# Searching for Powerful Supersaturated Designs

MARIA L. WEESE

*Miami University, Oxford, OH 45056*

BYRAN J. SMUCKER

*Miami University, Oxford, OH 45056*

DAVID J. EDWARDS

*Virginia Commonwealth University, Richmond, VA 23284*

An important property of any experimental design is its power, defined roughly as the ability to detect active factors. For supersaturated designs power is even more critical. We consider several popular supersaturated design construction criteria in the literature, propose several of our own, and perform a simulation study to evaluate them in terms of power. We use two analysis methods—forward selection and the Dantzig selector—and find that although the Dantzig selector clearly outperforms forward Selection, there is no clear winner among the design construction criteria.

Keywords: Bayesian D-optimal; Dantzig selector; $E(s^2)$; Forward selection; Model Robust

Dr. Weese is an Assistant Professor in the Department of Information Systems & Analytics. She is a senior member of ASQ. Her email address is weeseml@miamioh.edu.

Dr. Smucker is an Assistant Professor in the Department of Statistics. His email address is smuckerb@miamioh.edu

Dr. Edwards is an Associate Professor in the Department of Statistical Sciences & Operations Research. He is a member of ASQ. His email address is dedwards7@vcu.edu.

# 1 Introduction and Motivation

Screening experiments, using plans such as resolution III fractional factorials or Plackett-Burman designs, are frequently utilized in the early stages of process experimentation. The goal of screening is to distinguish the few important factors from the many unimportant ones. However, when there are severe temporal or economic constraints a typical screening design may require too many resources. Supersaturated designs (SSDs), which use $n < k+1$ runs to examine $k$ factors, are attractive alternatives. Clearly, SSDs have too few runs to allow for the estimation of all main effects and thus require the experimenter to rely heavily on the assumption of effect sparsity.

The origin of SSDs is attributed to Satterthwaite (1959) who proposed random balance experiments to identify a subset of important factors. Research in this area languished until the early 1990s when Lin (1993) and Wu (1993) provided two of the early methods for constructing $E(s^2)$-optimal SSDs. In particular, Lin (1993) constructed SSDs as half-fractions of Plackett-Burman (PB) designs and Wu (1993) proposed constructing $E(s^2)$-optimal SSDs by augmenting Hadamard matrices with two-factor interaction columns. See also Li and Wu (1997) who built $E(s^2)$-optimal SSDs based on a D-optimal design search by applying columnwise-pairwise algorithms. This method contrasted with Wu (1993) in that nonorthogonality does not mainly accumulate in the last factors. To date, $E(s^2)$ (to be discussed in section 2) has become the most commonly used criterion for constructing SSDs.

There are relatively few published case studies of the use of SSDs in practice. For a recent example (though in a context in which the number of model terms—not the number of factors—renders the experiment supersaturated), see Scinto et al. (2011). This experiment studied the effect of more than 70 model terms on the coefficient of friction of engine motor oil using just 28 runs. Another example, from Holcomb et al. (2007), concerns experiments to aid in the evaluation of various turbine engine designs. There are 27 factors of interest, each combination of which specifies an engine design, and the authors compare

the performance of several designs, including supersaturated designs of 12, 16, and 20 runs. They conclude that supersaturated designs are unlikely to produce definitive results.

Though an abundance of criteria have been proposed to evaluate SSDs, the power—defined as the average probability of detecting active effects for a specified set of effect sizes—when balanced by a controlled Type I error rate is the ultimate measure of a design's effectiveness. It also has a straightforward interpretation for experimenters and their sponsoring organizations. Gilmour (2006) notes that supersaturated "[d]esigns are usually built to optimize the $E(s^2)$ criterion, but this appears to be unrelated to the way in which the data are analyzed." Using power to evaluate and recommend a variety of supersaturated designs is the goal of the present study, and will help relate the design to its method of analysis and provide guidance to experimenters regarding which supersaturated designs to use. We calculate power by simulating data from supersaturated experiments and measuring the proportion of correctly identified active effects.

Marley and Woods (2010) measured power in this way and performed a simulation study that compared the power and type I error (defined as the proportion of incorrectly identified active effects) of two types of designs (Bayesian D-optimal and $E(s^2)$-optimal) and three analysis methods (forward selection, model averaging, and the Dantzig selector) over three supersaturated experiments of differing sizes and several experimental scenarios. They concluded that the analysis method had a large impact on power (the Dantzig selector was best; forward selection using $\alpha_{enter} = 0.05$ was worst) while the design construction criteria had no discernible effect. In this article, we focus and enlarge the study of Marley and Woods along the design selection axis, comparing six different SSD construction criteria over 12 different experiments. We use similar experimental scenarios as Marley and Woods, and utilize two versions of forward selection ($\alpha_{enter} = 0.05$ and AICc as two different stopping criteria) and the Dantzig selector as analysis methods. Our simulations, based on design construction criteria that are a mix of established and new approaches, confirm and deepen the conclusions of Marley and Woods. Even over this larger range of designs and experiments, no single design construction criteria distinguishes itself as superior to all

others in regards to power.

The existing SSD construction criteria that we consider in this article are Bayesian D-optimality (Jones et al., 2008), $E(s^2)$-optimality (Booth and Cox, 1962; Lin, 1993; Wu, 1993), and model-robust (Jones et al., 2009; **?**). We also motivate the consideration of several new criteria, including unbalanced $E(s^2)$-optimality, constrained $Var(s)$-optimality, and a criterion based upon an approximation of effect power. We test these design construction criteria over a suite of twelve SSDs, ranging from small and slightly supersaturated (e.g., 11 factors in 10 runs) to medium-sized and more severely supersaturated (e.g., 30 factors in 16 runs) to larger designs (e.g., 31 factors in 24 runs).

The statistical model we assume is standard. Let $\mathbf{D}$ represent the $n \times k$ supersaturated design matrix with $k$ factors with possible levels $\pm 1$, and $\mathbf{X} = [\mathbf{1}, \mathbf{D}]$ is the main effects model matrix where $\mathbf{1}$ is an $n \times 1$ vector of ones. Throughout the paper we consider a linear main effects model of the form

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \epsilon, \tag{1}$$

where $\mathbf{y}$ is an $n \times 1$ response vector, $\boldsymbol{\beta}$ is a vector of unknown parameters, and $\epsilon$ is the error vector with each element independent and $E(\epsilon_i) = 0$ and $Var(\epsilon_i) = \sigma^2$. For SSDs, $rank(\mathbf{X}) < k + 1$. Thus, $\mathbf{X}'\mathbf{X}$ is singular and no unique least squares estimate for $\beta$ can be obtained. This necessitates analysis methods that can exploit the assumed sparsity of effects (see section 3.1).

The rest of the paper is formatted as follows. Section 2 provides a description of the various design construction criteria (established and new) that we utilize. Section 3 includes a simulation study, using both forward selection and the Dantzig selector as SSD analysis methods, as well as a statistical analysis of the simulation results. We conclude in Section 4 with a discussion and conclusions.

# 2 Supersaturated design construction criteria

In this section we review several supersaturated design criteria from the literature. We also present several new criteria.

## 2.1 Established criteria

In the introduction to this article, we gave a brief review of some of the early revival in supersaturated designs and mentioned design construction methods. Here we give several leading supersaturated design criteria, and also review the methods by which they can be constructed. The criteria we consider are not exhaustive in the literature (see, for instance, Allen and Bernshteyn, 2003; Holcomb et al., 2003), but are popular and/or relatively easily constructed or obtained.

### 2.1.1 $E(s^2)$-optimality

Booth and Cox (1962) provided the first construction method utilizing the $E(s^2)$ criterion, which minimizes the average of all squared pairwise inner products producing near orthogonal designs. Let $s_{ij}$ be the $(i,j)th$ element of $\mathbf{X}'\mathbf{X}$. Then, the $E(s^2)$ criterion selects a design that minimizes

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

while keeping the design balanced (i.e. each column has the same number of $-1$ as $+1$). Marley and Woods (2010) extended the definition in equation (2) to include the intercept column of $\mathbf{X}$:

$$E(s^2) = \frac{2}{k(k+1)} \sum_{1 \le i < j} s_{ij}^2. \tag{3}$$

When the design is balanced the two criteria are essentially equivalent because balance

ensures that the intercept is uncorrelated with all main effects. The balanced $E(s^2)$ designs used in this paper were generated either by the algorithm of Ryan and Bulutoglu (2007) or by the nonorthogonal array (NOA) algorithm of Nguyen (1996). Both construction methods employ a secondary criterion of minimizing the $max_{i<j}s_{ij}^2$ when there are several designs that achieve the optimal lower bound on $E(s^2)$. Again, the $E(s^2)$ criterion is the most commonly used criterion in the literature on SSDs. See, for example, Cheng (1997), Liu et al. (2007), and Nguyen and Cheng (2008) for several other methods of constructing $E(s^2)$-optimal supersaturated designs.

### 2.1.2 Bayesian D-optimality

DuMouchel and Jones (1994) used a Bayesian approach to construct D-optimal designs with a reduced dependence on a user-specified model. Model terms are categorized as either primary or potential. Primary terms are assumed to be active, while potential terms may or may not appear in the true model.

Subsequently, Jones et al. (2008) applied this idea to SSDs, with the assumption that the intercept is primary and all main effects are potential terms. Under the assumption that primary terms have a diffuse prior and that potential terms have a prior mean of 0 and variance of $\tau^2$, the posterior variance-covariance matrix of $\beta$ is proportional to $(\mathbf{X'X} + \mathbf{K}/\tau^2)^{-1}$, where $\mathbf{K}/\tau^2$ is proportional to the prior variance-covariance matrix for $\beta$, and

$$\mathbf{K} = \begin{pmatrix} 0 & \mathbf{0_{1\times k}} \\ \mathbf{0_{k\times 1}} & \mathbf{I_{k\times k}} \end{pmatrix}. \tag{4}$$

Jones et al. (2008) suggested finding a supersaturated design that maximizes

$$\phi_D = |\mathbf{X'X} + \mathbf{K}/\tau^2|^{1/(1+k)}. \tag{5}$$

The Bayesian D-optimal designs used in this work were created using JMP software with

$\tau^2 = 1$. Jones et al. (2008) notes that these designs are relatively insensitive to the choice of $\tau^2$ and use $\tau^2 = 5$. This criterion can be used to produce designs with any number of factors at any run size and does not require the SSD to be balanced.

### 2.1.3 Model Robust Supersaturated Designs

Jones et al. (2009) developed a method of SSD construction based on the model-robust approach of Li and Nachtsheim (2000). Jones et al. (2009) begin by specifying a set of models, $\mathcal{F}_g$, which is composed of all models that include $g$ of the $k$ factors. They then seek model-robust supersaturated (MRSS) designs which maximize the proportion of estimable models in $\mathcal{F}_g$. This proportion, called the estimation capacity ($EC_g$), is the primary design criterion; a secondary criterion (the average D-efficiency across all models in $\mathcal{F}_g$) is also maximized, subject to $EC_g$ being maintained. The authors explore various combinations of $g$, $n$, and $k$ that result in MRSS designs that have 100% estimation capacity. The columnwise-pairwise exchange algorithm employed in Jones et al. (2009) requires column balance, unless the number of runs is odd in which case the design is made as balanced as possible.

For a given $k$ and $g$, there are $r = \binom{k}{g}$ models in $\mathcal{F}_g$. This quantity can grow large and becomes a computational bottleneck for Jones et al. (2009). Smucker and Drew (2014) consider the same supersaturated model space but overcome the computational challenges by choosing a subset of models from $\mathcal{F}_g$ and finding a design robust for the subset. They show that such designs give up little in terms of robustness with respect to the full model space, and can be constructed in a small fraction of the time it takes when the full model space is used.

Since in this article we consider experiments larger than those in Jones et al. (2009), we utilize an algorithm based on Smucker and Drew (2014) to construct these designs. We test designs using $g = \lceil n/3 \rceil$ and $g = \lceil 2n/3 \rceil$, to model varying degrees of effect sparsity. The algorithm is based upon coordinate exchange (Meyer and Nachtsheim, 1995) and thus does

not enforce balance.

## 2.2 New design construction criteria

We now present several original supersaturated design criteria.

### 2.2.1 Approximate Power

Under standard regression assumptions, power for each effect may be approximated using the noncentral F-distribution (Mee, 2009) as

$$\pi = 1 - F_{Dist}(F_{crit}, \nu_1, \nu_2, \lambda)$$

where $F_{crit} = F_{quantile}(1-\alpha, \nu_1, \nu_2)$, $\nu_1 = 1$ is the numerator degrees of freedom, $\nu_2$ is the denominator degrees of freedom and $\boldsymbol{\lambda} = (\lambda_0, \lambda_1, \ldots, \lambda_k)'$ is the $(k+1)$-vector of noncentrality parameters. Note that $\pi$ is also a $(k+1)$-vector and corresponds to the approximate power for each of the model parameters. The $i^{th}$ noncentrality parameter is calculated as

$$\lambda_i = \frac{\beta_i/\sigma^2}{c_{ii}}$$

where $c_{ii}$ is the $i^{th}$ diagonal element of $(\mathbf{X}'\mathbf{X}+\mathbf{K}/\tau^2)^{-1}$. We choose the signal to noise ratio, $\beta_i/\sigma^2$, to be 1 and set $\tau^2 = 5$. $K$ is as defined in equation (4). Although other possibilities exist, we construct approximate power optimal designs by maximizing the minimum power, $\pi_{maximin}$, where the minimum is taken over all non-intercept parameters.

Older versions of JMP software (e.g. version 8) provide this approximate power calculation in the custom design platform with $\nu_2 = 1$ for SSDs (in JMP 10 the experimenter is given a choice regarding $\nu_2$). Since there are not enough degrees of freedom to estimate all main effects—let alone to estimate the error term—one approach to a power approximation is JMP's old default: Conjure a single degree of freedom for error (i.e. $\nu_2 = 1$), with the implicit reliance on effect sparsity to provide it. However, since effect sparsity is a necessary

assumption when using a SSD, and Marley and Woods (2010) suggest "[t]he number of runs should be at least three times the number of active factors," we consider $\nu_2 = \lfloor n - n/3 \rfloor$, which better reflects the assumption of effect sparsity by supposing that the number of active effects will be at most $\lceil n/3 \rceil$.

### 2.2.2 Unbalanced $E(s^2)$-optimal

It has long been assumed in the literature, including by Booth and Cox (1962), that supersaturated designs should be balanced. However, more recent authors have realized possible advantages to relaxing this requirement. For instance, Jones et al. (2008) has demonstrated that relaxing the balance constraint in constructing Bayesian D-optimal designs produces designs with a lower value of $E(s^2)$ than the $E(s^2)$-optimal design. Indeed, Marley and Woods (2010) extend the definition of $E(s^2)$ as in (3) to include the intercept term, and this is only different from the standard definition when the design is unbalanced.

We use the definition in (3) and find designs that minimize this quantity, without the effect balance requirement. We construct these SSDs with an algorithm based upon coordinate exchange (see section 2.3). Table 1 compares several balanced $E(s^2)$-optimal designs with their unbalanced counterparts, in terms of (3), for three supersaturated experiments. It is clear that lower $E(s^2)$ values can be achieved if the balance requirement is removed.

Table 1: Comparison of balanced $E(s^2)$-optimal design values with the unbalanced $E(s^2)$-optimal for three design sizes, in terms of the criterion defined in (3).

| Design Size | Type | Criterion Value |
|---|---|---|
| $n = 12, k = 26$ | Balanced $E(s^2)$-optimal | 7.52 |
| | Unbalanced $E(s^2)$-optimal | 7.18 |
| $n = 14, k = 24$ | Balanced $E(s^2)$-optimal | 7.31 |
| | Unbalanced $E(s^2)$-optimal | 6.88 |
| $n = 18, k = 22$ | Balanced $E(s^2)$-optimal | 5.80 |
| | Unbalanced $E(s^2)$-optimal | 5.52 |

### 2.2.3 Constrained $Var(s)$-optimal

The traditional balanced $E(s^2)$ criterion was extracted from Booth and Cox (1962). However, Booth and Cox (1962) did not actually propose $E(s^2)$. Instead, they proposed $Var(s)$ as a criterion, though they apparently assumed that $E(s) = 0$, which would imply that $Var(s) = E(s^2)$. This assumption turns out not to be strictly true, whether in the design considered by Booth and Cox (1962) or in SSDs more generally.

Because of this, we examine $Var(s)$ as a criterion in this article. We calculate the variance of $s$ using

$$Var(s) = E(s^2) - E(s)^2 = \frac{2}{k(k+1)} \sum_{1 \le i < j} s_{ij}^2 - \left( \frac{2}{k(k+1)} \sum_{1 \le i < j} s_{ij} \right)^2 . \qquad (6)$$

A design chosen to minimize $Var(s)$ alone would allow very high $s$ values with little or no variation among them. To prevent this, we propose constrained $Var(s)$ designs, in which we minimize $Var(s)$ subject to a specified $E(s^2)$ efficiency. We define $E(s^2)$ efficiency for design $\mathbf{D}$ as

$$E(\mathbf{D}) = \frac{E(s^2)(\mathbf{D}^*)}{E(s^2)(\mathbf{D})}$$

where $\mathbf{D}^*$ is the $E(s^2)$-optimal design (balanced or unbalanced). After evaluating several choices, we have specified a lower bound on $E(s^2)$-efficiency of 80%.

## 2.3 Algorithms

For the designs constructed using the criteria described in Section 2.2, we utilized algorithms based upon coordinate exchange (Meyer and Nachtsheim, 1995). A sketch of our implementation is as follows, where $\phi$ represents one of the supersaturated criteria described above:

1. Randomly construct an $n \times k$ initial supersaturated design.

2. Iterate through the design coordinate by coordinate. At each coordinate, consider the effect on $\phi$ of multiplying the current value by $-1$.

   (a) If exchanging the current coordinate improves $\phi$, make the exchange.

   (b) Otherwise, continue to the next coordinate.

3. Continue iterating through the design until convergence.

In step 2, we update the $\mathbf{X}'\mathbf{X}$ matrix via the standard rank-1 update formula (see Meyer and Nachtsheim, 1995). This is a heuristic optimization algorithm which does not guarantee a globally optimal solution. Thus, each of multiple algorithm tries should begin from a different initial design and the best chosen. For the designs generated by these algorithms, we use 100 random starts.

# 3 Analysis and Simulations

## 3.1 Analysis Methods

While much has been published on the design of supersaturated experiments, much less has been written about their analysis (Dejaegher and Vander Heyden, 2008). Two methods of interest are forward selection and the Dantzig selector.

The Dantzig selector (Candes and Tao, 2007) chooses active factors by solving a very simple convex program that can be transformed into a linear program. It locates a vector of estimates, consistent with the data, that minimizes its $\ell_1$ norm. Phoa et al. (2009) utilize this procedure for SSDs, and Marley and Woods (2010) find that it is more effective in terms of power than forward selection or Bayesian variable selection.

When forward selection is used, Type I error rates can be quite high (Westfall et al., 1997). Abraham et al. (1999) indicates that forward selection can be negatively influenced by the way in which factors are assigned to columns, and suggest that all-subsets variable selection should be used instead, a recommendation with which Kelly and Voelkel (2000)

agree. Marley and Woods (2010) utilize forward selection with $\alpha_{enter} = 0.05$ in their simulation, and find it to be relatively ineffective. Another common version of forward selection, using $AICc = AIC + \frac{2k(k+1)}{n-k-1}$ as a stopping criterion, is available in popular software packages, like SAS and JMP.

In our simulation study we consider the Dantzig selector, given its effectiveness as demonstrated by Marley and Woods (2010), as well as both versions of forward selection. Despite its deficiencies, forward selection is simple, fast, and continues to be commonly used in practice, and if nothing else serves as a foil for the more effective Dantzig selector procedure.

Note that there are many other analysis techniques that have been proposed in the literature, including Bayesian approaches (Chipman et al., 1997; Beattie et al., 2002; Marley and Woods, 2010), genetic algorithms (Cela et al., 2001), approaches that use variants of standard least squares (Li and Lin, 2002; Zhang et al., 2006) and approaches that are elaborations of classical stepwise regression (Lu and Wu, 2004).

### 3.2 Simulation Procedure

The simulation procedure largely follows that of Marley and Woods (2010). In particular, we vary the following in our simulation:

1. Supersaturated design size. Twelve choices were used, some of them based on SSDs found in the literature. Specifically, we construct $n \times k$ SSDs of the following sizes: $8 \times 12$; $10 \times 11$; $10 \times 15$; $12 \times 22$; $12 \times 26$; $14 \times 23$; $14 \times 24$; $16 \times 30$; $18 \times 22$; $20 \times 24$; $24 \times 34$; $26 \times 31$;.

2. Design construction criteria. We investigate balanced $E(s^2)$-optimal, Bayesian D-optimal, model robust ($g = \lceil n/3 \rceil$), model robust ($g = \lceil 2n/3 \rceil$), approximate power-optimal (maximin), unbalanced $E(s^2)$-optimal, and constrained $Var(s)$-optimal supersaturated designs.

3. Simulation scenario. The number and magnitude of the active factors is varied across

four different scenarios. In particular, the coefficient for each of the $a$ randomly chosen active factors is drawn at random 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 each of four factors and $\mu = 2$ for five factors.

4. Analysis method. We apply both the Dantzig selector, forward selection using $\alpha_{enter} = 0.05$ and forward selection using AICc for each supersaturated design.

The simulation is carried out as follows. Following Marley and Woods (2010), in each of 10,000 iterations

1. From the columns $2, \ldots, k+1$ of $\mathbf{X}$, $a$ columns are randomly assigned as the active factors. The coefficients for the active factors are obtained by sampling from $N(\mu, 0.2)$ and a sign ($+$ or $-$) randomly applied.
2. The remaining effects (inactive effects) are randomly assigned a coefficient from $N(0, 0.2)$.
3. The response vector is generated from the model in equation (1) with errors ($\epsilon_i$) generated from $N(0, 1)$.
4. The active factors are determined by one of the model selection methods.

At the end of 10,000 iterations the average proportion of correctly identified active effects (i.e. power) and the average proportion of inactive effects identified as active (i.e., Type I error) was reported. The proportion of times an individual column was correctly identified (column power) was also calculated.

We note that the Dantzig selector requires specification of two parameters. The first, $\delta$, is a tuning parameter and is selected as in Marley and Woods (2010) via the usual BIC

statistic given by

$$BIC = n \log\left(\frac{(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})}{n}\right) + p \log(n), \tag{7}$$

where $p$ is the number of model terms and $\hat{\beta}$ is the least squares estimate found by regressing the response on the set of factors deemed active by the Dantzig selector. The second parameter, $\gamma$, is a threshold for coefficient estimates. That is, those factors chosen by the Dantzig selector whose absolute coefficients exceed $\gamma$ are then selected as active. We, again, follow Marley and Woods (2010) and choose $\gamma = 1.5$ to help control Type I error. Forward selection was performed using a significance level of $\alpha_{enter} = 0.05$ for entry into the model and forward selection using AICc as a stopping rule.

### 3.3 Simulation Results

Based on the analysis methods and simulation procedure described in the previous two sections, we display the results of our simulations in Figures 1, 2 and 3. These figures show, for both the Dantzig selector and forward selection, respectively, the power as a function of design criterion, design size, and simulation scenario. In Figure 1 (Dantzig selector), although we see some separation of the design construction criteria for the smaller SSDs (e.g., $8 \times 12$, $10 \times 11$), it appears evident that no design construction criterion dominates all others. We note though that the approximate power-optimal designs appear to distinguish themselves negatively overall. In Figures 2 and 3 (forward selection), there is even less notable differences between the construction criteria. See Tables A.1-A.6 in the appendix for complete tabular results, including Type I error rates.

In order to help formalize the conclusions from the simulation, we develop a statistical model for power by treating each of the simulation variables (i.e., number of experimental runs ($n$), number of factors ($k$), design construction criterion (Criteria), simulation scenario (Scenario), and analysis method (Method)) as experimental factors and proceed to fit a model that includes all main effects and two-factor interactions involving these five factors.

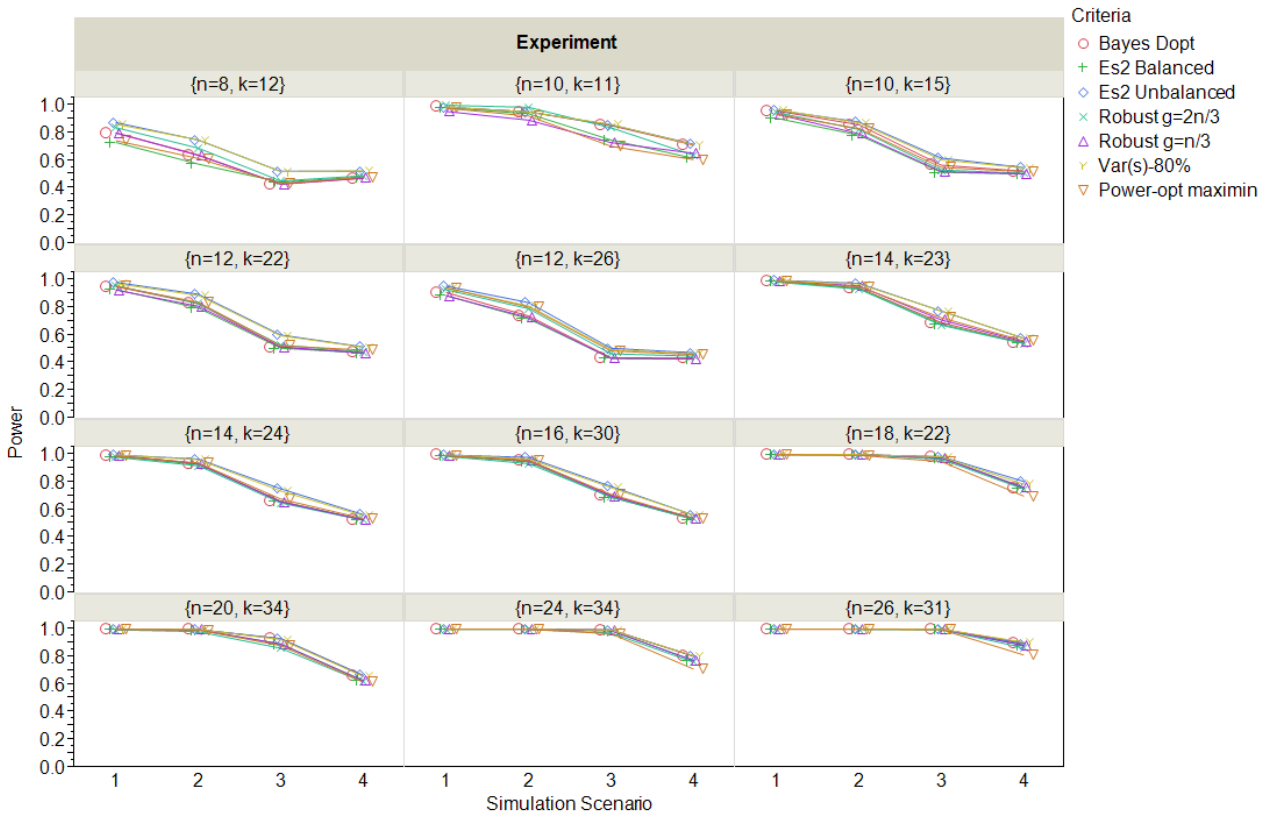



Figure 1: Power based on simulations with the Dantzig selector for each of seven design types, twelve experiments and four simulation scenarios. Simulation scenario 1 corresponds to $\{a = 3, \mu = 5\}$; 2 corresponds to $\{a = 4, \mu = 4\}$; 3 corresponds to $\{a = 6, \mu = 3\}$; 4 corresponds to $\{a = 9, \mu = \text{mixed}\}$.

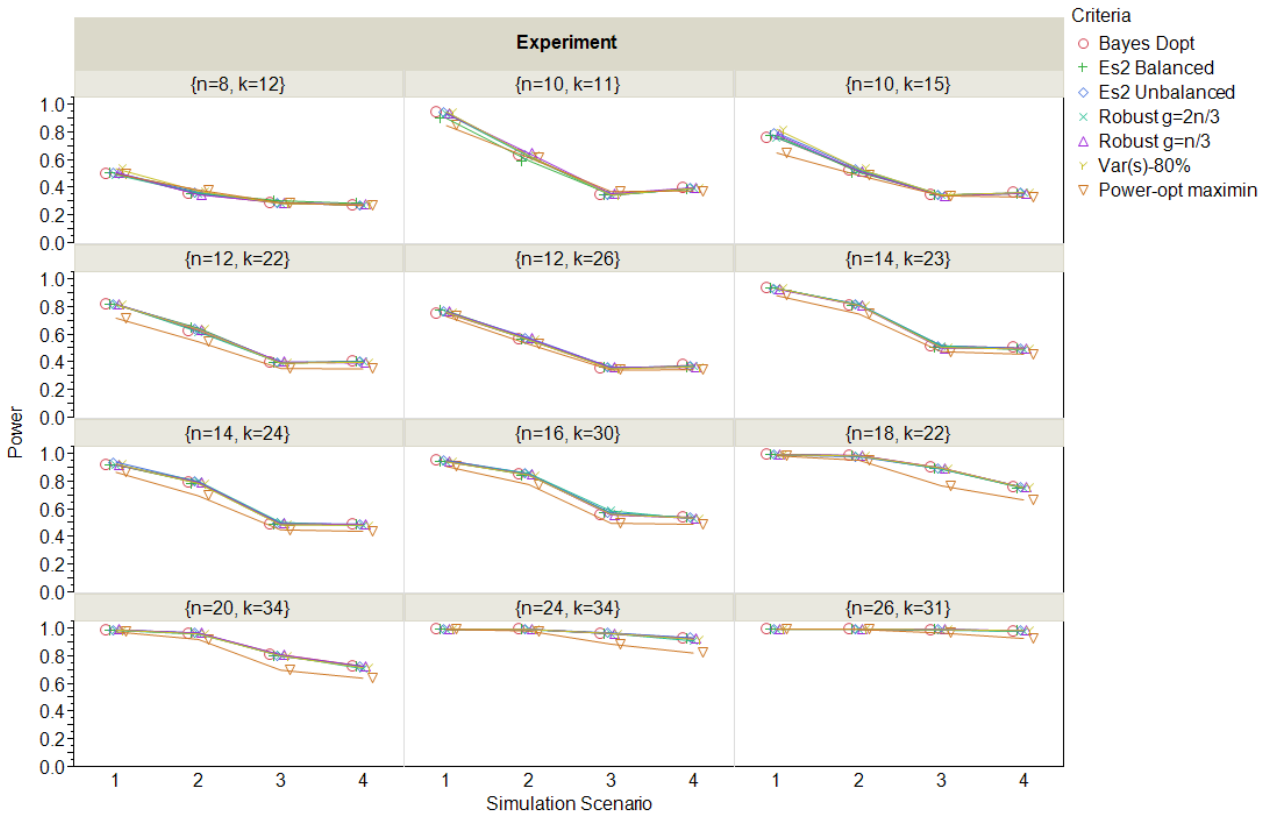



Figure 2: Power based on simulations with forward selection using AICc for each of seven design types, twelve experiments and four simulation scenarios. Simulation scenario 1 corresponds to $\{a = 3, \mu = 5\}$; 2 corresponds to $\{a = 4, \mu = 4\}$; 3 corresponds to $\{a = 6, \mu = 3\}$; 4 corresponds to $\{a = 9, \mu = \text{mixed}\}$.

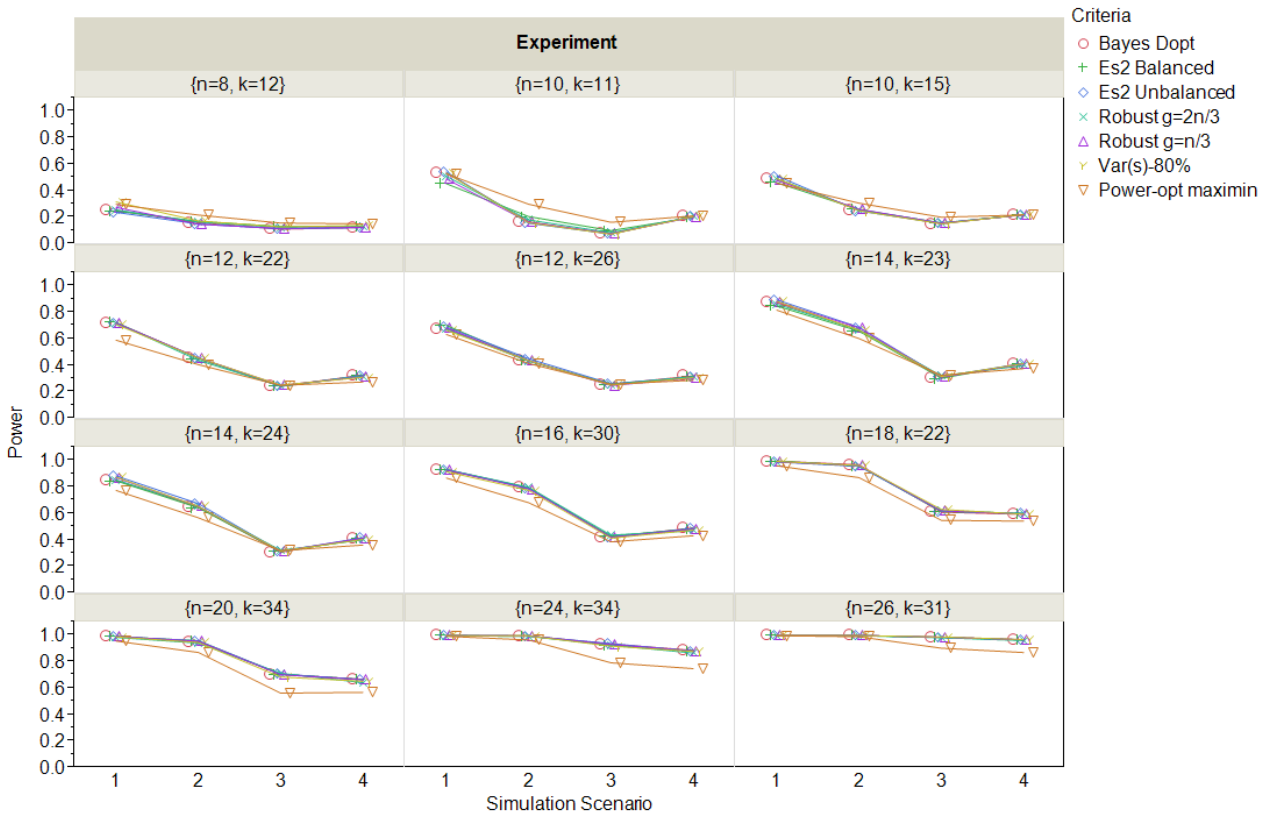



Figure 3: Power based on simulations with forward selection using $\alpha_{enter} = 0.05$ for each of seven design types, twelve experiments and four simulation scenarios. Simulation scenario 1 corresponds to $\{a = 3, \mu = 5\}$; 2 corresponds to $\{a = 4, \mu = 4\}$; 3 corresponds to $\{a = 6, \mu = 3\}$; 4 corresponds to $\{a = 9, \mu = \text{mixed}\}$.

Our goal in doing so is to more clearly emphasize the lack of differences among the various SSD design construction criteria as seen in Figures 1-3.

First, we note that using the raw power values from the simulation as the response variable resulted in an ill-fitting model as determined via usual residual diagnostics. Furthermore, typical variance stabilizing transformations (e.g., arcsine square root) on this response were unhelpful. As a consequence, we elected to model the number of truly active factors declared active. Diagnostics obtained after modeling this count variable indicated a good fit. Predicted values for power were then simply obtained by dividing the predicted count of active factors declared active by the number of active factors in a given scenario. The ANOVA table for this model is shown in Table 2 ($R^2 \approx 96\%$). Immediately, we take note that none of the interaction effects involving SSD construction criteria are significant. See Figure 4 for a graphical depiction. Although the main effect for construction criterion is statistically significant ($p \leq 0.0001$), a Tukey multiple comparison procedure indicates that only the approximate power-optimal SSDs (with lowest overall mean power) are significantly worse than some of the others. Thus, in general, the analysis suggests that no single design construction criteria distinguishes itself as superior to all others.

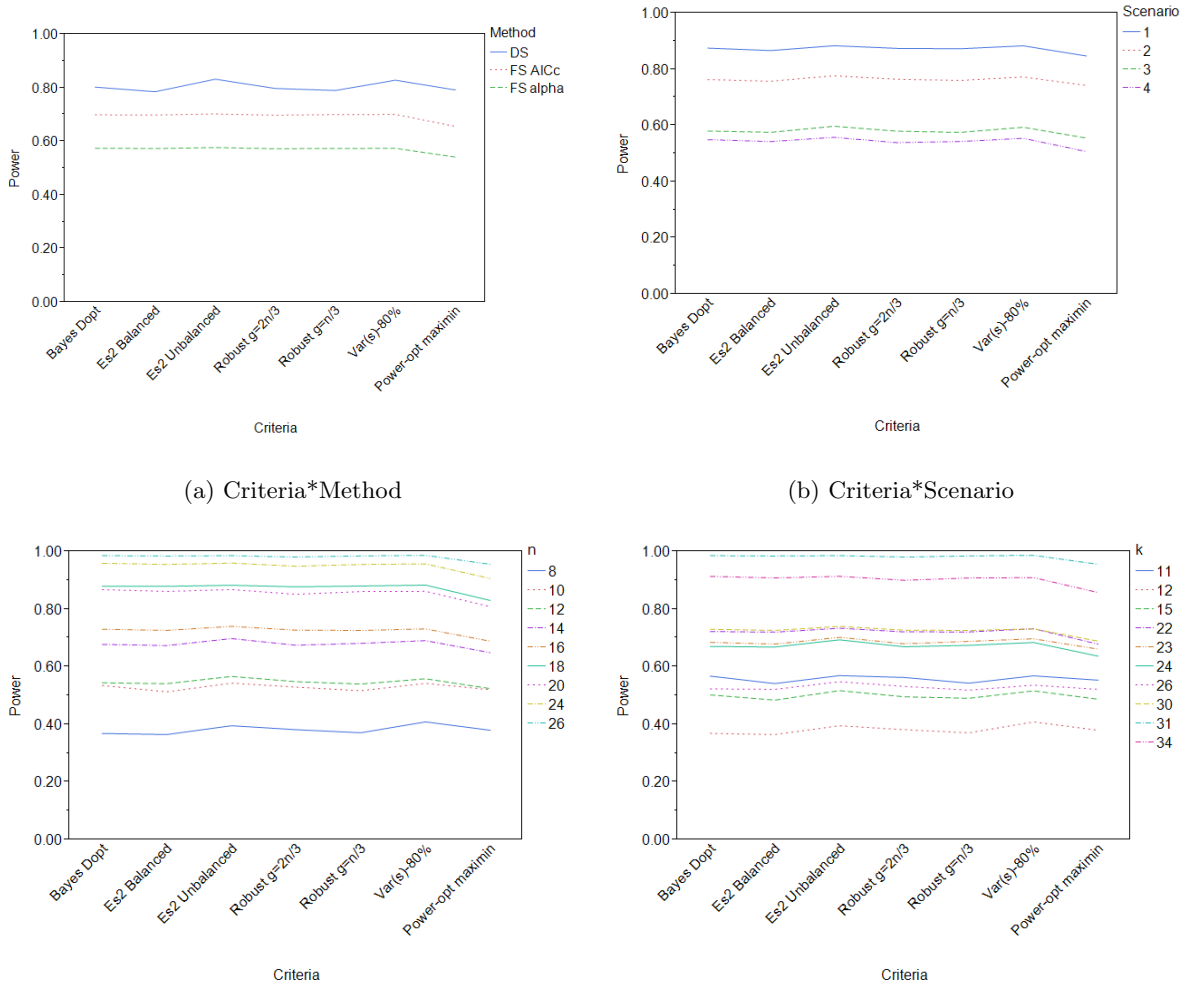
Figure 4a clearly displays the expected superiority of the Dantzig selector over forward selection. Additionally, we note the benefit of forward selection using AICc as a stopping criterion versus $\alpha_{enter} = 0.05$ with respect to power. Particularly, the Dantzig selector possesses an approximately 17% improvement in power versus forward selection with AICc and an approximate 41% improvement in power versus forward selection using $\alpha_{enter} = 0.05$. Figure 4a also serves to highlight the similarities of the design construction criteria with respect to power. A similar theme is depicted in Figures 4b,c,d in which simulation scenario, run size, and number of factors clearly have a significant impact (as is to be expected) on power. That is, higher values of power are seen as the SSD run size ($n$) increases (Figure 3c) while lower power is evident as the complexity of the model increases (Figure 3b). However, in each case, we continue to see that the effect of simulation scenario and SSD size on power are the same across the levels of design construction criteria. Briefly, we note that a similar

analysis for Type I error rates did not reveal differences among the SSD construction criteria. It is worth mentioning that, overall, the Dantzig Selector was indicated as possessing the lowest Type I error rate (mean of approximately 5%) while forward selection with AICc had the highest (mean of approximately 15%).

As stated previously, power is calculated in this article by simulating data from super-saturated experiments and measuring the proportion of correctly identified active effects. As an aside, we note that in addition to computing power in this manner, one might be interested in calculating power for an individual factor column. For instance, for a given effect size, if it is known that a specific column has a smaller power than others, one might be able to account for this in an analysis by, say, increasing the significance threshold for this specific term's entry into the model. In doing so, we are accepting an increased risk of making a Type I error due to the lower column power. For our simulations, we also investigated individual column power but again found no noticeable differences across the various design criteria.

Table 2: ANOVA Table for Analysis of Power Simulation Results

| Source | DF | Sum of Squares | F Ratio | P-value |
|---|---|---|---|---|
| Method | 2 | 246280.87 | 1405.24 | $< .0001$ |
| Criteria | 6 | 2683.26 | 5.10 | $< .0001$ |
| Scenario | 3 | 440192.83 | 1674.45 | $< .0001$ |
| $n$ (Number of Runs) | 1 | 242975.06 | 2772.75 | $< .0001$ |
| $k$ (Number of Factors) | 1 | 8418.56 | 96.07 | $< .0001$ |
| Criteria*Method | 12 | 1559.84 | 1.48 | 0.1242 |
| Criteria*Scenario | 18 | 738.90 | 0.47 | 0.9709 |
| Criteria*$n$ | 6 | 92.27 | 0.18 | 0.9835 |
| Criteria*$k$ | 6 | 259.90 | 0.49 | 0.8129 |
| Method*Scenario | 6 | 29798.99 | 56.68 | $< .0001$ |
| Method*$n$ | 2 | 5472.09 | 31.22 | $< .0001$ |
| Method*$k$ | 2 | 24771.59 | 141.34 | $< .0001$ |
| Scenario*$n$ | 3 | 80780.97 | 307.28 | $< .0001$ |
| Scenario*$k$ | 3 | 13015.67 | 49.51 | $< .0001$ |
| $n$*$k$ | 1 | 27065.41 | 308.86 | $< .0001$ |
| Error | 935 | 81933.60 | 87.60 | |
| Total | 1007 | 1889973.00 | | |




(a) Criteria*Method

(b) Criteria*Scenario

(c) Criteria*n

(d) Criteria*k

Figure 4: Interactions Involving Construction Criteria. Simulation scenario 1 corresponds to $\{a = 3, \mu = 5\}$; 2 corresponds to $\{a = 4, \mu = 4\}$; 3 corresponds to $\{a = 6, \mu = 3\}$; 4 corresponds to $\{a = 9, \mu = \text{mixed}\}$.

# 4 Discussion and Conclusions

In this article we have explored several well-established and new supersaturated design construction methods and evaluated the resulting designs via simulation in terms of their power to detect active factors. We have done this over 12 experiments that are of different

sizes and levels of supersaturation. We have largely followed the simulation protocol of Marley and Woods (2010) given their use of a wide range of scenarios involving the number and sizes of active factors.

Based on the simulation and the subsequent analysis, our general conclusion is that none of the tested supersaturated design construction criteria—whether established in the literature or proposed in this paper—are significantly and consistently better in terms of power than the others across the range of experiments and simulation scenarios considered. This deepens and confirms the results of Marley and Woods (2010), whose conclusions were based upon just three different experiments and two design construction methods. We also found, as in Marley and Woods (2010), that the Dantzig selector dominates forward selection as an analysis strategy. If forward selection is to be used, using AICc as a stopping rule will produce increased power over $\alpha_{enter} = 0.05$. We note again, however, that among the three analysis strategies forward selection with AICc as a stopping rule produced the highest Type I error rates on average.

We believe power to be the ultimate basis upon which to measure the effectiveness of a supersaturated design. Given our conclusion that most supersaturated design construction methods are indistinguishable in terms of this most important criterion, we believe that supersaturated designs can sensibly be chosen based upon convenience. For instance, the Bayesian D-optimal SSDs can be easily constructed in JMP and SAS for any reasonable number of runs and factors. On the other hand, software for constructing $E(s^2)$-optimal SSDs is not as accessible (though it does exist; see `http://www.designcomputing.net/gendex/`). The choice appears to have little significance.

The designs used in this article are available as supplementary material at `http://www.asq.org/pub/jqt/`.

# Appendix

Tables A.1-A.6 contain the tabulated results of our simulations as described in sections 3.2 and 3.3. Table A.7 contain the tabulated results of the column power simulation as described in section 3.2.

| | | | $a = 3$, $\mu = 5$ | | $a = 4$, $\mu = 4$ | | $a = 6$, $\mu = 4$ | | $a = 9$, $\mu =$ mixed | |
|---|---|---|---|---|---|---|---|---|---|---|
| $n$ | $k$ | Design Criterion | Power | Type I Error | Power | Type I Error | Power | Type I Error | Power | Type I Error |
| 8 | 12 | Balanced $E(s^2)$ | 0.7314 | 0.1272 | 0.5757 | 0.1751 | 0.4452 | 0.1803 | 0.4827 | 0.2550 |
| | | Bayes D-Optimal | 0.8007 | 0.0985 | 0.6419 | 0.1539 | 0.4304 | 0.1830 | 0.4731 | 0.2573 |
| | | Var(s)-80% | 0.8627 | 0.0767 | 0.7512 | 0.1176 | 0.5222 | 0.1627 | 0.5267 | 0.2324 |
| | | Unbalanced $E(s^2)$ | 0.8747 | 0.0653 | 0.7491 | 0.1138 | 0.5232 | 0.1699 | 0.5235 | 0.2483 |
| | | $\text{MRSS}_{n/3}$ | 0.7995 | 0.1092 | 0.6380 | 0.1630 | 0.4322 | 0.1863 | 0.4750 | 0.2614 |
| | | $\text{MRSS}_{2n/3}$ | 0.8385 | 0.0912 | 0.6907 | 0.1341 | 0.4538 | 0.1601 | 0.4911 | 0.2391 |
| | | Power-opt (maximin) | 0.7447 | 0.0994 | 0.6163 | 0.1469 | 0.4355 | 0.1920 | 0.4800 | 0.2600 |
| 10 | 11 | Balanced $E(s^2)$ | 0.9825 | 0.0080 | 0.9388 | 0.0354 | 0.7492 | 0.1516 | 0.6181 | 0.2887 |
| | | Bayes D-Optimal | 0.9872 | 0.0049 | 0.9522 | 0.0229 | 0.8538 | 0.0987 | 0.7130 | 0.2084 |
| | | Var(s)-80% | 0.9865 | 0.0051 | 0.9518 | 0.0235 | 0.8617 | 0.0900 | 0.7166 | 0.2135 |
| | | Unbalanced $E(s^2)$ | 0.9839 | 0.0061 | 0.9524 | 0.0228 | 0.8594 | 0.0904 | 0.7200 | 0.2121 |
| | | $\text{MRSS}_{n/3}$ | 0.9554 | 0.0171 | 0.8929 | 0.0519 | 0.7307 | 0.1755 | 0.6541 | 0.2953 |
| | | $\text{MRSS}_{2n/3}$ | 0.9999 | 0.0005 | 0.9850 | 0.0108 | 0.8376 | 0.0908 | 0.6401 | 0.2166 |
| | | Power-opt (maximin) | 0.9789 | 0.0181 | 0.9219 | 0.0508 | 0.7041 | 0.1830 | 0.6086 | 0.2943 |
| 10 | 15 | Balanced $E(s^2)$ | 0.9062 | 0.0395 | 0.7840 | 0.0791 | 0.5164 | 0.1371 | 0.5055 | 0.2081 |
| | | Bayes D-opt | 0.9592 | 0.0207 | 0.8587 | 0.0542 | 0.5676 | 0.1184 | 0.5241 | 0.1916 |
| | | Var(s)-80% | 0.9681 | 0.0155 | 0.8743 | 0.0492 | 0.6063 | 0.1298 | 0.5485 | 0.2016 |
| | | Unbalanced $E(s^2)$ | 0.9644 | 0.0173 | 0.8795 | 0.0497 | 0.6194 | 0.1192 | 0.5503 | 0.1979 |
| | | $\text{MRSS}_{n/3}$ | 0.9358 | 0.0276 | 0.7941 | 0.0724 | 0.5217 | 0.1318 | 0.5056 | 0.2010 |
| | | $\text{MRSS}_{2n/3}$ | 0.9475 | 0.0245 | 0.8256 | 0.0639 | 0.5327 | 0.1287 | 0.5094 | 0.1988 |
| | | Power-opt(maximin) | 0.9345 | 0.0351 | 0.8321 | 0.0633 | 0.5516 | 0.1114 | 0.5214 | 0.1842 |
| 12 | 22 | Balanced $E(s^2)$ | 0.9301 | 0.0179 | 0.7970 | 0.0463 | 0.5081 | 0.0876 | 0.4733 | 0.1447 |
| | | Bayes D-opt | 0.9522 | 0.0148 | 0.8290 | 0.0386 | 0.5132 | 0.0831 | 0.4753 | 0.1396 |
| | | Var(s)-80% | 0.9734 | 0.0100 | 0.8881 | 0.0276 | 0.5974 | 0.0709 | 0.5146 | 0.1254 |
| | | Unbalanced $E(s^2)$ | 0.9816 | 0.0070 | 0.8967 | 0.0249 | 0.6028 | 0.0704 | 0.5168 | 0.1255 |
| | | $\text{MRSS}_{n/3}$ | 0.9269 | 0.0196 | 0.8091 | 0.0421 | 0.5163 | 0.0784 | 0.4727 | 0.1367 |
| | | $\text{MRSS}_{2n/3}$ | 0.9574 | 0.0138 | 0.8306 | 0.0378 | 0.5301 | 0.0779 | 0.4810 | 0.1361 |
| | | Power-opt (maximin) | 0.9543 | 0.0201 | 0.8382 | 0.0379 | 0.5255 | 0.0706 | 0.4935 | 0.1220 |
| 12 | 26 | Balanced $E(s^2)$ | 0.8865 | 0.0233 | 0.7183 | 0.0500 | 0.4337 | 0.0789 | 0.4283 | 0.1300 |
| | | Bayes D-opt | 0.9083 | 0.0208 | 0.7396 | 0.0466 | 0.4390 | 0.0762 | 0.4350 | 0.1271 |
| | | Var(s)-80% | 0.9398 | 0.0150 | 0.8119 | 0.0342 | 0.4974 | 0.0681 | 0.4673 | 0.1153 |
| | | Unbalanced $E(s^2)$ | 0.9547 | 0.0121 | 0.8360 | 0.0310 | 0.5069 | 0.0687 | 0.4743 | 0.1161 |
| | | $\text{MRSS}_{n/3}$ | 0.8856 | 0.0250 | 0.7280 | 0.0471 | 0.4337 | 0.0752 | 0.4316 | 0.1255 |
| | | $\text{MRSS}_{2n/3}$ | 0.9287 | 0.0179 | 0.7918 | 0.0377 | 0.4672 | 0.0702 | 0.4502 | 0.1203 |
| | | Power-opt(maximin) | 0.9433 | 0.0174 | 0.8041 | 0.0355 | 0.4880 | 0.0681 | 0.4617 | 0.1168 |

Table A.1: Power and Type I error rates for Dantzig Selector and small supersaturated experiments.

| | | | $a = 3$, $\mu = 5$ | | $a = 4$, $\mu = 4$ | | $a = 6$, $\mu = 4$ | | $a = 9$, $\mu =$ mixed | |
|---|---|---|---|---|---|---|---|---|---|---|
| $n$ | $k$ | Design Criterion | Power | Type I Error | Power | Type I Error | Power | Type I Error | Power | Type I Error |
| 14 | 23 | Balanced $E(s^2)$ | 0.9911 | 0.0039 | 0.9381 | 0.0147 | 0.6797 | 0.0546 | 0.5421 | 0.1147 |
| | | Bayes D-opt | 0.9918 | 0.0035 | 0.9410 | 0.0139 | 0.6919 | 0.0495 | 0.5451 | 0.1063 |
| | | Var(s)-80% | 0.9955 | 0.0021 | 0.9701 | 0.0079 | 0.7763 | 0.0365 | 0.5799 | 0.0937 |
| | | Unbalanced $E(s^2)$ | 0.9979 | 0.0016 | 0.9750 | 0.0066 | 0.7749 | 0.0373 | 0.5786 | 0.0950 |
| | | $\text{MRSS}_{n/3}$ | 0.9945 | 0.0027 | 0.9563 | 0.0114 | 0.7113 | 0.0468 | 0.5532 | 0.1017 |
| | | $\text{MRSS}_{2n/3}$ | 0.9843 | 0.0055 | 0.9331 | 0.0155 | 0.6729 | 0.0503 | 0.5410 | 0.1070 |
| | | Power-opt (maximin) | 0.9908 | 0.0065 | 0.9515 | 0.0140 | 0.7269 | 0.0396 | 0.5605 | 0.0948 |
| 14 | 24 | Balanced $E(s^2)$ | 0.9857 | 0.0047 | 0.9290 | 0.0158 | 0.6607 | 0.0551 | 0.5299 | 0.1121 |
| | | Bayes D-opt | 0.9917 | 0.0035 | 0.9337 | 0.0150 | 0.6634 | 0.0529 | 0.5271 | 0.1097 |
| | | Var(s)-80% | 0.9965 | 0.0023 | 0.9631 | 0.0086 | 0.7370 | 0.0393 | 0.5606 | 0.0960 |
| | | Unbalanced $E(s^2)$ | 0.9969 | 0.0017 | 0.9675 | 0.0079 | 0.7533 | 0.0381 | 0.5674 | 0.0917 |
| | | $\text{MRSS}_{n/3}$ | 0.9945 | 0.0034 | 0.9352 | 0.0144 | 0.6585 | 0.0524 | 0.5290 | 0.1068 |
| | | $\text{MRSS}_{2n/3}$ | 0.9805 | 0.0063 | 0.9190 | 0.0175 | 0.6507 | 0.0531 | 0.5261 | 0.1077 |
| | | Power-opt (maximin) | 0.9887 | 0.0080 | 0.9390 | 0.0156 | 0.6791 | 0.0438 | 0.5382 | 0.0957 |
| 16 | 30 | Balanced $E(s^2)$ | 0.9903 | 0.0029 | 0.9477 | 0.0086 | 0.6915 | 0.0357 | 0.5320 | 0.0819 |
| | | Bayes D-opt | 0.9961 | 0.0019 | 0.9592 | 0.0072 | 0.7055 | 0.0323 | 0.5363 | 0.0764 |
| | | Var(s)-80% | 0.9969 | 0.0018 | 0.9681 | 0.0059 | 0.7614 | 0.0255 | 0.5641 | 0.0678 |
| | | Unbalanced $E(s^2)$ | 0.9973 | 0.0012 | 0.9789 | 0.0043 | 0.7705 | 0.0245 | 0.5638 | 0.0675 |
| | | $\text{MRSS}_{n/3}$ | 0.9931 | 0.0023 | 0.9548 | 0.0080 | 0.7007 | 0.0322 | 0.5412 | 0.0739 |
| | | $\text{MRSS}_{2n/3}$ | 0.9867 | 0.0037 | 0.9340 | 0.0104 | 0.6934 | 0.0330 | 0.5341 | 0.0787 |
| | | Power-opt(maximin) | 0.9922 | 0.0044 | 0.9556 | 0.0093 | 0.7128 | 0.0288 | 0.5411 | 0.0703 |
| 18 | 22 | Balanced $E(s^2)$ | 1.0000 | 0.0001 | 0.9995 | 0.0004 | 0.9765 | 0.0038 | 0.7598 | 0.0313 |
| | | Bayes D-opt | 1.0000 | 0.0000 | 1.0000 | 0.0002 | 0.9787 | 0.0031 | 0.7564 | 0.0282 |
| | | Var(s)-80% | 1.0000 | 0.0001 | 0.9999 | 0.0003 | 0.9836 | 0.0023 | 0.7863 | 0.0210 |
| | | Unbalanced $E(s^2)$ | 1.0000 | 0.0001 | 0.9989 | 0.0005 | 0.9852 | 0.0021 | 0.8062 | 0.0181 |
| | | $\text{MRSS}_{n/3}$ | 1.0000 | 0.0001 | 0.9997 | 0.0003 | 0.9768 | 0.0035 | 0.7633 | 0.0282 |
| | | $\text{MRSS}_{2n/3}$ | 0.9999 | 0.0002 | 0.9988 | 0.0007 | 0.9658 | 0.0050 | 0.7510 | 0.0305 |
| | | Power-opt (maximin) | 0.9995 | 0.0030 | 0.9941 | 0.0054 | 0.9497 | 0.0068 | 0.7001 | 0.0336 |
| 20 | 24 | Balanced $E(s^2)$ | 0.9991 | 0.0003 | 0.9946 | 0.0011 | 0.8932 | 0.0087 | 0.6309 | 0.0375 |
| | | Bayes D-opt | 0.9998 | 0.0002 | 0.9977 | 0.0007 | 0.9297 | 0.0053 | 0.6650 | 0.0267 |
| | | Var(s)-80% | 0.9998 | 0.0002 | 0.9976 | 0.0008 | 0.9266 | 0.0058 | 0.6602 | 0.0282 |
| | | Unbalanced $E(s^2)$ | 1.0000 | 0.0002 | 0.9969 | 0.0006 | 0.9325 | 0.0047 | 0.6680 | 0.0284 |
| | | $\text{MRSS}_{n/3}$ | 0.9996 | 0.0003 | 0.9957 | 0.0010 | 0.8925 | 0.0080 | 0.6274 | 0.0351 |
| | | $\text{MRSS}_{2n/3}$ | 0.9949 | 0.0012 | 0.9821 | 0.0028 | 0.8625 | 0.0105 | 0.6194 | 0.0390 |
| | | Power-opt (maximin) | 0.9989 | 0.0018 | 0.9940 | 0.0030 | 0.8815 | 0.0084 | 0.6200 | 0.0337 |
| 24 | 34 | Balanced $E(s^2)$ | 1.0000 | 0.0000 | 0.9999 | 0.0001 | 0.9887 | 0.0008 | 0.7745 | 0.0078 |
| | | Bayes D-opt | 1.0000 | 0.0000 | 1.0000 | 0.0001 | 0.9951 | 0.0003 | 0.8046 | 0.0050 |
| | | Var(s)-80% | 0.9999 | 0.0000 | 0.9999 | 0.0001 | 0.9938 | 0.0003 | 0.8086 | 0.0050 |
| | | Unbalanced $E(s^2)$ | 1.0000 | 0.0000 | 1.0000 | 0.0000 | 0.9948 | 0.0002 | 0.8073 | 0.0052 |
| | | $\text{MRSS}_{n/3}$ | 1.0000 | 0.0000 | 1.0000 | 0.0001 | 0.9889 | 0.0006 | 0.7741 | 0.0072 |
| | | $\text{MRSS}_{2n/3}$ | 0.9999 | 0.0002 | 0.9987 | 0.0004 | 0.9726 | 0.0014 | 0.7621 | 0.0101 |
| | | Power-opt (maximin) | 0.9994 | 0.0007 | 0.9988 | 0.0014 | 0.9675 | 0.0021 | 0.7128 | 0.0132 |
| 26 | 31 | Balanced $E(s^2)$ | 1.0000 | 0.0000 | 1.0000 | 0.0000 | 0.9994 | 0.0000 | 0.8848 | 0.0009 |
| | | Bayes D-opt | 1.0000 | 0.0000 | 1.0000 | 0.0000 | 0.9997 | 0.0000 | 0.9010 | 0.0005 |
| | | Var(s)-80% | 1.0000 | 0.0000 | 1.0000 | 0.0000 | 0.9997 | 0.0000 | 0.9062 | 0.0003 |
| | | Unbalanced $E(s^2)$ | 1.0000 | 0.0000 | 1.0000 | 0.0000 | 0.9997 | 0.0000 | 0.8944 | 0.0004 |
| | | $\text{MRSS}_{n/3}$ | 1.0000 | 0.0000 | 1.0000 | 0.0000 | 0.9996 | 0.0000 | 0.8826 | 0.0010 |
| | | $\text{MRSS}_{2n/3}$ | 1.0000 | 0.0000 | 0.9999 | 0.0001 | 0.9966 | 0.0003 | 0.8650 | 0.0021 |
| | | Power-opt (maximin) | 0.9999 | 0.0004 | 0.9995 | 0.0010 | 0.9960 | 0.0009 | 0.8141 | 0.0033 |

Table A.2: Power and Type I error rates for Dantzig Selector and larger supersaturated experiments.

| | | | $a = 3$, $\mu = 5$ | | $a = 4$, $\mu = 4$ | | $a = 6$, $\mu = 4$ | | $a = 9$, $\mu =$ mixed | |
|---|---|---|---|---|---|---|---|---|---|---|
| $n$ | $k$ | Design Criterion | Power | Type I Error | Power | Type I Error | Power | Type I Error | Power | Type I Error |
| 8 | 12 | Balanced $E(s^2)$ | 0.5097 | 0.1388 | 0.3644 | 0.1612 | 0.3108 | 0.1487 | 0.2903 | 0.1408 |
| | | Bayes D-opt | 0.5077 | 0.1214 | 0.3585 | 0.1413 | 0.2928 | 0.1528 | 0.2813 | 0.1330 |
| | | Var(s)-80% | 0.5419 | 0.1411 | 0.3791 | 0.1487 | 0.2910 | 0.1625 | 0.2823 | 0.1361 |
| | | Unbalanced $E(s^2)$ | 0.5125 | 0.1301 | 0.3735 | 0.1408 | 0.2939 | 0.1501 | 0.2787 | 0.1385 |
| | | $\text{MRSS}_{n/3}$ | 0.5161 | 0.1225 | 0.3547 | 0.1296 | 0.2957 | 0.1524 | 0.2847 | 0.1350 |
| | | $\text{MRSS}_{2n/3}$ | 0.4986 | 0.1244 | 0.3636 | 0.1424 | 0.2956 | 0.1493 | 0.2799 | 0.1341 |
| | | Power-opt (maximin) | 0.5025 | 0.1421 | 0.3888 | 0.1628 | 0.2969 | 0.1692 | 0.2761 | 0.1515 |
| 10 | 11 | Balanced $E(s^2)$ | 0.9059 | 0.1666 | 0.6005 | 0.0883 | 0.3555 | 0.1469 | 0.3974 | 0.0873 |
| | | Bayes D-opt | 0.9456 | 0.1822 | 0.6353 | 0.0947 | 0.3499 | 0.1052 | 0.4014 | 0.0845 |
| | | Var(s)-80% | 0.9520 | 0.1808 | 0.6373 | 0.0942 | 0.3530 | 0.1014 | 0.4004 | 0.0777 |
| | | Unbalanced $E(s^2)$ | 0.9500 | 0.1796 | 0.6360 | 0.0927 | 0.3552 | 0.1033 | 0.4014 | 0.0806 |
| | | $\text{MRSS}_{n/3}$ | 0.9393 | 0.1768 | 0.6539 | 0.0947 | 0.3620 | 0.1234 | 0.4050 | 0.0744 |
| | | $\text{MRSS}_{2n/3}$ | 0.9397 | 0.1620 | 0.6310 | 0.0875 | 0.3502 | 0.1392 | 0.3992 | 0.0870 |
| | | Power-opt (maximin) | 0.8533 | 0.1570 | 0.6224 | 0.1136 | 0.3772 | 0.1531 | 0.3821 | 0.1160 |
| 10 | 15 | Balanced $E(s^2)$ | 0.7802 | 0.1435 | 0.5160 | 0.1337 | 0.3519 | 0.1523 | 0.3633 | 0.1202 |
| | | Bayes D-opt | 0.7654 | 0.1537 | 0.5254 | 0.1404 | 0.3509 | 0.1612 | 0.3670 | 0.1228 |
| | | Var(s)-80% | 0.8266 | 0.1468 | 0.5419 | 0.1358 | 0.3529 | 0.1650 | 0.3699 | 0.1212 |
| | | Unbalanced $E(s^2)$ | 0.8009 | 0.1538 | 0.5411 | 0.1393 | 0.3517 | 0.1644 | 0.3701 | 0.1208 |
| | | $\text{MRSS}_{n/3}$ | 0.7901 | 0.1416 | 0.5246 | 0.1310 | 0.3450 | 0.1598 | 0.3648 | 0.1228 |
| | | $\text{MRSS}_{2n/3}$ | 0.7661 | 0.1527 | 0.5326 | 0.1411 | 0.3517 | 0.1578 | 0.3668 | 0.1250 |
| | | Power-opt (maximin) | 0.6562 | 0.1548 | 0.4936 | 0.1494 | 0.3468 | 0.1585 | 0.3380 | 0.1447 |
| 12 | 22 | Balanced $E(s^2)$ | 0.8208 | 0.1538 | 0.6521 | 0.1552 | 0.4035 | 0.1672 | 0.4102 | 0.1293 |
| | | Bayes D-opt | 0.8219 | 0.1529 | 0.6333 | 0.1584 | 0.3995 | 0.1677 | 0.4142 | 0.1277 |
| | | Var(s)-80% | 0.8224 | 0.1501 | 0.6432 | 0.1536 | 0.3995 | 0.1605 | 0.4055 | 0.1286 |
| | | Unbalanced $E(s^2)$ | 0.8249 | 0.1538 | 0.6459 | 0.1556 | 0.4051 | 0.1645 | 0.4094 | 0.1292 |
| | | $\text{MRSS}_{n/3}$ | 0.8251 | 0.1501 | 0.6425 | 0.1533 | 0.4087 | 0.1568 | 0.399 | 1 0.1297 |
| | | $\text{MRSS}_{2n/3}$ | 0.8252 | 0.1509 | 0.6252 | 0.1538 | 0.4009 | 0.1642 | 0.4086 | 0.1307 |
| | | Power-opt (maximin) | 0.7237 | 0.1502 | 0.5534 | 0.1475 | 0.3614 | 0.1503 | 0.3574 | 0.1376 |
| 12 | 26 | Balanced $E(s^2)$ | 0.7771 | 0.1423 | 0.5739 | 0.1476 | 0.3684 | 0.1596 | 0.3727 | 0.1322 |
| | | Bayes D-opt | 0.7595 | 0.1457 | 0.5675 | 0.1497 | 0.3585 | 0.1654 | 0.3848 | 0.1299 |
| | | Var(s)-80% | 0.7656 | 0.1429 | 0.5671 | 0.1474 | 0.3582 | 0.1572 | 0.3723 | 0.1308 |
| | | Unbalanced $E(s^2)$ | 0.7699 | 0.1456 | 0.5820 | 0.1489 | 0.3690 | 0.1600 | 0.3757 | 0.1299 |
| | | $\text{MRSS}_{n/3}$ | 0.7715 | 0.1397 | 0.5764 | 0.1451 | 0.3675 | 0.154 | 2 0.3682 | 0.1295 |
| | | $\text{MRSS}_{2n/3}$ | 0.7655 | 0.1439 | 0.5666 | 0.150 | 7 0.3661 | 0.1608 | 0.3750 | 0.1326 |
| | | Power-opt (maximin) | 0.7362 | 0.1390 | 0.5363 | 0.1433 | 0.3492 | 0.1496 | 0.3518 | 0.1290 |

Table A.3: Power and Type I error rates for the forward selection with AICc and small supersaturated experiments.

| | | | $a = 3$, $\mu = 5$ | | $a = 4$, $\mu = 4$ | | $a = 6$, $\mu = 4$ | | $a = 9$, $\mu =$ mixed | |
|---|---|---|---|---|---|---|---|---|---|---|
| $n$ | $k$ | Design Criterion | Power | Type I Error | Power | Type I Error | Power | Type I Error | Power | Type I Error |
| 14 | 23 | Balanced $E(s^2)$ | 0.9377 | 0.1697 | 0.8190 | 0.1594 | 0.5133 | 0.165 | 1 0.5026 | 0.1236 |
| | | Bayes D-opt | 0.9385 | 0.1707 | 0.8166 | 0.1632 | 0.5223 | 0.1660 | 0.5091 | 0.1189 |
| | | Var(s)-80% | 0.9392 | 0.1684 | 0.8161 | 0.1625 | 0.5125 | 0.1638 | 0.5008 | 0.1227 |
| | | Unbalanced $E(s^2)$ | 0.9357 | 0.1691 | 0.8235 | 0.1578 | 0.5247 | 0.1673 | 0.5064 | 0.1225 |
| | | $\text{MRSS}_{n/3}$ | 0.9342 | 0.1683 | 0.8146 | 0.1601 | 0.5078 | 0.1693 | 0.5041 | 0.1204 |
| | | $\text{MRSS}_{2n/3}3$ | 0.9390 | 0.1681 | 0.8205 | 0.1604 | 0.5246 | 0.1629 | 0.4945 | 0.1251 |
| | | Power-opt (maximin) | 0.8866 | 0.1644 | 0.7538 | 0.1564 | 0.4820 | 0.1554 | 0.4640 | 0.1333 |
| 14 | 24 | Balanced $E(s^2)$ | 0.9255 | 0.1673 | 0.7935 | 0.1599 | 0.4946 | 0.1698 | 0.4948 | 0.1226 |
| | | Bayes D-opt | 0.9259 | 0.1666 | 0.7952 | 0.1599 | 0.4950 | 0.1717 | 0.4964 | 0.1235 |
| | | Var(s)-80% | 0.9319 | 0.1647 | 0.7912 | 0.1596 | 0.4934 | 0.1651 | 0.4900 | 0.1236 |
| | | Unbalanced $E(s^2)$ | 0.9449 | 0.1669 | 0.8032 | 0.1596 | 0.5042 | 0.1704 | 0.4972 | 0.1243 |
| | | $\text{MRSS}_{n/3}$ | 0.9269 | 0.1658 | 0.8009 | 0.1600 | 0.5007 | 0.1675 | 0.4961 | 0.1237 |
| | | $\text{MRSS}_{2n/3}$ | 0.9265 | 0.1643 | 0.8068 | 0.1567 | 0.5107 | 0.1622 | 0.4860 | 0.1263 |
| | | Power-opt (maximin) | 0.8719 | 0.1600 | 0.7016 | 0.1568 | 0.4564 | 0.1565 | 0.4457 | 0.1353 |
| 16 | 30 | Balanced $E(s^2)$ | 0.9496 | 0.1732 | 0.8502 | 0.1634 | 0.5773 | 0.1718 | 0.5420 | 0.1315 |
| | | Bayes D-opt | 0.9593 | 0.1747 | 0.8537 | 0.1643 | 0.5647 | 0.1746 | 0.5480 | 0.1300 |
| | | Var(s)-80% | 0.9471 | 0.1716 | 0.8524 | 0.1617 | 0.5689 | 0.1687 | 0.5410 | 0.1297 |
| | | Unbalanced $E(s^2)$ | 0.9574 | 0.1726 | 0.8633 | 0.1622 | 0.5779 | 0.1714 | 0.5430 | 0.1323 |
| | | $\text{MRSS}_{n/3}$ | 0.9493 | 0.1734 | 0.8550 | 0.1625 | 0.5642 | 0.1725 | 0.5406 | 0.1319 |
| | | $\text{MRSS}_{2n/3}$ | 0.9501 | 0.1684 | 0.8631 | 0.1586 | 0.5928 | 0.1629 | 0.5354 | 0.1302 |
| | | Power-opt (maximin) | 0.9139 | 0.1582 | 0.7837 | 0.1539 | 0.5041 | 0.1588 | 0.4964 | 0.1358 |
| 18 | 22 | Balanced $E(s^2)$ | 0.9997 | 0.2164 | 0.9916 | 0.1938 | 0.8973 | 0.1637 | 0.7558 | 0.0973 |
| | | Bayes D-opt | 0.9998 | 0.2193 | 0.9935 | 0.1932 | 0.9033 | 0.1602 | 0.7603 | 0.0901 |
| | | Var(s)-80% | 0.9981 | 0.2190 | 0.9896 | 0.1949 | 0.9017 | 0.1640 | 0.7610 | 0.0956 |
| | | Unbalanced $E(s^2)$ | 0.9959 | 0.2202 | 0.9837 | 0.1974 | 0.8964 | 0.1657 | 0.7638 | 0.0992 |
| | | $\text{MRSS}_{n/3}$ | 0.9993 | 0.2194 | 0.9924 | 0.1956 | 0.9015 | 0.1558 | 0.7617 | 0.0923 |
| | | $\text{MRSS}_{2n/3}$ | 0.9985 | 0.2178 | 0.9891 | 0.1947 | 0.8941 | 0.1655 | 0.7583 | 0.1022 |
| | | Power-opt (maximin) | 0.9909 | 0.1954 | 0.9586 | 0.1865 | 0.7741 | 0.1634 | 0.6720 | 0.1233 |
| 20 | 24 | Balanced $E(s^2)$ | 0.9941 | 0.2051 | 0.9706 | 0.1850 | 0.8068 | 0.1731 | 0.7317 | 0.1338 |
| | | Bayes D-opt | 0.9923 | 0.2065 | 0.9694 | 0.1871 | 0.8158 | 0.1721 | 0.7327 | 0.1333 |
| | | Var(s)-80% | 0.9939 | 0.2000 | 0.9655 | 0.1829 | 0.8069 | 0.1694 | 0.7223 | 0.1336 |
| | | Unbalanced $E(s^2)$ | 0.9948 | 0.2064 | 0.9726 | 0.1865 | 0.8052 | 0.1736 | 0.7298 | 0.1339 |
| | | $\text{MRSS}_{n/3}$ | 0.9959 | 0.2044 | 0.9723 | 0.1849 | 0.8146 | 0.1701 | 0.7299 | 0.1331 |
| | | $\text{MRSS}_{2n/3}$ | 0.9907 | 0.1969 | 0.9636 | 0.1793 | 0.8125 | 0.1661 | 0.7149 | 0.1352 |
| | | Power-opt(maximin) | 0.9803 | 0.1721 | 0.9257 | 0.1644 | 0.7042 | 0.1563 | 0.6460 | 0.1299 |
| 24 | 34 | Balanced $E(s^2)$ | 1.0000 | 0.241 | 8 0.9984 | 0.2247 | 0.9685 | 0.1882 | 0.9292 | 0.1396 |
| | | Bayes D-opt | 0.9998 | 0.2451 | 0.9981 | 0.2243 | 0.9698 | 0.1880 | 0.9340 | 0.1376 |
| | | Var(s)-80% | 0.9995 | 0.2397 | 0.9981 | 0.2201 | 0.9691 | 0.1848 | 0.9274 | 0.1391 |
| | | Unbalanced $E(s^2)$ | 0.9999 | 0.2454 | 0.9987 | 0.2239 | 0.9729 | 0.1865 | 0.9367 | 0.1360 |
| | | $\text{MRSS}_{n/3}$ | 0.9996 | 0.2412 | 0.9982 | 0.2221 | 0.9673 | 0.1870 | 0.9301 | 0.1387 |
| | | $\text{MRSS}_{2n/3}$ | 0.9980 | 0.2323 | 0.9959 | 0.2168 | 0.9676 | 0.1834 | 0.9154 | 0.1412 |
| | | Power-opt (maximin) | 0.9984 | 0.1972 | 0.9867 | 0.1882 | 0.8918 | 0.1738 | 0.8277 | 0.1448 |
| 26 | 31 | Balanced $E(s^2)$ | 1.0000 | 0.2642 | 1.0000 | 0.2453 | 0.9964 | 0.2066 | 0.9886 | 0.1524 |
| | | Bayes D-opt | 1.0000 | 0.2657 | 0.9997 | 0.2452 | 0.9954 | 0.2078 | 0.9862 | 0.1528 |
| | | Var(s)-80% | 1.0000 | 0.2668 | 1.0000 | 0.2452 | 0.9968 | 0.2068 | 0.9892 | 0.1521 |
| | | Unbalanced $E(s^2)$ | 1.0000 | 0.2654 | 1.0000 | 0.2455 | 0.9962 | 0.2086 | 0.9903 | 0.1503 |
| | | $\text{MRSS}_{n/3}$ | 1.0000 | 0.2662 | 1.0000 | 0.2462 | 0.9978 | 0.2081 | 0.9902 | 0.1526 |
| | | $\text{MRSS}_{2n/3}$ | 0.9997 | 0.2587 | 0.9989 | 0.2405 | 0.9931 | 0.2060 | 0.9831 | 0.1550 |
| | | Power-opt (maximin) | 0.9997 | 0.2179 | 0.9977 | 0.2082 | 0.9721 | 0.1897 | 0.9328 | 0.1572 |

Table A.4: Power and Type I error rates for the forward selection with AICc and larger supersaturated experiments.

| | | | $a = 3$, $\mu = 5$ | | $a = 4$, $\mu = 4$ | | $a = 6$, $\mu = 4$ | | $a = 9$, $\mu =$ mixed | |
|---|---|---|---|---|---|---|---|---|---|---|
| $n$ | $k$ | Design Criterion | Power | Type I Error | Power | Type I Error | Power | Type I Error | Power | Type I Error |
| 8 | 12 | Balanced $E(s^2)$ | 0.2486 | 0.0385 | 0.1699 | 0.0576 | 0.1319 | 0.0500 | 0.1281 | 0.0514 |
| | | Bayes D-opt | 0.2532 | 0.0318 | 0.1611 | 0.0397 | 0.1108 | 0.0422 | 0.1244 | 0.0442 |
| | | Var(s)-80% | 0.3169 | 0.0465 | 0.1731 | 0.0490 | 0.1291 | 0.0506 | 0.1390 | 0.0473 |
| | | Unbalanced $E(s^2)$ | 0.2382 | 0.0326 | 0.1511 | 0.0390 | 0.1138 | 0.0425 | 0.1229 | 0.0455 |
| | | $\text{MRSS}_{n/3}$ | 0.2757 | 0.0268 | 0.1484 | 0.0321 | 0.1169 | 0.0412 | 0.1253 | 0.0426 |
| | | $\text{MRSS}_{2n/3}$ | 0.2565 | 0.0366 | 0.1659 | 0.0390 | 0.1241 | 0.0444 | 0.1321 | 0.0485 |
| | | Power-opt (maximin) | 0.2985 | 0.0560 | 0.2183 | 0.0651 | 0.1578 | 0.0658 | 0.1513 | 0.0617 |
| 10 | 11 | Balanced $E(s^2)$ | 0.4547 | 0.0234 | 0.2066 | 0.0078 | 0.1004 | 0.0233 | 0.2008 | 0.0252 |
| | | Bayes D-opt | 0.5351 | 0.0332 | 0.1661 | 0.0089 | 0.0750 | 0.0174 | 0.2074 | 0.0162 |
| | | Var(s)-80% | 0.5355 | 0.0370 | 0.1590 | 0.0101 | 0.0733 | 0.0153 | 0.2070 | 0.0159 |
| | | Unbalanced $E(s^2)$ | 0.5440 | 0.0380 | 0.1560 | 0.0090 | 0.0753 | 0.0156 | 0.2092 | 0.0163 |
| | | $\text{MRSS}_{n/3}$ | 0.4903 | 0.0297 | 0.1635 | 0.0127 | 0.0767 | 0.0211 | 0.2039 | 0.0134 |
| | | $\text{MRSS}_{2n/3}$ | 0.5106 | 0.0309 | 0.1808 | 0.0072 | 0.0843 | 0.0230 | 0.2071 | 0.0205 |
| | | Power-opt (maximin) | 0.5288 | 0.0396 | 0.2995 | 0.0261 | 0.1629 | 0.0392 | 0.2115 | 0.0371 |
| 10 | 15 | Balanced $E(s^2)$ | 0.4628 | 0.0495 | 0.2597 | 0.0419 | 0.1564 | 0.0438 | 0.2208 | 0.0540 |
| | | Bayes D-opt | 0.4904 | 0.0512 | 0.2562 | 0.0391 | 0.1534 | 0.0454 | 0.2213 | 0.0452 |
| | | Var(s)-80% | 0.4938 | 0.0488 | 0.2507 | 0.0371 | 0.1514 | 0.0452 | 0.2226 | 0.0461 |
| | | Unbalanced $E(s^2)$ | 0.5137 | 0.0527 | 0.2495 | 0.0353 | 0.1544 | 0.0448 | 0.2211 | 0.0476 |
| | | $\text{MRSS}_{n/3}$ | 0.4799 | 0.0457 | 0.2616 | 0.0377 | 0.1571 | 0.0474 | 0.2192 | 0.0510 |
| | | $\text{MRSS}_{2n/3}$ | 0.4929 | 0.0504 | 0.2568 | 0.0405 | 0.1588 | 0.0428 | 0.2206 | 0.0493 |
| | | Power-opt (maximin) | 0.4581 | 0.0648 | 0.3075 | 0.0654 | 0.2013 | 0.0654 | 0.2186 | 0.0679 |
| 12 | 22 | Balanced $E(s^2)$ | 0.7263 | 0.0944 | 0.4524 | 0.0749 | 0.2431 | 0.0705 | 0.3252 | 0.0775 |
| | | Bayes D-opt | 0.7231 | 0.0902 | 0.4585 | 0.0746 | 0.2444 | 0.0706 | 0.3283 | 0.0788 |
| | | Var(s)-80% | 0.7108 | 0.0878 | 0.4584 | 0.0788 | 0.2523 | 0.0744 | 0.3117 | 0.0776 |
| | | Unbalanced $E(s^2)$ | 0.7185 | 0.0914 | 0.4585 | 0.0789 | 0.2422 | 0.0738 | 0.3251 | 0.0773 |
| | | $\text{MRSS}_{n/3}$ | 0.7217 | 0.0938 | 0.4586 | 0.0771 | 0.2509 | 0.0710 | 0.3149 | 0.0778 |
| | | $\text{MRSS}_{2n/3}$ | 0.7229 | 0.0890 | 0.4399 | 0.0727 | 0.2442 | 0.0737 | 0.3206 | 0.0749 |
| | | Power-opt (maximin) | 0.5921 | 0.0866 | 0.4071 | 0.0826 | 0.2507 | 0.0795 | 0.2758 | 0.0841 |
| 12 | 26 | Balanced $E(s^2)$ | 0.7036 | 0.0972 | 0.4370 | 0.0869 | 0.2528 | 0.0830 | 0.3172 | 0.0931 |
| | | Bayes D-opt | 0.6813 | 0.1008 | 0.4358 | 0.0902 | 0.2587 | 0.0922 | 0.3218 | 0.0959 |
| | | Var(s)-80% | 0.6669 | 0.1006 | 0.4372 | 0.0890 | 0.2522 | 0.0870 | 0.3088 | 0.0935 |
| | | Unbalanced $E(s^2)$ | 0.6918 | 0.1022 | 0.4518 | 0.0931 | 0.2633 | 0.0887 | 0.3132 | 0.0938 |
| | | $\text{MRSS}_{n/3}$ | 0.6833 | 0.0982 | 0.4367 | 0.0884 | 0.2493 | 0.0815 | 0.3053 | 0.0916 |
| | | $\text{MRSS}_{2n/3}$ | 0.6757 | 0.0999 | 0.4324 | 0.0917 | 0.2585 | 0.0900 | 0.3150 | 0.0948 |
| | | Power-opt (maximin) | 0.6350 | 0.0929 | 0.4158 | 0.0895 | 0.2595 | 0.0886 | 0.2891 | 0.0902 |

Table A.5: Power and Type I error rates for the forward selection with $\alpha_{[}enter] = 0.05$ and small supersaturated experiments.

| | | | $a = 3$, $\mu = 5$ | | $a = 4$, $\mu = 4$ | | $a = 6$, $\mu = 4$ | | $a = 9$, $\mu =$ mixed | |
|---|---|---|---|---|---|---|---|---|---|---|
| $n$ | $k$ | Design Criterion | Power | Type I Error | Power | Type I Error | Power | Type I Error | Power | Type I Error |
| 14 | 23 | Balanced $E(s^2)$ | 0.8507 | 0.0971 | 0.6596 | 0.0833 | 0.3036 | 0.0664 | 0.4172 | 0.0701 |
| | | Bayes D-opt | 0.8759 | 0.1038 | 0.6696 | 0.0921 | 0.3114 | 0.0644 | 0.4147 | 0.0732 |
| | | Var(s)-80% | 0.8861 | 0.1036 | 0.6670 | 0.0887 | 0.3169 | 0.0667 | 0.4152 | 0.0738 |
| | | Unbalanced $E(s^2)$ | 0.8966 | 0.1038 | 0.6897 | 0.0888 | 0.3157 | 0.0661 | 0.4172 | 0.0699 |
| | | $\text{MRSS}_{n/3}$ | 0.8817 | 0.1034 | 0.6855 | 0.0911 | 0.3125 | 0.0680 | 0.4131 | 0.0710 |
| | | $\text{MRSS}_{2n/3}$ | 0.8635 | 0.1002 | 0.6700 | 0.0910 | 0.3227 | 0.0678 | 0.4006 | 0.0748 |
| | | Power-opt(maximin) | 0.8192 | 0.0984 | 0.6044 | 0.0858 | 0.3268 | 0.0761 | 0.3772 | 0.0766 |
| 14 | 24 | Balanced $E(s^2)$ | 0.8489 | 0.1013 | 0.6450 | 0.0888 | 0.3125 | 0.0739 | 0.4108 | 0.0763 |
| | | Bayes D-opt | 0.8543 | 0.1043 | 0.6502 | 0.0891 | 0.3095 | 0.0735 | 0.4156 | 0.0775 |
| | | Var(s)-80% | 0.8771 | 0.1024 | 0.6557 | 0.0898 | 0.3178 | 0.0727 | 0.4057 | 0.0777 |
| | | Unbalanced $E(s^2)$ | 0.8863 | 0.1058 | 0.6762 | 0.0927 | 0.3150 | 0.0716 | 0.4192 | 0.0782 |
| | | $\text{MRSS}_{n/3}$ | 0.8731 | 0.1057 | 0.6559 | 0.0932 | 0.3171 | 0.0732 | 0.4152 | 0.0762 |
| | | $\text{MRSS}_{2n/3}$ | 0.8559 | 0.1039 | 0.6558 | 0.0952 | 0.3239 | 0.0716 | 0.4028 | 0.0765 |
| | | Power-opt(maximin) | 0.7764 | 0.0970 | 0.5696 | 0.0911 | 0.3212 | 0.0839 | 0.3630 | 0.0818 |
| 16 | 30 | Balanced $E(s^2)$ | 0.9299 | 0.1253 | 0.7953 | 0.1207 | 0.4310 | 0.0983 | 0.4882 | 0.0977 |
| | | Bayes D-opt | 0.9337 | 0.1258 | 0.7964 | 0.1214 | 0.4251 | 0.0989 | 0.4926 | 0.0960 |
| | | Var(s)-80% | 0.9181 | 0.1173 | 0.7766 | 0.1173 | 0.4157 | 0.0938 | 0.4764 | 0.0942 |
| | | Unbalanced $E(s^2)$ | 0.9355 | 0.1259 | 0.7926 | 0.1201 | 0.4220 | 0.0990 | 0.4895 | 0.0999 |
| | | $\text{MRSS}_{n/3}$ | 0.9306 | 0.1269 | 0.7872 | 0.1210 | 0.4224 | 0.0984 | 0.4824 | 0.0958 |
| | | $\text{MRSS}_{2n/3}$ | 0.9339 | 0.1175 | 0.8001 | 0.1174 | 0.4388 | 0.0963 | 0.4722 | 0.0973 |
| | | Power-opt (maximin) | 0.8697 | 0.1062 | 0.6831 | 0.1008 | 0.3882 | 0.0922 | 0.4335 | 0.0939 |
| 18 | 22 | Balanced $E(s^2)$ | 0.9984 | 0.1032 | 0.9630 | 0.1023 | 0.6167 | 0.0624 | 0.6018 | 0.0434 |
| | | Bayes D-opt | 0.9978 | 0.1057 | 0.9705 | 0.1007 | 0.6121 | 0.0588 | 0.5959 | 0.0379 |
| | | Var(s)-80% | 0.9956 | 0.1045 | 0.9649 | 0.1007 | 0.6325 | 0.0637 | 0.5969 | 0.0409 |
| | | Unbalanced $E(s^2)$ | 0.9913 | 0.1059 | 0.9582 | 0.1031 | 0.6196 | 0.0662 | 0.6030 | 0.0462 |
| | | $\text{MRSS}_{n/3}$ | 0.9930 | 0.1062 | 0.9644 | 0.1005 | 0.6222 | 0.0597 | 0.5973 | 0.0429 |
| | | $\text{MRSS}_{2n/3}$ | 0.9950 | 0.1059 | 0.9630 | 0.1038 | 0.6219 | 0.0680 | 0.6023 | 0.0502 |
| | | Power-opt (maximin) | 0.9630 | 0.1007 | 0.8729 | 0.0964 | 0.5505 | 0.0758 | 0.5440 | 0.0660 |
| 20 | 24 | Balanced $E(s^2)$ | 0.9913 | 0.1346 | 0.9606 | 0.1365 | 0.7058 | 0.1204 | 0.6684 | 0.1094 |
| | | Bayes D-opt | 0.9919 | 0.1366 | 0.9534 | 0.1380 | 0.7072 | 0.1211 | 0.6681 | 0.1082 |
| | | Var(s)-80% | 0.9899 | 0.1323 | 0.9467 | 0.1313 | 0.6859 | 0.1162 | 0.6542 | 0.1044 |
| | | Unbalanced $E(s^2)$ | 0.9921 | 0.1344 | 0.9587 | 0.1342 | 0.7092 | 0.1203 | 0.6658 | 0.1068 |
| | | $\text{MRSS}_{n/3}$ | 0.9919 | 0.1352 | 0.9561 | 0.1339 | 0.7035 | 0.1201 | 0.6648 | 0.1046 |
| | | $\text{MRSS}_{2n/3}$ | 0.9852 | 0.1294 | 0.9425 | 0.1260 | 0.7144 | 0.1232 | 0.6436 | 0.1060 |
| | | Power-opt (maximin) | 0.9577 | 0.1052 | 0.8717 | 0.1074 | 0.5638 | 0.0956 | 0.5698 | 0.0930 |
| 24 | 34 | Balanced $E(s^2)$ | 0.9999 | 0.1339 | 0.9976 | 0.1318 | 0.9268 | 0.1228 | 0.8842 | 0.1122 |
| | | Bayes D-opt | 0.9996 | 0.1345 | 0.9976 | 0.1302 | 0.9345 | 0.1263 | 0.8876 | 0.1122 |
| | | Var(s)-80% | 0.9996 | 0.1302 | 0.9971 | 0.1330 | 0.9192 | 0.1234 | 0.8789 | 0.1164 |
| | | Unbalanced $E(s^2)$ | 1.0000 | 0.1314 | 0.9957 | 0.1306 | 0.9384 | 0.1240 | 0.8817 | 0.1152 |
| | | $\text{MRSS}_{n/3}$ | 0.9995 | 0.1289 | 0.9962 | 0.1282 | 0.9321 | 0.1252 | 0.8824 | 0.1131 |
| | | $\text{MRSS}_{2n/3}$ | 0.9972 | 0.1283 | 0.9931 | 0.1267 | 0.9268 | 0.1262 | 0.8673 | 0.1168 |
| | | Power-opt (maximin) | 0.9934 | 0.1108 | 0.9658 | 0.1151 | 0.7931 | 0.1104 | 0.7493 | 0.1003 |
| 26 | 31 | Balanced $E(s^2)$ | 1.0000 | 0.1220 | 0.9997 | 0.1203 | 0.9845 | 0.1181 | 0.9664 | 0.1093 |
| | | Bayes D-opt | 0.9998 | 0.1261 | 0.9994 | 0.1206 | 0.9869 | 0.1179 | 0.9677 | 0.1105 |
| | | Var(s)-80% | 0.9996 | 0.1226 | 0.9998 | 0.1200 | 0.9884 | 0.1191 | 0.9696 | 0.1119 |
| | | Unbalanced $E(s^2)$ | 1.0000 | 0.1221 | 0.9999 | 0.1230 | 0.9866 | 0.1175 | 0.9682 | 0.1088 |
| | | $\text{MRSS}_{n/3}$ | 1.0000 | 0.1234 | 0.9999 | 0.1217 | 0.9867 | 0.1170 | 0.9676 | 0.1080 |
| | | $\text{MRSS}_{2n/3}$ | 0.9989 | 0.1241 | 0.9981 | 0.1201 | 0.9831 | 0.1209 | 0.9612 | 0.1105 |
| | | Power-opt (maximin) | 0.9987 | 0.1142 | 0.9902 | 0.1134 | 0.9038 | 0.1124 | 0.8708 | 0.1034 |

Table A.6: Power and Type I error rates for the forward selection with $\alpha_{[}enter] = 0.05$ and larger supersaturated experiments.

## Acknowledgments

The authors would like to thank Phil Scinto for his encouragement, and for providing the impetus to begin this study. The authors also express gratitude to the editor for his helpful comments and suggestions.

## References

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

Allen, T. and M. Bernshteyn (2003). Supersaturated designs that maximize the probabitliy of identifying active factors. *Technometrics 45*(1), 90–97.

Beattie, S., D. K. Fong, and D. Lin (2002). A two-stage bayesian model selection strategy for suerpsaturated designs. *Technometrics 44*, 55–63.

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

Candes, E. and T. Tao (2007). The dantzig selector: Statistical estimations when $p$ is much larger than $n$. *The Annals of Statistics 35(6)*, 2313–2351.

Cela, R., E. Martinez, and A. Carro (2001). Supersaturated experimental designs: New approaches to building and using it, part ii: Solving supersaturated designs by genetic algorithms. *Chemometrics and Intelligent Laboratory Systems 57*, 75–92.

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

Chipman, H., M. Hamada, and C. Wu (1997). Bayesian variable selection for designed experiments with complex aliasing. *Technometrics 39*, 372–381.

Dejaegher, B. and Y. Vander Heyden (2008). Supersaturated designs: Set-ups, data interpretation, and analytical applications. *Annals of Bioanalytical Chemistry 390*, 1227–1240.

DuMouchel, W. and B. Jones (1994). A simple Bayesian modification of D-optimal designs to reduce dependence on an assumed model. *Technometrics 36(1)*, 37–47.

Gilmour, S. (2006). Factor screening via supersaturated designs. In A. Dean and S. Lewis (Eds.), *Screening: Methods for Experimentation in Industry, Drug Discovery, and Genetics*, pp. 169–190. Springer.

Holcomb, D., D. Montgomery, and W. Carlyle (2003). Analysis of supersaturated designs. *Journal of Qualtiy Technology 1*(35), 13–27.

Holcomb, D., D. Montgomery, and W. Carlyle (2007). The use of supersaturated experiments in turbine engine development. *Quality Engineering 19*, 17–27.

Jones, B., C. Nachtsheim, and D. Lin (2008). Bayesian D-optimal supersaturated desings. *Journal of Statistical Planning and Inference 138*, 86–92.

Jones, B., C. N. W. Li, and K. Ye (2009). Model robust supersaturated and partially supersaturated designs. *Journal of Statistical Planning and Inference 139*, 45–53.

Kelly, H. and J. Voelkel (2000). Asymptotic-power problems in the anlaysis of supersaturated designs. *Statistics and Probablity Letters 47*, 317–324.

Li, R. and D. Lin (2002). Data analysis in supersaturated designs. *Statistics and Probablity Letters 59*, 135–144.

Li, W. and C. J. Nachtsheim (2000). Model-robust factorial designs. *Technometrics 42(4)*, 345–352.

Li, W. and C. Wu (1997). Columnwise-pairwise alogrithims with applications to the construction of supersaturated designs. *Technometrics*.

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

Liu, Y., S. Ruan, and A. Dean (2007). Construction and analysis of E$s^2$ efficient supersaturated designs. *Journal of Statistical Planning and Inference 137*, 1516–1529.

Lu, X. and X. Wu (2004). A strategy of searching active factors in supersaturated screening experiments. *Journal of Quality Technology 36*, 392–399.

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

Mee, R. (2009). *A Comprehensive Guide to Factorial Two-Level Experimentation*. Springer.

Meyer, R. K. and C. J. Nachtsheim (1995). The coordinate-exchange algorithm for constructing exact optimal experimental designs. *Technometrics 37(1)*, 60–69.

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

Nguyen, N. and C. Cheng (2008). New e($s^2$)-optimal supersaturated designs constructed from incomplete block designs. *Technometrics 50*, 26–31.

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

Ryan, K. and D. Bulutoglu (2007). $e(s^2)$-optimal supersaturated designs with good minimax properties. *Journal of Statistical Planning and Inference 137*, 2250–2262.

Satterthwaite, F. (1959). Random balance experimentation. *Technometrics 1*, 111–137.

Scinto, P., R. Wilkinson, and D. Lin (2011). Screening for fuel economy: A case study of supersaturated designs in practice. *Quality Engineering 23*, 15–25.

Smucker, B. and N. Drew (2014). Approximate model spaces for model-robust experiment design. *Technometrics*, In Press.

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

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

Zhang, Q., R. Zhang, and M. Liu (2006). A method for screening active factors in super-saturated designs. *Journal of Statistical Planning and Inference 137*, 2068–2079.