A Criterion for Constructing Powerful Supersaturated Designs when Effect Directions are Known

Maria L. Weese[1], David J. Edwards[2], and Byran J. Smucker[3]

[1]Department of Information Systems and Analytics, Miami University, Oxford, OH
[2]Department of Statistical Sciences and Operations Research, Virginia Commonwealth University, Richmond, VA
[3]Department of Statistics, Miami University, Oxford, OH

Abstract

As a criterion for selecting supersaturated designs, we suggest minimizing the variance of the pairwise inner products of the information matrix, subject to a constraint on the $E(s^2)$-efficiency as well as a requirement that the average correlation between the columns is positive. We call these designs constrained positive $Var(s)$-optimal and argue that if the direction of the effects can be specified in advance, these designs are more powerful to detect active effects than other supersaturated designs while not substantially increasing Type I error rates. These designs are constructed algorithmically, using a coordinate-exchange algorithm that exploits the structure of the criterion to provide computational advantages. We also demonstrate that, for the simulation scenarios considered, misspecification of the effect directions will, at worst, result in power and Type I error rates in line with standard supersaturated designs.

Keywords biased designs, constrained $Var(s)$, coordinate exchange, Dantzig selector, forward selection, optimal design;

1 Introduction

Supersaturated experiments—classically defined as those in which the number of runs is no more than the number of factors—have been constructed using a wide variety of criteria, the most venerable being $E(s^2)$. These designs (Booth and Cox, 1962; Lin, 1993; Wu, 1993) are

pleasingly intuitive: they produce designs with small pairwise column correlations. They also minimize, for each non-intercept least squares parameter estimate, the bias due to the presence of other non-zero effects (Lin, 1995). Although the $E(s^2)$ criterion has received a majority of the attention in the supersaturated design literature, alternatives do exist that depart from a focus on pairwise dependencies (see, e.g., Deng et al. (1996) and Deng et al. (1999)).

Recent attention, both within academia (Marley and Woods, 2010; Draguljić et al., 2014; Weese et al., 2015) and from industry (Brenneman, 2014; Scinto et al., 2014), has focused researchers on answering the following: How does one know that the chosen supersaturated design will find the correct active factors? Motivated by this question, we have developed a class of supersaturated designs that exploit the bias due to other non-zero effects and demonstrate that under the condition that the effect directions are known in advance, our designs have an increased power to detect active effects (i.e. the factors that produce the largest change in the response) versus standard supersaturated design construction criteria.

Other studies (Marley and Woods, 2010; Weese et al., 2015) have suggested that when evaluated based on the power to detect active effects, neither $E(s^2)$ nor any of the other leading criteria, including Bayesian $\mathcal{D}$-optimality (Jones et al., 2008) and model-robustness approaches (Jones et al., 2009), distinguish themselves positively from the others. Based on what we know up to this point about the performance of the classical criteria, we are indifferent to the choice of supersaturated design criteria, except to the extent that some are more easily generated than others. This appears to be the case for experiments of varying size and for both traditional (forward selection) and modern (Dantzig selector) variable selection methods.

The $E(s^2)$ criterion was taken from Booth and Cox (1962), though these authors actually wrote about $Var(s)$ as a criterion, apparently assuming that $E(s) = 0$. Weese et al. (2015) explored constrained $Var(s)$-optimal designs, which minimize the variance of the off-diagonals of the information matrix subject to a specified level of $E(s^2)$-efficiency. They found that over a range of experimental scenarios they performed no better than other

supersaturated designs in terms of power. In this article, we revisit and refine this approach by imposing the additional constraint that the constrained $Var(s)$-optimal designs must have a positive average column correlation. We refer to these designs as constrained positive $Var(s)$-optimal, and argue that if the sign of the main effects can be specified in advance, the least squares estimates will have a positive bias. Via simulation, we then demonstrate that for a wide variety of experiments and effect sizes, across two versions of forward selection as well as the Dantzig selector, these designs are significantly more powerful than those designs constructed using standard criteria (i.e. Bayes D-optimal and $E(s^2)$-optimal). Furthermore, Type I error rates for the proposed designs are not any worse than those for other designs.

The requirement that the sign of the main effects must be specified is crucial to the increased power of the constrained positive $Var(s)$ designs. This is an assumption that has often been made in the group screening (e.g. Watson, 1961; Draguljić et al., 2014) and sequential bifurcation (e.g. Bettonvil and Kleijnen, 1996) literature. Indeed, in the planning stages of many studies, it is often the case that the experimenter has prior information about effect directions. Draguljić et al. (2014) notes that "[in] many experiments, for example, in engineering and chemistry, experts are often able to provide information on the 'direction' of each main effect based on scientific knowledge or previous experience." For instance, consider an example from the chemical manufacturing industry, in which several oxidation inhibitors are being investigated for their relationship with a measure of the viscosity increase of the lubricant they are added to. Since oxidation causes oil thickening and given that the inhibitors have been formulated expressly to reduce oxidation, it is plausible that the experimenter could identify that the presence of an oxidation inhibitor would suppress viscosity increase, even before the experiment is conducted. We show empirically, however, that even if signs are grossly misspecified, the constrained positive $Var(s)$-optimal designs do not fare any worse than other designs, even though the designs have more column correlation, on average. This work extends the work of Weese et al. (2015) by (1) introducing the constrained positive $Var(s)$ criterion and (2) establishing the superiority of the new constrained positive $Var(s)$ criterion in terms of power when the effect directions can be

specified in advance, at least for the range of simulations undertaken.

In the following section we define the setting, the criterion of interest, and the algorithm used to generate the designs. In Section 3 we provide arguments to justify our approach, including a simple theoretical explanation as well as one based upon properties of the new designs. Section 4 provides direct evidence that, for the wide variety of experimental scenarios simulated, the proposed designs enjoy increased power, have Type I error rates that are no worse then other designs, and are robust to the misspecification of the effect directions. Finally, in Section 5 we provide some discussion and conclusions.

2 The Constrained $Var(s)$ Criterion

2.1 Assumed Model

The implicit model of interest when executing a supersaturated experiment for k factors is the main effects model

$$Y = X\beta + \epsilon, \tag{1}$$

where Y and ϵ are $n \times 1$ vectors, β is a $(k+1) \times 1$ vector, X is $n \times (k+1)$, and the elements of the error vector are independent with $E(\epsilon_i) = 0$ and $Var(\epsilon_i) = \sigma^2$. Since $n < k + 1$, $M = X'X$ is singular and consequently the analysis must assume that the number of active effects is no greater than n. In this work we consider only two level designs with values 1 and -1.

2.2 Constrained Positive $Var(s)$ Criterion

Standard supersaturated design construction methods produce designs with off-diagonal elements of $X'X$ that are small in absolute value, which in turn produces relatively small correlations between columns. For such designs, there is an equal likelihood that the column correlations will be negative or positive. An example of such a standard criterion is $E(s^2)$,

i.e.,

$$E(s^2) = \frac{2}{k(k-1)} \sum_{2 \leq i < j} s_{ij}^2, \tag{2}$$

where s_{ij} is the $(i,j)th$ element of $X'X$. This criterion, when minimized, clearly produces s_{ij}^2 that are as small as possible on average.

For the designs in this paper, we take a different approach building upon the constrained $Var(s)$-optimal designs of Weese et al. (2015). Instead of minimizing the sum of the squares of the off-diagonal elements of $X'X$, we minimize the variance of the off-diagonal elements while forcing the average of those elements to be positive. This approach allows the average of the off-diagonal elements to be larger than that of the $E(s^2)$-optimal design, but we prevent the average from becoming too large by imposing a constraint on the design's $E(s^2)$-efficiency. Specifically, we choose a design D which minimizes

$$\begin{aligned} Var(s) &= E(s^2) - E(s)^2 \\ &= \frac{2}{k(k+1)} \sum_{1 \leq i < j} s_{ij}^2 - \left(\frac{2}{k(k+1)} \sum_{1 \leq i < j} s_{ij} \right)^2 \end{aligned} \tag{3}$$

subject to

$$E_{E(s^2)} = \frac{E(s^2)(D^*)}{E(s^2)(D)} > c \tag{4}$$

$$E(s) > 0$$

where D^* is the $E(s^2)$-optimal design and c is a user-specified efficiency that determines how close to $E(s^2)$-optimal the design must be. Note again that we choose a design to minimize this criterion while forcing $E(s) > 0$ and constraining the $E(s^2)$-efficiencies. This criterion bears some similarities to that proposed by Booth and Cox (1962), though they assumed (a) that the designs should be balanced; (b) $E(s) = 0$, which leads to the classical $E(s^2)$ criterion; and (c) the intercept column should not be included in the calculation.

Although we desire little variation among the off-diagonal elements of $X'X$, a design selected via minimization of $Var(s)$ alone would allow for extremely large column correla-

tions. For example, one could construct an unconstrained $Var(s)$-optimal design such that all column correlations equal 1. Hence, our proposed constraint on $E(s^2)$-efficiency. We use $c = 0.8$ in our simulations; a justification for this is given in Section 4.3.

To ensure an $E(s^2)$-efficiency greater than c, we must first find the $E(s^2)$-optimal design, D^*. We construct these designs algorithmically, using a similar procedure to that described in Section 2.3. Unlike the classical form of $E(s^2)$ given in (2), we include the intercept as a column to be considered and do not restrict these designs to be balanced in the sense that we allow effect columns to have different numbers of -1's and 1's. Note that a recent paper (Jones and Majumdar, 2014) studies these unbalanced $E(s^2)$-optimal designs and concludes that they generally produce more precise estimates of main effects than their balanced predecessors. We also note, however, that Weese et al. (2015) did not find these unbalanced $E(s^2)$-optimal designs to be superior in terms of power to detect active effects.

Further justification for the proposed criterion, particularly in light of the requirement that effect directions be specified in advance, will be provided in Section 3.

2.3 Algorithm

To construct the constrained positive $Var(s)$ designs, we use an algorithm following coordinate exchange (Meyer and Nachtsheim, 1995), that goes roughly as follows:

1. Randomly construct a $n \times k$ initial designs until one is found for which $E(s) > 0$.

2. Construct an initial design that satisfies (4). Iterate from one coordinate to the next, row by row, and if the last row is reached begin at the first row again. At each coordinate consider the impact on the design when the current coordinate value is exchanged (i.e. multiplied by -1, since we are considering only two levels). Make an exchange when $E_{E(s^2)}$ is increased while $E(s) > 0$ is maintained. When $E_{E(s^2)} \geq c$, go to Step 3.

3. Iterate from one coordinate to the next, row by row, and if the last row is reached begin

at the first row again. At each coordinate, consider the change in the $Var(s)$ criterion, the $E(s^2)$-efficiency, and $E(s)$ when the current coordinate value is multiplied by -1.

(a) If exchanging the coordinate improves $Var(s)$ while maintaining $E_{E(s^2)} \geq c$ and $E(s) > 0$, make the exchange.

(b) Otherwise, continue to the next coordinate.

Continue with this step, iterating through the design until convergence to a local optimum.

Global optimality is not guaranteed for a single execution of this algorithm, so several algorithm tries should be run from different initial designs and the best chosen. In this work we use 100 random starts, in an attempt to maintain a balance between near-optimality and time-efficiency.

For steps 2 and 3, instead of using a standard rank-1 updating formula for M as is typical in coordinate exchange algorithms (Meyer and Nachtsheim, 1995), we can update $E(s^2)$ and $Var(s)$ via arithmetic operations as follows. Recall that in this article we consider only two-level designs with levels -1 and 1. Let X and M be the design matrix and information matrix for an arbitrary two-level design, and $s_{ij} = [M]_{ij}$. Suppose $x_{ij} = [X]_{ij}$ is updated such that $x_{ij}^{\text{up}} = (-1)x_{ij}$. Denote M^{up} as the information matrix for X^{up}. Let $T = \sum_{i'<j'} s_{i'j'}$ for M so that $T = \binom{k+1}{2} E(s)$, and T^{up} be the analogous sum for M^{up}.

When x_{ij} is replaced by x_{ij}^{up}, only elements in row j and column j of M will change. Note that an arbitrary element of the information matrix is $[M]_{uv} = \sum_{i'=1}^{n} x_{i'u}x_{i'v}$ so that each element of M is composed of n terms and for column $[M]_{.j}$, a single one of those terms includes x_{ij}. That particular term, in each element of column j of M (the exception is $M_{jj} = \sum_{i'} x_{i'j}^2$, which will be unchanged because $x_{ij}^2 = (-x_{ij})^2$) determines the change in T as follows:

$$T^{\text{up}} = T + 2x_{ij}^{\text{up}} \sum_{k \neq j} x_{ik}. \tag{5}$$

That is, if x_{ij}^{up} has the same sign as x_{ik}, T increases by 2; if the sign is not the same, T decreases by the same amount. Then,

$$[E(s)]^{\text{up}} = \frac{T^{\text{up}}}{\binom{k+1}{2}}. \tag{6}$$

To update $E(s^2)$ if x_{ij} is exchanged for $x_{ij}^{\text{up}} = (-1)x_{ij}$, let $Q = \sum_{i'<j'} s_{i'j'}^2$ and Q^{up} be the analogous sum for M^{up}. Then,

$$Q^{\text{up}} = Q + \left[\sum_{i' \neq j} (s_{i'j} + c_{i'})^2 - \sum_{i' \neq j} s_{i'j}^2 \right], \tag{7}$$

where $c_{i'} = 2x_{ij}^{\text{up}} x_{ii'}$ is the change in T due to the i'^{th} element of $[X]_{i.}$. Then $[E(s^2)]^{\text{up}} = \frac{Q^{\text{up}}}{\binom{k+1}{2}}$ and

$$[Var(s)]^{\text{up}} = [E(s^2)]^{\text{up}} - \{[E(s)]^{\text{up}}\}^2. \tag{8}$$

3 Justification of the Constrained Positive $Var(s)$ Criterion

Lin (1995) shows that the expected values of the least squares estimates obtained via componentwise regression are related to the column correlations in a simple way. That is,

$$E(\hat{\beta}_i) = \beta_i + \sum_{i \neq j} r_{ij} \beta_j, \tag{9}$$

where $\hat{\beta}_i$ is the least-squares estimate of β_i, r_{ij} is the correlation between columns i and j.

Now, the constrained positive $Var(s)$-optimal designs have, on average, $E(r_{ij}) > 0$ by construction. (Note that by forcing the average to be positive we are implicitly encouraging the individual elements to be positive, although this is not explicitly required.) Furthermore, we assume that the sign of each effect can be specified, so that each column in X can be appropriately transformed such that the value of each parameter is positive. Thus,

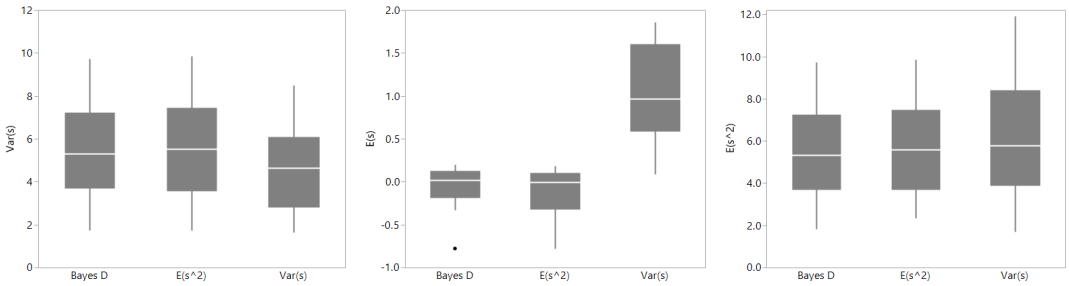
(9) clearly shows that parameter estimates under these conditions will be inflated in the correct direction rather than being accidentally shrunk. Lin (1995) remarks that in order for an active factor to be identified, it "must have an effect too large to be masked by the experimental error and the combined effects of unimportant factors." Here, we use other active factors to make the estimate larger, in order to exaggerate its effect and increase the chance of identifying it as active.

Table 1 and Figure 1 display summaries of design characteristics (i.e., $Var(s)$, $E(s)$, $E(s^2)$, mean absolute correlation, maximum absolute correlation) across different design construction criteria for all supersaturated designs studied (Table 2). By design, constrained positive $Var(s)$-optimal SSDs have smaller $Var(s)$ and larger $E(s)$ than their Bayesian D-optimal and $E(s^2)$-optimal counterparts (Figure 1a and 1b). Interestingly, while the $Var(s)$-optimal designs have slightly larger $E(s^2)$ values, differences are less apparent among the designs constructed from the three criteria with regards to mean $|r_{ij}|$ or the max $|r_{ij}|$ (Figures 1c, 1d, and 1e). Note also that in this and subsequent comparisons of supersaturated designs, we used traditional balanced $E(s^2)$-optimal designs. This explains why Table 1 shows that the Bayesian D-optimal designs have smaller $E(s^2)$ values, on average, than the $E(s^2)$-optimal designs.

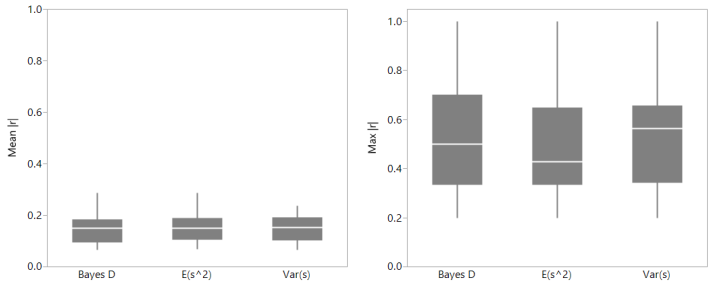
Table 1: Average values of various design characteristic measures, averaged over the 20 designs in Table 2.

	Design Criterion		
Measure	Bayesian D-optimal	$E(s^2)$-optimal	$Var(s)$-optimal
$Var(s)$	5.350	5.590	4.670
$E(s)$	−0.040	−0.110	1.020
$E(s^2)$	5.400	5.670	6.040
$E(\|r\|)$	0.149	0.156	0.151
$Max(\|r\|)$	0.512	0.499	0.525

To further strengthen our justification for the constrained positive $Var(s)$ criterion, we investigate the Frobenius norm of the alias matrix (i.e., $\|A\|_F = \sqrt{trace(A^T A)}$) across projections of supersaturated designs of the various sizes given in Table 2. For each design size, projections of up to half the run size are considered. This choice is reasonable given the



(a) $Var(s)$ (b) $E(s)$ (c) $E(s^2)$



(d) Mean $|r|$ (e) Max $|r|$

Figure 1: Comparison of design characteristics for Bayesian D-optimal, $E(s^2)$-optimal, and constrained positive $Var(s)$-optimal designs

guideline of Marley and Woods (2010) that the number of runs should be at least three times the number of active factors. Letting X_1 denote a particular design projection (plus the intercept column) and X_2 consist of all remaining columns of the design matrix, we compute $A = (X_1'X_1)^{-1}X_1'X_2$ and $\|A\|_F$ for each projection. Figure 2 shows the Frobenius norm (up to 5 factor projections) for Bayesian D-optimal, $E(s^2)$-optimal, and constrained positive $Var(s)$-optimal designs across all design sizes. From this, it is clear that the constrained positive $Var(s)$ designs produce alias matrices with larger Frobenius norms and thus, as desired, more positively biased effect estimates on average.

Of course, two objections may be raised: (1) while the estimates of active effects are inflated so are the estimates of inactive effects, and this should increase Type I error rates; and (2) since column correlations are larger, effect estimate standard errors are slightly larger on average. In the following section we demonstrate, for the simulation conditions considered, that the bump in the estimates for constrained positive $Var(s)$ designs over-

shadows this increase of standard errors and does indeed result in greater power to detect active effects. Furthermore, Type I error rates are not significantly higher.

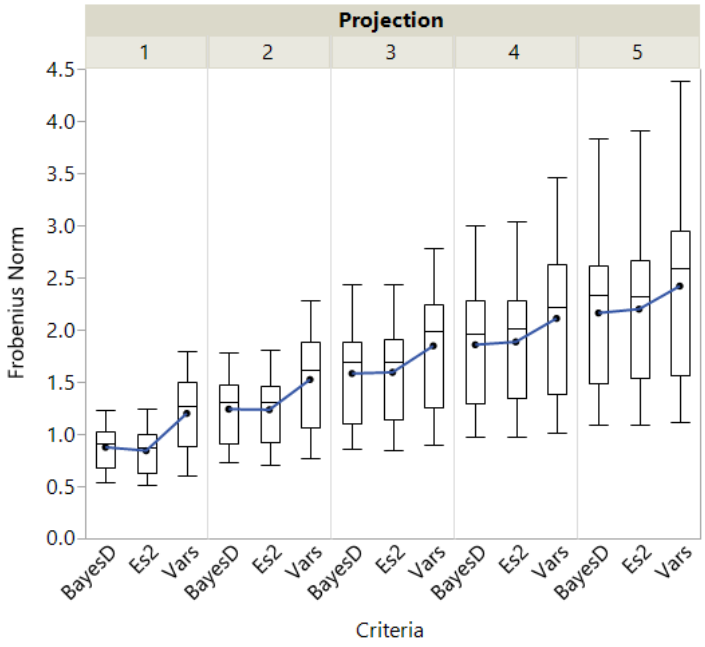


Figure 2: Frobenius norm of alias matrix across design projections

4 Simulations

In this section we use simulation to study constrained positive $Var(s)$-optimal designs and compare them to standard supersaturated designs. We consider a wide variety of supersaturated experimental situations, assuming the experimenter can specify with certainty all effect directions. Then, we study the effects of misspecifying the sign of the effects. First, however, we describe several analysis methods that we use to judge the power of various designs in our simulations, and detail the general simulation protocol that was used.

4.1 Analysis Methods

We consider three analysis methods: the Dantzig selector and two versions of forward selection. The Dantzig selector (Candes and Tao, 2007) chooses as active factors those terms

that are consistent with the data and minimize the ℓ_1 norm of the vector of estimates. These estimates can be obtained via linear programming. Phoa et al. (2009) use this procedure to analyze supersaturated designs, and evidence is growing (Marley and Woods, 2010; Draguljić et al., 2014; Weese et al., 2015) that this method dominates traditional forward selection methods, in terms of power to detect active effects. Note that the Dantzig selector requires specification of two tuning parameters, δ and γ. As in Marley and Woods (2010), we use the usual BIC statistic to select δ and choose $\gamma = 1.5$ to help control Type I error.

We also consider two versions of forward selection because they are fast, easy-to-understand, and commonly used in practice. We acknowledge that the literature (Westfall et al., 1997; Abraham et al., 1999; Marley and Woods, 2010, for example)—and our experience Weese et al. (2015)—points to a myriad of deficiencies with this general approach. However, we note that this article's objective is not to adjudicate analysis methods but to demonstrate that the constrained positive $Var(s)$ designs are better for a variety of analysis methods, when the effect directions can be accurately specified. Thus, we use forward selection with two different stopping criteria: $\alpha_{enter} = 0.05$ and the corrected version of Akaike's Information Criterion, $AICc = AIC + \frac{2k(k+1)}{n-k-1}$ (Hurvich and Tsai, 1989).

4.2 Simulation Protocol

We wish to test constrained positive $Var(s)$-optimal designs against Bayesian D-optimal and $E(s^2)$-optimal designs, in a wide variety of experimental scenarios. Consequently, we vary the following in our simulations:

1. Design size,(n, k). We considered twenty different sizes, from small to moderately large and with a variety of levels of supersaturation. See Table 2.

2. Experimental scenario. Following Marley and Woods (2010), we vary the number and magnitude of active factors over four scenarios. Specifically, each scenario has a active factors and for each their regression coefficient is drawn from $N(\mu, 0.2)$ for the following:

- $a = 3$, $\mu = 5$;
- $a = 4$, $\mu = 4$;
- $a = 6$, $\mu = 3$;
- $a = 9$, $\mu = 3, 5, 8, 10$ for four factors and $\mu = 2$ for five factors.

If $k < a$, the scenario is skipped for that design.

3. Design construction criteria. Here we include the classical balanced $E(s^2)$-optimal (or nearly balanced for odd n), Bayesian D-optimal, and the new constrained positive $Var(s)$-optimal designs.

4. Analysis method. As mentioned in Section 4.1, we will use the Dantzig selector as well as two versions of forward selection.

The constrained positive $Var(s)$-optimal designs were generated as described in section 2.3. The Bayesian D-optimal designs were generated using 100 random starts in JMP Pro 11.2 and a prior variance of $\tau^2 = 1$. The $E(s^2)$-optimal designs were constructed using the nonorthogonal array (NOA) algorithm of Nguyen (1996) in Gendex (`www.designcomputing.net/gendex/`). Note that in Weese et al. (2015) the unbalanced $E(s^2)$-optimal designs were not significantly better than the classical balanced $E(s^2)$-optimal designs; hence their absence in our simulation study.

The simulation also largely follows Weese et al. (2015). In each of 10,000 iterations,

1. From the columns $2, ..., k + 1$ of X, a columns are randomly assigned as the active factors. The coefficients are obtained by sampling from $N(\mu, 0.2)$.

2. The remaining $(k - a)$ columns (inactive effects) are assigned a coefficient from $abs(N(0, 0.2))$. (Note that we also performed simulations under other inactive effect conditions, including setting the inactive effects to be exactly zero, with similar and favorable results for the constrained positive $Var(s)$-optimal designs. Ultimately, we elected to assign inactive effects a small value from a normal distribution in order

Table 2: Supersaturated design sizes considered in our simulations, and a measure of supersaturation.

n	k	k/n
5	10	2.00
6	10	1.67
6	11	1.83
7	8	1.14
8	12	1.50
9	12	1.33
9	18	2.00
10	11	1.10
10	15	1.50
12	26	2.17
14	23	1.64
14	24	1.77
16	30	1.88
17	18	1.05
18	22	1.22
19	23	1.21
20	34	1.70
24	34	1.42
26	31	1.19
31	33	1.06

to create a more realistic screening setting. The alternative simulation results are available in the supplementary materials at `www.asq.org/jqt`.)

3. The signs of the all of the coefficients were determined by sampling from $Bernoulli(p)$ where p is chosen to be either 0, 0.25, or 0.5. For simulations testing the effectiveness of designs when the signs of all coefficients are correctly specified, $p = 0$. The other values of p are used for the simulations testing the robustness of the constrained positive $Var(s)$-optimal designs to effect direction misspecification. For instance, $p = 0.25$ indicates that each coefficient had a 25% chance of being changed from positive to negative.

4. The response vector is generated from the model in (1) with errors (ϵ_i) generated from $N(0, 1)$.

5. The declared "active" factors are determined by one of the analysis methods.

At the end of 10,000 iterations the proportion of correctly identified active effects (power), the proportion of the inactive effects identified as active (Type I error) and the proportion of times all of the truly active effects were chosen (coverage) were reported.

4.3 The choice of c for Constructing Constrained Positive $Var(s)$-optimal designs

To illustrate the differences between the choices of c, consider Table 3. It is clear that $Var(s)$ is smaller, on average, as c decreases. However, this comes at the expense of larger values for $E(s^2)$. We performed simulations to compare the power and Type I error rates of constrained positive $Var(s)$-optimal designs for $c = (0, 0.2, 0.4, 0.6, 0.8, 1)$. Our analysis showed no statistical difference in Type I error rates for the differing values of c, but found that $c = 1$ or $c = 0$ was significantly inferior to designs constructed with c between 0.2 and 0.8 in terms of power. The choice of $c = 0.8$ provides a balance between Type I error and power.

Table 3: Objective function values and mean and maximum absolute correlations across all design sizes for various values of c

	Criterion Value				
c	$Var(s)$	$E(s)$	$E(s^2)$	Mean $\lvert r\rvert$	Max $\lvert r\rvert$
0.000	3.394	4.141	27.776	0.378	0.701
0.200	3.810	3.242	17.185	0.288	0.698
0.400	4.246	2.503	11.231	0.225	0.574
0.600	4.621	1.745	8.241	0.175	0.527
0.800	4.948	1.003	6.272	0.149	0.509
1.000	5.226	0.370	5.391	0.138	0.517

4.4 Simulation Results when Effect Directions are Known ($p = 0$)

We first analyze the optimal case of no sign misspecification ($p = 0$); that is, all of the coefficient signs are known. Note that this is equivalent to all coefficient signs being positive. Figure 3a displays the results of these simulations for all designs sizes, across all scenarios and analysis methods. This figure clearly shows the dominance of the constrained positive

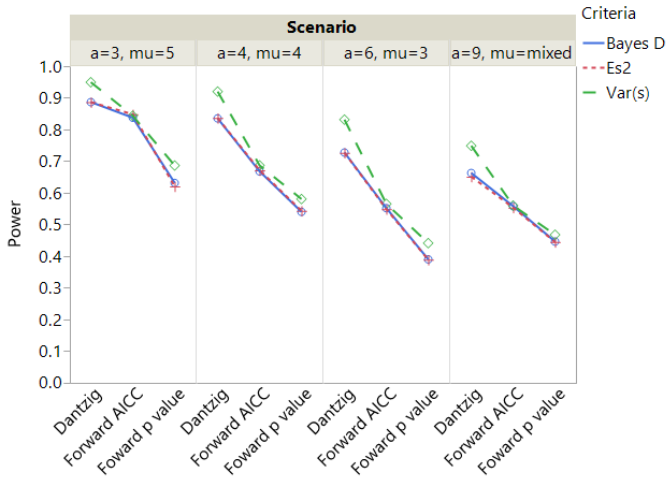
$Var(s)$-optimal designs with regards to power, and the results are clearest when the analysis method is either the Dantzig selector or Forward selection using $\alpha_{enter} = 0.05$. Constrained positive $Var(s)$-optimal designs perform particularly well for the last two, more difficult, scenarios ($[\mu = 3, a = 6]$ and $[\mu = \text{mixed}, a = 9]$) using the Dantzig selector.

Figure 3b displays a similar plot for the average Type I error rate across all design sizes. There appears to be little difference between the Type I error rates of the design criteria with the exception that the combination of the constrained positive Var(s)-optimal designs have lower Type I error using the Dantzig selector than the other two design types. Based on the simulation results, Forward selection using AICc has the overall highest Type I error rate compared to the other analysis methods.

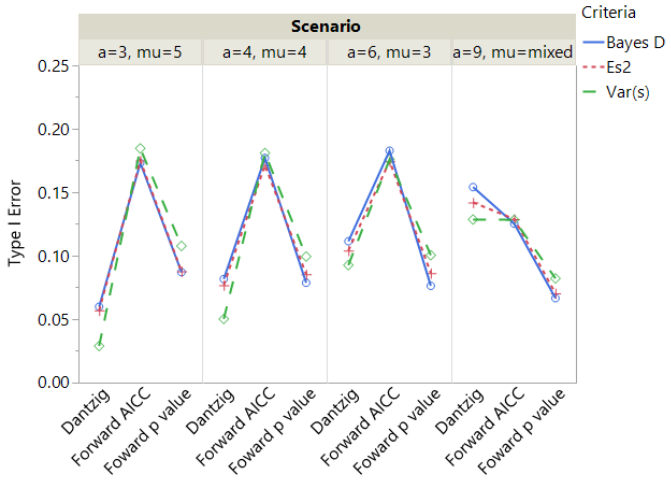
The difference between the constrained positive $Var(s)$-optimal designs and the standard designs persists when coverage is considered. Figure 4 shows the average coverage across all twenty design sizes for each analysis method and scenario. The constrained positive $Var(s)$ criteria leads to higher average coverage and this is especially evident in the first three scenarios using the Dantzig selector.

To strengthen the conclusions from the simulation, we consider a formal statistical analysis using a measure of the power and Type I error as responses and treating the simulation variables (number of experimental runs, n; number of factors, k; design construction criteria, Criteria; simulation scenario, Scenario; and analysis method, Method) as experimental factors. We note that using the raw power and Type I error proportions from the simulations resulted in ill-fitting models as determined via usual residual diagnostics. Consequently, we fit a typical regression model using the square root of the count values of either the truly active factors declared active (power count) or the truly inactive factors declared active (Type I error count) as the response. Model diagnostics for both models indicate a reasonably good fit.

Figure 5 displays the Criteria main effect (p-value < 0.0001) from the model for power counts ($R^2 \approx 95\%$). A Tukey multiple comparison procedure indicates that the constrained



(a) Power



(b) Type I Error

Figure 3: Average power and Type I error based on simulations when $p = 0$ over all twenty designs sizes, shown for each of the three analysis methods and each of the four scenarios.

positive $Var(s)$-optimal designs have the highest overall power, significantly better than both Bayesian D-optimal and $E(s^2)$-optimal designs. We note that the only significant interaction involving the design construction criteria is the interaction with analysis method; see Figure 5. Using a Tukey p-value adjustment to compare means, we find that, as seen previously in Figure 3a, the constrained positive $Var(s)$-optimal designs using the Dantzig

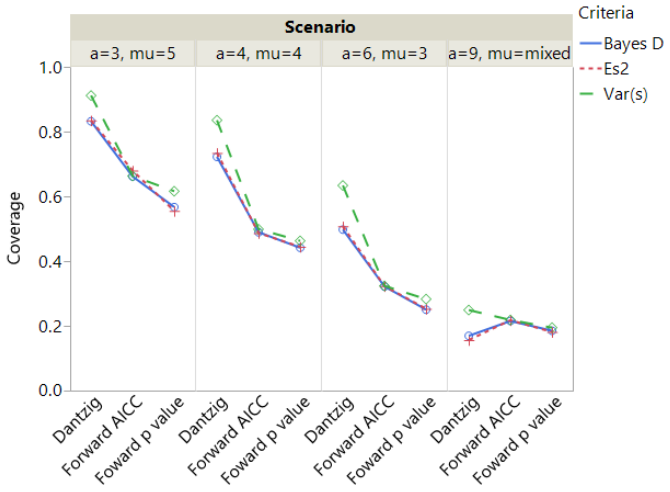
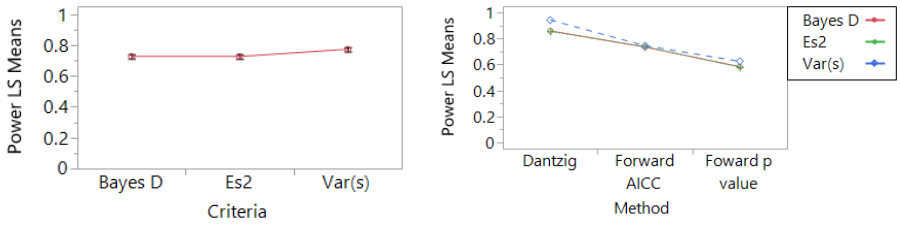


Figure 4: Average coverage based on simulations when $p = 0$ over all twenty designs sizes, shown for each of the three analysis methods and each of the four scenarios.

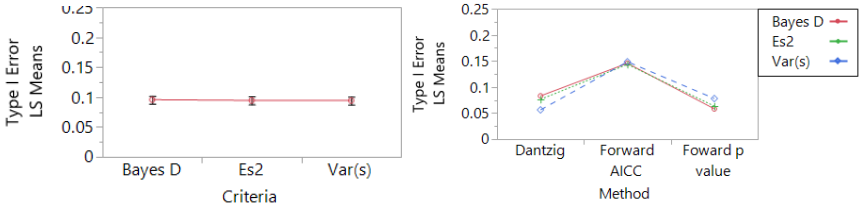


(a) Criteria Main Effect Plot (b) Criteria*Method Interaction

Figure 5: Criteria main effect and interaction plot of power when effect signs are known. Note the y-axis is adjusted to show proportions.

selector have a significantly higher average power than any other criteria/method combination.

A formal analysis of the Type I error rate counts ($R^2 \approx 93\%$) fails to produce a significant main effect of design construction criteria. There is a significant two-factor interaction involving the design construction criteria and analysis method; see Figure 6b. A Tukey analysis shows that Forward Selection using AICc for all criteria has the highest Type I error rate.



(a) Criteria Main Effect Plot (b) Criteria*Method Interaction

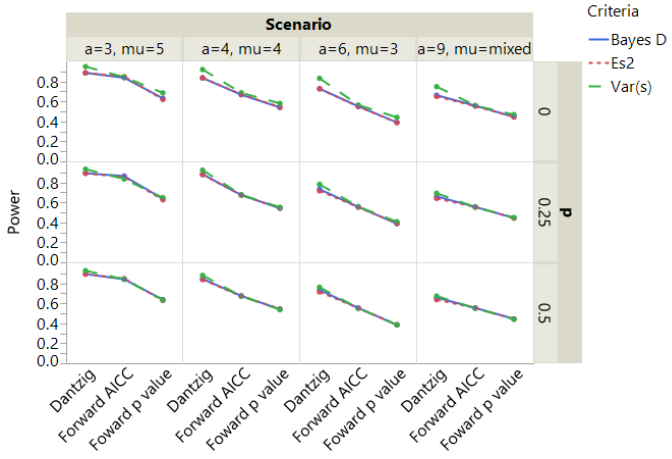
Figure 6: Criteria main effect and interaction plot for Type I error rate when effect signs are known. Note the y-axis is adjusted to show proportions.

4.5 Effect of Misspecifying Effect Directions

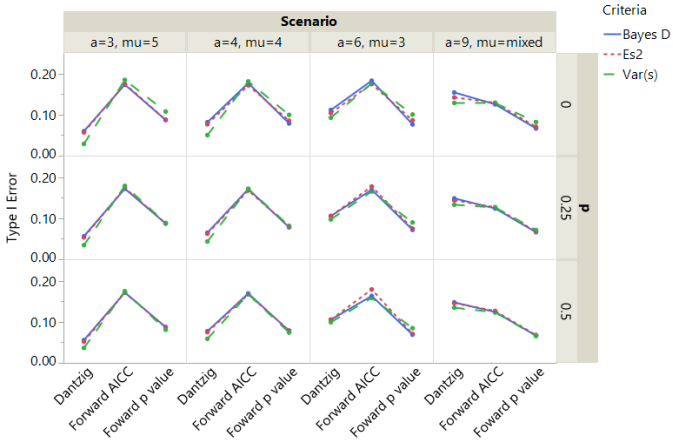
Based on the simulations in the previous section, the constrained positive $Var(s)$ criterion produces designs with higher power and similar Type I error rates when all of the effect directions are correctly specified. Previous results (Weese et al., 2015) suggest that when effect directions are randomly assigned, the constrained positive $Var(s)$ criterion loses its advantage, though in Weese et al. (2015) the positive $E(s)$ constraint had not been imposed. To clarify further, we perform simulations in which we systematically vary the proportion of misspecified effects.

Figures 7a and 7b show the average power and Type I error for each of the criteria, for each value of p and each scenario across all design sizes. When comparing the rows of Figure 7a, we notice that the advantage of the constrained positive $Var(s)$ designs are most pronounced for $p = 0$, still evident for $p = 0.25$, but mostly disappear when $p = 0.5$. In other words, as long as most of the effects are specified correctly, the difference between the design criteria appears to exist, but even if the experimenter simply guesses, the constrained positive $Var(s)$ designs will still perform as well as the standard supersaturated designs.

A two-factor interaction model was fit to the power simulation data depicted in Figure 7a, where again we take as the response the square root of the counts of correctly identified effects ($R^2 \approx 94\%$). We use as factors: p, Criteria, Scenario, Method, n, and k. Figure 8 shows the main effect plot of Criteria (p-value < 0.0001) as well as the interaction plot between Criteria and p (p-value $= 0.0032$). These results confirm what Figure 7a suggested:



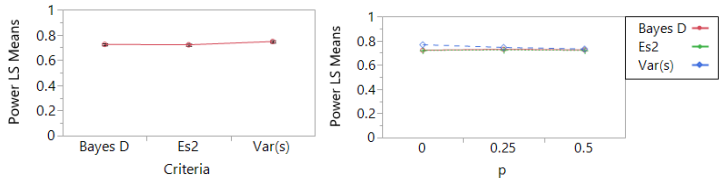
(a) Power



(b) Type I Error

Figure 7: Average power and Type I error by design criteria over all twenty designs sizes, shown for each of the three analysis methods, every value of p, and each of the four scenarios.

when the effect signs are correctly or mostly correctly specified the constrained positive $Var(s)$ designs are more powerful, but even when signs are misspecified the constrained positive $Var(s)$ designs are no worse than standard designs (Figure 8b).



(a) Criteria Main Effect Plot (b) Criteria*p Interaction

Figure 8: Criteria main effect and interaction plot of power over all values of p. Note the y-axis is adjusted to show proportions.

5 Conclusions

Supersaturated designs are typically employed to screen a few factors from many candidates. They have great potential to aid in discoveries in a resource-efficient manner, but this potential is limited by their power to detect the effects that are active. In this article, we have presented a criterion that produces designs with markedly higher power than traditional supersaturated designs, when effect directions can be specified in advance. Even when effect directions are misspecified, however, the proposed designs are not inferior. Further, we have presented an algorithm that takes advantage of the structure of these SSDs for efficient design construction.

In a screening experiment, it is more important to identify the active effects than to ensure that no spurious factors are included. Thus, the combination of a design and an analysis method that increases power is of interest in this setting, even at the expense of somewhat elevated Type I error rates. However, for the range of experimental scenarios explored in this article, the constrained positive $Var(s)$-optimal designs have produced increased power without any discernible rise in the rate of false positives.

As noted by a referee, an alternative to our criterion as presented in Section 2.2 (equations (3) and (4)) could be to minimize $Var(s)$ subject to a constraint on the maximum absolute

correlation between design columns. In other words, select a design D which minimizes

$$\begin{aligned} Var(s) &= E(s^2) - E(s)^2 \\ &= \frac{2}{k(k+1)} \sum_{1 \le i < j} s_{ij}^2 - \left(\frac{2}{k(k+1)} \sum_{1 \le i < j} s_{ij} \right)^2 \end{aligned} \tag{10}$$

subject to

$$max\,|r_{ij}| < c, \tag{11}$$

where r_{ij} is the correlation between columns i and j. An initial reaction is that this suggestion would produce similar designs to our proposed criterion. The $max\,|r_{ij}| < c$ constraint would seemingly accomplish something similar to the $E(s^2)$-efficiency constraint, though the fact that the maximum correlation is being limited might produce interesting differences. To ensure that the average correlation is positive (equation (10)), it is likely that an additional constraint similar to $E(r_{ij}) > 0$ would need to be imposed. Further research is recommended.

We do note the limitations of a simulation study; it is possible that scenarios could be devised for which the Type I error rates of the proposed designs are problematically high. Further research, using additional scenarios informed by real supersaturated experiments, might shed additional light on the effectiveness of these designs. Interesting future work would also include a focus on increasing the ability to detect interaction effects, as well as a comparison with other analysis strategies which include the known effect direction assumption, such as group screening and sequential bifurcation.

Supplementary Materials

Supplementary material for "Powerful Supersaturated Designs when Effect Directions are Known" is included online and consists of the following:

A_Designs A folder which includes all constrained positive $Var(s)$-optimal designs, along

with the balanced $E(s^2)$-optimal designs

B_Code A folder containing Matlab code used to generate the constrained positive $Var(s)$-optimal designs.

C_Additional Simulations A document providing simulation results when only active factor signs are misspecified.

Acknowledgments

The authors thank Phil Scinto for his suggestion to begin a study of supersaturated designs, as well as anonymous referees who provided excellent feedback to earlier versions of this article. This work was partially supported by a grant from the Simons Foundation (award no. 244759 to David Edwards).

References

Abraham, B.; Chipman, H.; and Vijayan, K. (1999). "Some risks in the construction and analysis of supersaturated designs". *Technometrics* 41, 2, pp. 135–141.

Bettonvil, B. and Kleijnen, J. P. C. (1996). "Searching for important factors in simulation models with many factors: Sequential bifurcation". *European Journal of Operational Research* 96, 1, pp. 180–194.

Booth, K. H. V. and Cox, D. R. (1962). "Some systematic supersaturated designs". *Technometrics* 4, 4, pp. 489–495.

Brenneman, W. A. (2014). "Comment: Simulation Used to Solve Tough Practical Problems". *Technometrics* 56, 1, pp. 19–20.

Candes, E. and Tao, T. (2007). "The Dantzig Selector: Statistical Estimations when p is Much Larger than n". *The Annals of Statistics* 35, 6, pp. 2313–2351.

Deng, L.-Y.; Lin, D. K. J.; and Wang, J. (1996). "A measurement of multi-factor orthogonality". *Statistics & Probability Letters* 28, 3, pp. 203–209.

Deng, L.-Y.; Lin, D. K. J.; and Wang, J. (1999). "A resolution rank criterion for supersaturated designs". *Statistica Sinica* 9, pp. 605–610.

Draguljić, D.; Woods, D. C.; Dean, A. M.; Lewis, S. M.; and Vine, A.-J. E. (2014). "Screening Strategies in the Presence of Interactions". *Technometrics* 56, 1, pp. 1–16.

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

Jones, B.; Li, W.; Nachtsheim, C. J.; and Ye, K. Q. (2009). "Model robust supersaturated and partially supersaturated designs". *Journal of Statistical Planning and Inference* 139, 1, pp. 45–53.

Jones, B.; Lin, D. K. J.; and Nachtsheim, C. J. (2008). "Bayesian D-optimal supersaturated designs". *Journal of Statistical Planning and Inference* 138, 1, pp. 86–92.

Jones, B. and Majumdar, D. (2014). "Optimal supersaturated designs". *Journal of the American Statistical Association* 109, 508, pp. 1592–1600.

Lin, D. K. J. (1993). "A new class of supersaturated designs". *Technometrics* 35, pp. 28–31.

Lin, D. K. J. (1995). "Generating systematic supersaturated designs". *Technometrics* 37, pp. 213–225.

Marley, C. J. and Woods, D. C. (2010). "A Comparison of design and model selection methods for supersaturated experiments". *Computational Statistics and Data Analysis* 54, pp. 3158–3167.

Meyer, R. K. and Nachtsheim, C. J. (1995). "The Coordinate-Exchange Algorithm for Constructing Exact Optimal Experimental Designs". *Technometrics* 37, 1, pp. 60–69.

Nguyen, N. (1996). "An alogrithmic approach to constructing supersaturated designs". *Technometrics* 38, pp. 69–73.

Phoa, F.; Pan, Y.-H.; and Xu, H. (2009). "Analysis of supersaturated designs via the Dantzig Selector". *Journal of Statistical Planning and Inference* 139, pp. 2362–22372.

Scinto, P. R.; Wilkinson, R. G.; Wang, Z.; and Rose, A. (2014). "Comment: Need for Guidelines on Appropriate Screening Designs for Practitioners". *Technometrics* 56, 1, pp. 23–24.

Watson, G. S. (1961). "A Study of the Group Screening Method". *Technometrics* 3, 3, pp. 371–388.

Weese, M. L.; Smucker, B. J.; and Edwards, D. J. (2015). "Searching for Powerful Supersaturated Designs". *Journal of Quality Technology* 47, 1, pp. 66–84.

Westfall, P. H.; Young, S. S.; and Lin, D. K. J. (1997). "Forward selection error control in the analysis of supersaturated designs". *Statistica Sinica* 8, pp. 101–117.

Wu, C. F. J. (1993). "Construction of supersaturated designs through partially aliased intercations". *Biometrika* 80, pp. 661–669.