Analysis of Supersaturated Designs via the Dantzig Selector

Frederick K. H. Phoa, Yu-Hui Pan and Hongquan Xu[1]
Department of Statistics, University of California, Los Angeles, CA 90095-1554, U.S.A.
October 23, 2008

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 experiments, where the goals are to identify sparse and dominant active factors with low cost. In this paper, we study a variable selection method via the Dantzig selector, proposed by Candes and Tao (2007), to screen important effects. A graphical procedure and an automated procedure are suggested to accompany with the method. Simulation shows that this method performs well compared to existing methods in the literature and is more efficient at estimating the model size.

MSC: primary 62K15; secondary 62J05; 62J07

Keywords: Akaike Information Criterion; Dantzig Selector; Factor 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 their run size economy and mathematical novelty.

The construction of supersaturated designs dates 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), 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 *factor screening*. There are usually a large number of factors to be investigated in a screening experiment, but it is believed that only a few of them are active, or explicitly speaking, have significant impact on the response. This phenomenon is commonly recognized as *factor sparsity* (Box and Meyer 1986). 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 screening experimentation due to their run-size economy (Lin 1999).

Some analysis methods were developed in recent years. Lin (1993) used stepwise regression for

[1]Corresponding author. Tel.: +1 310 206 0035
E-mail address: hqxu@stat.ucla.edu (H. Xu)

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 supersaturated designs via the Dantzig selector proposed by Candes and Tao (2007). The Dantzig selector chooses the best subset of variables or active factors by solving a 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 and has been successfully used in biomedical imaging, analog to digital conversion and sensor networks, where the goals are to recover some sparse signals from some massive data. 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 the Dantzig selector and discuss how to implement the Dantzig selector in practice. Section 3 suggests a graphical procedure using a *profile plot* in analyzing the results from the Dantzig selector. Three real-life experiments are used to examine the efficiency of the method. The results show that the profile plot is efficient at identifying important factors in experiments, even if there are mixed-level factors. Section 4 suggests an automatic variable selection procedure to accompany 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, simulations are performed to show how efficient the Dantzig selector method is when it is compared to existing methods in the literature. Section 6 gives some concluding remarks.

2 The Dantzig Selector

Consider a linear regression model $y = X\beta + \epsilon$ where y is an $n \times 1$ vector of observations, X is an $n \times k$ model matrix, β is a $k \times 1$ vector of unknown parameters, and ϵ is an $n \times 1$ vector of random errors. Assume that $\epsilon \sim N(\mathbf{0}, \sigma^2 I_n)$ is a vector of independent normal random variables. Candes and Tao (2007) proposed a new estimator called the *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

$$\min_{\hat{\beta} \in R^k} \|\hat{\beta}\|_{l_1} \text{ subject to } \|X^t r\|_{l_\infty} \leq \delta \tag{1}$$

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 the 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 $(UX)^t(Uy - UX\hat{\beta}) = X^t(y - X\hat{\beta})$, 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.

$$\min \sum_i u_i \text{ subject to } -u \leq \hat{\beta} \leq u \text{ and } -\delta \mathbf{1}_k \leq X^t(y - X\hat{\beta}) \leq \delta \mathbf{1}_k \tag{2}$$

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

$$\min c^t x \text{ subject to } Ax \geq b \text{ and } x \geq 0 \tag{3}$$

where

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

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 the columns of X are orthogonal and have unit length, the Dantzig selector $\hat{\beta}$ is the l_1-minimizer subject to the constraint $\|X^ty - \hat{\beta}\|_{l_\infty} \leq \delta$. This implies that $\hat{\beta}$ is simply the soft-thresholded version of X^ty at level δ, thus

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

where $(X^ty)_i$ is the ith component of X^ty. In other words, X^ty is shifted toward the origin. For an arbitrary X, 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 $\hat{I} = \{i : |\hat{\beta}_i| > \gamma\}$ for some $\gamma \geq 0$ with $\hat{\beta}$ as in the solution to the l_1-regularization problem (1). Second, construct the estimator $\hat{\beta}_{\hat{I}} = (X^t_{\hat{I}}X_{\hat{I}})^{-1}X^t_{\hat{I}}y$ and set the other coordinates to zero, where $X_{\hat{I}}$ is the corresponding model matrix for model $\hat{I}$. 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 to this estimator 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 (1) 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 the predictors. This leads to the insignificance of all predictors towards the change in response, so we drop all of the predictors. On the other hand, 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 the variation of the response. Therefore, an appropriate value of δ is essential in identification of the active factors.

3 A Procedure for Analyzing Supersaturated Designs

A proper choice of the tuning parameter δ is crucial for the Dantzig selector. Candes and Tao (2007) suggested the choice of $\delta = \lambda\sigma$ when X is unit length normalized, where $\lambda = \sqrt{2\log k}$ and σ is the standard deviation of the random error. However, we do not know σ in practice and it is a difficult task itself to estimate σ accurately for supersaturated designs. Furthermore, even if we know σ (as in simulation), this choice of δ does not always lead to the best performance. Cai and Lv (2007) argued that it might be possible that $\lambda = \sqrt{2\log k}$ overshrinks the $k \times 1$ vector of unknown parameters β and underestimates the nonzero coordinates when k is much larger than n.

One can borrow ideas from ridge regression or other shrinkage methods. A popular practice in ridge regression is to determine the tuning parameter by inspection of the ridge trace (Hoerl and Kennard 1970). Adopting this we suggest the following procedure for screening important effects for supersaturated designs.

1. Standardize data so that y has mean 0 and columns of X have equal lengths. Compute $\delta_0 = \max |x_i^t y|$, where x_i is the ith column of X.

2. Solve the linear program (2) or (3) to obtain the Dantzig selector $\hat{\beta}$ for some values of δ ranging from 0 to δ_0.

3. Make a profile plot of the estimates by plotting $\hat{\beta}$ against δ.

4. Identify important effects by inspection of the profile plot.

It follows from (1) that all of the estimates $\hat{\beta}_i$ are 0 if $\delta \geq \delta_0$. Thus it is sufficient to choose δ from the interval $[0, \delta_0]$ in step 2. For convenience, we use up to hundreds evenly spaced δ to make a profile plot, which is very fast in computation. Then we rank and identify important effects by examining relative magnitudes of the effects as well as how slowly they decay to zero. A long lasting effect ought to be viewed as important. This procedure does not require a precise choice of δ and also allows other information to be used to determine the significance of the effects (see Example 1 below for an example). However, the effects identified as important should be treated as tentative and, as a precaution, follow-up experiments are recommended to validate the results.

We illustrate the procedure on three real data in the literature.

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. Figure 1 shows the profile plot of the estimates. Note that the trajectories are parallel lines because the columns of X are orthogonal. All estimates are 0 when $\delta > \delta_0 = 5.5$. If δ is chosen between 3.1 and 5.5, only F is nonzero; and if δ is chosen between 1.9 and 3.0, both F and D are nonzero. The conclusion is that F and possibly D are important. This 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 two or three important effects that decay slowly towards 0. Among the three effects, AE is less significant than F and FG, which 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, Section 8.4).

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 the Dantzig selector to a model with 15 terms. The first column corresponds to the two-level factor A. The next 7 columns correspond to the linear contrasts of the 7 three-level factors from B to H. The last 7 columns correspond to the quadratic contrasts of the 7 three-level factors. The coding of linear and quadratic contrasts is:

$$\begin{aligned} \text{Linear Contrast: } &\begin{pmatrix} 0 & 1 & 2 \end{pmatrix} \to \begin{pmatrix} +1 & 0 & -1 \end{pmatrix} \\ \text{Quadratic Contrast: } &\begin{pmatrix} 0 & 1 & 2 \end{pmatrix} \to \begin{pmatrix} +1 & -2 & +1 \end{pmatrix} \end{aligned}$$

The model matrix X is then normalized to have unit length for each column. Figure 3 shows the profile plot, where the trajectories are parallel lines as expected. The big gap between the two leading effects and the rest suggests that E_q and F_q are the only two important effects. The 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 and consider a model with 15 linear and quadratic terms and 98 two-factor interaction terms. The model matrix X is normalized to have unit length for each column. The profile plot (Figure 4) suggests two long lasting 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), which are most significant. However, it is not obvious whether $(AH)_{lq}$ should be viewed as significant. Using a Bayesian approach, Chipman et al. (1997, Table 7) identified top 10 models; see also Wu and Hamada (2000, Table 8.3). All 10 models include $(BH)_{lq}$ and $(BH)_{qq}$, and one of the models includes $(AH)_{lq}$.

Example 3. In this example, we apply the Dantzig selector to a supersaturated design demonstrated first by Lin (1993). The original dataset has 24 factors but two factors (13 and 16) are identical. As Beattie et al. (2002), we delete factor 13 and rename factors 14–24 as 13–23. The design matrix and response data are given in Table 3. The profile plot (Figure 5) suggests that only X_{14} appears to be important in this data.

The same data were previously analyzed by several authors. Westfall et al. (1998) highlighted X_{14}, X_{12}, X_{19}, X_4, X_{10} and X_{11} as important, among which X_{14} is the only significant variable at 5% significance level and X_4 is marginally significant. Beattie et al. (2002) compared several model selection methods; only X_{14} is identified as important in every method. Both Li and Lin (2003) and Zhang et al. (2007) suggested X_{14}, X_{12}, X_{19} and X_4 as active factors.

The difference is not surprising when we look at the trajectories of $\hat{\beta}$ in Figure 5. Almost all effects, except X_{14}, are noisy and the magnitudes are small enough to be considered 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. Different approaches may provide different answers on the list of active factors and X_{14} is probably the only common active factor found in different approaches.

4 Automatic Variable Selection

The preceding graphical procedure is simple and easy to use. Nevertheless, it is sometimes desirable, for instance in simulation, to have a procedure for choosing the tuning parameter δ automatically. Here we propose a general procedure for choosing δ based on a model selection criterion. First we obtain the Dantzig selector $\hat{\beta}$ for some values of δ ranging from 0 to δ_0 as in step 2 of the previous section. Then for a fixed $\gamma \geq 0$, we obtain a list of models $\hat{I} = \{i : |\hat{\beta}_i| > \gamma\}$, compare these models according to a criterion and choose a δ that yields the best model.

Akaike information criterion (AIC) is popular for model selection. For linear models, it is defined as

$$\text{AIC} = n \log(RSS/n) + 2p$$

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. It is known that AIC tends to overfit the model when the sample size is small. Hurvich and Tsai (1989) proposed a bias correction by adding an additional penalty term to AIC. Their modified AIC is defined as

$$\text{cAIC} = \text{AIC} + 2(p+1)(p+2)/(n-p-2).$$

The cAIC typically chooses a smaller model than AIC. However, it still tends to overfit the model for supersaturated designs.

Factor sparsity is an important assumption for the use of supersaturated designs and the Dantzig selector. Based on this assumption we impose a heavy penalty on the model complexity and propose a new modified AIC for supersaturated designs as follows:

$$\text{mAIC} = n \log(RSS/n) + 2p^2. \tag{4}$$

The difference between our modified AIC and AIC or cAIC is the penalty of model complexity p. The penalty on p in mAIC is quadratic whereas that in AIC is linear; therefore, mAIC chooses more parsimonious model than AIC. The penalty in cAIC is complicated. It is nearly quadratic on p when p is close to n and nearly linear when p is close to $n/2$. As will be seen later, this modification in (4) works well for our examples and simulations. It remains to be seen whether it works for other situations.

The parameter γ can be viewed as a threshold between signal and noise and a relatively small γ should be chosen. One can choose γ according to the profile plot or information on the magnitude of effects or noise. It is recommended that the procedure be repeated with a few choices of γ. When the signal and noise ratio is large, the choice of γ is not crucial. On the other hand, if the result is sensitive to the choice of γ, one should be cautious about the procedure and the result. In the simulation, we do a coarse grid search for γ. We find that the modified AIC defined in (4) tends to produce more robust results against different choices of γ than AIC or cAIC.

Example 4. We illustrate the automatic procedure with the cast fatigue experiment in Example 1. We fix $\gamma = 0$ and choose δ according to the three information criteria. When entertaining the main effects only, AIC chooses a 2-factor model (F, D) while both cAIC and mAIC choose a 1-factor model (F). When entertaining the two-factor interactions, AIC chooses a model with 9 terms $(F, FG, AE, D, EF, AD, DG, A, AB)$, cAIC chooses a model with 5 terms (F, FG, AE, D, EF) and mAIC chooses a model with 2 terms (F, FG). Table 4 lists RSS, R^2, and the AIC, cAIC and mAIC values for these 5 models and an additional model with 3 terms (F, FG, AE). The model with 3 terms (F, FG, AE) has a slightly larger mAIC value than the model with 2 terms (F, FG). It is evident here that mAIC performs the best among the three criteria. The mAIC works well with other choices of γ. For instance, with $\gamma = 0.1$ (roughly 25% of $\max|\hat{\beta}_i|$), both AIC and cAIC choose the 5-term model while mAIC still chooses the 2-term model when entertaining the two-factor interactions.

5 Simulation

In this section, we investigate the performance of the Dantzig selector approach via simulation. Example 5 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). Our simulations are conducted in R using package "lpSolve".

Example 5. 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) and Zhang et al. (2007). We generate data from the linear model

$$y = X\beta + \epsilon \tag{5}$$

where X is the 14×23 matrix given in Table 3 and the random error $\epsilon \sim N(0, 1)$. 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, $\beta_1 = -15$, β_5=8, $\beta_9 = -2$, and all other components of β equal 0;

Case III: Five active factors, $\beta_1 = -15$, β_5=12, $\beta_9 = -8$, $\beta_{13} = 6$, $\beta_{17} = -2$, and all other components of β equal 0.

We run the simulations 1,000 times by fixing $\gamma = 1$ (corresponding to 10% or 6.7% of $\max |\beta_i|$) and choosing δ automatically using mAIC. Table 5 compares the Dantzig selector method 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.

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 cases II and III, the probability of getting the smallest effect with the Dantzig selector method is less than that of SCAD and PLSVS. In terms of the model size, the Dantzig selector method performs the best. The average model size is closer to the true model size than those resulted from the other methods. In this sense our method is more efficient.

We also evaluate the performance of the Dantzig selector with different choices of γ and different criteria. Table 6 summarizes simulation results using AIC, cAIC and mAIC with $\gamma = 1.25$, 1.00, 0.75, and 0.50. It is evident that mAIC performs the best and AIC performs the worst among all cases; mAIC produces the most stable and accurate results with different choices of γ.

In the next example, we randomly generate some models and evaluate the performance of the Dantzig selector via simulations.

Example 6. As in Example 5, we generate data from (5) where X is the 14×23 matrix given in Table 3 and $\epsilon \sim N(0, 1)$. We consider five cases for β. There are i active factors for case i, $1 \leq i \leq 5$. For each case, we generate 500 models where 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 model, we generate data 100 times according model (5) and obtain the true model identification rate (TMIR) and the average model size. Table 7 gives the summary statistics of these two quantities among 500 models. In the simulations we fix $\gamma = 1$ and choose δ according to mAIC.

The Dantzig selector method is very effective in identifying 1 active factor; the TIMR ranges from 96% to 100% and the average model size ranges from 1 to 1.04. The method is still effective in

identifying 2 active factors; the TMIR ranges from 78% to 100% and the average model size ranges from 1.86 to 2.08. The method is less effective in identifying 3 or more active factors. The median TMIR values are still very high and the median model sizes are still accurate for 3 or 4 active factors. However, the minimum TMIR values are 0%, suggesting that the method fails to identify a few models. The situation becomes worse for 5 active factors; although the median TMIR is 86%.

The Dantzig selector method does an excellent job in identifying one active factor in both simulations. This is supported by the theory developed by Candes and Tao (2007), which roughly says that the probability of correctly identifying one active factor is high when the correlations between the variables are small. However, their theory says nothing about the performance of the Dantzig selector when there are more than one active factors, because the supersaturated design does not meet the required uniform uncertainty condition. This partially explains why the method fails to identify some models with 3 or more active factors, where factor sparsity would be questionable with 23 factors and only 14 runs.

6 Concluding Remarks

This paper studies the Dantzig selector for selecting active effects in supersaturated designs. We propose a graphical procedure and an automatic variable selection method to accompany with the Dantzig selector. The graphical procedure is recommended in practice and the automatic method, like other automatic methods, should be used with caution. Simulation shows that the Dantzig selector method performs well compared to existing methods in the literature and is more efficient at estimating the model size.

A modified AIC is proposed for model selection. It works well for the examples and 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 the Dantzig selector are as follows. First, the Dantzig selector has a profound theory. Candes and Tao (2007) proved that the Dantzig selector is able to perform an ideal model selection when some uniform uncertainty conditions are fulfilled. Second, the Dantzig selector 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. Linear programming algorithms are available in many software and packages, like R, Matlab, Mathematica, etc., making it easy to program and use the Dantzig selector. Third, the Dantzig selector is able to handle a large number of factors in two-level, multi-level and mixed-level experiments. Candes and Tao (2007) applied the Dantzig selector to an experiment with up to 200 active factors among 5,000 binary factors and 1,000 observations.

Acknowledgments

The research was supported by National Science Foundation grants DMS-0505728 and DMS-0806137. The authors thank an associate editor and two referees for their criticisms and constructive comments that lead to an improvement of the paper.

References

[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] Cai, T.T., Lv, J., 2007. Discussion: The Dantzig selector: statistical estimation when p is much larger than n. Annals of Statistics 35, 2365–2369.

[10] Candes, E.J., Tao, T., 2007. The Dantzig selector: statistical estimation when p is much larger than n. Annals of Statistics 35, 2313–2351.

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

[12] 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.

[13] 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.

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

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

[16] Hoerl, A.E., Kennard, R.W., 1970. Ridge regression: applications to nonorthogonal problems. Technometrics 12, 69-82.

[17] Holcomb, D.R., Montgomery, D.C., Carlyle, W.M., 2003. Analysis of supersaturated designs. Journal of Quality Technology 35, 13-27.

[18] Hurvich, C.M., Tsai, C.-L., 1989. Regression and time series model selection in small samples. Biometrika 76(2), 297–307.

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

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

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

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

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

[24] Lin, D.K.J., 1999. Supersaturated designs. Encyclopedia of Statistical Sciences, updated volume 3, Wiley, New York, 727-731.

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

[26] 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.

[27] 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.

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

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

[30] 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.

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

[32] 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.

[33] 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.

[34] 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.

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

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

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

[38] 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.

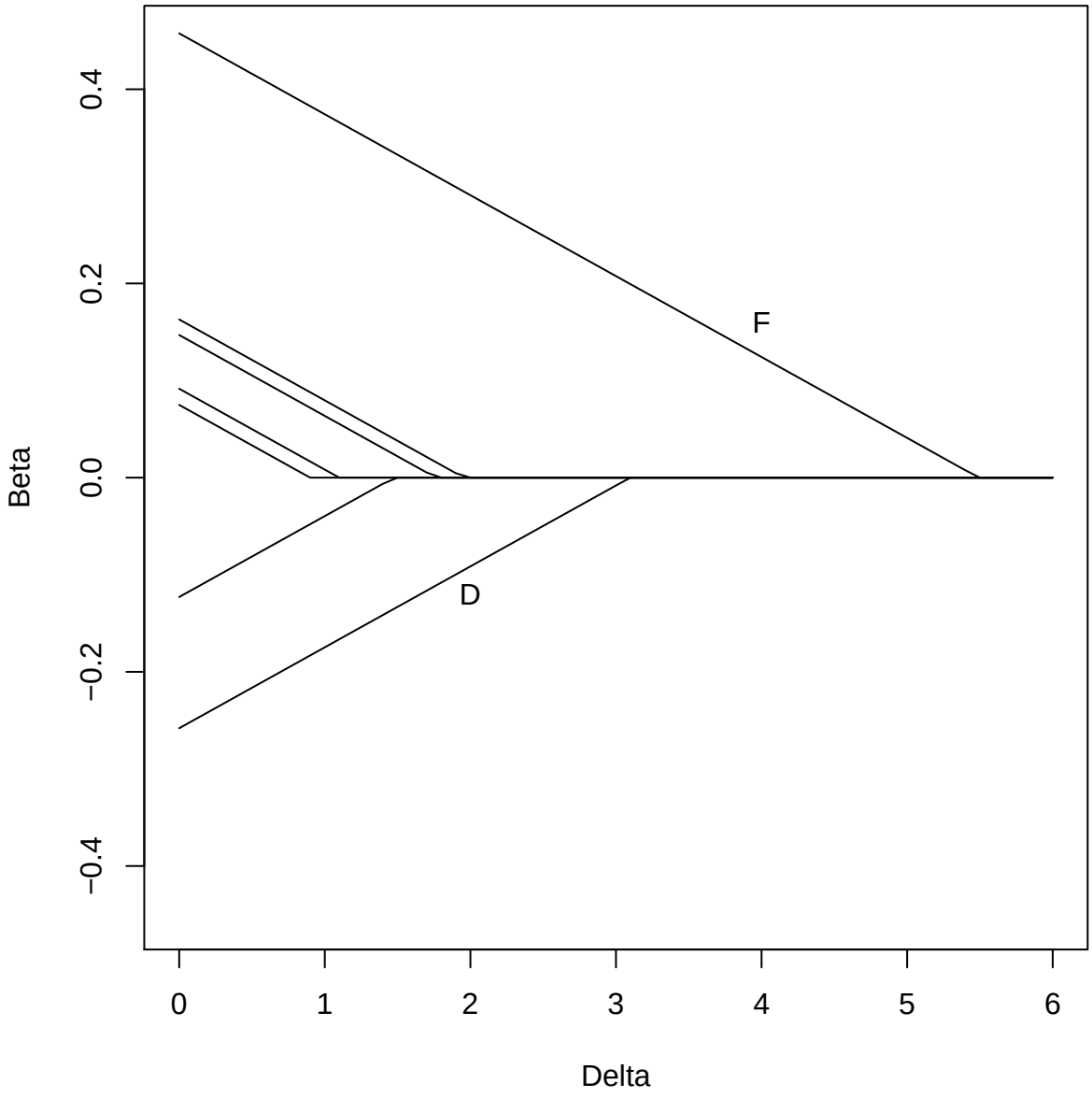

Figure 1: Profile plot for the cast fatigue experiment without interactions. The model includes 7 main effects.

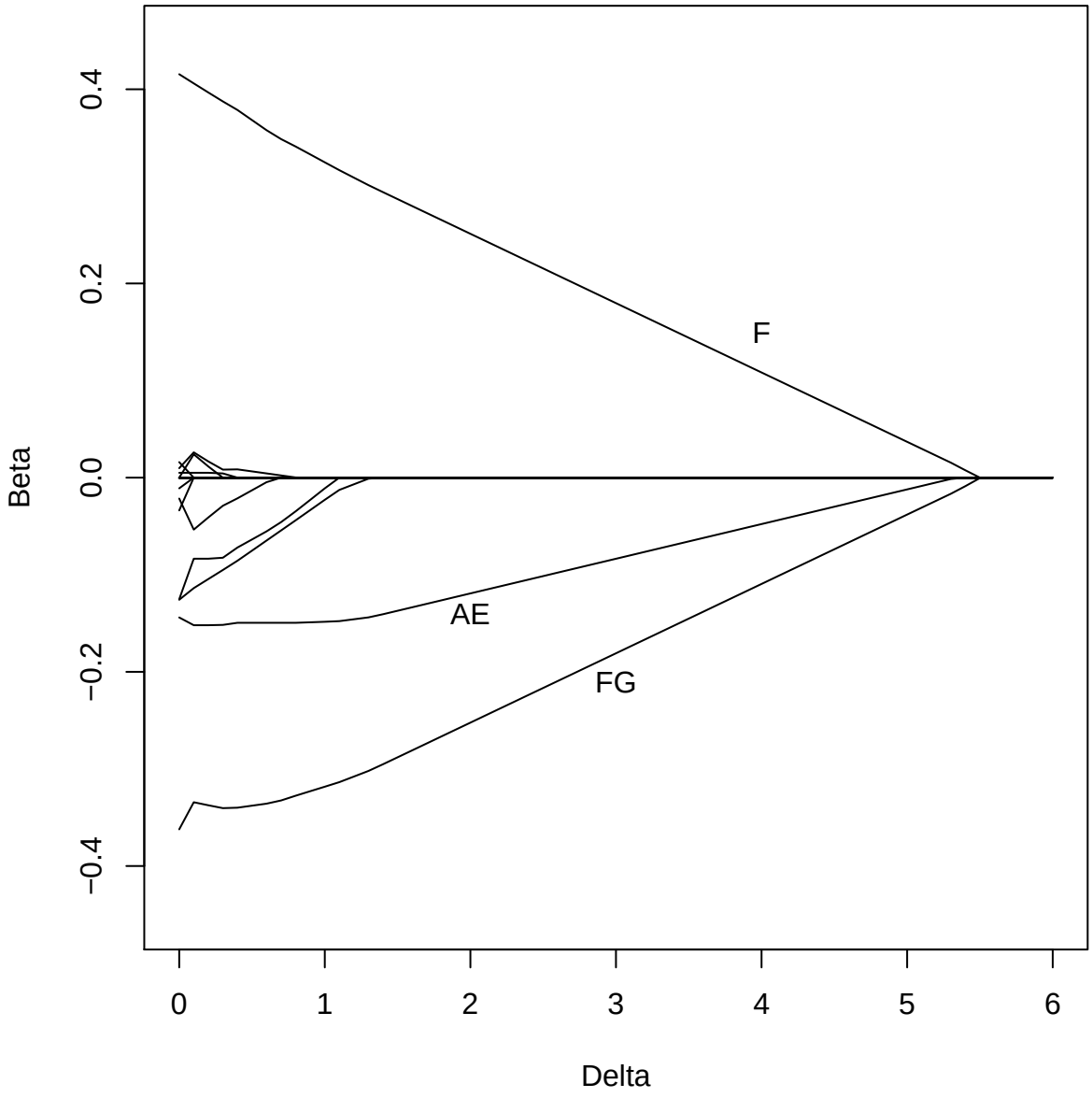

Figure 2: Profile plot for the cast fatigue experiment with interactions. The model contains 7 main effects and 21 two-factor interactions.

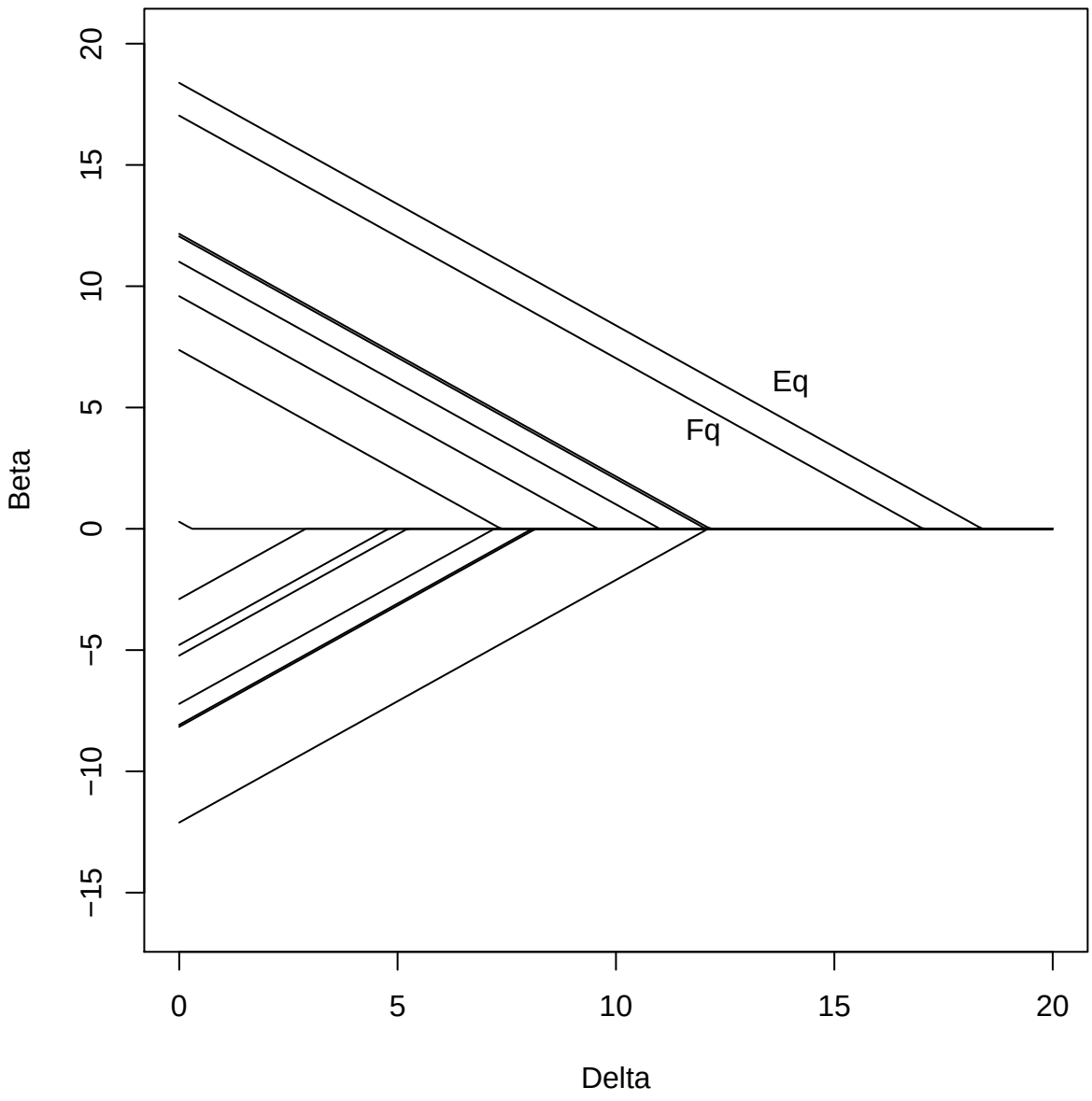

Figure 3: Profile plot for the blood glucose experiment without interactions. The model contains 15 linear and quadratic terms.

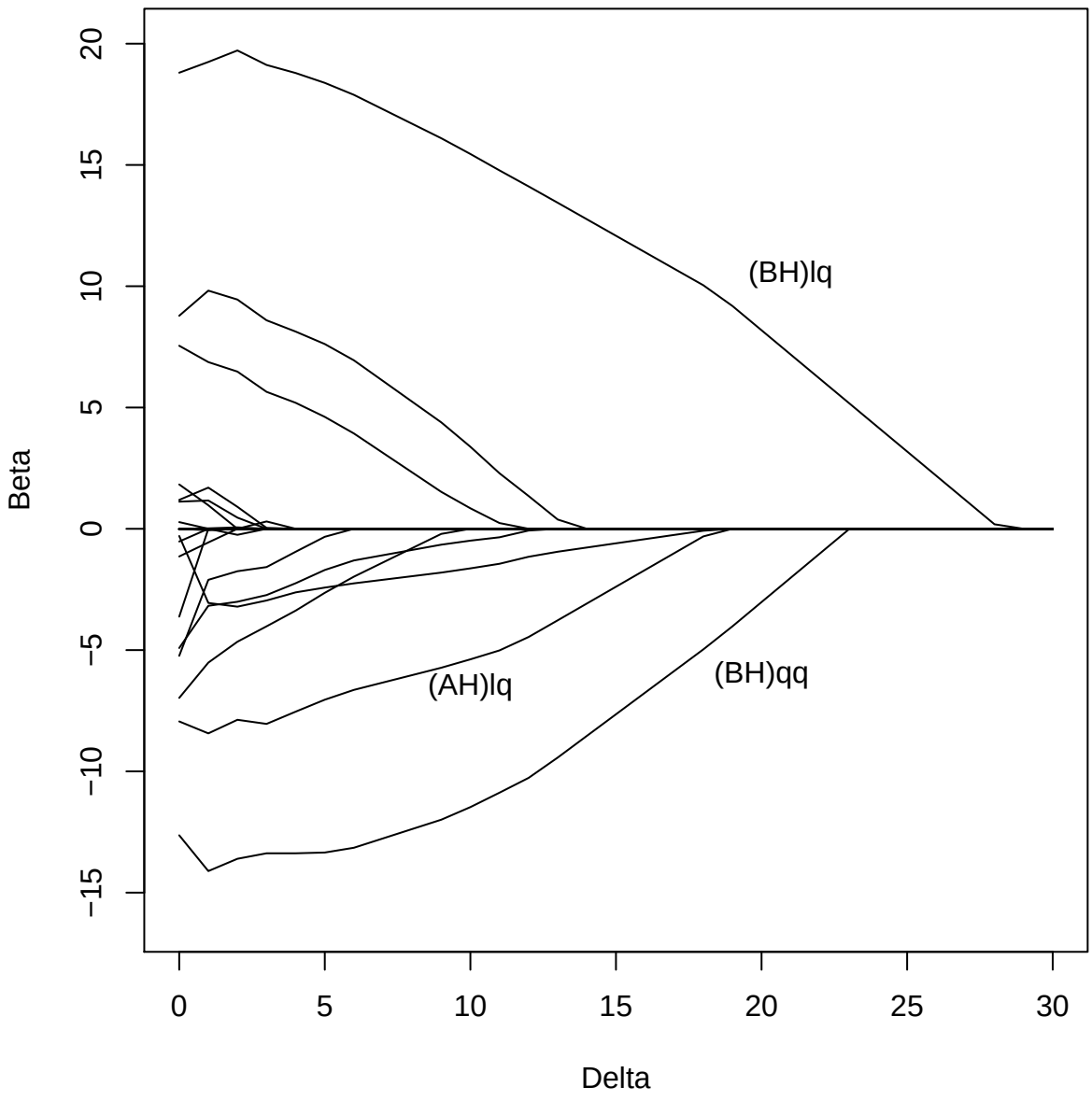

Figure 4: Profile plot for the blood glucose experiment with interactions. The model contains 15 linear and quadratic terms and 98 two-factor interaction terms.

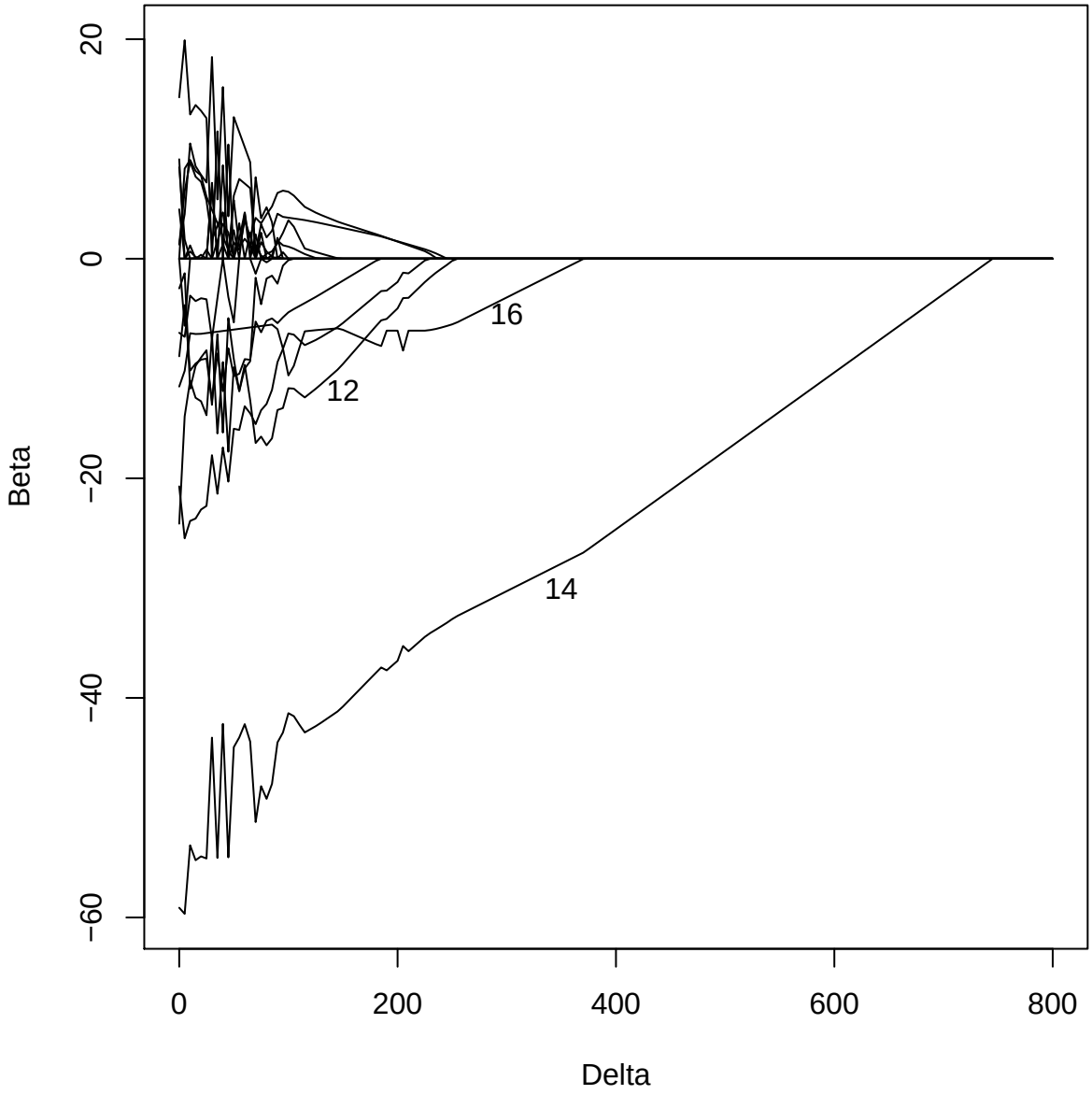

Figure 5: Profile plot for the Lin (1993) data. The model contains 23 main effects.

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	100.58
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).

	Factors																							Response
Run	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	Y
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: Comparison of information criteria in Example 4

Model	Terms	p	RSS	R^2	AIC	cAIC	mAIC
1	F	1	3.132	44.5%	−14.12	−12.79	−14.12
2	F, D	2	2.333	58.7%	−15.65	−12.65	−11.65
3	F, FG	2	0.6066	89.3%	−31.82	−28.82	−27.82
4	F, FG, AE	3	0.2673	95.3%	−39.65	−33.94	−27.65
5	F, FG, AE, D, EF	5	0.03568	99.4%	−59.82	−43.02	−19.82
6	F, FG, AE, D, EF AD, DG, A, AB	9	0.001167	99.98%	−92.86	127.14	51.14

Table 5: Comparison of simulation results in Example 5

Case	Method	TMIR	SEIR	Median	Mean
I	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 ($\gamma = 1$)	99.4%	100%	1	1.0
II	SSVS(1/10,500)	8.6%	30.0%	3	4.7
	SSVS(1/10,500)/IBF	8.0%	28.0%	3	4.2
	SCAD	75.6%	98.5%	3	3.3
	PLSVS (m=1)	76.4%	100%	3	3.3
	Dantzig Selector ($\gamma = 1$)	84.4%	85.3%	3	2.9
III	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 ($\gamma = 1$)	79.1%	91.2%	5	5.1

Table 6: Summary of simulation results in Example 5

		TMIR			Average Size		
Case	γ	AIC	cAIC	mAIC	AIC	cAIC	mAIC
I	1.25	99.9%	99.9%	100%	1.001	1.001	1.000
	1.00	99.0%	99.1%	99.4%	1.010	1.009	1.006
	0.75	90.2%	91.3%	94.7%	1.105	1.091	1.054
	0.50	43.9%	50.7%	71.8%	1.843	1.701	1.310
II	1.25	68.9%	68.9%	69.1%	2.769	2.768	2.721
	1.00	79.8%	81.4%	84.4%	3.036	3.015	2.901
	0.75	64.8%	74.9%	85.6%	3.418	3.274	3.012
	0.50	21.3%	42.8%	69.6%	4.538	3.857	3.086
III	1.25	69.4%	79.8%	81.2%	5.263	5.037	4.967
	1.00	54.7%	77.1%	79.1%	5.709	5.263	5.143
	0.75	32.2%	59.9%	63.9%	6.342	5.550	5.372
	0.50	8.1%	32.2%	37.1%	7.573	6.131	5.743

Table 7: Summary of simulation results in Example 6

Case		Min	1st Quartile	Median	Mean	3rd Quartile	Max
I	TMIR	96%	99%	100%	99.5%	100%	100%
	Size	1.00	1.00	1.00	1.005	1.01	1.04
II	TMIR	78%	99%	100%	99.3%	100%	100%
	Size	1.86	2.00	2.00	2.004	2.01	2.08
III	TMIR	0%	99%	100%	95.6%	100%	100%
	Size	2.22	3.00	3.00	3.001	3.01	3.83
IV	TMIR	0%	88%	98%	84.0%	100%	100%
	Size	1.36	3.97	4.00	3.850	4.01	4.94
V	TMIR	0%	8.8%	86%	64.0%	98%	100%
	Size	1.23	4.05	4.89	4.395	5.00	6.29