Center for Quality and Productivity Improvement
UNIVERSITY OF WISCONSIN
610 Walnut Street
Madison, Wisconsin 53705

(608) 263–2520
(608) 263–1425 FAX

Report No. 132

A Comparison of Dispersion Effect Identification Methods for Unreplicated Two-Level Factorials

Howard T. Fuller and Søren Bisgaard

February 1996

C Q P I
UNIVERSITY OF WISCONSIN

Report No. 132

Report Series
in Quality and Productivity

A Comparison of Dispersion Effect Identification Methods for Unreplicated Two-Level Factorials

Howard T. Fuller

Center for Quality and Productivity Improvement

Univeristy of Wisconsin

Madison, Wisconsin

Søren Bisgaard

Center for Quality and Productivity Improvement

Univeristy of Wisconsin

Madison, Wisconsin

ABSTRACT

Identifying factors that affect the variability of a process has become an important step in improving quality. Over the years many methods based on unreplicated two-level factorial experiments have been proposed for identifying such factors. In this article we present a comparison of statistical power associated with several alternative methods. To this end we also present a method of operationalizing the half-normal probability plot which can be used to assess the relative "power" associated with it.

KEYWORDS: Location effects; Dispersion effects,; Power; Operationalization

CQPI Report No. 132, February 1996

A Comparison of Dispersion Effect Identification Methods for Unreplicated Two-Level Factorials

Howard T. Fuller and Søren Bisgaard

Identifying factors that affect the variability of a process has become an important step in improving quality. Over the years many methods based on unreplicated two-level factorial experiments have been proposed for identifying such factors. In this article we present a comparison of statistical power associated with several alternative methods. To this end we also present a method of operationalizing the half-normal probability plot which can be used to assess the relative "power" associated with it.

1. INTRODUCTION

Unreplicated two-level factorials and fractional factorials have until recently primarily been used to investigate the effect of experimental factors on the *mean level* of some quality characteristic. However, Box and Meyer (1986) recently showed that they also can be used to identify factors that affect variability known as *dispersion effects*. Since the publication of their paper, a number of alternative methods and related discussions have appeared, see e.g. Nair and Pregibon (1988), Wang (1989), Montgomery (1991), Miller and Wu (1993) and Bergman and Hynen (1995).

The objective of this article is to examine and compare several alternative methods including those referred to above. Specifically we will compare measures of statistical power either through simulation or when possible by analytic means. As a result we will show that none of the methods are uniformly most powerful. However, a new method proposed by Bergman and Hynen (1995) seems to perform the best overall.

To introduce the concept of dispersion effects and its practical significance, let us consider Montgomery's (1991) experiment on an injection molding process. The objective was to reduce excessive shrinkage of plastic parts. The original experiment was an unreplicated 2_{IV}^{7-3} design augmented by four center points. However we will only use the factorial part. The seven factors were *A: temperature, B: screw speed, C: holding time, D: gate size, E: cycle time, F: moisture content* and *G: holding pressure*. The design matrix and responses for the first 16 factorial runs are given in Table 1.

Run	*A*	*B*	*C*	*D*	*E*	*F*	*G*	Shrinkage (x10)
1	−	−	−	−	−	−	−	6
2	+	−	−	−	+	−	+	10
3	−	+	−	−	+	+	−	32
4	+	+	−	−	−	+	+	60
5	−	−	+	−	+	+	+	4
6	+	−	+	−	−	+	−	15
7	−	+	+	−	−	−	+	26
8	+	+	+	−	+	−	−	60
9	−	−	−	+	−	+	+	8
10	+	−	−	+	+	+	−	12
11	−	+	−	+	+	−	+	34
12	+	+	−	+	−	−	−	60
13	−	−	+	+	+	−	−	16
14	+	−	+	+	−	−	+	5
15	−	+	+	+	−	+	−	37
16	+	+	+	+	+	+	+	52

Table 1. The 2_{IV}^{7-3} design matrix and responses for Montgomery's injection molding experiment.

A normal probability plot of the location effect estimates is given in Figure 1a. From it we see that factors *A: temperature* and *B: screw speed* and their interaction appear to have significant effects on shrinkage. Thus we may fit the model $\hat{y} = \hat{\beta}_{Avg} + \hat{\beta}_A x_A + \hat{\beta}_B x_B + \hat{\beta}_{AB} x_A x_B$, compute residuals, and check for violations of the standard assumptions. Figure 1b shows a normal plot of the residuals, and Figure 1c a plot of the residuals versus the predicted values. Based on these plots, there does not appear to be any abnormalities. However, plotting the residuals against the levels of each of the 15 main effects and interactions revealed that, as we see in Figure 2, the residual variability seems to depend on the levels of factor *C*. Thus, it appears that this factor, the holding time, has a dispersion effect.

Technologically this might be an important finding. First, since factor *C* does not appear to have a location effect we may set that factor at whichever level reduces variation. In turn this might likely reduce manufacturing problems associated with excess variability. Second, by identifying holding time as a critical factor for variability, engineers familiar with the technical aspects of the process might be able to use this new understanding to reason about possible ways to further reduce variability. The former is usually referred to as *empirical feedback* and the latter *scientific feedback*, see Box and Draper (1969, p.153).

Unfortunately plotting residuals against all contrasts is laborious, and making relative comparisons between the plots can be difficult. It is for this reason methods for dispersion effects have gained popularity. However, to avoid confusion about the objective we hasten to point out that our focus is on *identification* of a dispersion effect, not *estimation* of already identified effects, a distinction expounded by Box and Jenkins (1976, p. 19). We will now provide a brief overview of some of the many methods proposed for identifying dispersion effects.

2. OVERVIEW OF COMPETING METHODS FOR IDENTIFYING DISPERSION EFFECTS

The recent interest in methods for identifying dispersion effects in the context of quality improvement and robust design dates back to Box and Meyer (1986). Inspired by a careful analysis of an experiment by Taguchi and Wu (1985) that contained a dispersion effect, they proposed the use of the ratio of the sum of the squared residuals at each of the factor's levels as a method for discovering such effects. Because of this statistic's similarity to an *F* ratio, they further suggested that the log of this ratio will approximately follow a standard normal distribution.

Our literature search shows, however, that Daniel (1976, pp.138, 286), most probably out of a concern for checking statistical assumptions and not product robustness, likely is the first to have discussed how to identify dispersion effects for unreplicated two-level factorial experiments. In an inconspicuous fashion he proposed to first remove the location effects and then re-analyze the experiment using the absolute value of the residuals as the response. This proposal is akin to one later made by Hoaglin, Mosteller, and Tukey (1983) in the general regression situation. They suggested using either the absolute or the squared residuals as the response. However, they did not propose a particular test for significance.

Following the publication of Box and Meyer (1986), Nair and Pregibon (1988) in the case of replication, then compared Bartlett and Kendall's (1946) method of taking the log of the within-trial sample variance to their modified version of Box and Meyer's (1986) method for replicated experiments. Nair and Pregibon's (1988) conclusion was that taking the log of the within-trial sample variance in general was a more powerful technique than their modification of Box and Meyer's (1986). However, this in some sense involves comparing incommensurables since Box and Meyer intended theirs for unreplicated factorials only. Furthermore, Nair and Pregibon (1988) showed that both approaches when used for estimation are special cases of Maximum Likelihood Estimation (MLE) under specific model assumptions.

More recently two new methods by Wang (1989) and by Bergman and Hynen (1995) have been proposed. Building on the results of Cook and Weisberg (1983), Wang proposed using a Fisher score statistic for each factor and interaction to identify dispersion effects. Moreover Wang (1989) indicated that the approximate significance can be determined using a $\chi^2(1)$ distribution. Bergman and Hynen (1995) alternatively proposed using a test statistic based on the data instead of the residuals setting this method apart from all the other proposals.

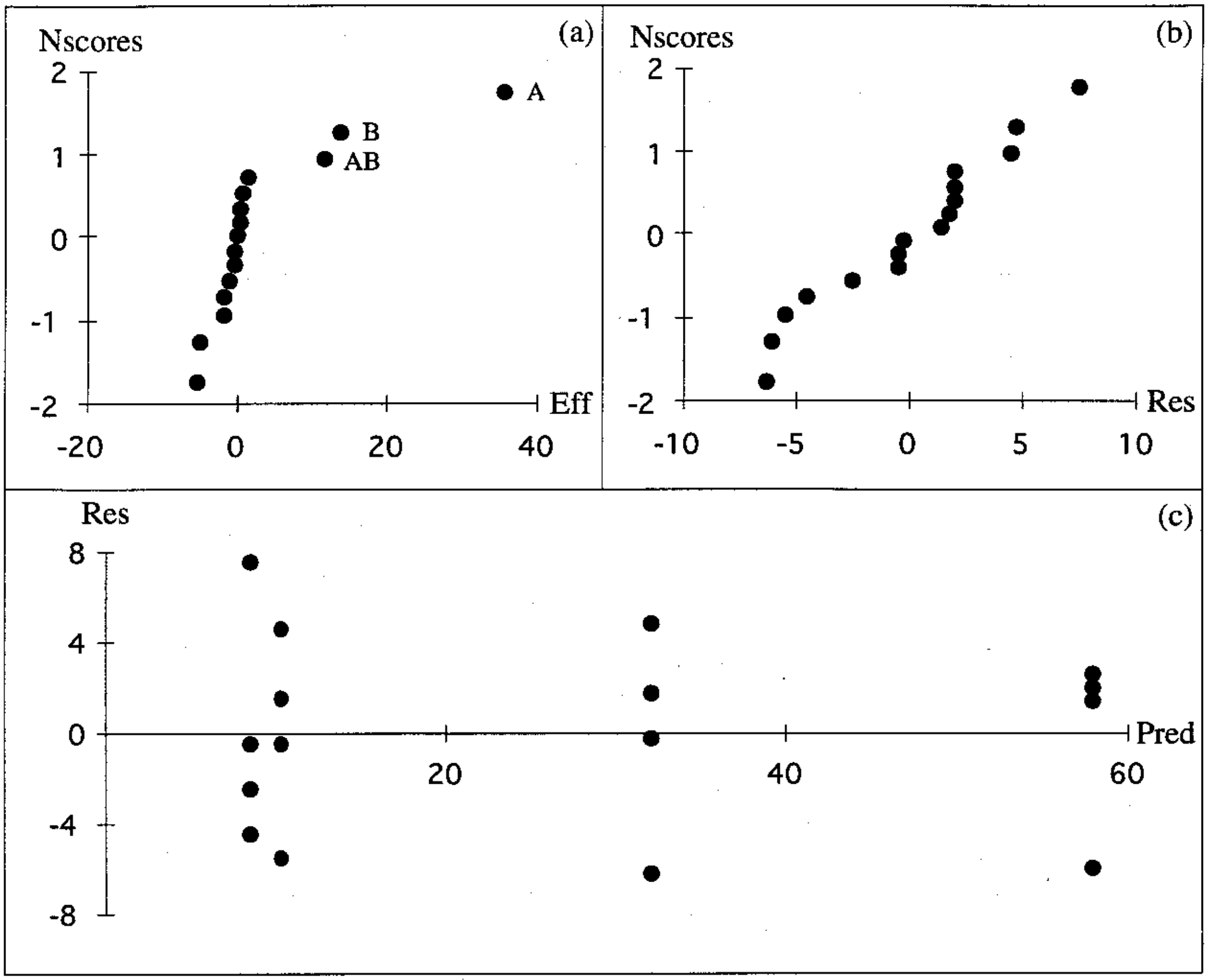


Figure 1. Normal probability plots of a) location effects, b) residuals and c) the residuals vs. predicted values.

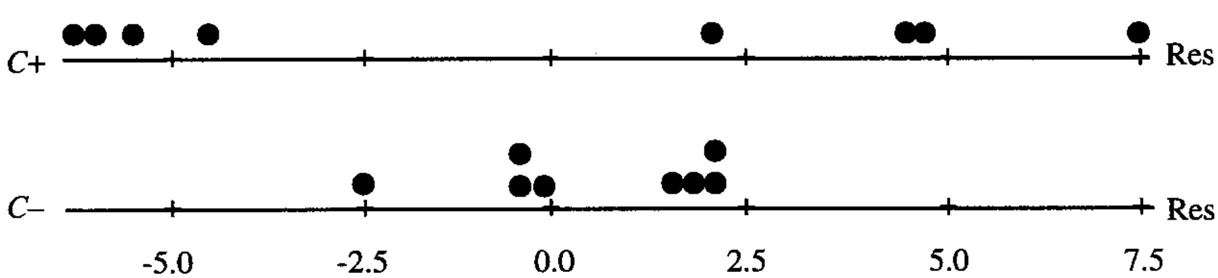


Figure 2. The dot plot of residuals versus the levels of factor *C*.

Conditional on the level of the factor or interaction one wishes to test for a dispersion effect, all the location effects assumed to be inert are first computed. A test statistic is then formed by taking the ratio of the sum the squared effects computed at each of the factor's levels. A major advantage of their method is that under the null hypothesis the test statistic is F distributed.

A key problem for most of the methods described so far, except for Bergman and Hynen's (1995), is the lack of an exact reference distribution under the null hypothesis. To avoid this problem Montgomery (1991) presented an interesting innovation by using a normal probability plot of the log of Box and Meyer's (1986) test statistic. His example used for demonstration was the injection molding experiment discussed earlier. In a similar situation, Miller and Wu (1993) used a half normal plot of test statistics estimated from the log absolute residuals.

The various methods for identifying dispersion effects have thus far not been systematically compared causing some confusion as to which method to prefer. In the remainder of this article we therefore will attempt such a comparison. First we present a definition of a dispersion effect and then define the different test statistics to be examined in this article. Using the different tests for significance we then compute measures of power associated with those tests. However, since several of the methods rely on using normal plots and not tests in the traditional sense this poses a problem. Thus we have developed a criterion, not perfect but useful, to evaluate the "power" for such tests. Finally in Section 6 we present our comparison of the proposed methods.

3. THE DISPERSION EFFECT MODEL

We will now introduce the dispersion effect model and methods for their identification. Suppose a standard unreplicated two-level factorial or fractional factorial experiment has been conducted and that $\mathbf{y}$ is the $n \times 1$ vector of observations. Further let $\mathbf{X}_n$ be the full rank model matrix. Most of the methods for identifying dispersion effects to be discussed below are based on first identifying and removing the location effects. Thus we first fit the full location effects model

$$\mathbf{y} = \mathbf{X}_n \beta_n + \varepsilon \qquad (1)$$

where it is assumed that $V\{\varepsilon\} = \mathbf{I}_n \sigma^2$. The least squares estimates of all the possible location effects are then $2\hat{\beta}_n = 2(\mathbf{X}_n'\mathbf{X}_n)^{-1}\mathbf{X}_n'\mathbf{y}$, and a normal probability plot of the elements of the vector $2\hat{\beta}_n$, except the first associated with the mean, can then be used to identify active location effects.

Having identified the active location effects, let $\mathbf{X}_{\mathcal{L}}$ be the matrix of column vectors corresponding to those effects. Then replacing $\mathbf{X}_n$ by $\mathbf{X}_{\mathcal{L}}$ in (1) gives a reduced model

$$\mathbf{y} = \mathbf{X}_{\mathcal{L}} \beta_{\mathcal{L}} + \varepsilon \qquad (2)$$

where again it is assumed that $V\{\varepsilon\} = \mathbf{I}_n \sigma^2$. The residuals are then given by

$$\mathbf{r} = \mathbf{y} - \mathbf{X}_{\mathcal{L}} \hat{\beta}_{\mathcal{L}}. \qquad (3)$$

where $\hat{\beta}_{\mathcal{L}} = (\mathbf{X}_{\mathcal{L}}\mathbf{X}_{\mathcal{L}}')^{-1}\mathbf{X}_{\mathcal{L}}'\mathbf{y}$.

Typically plots of the residuals versus the predicted values, and normal plots of the residuals are used to check for normality, dependence of the variance on the mean, no outliers, and so on. However, when checking for dispersion effects we need to check the assumption of $V\{\varepsilon\} = \mathbf{I}_n \sigma^2$ in (2) against the more general assumption of $V\{\varepsilon\} = \Sigma_\varepsilon = diag\{\sigma_{ii_\varepsilon}\}$ where σ_{ii_ε} are not necessarily all equal but depend on the levels of some experimental factors. Specifically a factor f is said to have a dispersion effect if when changed from its low to its high level, the variance is changed by a factor $k_f^2 \neq 1$ (Box and Meyer, 1986).

All but one of the test statistics we will examine in this article are essentially functions of absolute residuals raised to some power λ. We will therefore use the short hand notation

$$|r_i|^{(\lambda)} = \begin{cases} |r_i|^\lambda & \lambda > 0 \\ \log|r_i| & \lambda = 0 \end{cases}.$$

To identify dispersion effects Box and Meyer (1986), Bergman and Hynen (1995), and Wang (1989) have proposed test statistics which are some function of $|r_i|^{(\lambda)}$. These test statistics will be denoted by $\hat{D}_f^{BM}$, $\hat{D}_f^{BH}$ and $\hat{D}_f^{W}$ respectively. Those as well as others can conveniently be defined as

$$\hat{D}_f^\lambda = \begin{cases} n^{-1}\left(\sum_{i=f+}^{n/2}|r_i|^{(\lambda)} - \sum_{i=f-}^{n/2}|r_i|^{(\lambda)}\right) & \lambda \geq 0 \\ \log\left(\sum_{i=f+}^{n/2} r_i^2 \Big/ \sum_{i=f-}^{n/2} r_i^2\right) & \lambda = BM \\ \left(2n\hat{\sigma}^2\right)^{-1}\left(\sum_{i=f+}^{n/2} r_i^2 - \sum_{i=f-}^{n/2} r_i^2\right)^2 & \lambda = W \\ \sum_{i=1}^{v}\left(eff|f+\right)_i^2 \Big/ \sum_{i=1}^{v}\left(eff|f-\right)_i^2 & \lambda = BH \end{cases} \qquad (5)$$

where $i = f+$ and $i = f-$ are summation indicators over the high and low levels of factor f, $\left(eff|f+\right)_i$ and $\left(eff|f-\right)_i$ for $i = 1,2,\ldots,v$ are conditional effects computed using only the data at the high or low level of factor f, and $\hat{\sigma}^2 = n^{-1}\sum_{i=1}^{n} r_i^2$. Thus Daniel's (1976) proposal corresponds to $\lambda = 1$ and Miller and Wu's (1993) to $\lambda = 0$. Note that in our definition of $\hat{D}_f^\lambda$ we excluded negative powers of λ. This is because the residuals frequently are close to zero causing test statistics based on negative powers to be unstable.

Before we continue, let us briefly comment on the computational aspects of the above test statistics. Since most practitioners have access to standard least squares software packages the test statistics defined by $\hat{D}_f^\lambda$ for $\lambda = W$ and $\lambda \geq 0$ are extremely simple to compute and require no special software development. On the other hand the test statistics defined by $\hat{D}_f^\lambda$ for $\lambda = BH$ and BM although not complicated, are not straight forward either. For completeness it should also be noted that Box and Meyer (1986) provided a modification to their test statistic to compensate for the fitting of location factors before computing the residuals. However, because the argument for the log function frequently can become negative the test statistics $\hat{D}_f^{BM}$ may be undefined. Since our simulation studies furthermore indicate that the benefits from using their modified test statistics are largely offset by an increase in the variance due to the modification we will limit our study to the test statistic as defined in (5).

One unusual aspect of the problem of comparing the test statistics defined in (5) is the different operationalizations of the tests of significance. Bergman and Hynen's (1995) test statistic is the only one with an exact test following the F distribution with v degrees of freedom for both the numerator and denominator. Wang (1989) suggested for his test to use the $\chi^2(1)$ distribution to approximate the null distribution of $\hat{D}_f^W$, and Box and Meyer (1986) using asymptotic arguments suggested that their test statistics $\hat{D}_f^{BM}$ under the null hypothesis is approximately $N(0,1)$.

To check the adequacy of the approximations used by Wang (1989) and by Box and Meyer (1986) we have simulated the distributions of $\hat{D}_f^W$ and $\hat{D}_f^{BM}$ for various combinations of location and dispersion effects. From this we have found that Wang's (1995) use of the $\chi^2(1)$ distribution appears to perform relatively well. However, using the $N(0,1)$ distribution to test the significance of $\hat{D}_f^{BM}$ frequently gives misleading results. For example, factors that have dispersion effects can greatly increase the apparent significance of other factors that do not.

As already indicated Montgomery (1991) used a normal probability plot of $\hat{D}_f^{BM}$ rather than using the $N(0,1)$ reference distribution. Our investigations shows that this alternative provides for a more reliable test so we will base our comparisons of $\hat{D}_f^{BM}$ as well as $\hat{D}_f^\lambda$ for $\lambda \geq 0$ on the use of probability plots instead of the $N(0,1)$ reference distribution.

Comparing the power associated with the test statistics defined in (5) will require comparisons between an exact reference distribution test, an approximate reference distribution test, and several informal tests associated with the use of half or full normal probability plot. This of course creates some methodological problems that will necessitate a rather non-standard approach. Below we will present two types of comparisons. In the section that follows we will apply many of the proposed tests to the same data set given in Table 1, and in Section 5 we present a comparison of the power functions, some developed using ad hoc constructions, for the different tests.

4. COMPARATIVE ANALYSIS USING AN EXAMPLE

As a preliminary assessment of the performance of the different methods it is interesting to see how they all perform side by side for a particular data set containing a dispersion effect. In Figure 3 half-normal probability plots of $\hat{D}_f^\lambda$ for $\lambda = BM$, 0, 1/2, 1, and 2 are given. However, the tests of significance associated with $\hat{D}_f^{BH}$ and $\hat{D}_f^W$ are based on known reference distributions and thus are presented in tabular form. From Figure 3 we see that for detecting the large dispersion effect associated with factor C, all methods provide the same conclusion. Several of the methods, however, also seem to indicate a small or moderate dispersion effect associated with the AF interaction. To obtain more evidence about this possibility, we have in Figure 4 plotted on the same scale the residuals for C and AF. From these plots we see that the residuals are more concentrated about zero at the low level of the AF contrast and are more dispersed at the high level. However, compared to the dispersion effect associated with C, there appears to be less evidence for the AF dispersion effect.

Although all the methods in Figure 3 indicate a large dispersion effect associated with factor C it is natural to ask which method is best. To provide an answer to this we will in the next sections compare measures of power associated with the test statistics defined in (5).

5. MEASURES OF POWER

We will now develop measures of power for the various tests. The null hypothesis, H_o, for testing for a dispersion effect associated with factor f amounts to testing to see if changing factor f from its low to high level changes the variance of the response. Thus let the variance at the low and high levels be σ^2 and $k_f^2\sigma^2$ respectively. More formally the test is then $H_o: k_f = 1$ versus $H_a: k_f \neq 1$. Given a particular test statistic $\delta(\mathbf{y})$ and a rejection criterion, say $\delta(\mathbf{y}) > c$, the power function $\beta(k_f|\lambda)$ is then

$$\beta(k_f) = \Pr\{\delta(\mathbf{y}) > c \,|k_f, k_i = 1, i \neq f\}. \tag{6}$$

Note that with this definition we have limited out analysis to the power of detecting a single dispersion effect associated with factor f.

When conducting a formal test a specific well defined criterion is usually given for $\delta(\mathbf{y})$.. Moreover, critical values and the power are computed based on the probability distribution of $\delta(\mathbf{y})$. However, since some of the tests to be evaluated here are based on probability plots we need to proceed a little differently. Before we discuss how to deal with the probability plot based test let us consider the two tests by Bergman and Hynen (1995) and Wang (1989).

5.1 COMPUTING THE POWER FOR $\lambda = BH$ and W

First we will derive the distribution of $\hat{D}_f^{BH}$.. Let the sums of the variance of responses at the high and low levels of f be denoted by $\sum_{i=1}^{N/2} V\{y_i|f+\}$ and $\sum_{i=1}^{N/2} V\{y_i|f-\}$ respectively. Under the null hypothesis it can then readily be seen that

$$V\{(eff|f+)_i\} = V\{(eff|f-)_i\} = 4N^{-2}\sum_{i=1}^{N/2} V\{y_i|f+\}$$

and

$$E\{(eff|f+)_i\} = E\{(eff|f-)_i\} = 0 \text{ for}$$

$i = 1, 2, \ldots, v+1$. If we then add to the assumptions given in (3) that $\varepsilon \sim N(\mathbf{0}, \Sigma_\varepsilon)$ then it follows that

$$A = N^2 \sum_{i=1}^{v} (eff|f+)_i^2 \Big/ 4\sum_{i=1}^{v} V\{y_i|f+\} \tag{8}$$

and

$$B = N^2 \sum_{i=1}^{v} (eff|f-)_i^2 \Big/ 4\sum_{i=1}^{v} V\{y_i|f-\} \tag{9}$$

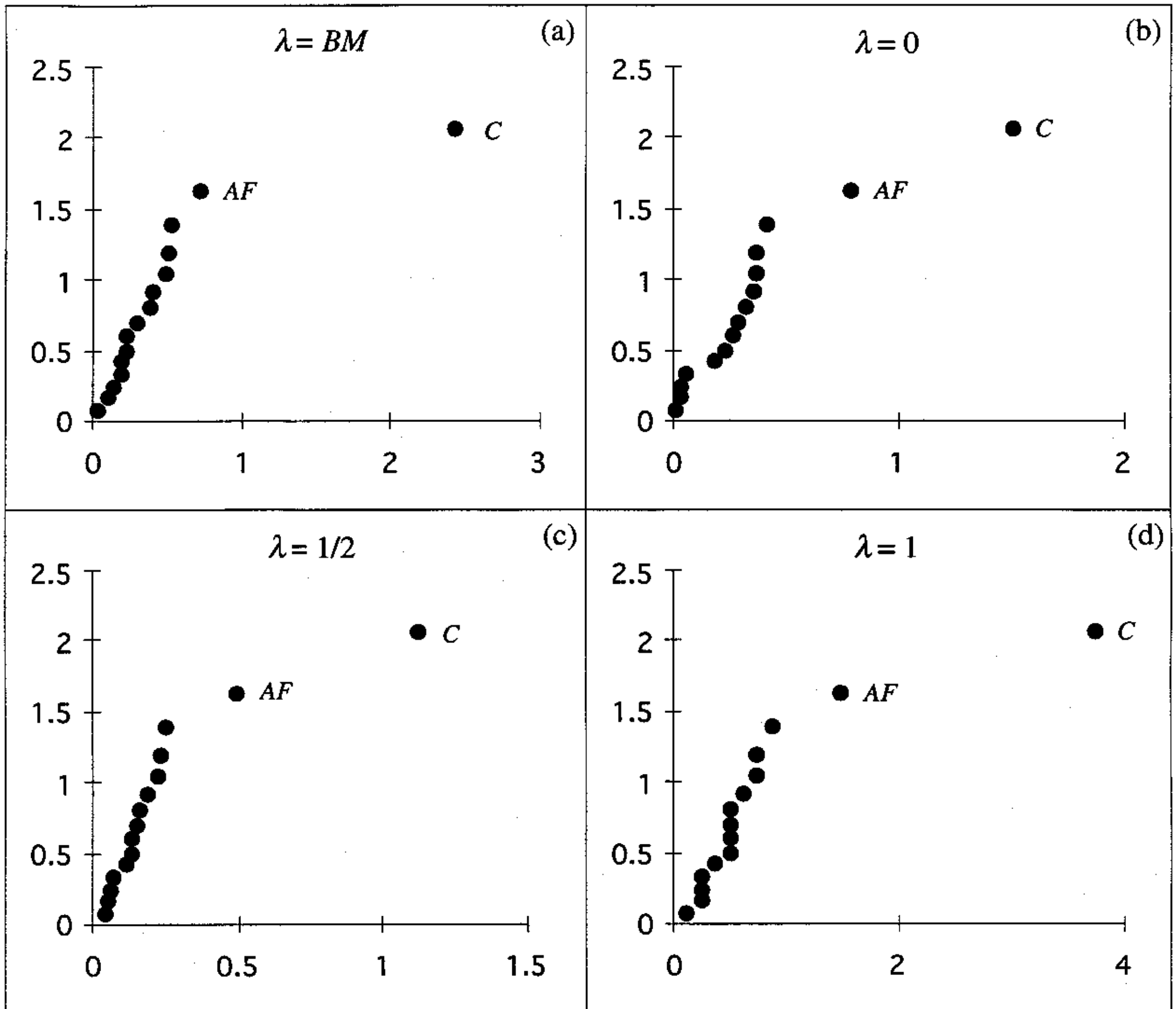
both are $\chi^2(v)$ distributed. Hence, $W = A/B$ is $F_{v,v}$ distributed. Now let

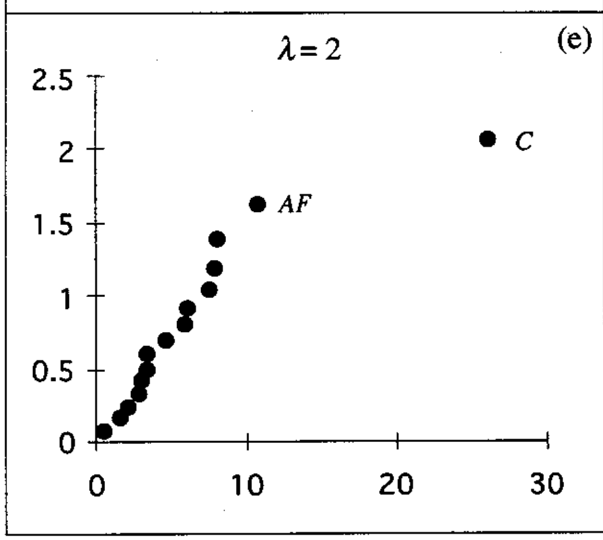
$$u = \sum_{i=1}^{n/2} V\{y_i|f+\} \Big/ \sum_{i=1}^{n/2} V\{y_i|f-\}.$$

Then it follows that

$$\Pr\{\hat{D}_f^{BH} \leq c\} = \Pr\{uW \leq c\} = \Pr\{W \leq u^{-1}c\} = F_{v,v}(u^{-1}c).$$

Thus Bergman and Hynen's (1995) test statistic follows a scaled $F_{v,v}$ distribution which implies that the power of their test can be computed analytically. Now let us turn to Wang's (1995) test. For $\delta(\mathbf{y}) = \hat{D}_f^W$ he suggests that the $\chi^2(1)$ distribution be





(f)

f	λ=BH	ν	P	λ=W	P
A	0.68	6	0.65	1.13	0.29
B	0.83	6	0.83	0.28	0.60
C	35.75*	4	0.00	22.48*	0.00
D	2.86	4	0.33	1.89	0.17
E	0.96	4	0.97	0.01	0.92
F	1.26	4	0.91	0.73	0.39
G	0.89	4	0.83	0.16	0.69
AB	1.11	6	0.90	0.09	0.76
AC	0.64	4	0.68	1.22	0.27
AD	1.56	4	0.68	0.39	0.53
AE	0.79	4	0.82	0.39	0.53
AF	3.59	4	0.24	3.78	0.05
AG	2.40	4	0.42	2.02	0.16
BD	0.20	4	0.15	0.29	0.59
ABC	3.05	4	0.31	2.09	0.15

Fig. 3. Half-normal plots of (a) $\hat{D}_f^{BM}$, (b) $\hat{D}_f^{o}$, (c) $\hat{D}_f^{1/2}$, (d) $\hat{D}_f^{1}$, (e) $\hat{D}_f^{2}$ and (f) the test statistics $\hat{D}_f^{BH}$ and $\hat{D}_f^{W}$ where ν = numerator and denominator degrees of freedom associated with each $\hat{D}_f^{BH}$ and p = p – value for each statistic.

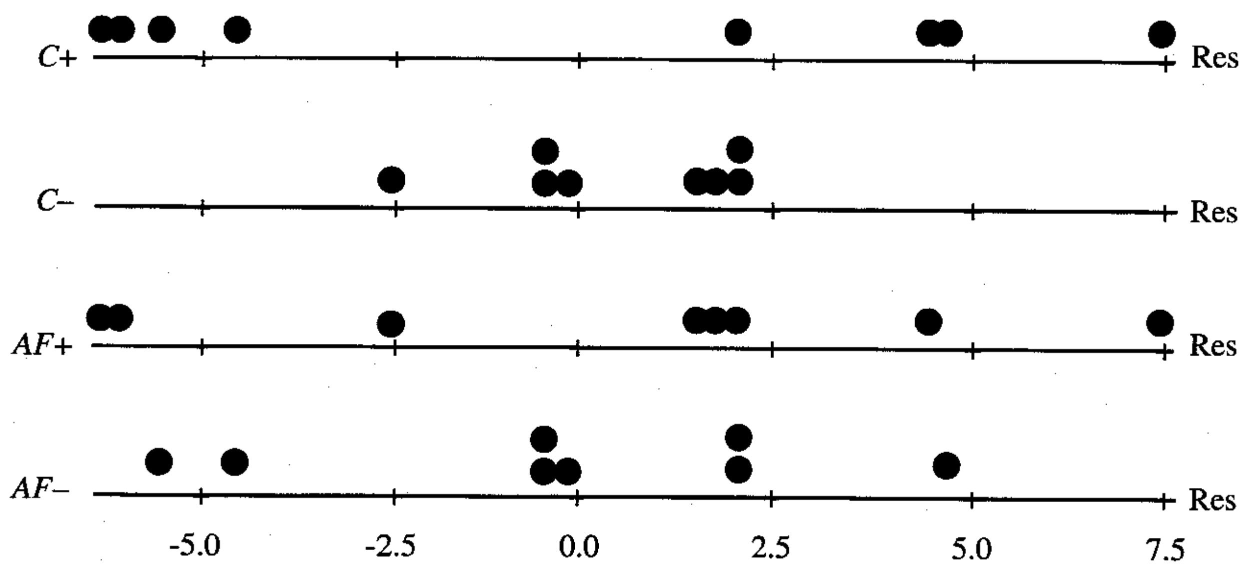


Figure 4. The dot plots of residuals versus the levels of the C and AF contrasts.

used to approximate the null distribution. Thus for our power computations we will as a rejection criterion use $\hat{D}_f^W > \chi^2(0.05, 1)$ and $\beta(k_f|\lambda = W)$ will be obtained for different values of k_f by simulating the distribution of $\hat{D}_f^W$ based 100,000 replications. For the remaining test statistics which use probability plots we first need to operationalize the rejection criterion.

5.2 POWER BASED ON OPERATIONALIZING THE HALF-NORMAL PLOT

In practice when using normal plots the "critical value" is chosen subjectively by visual inspection of the plot. Therefore different data analysts uses different criteria. We will therefore proceed on the following premise. If we develop an operational criterion for rejecting H_0 that more or less mimics the visual "rule" used by several data analysts, and we use that *same* criterion to compare *all* the tests then at least the comparison is fair. Moreover, if the criterion used is close to what several reasonable people agree on, then the ranking of the performance of the tests will be informative to practitioners in terms of which method to prefer. However, we hasten to say that we do not claim that the "power" obtained this way necessarily can be compared to power functions obtained in the usual sense. We only claim that such comparisons are indicative of relative performance.

To simplify the determination of the criterion somewhat, we will operationalize the half-normal probability plot because it has more structure than the full normal probability plot. However, this does not mean that we necessarily recommend that practitioners use half normal plots rather than full normal plots for discovering dispersion effects, and we certainly do not want to suggest that our criterion be used in practice instead of either of them.

A half-normal probability plot is constructed by plotting the ordered absolute test statistics with respect to the standard half-normal scores and then graphically drawing a line, ℓ, through the origin and the points on the plot assumed to be associated with inert effects. Figures 5a and 5b below show examples of half-normal probability plots. Note that for our operationalization we have switched the axes of the normal plot so that the normal scores are on the x-axis. The reason for this is that we are going to fit, with least squares, a line and want the random element to be on the y-axis. However, for the practical application we agree with Daniel (1983, pp. 565-568) that it is more natural to plot the test

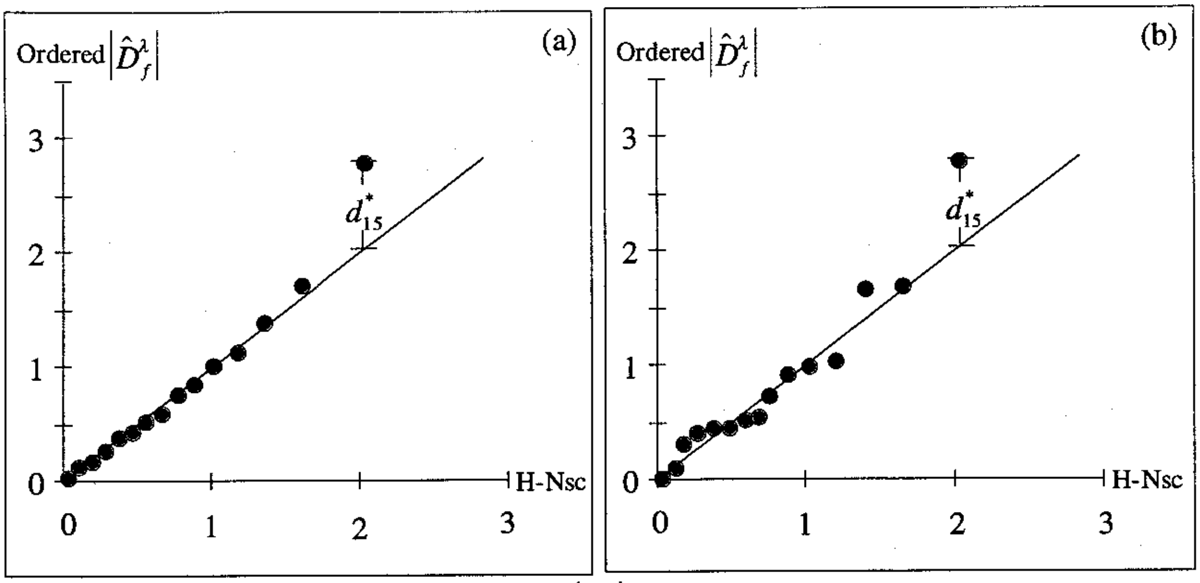


Figure 5. Examples of half-normal plots of ordered $|\hat{D}_f^\lambda|$ for $f = 1, 2, \ldots, 15$ where d_{15}^* represents the same vertical deviation of $|\hat{D}_{15}^\lambda|$ from ℓ in both plots. The line ℓ is a regression line with no intercept fitted to the first twelve ordered statistics.

statistics on the x-axis and the normal score on the y-axis.

To determine ℓ for an n-run two-level factorial experiment we will regress a line with no intercept on the $[0.75n]$ smallest $|\hat{D}_f^\lambda|$'s and their corresponding half-normal scores. This procedure is illustrated in Figures 5a and 5b. We have found that using 75% of the test statistics to determine the line is reasonable since we will assume both location and dispersion effect sparsity. From the regression we then determine the residual mean square $MS_{res}(\hat{D}_f^\lambda)$ and the vertical deviations, d_f^*, between $|\hat{D}_f^\lambda|$ and the fitted line ℓ. Thus for contrast f, d_f^*, represents the deviation of $|\hat{D}_f^\lambda|$ from its expected value. However, the magnitude of d_f^* should be judged relative to $MS_{res}(\hat{D}_f^\lambda)$. For our significance criterion we therefore will use d_f, defined as

$$d_f = d_f^* \Big/ \sqrt{MS_{res}(\hat{D}_f^\lambda)}. \qquad (10)$$

Note that one advantage of d_f as defined in (10) is that it is invariant to a location-scale transformation of the original data.

The test described in Section 5.1 based on an explicit criterion for each factor and interaction are tested separately. However, tests based on the appearance of the pattern of a probability plot tests all the factors simultaneously. Thus to be consistent with (7) and make the comparison fair, we need to develop a criterion for use with the probability plot that searches for a dispersion effect associated with f only. The significance of other test statistics $|\hat{D}_i^\lambda|$, $i \neq f$ not associated with the particular effect f under consideration will, therefore, be ignored to the extent that they are only considered when they impact the significance of $|\hat{D}_f^\lambda|$.

Another important yet subtle aspect of developing an operational rule for declaring a factor significant is the individual rank order of all the test statistics. First note that for a particular probability plot, $MS_{res}(\hat{D}_f^\lambda)$ is constant. Therefore from (10) it follows that $d_h < d_f$ if and only if $d_h^* < d_f^*$. However, it is possible that for some factor h for which $|\hat{D}_h^\lambda| > |\hat{D}_f^\lambda|$, the associated distances from the line are related as $d_h^* < d_f^*$. In such cases, we believe that $|\hat{D}_f^\lambda|$ should be declared significant only if for all h such that $|\hat{D}_h^\lambda| > |\hat{D}_f^\lambda|$ the factors associated with $|\hat{D}_h^\lambda|$ are also declared significant.

To formalize the above, let the set of all possible contrasts from a standard two-level factorial or fractional factorial experiment be denoted by $\mathcal{F}$. Also let the set of contrasts with absolute test statistics greater than or equal to that of contrast f be the set $\mathcal{Z}_f = \left\{ i \in \mathcal{F} \;\middle|\; |\hat{D}_i^\lambda| \geq |\hat{D}_f^\lambda| \right\}$. Given some critical value, c, the test statistic associated with f will then be declared significant if $\forall\, j \in \mathcal{Z}_f : d_j > c$. With this additional rule we can now write the power function for a given λ as

$$\beta(k_f|\lambda) = \Pr\left\{ \forall\, j \in \mathcal{Z}_h : d_j > c_\alpha \;\middle|\; k_f, k_i = 1, i \neq f \right\} \tag{11}$$

where c_α is the critical value obtained by simulation as the solution to

$$\Pr\left\{ \forall\, j \in \mathcal{Z}_f : d_j > c_\alpha \;\middle|\; k_i = 1, i = 1, \ldots n-1 \right\} = \alpha.$$

6. EXAMPLES OF POWER COMPARISONS

In this section we will compare the power curves of $\beta(k_f|\lambda)$ for $\lambda = BH, BM, W, 0, 1/2, 1,$ and 2 for different sets of location and dispersion effects. We will let $\mathcal{L}$ denote the set of location effects and $\mathcal{D}$ the set of dispersion effects. For each value of k_f the comparison will be based on 100,000 unreplicated 16-run two-level factorial experiments where the standard four basic column vectors are labeled A, B, C and D and we will always assume the mean is included in the location model. For each of the 100,000 experiments we then remove the location effects, compute the residuals, estimate the test statistics for the dispersion effects, fit the line for the normal plot, and use the criterion discussed above. For a given λ the power is determined for $k_f = 1, 2, \ldots, 8$ and the power curve estimated by spline interpolation between these points.

6.1 POWER WHEN $\mathcal{L} = \{A\}$ AND $\mathcal{D} = \{A\}$

In our first example we will examine $\beta(k_f|\lambda)$ with the assumption of one active location effect and a single dispersion effect associated with the same factor. Since the choice of column vector clearly is irrelevant we will use $\mathcal{L} = \{A\}$ and $\mathcal{D} = \{A\}$. Following the methods for computing power curves as described in Section 5, we can estimate the power curves as shown in Figure 6 below.

Figure 6 shows that of the statistics considered the power associated with $\hat{D}_A^{BH}$, $\hat{D}_A^W$ or $\hat{D}_A^{1/2}$ appears to be larger than the power associated with the test statistics recommended by Daniel (1976), $\hat{D}_A^1$, Box and Meyer (1986), $\hat{D}_A^{BM}$, and Hoaglin, Mosteller, and Tukey (1983), $\hat{D}_f^1$ or $\hat{D}_A^2$. We also see that the method based on $\hat{D}_A^W$ is considerably more powerful than the method based on $\hat{D}_A^2$ even though the test statistics differ by only a constant.

This disparity between $\hat{D}_A^W$ and $\hat{D}_A^2$ is attributable to the different approaches to selecting the critical value for the two tests. When using $\hat{D}_A^W$ the reference distribution is constant given by the chi-square distribution and does not change regardless of the magnitude of the dispersion effect. However, since the "reference distribution" on a half-normal plot is constructed from the test statistics assumed inert, it will change if the test statistics are influenced by the magnitude of the dispersion effect being tested for. Using results from Johnson and Kotz (1972, p. 91) the variance $V\{\hat{D}_{f \neq A}^2\}$ of the test statistics $\hat{D}_i^2$ not associated with factor A, can be shown to be $3\sigma^2/64\,(k_A^2+1)$ and the expectation $E\{\hat{D}_A^2\}$, is $7\sigma^2/16\,(k_A^2-1)$. Thus the slope of the line drawn on the probability plot will likely decrease almost as fast as the increase in the expectation of $\hat{D}_A^2$.

Note that this in turn suggest that when using Wang's method the Type I error rate of test statistics associated with inert dispersion effects should increase along with the magnitude of k_A^2. This does in fact occur. However, as stated earlier, in a separate simulation study we found the increases to be marginal. Furthermore, Wang provided an additional check for significance during the estimation stage which most likely will screen out any dispersion effects mis-identified as active.

POWER WHEN $\mathcal{L} = \{A\}$ AND $\mathcal{D} = \{B\}$

We will now examine $\beta(k_f|\lambda)$ under the assumption of one active location effect associated with some

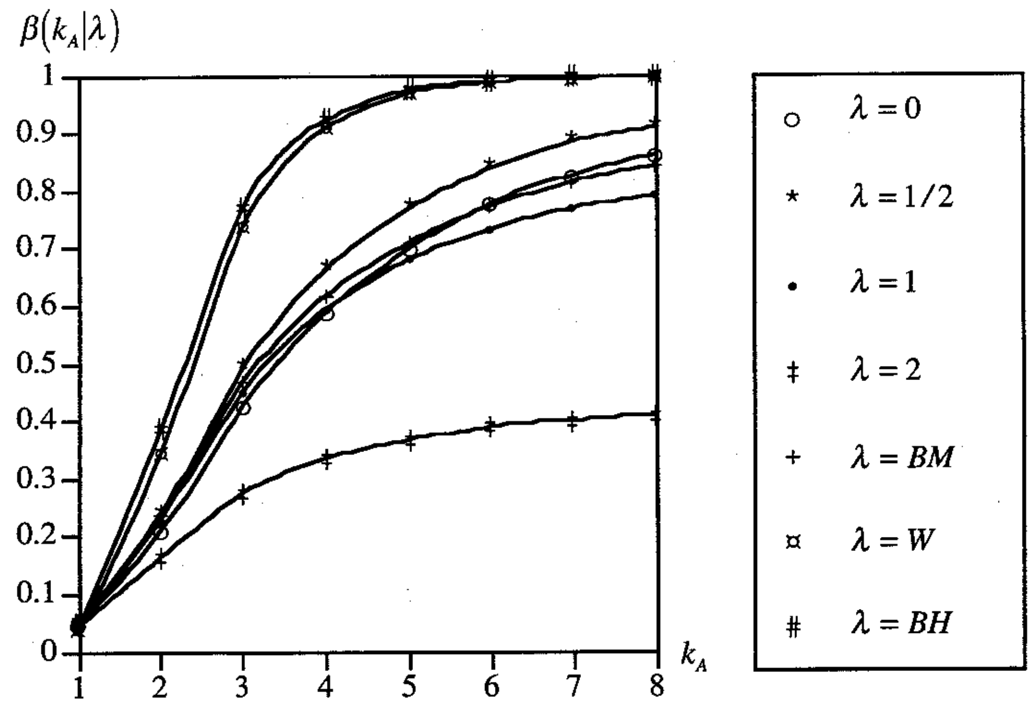


Figure 6. Power curves for $\lambda = \text{BH}, \text{BM}, \text{W}, 0, 1/2, 1,$ and 2 based on a 16-run two-level factorial experiment with $\mathcal{L} = \{A\}$ and $\mathcal{D} = \{A\}$.

basic column vector and a single dispersion effect associated with some other basic column vector. Any two basic column vectors will do so we will select $\mathcal{L} = \{A\}$ and $\mathcal{D} = \{B\}$. Thepower curves for $\lambda = \text{BH}, \text{BM}, \text{W}, 0, 1/2, 1,$ and 2 are shown below in Figure 7. Comparing Figure 6 to Figure 7 we see that all the curves show less power than before. However, the power curve associated with $\hat{D}_B^{BH}$ seems to have decreased the least. In the next section we will compare the power of the different methods when there are multiple location effects.

6.3 POWER WHEN $\mathcal{L} = \{A, B, AB\}$ AND $\mathcal{D} = \{A\}$

For the remainder of the comparisons we will examine $\beta(k_f|\lambda)$ when there are three location effects; two associated with basic column vectors and the third with their interaction. Thus let $\mathcal{L} = \{A, B, AB\}$. For $\mathcal{D} = \{A\}$ the power curves under these conditions are shown in Figure 8.

Figure 8 appears to show, once again, that the most powerful tests are those based on $\hat{D}_A^W$, $\hat{D}_A^{BH}$ or $\hat{D}_A^{1/2}$. Furthermore, the power curves associated with all the test statistics are similar to those in Figure 6. However, one noticeable difference is that Wang's method appears more powerful for small k_A in this example. Thus, we see that a uniformly most powerful test does not seem to exist among the methods considered. This may be due, in part, to the approximate test of significance associated with Wang's method. Next we show an example where the dispersion effect is associated with some contrast other than those associated with location effects.

6.4 POWER WHEN $\mathcal{L} = \{A, B, AB\}$ AND $\mathcal{D} = \{C\}$

The power curves for $\mathcal{L} = \{A, B, AB\}$ and $\mathcal{D} = \{C\}$ are shown in Figure 9. Comparing Figure 8 and Figure 9 we see that there is a considerable dif-

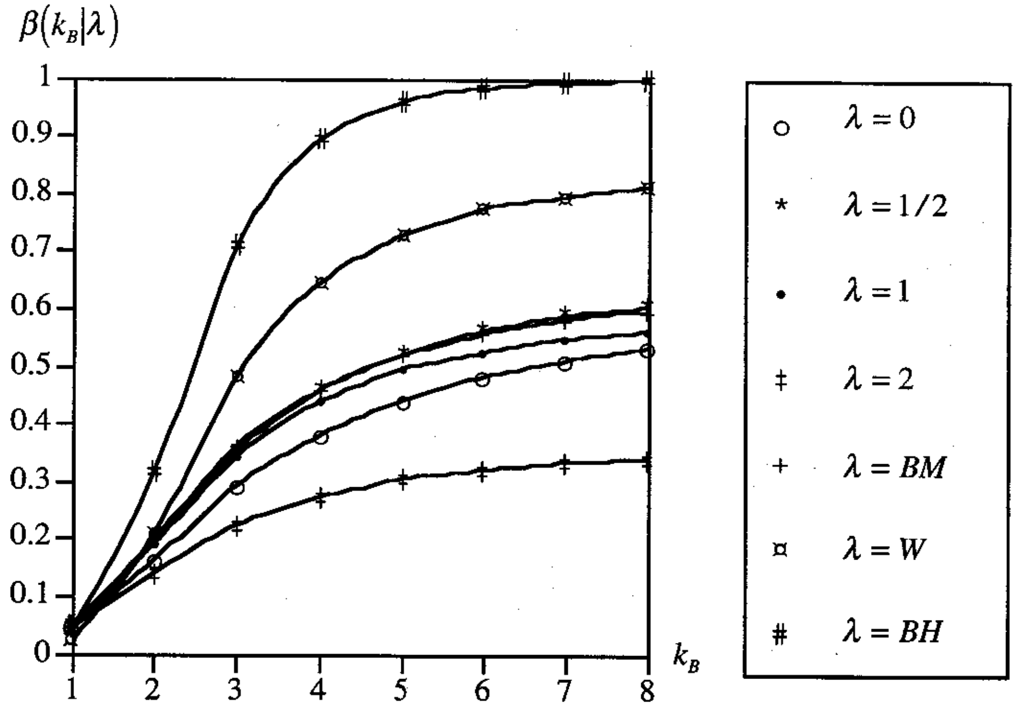


Figure 7. Power curves for $\lambda = BH, BM, W, 0, 1/2, 1,$ and 2 based on a 16-run two-level factorial.

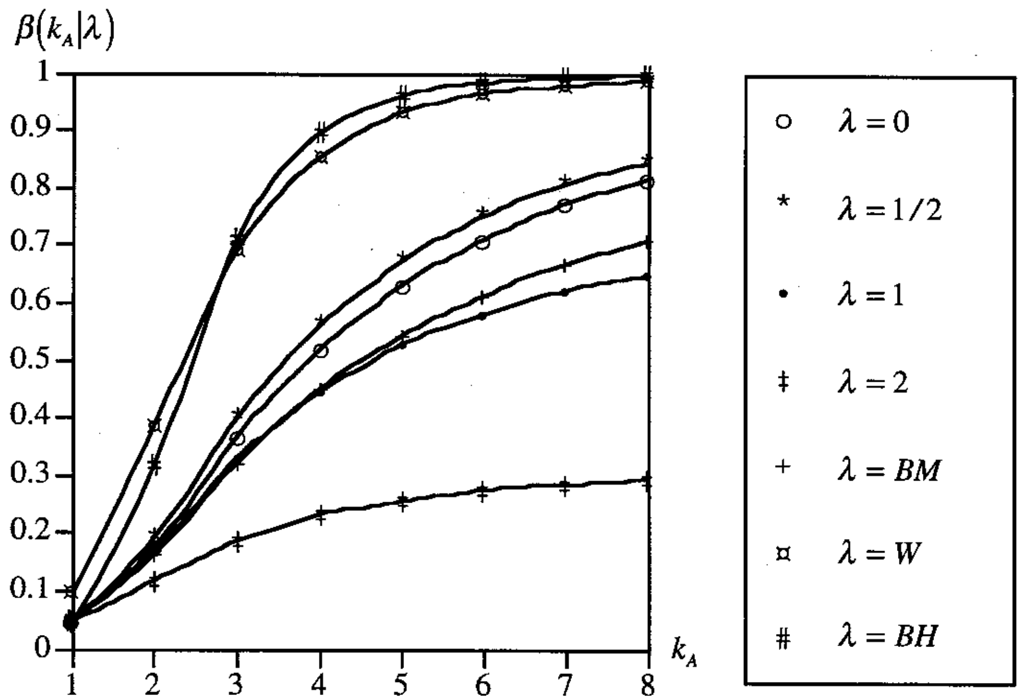


Figure 8. Power curves for λ = BH, BM, W, 0, 1/2, 1, and 2 based on a 16-run two-level factorial experiment with $\mathcal{L} = \{I, A, B, AB\}$ and $\mathcal{D} = \{A\}$.

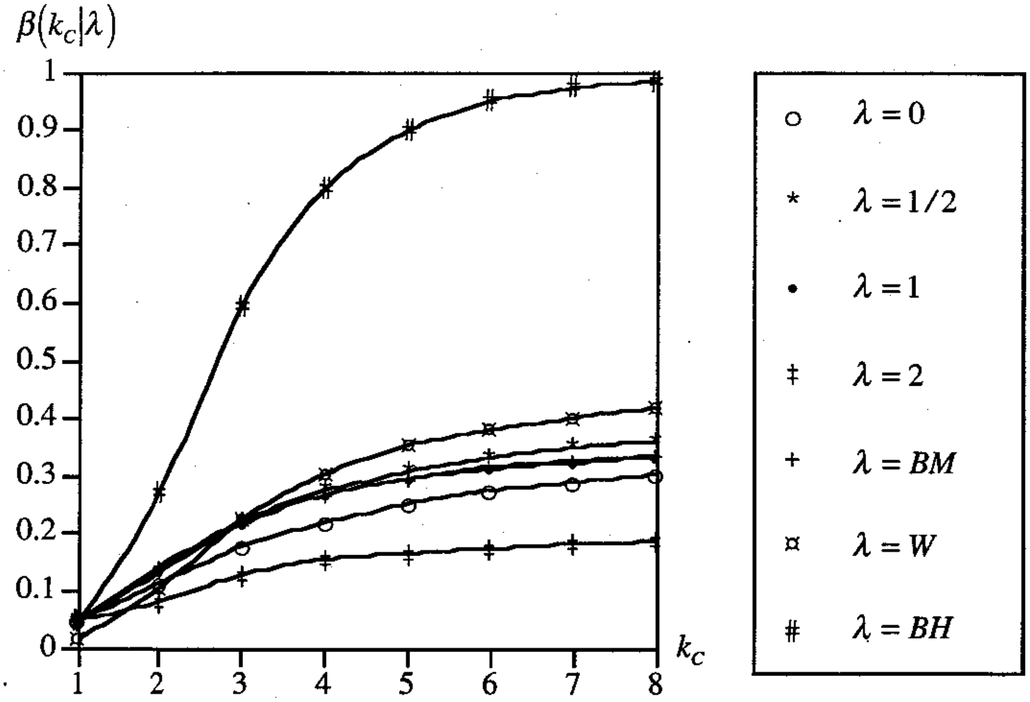


Figure 9. Power curves for λ = BH, BM, W, 0, 1/2, 1, and 2 based on a 16-run two-level factorial experiment with $\mathcal{L} = \{I, A, B, AB\}$ and $\mathcal{D} = \{C\}$.

ference in the detectability of a dispersion effect associated with factor A as compared to one associated with factor C for all the methods except Bergman and Hynen's (1995). Thus for $\lambda \neq BH$ the influence of the location effects on $\hat{D}_C^{\lambda}$ appears to be much more harmful than the influence of the location effects on $\hat{D}_A^{\lambda}$. This comparison indicates that Bergman and Hynen's (1995) method has an advantage over all others being considered in this article.

The conclusions drawn from these examples are consistent with a larger study we conducted. In every case Bergman and Hynen's (1995) method was either among the most powerful methods or significantly more powerful than all other methods considered. Two other relatively powerful methods are those proposed by Wang based on $\lambda = W$ and the one proposed by us based on $\lambda = 1/2$. This latter method has the advantage that it essentially amounts to repeating the same type of analysis performed on the location effects but using $|r_i|^{1/2}$ for $i = 1, 2, \ldots, n$ as the response. Thus, it is exceedingly simple to apply in connection with any software set up for analyzing two-level factorials. Since this method performed relatively well for most of the cases we have studied we think it merits serious consideration by practitioners appreciating simplicity.

7. SUMMARY AND CONCLUSIONS

In this paper we have examined several methods for identifying dispersion effects in unreplicated factorials. To do so necessitated that we operationalized the visual decision procedure associated with the half-normal probability plot. Strictly speaking the power curves computed using this operationalization should not be plotted with the others. However, by doing so we have provided an indication of which methods are likely the best with respect to power. Thus, through analytic results and simulation we have shown that, of the methods considered, the three most powerful methods are $\hat{D}_A^{BH}$, $\hat{D}_A^{W}$ and $\hat{D}_A^{1/2}$. Furthermore, in terms of power Bergman and Hynen's method is the preferred technique but on balancing simplicity and power our method based on $\hat{D}_A^{1/2}$ is a close contender.

8. ACKNOWLEDGMENT

The work on this manuscript was supported by a grant from the Alfred P. Sloan Foundation, the Samuel C. Johnson Distinguished Fellowship Fund, and by the National Science Foundation (NSF) Grant No. ECD8721545. The authors would like to thank the CQPI reports committee for their helpful comments on a previous draft of this manuscript.

9. REFERENCES

Bartlett, M. S. and Kendall, D. G. (1946). "The Statistical Analysis of Variance – Heterogeneity and the Logarithmic Transformation," *Journal of the Royal Statistical Society*, Supplement Vol. VIII, No. 1, 128-138.

Bergman, B. and Hynen, A. (1995). "Dispersion Effects from Unreplicated Designs in the 2^{k-p} Series," *Technometrics*, to appear.

Box, G. E. P. and Draper, N. R. (1969). Evolutionary Operation, John Wiley and Sons, N. Y.

Box, G. E. P. and Jenkins, G. M., (1976). Time Series Analysis: Forecasting and Control, Holden Day, Inc.

Box, G. E. P. and Meyer, D. (1986). "Dispersion Effects from Fractional Designs," *Technometrics*, Vol. 28, No. 1., 19-27.

Cook, R. D. and Weisberg, S. (1983). "Diagnostics for Heteroscedasticity in Regression," *Biometrika*, Vol. 70, No. 1, 1-10.

Daniel, C. (1976). Applications of Statistics to Industrial Experimentation, John Wiley and Sons, N. Y.

Daniel, C. (1983). "Half – Normal Plots," Encyclopedia of Statistics, edited by Kotz, S. and Johnson, N. L., John Wiley and Sons, N. Y, 565-568.

Hoaglin, D. C., Mosteller, F., and Tukey, J. W. (1983). Understanding Robust and Exploratory Data Analysis, John Wiley & Sons, Inc., NY.

Johnson, N. L. and Kotz, S. (1972) Distributions in Statistics, Continuous Multivariate Distributions, John Wiley and Sons, Inc., NY.

Miller, A. and Wu, C. F. J. (1993). "Improving a Calibration System through Designed Experiments," to appear.

Montgomery, D. C. (1991). "Using Fractional Factorial Designs for Robust Process Development," *Quality Engineering*, Vol. 3, No. 2, 193-205.

Nair, V. N. and Pregibon, D. (1988). "Analyzing Dispersion Effects form Replicated Factorial Experiments," *Technometrics*, Vol. 30, No. 3., 247-257.

Taguchi, G. and Wu, Y. (1985). Introduction to Off-line Quality Control, Nagoya, Japan: Central Japan Quality Control Association.

Wang, P. C. (1989). "Tests for Dispersion Effects from Orthogonal Arrays," *Computational Statistics and Data Analysis*, Vol. 8 , 109-117.