHIM and LASSO, respectively. Figure 2(b) demonstrates a similar pattern for Case 2, where the model is similar to Case 1 except a smaller coefficient for the interaction term. For Case 3 where strong heredity holds, Figure 2(c) shows that both SHIM and WHIM outperform LASSO.

In summary, we believe that both SHIM of Choi, Li, and Zhu (2010) and WHIM proposed here provide attractive alternatives for analyzing models following heredity properties.

## ACKNOWLEDGMENTS

We thank Huanan Zhang of the University of Minnesota for computational help. This work was supported by the Dean's Small Research Grant at the Carlson School of Management of the University of Minnesota.

## ADDITIONAL REFERENCES

Choi, N. H., Li, W., and Zhu, J. (2010), "Variable Selection With the Strong Heredity Constraint and Its Oracle Property," *Journal of American Statistical Association*, 105, 354–364. [21,22]