Nonparametric Statistics, 2002, Vol. 14(6), pp. 699–714
Taylor & Francis
Taylor & Francis Group

A NONPARAMETRIC DISPERSION TEST FOR UNREPLICATED TWO-LEVEL FRACTIONAL FACTORIAL DESIGNS

RICHARD N. MCGRATH[a,*] and DENNIS K. J. LIN[b,†]

[a]*Department of Applied Statistics and Operations Research, Bowling Green State University, Bowling Green, OH 43403;* [b]*Department of Management Science and Information Systems, The Pennsylvania State University, University Park, PA 16802*

(Received February 1999; In final form October 2000)

A consistent product/process will have little variability, *i.e.* dispersion. The widely-used unreplicated two-level fractional factorial designs can play an important role in detecting dispersion effects with a minimum expenditure of resources. In this paper we develop a nonparametric dispersion test for unreplicated two-level fractional factorial designs. The test statistic is defined, critical values are provided, and large sample approximations are given. Through simulations and examples from the literature, the test is compared to general nonparametric dispersion tests and a parametric test based on a normality assumption. These comparisons show the test to be the most robust of those studied and even superior to the normality-based test under normality in some situations. An example is given where this new test is the only one of those studied that does not incorrectly detect a spurious dispersion effect.

Keywords: Dispersion; Fractional factorial; Nonparametric; Scale

1 INTRODUCTION

Unreplicated two-level fractional factorial designs are often used in industry to study factors' effects on the mean of a response. If k factors, each with two levels, are studied by obtaining a single observation at each of the 2^k factor-level combinations, then the design is called a 2^k factorial design. (The two levels of each factor are commonly labeled $+1$ and -1). In a 2^k experiment, $n = 2^k$ independent location effects can be estimated: the overall mean, and $\binom{k}{j}$ j-factor interactions, $j = 2, \ldots, k$. The effect matrix is the resulting $n \times n$ matrix of $+1$s and -1s. Table I shows an effect matrix for a 16 run experiment.

Replicating the experiment, *i.e.* obtaining $r \geq 2$ observations at each design point, allows estimation of variance at each design point as well as an error term for testing for active location effects. Often, however, replication is too expensive due to limited resources. Sometimes even an unreplicated factorial may not be feasible. A 2^{k-p} fractional factorial design is created by assigning the k factors to the columns of a 2^{k-p} effect matrix. In these designs,

* Corresponding author. E-mail: rnmcgra@cba.bgsu.edu
† E-mail: DKL5@psu.edu

ISSN 1048-5252 print; ISSN 1029-0311 online
DOI: 10.1080/1048525021000008518

TABLE I Experimental Designs and Responses.

M		*A*	*B*	*C*	*D*							*E*		*G*	*F*			
D		*A*	*B*	*C*	*D*	*AB*	*AC*	*AD*	*BC*	*BD*	*CD*	*DE*	*CE*	*BE*	*AE*	*E*		
	0	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*	y_D	y_M
1	1	−1	−1	−1	−1	1	1	1	1	1	1	−1	−1	−1	−1	1	201.5	6
2	1	1	−1	−1	−1	−1	−1	−1	1	1	1	1	1	1	−1	−1	178.0	10
3	1	−1	1	−1	−1	−1	1	1	−1	−1	1	1	1	−1	1	−1	183.5	32
4	1	1	1	−1	−1	1	−1	−1	−1	−1	1	−1	−1	1	1	1	176.0	60
5	1	−1	−1	1	−1	1	−1	1	−1	1	−1	1	−1	1	1	−1	188.5	4
6	1	1	−1	1	−1	−1	1	−1	−1	1	−1	−1	1	−1	1	1	178.5	15
7	1	−1	1	1	−1	−1	−1	1	1	−1	−1	−1	1	1	−1	1	174.5	26
8	1	1	1	1	−1	1	1	−1	1	−1	−1	1	−1	−1	−1	−1	196.5	60
9	1	−1	−1	−1	1	1	1	−1	1	−1	−1	−1	1	1	1	−1	255.5	8
10	1	1	−1	−1	1	−1	−1	1	1	−1	−1	1	−1	−1	1	1	240.5	12
11	1	−1	1	−1	1	−1	1	−1	−1	1	−1	1	−1	1	−1	1	208.5	34
12	1	1	1	−1	1	1	−1	1	−1	1	−1	−1	1	−1	−1	−1	244.0	60
13	1	−1	−1	1	1	1	−1	−1	−1	−1	1	1	1	−1	−1	1	274.0	16
14	1	1	−1	1	1	−1	1	1	−1	−1	1	−1	−1	1	−1	−1	257.5	5
15	1	−1	1	1	1	−1	−1	−1	1	1	1	−1	−1	−1	1	−1	256.0	37
16	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	274.5	52

p denotes the degree of fractionization. For example, if 7 factors are studied in $n = 16$ runs, the design is an unreplicated 2^{7-3}, or a $1/2^3 = 1/8$ fraction of the 2^7 factorial. In unreplicated designs, an assumption of homogeneity of variance is often made in order to study location effects.

Suppose the experimenter wishes to study seven factors labeled A, B, C, D, E, F, and G, *i.e.* a 2^{7-3}. Row M in Table I shows these labels. The experiment would be run by setting each factor at the level shown and observing the response for a given row. For example, the first combination has all factors at their "low" or -1 level. In practice, the 16 combinations would be performed in random order in an attempt to balance out any noise due to a time effect, for example.

The other columns in the matrix that are not labeled represent interactions between the factors being studied. In a fractional factorial design, each column actually represents a string of effects. In other words, main (factor) effects are confounded (confused) with one or more interactions and all interactions are confounded with other interactions. For example, the column used to estimate the interaction between factors A and B, AB, can be obtained by multiplying the A and B columns resulting in column 5. One can verify that the CE and FG interactions also appear in this column. Thus, a fractional factorial creates a confounding or alias relationship among effects.

Ordinary least squares (OLS) analysis is generally used to estimate the location effects associated with each column. Using $i = 1, \ldots, n$ to index the $n = 2^{k-p}$ rows of the design matrix and $j = 0, \ldots, n-1$ to index the columns, column j then is $x_j = (x_{1j}, x_{2j}, \ldots, x_{nj})'$ and the OLS estimate (location effect estimate) for column j is $\hat{\beta}_j = \sum_{i=1}^{n} x_{ij}y_i/n$. When effects in all columns are of interest (common in these designs), there is no error term available for traditional hypothesis testing and alternate methods of detecting these location effects are necessary. Daniel [2, 3] suggested normal probability plotting of the coefficients. Many other approaches have since been developed. Examples include Box and Meyer [4] and Lenth [5]. See Hamada and Balakrishnan [6] for an overview and comparison of methods. All of these methods assume homogeneous variances. As this paper is focused on dispersion effects, we will not dwell on these methods. In order to properly study dispersion effects, however, location effects must first be identified. We will assume that some method has been used to identify location effects considered to be active (nonzero).

Studying dispersion effects is also important in industrial applications. Knowledge of a factor's effect on variation can be used to make a more consistent product or process. In a 2^{k-p} design, a dispersion effect occurs when the variance of the response is different at the low level (−1) of a factor than at its high level (+1). Let the true response in row i be $Y_i = \sum_{j=0}^{n-1} x_{ij}\beta_j + \varepsilon_i$, $\mathrm{Var}(\varepsilon_i) = \sigma_i^2$. Define $\sigma_{d+}^2 = (2/n)\sum_i(\sigma_i^2 \mid x_{id} = +1)$ and $\sigma_{d-}^2 = (2/n)\sum_i(\sigma_i^2 \mid x_{id} = -1)$. Then a dispersion effect occurs if $\sigma_{d+}^2 \neq \sigma_{d-}^2$. Informally, plots of residuals from the fitted location model against the levels of each factor can be used to subjectively assess any difference in variance. Box and Meyer [7] and Montgomery [8] are among those who have developed more formal methods for studying dispersion (variance) effects in addition to location (mean) effects in these designs. Bergman and Hynén [9] and McGrath and Lin [10] showed that the traditional F test is appropriate to test for a dispersion effect in a design column of a 2^{k-p} experiment under a specific condition. This condition results in all residuals at the +1 level of a column being uncorrelated with all residuals at the −1 level of the same column, although residuals are still correlated within the +1 and −1 levels. Under the assumption of normally and independently distributed errors, the +1 residuals are independent of the −1 residuals. Thus the ratio of the sample variances at the +1 and −1 levels of the column to be tested can be compared to an F distribution. However, it is well known that this test is very sensitive to the normality assumption. Hence the motivation for a non-parametric dispersion test for unreplicated 2^{k-p} designs.

Many have used nonparametric approaches to study location effects in factorial designs. The Kruskal-Wallis and Friedman tests are rank tests used for detecting location effects in one-factor and two-factor designs respectively. Groggel and Skillings [11] developed a location effect test for multifactor designs. We have not found any work in the literature that uses a nonparametric approach to study dispersion in unreplicated fractional factorial designs. This idea is apparently new.

This paper is organized as follows. In Section 2 we develop a nonparametric dispersion test designed specifically for 2^{k-p} factorial designs. The distribution of the test statistic is given and large sample approximations using the normal and beta distributions are discussed. A simple example of the test is then given. In Section 3, simulations are used to compare this test to the aforementioned F test. In addition, the new test is compared to the nonparametric tests of Mood [12], Ansari–Bradley [13], Siegel–Tukey [14], Klotz [15], Conover [16] and Pan [17]. Here it is shown that the proposed test performs well compared to all of these existing tests. Section 4 provides application of the new test to two examples from the literature. Finally, a summary is provided in Section 5.

2 A NONPARAMETRIC DISPERSION TEST

By selecting a column x_d to test for a dispersion effect, we can form $n/2$ pairs of columns $(x_j, x_{j'})$ such that $x_{ij}x_{ij'} = x_{id}\ \forall i$. The OLS estimates associated with each pair of these columns, $(\hat{\beta}_j, \hat{\beta}_{j'})$ will be referred to as "alias pairs" based on $I = x_d$. (See Appendix A for a brief overview of this concept or Box *et al.* [1, pages 374–385] for a thorough discussion.) McGrath and Lin [10] showed that in order to have the residuals at the +1 level of a column uncorrelated with the residuals at the −1 level of the same column, it is necessary to adapt the assumed "true" location model. The adapted model must include exactly the following terms: all active location effects, the location effect of the dispersion column to be tested, and all of their aliases. The effects that are not in this model can be grouped into g alias pairs $(\hat{\beta}_j^{(f)}, \hat{\beta}_{j'}^{(f)}), f = 1, \ldots, g$. Here, $\hat{\beta}_j^{(f)}$ and $\hat{\beta}_{j'}^{(f)}$ are the OLS estimates associated with columns

x_j and $x_{j'}$. McGrath and Lin [10] showed that Bergman and Hynén's D_{BH} statistic is a function of these alias pairs,

$$D_{BH} = F = \frac{\sum_{f=1}^{g}(\hat{\beta}_j^{(f)} + \hat{\beta}_{j'}^{(f)})^2}{\sum_{f=1}^{g}(\hat{\beta}_j^{(f)} - \hat{\beta}_{j'}^{(f)})^2}. \tag{1}$$

This statistic has an F distribution with g degrees of freedom in both the numerator and denominator.

It is well known that this test is quite sensitive to the normality assumption. To make this a non-parametric test, we first let $\hat{R}_j^{(f)}$ and $\hat{R}_{j'}^{(f)}$ be the ranks of $\hat{\beta}_j^{(f)}$ and $\hat{\beta}_{j'}^{(f)}$ respectively with the ranks being assigned among all $2g$ $\hat{\beta}_j$s that are not in the model. We could then substitute $\hat{R}_j^{(f)}$ for $\hat{\beta}_j^{(f)}$ in (1). However, when using ranks, the numerator and denominator are perfectly negatively correlated because their sum is a fixed constant (twice the sum of squares of the positive integers less than or equal to $2g$). Hence, we will use the denominator as the test statistic, the sum of squared differences in ranks (*SSDR*). Then small values of *SSDR* imply $\sigma_{d-}^2 < \sigma_{d+}^2$ and large values imply $\sigma_{d-}^2 > \sigma_{d+}^2$. Thus, by studying the differences in ranks within alias pairs we can perform a two-sided dispersion test. If we let $D_f = |\hat{R}_j^{(f)} - \hat{R}_{j'}^{(f)}|$, then

$$SSDR = \sum_{j=1}^{g}(\hat{R}_j^{(f)} - \hat{R}_{j'}^{(f)})^2 = \sum_{f=1}^{g} D_f^2. \tag{2}$$

The distribution of *SSDR* is rather complicated. However, it is straightforward to simulate it by sampling permutations of the integers $1, \ldots, 2g$. Table II shows critical values of *SSDR*

TABLE II Critical Values of *SSDR*.

	Nominal significance levels							
g	*0.005*	*0.01*	*0.025*	*0.05*	*0.95*	*0.975*	*0.990*	*0.995*
4	NA	4	10	12	80	82	84	NA
5	5	13	19	29	151	155	159	163
	11	17	21	31	153	157	161	165
6	24	30	42	54	250	260	270	274
	26	32	44	56	252	262	272	276
7	45	57	77	97	387	405	421	429
	47	59	79	99	389	407	423	431
8	80	98	127	158	568	603	638	660
	81	99	128	159	569	604	639	661
9	129	155	197	239	779	819	857	879
	131	157	199	241	781	821	859	881
10	190	228	288	344	1048	1102	1156	1188
	192	230	290	346	1050	1104	1158	1190
11	279	329	405	481	1369	1439	1513	1557
	281	331	407	483	1371	1441	1515	1559
12	388	448	550	646	1746	1834	1928	1982
	390	450	552	648	1748	1836	1930	1984
13	529	607	731	845	2189	2301	2421	2497
	531	609	733	847	2191	2303	2423	2499
14	692	790	940	1078	2700	2836	2978	3070
15	879	999	1187	1359	3285	3451	3629	3741
16	1128	1268	1486	1684	3944	4136	4346	4482
17	1397	1561	1819	2047	4685	4907	5157	5319
18	1718	1906	2208	2468	5504	5768	6052	6230
19	2087	2297	2645	2955	6425	6733	7059	7281
20	2504	2750	3142	3498	7432	7782	8168	8422

for $g = 4, \ldots, 20$ based on 200,000 simulations for each value of g. Note that while *SSDR* can be calculated for $g = 2$ or 3, its discreteness results in too few possible values to be used as a practical test statistic. For $g \leq 13$, two *SSDR* values are given. The top number is associated with a significance level less than nominal and the bottom is associated with a value greater than nominal. For $g > 13$ only a single value is needed as all are very close to the specified nominal.

The expected value of *SSDR* can be found as follows. The range of D_f is $[1, 2, \ldots, 2g-1]$. For $D_f = 1$, there are $2g-1$ possible values of $\min(\hat{R}_j^{(f)}, \hat{R}_{j'}^{(f)})$, *i.e.* $1, 2, \ldots, 2g-1$. For $D_f = 2$, there are $2g-2$ possible values of $\min(\hat{R}_j^{(f)}, \hat{R}_{j'}^{(f)})$, *i.e.* $1, 2, \ldots, 2g-2$. In general, $D_f = i$ has $2g-i$ possible values of $\min(\hat{R}_j^{(f)}, \hat{R}_{j'}^{(f)})$, *i.e.* $1, 2, \ldots, 2g-i$. The pair $(\min(\hat{R}_j^{(f)}, \hat{R}_{j'}^{(f)}), D_f)$ fully defines the alias pairs' rankings. There are $\binom{2g}{2} = g(2g-1)$ possible pairs of ranks. Thus, $P(D_f = i) = P(D_f^2 = i^2) = (2g-i)/g(2g-1)$, $i = 1, 2, \ldots, 2g-1$. From this it follows that $E(D_f^2) = \sum_{i=1}^{2g-1} i^2(2g-i)/g(2g-1) = g(2g+1)/3$. Then $E(SSDR) = E(\sum_{f=1}^{g} D_f^2)$ and

$$E(SSDR) = \frac{g^2(2g+1)}{3} \tag{3}$$

In a similar manner, it can be shown that $\mathrm{Var}(D_f^2) = g(g-1)(2g+1)(14g+9)/45$. However, the D_f are not independent. For $g = 2$ and $g = 3$, all possible rankings were determined leading to the exact joint distribution of the D_f. In addition, simulations of *SSDR* for $g = 4, \ldots, 20$ were used. Based on these exact and simulated distributions the following covariance and correlation were empirically derived: $\mathrm{Cov}(D_f^2, D_{f'}^2) = -g(2g+1)(4g+3)/45$ and $\mathrm{Corr}(D_f^2, D_{f'}^2) = -(4g+3)/[(g-1)(14g+9)]$. Finally, we have

$$\mathrm{Var}(SSDR) = \frac{2g^2(g-1)(2g+1)(5g+3)}{45} \tag{4}$$

As *SSDR* is a sum, there is a central limit effect and a normal approximation is reasonable for large g. Given g and the above formulas, we can obtain critical values using $E(SSDR) \pm Z_{\alpha/2}(\mathrm{Var}(SSDR))^{1/2}$ where $\alpha =$ the specified P (Type I error) and $Z_{\alpha/2}$ is the appropriate standard normal quantile. (Although *SSDR* is discrete, not much is gained by a continuity correction as the values of *SSDR* are quite large for large g.) However, the distribution of *SSDR* is light-tailed. For moderately large g, say $20 \leq g \leq 30$, a symmetric beta distribution provides a better approximation than the normal. This is intuitive as the beta distribution has a finite range, as does *SSDR*. A symmetric beta distribution has a single shape parameter, say b. Letting

$$B = \frac{3}{2g^2(2g+1)} SSDR,$$

it is shown in Appendix B that

$$B \mathrel{\dot\sim} \mathrm{Beta}(b, b) \quad \text{where } b = \frac{1}{2}\left(\frac{5g^2(2g+1)}{2(5g+3)(g-1)} - 1\right) \approx \frac{g-1}{2} \text{ for large } g.$$

In this paper we will study the most common design sizes, *i.e.* $n = 16$ and $n = 32$, where these approximations are not necessary due to Table II. They are provided for use with larger designs.[1] Note that the *SSDR* statistic, along with other nonparametric statistics discussed later, assume the response is continuous. So while, theoretically, there are no ties among

[1] As Table II includes values of g up to 20, even some 64 run designs can be studied from the table. For example, it is possible that when $n = 64$, $g = 20$ with as few as 11 ($\approx$17%) active location effects

the responses (or residuals or estimates), ties do occur relatively often in practice. Ties will be addressed using an example in Section 4.

As an example of calculating *SSDR*, we study data originally analyzed by Davies [18] and subsequently by Bergman and Hynén [9]. The effect of five factors on the quality of a dyestuff was studied in an unreplicated 2^{5-1} design. The five factors were temperature (A), starting material (B), reduction pressure (C), oven drying pressure (D), and vacuum leak (E). Table I shows the design matrix and responses where the y_D column contains the responses and the D row labels the factors and the two-factor interactions. The regression coefficients (OLS estimates) are shown below. As there are no degrees of freedom for error, we can not calculate standard errors to accompany these point estimates.

Intercept	*A*	*B*	*C*	*D*	*AB*	*AC*	*AD*
217.9688	0.21875	−3.78125	7.03125	33.34375	8.34375	1.53125	2.59375
BC	*BD*	*CD*	*DE*	*CE*	*BE*	*AE*	*E*
4.15625	−1.78125	7.15625	0.03125	2.34375	−3.84375	1.15625	−1.96875

Inspection of these coefficients shows that D has a large impact on location, *i.e.* the mean dyestuff quality. This agrees with the results using Lenth's [5] method and the findings of Davies [18] and Bergman and Hynén [9] that the only location effect is D. Bergman and Hynén also found a dispersion effect due to E. So while each of the columns 1–15 may be tested for a dispersion effect, we will calculate *SSDR* for column 15 ($SSDR_E$) for illustrative purposes. To ensure uncorrelated residuals between the two levels of E, we adapt the model with the overall mean and $\hat{\beta}_D$ to also include $\hat{\beta}_E$ and $\hat{\beta}_{DE}$ because $(\hat{\beta}_D, \hat{\beta}_{DE})$ is an alias pair for $I = E$.

$$\hat{y}_i = 217.9688 + 33.34375x_{iD} - 1.96875x_{iE} + 0.03125x_{iDE}.$$

To calculate $SSDR_E$, we use the coefficients that are not included in this model. These 12 coefficients are ranked from 1 to 12 in increasing order. These rankings are $BE = 1, B = 2, BD = 3, A = 4, AE = 5, AC = 6, CE = 7, AD = 8, BC = 9, C = 10, CD = 11, AB = 12$. These columns fall into $g = 6$ alias pairs: *A*:*AE*, *B*:*BE*, *C*:*CE*, *AB*:*CD*, *AC*:*BD*, *AD*:*BC*. Notice that the product of the two columns in each pair results in column E. From (2) we see that $SSDR_E$ is then the sum of the difference in the ranks within each pair, *i.e.*,

$$SSDR_E = (4-5)^2 + (2-1)^2 + (10-7)^2 + (12-11)^2 + (6-3)^2 + (8-9)^2 = 22.$$

(Calculation of *SSDR* for the other columns can be done in the same straightforward manner.) From Table II, with $g = 6$, we see that $SSDR_E$ leads to a two-sided p-value < 0.01. In fact, from the simulated reference distribution (available upon request from the first author), we have a p-value of 0.007. This obviously agrees quite well with Bergman and Hynén's finding for column E of $F = 11.51$ and p-value $= 0.009$ based on (1). However, *SSDR* and the F test will not always provide similar conclusions. This is shown in the next section where we compare *SSDR* to the F and several other tests, and also in Section 4 where we study the other columns from this example.

3 COMPARISON OF DIFFERENT METHODS

We compare *SSDR* to six other dispersion tests. All of these tests are based on comparing two samples: the residuals from rows $x_{id} = +1$ and $x_{id} = -1$ respectively. To describe the other tests, we first define some notation. Let $P = \{i: x_{id} = +1\}$ where x_d is the column being tested for dispersion. Let e_i be the residual from observation i from the adapted model that includes all location effects, the location effect of the column being tested for dispersion, and their aliases, and let r_i be the rank of this residual among all n residuals. All of the other nonparametric tests to be studied are general two sample tests in that they allow unequal sample sizes. They do, however, assume independent observations. When studying residuals from an $n = 2^{k-p}$ design, we have equal sample sizes ($n/2$) but will not have independent residuals. Thus, other nonparametric tests violate this assumption while *SSDR* is designed specifically for dispersion effect testing from 2^{k-p} designs. We might expect that the normality-based tests will outperform *SSDR* (and all other nonparametric tests) under the normality assumption.

Numerous simulations were performed to compare these dispersion tests. For each case studied, 10,000 data sets were simulated. Most of these were performed with $n = 16$. Four different general comparisons were made. The power and P (Type I error) of dispersion tests were studied under the normality assumption in three general scenarios: power with no location effects, P (Type I error) with one or more active location effects, and P (Type I error) with two active location effects undetected in the location model. Finally, each test was compared based on its ability to hold a specified P (Type I error) under different (non-normal) distributional assumptions. For a fair comparison among tests, the adapted model (see Section 2) was used.

Mood [12] apparently was one of the first to develop a nonparametric dispersion test. In our notation, the test statistic is $\sum_{i\in P}(r_i - n/2)^2$. Ansari and Bradley [13] used the statistic $\sum_{i\in P} r_i$. Siegel and Tukey [14] developed a dispersion test that is a variation of the Wilcoxon rank sum test, for location differences. In the Siegel–Tukey test, the smallest ranks are assigned alternately between the largest and smallest observations. For example, if $n = 8$ observations have been sorted from smallest to largest, the ranks assigned would be $1, 4, 5, 8, 7, 6, 2, 3$. Denoting the rank of residual i using this method as t_i, the test statistic is then $\sum_{i\in P} t_i$. Klotz [15] modified Mood's statistic to be $\sum_{i\in P}[\Phi^{-1}(r_i/(n+1))]^2$ where $\Phi[y]$ is the cumulative standard normal distribution evaluated at y. A squared ranks test (see, for example, Conover [16]) uses the statistic $\sum_{i\in P} r_i^2$. We will compare *SSDR* to all of the above.

Fligner and Killeen [19] developed analogs of some of the above tests. More recently, Hall and Padmanabhan [20] used a bootstrapping approach to the problem. Pan [17] modified the popular test of Levene [21] and compared it to the Fligner and Killeen and Hall and Padmanabhan tests. Therefore, of the more recent tests, we will compare *SSDR* to Pan's test.

3.1 No Location Effects, Normal Errors

We first study the Type I error rate and power of the test for dispersion effects in the case of no location effects and normally distributed errors. With $n = 16$ and no location effects, the adapted model consists solely of the overall mean and the location effect of the column to be tested for dispersion. This leads to 14 effects left out of the model, or $g = 7$ alias pairs. Table III shows the results for various magnitudes of dispersion. Here, $\Delta = \sigma^2_{d+}/\sigma^2_{d-}$ designates the dispersion effect magnitude with $\Delta = 1$ being the null case. The dispersion effect was created by generating 16 standard normal variates and then multiplying half of these (where $x_{id} = +1$) by $\Delta^{1/2}$. As expected, Table III shows the standard F test holds the nominal value of α and has the best power under these conditions. *SSDR* and the tests of Ansari–Bradley, Siegel–Tukey and Pan all hold the nominal α reasonably well. Of these tests, Pan's generally has the best power followed by *SSDR*. The squared ranks test and the tests

TABLE III Power Under Normality Assumption ($n = 16$, $g = 7$).

Δ	*F*	*SSDR*	*Mood*	*Klotz*	*Ansari–Bradley*	*Siegel–Tukey*	*Squared Ranks*	*Pan*
					$\alpha = 0.01$			
1^2	0.0101	0.0136	0.0141	0.0159	0.0080	0.0086	0.0102	0.0087
2^2	0.1623	0.1062	0.1144	0.1300	0.0718	0.0837	0.0962	0.1125
3^2	0.5130	0.3221	0.2944	0.3294	0.1961	0.2227	0.2754	0.3665
4^2	0.7772	0.5312	0.4533	0.4977	0.3227	0.3563	0.4489	0.5977
5^2	0.9055	0.6778	0.5688	0.6098	0.4236	0.4598	0.5727	0.7561
					$\alpha = 0.05$			
1^2	0.0502	0.0529	0.0748	0.0720	0.0533	0.0632	0.0634	0.0429
2^2	0.3978	0.3063	0.3386	0.3491	0.2575	0.2790	0.3461	0.3155
3^2	0.7776	0.6192	0.6233	0.6422	0.5050	0.5315	0.6404	0.6596
4^2	0.9294	0.7931	0.7781	0.7951	0.6642	0.6828	0.8012	0.8473
5^2	0.9754	0.8850	0.8627	0.8794	0.7626	0.7785	0.8819	0.9336
					$\alpha = 0.10$			
1^2	0.0988	0.1046	0.1321	0.1373	0.0863	0.0976	0.1217	0.0871
2^2	0.5341	0.4384	0.4638	0.4936	0.3400	0.3595	0.4760	0.4561
3^2	0.8644	0.7432	0.7403	0.7717	0.6047	0.6259	0.7582	0.7806
4^2	0.9601	0.8806	0.8648	0.8904	0.7481	0.7635	0.8843	0.9215
5^2	0.9882	0.9411	0.9257	0.9386	0.8352	0.8474	0.9392	0.9693

of Mood and Klotz have inflated true α values. (With nominal $\alpha = 0.10$, the standard error is approximately 0.003, for example.)

3.2 One or More Location Effects, Normal Errors

We next estimate the Type I error rate assuming that one or more location effects are present under the normality assumption. Here we use $n = 32$ to more fully study the impact of many location effects. Recall that for each location effect included in the "true" model we must add its alias based on $I = x_d$, thus reducing g. Figure 1 plots the estimates of the true α for nominal $\alpha = 0.05$. (Appendix C shows estimates of the true α for nominal $\alpha = 0.01$, 0.05, and 0.10.) Only *SSDR* and *F* hold the specified value of α. The other tests are severely impacted by the independence assumption as g decreases with respect to n. So in general, the only viable tests under the normality assumption are *SSDR* and *F*.

3.3 Two Undetected Location Effects, Normal Errors

Next we study the Type I error rate in the case where there are two active location effects that are not identified and hence mistakenly left out of the model. McGrath and Lin [22] showed that these unidentified location effects create an expected difference between s^2_{d+} and s^2_{d-}, the sample variance of the residuals where $x_{id} = +1$ and $x_{id} = -1$ respectively, leading to possible detection of a spurious dispersion effect. As in Section 3.1, we assume normality with $n = 16$ and $g = 7$. Several combinations of location effect magnitudes (from 0.5σ to 2σ) were simulated assuming no dispersion effects. Figure 2 show the results (see Appendix D for details). All residual-based nonparametric tests have poor performance. As the magnitude of the unidentified effects increases, the *P* (Type I error) greatly increases. The performance of *F* is also poor but not as much so. *SSDR*, while having a slightly inflated *P* (Type I error) does much better than all of the others. Thus, *SSDR* is the only test of those studied that provides reasonable protection against identifying spurious dispersion effects caused by a pair of unidentified location effects. This is a very valuable property as will be seen in Section 4.

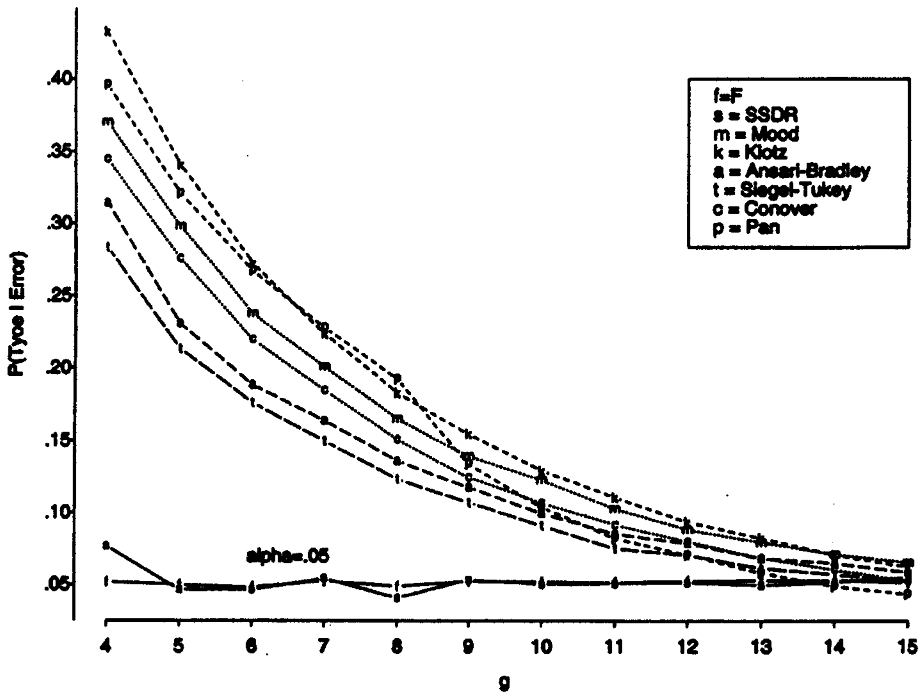

FIGURE 1 P (Type I Error), Normal Distribution ($n = 32$)

So under the normality assumption, the common F test (based on the adapted model) is preferred if all location effects are identified. However, it is not uncommon that one or more active location effects remain undetected in unreplicated fractional factorial designs. If two or more location effects are left unidentified, then the F test can be grossly misleading. So even though *SSDR* is a nonparametric test, it is preferred over a test correctly assuming normality in some situations. Additionally, the F test is quite sensitive to the normality assumption. So our next discussion is the robustness of all tests with respect to non-normal distributions.

3.4 Non-normal Distributions

Along with the normal distribution, four additional error distributions were examined; uniform(0, 1), beta(1, 2), $t(5)$, and exponential(1). These were initially studied with $n = 16$, $g = 7$, and $\alpha = 0.01, 0.05$ and 0.10. Of the symmetric distributions, the t distribution is heavy-tailed while the uniform may be considered light-tailed due to its finite range. The beta distribution studied is slightly skewed and light-tailed. The exponential distribution is asymmetric (positively skewed). Table IV shows the results.

For the uniform distribution, the *SSDR*, *F*, and Pan tests have substantially conservative P (Type I error)s. The Mood and Klotz tests have inflated (liberal) P (Type I error)s but the Ansari–Bradley, Siegel–Tukey, and squared ranks tests hold α reasonably well. For the beta distribution, again the *SSDR*, *F*, and Pan tests are conservative, but only slightly so. The Mood and Klotz tests have quite liberal P (Type I error)s. The Ansari–Bradley, Siegel–Tukey, and squared ranks tests seem to hold α at $\alpha = 0.01$ and 0.05 but are liberal for $\alpha = 0.10$. For the t distribution, all tests except Pan, Ansari–Bradley, and Siegel–Tukey

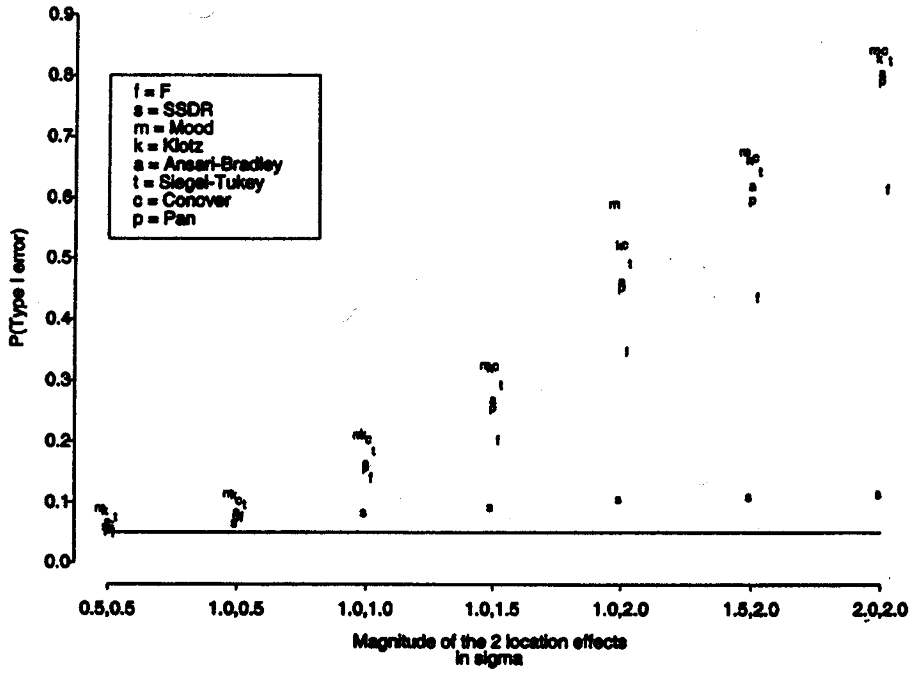

FIGURE 2 Dispersion Effect P (Type I Error) (nominal = .05) with Two Unidentified Location Effects

have unacceptably high P (Type I error)s. So for a symmetric distribution, the Pan, Ansari–Bradley and Siegel–Tukey tests appear to be the most robust to distributional assumptions when there are no location effects.

To test performance when location effects are present (a more realistic situation), we let $g = 5$ with $n = 16$. This combination can occur when there are between two and five active location effects. The results are shown in the bottom of Table IV. Here we see that *SSDR* outperforms all others. For the normal, uniform, beta, and t distributions, *SSDR* holds α better than the other tests with few exceptions.

For the exponential distribution, all tests performed poorly for both $g = 5$ and $g = 7$. Other skewed distributions were studied and the results (not shown) were also poor. But the simulations were performed assuming that location effects were identified and removed correctly. With a heavily skewed distribution, it is quite likely that extreme values will occur and actually make one or more location effects appear active. It is conjectured that if these spurious location effects were removed, the dispersion effect test results may improve. As is well-known, a proper transformation may be helpful to bring the error distribution closer to normality. Even if the errors are not normally distributed, tests will behave better if based on a reasonably symmetric distribution (*e.g.* the beta distribution tested here) as opposed to a heavily skewed one.

So the simulations have shown that no dispersion test studied is truly "distribution free" in 2^{k-p} designs as none is able to hold its P (Type I error) for all non-normal distributions. Additionally, unlike location effect estimates in fractional factorial designs, dispersion effect estimates are not independent. However, it seems that *SSDR* is the most robust dispersion test for these designs. It holds α under more conditions and protects against spurious dispersion effects caused by undetected location effects far better than any other test. We will demonstrate these properties with two examples from the literature.

TABLE IV P (Type I Error), $n = 16$.

α	*Distribution*	*F*	*SSDR*	*Mood*	*Klotz*	*Ansari–Bradley*	*Siegel–Tukey*	*Squared ranks*	*Pan*
				Case: $g=7$					
0.01	Normal	0.0107	0.0087	0.0154	0.0156	0.0082	0.0097	0.0089	0.0082
	Uniform	0.0018	0.0026	0.0160	0.0163	0.0096	0.0112	0.0102	0.0042
	Beta	0.0080	0.0089	0.0288	0.0287	0.0136	0.0159	0.0168	0.0100
	$t(5)$	0.0345	0.0252	0.0189	0.0176	0.0102	0.0122	0.0136	0.0102
	Exponential	0.0957	0.0581	0.0775	0.0646	0.0471	0.0561	0.0670	0.0289
0.05	Normal	0.0469	0.0498	0.0648	0.0722	0.0507	0.0618	0.0635	0.0405
	Uniform	0.0124	0.0163	0.0696	0.0743	0.0575	0.0690	0.0615	0.0276
	Beta	0.0392	0.0430	0.1075	0.1127	0.0849	0.0966	0.0900	0.0465
	$t(5)$	0.1121	0.0995	0.0710	0.0756	0.0565	0.0660	0.0727	0.0552
	Exponential	0.2214	0.1805	0.2397	0.2349	0.1989	0.2145	0.2081	0.0936
0.10	Normal	0.0981	0.0959	0.1330	0.1365	0.0841	0.0963	0.1184	0.0844
	Uniform	0.0326	0.0415	0.1332	0.1312	0.0907	0.1015	0.1162	0.0588
	Beta	0.0829	0.0872	0.1887	0.1883	0.1299	0.1406	0.1569	0.0871
	$t(5)$	0.1872	0.1757	0.1379	0.1477	0.0872	0.0978	0.1313	0.1057
	Exponential	0.3085	0.2796	0.3605	0.3543	0.2668	0.2810	0.3049	0.1540
				Case: $g=5$					
0.01	Normal	0.0094	0.0104	0.0350	0.0354	0.0212	0.0222	0.0248	0.0318
	Uniform	0.0059	0.0063	0.0382	0.0383	0.0264	0.0277	0.0281	0.0297
	Beta	0.0097	0.0100	0.0371	0.0374	0.0234	0.0254	0.0278	0.0321
	$t(5)$	0.0229	0.0161	0.0428	0.0445	0.0271	0.0296	0.0330	0.0469
	Exponential	0.0595	0.0300	0.0731	0.0729	0.0478	0.0552	0.0647	0.0725
0.05	Normal	0.0459	0.0473	0.1123	0.1189	0.0876	0.1016	0.1096	0.0968
	Uniform	0.0293	0.0310	0.1136	0.1186	0.0930	0.1037	0.1090	0.0872
	Beta5	0.0447	0.0481	0.1222	0.1279	0.0929	0.1060	0.1158	0.0946
	$t(5)$	0.0924	0.0745	0.1348	0.1457	0.1017	0.1140	0.1353	0.1330
	Exponential	0.1595	0.1096	0.2064	0.2186	0.1599	0.1771	0.1043	0.1820
0.10	Normal	0.0974	0.0967	0.1882	0.2078	0.1248	0.1349	0.1828	0.1674
	Uniform	0.0605	0.0650	0.1798	0.1875	0.1300	0.1392	0.1702	0.1429
	Beta	0.0882	0.0945	0.1979	0.2121	0.1344	0.1436	0.1841	0.1594
	$t(5)$	0.1633	0.1391	0.2189	0.2381	0.1464	0.1804	0.2156	0.2064
	Exponential	0.2478	0.1894	0.2980	0.3187	0.2162	0.2297	0.2906	0.2681

4 EXAMPLES

In the example from Section 2, we tested for a dispersion effect in column E only. We now discuss testing for dispersion effects in all of the other columns using the F test and *SSDR*. Table V shows the statistics and p-values for each. The F test results show one highly significant dispersion effect (E), and two mildly significant effects, D and DE. The *SSDR* results show E as being highly significant but two different mildly significant effects (C and AE). The interpretation here is not clear. The simulations of Section 3 showed that non-normal distributions, unidentified location effects, or other dispersion effects can create spurious dispersion effects. Thus, from a practical view, one might only consider dispersion effects to be active if the p-value is quite small. In this example we would conclude that only factor E has a dispersion effect.

We now provide an example where the *SSDR* test is the only one of those studied that does not incorrectly detect a spurious dispersion effect. The data in this example were originally analyzed by Montgomery [8] and subsequently by McGrath and Lin [22]. The data are from an injection molding experiment where the response to be optimized was shrinkage. The factors studied were mold temperature (A), screw speed (B), holding time (C), gate size (D),

TABLE V Dyestuff Dispersion Effect Statistics.

Column	*g*	s^2_{d-}	s^2_{d+}	*F*	*p-value*	*SSDR*	*p-value*
A	6	391	141	0.361	0.241	250	0.100
B	6	133	376	2.827	0.232	112	0.505
C	6	231	86	0.373	0.255	260	0.049
D	7	100	447	4.474	0.066	115	0.151
AB	6	228	148	0.651	0.615	198	0.513
AC	6	115	393	3.417	0.160	54	0.089
AD	6	377	157	0.417	0.311	234	0.202
BC	6	346	160	0.462	0.370	224	0.277
BD	6	124	384	3.100	0.194	74	0.193
CD	6	216	102	0.471	0.381	200	0.487
DE	6	86	455	5.292	0.062	74	0.193
CE	6	275	101	0.368	0.249	248	0.109
BE	6	148	361	2.441	0.302	82	0.247
AB	6	409	96	0.235	0.102	264	0.034
E	6	43	495	11.513	0.009	22	0.007

cycle time (E), moisture content (F), and holding pressure (G). The design is a 2^{7-3} fractional factorial, a $1/2^3$ fraction of a 2^7 design. The factors are labeled in the row marked M and the responses are in the column marked y_M in Table I. The location effect estimates are shown below. Due to the complex confounding pattern associated with this design, we have merely labeled the coefficients by their column numbers.

0	1	2	3	4	5	6	7
27.2125	6.9375	17.8125	−0.4375	0.6875	5.9375	−0.8125	−2.6875
8	**9**	**10**	**11**	**12**	**13**	**14**	**15**
−0.9375	−0.0625	−0.0625	0.1875	0.0625	−2.4375	0.1875	0.3125

Based on a normal probability plot, Montgomery [8] found active location effects in columns 1, 2, and 5 (which agrees with Lenth's procedure). Fitting this model,

$$\hat{y}_i = 27.3125 + 6.9375x_{i1} + 17.8125x_{i2} + 5.9375x_{i5},$$

Montgomery identified dispersion effects by calculating $\ln(s^2_{d+}/s^2_{d-})$ for each column and producing a normal probability plot. (Here, s^2_{d+} and s^2_{d-} are the sample variances of residuals for $x_{id} = +1$ and $x_{id} = -1$ respectively.) Based on this analysis, column 3 was found to have a dispersion effect with $\ln(s^2_{3+}/s^2_{3-}) = \ln(32.44/2.66) = \ln(12.20) = 2.50$. To make the residuals in the $+1$ rows of column 3 uncorrelated with the residuals of the -1 rows, we adapt the model as discussed in Section 2 resulting in

$$\hat{y}_i = 27.3125 + 6.9375x_{i1} + 17.8125x_{i2} - 0.4375x_{i3} + 5.9375x_{i5} \\ - 0.08125x_{i6} - 0.9375x_{i8} + 0.1875x_{i11}.$$

This leaves eight terms out of the model and $g = 4$. The alias pairs are 4:10, 7:13, 9:14, and 12:15. Looking at these coefficients, we note that two are tied, namely $\hat{\beta}_9 = \hat{\beta}_{10} = -0.0625$. We use two methods to address ties. First, we assign the arithmetic mean of the ranks that would be assigned without ties. Noting there are two coefficients with lower values than the tied coefficients, we assign rank 3.5 to each. This leads to $SSDR = 31.5$ with an approximate p-value of 0.576 (from interpolation of the reference distribution). The second approach is to calculate *SSDR* under all possible rankings. Here, there are only two possible rankings, $(\hat{R}_9 = 3, \hat{R}_{10} = 4)$ and $(\hat{R}_9 = 4, \hat{R}_{10} = 3)$, but the method can be extended for three or more tied values or multiple groups of ties. These rankings result in *SSDR* values of 34 and 30 with p-values of 0.648 and 0.533 respectively. The minimum and maximum p-values can be used as liberal and conservative significance figures. Analysis of other experiments has shown that, in most cases, the mean rank method is satisfactory as the p-values do not generally change greatly with the other ranking combinations. Obviously any ranking assignment in this example leads to a conclusion of no significant dispersion effect in column 3.

The other tests provided the following test statistics for column 3: $F = 35.75$ ($p = 0.004$), Pan $= 8.54$ ($p < 0.001$), Mood $= 37.5$, Klotz $= 0.855$, Ansari–Bradley $= 53$, Siegel–Tukey $= 18$, and Conover $= 1286$. The p-values of the nonparametric tests are all < 0.01. Thus all tests but *SSDR* find a significant dispersion effect in column 3. Recall that two unidentified location effects create a spurious dispersion effect in their interaction column. *SSDR* is relatively insensitive to this spurious effect while all other tests studied have a relatively high probability of incorrectly detecting this effect (see Fig. 2). Of the effects not identified by Montgomery, columns 7 and 13 are by far the largest. A normal probability plot of the effects (not shown) does provide some evidence that these two effects are indeed active. Note that these columns form an alias pair. Including these two location effects in the location model and testing again ($g = 3$) we find that no test shows significance (very high p-values across the board). This very large change in p-values for all tests but *SSDR* indicates that the other tests were greatly affected by the two location effects left out of the model while *SSDR* was not. Thus it seems that *SSDR* provided the most reasonable conclusion of no dispersion effect in column 3 along with location effects in columns 7 and 13.

5 SUMMARY

With today's short product life cycles, product designers and engineers are pressed to develop high quality products and processes in a short time frame. A consistent product/process will have little variability. When a factor has a dispersion effect on a response, the level of the factor can be chosen to reduce the response variation resulting in improved consistency. Unreplicated 2^{k-p} designs can play an important role in detecting dispersion effects with a minimum expenditure of resources.

We compared *SSDR* to the F test which is sensitive to the normality assumption. We showed that when two active location effects are mistakenly left out of the model, the other nonparametric tests and the F test have a high probability of detecting the spurious dispersion effect created. *SSDR* provides protection against this misidentification even under normality.

As for the other nonparametric tests studied, they assume independent observations. When residuals from a location model are studied, the residuals are correlated regardless of which model is fit or what distributional assumptions are made. Accordingly, general nonparametric dispersion tests do not hold the nominal $\alpha = P$ (Type I error). *SSDR*, on the other hand is based on the regression coefficients left out of the model. Under the null hypothesis of no dispersion, these quantities are uncorrelated regardless of distributional assumptions and independent under normality. Thus *SSDR* is a very robust dispersion effect test for unreplicated fractional factorial designs.

After deriving the distribution of *SSDR*, it was discovered that *SSDR* is similar to a statistic derived by Shirahata [23]. Shirahata's test was developed as a nonparametric test for an intraclass correlation coefficient while ours is derived from the common F test.

Finally, we note that there are difficulties in studying dispersion effects from unreplicated 2^{k-p} designs (see Pan [24]). Regardless of distributional assumptions (and unlike location effect estimates), dispersion effect estimates are not independent of one another. As shown by McGrath and Lin [25], two dispersion effects actually create a spurious dispersion effect in their interaction column. They developed a normality-based method that is applicable when multiple dispersion effects are present. Further research is required to develop a nonparametric test applicable when multiple dispersion effects are present.

Acknowledgements

We thank two anonymous referees for their comments that helped improve the paper. Dennis Lin is partially supported by National Science Foundation via Grant DMS-9704711 and National Science Council of ROC via Contract NSC 87-2119-M-001-007.

References

[1] Box, G. E. P., Hunter, W. G. and Hunter, J. S. (1978). *Statistics for Experimenters*. John Wiley, New York.
[2] Daniel, C. (1959). Use of half-normal plots in interpreting factorial two-level experiments. *Technometrics*, **1**, 311–341.
[3] Daniel, C. (1976). *Applications of Statistics to Industrial Experimentation*. John Wiley, New York.
[4] Box, G. E. P. and Meyer, R. D. (1986). An analysis for unreplicated fractional factorials. *Technometrics*, **28**, 11–18.
[5] Lenth, R. (1989). Quick and easy analysis of unreplicated factorials. *Technometrics*, **31**, 469–473.
[6] Hamada, M. and Balakrishnan, N. (1998). Analyzing unreplicated factorial experiments: a review with some new proposals. *Statistica Sinica*, **8**, 1–38 (with discussion).
[7] Box, G. E. P. and Meyer, R. D. (1986). Dispersion effects from fractional designs. *Technometrics*, **28**, 19–27.
[8] Montgomery, D. C. (1990). Using fractional factorial designs for robust process development. *Quality Engineering*, **3**, 193–205.
[9] Bergman, B. and Hynén, A. (1997). Dispersion effects from unreplicated designs in the 2^{k-p} series. *Technometrics*, **39**, 191–198.
[10] McGrath, R. N. and Lin, D. K. J. (1999). Analysis of location and dispersion effects in unreplicated fractional factorials, *Technical Report 99-01*, Department of Statistics, Pennsylvania State University, 583–621.
[11] Groggel, D. J. and Skillings, J. H. (1986). Distribution-free tests for main effects in multifactor designs. *American Statistician*, **40**, 99–102.
[12] Mood, A. M. (1954). On the asymptotic efficiency of certain nonparametric two-sample tests. *Annals of Mathematical Statistics*, **25**, 514–522.
[13] Ansari, A. R. and Bradley, R. A. (1960). Rank sum tests for dispersions. *Annals of Mathematical Statistics*, **31**, 1174–1189.
[14] Siegel, S. and Tukey, J. W. (1960). A nonparametric sum of ranks procedure for relative spread in unpaired samples. *Journal of the American Statistical Association*, **55**, 429–445.
[15] Klotz, J. (1962). Nonparametric tests for scale. *Annals of Mathematical Statistics*, **33**, 498–512.
[16] Conover, W. J. (1908). *Practical Nonparametric Statistics*, 2nd ed. John Wiley, New York, pp. 239–248.
[17] Pan, G. (1999). On a Levene type test for the equality of two variances. *Journal of Statistical Computing and Simulation*, **63**, 59–71.
[18] Davies, O. L. (Ed.) (1956). *Design and Analysis of Industrial Experiments*, 2nd ed. Oliver and Boyd, London.
[19] Fligner, M. A. and Killeen, T. J. (1976). Distribution-free two-sample tests for scale. *Journal of the American Statistical Association*, **71**, 210–213.
[20] Hall, P. and Padmanabhan, A. R. (1997). Adaptive inference for the two-sample scales problem. *Technometrics*, **39**, 412–422.
[21] Levene, H. (1960). Robust tests for equality of variances. In: Olkin, I. (Ed.), *Contributions to Probability and Statistics*. Stanford University Press, Palo Alto, California, pp. 278–292.
[22] McGrath, R. N. and Lin, D. K. J. (2001). Confounding of Location and Dispersion Effects in Unreplicated Fractional Factorials. *Journal of Quality Technology*, **33**, 129–139.
[23] Shirahata, S. (1981). Intraclass rank tests for independence. *Biometrika*, **68**, 451–456.

[24] Pan, G. (1999). The impact of unidentified location effects on dispersion-effects identification from un-replicated factorial designs. *Technometrics*, **41**, 313–326.
[25] McGrath, R. N. and Lin, D. K. J. (2001). Testing multiple dispersion effects in unreplicated fractional factorials. *Technometrics*, **43**, 406–414.

APPENDIX A – BRIEF OVERVIEW OF ALIASING IN 2^{k-p} DESIGNS

An example 2^3 design matrix including all interactions is shown below. This is a full factorial.

I	A	B	C	AB	AC	BC	ABC
1	−1	−1	−1	1	1	1	−1
1	1	−1	−1	−1	−1	1	1
1	−1	1	−1	−1	1	−1	1
1	1	1	−1	1	−1	−1	−1
1	−1	−1	1	1	−1	−1	1
1	1	−1	1	−1	1	−1	−1
1	−1	1	1	−1	−1	1	−1
1	1	1	1	1	1	1	1

Here the I column represents the intercept or overall mean. When a large dispersion effect exists, the effect is to greatly reduce the impact of half of the observations, resulting in a near half fraction of the original design. Suppose, instead of the full factorial, a $1/2$ fraction, a 2^{3-1} is used by only running the rows where ABC = 1 above. This results in the following design matrix.

I	A	B	C	AB	AC	BC	ABC
1	1	−1	−1	−1	−1	1	1
1	−1	1	−1	−1	1	−1	1
1	−1	−1	1	1	−1	−1	1
1	1	1	1	1	1	1	1

Notice the following equalities: $I = ABC$, $A = BC$, $B = AC$, $C = AB$. In this paper, these are the alias pairs we would use to study the dispersion effect in column ABC of the original design. Although this example used $n = 8$ in a full factorial, the idea is applicable to any 2^{k-p} design. To study a dispersion effect in, say column D of a 2^{k-p} design, we find the alias pairs formed in a $1/2$ fraction of the original 2^{k-p}, a 2^{k-p-1} where $I = D$. For a more thorough discussion of 2^{k-p} designs, see Box *et al.* [1, pages 374–385].

APPENDIX B

SSDR has a finite range and is roughly symmetric for large g so a scaled symmetric beta distribution may form a reasonable approximation. If $B \sim \text{Beta}(b, b)$ then $E(B) = 1/2$ and $\text{Var}(B) = 1/4(2b+1)$. To find an appropriate scaling factor, we define $W = aB$ and match the first two central moments of W and *SSDR* to solve for a and b. $E(W) = a/2 = E(SSDR) = g^2(2g+1)/3$ so $a = 2g^2(2g+1)/3$. Additionally, $\text{Var}(W) = a^2/4(2b+1) = g^4(2g+1)^2/9(2b+1) = \text{Var}(SSDR) = 2g^2(g-1)(2g+1)(5g+3)/45$. Solving for b we have $b = 1/2((5g^2(2g+1)/(2(5g+3)(g-1)) - 1)$. Thus, $W = 2g^2(2g+1)B/3$ and *SSDR* have equal expected values and variances implying that $B = 3SSDR/2g^2(2g+1) \mathrel{\dot\sim} \text{Beta}(b, b)$. Simulations using $20 \leq g \leq 30$ have shown the approximation works rather well.

APPENDIX C

P (Type I Error) Under Normality Assumption, $n = 32$.

g	*F*	*SSDR*	*Mood*	*Klotz*	*Ansari–Bradley*	*Siegel–Tukey*	*Squared ranks*	*Pan*
				$\alpha = 0.01$				
15	0.0093	0.0114	0.0153	0.0153	0.0128	0.0122	0.0107	0.0088
14	0.0117	0.0117	0.0174	0.0183	0.0150	0.0143	0.0141	0.0126
13	0.0103	0.0116	0.0209	0.0225	0.0174	0.0169	0.0164	0.0146
12	0.0104	0.0120	0.0257	0.0281	0.0193	0.0187	0.0196	0.0180
11	0.0091	0.0110	0.0305	0.0336	0.0222	0.0222	0.0229	0.0220
10	0.0103	0.0107	0.0390	0.0432	0.0297	0.0287	0.0324	0.0313
9	0.0114	0.0115	0.0507	0.0570	0.0384	0.0370	0.0414	0.0494
8	0.0119	0.0058	0.0645	0.0742	0.0489	0.0476	0.0556	0.0899
7	0.0113	0.0111	0.0831	0.0961	0.0647	0.0628	0.0728	0.1133
6	0.0102	0.0103	0.1094	0.1302	0.0792	0.0763	0.0954	0.1447
5	0.0095	0.0099	0.1421	0.1704	0.1094	0.1057	0.1273	0.1868
4	0.0124	0.0190	0.2139	0.2530	0.1795	0.1795	0.1969	0.2642
				$\alpha = 0.05$				
15	0.0518	0.0555	0.0658	0.0626	0.0588	0.0529	0.0531	0.0436
14	0.0532	0.0519	0.0712	0.0709	0.0652	0.0573	0.0606	0.0492
13	0.0533	0.0498	0.0796	0.0826	0.0684	0.0615	0.0687	0.0583
12	0.0527	0.0516	0.0884	0.0937	0.0788	0.0708	0.0802	0.0709
11	0.0518	0.0509	0.1029	0.1109	0.0854	0.0753	0.0917	0.0823
10	0.0524	0.0505	0.1227	0.1293	0.0996	0.0911	0.1070	0.1039
9	0.0526	0.0535	0.1393	0.1546	0.1178	0.1074	0.1247	0.1338
8	0.0490	0.0410	0.1651	0.1828	0.1360	0.1233	0.1508	0.1922
7	0.0526	0.0545	0.2007	0.2235	0.1638	0.1498	0.1848	0.2277
6	0.0484	0.0466	0.2377	0.2721	0.1884	0.1761	0.2197	0.2669
5	0.0500	0.0467	0.2980	0.3408	0.2308	0.2133	0.2757	0.3207
4	0.0521	0.0768	0.3705	0.4327	0.3147	0.2836	0.3448	0.3961
				$\alpha = 0.10$				
15	0.1023	0.1080	0.1249	0.1226	0.1129	0.1009	0.1111	0.0884
14	0.1003	0.1012	0.1341	0.1337	0.1052	0.1089	0.1203	0.0985
13	0.1004	0.0998	0.1448	0.1462	0.1091	0.1131	0.1289	0.1107
12	0.0999	0.1018	0.1537	0.1610	0.1189	0.1250	0.1410	0.1248
11	0.1013	0.0988	0.1784	0.1793	0.1455	0.1408	0.1591	0.1459
10	0.1045	0.1017	0.1961	0.2053	0.1473	0.1509	0.1803	0.1711
9	0.1059	0.1037	0.2202	0.2312	0.1655	0.1729	0.2072	0.2056
8	0.1026	0.0929	0.2464	0.2658	0.1890	0.1940	0.2318	0.2729
7	0.1013	0.1055	0.2992	0.3151	0.2243	0.2291	0.2709	0.3096
6	0.1008	0.1046	0.3257	0.3639	0.2471	0.2516	0.3134	0.3523
5	0.0993	0.0927	0.3874	0.4343	0.2938	0.2988	0.3689	0.4046
4	0.1006	0.1157	0.4640	0.5370	0.3633	0.3633	0.4418	0.4776

APPENDIX D

P (Type I error) of Spurious Dispersion Effects When Two Location Effects are Unidentified (Nominal = 0.05, $n = 16$).

Location effects (in σ)	*F*	*SSDR*	*Mood*	*Klotz*	*Ansari–Bradley*	*Siegel–Tukey*	*Squared ranks*	*Pan*
0.5, 0.5	0.0485	0.0574	0.0893	0.0859	0.0655	0.0754	0.0622	0.0509
0.5, 1.0	0.0760	0.0646	0.1136	0.1127	0.0825	0.0962	0.1019	0.0727
1.0, 1.0	0.1411	0.0818	0.2099	0.2118	0.1621	0.1841	0.2029	0.1555
1.0, 1.5	0.2023	0.0914	0.3241	0.3214	0.2673	0.2928	0.3224	0.2536
1.5, 1.5	0.3472	0.1046	0.5885	0.5218	0.4603	0.4910	0.5212	0.4495
1.5, 2.0	0.4363	0.1086	0.6738	0.6627	0.6179	0.6422	0.6643	0.5938
2.0, 2.0	0.6137	0.1131	0.8410	0.8287	0.8030	0.8246	0.8404	0.7891