Statistica Sinica **13**(2003), 727-737

ON ADAPTIVE ESTIMATION IN ORTHOGONAL SATURATED DESIGNS

Weizhen Wang and Daniel T. Voss

Wright State University

Abstract: A simple method is provided to construct a general class of individual and simultaneous confidence intervals for the effects in orthogonal saturated designs. These intervals use the data adaptively, maintain the confidence levels sharply at $1-\alpha$ at the least favorable parameter configuration, work effectively under effect sparsity, and include the intervals by Wang and Voss (2001) as a special case.

Key words and phrases: Effect sparsity, factorial design, minimum function, symmetric unimodal distribution.

1. Introduction

Unreplicated factorial designs are extremely useful in industrial experimentation to identify active effects at low costs. Often the number of observations is just enough to estimate parameters for mean response, so one can obtain an estimator for each effect but have no degrees of freedom to estimate the variance. For example, consider a single replicate or orthogonal fraction of a 2^k factorial design yielding observations $Y_1, \ldots, Y_n$, assumed to be independently normally distributed with variance σ^2. The design is said to be *saturated* if the factorial effect contrasts, $\mu_1, \ldots, \mu_p$ say, are estimable and $n = p + 1$. We then have n observations and want to make inferences on $n-1$ parameters of interest $\mu_1, \ldots, \mu_p$, with μ_0 and σ^2 as nuisance parameters. Henceforth we refer to the factorial effect contrasts μ_i, $1 \le i \le p$, simply as "effects". Let X_i denote the least squares estimator of μ_i. The design is said to be *orthogonal* if the estimators $X_1, \ldots, X_p$ of the effects are uncorrelated. In most cases, unreplicated factorial designs are orthogonal and saturated. Under normality the estimators X_i are independent. Furthermore, $X_i \sim N(\mu_i, a^2\sigma^2)$ for known constant a. Without loss of generality, we take $a^2 = 1$. In a more general setting, let f_i be the pdf of a continuous, unimodal distribution which is symmetric about zero with finite variance, $1 \le i \le p$. Assume independent estimators $X_1, \ldots, X_p$, where

$$X_i \sim \frac{1}{\sigma} f_i\left(\frac{x_i - \mu_i}{\sigma}\right) \tag{1}$$

for unknown $\mu_1, \ldots, \mu_p$ and σ. The goal of this paper is to construct confidence intervals for the effects, $\mu_1, \ldots, \mu_p$ under the model (1). Lacking an independent variance estimator, the analysis is based solely on $X_1, \ldots, X_p$. This can be done by assuming *effect sparsity* — namely, most of the effects μ_i are zero (or negligible).

There are two primary concerns about the desired confidence intervals: (i) control of the error rate, and (ii) effective use of the data. We say intervals control the error rate at level $1-\alpha$ if the minimum or infimum over all parameter configurations of the coverage probability of the intervals is $1-\alpha$. Hochberg and Tamhane (1987, p.3) call this *strong control of error rates*. Due to effect sparsity, most of X_i's have mean $\mu_i = 0$. Effective use of the data means many of the X_i's which have mean zero go into the estimation of σ, though which ones and how many to use are unknown. Intervals are called *adaptive* if they use the data to determine which and how many X_i's should be used to estimate σ. Such adaptive intervals are typically narrower than those using a fixed number of X_i's to estimate σ, they are more efficient.

Many confidence intervals have been proposed in orthogonal saturated designs. See Voss (1999), Voss and Wang (1999), Lenth (1989) Juan and Peña (1992), Dong (1993) and Haaland and O'Connell (1995). The first two papers propose intervals controlling the error rate but do not use the data adaptively, while the others obtain intervals that use the data adaptively but do not show that the error rate is controlled at level $1-\alpha$. For more results on this topic, see the extensive reviews by Hamada and Balakrishnan (1998) and Kinateder, Voss and Wang (2000). Wang and Voss (2001) derived intervals that control the error rate and use the data adaptively by constructing an estimator of σ^2 on each set of a partition of the sample space. Constants are chosen so that the resultant estimator is monotone increasing in each of the $|X_i|$'s. However, their method depends heavily on the initial guess on the number of X_i's used to estimate σ. If one knows from past experience that it is very likely that either 8 or 12 out of the total 15 effects are zero, for example, Wang and Voss's (2001) interval cannot utilize such information well.

In this paper we provide a class of confidence intervals, both individual and simultaneous, which control error rates and use data adaptively for the analysis of orthogonal saturated designs. These intervals overcome the problem mentioned above, Wang and Voss's intervals are included as a special case, and they can be constructed easily. Individual and simultaneous confidence intervals are derived in Sections 2 and 3, respectively. Individual confidence intervals are illustrated in Section 4. Finally, competing methods are compared with respect to power in Section 5.

2. Individual Confidence Intervals

In this section, we discuss how to construct the individual confidence interval for each effect μ_i, without loss of generality μ_p. Intuitively, one should estimate μ_p by X_p and estimate σ by combining X_1 through X_{p-1}. Denote the vector of effects by $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_p)$, with $\boldsymbol{\mu}_0 = (0, \ldots, 0)$ representing the null case. A function $G(x_1, \ldots, x_{p-1})$ is symmetric about zero if $G(x_1, \ldots, x_{p-1}) = G(|x_1|, \ldots, |x_{p-1}|)$.

Theorem 1. *Suppose $F(x_p)$ and $G(x_1, \ldots, x_{p-1})$ are nonnegative functions satisfying*

(i) $G(x_1, \ldots, x_{p-1})$ *is symmetric about zero,* $G(x_1, \ldots, x_{p-1}) = G(|x_1|, \ldots, |x_{p-1}|)$, *and nondecreasing in $|x_i|$ for each $1 \le i \le p-1$ when the variables x_j $(j \neq i)$ are held fixed;*

(ii) $F(ax_p)/G(ax_1, \ldots, ax_{p-1}) = F(x_p)/G(x_1, \ldots, x_{p-1})$, *for any $a > 0$.*

Then for any positive constant d, $P_{\mu,\sigma}(F(X_p - \mu_p)/G(X_1, \ldots, X_{p-1}) \geq d)$ depends on its parameters through $\mu_1/\sigma, \ldots, \mu_{p-1}/\sigma$, and is non-increasing in each $|\mu_i/\sigma|$ when the others are fixed. Therefore $P_{\mu_o,\sigma}(F(X_p - \mu_p)/G(X_1, \ldots, X_{p-1}) \geq d) = \sup_{\mu,\sigma} P_{\mu,\sigma}(F(X_p - \mu_p)/G(X_1, \ldots, X_{p-1}) \geq d) = \alpha$ say, so that

$$\left\{\mu_p : \frac{F(X_p - \mu_p)}{G(X_1, \ldots, X_{p-1})} \leq d\right\} \tag{2}$$

is a confidence set for μ_p with confidence coefficient $1 - \alpha$.

Proof. It is clear that the distribution of

$$Q = \frac{F(X_p - \mu_p)}{G(|X_1|, \ldots, |X_{p-1}|)} = \frac{F((X_p - \mu_p)/\sigma)}{G(|X_1|/\sigma, \ldots, |X_{p-1}|/\sigma)}$$

depends on the parameters through $|\mu_1/\sigma|, \ldots, |\mu_{p-1}/\sigma|$ because of (ii) and conditions on the f_i. Since $X_1, \ldots, X_p$ are independent, Q is non-increasing as a function of $|x_i|$ for each $i < p$, and each $|X_i|/\sigma$ $(i < p)$ is stochastically nondecreasing in $|\mu_i/\sigma|$, the distribution of Q is stochastically non-increasing in each $|\mu_i/\sigma|$.

Theorem 2. *Suppose $F(x_p)$ and $G_j(x_1, \ldots, x_{p-1})$ for $1 \le j \le p-1$ are nonnegative functions. Let*

$$G = \min_{1 \le j \le p-1} G_j. \tag{3}$$

If each pair (F, G_j) satisfies conditions (i) *and* (ii) *in Theorem* 1*, so does the pair (F, G). Therefore, a confidence set for μ_p with confidence coefficient $1 - \alpha$ is given by* (2).

Proof. Since G is the minimum function and each G_j is nondecreasing in $|x_i|$, G is nondecreasing in $|x_i|$ as well. It is clear that F and G satisfy the rest of conditions in Theorem 1, and we establish the claim.

Typically each G_j is an estimator of σ or σ^2 using a fixed number of X_i's—it is not an adaptive one. The minimum function compares all G_j's and chooses the smallest, which most likely only involves those X_i's with mean 0. Therefore G uses the data adaptively, as shown in the following examples. Let $|X|_{(i)}$ be the ith order statistic of $|X_1|, \ldots, |X_{p-1}|$.

Example 1. Let

$$SS_j = \sum_{h=1}^{j} |X|^2_{(h)} \tag{4}$$

denote the sum of squares of the j smallest of these order statistics, with observed value $ss_j = \sum_{h=1}^{j} |x|^2_{(h)}$ for $1 \le j \le p-1$. Define

$$F(x_p) = x_p^2, \quad G_j(x_1, \ldots, x_{p-1}) = \frac{ss_j}{K_j}, \tag{5}$$

where K_j's are nonnegative constants. Then the functions F, G_j and $G_{SN} = \min_{1 \le j \le p-1} G_j$ satisfy the conditions in Theorem 2. The confidence set for μ_p in (2) reduces to a confidence interval of the form:

$$X_p \pm \sqrt{dG_{SN}(X_1, \ldots, X_{p-1})}. \tag{6}$$

This interval should be used if X_i's are i.i.d. standard normal.

Each G_j in (5) is exchangeable in the components x_i. Suppose in addition the same functions G_j are used to obtain the confidence interval for each effect μ_i. These conditions are sufficient for the p confidence intervals for $\mu_1, \ldots, \mu_p$ to be consistent in the following sense — if $|x_i| > |x_j|$ and the confidence interval for μ_i contains zero, then the confidence interval for μ_j contains zero.

The larger K_j is in (5), the larger chance G_{SN} has to be G_j, which should be used when there are exactly j negligible effects. This provides a guide to choosing the K_j's based on any existing knowledge concerning the likely number of negligible effects. If one wants to be able to use each G_j, i.e., if $P(G_{SN} = G_j)$ is to be positive for each j, then necessarily $K_{j+1} \ge K_j(1 + 1/j)$. Let $D_j = \{(x_1, \ldots, x_{p-1}) : G_j < G_i \ \forall\ i \ne j\}$ for $1 \le j < p$. Then $G_{SN} = G_j$ on D_j.

Wang and Voss (2001) provide an adaptive estimator G_{WV} for σ^2 where

$$G_{WV}(x_1, \ldots, x_{p-1}) = \frac{ss_m}{1 + (m - \nu)c_\nu}, \tag{7}$$

$$m = \begin{cases} p-1, \text{ if } |x|^2_{(i+1)} < c_i ss_i, \quad \forall\ i = \nu, \ldots, p-2 \\ \min\{i : i \ge \nu,\ |x|^2_{(i+1)} \ge c_i ss_i\}, \text{ otherwise,} \end{cases}$$

for $c_i = c_\nu/[1 + (i - \nu)c_\nu]$ and for ν a positive integer and c_ν a positive constant. Here it is anticipated that at least ν effects are negligible. Roughly speaking, Wang and Voss (2001) compare SS_j only with SS_{j-1} and SS_{j+1} at best. In contrast, SS_j is compared with all SS_i's in this paper. Let $A_\nu = \{(x_1, \ldots, x_{p-1}) : |x|^2_{(\nu+1)} \geq c_\nu ss_\nu\}$, $A_j = \{(x_1, \ldots, x_{p-1}) : |x|^2_{(i+1)} < c_i ss_i\}$ for $\nu < j < p-1$, and $A_{p-1} = \{(x_1, \ldots, x_{p-1}) : |x|^2_{(i+1)} < c_i ss_i\}$. The A_j's, $\nu \leq j \leq p-1$, form a partition of R^{p-1} and $G_{WV} = ss_j/[(1+(j-\nu)c_\nu)]$ on A_j. The methods of this paper include those of Wang and Voss (2001) as a special case, as established by the following result.

Theorem 3. *If we define* $K_j = 0$ *for* $1 \leq j < \nu$ *and* $K_j = (1+(j-\nu)c_\nu)$ *for* $\nu \leq j \leq p-1$, *then* A_j *is contained in* $\bar{D}_j$, *the closure of* D_j, *for* $\nu \leq j \leq p-1$, *and* $G_{SN} = G_{WV}$.

Proof. Note that D_j is empty if $j < \nu$. It is clear that $\bar{D}_j = \{(x_1, \ldots, x_{p-1}) : G_j \leq G_i \forall i \neq j\}$. For $\nu \leq j < p-2$, fix $(x_1, \ldots, x_{p-1}) \in A_j$. For $i > j$, $G_i = (ss_j + \sum_{h=j+1}^{i} |x|^2_{(h)})/(1+(i-\nu)c_\nu) \geq (ss_j + (i-j)|x|^2_{(j+1)})/(1+(i-\nu)c_\nu) \geq (ss_j(1+(i-j)c_j))/(1+(i-\nu)c_\nu) = G_j$; For $i \leq j$, $G_i = (ss_{i-1} + |x|^2_{(i)})/(1+(i-\nu)c_\nu) \leq (ss_{i-1} + c_{i-1}ss_{i-1})/(1+(i-\nu)c_\nu) = G_{i-1}$, and then $G_j \leq G_{j-1} \leq \ldots \leq G_i$. Therefore, A_j is a subset of $\bar{D}_j$. Similarly, one can show that A_{p-1} is a subset of $\bar{D}_{p-1}$. Since on each A_j, $G_{SN} = ss_j/K_j = G_{WV}$ and all A_j's form a partition, we conclude that $G_{SN} = G_{WV}$.

In Wang and Voss's (2001) interval, one can only choose one constant c_ν, and the constant d is determined by the confidence level—the method provides little flexibility. For example, if $p = 15$ and we believe that either 8 or 12 effects are negligible but are not sure which is the case, we can choose $\nu = 8$ and a large c_8 (or a large K_8), so G_{WV} has a big chance to be SS_8/K_8. However, c_{12} (or K_{12}) is determined by c_8 (or K_8) and cannot be large enough for G_{WV} to have a big chance to be SS_{12}/K_{12}, which it should, and so the resultant confidence interval tends to be wider. For the current interval, since K_8 and K_{12} are functionally unrelated, one can choose K_8 and K_{12} to balance between the chances of G_{SN} being SS_8/K_8 or SS_{12}/K_{12} as one sees fit.

In fact, the current interval can handle even more complicated cases and can also be considered as a Bayesian approach in which one has a prior distribution π on the true number N of zero effects. More precisely, let $\pi_j = P(N = j)$ for $1 \leq j \leq p-1$. Determine the K_j's by solving $P_{\mu_o}(D_j) = \pi_j$, $1 \leq j \leq p-1$. This is not easily done, however.

Alternatively, here is a frequentist approach for selecting the K_j's. Anticipate that ν of effects are negligible−typically, ν is at least $(p+1)/2$ — and let $K_j = 0$ for $j < \nu$. One can then determine K_j for $j \geq \nu$ by solving $E_{\mu_o} G_j = \sigma^2$. Thus, each G_j for $j \geq \nu$ is an unbiased estimator of σ^2 under the null case. Variations on this approach are considered in the power study in Section 4.

Example 2. Define $F(x_p) = |x_p|$, $G_j(x_1, \ldots, x_{p-1}) = |x|_{(j)}/K_j$, where K_j's are nonnegative constants. Then the functions F, G_j and $G_U = \min_{1 \le j \le p-1} G_j$ satisfy the conditions in Theorem 2. The confidence set in (2) reduces to a confidence interval of the form $X_p \pm dG_U(X_1, \ldots, X_{p-1})$. This interval should be used if X_i's are from uniform distributions on intervals $[\mu_i - \sigma, \mu_i + \sigma]$. If a specific combination of ν of the μ_i's were known to be zero, the MLE for σ would be the maximum of the corresponding ν absolute effect estimates. This motivates the choice of G_U, not knowing which or how many effects are zero.

Example 3. Define $F(x_p) = |x_p|$, $G_j(x_1, \ldots, x_{p-1}) = \sum_{h=1}^{j} |x|_{(h)}/K_j$, where K_j's are nonnegative constants. Then the functions F, G_j and $G_{DE} = \min_{1 \le j \le p-1} G_j$ satisfy the conditions in Theorem 2. The confidence set in (2) reduces to a confidence interval of the form $X_p \pm dG_{DE}(X_1, \ldots, X_{p-1})$. This interval should be used if X_i's are from double exponential distributions, $f_i(x) = (1/2)e^{-|x|}$. This choice of G_{DE} is reasonable because, if a specific combination of ν of the μ_i's were known to be zero, the MLE for σ would be the mean of the corresponding ν absolute effect estimates.

3. Simultaneous Confidence Intervals

To construct simultaneous confidence intervals for $\{\mu_1, \ldots, \mu_p\}$, we follow the method of Voss and Wang (1999), omitting the proof.

Let $\hat{\underline{x}}_i = (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_p)$, for $1 \le i \le p$. Note that $(x_1, \ldots, x_{p-1}) = \hat{\underline{x}}_p$ and $G(X_1, \ldots, X_{p-1}) = G(\hat{\underline{X}}_p)$.

Theorem 4. *Suppose $F(x_i)$ and $G_j(\hat{\underline{x}}_i)$ for $1 \le j \le p-1$ are nonnegative functions. Each pair (F, G_j) satisfies conditions* (i) *and* (ii) *in Theorem* 1. *Define G as in* (3), *$V_i = F(X_i - \mu_i)/G(\hat{\underline{X}}_i)$ for $1 \le i \le p$, and let $W = \max_{1 \le i \le p} V_i$. Then $P_{\mu_o,\sigma}(W \ge d') = \sup_{\mu,\sigma} P_{\mu,\sigma}(W \ge d') = \alpha$ say, where d' is a constant, and*

$$\Big\{\mu_i : \frac{F(X_i - \mu_i)}{G(\hat{\underline{X}}_i)} \le d'\Big\}, \tag{8}$$

$1 \le i \le p$, are simultaneous confidence sets for $\mu_1, \ldots, \mu_p$ with simultaneous confidence coefficient $1 - \alpha$.

The simultaneous confidence sets (8) reduce to confidence intervals if the underlying distribution f_i is any of the examples in the previous section. Furthermore, if the same exchangeable functions G_j are used for each effect μ_i, then the simultaneous confidence intervals are consistent, as were the individual confidence intervals.

4. An Example

We illustrate the proposed methodology using a 2^4 experiment from Davies (1954), which served as "Example IV" in the papers of Box and Meyer (1986) and Lenth (1989). The four factors are acid strength (S), acid amount (A), time (M) and temperature (T), and the response measured is the yield of isatin. Table 1 contains the design, data and some statistics, with the estimates and squared estimates sorted by magnitude.

Table 1. A 2^4 experiment: Example IV of Box and Meyer (1986).

S	A	M	T	Yield	Effect	Estimate	(Estimate)^2	ss_j
-1	-1	-1	-1	0.08	S*M	-0.00125	0.00000156	
1	-1	-1	-1	0.04	S*A*T	-0.00625	0.00003906	
-1	1	-1	-1	0.53	S*A*M*T	0.01875	0.00035156	
1	1	-1	-1	0.43	M	-0.02125	0.00045156	
-1	-1	1	-1	0.31	A*T	-0.02625	0.00068906	
1	-1	1	-1	0.09	S*A	0.03375	0.00113906	
-1	1	1	-1	0.12	A*M	-0.06625	0.00438906	
1	1	1	-1	0.36	A	-0.07625	0.00581406	$ss_8 = 0.0128750$
-1	-1	-1	1	0.79	S*M*T	-0.10125	0.01025156	
1	-1	-1	1	0.68	A*M*T	0.12375	0.01531406	
-1	1	-1	1	0.73	S*A*M	0.14875	0.02212656	
1	1	-1	1	0.08	S*T	-0.16125	0.02600156	$ss_{12} = 0.0865687$
-1	-1	1	1	0.77	S	-0.19125	0.03657656	
1	-1	1	1	0.38	M*T	-0.25125	0.06312656	
-1	1	1	1	0.49	T	0.27375	0.07493906	
1	1	1	1	0.23				

Apply the methodology of Example 1, using $K_8 = 1.8495$, $K_{12} = 6.9898$ and $K_j = 0$ otherwise. (The values of K_ν for $\nu = 8, 12$ were obtained as the average value of ss_ν in (4) computed for 100,000 pseudo-random samples of size 14.) For the three effects with largest estimates, $G_{SN} = \min\{ss_8/K_8, ss_{12}/K_{12}\} = \min\{0.0128750/1.8495, 0.0865687/6.9898\} \approx \min\{0.006961, 0.01239\} = 0.006961$. For individual 95% confidence intervals, the critical value d in equation (6) is the 95th percentile of the null distribution of $F(X_{15})/G_{SN}(X_1, \ldots, X_{14}) = X_{15}^2/\min\{SS_8/1.8495, SS_{12}/6.9898\}$, and we obtained the estimate $d = 6.1639$ based on 99,999 pseudo-random samples. The minimum significant difference for the confidence interval in equation (6) becomes $\sqrt{dG_{SN}} = \sqrt{(6.1639)(0.006961)} \approx 0.2071$. Thus, the main effect of T is significantly positive and the M*T interaction effect is significantly negative, but the main effect of S is not significant. Because the method is consistent, no other effects will be significantly nonzero. Note that if more effects were to be considered, the values of ss_8 and ss_{12} would be larger, as they would be computed from the other 14 estimates−namely, excluding the estimate of the effect for which the confidence interval is being constructed.

For sake of comparison, also apply Lenth's (1989) method to these data. His method yields the same initial and adaptive estimate of the standard deviation of the estimators — $\hat{\sigma} = (1.5)(0.07625) \approx 0.1144$. The minimum significant difference for each 95% confidence interval is then $2.12053\hat{\sigma} \approx 0.2425$. Here the critical value 2.12053 is an estimate of the upper 95th percentile of the null distribution of $|x_p|/\hat{\sigma}$ based on 99,999 pseudo-random samples generated under the null distribution. The same two effects are significantly nonzero.

5. Power Study

In this section, three variations on the method of this paper are compared for power with competing adaptive and non-adaptive methods from the literature. Power was estimated by simulation for $p = 15$ estimators, as one would have for example in the analysis of a regular orthogonal 2^{15-11}_{III} fraction. Included were 42 parameter configurations, including from one to seven non-zero effects each of the same size, with effect sizes from one to six standard deviations of the estimators. For each of these 42 parameter configurations, 100,000 samples of size 15 were generated. For each sample, each of 11 methods was used to construct an individual 95% confidence interval for the nonzero effect μ_p. The power estimate for each method and parameter configuration is the fraction of confidence intervals excluding zero. The methods compared will now be described.

Consider first the variations on the method of this paper. The basic method is outlined in Example 1 and requires only the specification of the constants K_j of equation (5). The variation labeled WV2:u2 uses values of K_8 and K_{12} chosen so that SS_8/K_8 and SS_{12}/K_{12} are each unbiased for the estimator variance σ^2 under the null distribution, with $K_j = 0$ otherwise. Thus, the denominator adaptively chooses between the use of the 8 or 12 smallest sums of squares. The variation labeled WV2:u7 is similar but uses values of K_j chosen so that SS_j/K_j is unbiased for σ^2 for each $j \geq 8$ under the null distribution, with $K_j = 0$ for $j < 8$. The method labeled WV2:b7 is a variation on WV2:u7, multiplying the terms $SS_8/K_8, \ldots, SS_{14}/K_{14}$ of WV2:u7 by the factors $1.0, 1.1, \ldots, 1.6$, respectively, to bias the method in favor of using denominator SS_j/K_j for smaller j.

WV1 denotes the method of Wang and Voss (2001), which is a restricted case of the method of this paper given in (7).

V:8, V:12 and V:14 denote the non-adaptive method of Voss (1999). Specifically, V:ν is the method of Example 1, with $K_\nu = \nu$ for $j = \nu$ and $k_j = 0$ otherwise in equation (5). For V:14, the confidence interval in equation (6) is precisely the standard t-interval with 14 degrees of freedom.

"Lenth" denotes the popular method of Lenth (1989), for which σ is initially estimated by $\hat{\sigma}_0 = 1.5 \times$ median $\{|x_i|\}$ using all 15 absolute estimates, then one obtains and uses the adaptive *pseudo standard error* $\hat{\sigma} = 1.5\times$ median $\{|x_i| \ : \ |x_i| \leq 2.5\hat{\sigma}_0\}$. The confidence interval for μ_p is $x_p \pm c_\alpha \hat{\sigma}$, where the critical value c_α is obtained by simulation under the null distribution. "Lenth:I" denotes a variation on this in which $\hat{\sigma}$ is computed from $x_1, \ldots, x_{p-1}$, so X_p and $\hat{\sigma}$ are independent.

"Dong" denotes the method of Dong (1993). Dong uses the same initial estimate of σ as does Lenth, but then an adaptive estimate of σ^2 is computed as $\hat{\sigma}^2 = SS_\nu/\nu$, where $\nu = |\{x_i : |x_i| \leq 2.5\hat{\sigma}_0\}|$. Again, the critical value is computed by simulation under the null distribution. "Dong:I" denotes a variation on this in which $\hat{\sigma}^2$ is computed from $x_1, \ldots, x_{p-1}$, so X_p and $\hat{\sigma}$ are independent.

The results of the power study are summarized in Table 2. Marginal mean power is given for each effect size averaging over the number of active (i.e., nonzero) effects, and for each number of active effects averaging over effect sizes. The overall mean power averages over all 42 parameter configurations. The methods are sorted by their values of *maximum percentage power loss*, computed as follows. For each of the 42 parameter configurations, the percentage power loss of a given method was computed from its power and the power of the best method as (("best power" − "power")/"best power"). For each method, the maximum of the corresponding 42 values is reported. Thus, the WV2:u2 method is minimax of the 11 methods considered−namely, it minimizes the maximum loss of power over the 42 parameter configurations, suffering only a 10.3% power loss at worst.

Some further observations can be made from Table 2. The first six methods listed are all competitive in terms of average power. Not surprisingly, the non-adaptive methods V:12 and V:14 are best (or essentially best) for three and one active effects, respectively, but the methods break down for more active effects. The non-adaptive method V:8 does surprisingly well even when the number of active effects is small. It is interesting that WV2:u2 mixes V:8 and V:12 so as to maintain the good overall mean power of V:8 but with improved maximum percentage power loss.

While the reported simulation results are condensed, the complete results provide further insight. The top four methods with respect to maximum percentage power loss−WV2:u2, WV2:b7, V:8, and WV2:u7−maintain good power across parameter configurations. The Lenth and Lenth:I methods are comparable to one another and perform very well when there are at least four active effects of size three or more. Surprisingly, the WV1 method breaks down when there are seven active effects of size at least three, though it does very well anytime the number of active effects is at most five (covering most cases of typical interest) or the effect size is at most two. The Dong and Dong:I methods apparently suffer some from the inclusion of too many terms in the denominator, though they do very well when there are up to three large effects.

A few summarizing comments are now in order. We have attempted to compare the methods fairly, in the sense that each of the methods have a natural common breakdown point of eight or more large active effects, (except V:12 and V:14 which break down sooner). Of the adaptive methods considered, the WV1 and WV2:ν methods are known to control error rates over all parameter configurations, whereas it remain an open problem to show that the methods of Lenth (1989) and Dong (1993) enjoy the same property. In view of this, and since the WV2:u2 method has competitive overall mean power and is minimax in the sense discussed, it is reasonable to advocate use of the WV2:u2 method or similar methods.

Table 2. Power comparison of 11 adaptive and non-adaptive methods.

Method	Max % Power Loss	Mean Power													
			Effect Size						Number of Active Effects						
		Overall	1	2	3	4	5	6	1	2	3	4	5	6	7
WV2:u2	0.103	0.553	0.11	0.25	0.47	0.69	0.85	0.94	0.71	0.68	0.64	0.58	0.52	0.44	0.31
WV2:b7	0.124	0.556	0.11	0.25	0.47	0.70	0.86	0.95	0.70	0.67	0.64	0.69	0.53	0.45	0.32
V:8	0.132	0.556	0.11	0.25	0.47	0.70	0.86	0.95	0.69	0.67	0.63	0.59	0.53	0.45	0.33
WV2:u7	0.149	0.550	0.11	0.26	0.47	0.69	0.85	0.93	0.71	0.68	0.64	0.59	0.52	0.43	0.30
Lenth	0.186	0.552	0.11	0.25	0.47	0.70	0.85	0.93	0.69	0.67	0.64	0.60	0.54	0.44	0.28
LenthI	0.191	0.559	0.11	0.24	0.47	0.71	0.88	0.95	0.68	0.66	0.64	0.60	0.55	0.47	0.33
DongI	0.575	0.525	0.11	0.25	0.45	0.65	0.80	0.89	0.71	0.68	0.64	0.59	0.50	0.36	0.19
Dong	0.624	0.510	0.12	0.25	0.44	0.63	0.77	0.86	0.72	0.68	0.64	0.57	0.47	0.33	0.17
WV1	0.685	0.527	0.12	0.26	0.46	0.66	0.79	0.87	0.71	0.68	0.65	0.60	0.52	0.39	0.14
V:14	0.988	0.343	0.12	0.24	0.36	0.42	0.45	0.47	0.73	0.63	0.46	0.29	0.16	0.09	0.05
V:12	0.998	0.410	0.12	0.26	0.42	0.53	0.57	0.58	0.72	0.69	0.64	0.47	0.22	0.09	0.04

The Lenth:I and Dong:I variations of the respective methods of Lenth (1989) and Dong (1993) were considered for the following reason. They are based on a pivotal quantity for which the numerator and denominator are independent. This property makes the problem of establishing strong control of error rates more tractable, though this remains an open problem for these methods. In view of this, it is interesting to note that the operating characteristics of both methods are little affected by this variation.

References

Box, G. E. P. and Meyer, R. D. (1986). An analysis for unreplicated fractional factorials, *Technometrics* **28**, 11-18.

Davies, O. L. (ed.) (1954), *The Design and Analysis of Industrial Experiments*, Oliver and Boyd, London.

Dong, F. (1993). On the identification of active contrasts in unreplicated fractional factorials. *Statist. Sinica* **3**, 209-217.

Haaland, P. D. and O'Connell, M. A. (1995). Inference for effect-saturated fractional factorials. *Technometrics* **37**, 82-93.

Hamada, M. and Balakrishnan, N. (1998). Analyzing unreplicated factorial experiments: A review with some new proposals. *Statist. Sinica* **8**, 1-41.

Hochberg, Y. and Tamhane, A. C. (1987). *Multiple Comparison Procedures*. Wiley, New York.

Juan, J. and Peña, D. (1992). A simple method to identify significant effects in unreplicated two-level factorial designs. *Comm. Statist. Theory Methods* **21**, 1383-1403.

Kinateder, K. K. J., Voss, D. T. and Wang, W. (2000). Analysis of saturated and super-saturated factorial designs: A review. In *Proceedings of the Indian International Statistical Association* 1998 *International Conference* (Edited by N. Balakrishnan), 325-347. Gordon and Breach, New Jersey.

Lenth, R. V. (1989). Quick and easy analysis of unreplicated factorials. *Technometrics* **31**, 469-473.

Voss, D. T. (1999). Analysis of orthogonal saturated designs. *J. Statist. Plann. Inference* **78**, 111-130.

Voss, D. T. and Wang, W. (1999). Simultaneous confidence intervals in the analysis of orthogonal saturated designs. *J. Statist. Plann. Inference* **81**, 383-392.

Wang, W. and Voss, D. T. (2001). Control of Error Rates in Adaptive Analysis of Orthogonal Saturated Designs. *Ann. Statist.* **29**, 1058-1065.

Department of Mathematics and Statistics, Wright State University, Dayton, OH 45435, U.S.A.

E-mail: wwang@math.wright.edu

Department of Mathematics and Statistics, Wright State University, Dayton, OH 45435, U.S.A.

E-mail: dan.voss@wright.edu

(Received January 2002; accepted April 2003)