Title
Analysis of Supersaturated Designs via Dantzig Selector

Authors
Frederick K.H. Phoa
Yu-Hui Pan
Hongquan Xu

Publication Date
2011-10-25

Analysis of Supersaturated Designs via Dantzig Selector

FREDERICK K.H. PHOA, YU-HUI PAN and HONGQUAN XU*
Department of Statistics, University of California, Los Angeles, CA 90095, USA

Abstract

A supersaturated design is a design whose run size is not enough for estimating all the main effects. It is commonly used in screening experiment, where the goal is to identify sparse and dominant active effects with low cost. In this paper, we study a variable selection method via Dantzig selector, proposed by Candes and Tao (2007), to screen active effects. A graphical procedure and an automated procedure are suggested to accompany with the method. Simulation studies show that this method is effective over the existing data analysis methods in the literature.

Key words: Akaike Information Criterion; Dantzig Selector; Effect Sparsity; Linear Programming; Profile Plot; Screening Experiment; Supersaturated Design

1. Introduction

As science and technology have advanced to a higher level nowadays, investigators are becoming more interested in and capable of studying large-scale systems. Typically these systems have many factors that can be varied during design and operation. The cost of probing and studying a large-scale system can be extremely expensive. Building prototypes is time-consuming and costly, even using the best computer system with the best algorithms. To address the challenges posed by this technological trend, research in experimental design has lately focused on the class of *supersaturated designs* for its run size economy and mathematical novelty.

The construction of supersaturated designs dated back to Satterthwaite (1959) and Booth and Cox (1962). The former suggested the use of random balanced designs and the latter proposed an algorithm to construct systematic supersaturated designs. Many methods have been proposed for constructing supersaturated designs in the last 15 years, for examples, among others, Lin (1993, 1995), Wu (1993), Nguyen (1996), Cheng (1997), Li and Wu (1997), Tang and Wu (1997), Fang et al. (2000), Butler et al. (2001), Bulutoglu and Cheng (2004), Liu and Dean (2004), Xu and Wu (2005),

*Corresponding Author. Tel: 1-310-206-0035;
Email address: hqxu@stat.ucla.edu (Hongquan Xu).

Georgiou et al. (2006), Ai et al. (2007), Bulutoglu (2007), Liu, Liu and Zhang (2007), Liu, Ruan and Dean (2007), Ryan and Bulutoglu (2007) and Tang et al. (2007).

A common application of supersaturated designs is the *screening experiment*. There are usually a large number of factors to be investigated in the screening experiments, but it is believed that only a few of them will be active, or explicitly speaking, have significant influence on the response. This phenomenon is commonly recognized as *effect sparsity* (Box and Meyer 1986, Wu and Hamada 2000 section 3.5). The purpose of screening experiments is to identify the active factors correctly and economically. The inactive factors will be discarded, while the active factors will be investigated further in some follow-up experiments. Supersaturated designs are particularly useful in the screening experiments due to their run-size economy (Lin 1999).

Some analysis methods were developed in recent years. Lin (1993) used stepwise regression for selecting active factors. Chipman et al. (1997) proposed a Bayesian variable-selection approach for analyzing experiments with complex aliasing. Westfall et al. (1998) proposed an error control skill in forward selection. Beattie et al. (2002) proposed a two-stage Bayesian model selection strategy for supersaturated experiments. Li and Lin (2002, 2003) proposed a method based on penalized least squares. Holcomb et al. (2003) proposed contrast-based methods. Lu and Wu (2004) proposed a modified stepwise selection based on an idea of staged dimensionality reduction. Zhang et al. (2007) proposed a method based on partial least squares.

In this paper, we consider searching active factors in the supersaturated designs via Dantzig selector, a new estimator proposed by Candes and Tao (2007). The Dantzig selector chooses the best subset of variables or active factors by solving a very simple convex program, which can be recast as a convenient linear program. Candes and Tao (2007) showed that the Dantzig selector has some remarkable properties under some conditions. Our simulation also demonstrates that the Dantzig selector is powerful for analyzing supersaturated designs.

This paper is organized as follows. In Section 2, we introduce Dantzig selector, and discuss how to implement the Dantzig selector in practice. Section 3 suggests a graphical procedure, called *Profile Plot*, in analyzing the results from the Dantzig selector method. Three real-life experiments are used to examine the efficiency of the profile plots. The results show that the profile plot is efficient at identifying active factors in the experiments, even if there are mixed-level factors. Section 4 suggests an automatic variable selection procedure to accompany with the Dantzig selector method. A new criterion modified from traditional AIC is suggested. Real-life experiments are used again to show the efficiency of the numerical method. In section 5, two simulations are performed to show how efficient Dantzig selector method is

when it is compared to existing approaches in the literature. A final conclusion is given in Section 6.

2. Dantzig Selector

Consider a linear regression model $y = X\beta + \varepsilon$ where y is an $n \times 1$ vector of observations, X is an $n \times k$ model matrix, β is the $k \times 1$ vector of unknown parameters, and ε is an $n \times 1$ vector of random errors. Assume that $\varepsilon \sim N(\mathbf{0}, \sigma^2 \boldsymbol{I_n})$ is a vector of independent normal random variables. Candes and Tao (2007) proposed a new estimator called *Dantzig selector* to estimate the vector of parameters β under the situation of supersaturated experiments (i.e., the number of variables is greater than the number of observations). This estimator is the solution to the l_1-regularization problem

$$\min_{\hat{\beta} \in R^k} \left\| \hat{\beta} \right\|_{l_1} \text{ subject to } \left\| X^t r \right\|_{l_\infty} \leq \delta,$$

where r is the residual vector $r = y - X\hat{\beta}$, δ is a tuning parameter and for a vector a, $\|a\|_{l_1} = \sum |a_i|$ and $\|a\|_{l_\infty} = max|a_i|$. In other words, an estimator with minimum complexity measured by the l_1-norm is searched among all estimators that are consistent with the data.

According to Candes and Tao (2007), there are some reasons to restrict the correlated residual vector $X^t r$ rather than the size of the residual vector r. One of the reasons is that the estimation procedure using correlated residual vector is invariant with respect to orthonormal transformations applied to the data vector since the feasible region is invariant. Suppose an orthonormal transformation is applied to the data, giving $y' = Uy$, then $(UX)^t \left(Uy - UX\hat{\beta}\right) = X^t \left(y - X\hat{\beta}\right)$, which shows the invariant. This implies that the estimation of β does not depend upon U.

The Dantzig selector can be recast as a linear program

$$min \sum_i u_i \text{ subject to } -u \leq \hat{\beta} \leq u \text{ and } -\delta \mathbf{1_k} \leq X^t \left(y - X\hat{\beta}\right) \leq \delta \mathbf{1_k},$$

where the optimization variables are u, $\hat{\beta} \in R^k$ and $\mathbf{1_k}$ is a vectors of k ones. This is equivalent to the standard linear program

$$min\, c^t x \text{ subject to } Ax \geq b \text{ and } x \geq 0,$$

where

$$c = \begin{pmatrix} \mathbf{1_k} \\ \mathbf{0_k} \end{pmatrix}, \quad A = \begin{pmatrix} X^t X & -X^t X \\ -X^t X & X^t X \\ 2\boldsymbol{I_k} & -\boldsymbol{I_k} \end{pmatrix}, \quad b = \begin{pmatrix} -X^t y - \delta \mathbf{1_k} \\ X^t y - \delta \mathbf{1_k} \\ \mathbf{0_k} \end{pmatrix}, \quad x = \begin{pmatrix} u \\ u + \beta \end{pmatrix}.$$

Candes and Tao (2007) showed that under certain conditions on the model matrix X which roughly guarantee that the model is identifiable, the Dantzig selector can correctly identify the active variables with large probability. Unfortunately, the

conditions are too strong and most supersaturated designs in the literature do not satisfy these conditions.

When X is an orthogonal matrix and has unit length for each column, the Dantzig selector $\hat{\beta}$ is then the l_1-minimizer subject to the constraint $\left\|X^t y - \hat{\beta}\right\|_{l_\infty} \leq \delta$. This implies that $\hat{\beta}$ is simply the soft-thresholded version of $X^t y$ at level δ, thus

$$\hat{\beta}_i = \begin{cases} (X^t y)_i - \delta & if\ (X^t y)_i \geq \delta \\ (X^t y)_i + \delta & if\ (X^t y)_i \leq -\delta \\ 0 & \text{otherwise} \end{cases}$$

where $(X^t y)_i$ is the i^{th} component of $X^t y$. In other words, $X^t y$ is shifted toward the origin if X is an orthogonal matrix. For arbitrary X's, the method continues to exhibit a soft-thresholding type of behavior and as a result, may slightly underestimate the true value of the nonzero parameters.

There are several simple methods to correct for this bias and increase performance in practical settings. Candes and Tao (2007) suggested a two-stage procedure. First, estimate $I = \{i: \beta_i \neq 0\}$ with $\hat{I} = \left\{i: \left|\hat{\beta}_i\right| > \gamma\right\}$ for some $\gamma \geq 0$ with $\hat{\beta}$ as in the solution to the l_1-regularization problem. Second, construct the estimator $\hat{\beta}_{\hat{I}} = \left(X_{\hat{I}}^t X_{\hat{I}}\right)^{-1} X_{\hat{I}}^t y$ and set the other coordinates to zero. Hence, we rely on the Dantzig selector to estimate the model I by $\hat{I}$, and construct a new estimator by regressing y onto the model $\hat{I}$. Candes and Tao (2007) referred this variation as the *Gauss-Dantzig selector*. This estimator centralizes the estimates and generally yields higher statistical accuracy.

The tuning parameter (δ) in the l_1-regularization problem has a significant impact on the results of the estimates. If δ is set to be too high, or in other words, we allow a large range of residuals to take part in the regression equation, the residuals are able to explain all the variations of the response themselves without considering any changes in predictors. This leads to the insignificance of all predictors towards the change in response, so we drop all of the predictors. Oppositely, if δ is set to be too low, or in other words, we minimize the variation of the residuals, the variation of the response has to be explained by the predictors, so some inactive factors with small magnitudes of coefficients are falsely included to help explaining in the variation of the response. Therefore, an appropriate value of δ is essential to the active-factor identification.

The threshold (γ) in the model selection may have a significant impact on the results of the final model. If γ is set to be too large, some true active factors are falsely identified as inactive, and then the final model is different from the true model because it misses some true active factors. If γ is set to be too small, some inactive factors are falsely identified as active, and then the final model is different from the

true model because it includes some inactive factors. Therefore, an appropriate value of γ is also essential to the active-factor identification. For convenience, we define the noise level to be the range between $\pm\gamma$. Unless specified, we select γ to be $\frac{1}{10}$ of the largest $|\hat{\beta}_i|$ in the model when $\delta = 0$. The choice $\frac{1}{10}$ is arbitrary.

3. A Procedure for Analyzing Supersatuated Designs

Candes and Tao (2007) suggested the choice of $\delta = \lambda\sigma$ when *X* is unit length normalized, where $\lambda = \sqrt{2logk}$ and σ is the standard deviation of the random error. However, we do not know σ in practice. Furthermore, even if we know σ (as in simulation), the choice of $\lambda = \sqrt{2logk}$ is not appropriate for the supersaturated designs we consider.

Here we propose a simple approach for analyzing supersaturated designs:

1. Standardize data so that *y* has mean 0 and columns of *X* have equal length.
2. Use linear program to obtain the Dantzig selector $\hat{\beta}$ for a range of δ.
3. Make a profile plot of the estimates by plotting the $\hat{\beta}$ against δ.
4. Choose a proper δ and the active effects according to the profile plot.
5. Obtain new estimates by regressing *y* on the active effects selected in step 4.

Here are three examples on real data.

Example 1. Consider the cast fatigue experiment (Wu and Hamada 2000, section 7.1), a real data set consisting of 7 two-level factors. The design matrix and the response data are given in Table 1. We first consider the main effects model, where each column corresponds to a two-level factor. The profile plot (Figure 1) suggests that *F* does not decay to noise level even if we choose a large value of δ. In addition, *D* decays to noise level if we choose a small to medium value of δ. This implies that *F* is strongly significant and *D* is moderately significant. Our result is consistent to the analysis using half-normal plot in Wu and Hamada (2000, Figure 8.1)

We further investigate potential active two-factor interactions. We consider a model with 7 main effects and all 21 two-factor interactions so that the model is supersaturated. The profile plot (Figure 2) suggests that there are three significant effects, *F*, *FG* and *AE*, which do not decay to noise range even if we select a large value of δ. Other effects decay to zero even though only a small value of δ is selected. Among these three factors, *F* and *FG* are strongly significant and *AE* is weakly significant when we entertain the two-factor interactions. This agrees with the result in Westfall et al. (1998). Note that the significance of *AE* without its parent main

effects violates the effect heredity principle (Wu and Hamada 2000, section 3.5), so one might accept a model with *F* and *FG* only, which is recommended by Wu and Hamada (2000).

Example 2. Consider the blood glucose experiment (Wu and Hamada 2000, section 7.1), a real data set consisting of 1 two-level and 7 three-level factors. The design matrix and the response data are given in Table 2. We first apply Dantzig selector to a main effects model with a 18×15 model matrix. The first column corresponds to the two-level factor *A*. The next 7 columns correspond to the linear contrast of the 7 three-level factors from *B* to *H*. The last 7 columns correspond to the quadratic contrast of the 7 three-level factors. The coding of linear and quadratic contrasts is:

$$\text{Linear Contrast: } (0 \quad 1 \quad 2) \rightarrow (+1 \quad 0 \quad -1)$$
$$\text{Quadratic Contrast: } (0 \quad 1 \quad 2) \rightarrow (+1 \quad -2 \quad +1)$$

The model matrix *X* is normalized to have unit length for each column. The profile plot (Figure 3) suggests that E_q and F_q do not decay to noise level even if we select a large value of δ. Even though there are several factors that do not decay to 0 for large residual variations, it is difficult to distinguish them from the noise level. This implies that E_q and F_q are strongly significant. Our result is consistent to the analysis using half-normal plot in Wu and Hamada (2000, Figure 8.2)

We also include two-factor interaction terms in the analysis. We consider a model with 15 main effects and 98 two-factor interaction effects. The model matrix *X* is normalized to have unit length for each column. The profile plot (Figure 4) suggests that there are two significant effects, $(BH)_{lq}$ (the interaction between the linear contrast of *B* and the quadratic contrast of *H*) and $(BH)_{qq}$ (the interaction between the quadratic contrasts of *B* and *H*), that do not decay to noise range even if we select a large value of δ. Other effects decay to noise level even though a small to moderate value of δ is selected. Our result does not completely match Equation 8.10 of Wu and Hamada (2000), but it is consistent to the top model identified by a Bayesian approach in Wu and Hamada (2000, Table 8.3).

Example 3. In this example, we apply Dantzig selector to the supersaturated design demonstrated first by Lin (1993). The design matrix and response data are given in Table 3. The profile plot (Figure 5) suggests that there is only one strong significant effect, X_{15}, in this data. X_{17} is barely considered as a moderate significant effect, depending on the range of the noise level we set.

Westfall et al. (1998), Beattie et al. (2002) and Li and Lin (2003) have done some analyses on the same design. The results of forward selection in Westfall et al. (1998) highlights $X_{15,}$ X_{12}, X_{20}, X_4, X_{10} and X_{11} as important variables. Among them,

X_{15} is the only significant variable at 5% significance level, and X_4 is marginally significant. Beattie et al. (2002) summarizes the important factors identified from different model selection method. Factor 14, which is our X_{15}, is identified as important in every model selection. Li and Lin (2003) suggests $X_{15,}$ X_{12}, X_{20}, X_4 and X_{10} as active effects.

We compare our result to the results obtained in the above methods. The difference is not surprising when we look at the trajectories of different $\hat{\beta}$ in Figure 5. Almost all effects, except X_{15}, are noisy and the magnitudes are small enough to be within the noise level. We agree with Abraham et al. (1999) that it is not clear the correct answers on which the active factors are, and different approaches may provide different answers on the list of active factors. X_{15} is probably the only active factors found in different approaches.

4. Automatic Variable Selection

Here is a general procedure to select an appropriate δ. For a fixed γ, we obtain a list of active factors $\hat{I} = \{i: |\hat{\beta}_i| > \gamma\}$ for different δ. Then we compare different linear models suggested by $\hat{I}$ according to some criteria. The popular model selection criterion is the *Akaike Information Criterion* (AIC), which is equivalent to

$$AIC = nlog\left(\frac{RSS}{n}\right) + 2p$$

for linear regression where $RSS = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$ is the residual sum of squares and p is the number of parameters in the model. Traditional AIC does not work properly for supersaturated designs, so we suggest a modified AIC criterion for model comparison:

$$mAIC = \frac{n}{p}log\left(\frac{RSS}{n}\right) + \frac{2p^2}{\sqrt{n}}$$

The main differences between our criterion and the conventional version are as follows. First, we standardize the conventional AIC by dividing the number of active factors and include it in the first term. This minimizes the effect of the number of factors that commonly affects AIC. Second, we propose a penalized multiplier that includes both the model complexity and the number of observations, rather than either a constant or observational-dependent only. It is important because it is preferred to have small number of active factors in the screening experiment. The modification of the penalty makes our criterion more efficient than the conventional criterion.

We illustrate the above procedure on example 1. Automatic variable selection suggests that the model containing a strongly significant effect F has the minimum mAIC = -15.54 when δ is between 2.62 and 5.01 and γ is fixed at 0.0458. For the

model with the two-factor interactions, the model that contains two strongly significant effects *F* and *FG* has the minimum mAIC = -15.60 when δ is between 4.22 and 4.95 and γ is fixed at 0.0416.

5. Simulations

In this section, we investigate the performance of the Dantzig selector approach via simulation. Example 4 compares the performance of the Dantzig selector method with four different approaches suggested in the literature, and they are (i) SSVS, the Bayesian variable selection procedure proposed by George and McMulloch (1993) and extended for supersaturated designs by Chipman et al. (1997); (ii) SSVS/IBF, the two stage Bayesian procedure by Beattie et al. (2002); (iii) SCAD, the penalized least squares approach proposed by Li and Lin (2003); and (iv) PLSVS, the partial least square regression technique by Zhang et al. (2007). Example 5 explores the performance of the Dantzig Selector method in a design with a large number of factors. All simulations are conducted using R codes.

Example 4. To compare the performance of the Dantzig selector method with that of the four methods by simulation, we consider the same models as Li and Lin (2003).

Consider the design matrix X displayed in Table 3. We generate data from the linear model

$$y = X\beta + \varepsilon$$

where the random error $\varepsilon \sim N(0,1)$. We consider the following three cases for β:

Case I: One active factor, $\beta_1 = 10$, and all other components of β equal 0;

Case II: Three active factors, $\beta_1 = -15, \beta_5 = 8, \beta_9 = -2$, and all other components of β equal 0;

Case III: Five active factors, $\beta_1 = -15, \beta_5 = 12, \beta_9 = -8, \beta_{13} = 6, \beta_{17} = -2$, and all other components of β equal 0;

Simulation results for Dantzig selector based on 1000 replicates are summarized in Table 4 and are compared with the other four methods. In this table, "TMIR" stands for True Model Identified Rate, "SEIR" stands for Smallest Effect Identified Rate, and "Median" and "Mean" are the median and mean sizes of the models. In the simulation, we fix γ=1 and choose δ according to the mAIC criterion.

The Dantzig selector method identifies the true model with the highest probabilities among all five methods. In case I, the Dantzig selector shares 100% perfect identification rates with SCAD and PLSVS in identifying the smallest effect. In case II and III, the probability of getting the smallest effect with the Dantzig selector method is a little lower than that of the SCAD and PLSVS. In terms of the

model size, the Dantzig selector method performs better than others. The average model size is closer to the true model size than those resulted from the other four methods. In this sense our method is more efficient than the other four.

Example 5. In this example, we consider a systematically generated design in Lin (1995). The design has 12 runs and 66 factors, which is bigger than the model in Example 4. We generate data from the same linear model as in Example 4, and we consider five different cases. There are i active factors in case i, where i ranges from 1 to 5. The selection of active factors is random without replacement. The signs of the active factors are randomly selected from either positive or negative, and the magnitudes are randomly selected from 2 to 10 with replacement. For each case, 400 trials are performed and each trial consists of 100 replicates. We fix γ=0.75 and choose δ according to the mAIC criterion. For each trial, we obtain the true model identification rate and the average model size of 100 replicates. Table 3 summarizes the results of 400 trials. In this table, "TMIR" and "Size" correspond to the true model identification rate and the size of the model under the optimal value of mAIC. The summary statistics of these two quantities are presented.

The Dantzig selector method is very efficient in identifying 1 active factor in the simulation. The probability of correctly identifying the active factors in true model is high (87 – 99%) and the average model size (1.01 – 1.15) is very near to the true number of active factors. The method is still efficient in identifying 2 to 4 active factors from 66 factors in the simulation. The probabilities of correctly identifying the active factors in true model are still reasonable with mean ranging from 46.9% to 66.6% and the average model sizes are near to the true number of active factors. The method is not good in identifying 5 active factors. Even though the ability of identification decreases when the number of active factors increases, this method is still useful to identify small number of dominant active effects.

6. Concluding Remarks

This paper studies the Dantzig selector for selecting active effects in supersaturated designs. We suggest a graphical procedure and an automatic variable selection method to accompany with the Dantzig selector. A modified AIC criterion is proposed for model selection. Simulations demonstrate that the Dantzig selector method is efficient over the existing data analysis methods in the literature. This ability of active-factor identification raises tantalizing opportunities in various fields. Candes and Tao (2007) suggested several applications in biomedical imaging, analog to digital conversion and sensor networks.

The modified AIC criterion works for the simulations conducted here, but may not work well for other situations. Nevertheless, it demonstrates that supersaturated designs are useful when properly analyzed and that the Dantzig selector is a good tool.

The advantages of Dantzig selector method are as follows. First, Dantzig selector method is relatively fast, easy and simple to use. It is basically a linear program, which is widely considered as a fast and efficient algorithm to perform massive computation. In addition, the linear programming algorithm is available in many software and packages, like R, Matlab, Mathematica, etc. The computation of this paper is done by using R package "lpSolve". These software and packages make Dantzig selector easy to program and use.

Second, Dantzig selector method is able to handle a large number of factors in two-levels, multi-levels and mixed-levels. Candes and Tao (2007) applied Dantzig selector to an experiment with up to 200 active factors among 5000 binary factors and 1000 observations. In addition, Example 2 demonstrates that Dantzig selector is readily implemented for multi-level designs via the method of orthogonal contrasts.

Third, Candes and Tao (2007) provides good theoretical justifications to Dantzig selector. When some uniform uncertainty conditions are fulfilled, they proved that Dantzig selector is able to perform an ideal model selection.

ACHKNOWLEDGMENT

The research was partially supported by National Science Foundation Grant DMS-0505728.

REFERENCE

[1] Abraham, B., Chipman, H., Vijayan, K., 1999. Some risks in the construction and analysis of supersaturated designs. Technometrics 41, 135-141.

[2] Ai, M., Fang, K.T., He, S., 2007. E(s^2)-optimal mixed-level supersaturated designs. Journal of Statistical Planning and Inference 137(1), 306-316.

[3] Beattie, S.D., Fong, D.K.H., Lin, D.K.J., 2002. A two-stage Bayesian model selection strategy for supersaturated designs. Technometrics 44, 55-63.

[4] Booth, K.H.V., Cox, D.R., 1962. Some systematic supersaturated designs. Technometrics 4, 489-495.

[5] Box, G.E.P., Meyer, R.D., 1986. An analysis for unreplicated fractional factorials. Technometrics 28, 11-18.

[6] Bulutoglu, D.A., 2007. Cyclicly constructed E(s^2)-optimal supersaturated designs. Journal of Statistical Planning and Inference 137(7), 2413-2428.

[7] Bulutoglu, D.A., Cheng, C.S., 2004. Construction of E(s^2)-optimal supersaturated designs.

Annual of Statistics 32, 1662-1678.

[8] Butler, N.A., Mead, R., Eskridge, K.M., Gimour, S.G., 2001. A general method of constructing $E(s^2)$-optimal supersaturated designs. Journal of the Royal Statistical Society. Series B. Statistical Methodology 63, 621-632.

[9] Candes, E.J., Tao, T., 2007. The Dantzig selector: statistical estimation when p is much larger than n. To appear in Annals of Statistics. Available on the ArXiV preprint server: math.ST/0506081.

[10] Cheng, C.S., 1997. $E(s^2)$-optimal supersaturated designs. Statistica Sinica 7, 929-939.

[11] Chipman, H., Hamada, H., Wu, C.F.J., 1997. A Bayesian variable-selection approach for analyzing designed experiments with complex aliasing. Technometrics 39, 372-381.

[12] Fang, K.T., Lin, D.K.J., Ma, C.X., 2000. On the construction of multi-level supersaturated designs. Journal of Statistical Planning and Inference 86, 239-252.

[13] George, E.I., McMulloch, R.E., 1993. A statistical view of some chemometrics regression tools. Technometrics 35, 109-148.

[14] Georgiou, S., Koulouvinos, C., Mantas, P., 2006. On multi-level supersaturated designs. Journal of Statistical Planning and Inference 136(8), 2805-2819.

[15] Holcomb, D.R., Montgomery, D.C., Carlyle, W.M., 2003. Analysis of Supersaturated Designs. Journal of Quality Technology 35, 13-27.

[16] Li, R., Lin, D.K.J., 2002. Data analysis in supersaturated designs. Statistics & Probability Letters 59, 135-144.

[17] Li, R., Lin, D.K.J., 2003. Analysis methods for supersaturated design: some comparisons. Journal of Data Science 1, 249-260.

[18] Li, W.W., Wu, C.F.J., 1997. Columnwise-pairwise algorithms with applications to the construction of supersaturated designs. Technometrics 39, 171-179.

[19] Lin, D.K.J., 1993. A new class of supersaturated designs. Technometrics 35, 28-31.

[20] Lin, D.K.J., 1995. Generating systematic supersaturated designs. Technometrics 37, 213-225.

[21] Lin, D.K.J., 1999. Supersaturated design (Update). Encyclopedia of Statistical Science (Update) 3, 727-731.

[22] Liu, Y., Dean, A., 2004. k-circulant supersaturated designs. Technometrics 46, 32-43.

[23] Liu, Y., Liu, M., Zhang, R., 2007. Construction of multi-level supersaturated design via Kronecker product. Journal of Statistical Planning and Inference 137(9), 2984-2992.

[24] Liu, Y., Ryan, S., Dean, A.M., 2007. Construction and analysis of $E(s^2)$ efficient supersaturated designs. Journal of Statistical Planning and Inference 137(5), 1516-1529.

[25] Lu, X., Wu, X., 2004. A strategy of searching active factors in supersaturated screening experiments. Journal of Quality Technology 36, 392-399.

[26] Nguyen, N.K., 1996. An algorithmic approach to constructing supersaturated designs. Technometrics 38, 69-73.

[27] Ryan, K.J., Bulutoglu, D.A., 2007. $E(s^2)$-optimal supersaturated designs with good minimax

properties. Journal of Statistical Planning and Inference 137(7), 2250-2262.

[28] Satterthwaite, F.E., 1959. Random balance experimentation. Technometrics 1, 111-137.

[29] Tang, B., Wu, C.F.J., 1997. A method for constructing supersaturated designs and its $E(s^2)$ optimality. Canadian Journal of Statistics 25, 191-201.

[30] Tang, Y., Ai, M., Ge, G., Fang, F.T., 2007. Optimal mixed-level supersaturated designs and a new class of combinational designs. Journal of Statistical Planning and Inference 137(7), 2294-2301.

[31] Westfall, P.H., Young, S.S., Lin, D.K.J., 1998. Forward selection error control in the analysis of supersaturated designs. Statistica Sinica 8, 101-117.

[32] Wu, C.F.J., 1993. Construction of supersaturated designs through partially aliased interactions. Biometrika 80, 661-669.

[33] Wu, C.F.J., Hamada, M., 2000. Experiments: Planning, Analysis, and Parameter Design Optimization. Wiley, New York.

[34] Xu, H., Wu, C.F.J., 2005. Construction of optimal multi-level supersaturated designs. Annals of Statistics 33, 2811-2836.

[35] Zhang, Q.Z., Zhang, R.C., Liu, M.Q., 2007. A method for screening active effects in supersaturated designs. Journal of Statistical Planning and Inference 137(9), 235-248.

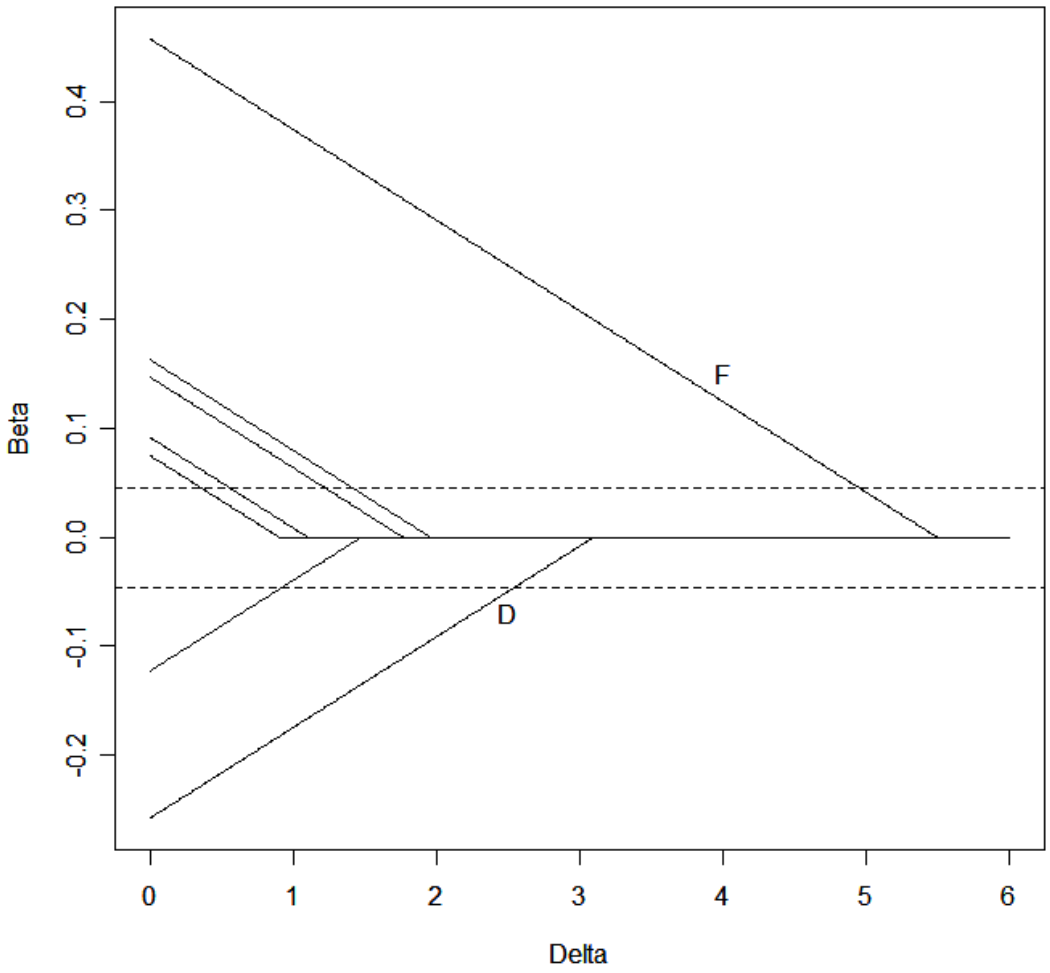

Figure 1. Profile plot for the cast fatigue experiment without interactions. The model includes 7 main effects. $\hat{\beta}_F$ and $\hat{\beta}_D$ decay slowly to noise level (γ=0.0458, dotted lines) when δ increases.

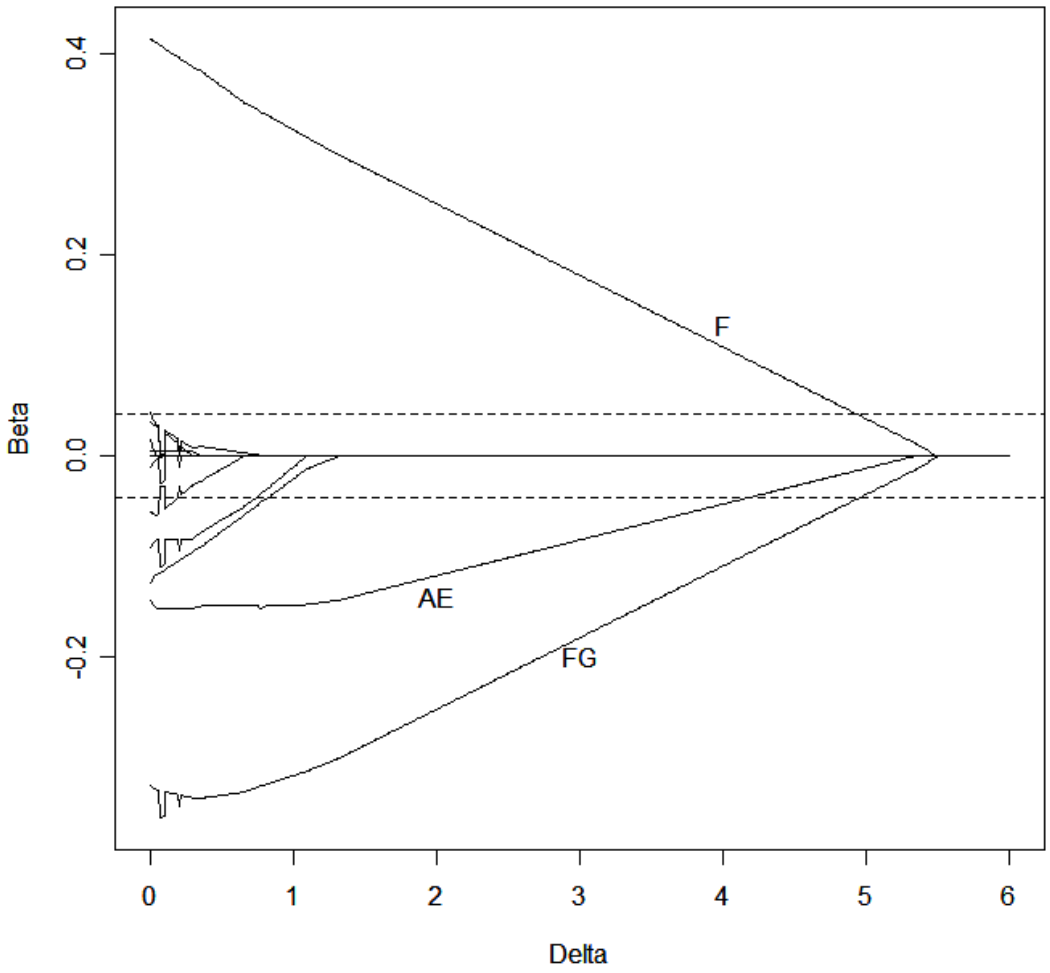

Figure 2. Profile plot for the cast fatigue experiment with interactions. The model contains 7 main effects and 21 two-factor interactions. $\hat{\beta}_F$, $\hat{\beta}_{FG}$ and $\hat{\beta}_{AE}$ decay slowly to noise level (γ=0.0416, dotted lines) when δ increases.

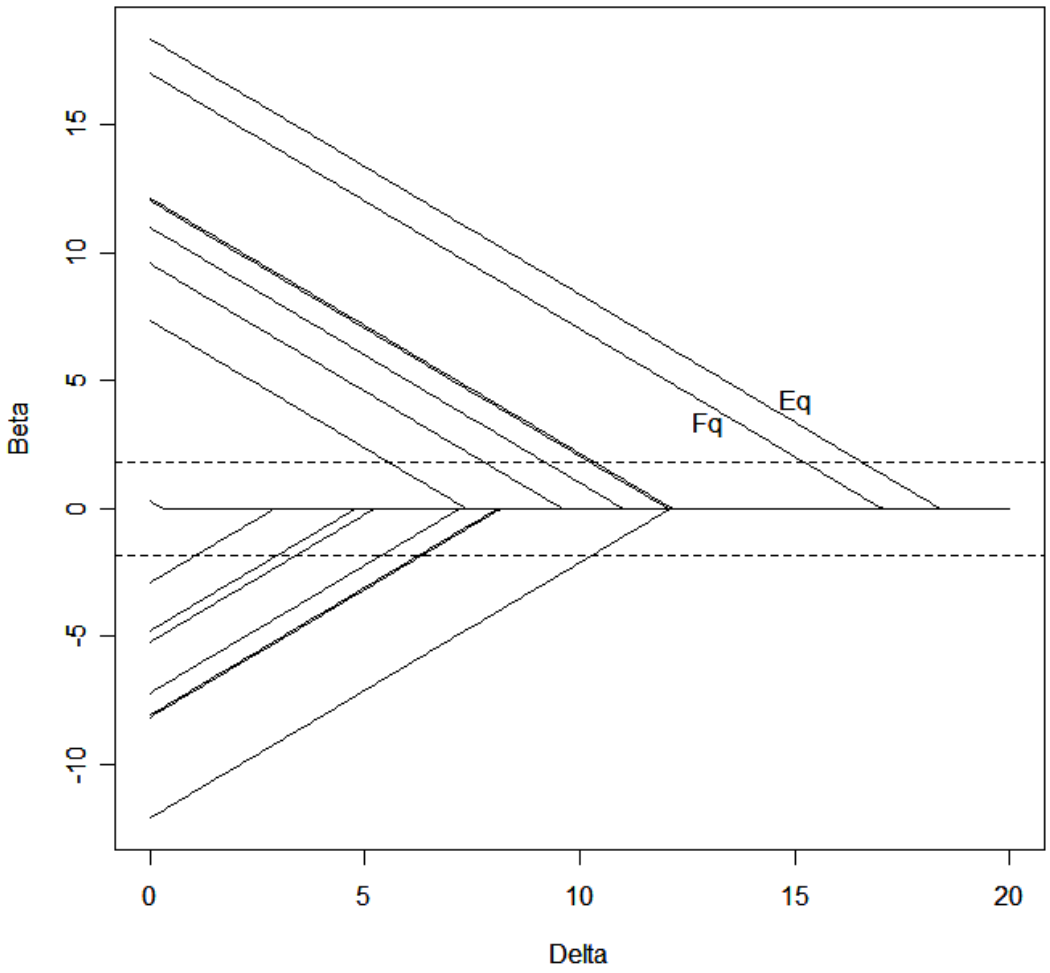

Figure 3. Profile plot for the blood glucose experiment without interactions. The model contains 15 main effects. $\hat{\beta}_{E_q}$ and $\hat{\beta}_{F_q}$ decay slowly to noise level (γ=1.8384, dotted lines) when δ increases.

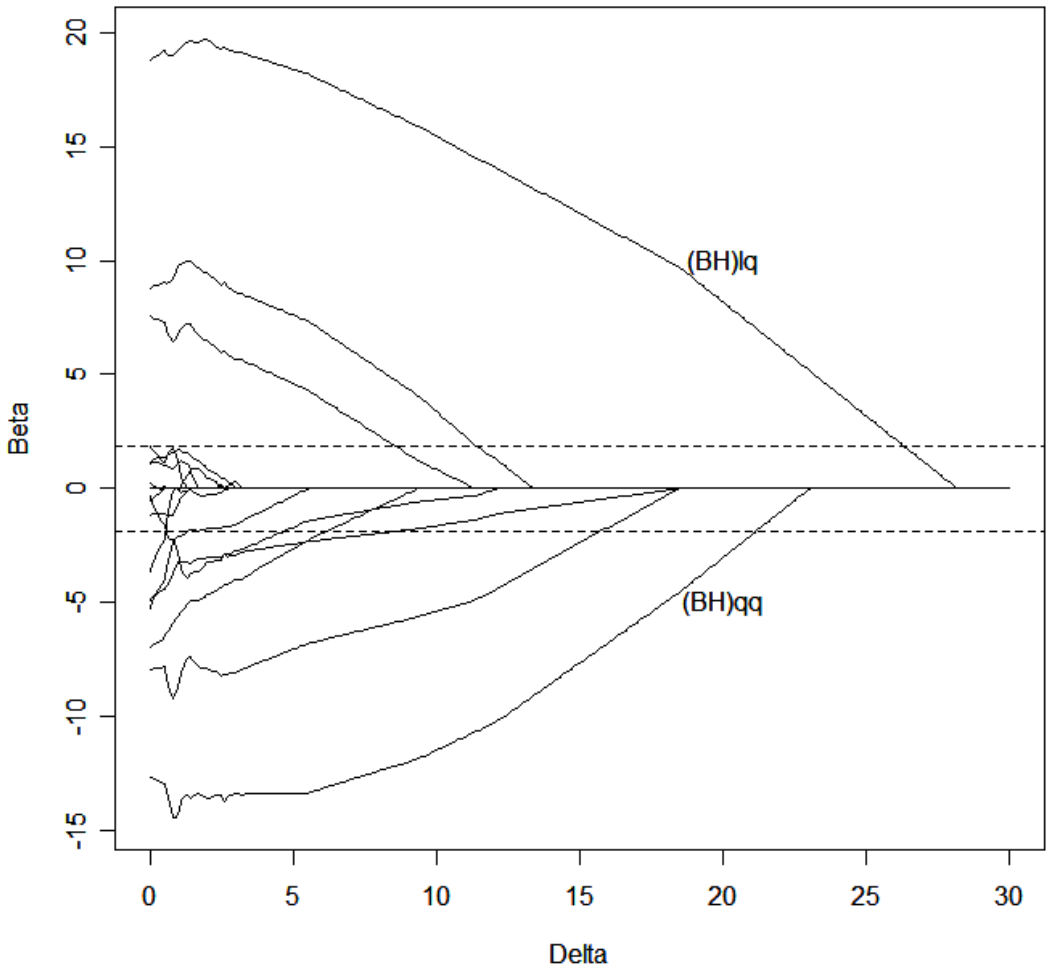

Figure 4. Profile plot for the blood glucose experiment with interactions. The model contains 15 main effects and 98 two-factor interaction effects. $\hat{\beta}_{(BH)_{lq}}$ and $\hat{\beta}_{(BH)_{qq}}$ decay slowly to noise level (γ=1.8803, dotted lines) when δ increases.

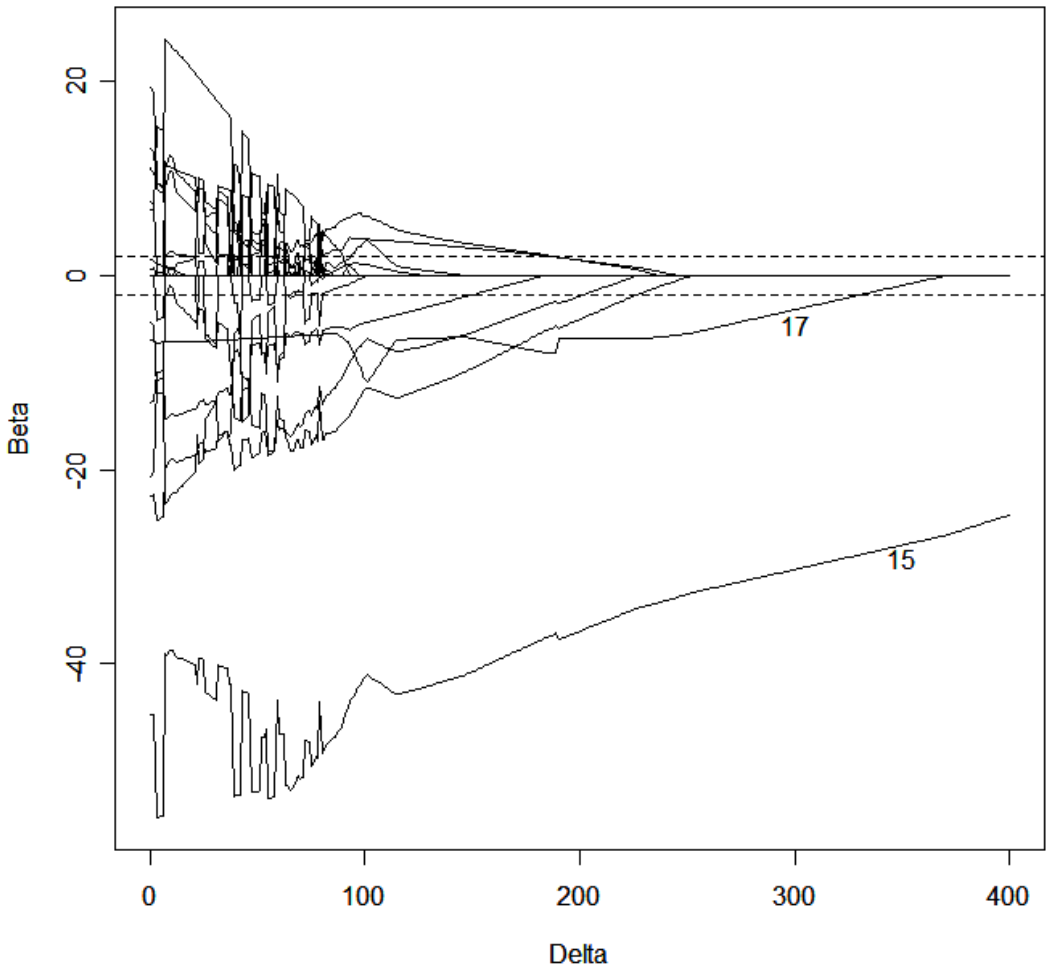

Figure 5. Profile plot for the Lin (1993) data. The model contains 23 main effects. $\hat{\beta}_{15}$ decays very slowly when δ increases, and $\hat{\beta}_{17}$ decays to noise level (γ=4.5214, dotted lines) when δ increases.

Table 1. Design Matrix and Response Data, Cast Fatigue Experiment

Run	A	B	C	D	E	F	G	Response
1	+	+	–	+	+	+	–	6.058
2	+	–	+	+	+	–	–	4.733
3	–	+	+	+	–	–	–	4.625
4	+	+	+	–	–	–	+	5.899
5	+	+	–	–	–	+	–	7.000
6	+	–	–	–	+	–	+	5.752
7	–	–	–	+	–	+	+	5.682
8	–	–	+	–	+	+	–	6.607
9	–	+	–	+	+	–	+	5.818
10	+	–	+	+	–	+	+	5.917
11	–	+	+	–	+	+	+	5.863
12	–	–	–	–	–	–	–	4.809

Table 2. Design Matrix and Response Data, Blood Glucose Experiment.

Run	A	B	C	D	E	F	G	H	Response
1	0	0	0	0	0	0	0	0	97.94
2	0	1	1	1	1	1	0	1	83.40
3	0	2	2	2	2	2	0	2	95.88
4	0	0	0	1	1	2	1	2	88.86
5	0	1	1	2	2	0	1	0	106.58
6	0	2	2	0	0	1	1	1	89.57
7	0	0	1	0	2	1	2	2	91.98
8	0	1	2	1	0	2	2	0	98.41
9	0	2	0	2	1	0	2	1	87.56
10	1	0	1	2	1	1	0	0	88.11
11	1	1	2	0	2	2	0	1	83.81
12	1	2	0	1	0	0	0	2	98.27
13	1	0	2	2	0	2	1	1	115.52
14	1	1	0	0	1	0	1	2	94.89
15	1	2	1	1	2	1	1	0	94.70
16	1	0	2	1	2	0	2	1	121.62
17	1	1	0	2	0	1	2	2	93.86
18	1	2	1	0	1	2	2	0	96.10

Table 3. A two-level supersaturated design (Lin 1993)

X		Factors																						Response Y
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	17	18	19	20	21	22	23	24	
Runs	1	+	+	+	–	–	–	+	+	+	+	+	–	+	–	–	+	–	–	+	–	–	–	+	133
	2	+	–	–	–	–	–	+	+	+	–	–	–	+	+	+	–	+	–	–	+	+	–	–	62
	3	+	+	–	+	+	–	–	–	–	+	+	+	+	+	+	+	–	–	–	–	+	+	–	45
	4	+	+	–	+	–	+	–	–	–	+	+	–	+	–	+	–	+	+	+	–	–	–	+	52
	5	–	–	+	+	+	+	–	+	+	–	–	–	+	–	+	+	–	–	+	–	+	+	–	56
	6	–	–	+	+	+	+	+	–	+	+	+	–	–	+	+	+	+	+	+	+	+	–	+	47
	7	–	–	–	–	+	–	–	+	–	+	–	+	+	+	–	+	+	+	+	+	–	–	–	88
	8	–	+	+	–	–	+	–	+	–	+	–	–	–	–	–	–	–	+	–	+	+	+	+	193
	9	–	–	–	–	–	+	+	–	–	–	+	+	–	–	+	+	+	–	–	–	–	+	+	32
	10	+	+	+	+	–	+	+	+	–	–	–	+	–	+	+	+	–	+	–	+	–	–	+	53
	11	–	+	–	+	+	–	–	+	+	–	+	–	–	+	–	–	+	+	–	–	–	+	+	276
	12	+	–	–	–	+	+	+	–	+	+	+	+	+	–	–	–	–	+	–	+	+	+	+	145
	13	+	+	+	+	+	–	+	–	+	–	–	+	–	–	–	–	+	–	+	+	–	+	–	130
	14	–	–	+	–	–	–	–	–	–	–	+	+	–	+	–	–	–	–	+	–	+	–	–	127

Table 4. Summary of simulation results in Example 4.

Method	TMIR	SEIR	Median	Mean
Case I: One Active Effect				
SSVS (1/10,500)	40.5%	99.0%	2	3.1
SSVS (1/10,500)/IBF	61.0%	98.0%	1	2.5
SCAD	75.6%	100%	1	1.7
PLSVS (m=1)	61.0%	100%	1	1.5
Dantzig Selector	99.0%	100%	1	1.01
Case II: Three Active Effects:				
SSVS (1/10,500)	8.6%	30.0%	3	4.7
SSVS (1/10,500)/IBF	8.0%	28.0%	3	4.2
SCAD	74.7%	98.5%	3	3.3
PLSVS (m=1)	76.4%	100%	3	3.3
Dantzig Selector	91.3%	91.3%	3	3.01
Case III: Five Active Effects:				
SSVS (1/10,500)	36.4%	84.0%	6	8.0
SSVS (1/10,500)/IBF	40.7%	75.0%	5	5.6
SCAD	69.7%	99.4%	5	5.4
PLSVS (m=1)	73.6%	95.0%	5	5.2
Dantzig Selector	89.9%	89.9%	5	5.02

Table 5. Summary of simulation results in Example 5.

Case		Average results for optimal mAIC					
		Min	1Q	Median	Mean	3Q	Max
I	TMIR	87.0%	91.0%	93.0%	92.78%	94.0%	99.0%
	Size	1.010	1.060	1.070	1.076	1.090	1.150
II	TMIR	0.0%	17.0%	92.0%	66.6%	96.0%	100%
	Size	2.000	2.040	2.080	2.678	3.370	5.370
III	TMIR	0.0%	31.0%	42.0%	52.5%	90.5%	100%
	Size	2.570	3.000	3.340	3.409	3.760	5.650
IV	TMIR	0.0%	29.0%	44.0%	46.9%	60.0%	100%
	Size	2.640	3.340	3.650	3.622	3.913	5.240
V	TMIR	0.0%	2.0%	6.0%	10.7%	14.0%	100%
	Size	2.810	3.530	3.780	3.746	3.970	5.030