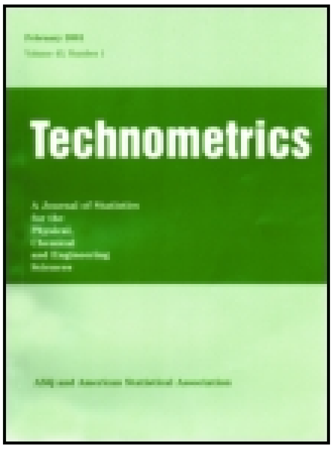

Technometrics
Publication details, including instructions for authors and subscription information:
http://amstat.tandfonline.com/loi/utch20

Adaptive Inference for the Two-Sample Scale Problem

Peter Hall [a] & A. R. Padmanabhan [b]

[a] Centre for Mathematics and its Applications Australian National University , Canberra , ACT , 2601 , Australia

[b] Department of Mathematics , Monash University , Clayton , Victoria , 3168 , Australia
E-mail:

Published online: 12 Mar 2012.

Adaptive Inference for the Two-Sample Scale Problem

Peter HALL
Centre for Mathematics
and its Applications
Australian National University
Canberra, ACT 2601
Australia

A. R. PADMANABHAN
Department of Mathematics
Monash University
Clayton, Victoria 3168
Australia
(appaswamy.padmanabhan.@sci.monash.edu.au)

Some adaptive tests are proposed (a) for the general two-sample scale problem and (b) for the restricted two-sample scale problem, in which the random variables are known to be right-skewed and also positive valued. These are based on either the ratio of the adaptively trimmed sample variances or adaptive rank tests of scale, whose quantiles are estimated by means of the bootstrap. Simulations show that, for practical sample sizes, these tests tend to perform better than their competitors. The methodology is illustrated by an application to a problem considered by Nair regarding the breakdown times of an insulating fluid under elevated voltage stress.

KEY WORDS: Bootstrap; Monte Carlo studies; Tests of homogeneity of variances.

The problem of testing for scale arises in a variety of contexts, including quality control and analysis of outer continental shelf bidding on oil and gas (Conover, Johnson, and Johnson 1981), chemistry (Bethea, Duran, and Boullion 1975, p. 153), and engineering (Hald 1967, p. 290; Mendenhall, Wackerly, and Scheaffer 1990, p. 464; Nair 1984, p. 824). Therefore, it is worthwhile to develop an effective methodology for this problem. In this connection, Conover et al. (1981) undertook an extensive simulation study of the tests of scale and made some recommendations. Subsequently, Good and Chernick (1993) introduced a test based on the ratio of sample variances. For the preceding problem, this article proposes two adaptive tests, the first based on the ratio of adaptively trimmed sample variances and the second test based on an adaptive rank test of scale.

The restricted two-sample scale problem, in which the underlying random variables are known to be right-skewed and are positive valued, arises in connection with life-testing (Schafer and Sheffield 1976), study of amplitude of sound derived from many independent sources (Qureishi 1964), target detection (Woinsky 1972), and atmospheric sciences (Simpson 1972). For this problem, this article proposes another adaptive rank test.

The quantiles of these adaptive statistics are estimated by means of the bootstrap. Monte Carlo studies show that the corresponding tests have, for practical sample sizes, either greater robustness of validity (in terms of level) or greater robustness of efficiency (in terms of power) or both, than their competitors.

Section 1 describes the test statistics, the bootstrap, and the adaptive procedures. Section 2 contains the Monte Carlo studies and makes some recommendations for the practicing statistician. Section 3 illustrates the methodology by applying it to a problem in electrical engineering.

1. SOME PRELIMINARIES

1.1 The Scale Statistics

Let F be a distribution with median 0, $\mathbf{X}_i = (X_{i1}, \ldots, X_{in_i})$ denote independent samples from distributions F_i, where $F_i(u) = F((u-m_i)/\sigma_i), i = 1, 2$ (so that F_i has median m_i) with m_i and σ_i unknown, and $\Delta = \sigma_2/\sigma_1$ be the ratio of the scale parameters. For $0 < \alpha < .5$, let s_i^2 and $s_{i,\alpha}^2$ denote the sample variances of the $\mathbf{X}_i$ sample and the symmetrically α-trimmed $\mathbf{X}_i$ samples, respectively, $i = 1, 2$. We define the trimmed variance ratio V_α as $V_\alpha = s_{2,\alpha}^2/s_{1,\alpha}^2$. Because s_2^2/s_1^2 is the APF statistic of Shorack (1969), we call V_α the trimmed APF statistic.

For $i = 1, 2$ and $l = 1, 2, \ldots, n_i$, let $X_{i,\text{med}}$ denote the median of the sample $\mathbf{X}_i, Y_{il} = X_{il} - X_{i,\text{med}}, \mathbf{Y}_i = (Y_{i1}, \ldots, Y_{im_i}), B_{il} = |Y_{il}|$, and $\mathbf{B}_i = (B_{i1}, \ldots, B_{in_i})$. Denote by $\mathbf{B}_{i.}$ the mean of $\mathbf{B}_i, \mathbf{B}_{..}$, the mean of the combined sample $(\mathbf{B}_1, \mathbf{B}_2)$, and $N = n_1 + n_2$. Then the Lev1:med statistic is

$$\frac{(N-2)(n_1(\mathbf{B}_{1.} - \mathbf{B}_{..})^2 + n_2(\mathbf{B}_{2.} - \mathbf{B}_{..})^2)}{\sum_{r=1}^{n_1}(B_{1r} - \mathbf{B}_{1.})^2 + \sum_{t=1}^{n_2}(B_{2t} - \mathbf{B}_{2.})^2}$$

(Brown and Forsythe 1974).

The Good–Chernick procedure is also based on s_2^2/s_1^2. For estimating the quantiles of s_2^2/s_1^2, however, it uses a special kind of bootstrap (Good and Chernick 1993), whereas the APF test uses the quantiles of a certain $\mathcal{F}$ distribution.

Denote by R_{2j} the rank of B_{2j} in $(\mathbf{B}_1, \mathbf{B}_2)$. For some set of scores $a_N(1), \ldots, a_N(N)$, define

$$h = \sum_{j=1}^{n_2} a_N(R_{2j}). \tag{1}$$

Some examples of h are given in Table 1, where $\mathbf{\Phi}^{-1}$ denotes the inverse of the cumulative normal distribution function.

The first three are modifications of the proposals of Fligner and Killen (1976, p. 210), and the fourth is the F-

K:med statistic of Conover et al. (1981). The first and fourth statistics are centered forms of a statistic given by Hajek and Sidak (1967, p. 74).

Remark 1.1. Conover et al. (1981) recommended the use of the F-K:med and Lev1:med tests.

Remark 1.2. A slight modification is required when we are also interested in confidence intervals. The endpoints of a confidence interval (CI) will be the ratios of the centered observations y_{2j}/y_{1i} for some i and j. For these ratios to make sense, the centered values must be nonzero. When n_i is odd, we ensure this by centering at $((n_i-1)/n_i)\times$ the sample median. Let $\mathbf{Y}_i'$ be the resulting sample and h_i' the corresponding statistic. Monte Carlo studies show that, for $n_i \geq 10$, the h_i' test is virtually identical to the h_i test regarding level and power. Even after this modification, however, there may be one or more zero values, especially in small samples. For example, let $n_1 = n_2 = 5$ and $X_1 = (.3, .4, .5, .6, .7)$. Then $Y_1' = (-.1, .0, .1, .2)$. In such cases, we discard the zero values and construct the CI using the remaining values.

Next consider the following restricted two-sample problem. Suppose that $\mathbf{X}_i = (X_{i1}, \ldots, X_{in_i}), i = 1, 2$, are independent samples from F_i, where, for some location parameter $b_i, F_i(u) = F((u-b_i)/\sigma_i), b_i \geq 0, \sigma_i > 0$, and F is right-skewed and also positive. Let $X_{i,\min}$ be the minimum of the $\mathbf{X}_i$-sample, and write $\tilde{Y}_{il} = X_{il} - X_{i,\min}, l = 1, 2, \ldots, n_i$, and $\tilde{\mathbf{Y}}_i = (\tilde{Y}_{i1}, \ldots, \tilde{Y}_{in_i})$. Denote by $\tilde{R}_{2j}$ the rank of $\tilde{Y}_{2j}$ in $(\tilde{\mathbf{Y}}_1, \tilde{\mathbf{Y}}_2)$. For some set of scores $\tilde{a}_N(1), \ldots, \tilde{a}_N(N)$, define

$$\tilde{h} = \Sigma \tilde{a}_N(\tilde{R}_{2j}). \tag{2}$$

Some examples of scores are given in Table 2. The statistic $\tilde{h}_1$ is the sum of the squared ranks statistic (Duran and Mielke 1968), and $\tilde{h}_2$ and $\tilde{h}_3$ are the Wilcoxon and Savage statistics, respectively (Hajek and Sidak 1967).

Remark 1.3. Note that, unlike $h_1 - h_4$, which are based on samples centered at the sample medians, $\tilde{h}_1 - \tilde{h}_3$ are based on samples centered at the sample minima. The latter three statistics are defined because centering at the sample medians seriously affects the power of the test and also the performance of the bootstrap.

Remark 1.4. Once again, a slight modification is required when we are also interested in CI's. The endpoints of a CI will be ratios of the centered observations $\tilde{y}_{2j}/\tilde{y}_{1i}$. Because the values are centered at the sample minima, however, in each (centered) sample, there will be a zero value. To resolve this problem, we center the sample values at $((n_i-1)/n_i)\times$ the sample minimum, n_i being the sample size; that is, we work with $\tilde{Y}_{il}' = X_{il} - ((n_i - 1)/n_i)X_{i,\min}, l = 1, 2, \ldots, n_i, i = 1, 2$. Note that $\tilde{Y}_{il}'$ will be strictly positive for any sample size because any value in the X_i sample will be greater than or equal to $X_{i,\min}$ and hence strictly greater than $((n_i-1)/n_i)X_{i,\min}$. For $j = 1-3$, let $\tilde{h}_j'$ be obtained from $\tilde{h}_j$ by replacing $\tilde{Y}_{il}$ by $\tilde{Y}_{il}'$. Monte Carlo studies show that for $n_i \geq 10$, the $\tilde{h}_j'$ test is virtually identical to the $\tilde{h}_j$ test regarding level and power.

Table 1. Some Statistics for the General Two-Sample Scale Problem

$a_N(i), i = 1, \ldots, N$	Statistic
$(\Phi^{-1}((N+i)/(2N+1)))^2$	h_1
$(i/N+1)^2$	h_2
$(i/N+1)$	h_3
$\Phi^{-1}((N+i)/(2N+1))$	h_4

Table 2. Some Statistics for the Restricted Two-Sample Scale Problem

$\tilde{a}_N(i), i = 1, 2, \ldots, N$	Statistic
$(i/N+1)^2$	$\tilde{h}_1$
$(i/N+1)$	$\tilde{h}_2$
$\frac{1}{N+1-i} + \frac{1}{N+2-i} + \cdots + \frac{1}{N}$	$\tilde{h}_3$

1.2 The Bootstrap

Let us begin with the general two-sample scale problem. The asymptotic distribution of V_α is intractable even for $n_1 = n_2 = 60$ (Hall and Padmanabhan 1990). Therefore, its quantiles are estimated using the following classical bootstrap or the percentile method (Efron and Tibshirani 1986, p. 68). Write $Z_r = (X_{1r} - \bar{\mathbf{X}}_1)/s_1$, and $Z_{t+n_1} = (X_{2t} - \bar{\mathbf{X}}_2)/s_2, r = 1, \ldots, n_1, t = 1, 2, \ldots, n_2$. Let $Z_1, Z_2, \ldots, Z_{2,000}$ be the values of V_α for these samples and $Z_{(1)} \leq Z_{(2)} \leq \cdots \leq Z_{(2,000)}$ the corresponding ordered values. Then $Z_{(1,901)}$ is an estimate of the 95% quantile of V_α.

Let h and $a_N(i)$ be as in (1) and $\bar{a}_N = (1/N)\Sigma a_N(i)$. Suppose that the null hypothesis of equal scales is valid and $\min(n_1, n_2) \to \infty$. Then the following results (a) and (b) can be proved by the methods of Fligner and Hettmansperger (1979) (Hettmansperger 1992, personal communication).

(a) For symmetric F, the asymptotic null distribution of h is normal with mean $n_2\bar{a}_N$ and variance

$$\frac{n_1 n_2}{(N-1)} \sum (a_N(i) - \bar{a}_N)^2. \tag{3}$$

(b) For skewed F, the asymptotic null distribution of h is normal with the same mean but variance generally larger than (3). Moreover, this variance will involve the unknown F.

Note that in particular (a) and (b) apply to $h_1' - h_3'$ and the F-K:med statistic h_4'. Therefore, the quantiles of all these statistics are estimated by the following "smooth" bootstrap, which is a modification of an earlier proposal due to Collings and Hamilton (1988). This technique works well for all standard rank statistics of scale, including the Mood, Ansari–Bradley, quartile, and Klotz normal scores statistics.

Let h' denote any of the preceding four statistics, and, for $i = 1, 2$, let MADi denote the median of $\mathbf{B}_i = |\mathbf{Y}_i| = (|Y_i|, \ldots, |Y_{im_i}|)$—that is, the median of the absolute deviations from the median of the $\mathbf{X}_i$ sample. Consider the pooled sample

$$\left(\frac{\mathbf{Y}_1}{\text{MAD1}}, \frac{\mathbf{Y}_2}{\text{MAD2}}\right) = \left(\frac{Y_{11}}{\text{MAD1}}, \ldots, \frac{Y_{1n_1}}{\text{MAD1}}, \frac{Y_{21}}{\text{MAD2}}, \ldots, \frac{Y_{2n_2}}{\text{MAD2}}\right).$$

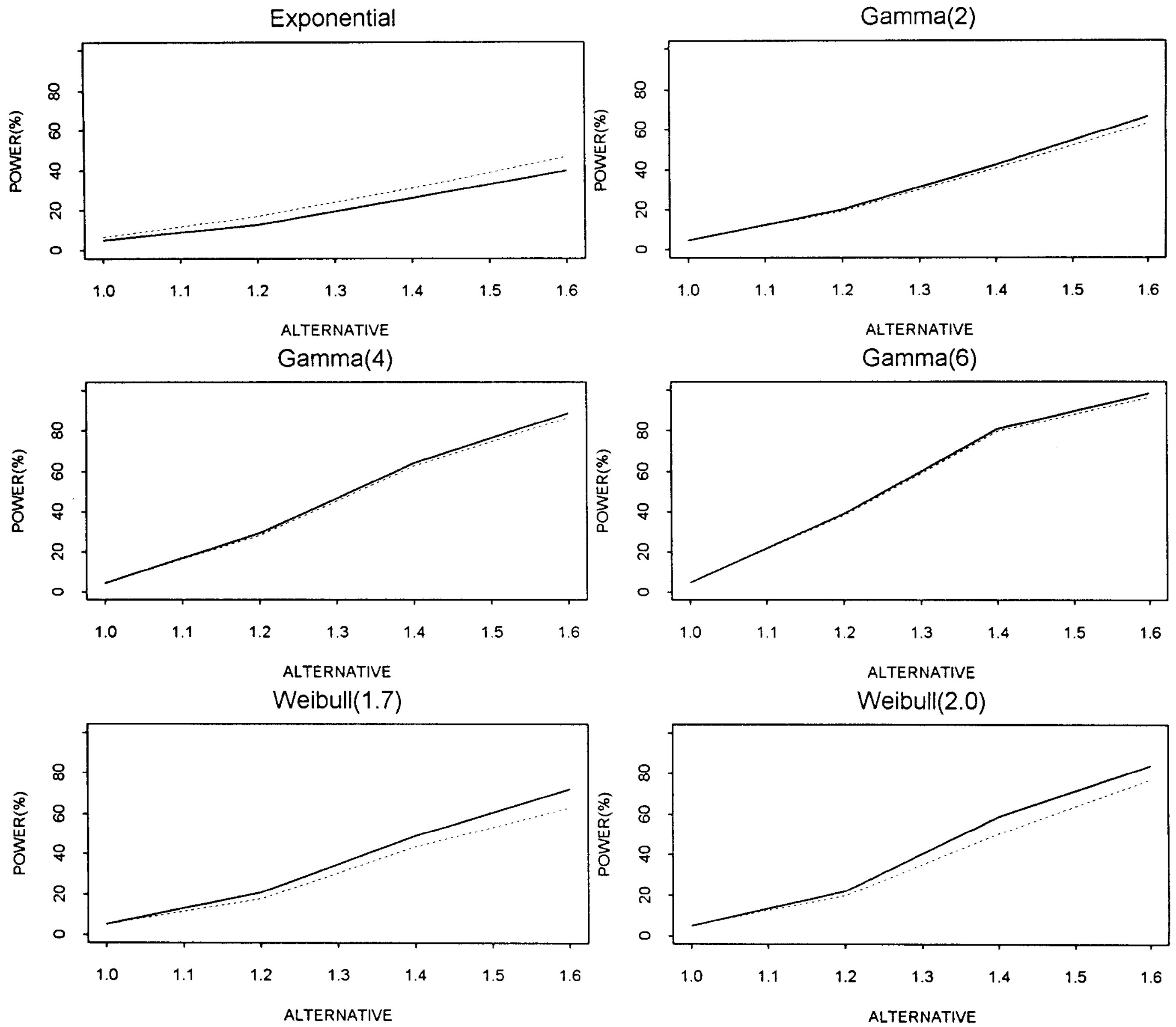

Figure 1. Simulated Powers of Tests for the Restricted Two-Sample Scale Problem: Sample Size (n1, n2) = (20, 20); ———, Adaptive; . . . ., Woinsky.

Denote by $U_{(1)} \leq U_{(2)} \leq \cdots \leq U_{(n_1+n_2)}$ the order statistics of the combined sample. Define $U_{(n_1+n_2+1)} = 2U_{(n_1+n_2)} - U_{(n_1+n_2-1)}$. Let F^* denote the continuous distribution function that assigns uniformly the probability $1/(n_1 + n_2)$ to the interval $(U_{(k)}, U_{(k+1)}), k = 1, 2, \ldots, n_1 + n_2$. First, a bootstrap sample of size $n_1 + n_2$, say $z_1^*, \ldots, z_{n_1}^*, z_{n_1+1}^*, \ldots, z_{n_1+n_2}^*$ is drawn. Write $x_{11}^* = z_1^*, \ldots, x_{1n_1}^* = z_{n_1}^*, \mathbf{x}_1^* = (x_{11}^*, \ldots, x_{1n_1}^*), x_{21}^* = z_{n_1+1}^*, \ldots, x_{2n_2}^* = z_{n_1+n_2}^*$, and $\mathbf{x}_2^* = (x_{21}^*, \ldots, x_{2n_2}^*)$. For $i = 1, 2$, let $x_{i,\text{med}}^*$ denote the median of the $\mathbf{x}_i^*$ sample and $\mathbf{y}_i^*$ denote the centered sample $\mathbf{x}_i^* - (n_i - 1/n_i)x_{i,\text{med}}^*$. Suppose that $h'(\mathbf{y}_1^*, \mathbf{y}_2^*) = h_1'^*$. This process is repeated 1,999 more times. Suppose that these 2,000 values are $h_1'^*, h_2'^*, \ldots, h_{2,000}'^*$ and the corresponding ordered values are $h_{(1)}'^* \leq h_{(2)}'^* \leq \cdots \leq h_{(2,000)}'^*$. Then $h_{(1,901)}'^*$ is an estimate of the 95% (null) quantile of h'.

Let us now turn to the restricted two-sample problem. For the corresponding statistics $\tilde{h}_1' - \tilde{h}_3'$, the only change in bootstrap is that now each bootstrap sample has to be centered at $(n_i - 1/n_i)$ times the sample minimum, n_i being the sample size, $i = 1, 2$.

Remark 1.5. At present, the bootstrap seems to be the best way of estimating the quantiles of the rank statistics. Permutation procedures take up much more time and are much less efficient. For example, let $n_1 = n_2 = 20$, the nominal level be 5% and the null distribution be exponential. The empirical level of the (one-sided) Ansari–Bradley test based on 2,000 bootstrap samples (for each sample) was about 7%. The corresponding level based on 10,000 permutations (for each sample) was over 10%, although the latter took up much more computing time.

1.3 The Adaptive Procedures

For $i = 1, 2$, a measure that has been associated with

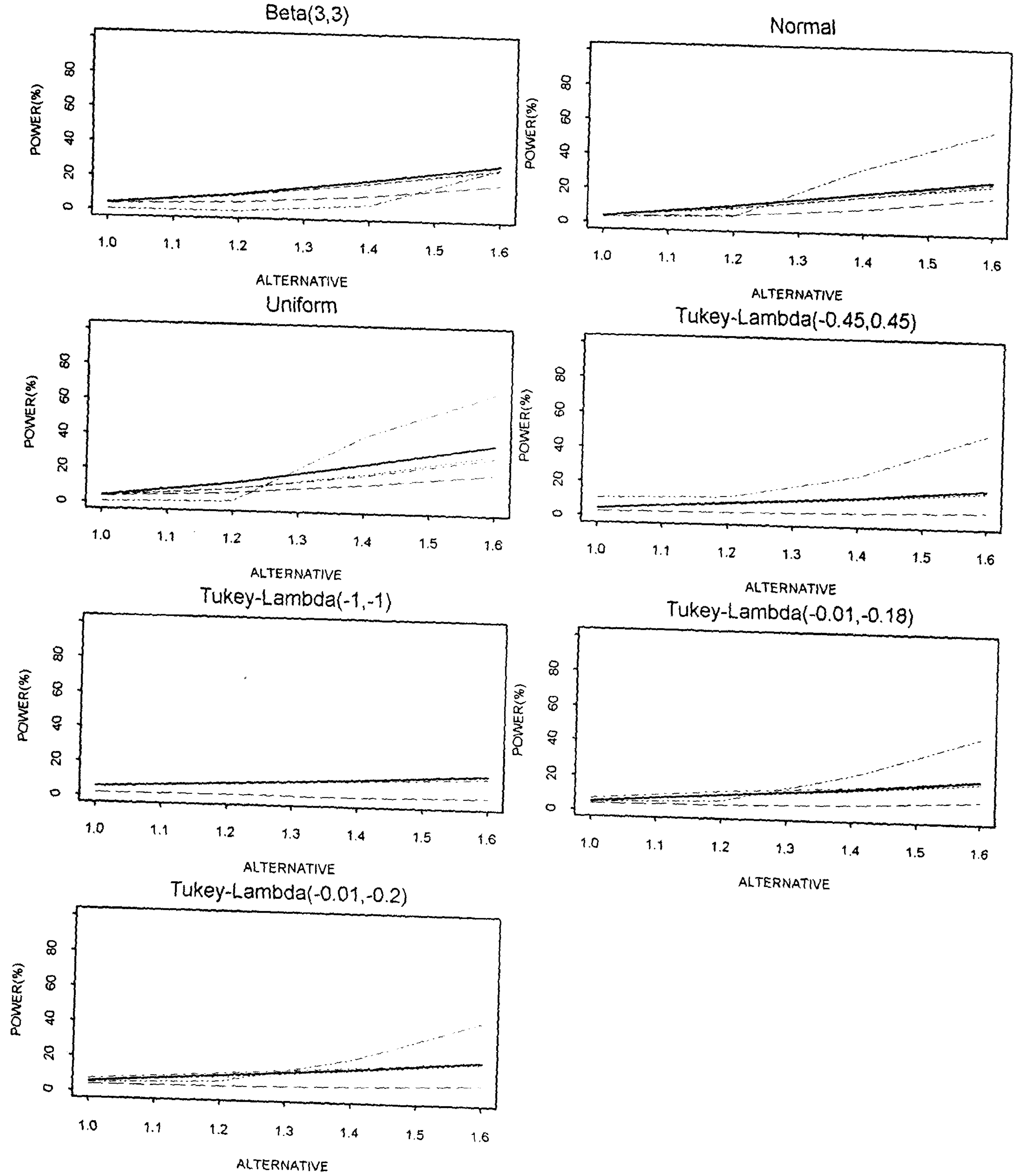

Figure 2. Simulated Powers of Tests for the General Two-Sample Scale Problem: Sample Size (n1, n2) = (10, 10): Good–Chernick—not Computable for Tukey–Lambda (−1, −1) Because Some of the Associated Quantities Become Excessively Large: ———, Adaptive II; . . . ., F.K:med; —·—, Adaptive I; - - -, Levl:med; —··—, Good–Chernick.

the sample $\mathbf{X}_i$ for the purpose of adaptation is the tail-weight Q_{2i} (Hogg, Randles, and Fisher 1975). For our problem, we combine them into the weighted average $Q_2 = (n_1Q_{21} + n_2Q_{22})/(n_1 + n_2)$. Let $\bar{\mathbf{X}}_i, s_i^2$ and $c_i = \bar{X}_i/s_i$ be, respectively, the sample mean, the sample variance, and the reciprocal of the (estimated) coefficient of variation and $c = (n_1c_1 + n_2c_2)/(n_1 + n_2)$. Adaptive Procedures I and II are designed for the general two-sample problem, but Adaptive Procedure III is designed for the restricted two-sample problem.

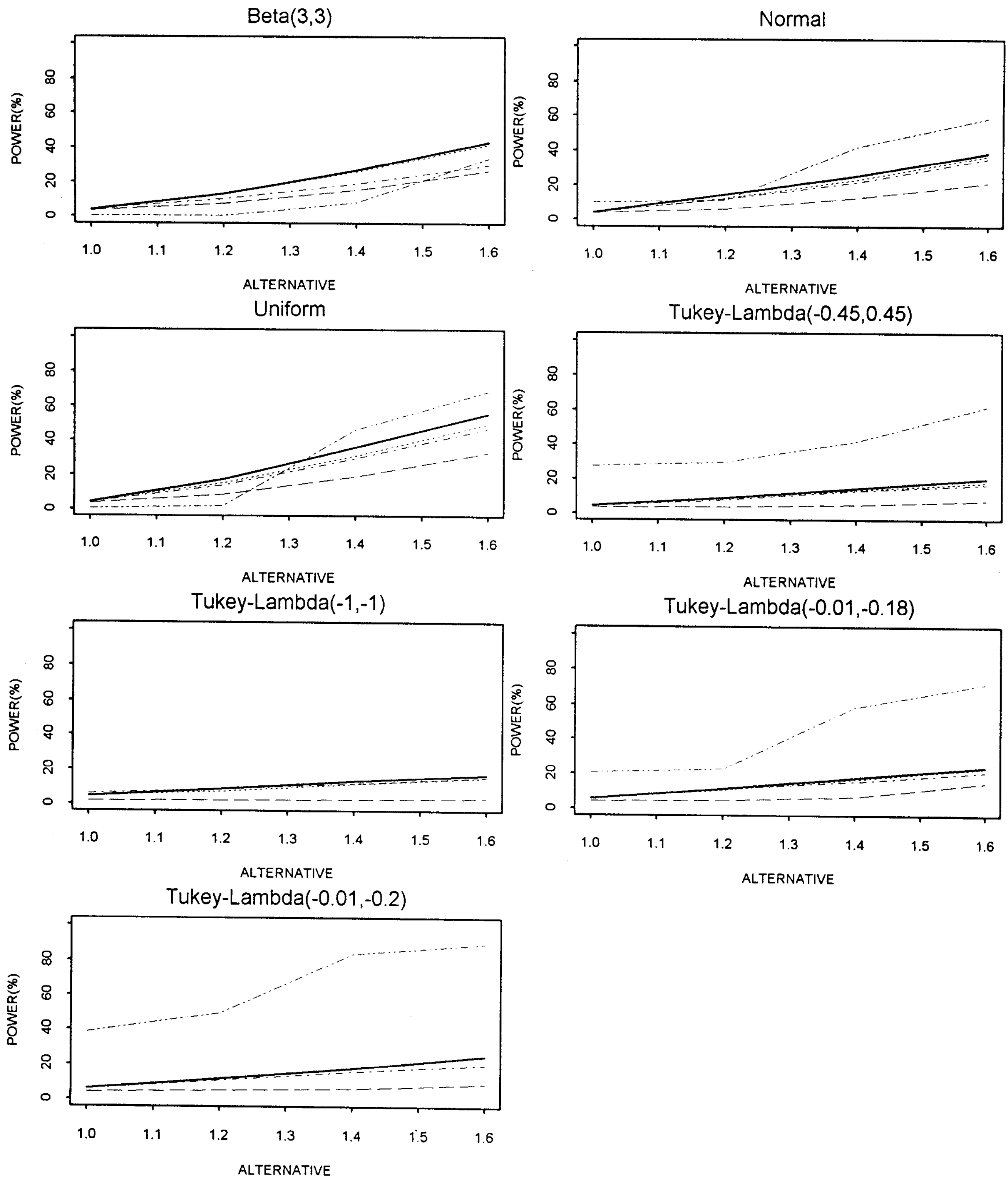

Figure 3. Simulated Powers of Tests for the General Two-Sample Scale Problem: Sample Size (n1, n2) = (10, 20): Good–Chernick—not Computable for Tukey–Lambda (−1, −1) (cf. Fig. 2): ———, Adaptive II; . . . ., F.K: med; —·—·, Adaptive I; - - -, Levl: med; —··—··, Good–Chernick.

Adaptive Procedure I: Use $V_{.10}$ if $Q_2 < 2.5$, $V_{.25}$ otherwise.

Adaptive Procedure II: Use h_1' if $Q_2 < 3, h_3'$ if $Q_2 \geq 5, h_2'$ otherwise.

Adaptive Procedure III: Use $\tilde{h}_3'$ if $c < 2.24, \tilde{h}_2'$ if $c \geq 3.22, \tilde{h}_1'$ otherwise.

Let $T_1 = c \log(\bar{x}_2/\bar{x}_1)$ and T_2 be the Wilcoxon statistic based on the original (and not on the centered) sample values. For the restricted two-sample scale problem, Woinsky (1972, pp. 71–73) proposed the statistic, say W, given by

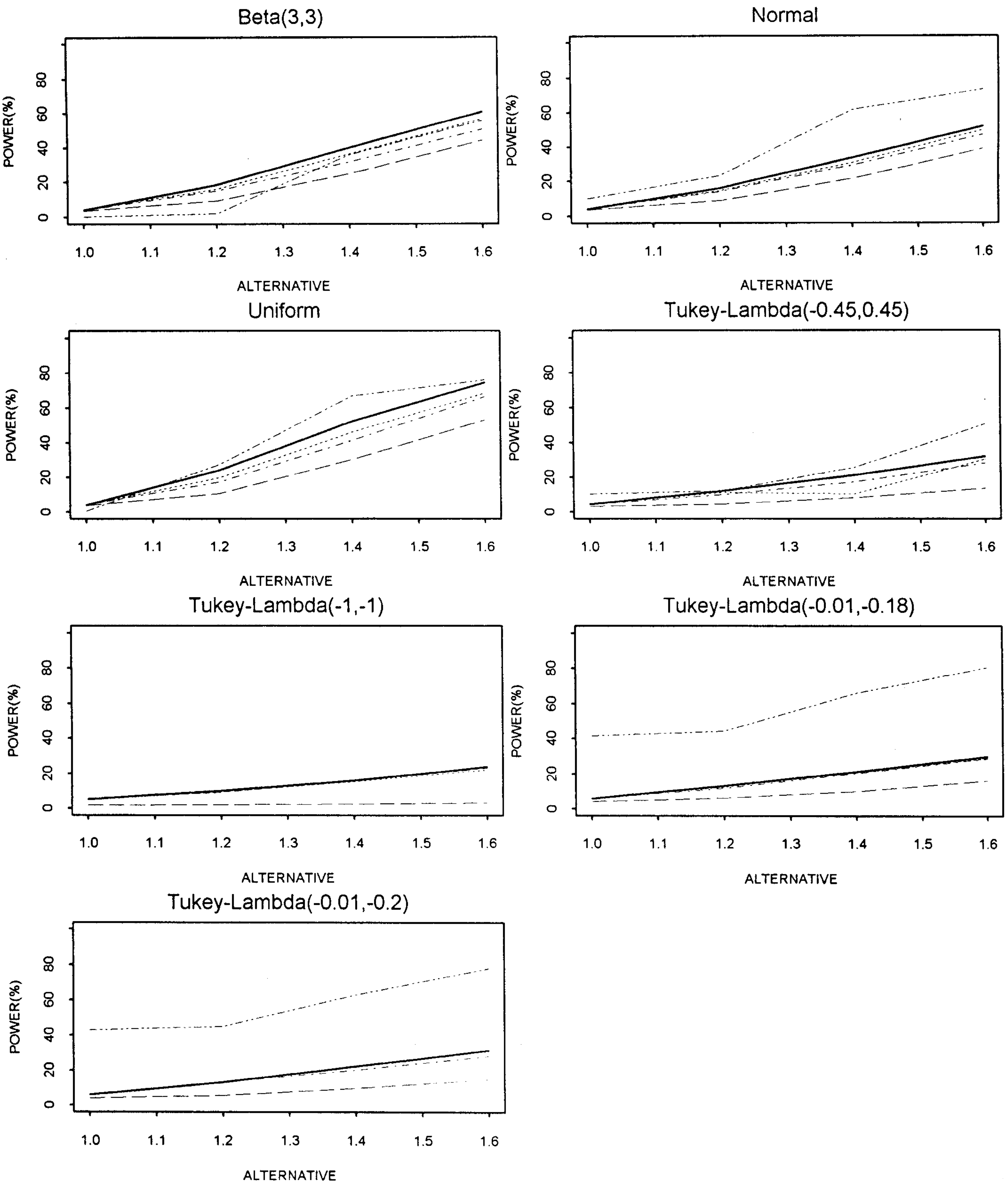

Figure 4. Simulated Powers of Tests for the General Two-Sample Scale Problem: Sample Size (n1, n2) = (20, 20): Good–Chernick—not Computable for Tukey–Lambda (−1, −1) (cf. Fig. 2): ———, Adaptive II; . . . ., F.K : med; —·—·, Adaptive I; - - -, Levl : med; —··—··, Good–Chernick.

$$
\begin{aligned}
W &= T_1, \quad c < 1.414 \\
&= T_2, \quad \text{otherwise.}
\end{aligned}
$$

The null quantiles of W can be estimated as by Woinsky (1972, p. 72). The W test requires the stronger assumption that $b_1 = b_2 = 0$ (Woinsky, 1972, p. 65), whereas Adaptive III requires only $b_1 \geq 0$ and $b_2 \geq 0$. To overcome this limitation, we may modify the W test by basing its constituent statistics not on the original samples but on the samples centered at $(n_i - 1)/n_i$ times the sample minimum and es-

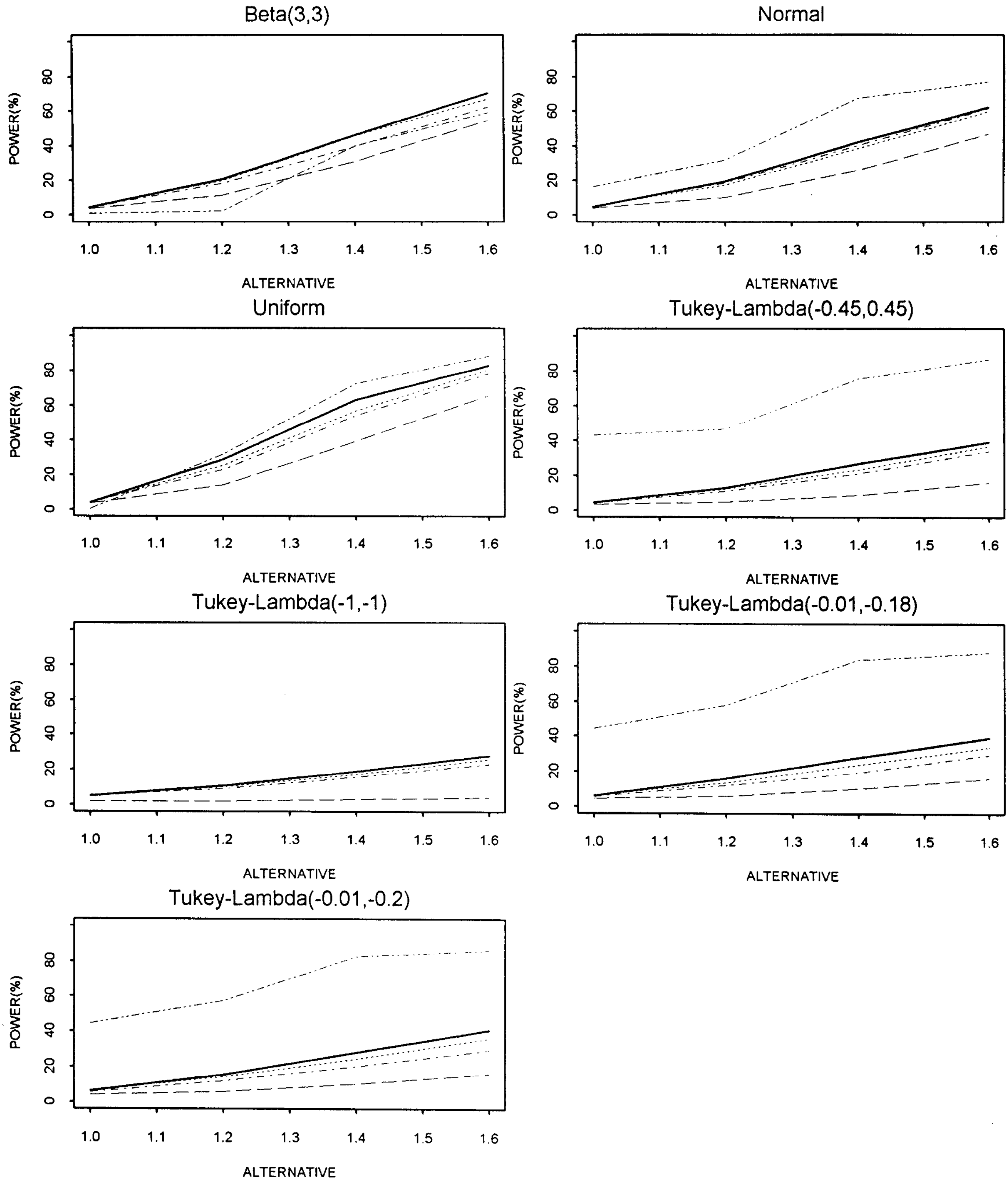

Figure 5. Simulated Powers of Tests for the General Two-Sample Scale Problem: Sample Size (n1, n2) = (20, 30): Good–Chernick—not Computable for Tukey–Lambda (−1, −1) (cf. Fig. 2): ———, Adaptive II; . . . ., F.K: med; —·—·, Adaptive I; - - -, Levl: med; —··—··, Good–Chernick.

timate the null quantiles of this modified statistic, using the bootstrap. The resulting test, however, lacks robustness of validity (cf. Sec. 2).

Computation of the estimators and confidence intervals is described in the Appendix.

Remark 1.6. For estimation in the location and regression models, an adaptive procedure, based on the sum H of a tailweight indicator H_2 and an indicator H_3 of peakedness was proposed by Yuh and Hogg (1988). For hypothesis testing, however, this procedure could have some drawbacks. This result, combined with theory and some simulations, shows that our procedures are superior to that of

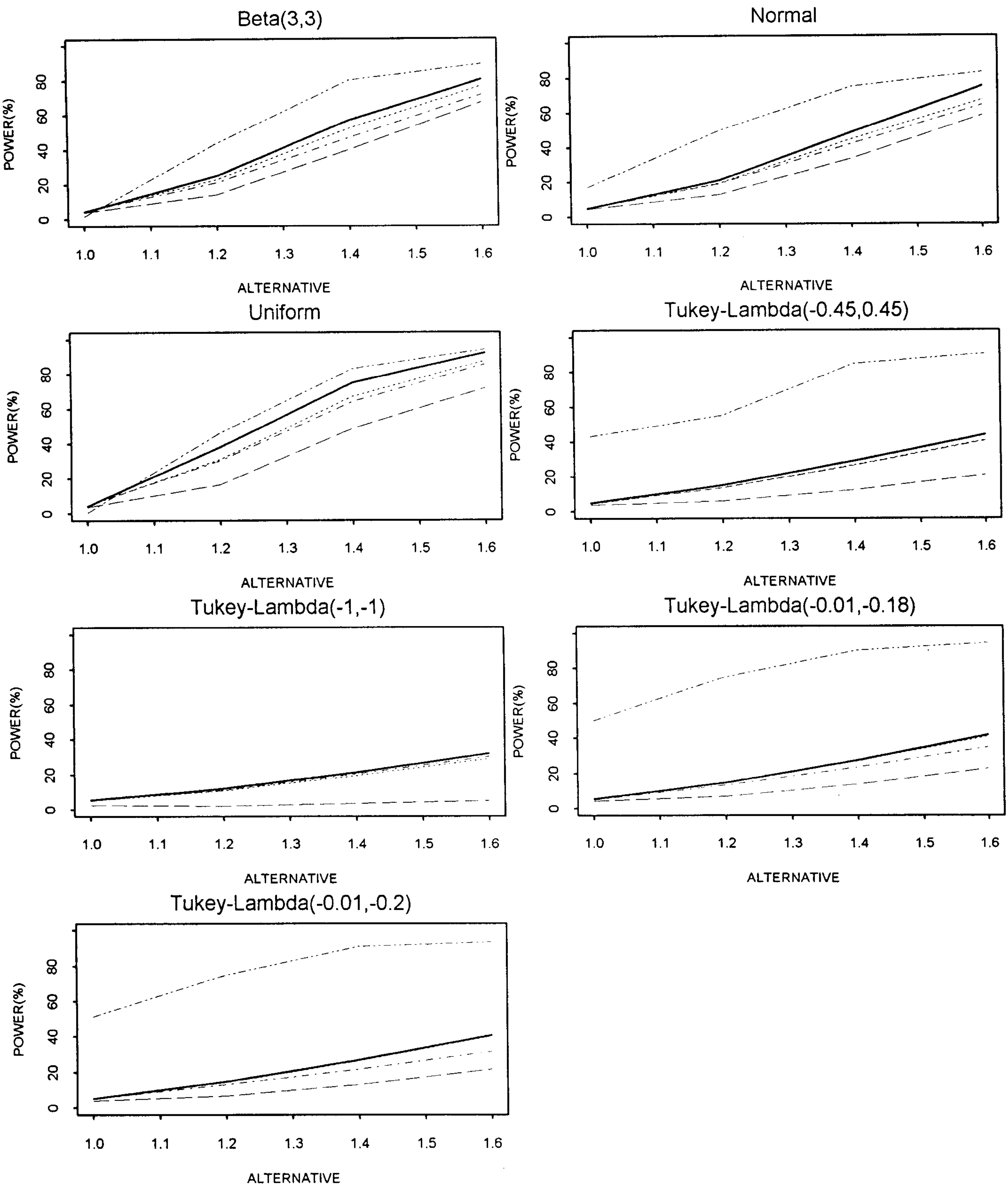

Figure 6. Simulated Powers of Tests for the General Two-Sample Scale Problem: Sample Size (n1, n2) = (30, 30): Good–Chernick—not Computable for Tukey–Lambda (−1, −1) (cf. Fig. 2): ———, Adaptive II; . . . ., F.K: med; –·–·, Adaptive I; - - -, Levl: med; –··–··, Good–Chernick.

Yuh and Hogg in terms of robustness of validity as well as robustness of efficiency. To save computing time, peaked distributions have been excluded in the detailed simulation studies of Section 2. Nevertheless, theory and Monte Carlo studies show that our procedures work well for peaked distributions and in particular the double exponential. For brevity, the details regarding the preceding assertions are omitted.

2. THE MONTE CARLO STUDIES

2.1 The Restricted Two-Sample Scale Problem

For $r \geq 1$, let Weibull(r) and gamma(r) denote, respectively, the standard Weibull and gamma distributions with shape parameter r. The simulation studies involved Adaptive III and the Woinsky test for the distributions (standard) exponential, gamma(2), gamma(4), gamma(6), Weibull(1.7), and Weibull(2.0). These will be called distributions 1, 2, 3, 4, 5, and 6, respectively. The last two were chosen because of their importance in life-testing and the study of the amplitude of sound derived from many independent sources (Qureishi 1964).

Ten thousand samples of sizes $(n_1, n_2) = (20, 20)$ were drawn from each distribution on a VAX23 computer using IMSL subroutines. The nominal level was kept at 5%. For each statistic, the proportion of times it was at least as large as its estimated 95% quantile constituted its empirical level ($\Delta = 1$). Then all values in the second sample were multiplied by 1.2, and the preceding process repeated the empirical powers corresponding to $\Delta = 1.2$. Empirical powers corresponding to $\Delta = 1.4$ and 1.6 were computed similarly and are displayed in Figure 1, page 414.

Figure 1 shows that Adaptive III has robustness of validity because the empirical level is close to the nominal level. By contrast, the empirical level of the Woinsky test for the exponential is slightly over 6%, which is statistically significantly different from the nominal level. Hence it is clear that, for the gamma and Weibull distributions, Adaptive III has also greater robustness of efficiency.

In the preceding study, the locations b_1 and b_2 were assumed to be 0. For at least the Weibull distributions, however, it is more realistic to assume that $b_1 \geq 0$ and $b_2 \geq 0$ and b_1 and b_2 are possibly unequal (cf. Schafer and Sheffield 1976). Being unaffected by location changes, Adaptive III continues to apply and gives the same results. Woinsky's test, however, has to be modified by basing it on the observations centered at the respective sample minima. This modified version lacks robustness of validity; its empirical levels (in percent) for distributions $1-5$ are 1.30, 1.45, 3.95, 3.1, and 3.1, respectively.

2.2 The General Two-Sample Scale Problem

Let U be a standard rectangular variate. Then for arbitrary real numbers (λ_1, λ_2), Tukey–lambda (λ_1, λ_2) will denote the distribution of $U^{\lambda_1} - (1-U)^{\lambda_2}$ (cf. Moberg, Randles, and Ramberg 1980). The simulation studies compared Adaptive I, Adaptive II, $F-K$:med, Lev1:med, and Good–Chernick (GC) tests for the distributions beta(3, 3), the standard normal and Tukey–lambda (λ_1, λ_2) for $(\lambda_1, \lambda_2) = (-.45, -.45), (-1, -1), (-.01, -.18)$, and $(-.01, -.20)$. The first five distributions are symmetric and the last two are highly skewed to the right. The fourth and fifth distributions have heavy and very heavy tails, respectively. Ten thousand samples of sizes (n_1, n_2) were drawn from each distribution on a VAX23 computer using IMSL subroutines. The configurations for (n_1, n_2) were (10, 10), (10, 20), (20, 20), (20, 30), and (30, 30). Once again, the nominal level was kept at 5% ($\Delta = 1$), and power was calculated at $\Delta = 1.2$, 1.4, and 1.6.

Figures 2–6, pages 415–419, show that Adaptive I is good in terms of robustness of validity and robustness of efficiency. Besides, the resulting estimator and confidence intervals are easy to compute. Adaptive II is the best in terms of robustness of efficiency and is good in terms of robustness of validity. The $F-K$:med is second only to Adaptive II in terms of robustness of efficiency. It is also good in terms of robustness of validity. The Lev1:med is conservative, with the resulting loss of power, especially for heavy-tailed distributions. This is reasonably consistent with the findings of Brown and Forsythe (1974) and Miller (1986, p. 169). The GC procedure is quite liberal even for the normal distribution and extremely liberal for skewed distributions and symmetric heavy-tailed distributions, sometimes as high as 50% for Tukey–lambda $(-.01, -.18), n_1 = n_2 = 30$. The high powers of the GC method are therefore misleading. In fact, the APF test (Shorack 1969, pp. 993–1013) has far greater robustness of validity than the GC test, although both procedures use the ratio of the sample variances.

In the light of these findings, we make the following recommendations: For the general two-sample scale problem, Adaptive I and Adaptive II both provide good performance; Adaptive I is computationally simpler. For the restricted two-sample scale problem, Adaptive III is preferred.

3. AN ILLUSTRATION

Table 3 (Nair 1984, p. 824) gives the times (in minutes) to breakdowns of an insulating fluid under elevated voltage stresses of 32 KV (= the $\mathbf{X}_1$ sample) and 36 KV (= the $\mathbf{X}_2$ sample).

We shall test H_0 against the obvious two-sided alternative at the 5% level by using CI's. Because life-times tend to be right-skewed, Adaptive III also applies. Let h be an adaptive statistic. For Adaptive I, $h = V_{.25}$, and the corresponding CI is (.008, .225). For Adaptive II, $h = h_2'$ and its (estimated) 2.5% and 97.5% quantiles are 3,199 and 6,156, respectively. The corresponding quantiles of h_5' are, respectively, 62 and 163. Hence the CI based on h_5' is $[a(62), a(163)] = [.021, .202]$, where $a(1) \leq a(2) \leq \cdots \leq a(225)$ are the ordered ratios $y_{2j}'/y_{1i}', 1 \leq i, j \leq 15$. Starting with this and performing a few iterations as described in the Appendix, we get the CI corresponding to h_2' as [.0226, .199].

For Adaptive III, $c = .18, h = \tilde{h}_3'$. The quantiles of $\tilde{h}_3'$ are 9 and 21, respectively, and those of $\tilde{h}_2'$ are 63 and

Table 3. Times to Breakdowns

Voltage stress	Time in minutes						
$x_1 = 32$	.27	.40	.69	.79	2.75	3.91	9.88
	13.95	15.93	27.30	53.24	82.85		
	89.25	100.58	215.50				
$x_2 = 36$	.35	.59	.96	.99	1.69	1.97	
	2.07	2.58	2.71	2.90	3.67		
	3.99	5.35	13.77	25.50			

162, respectively. Hence the CI corresponding to $\tilde{h}'_2$ is $(\tilde{a}(63), \tilde{a}(162)) = (.021, .233)$. Starting with this and performing a few iterations (as in Adaptive II), we get the CI corresponding to $\tilde{h}'_3$ as (.023, .209). Because none of the CI's contain 1, all of these procedures reject H_0.

ACKNOWLEDGMENTS

We are deeply grateful to the editor, past editor S. Vardeman, the associate editor, and the referees for valuable suggestions and comments, and to Professors Hettmansperger, Sethuraman, Koul, and Rao S. Jammalamadaka for the asymptotics in the earlier versions, to Professor Kochar for information on the restricted two-sample scale problem, and to Professor Lenth and Dr. McCloud for the high-quality graphics.

APPENDIX: COMPUTATION OF THE ESTIMATORS AND CONFIDENCE INTERVALS

In the case of Adaptive I, let V be the adaptive statistic and $V^*_{.025}$ and $V^*_{.975}$ be its estimated 2.5% and 97.5% quantiles, respectively. Then an estimate of Δ is simply $\sqrt{V}$ and a 95% confidence interval for Δ is (a_L, a_U), where $a_L^2 = V/V^*_{.975}$ and $a_U^2 = V/V^*_{.025}$. In the two-sample location model, the estimator and the CI's corresponding to the Wilcoxon statistic are known in closed forms (Hollander and Wolfe 1973, p. 78). For h'_3 and $\tilde{h}'_2$, the estimators and the CI's are appropriate scale analogs of those for h_W and hence can be calculated similarly, as will be explained. In almost all other cases, they are computed by the algorithm of Bauer (1972), which could require many iterations even for moderate samples. Therefore, we propose the following shortcut. Let h' be the statistic based on Adaptive II and $[\Delta_L, \Delta_U]$ the corresponding 95% CI. Let $h'_{.025}$ and $h'_{.975}$ be its estimated 2.5% and 97.5% (null) quantiles, respectively. The 95% CI, say $[\Delta_{L_3}, \Delta_{U_3}]$ corresponding to h'_3, can be computed as in the next paragraph. Taking $\Delta_{U_3}(\Delta_{L_3})$ as a starting point, we can compute $\Delta_U(\Delta_L)$ in a few iterations. The estimator (corresponding to h') is exactly or very nearly the midpoint of $[\Delta_L, \Delta_U]$. Next let $\tilde{h}'$ be the statistic based on Adaptive III and $[\tilde{\Delta}_L, \tilde{\Delta}_U]$ the corresponding 95% CI. Now the 95% CI corresponding to $\tilde{h}'_2$, say, $[\tilde{\Delta}_{L_2}, \tilde{\Delta}_{U_2}]$ is known in a closed form. Starting with $\tilde{\Delta}_{U_2}(\tilde{\Delta}_{L_2})$, we can compute $\tilde{\Delta}_U(\tilde{\Delta}_L)$ in a few iterations. The precise details are given in the following paragraphs.

Write $Y'_{kl} = X_{kl} - ((n_k-1)/n_k)X_{k,\text{med}}, l = 1, \ldots, n_k, k = 1, 2$. Let $a(1) \le a(2) \le \cdots \le a(n_1 n_2)$ denote the $n_1 n_2$ ratios $Y'_{2j}/Y'_{1i}, (j = 1, \ldots, n_2, i = 1, \ldots, n_1)$ arranged in the ascending order. Then h'_3 is simply the Wilcoxon statistic based on $(|Y'_{11}|, \ldots, |Y'_{1n_1}|; |Y'_{21}|, \ldots, |Y'_{2n_2}|)$. Hence, $h'_5 = h'_3 - \min h'_3 = h'_3 - n_2(n_2+1)/2$ is simply the corresponding Mann–Whitney statistic. Let $L'_3(L'_5)$ and $U'_3(U'_5)$ be, respectively, the (estimated) 2.5% and 97.5% quantiles of $\tilde{h}'_3(h'_5)$. Then $L'_5 = L'_3 - n_2(n_2+1)/2$ and $U'_5 = U'_3 - n_2(n_2+1)/2$. An appropriate modification of the arguments in the location model shows that the 95% CI corresponding to h'_5 is simply $[a(L'_5), a(U'_5)]$.

Starting with $a(U'_5)$, we can compute Δ_U in a few steps as follows: For any $\rho > 0$, write $(y'_1, y'_2/\rho) = (y'_{11}, \ldots, y'_{1n_1}; y'_{21}/\rho, \ldots, y'_{2n_2}/\rho)$. Considered as a function of $\rho, h'(y'_1, y'_2/\rho)$ has the following properties:

1. It is a constant within each of the open intervals $(a(i), a(i+1)), i = 1, 2, \ldots, n_1 n_2 - 1$.
2. It is decreasing as ρ increases; that is, its value in $(a(i+1), a(i+2))$ is less than or equal to its value in $(a(i), a(i+1)), 1 \le i \le n_1 n_2 - 2$.

For $1 \le j \le n_1 n_2 - 1$, let $b(j)$ be the midpoint of $(a(j), a(j+1))$. To obtain the value of $h'(y'_1, y'_2/\rho)$ in the preceding interval, it suffices to compute its value when $\rho = b(j)$ (in view of property 1). Now Δ_U is characterized by the property that for any $\rho < \Delta_U, h'(y'_1, y'_2/\rho) \ge h'_{.025}$ and for any $\rho > \Delta_U, h'(y'_1, y'_2/\rho) \le h'_{.025}$. To compute Δ_U, we start with $a(U'_5)$. Recall that $b(i)$ is the midpoint of $(a(i), a(i+1))$, and in particular $b(Y'_5 - 1)$ is the midpoint of $(a(U'_5 - 1), a(U'_5))$. Compute $h'(y'_1, y'_2/\rho)$ when $\rho = b(U'_5 - 1)$. Suppose that it is $\le h'_{25}$. Then we consider the value when $\rho = b(U'_5 - 2)$ and so forth. Proceeding thus, we can find $b(l)$ and $b(l+1)$ such that $h'(y'_1, y'_2/b(l+1)) \le h'_{.025}$ and $h'(y'_1, y'_2/b(l)) \ge h'_{.025}$. Then $\Delta_U = a(l+1)$.

Next suppose that at $\rho = b(U'_5 - 1), h(y'_1, y'_2/\rho)$ is $\ge h'_{.025}$. Then we consider the value when $\rho = b(Y'_5)$ and so forth. Proceeding thus, we can find $b(k)$ and $b(k+1)$ such that $h'(y'_1, y'_2/b(k+1)) \le h'_{.025}$ and $h'(y'_1, y'_2/b(k)) \ge h'_{.025}$. Then $\Delta_U = a(k+1)$. Normally at most four or five iterations will be required. Similarly starting with $a(L'_5), \Delta_L$ can be computed using the following characterization of Δ_L: $h(y'_1, y'_2/\rho) \ge h'_{.975}$ or $\le h'_{.975}$ according as $\rho < \Delta_L$ or $\rho > \Delta_L$.

We now turn to $[\tilde{\Delta}_L, \tilde{\Delta}_U]$, the CI corresponding to $\tilde{h}'$ based on Adaptive III. We indicate the details very briefly in view of their similarity to those of Adaptive II. Let $\tilde{h}'_{.025}$ and $\tilde{h}'_{.975}$ be the null quantiles of the adaptive statistic $\tilde{h}'$. Write $\tilde{Y}'_{kl} = Y_{kl} - ((n_k - 1)/n_k)X_{k,\min}, l = 1, \ldots, n_k, k = 1, 2$. Let $\tilde{a}(1) \le \tilde{a}(2) \le \cdots \le \tilde{a}(n_1 n_2)$ denote the $n_1 n_2$ ratios $\tilde{y}'_{2j}/\tilde{y}'_{1i}, (j = 1, \ldots, n_2, i = 1, \ldots, n_1)$. Then $\tilde{h}_2$ is simply the Wilcoxon statistic based on $(\tilde{y}'_{11}, \ldots, \tilde{y}'_{1n_1}; \tilde{y}'_{21}, \ldots, \tilde{y}'_{2n_2})$. Hence $\tilde{h}'_5 = \tilde{h}'_2 - \min \tilde{h}'_2 = \tilde{h}'_2 - n_2(n_2+1)/2$ is the corresponding Mann–Whitney statistic. Let $\tilde{L}'_{2,.0225}(\tilde{L}'_{5,.025})$ and $\tilde{U}'_{2,.975}(\tilde{U}'_{5,.975})$ be, respectively, the (estimated) 2.5% and 97.5% quantiles of $\tilde{h}'_2(\tilde{h}'_5)$. Then $\tilde{L}'_5 = \tilde{L}'_2 - n_2(n_2+1)/2$ and $\tilde{U}'_5 = \tilde{U}'_2 - n_2(n_2+1)/2$. Arguments similar to the location model show that the 95% CI corresponding to $\tilde{h}'_5$ is simply $[\tilde{a}(\tilde{L}'_5), \tilde{a}(\tilde{U}'_5)]$. This is also the CI corresponding to $\tilde{h}'_2$ because of the equivalence of $\tilde{h}'_2$ and $\tilde{h}'_5$. Starting with $[\tilde{a}(\tilde{L}'_5), \tilde{a}(\tilde{U}'_5)]$, we can compute $[\tilde{\Delta}_L, \tilde{\Delta}_U]$ in a few steps.

[Received September 1992. Revised February 1997.]

REFERENCES

Bauer, D. F. (1972), "Constructing Confidence Sets Using Rank Statistics," *Journal of the American Statistical Association*, 67, 687–690.

Bethea, R. M., Duran, B. S., and Boullion, T. L. (1975), *Statistical Methods for Engineers and Scientists*, New York: Marcel-Dekker.

Brown, M. B., and Forsythe, A. B. (1974), "Robust Tests for the Equality of Variances," *Journal of the American Statistical Association*, 69, 364–

367.

Collings, B. J., and Hamilton, M. A. (1988), "Estimating the Power of the Two-Sample Wilcoxon Test for Shift," *Biometrics*, 44, 847–876.

Conover, W. J., Johnson, M. E., and Johnson, M. M. (1981), "A Comparative Study of the Tests for Homogeneity of Variances, With Applications to the Outer Continental Shelf Bidding Data," *Technometrics*, 23, 351–361.

Duran, B. S., and Mielke, P. W. (1968), "Robustness of Sum of Squared Ranks Test," *Journal of the American Statistical Association*, 63, 338–344.

Efron, B., and Tibshirani, R. (1986), "Bootstrap Methods for Standard Errors, Confidence Intervals and Other Measures of Statistical Accuracy," *Statistical Science*, 1, 54–75.

Fligner, M. A., and Hettmansperger, T. P. (1979), "On the Use of Conditional Asymptotic Normality," *Journal of the Royal Statistical Society*, Ser. B, 41, 178–183.

Fligner, M. A., and Killen, T. J. (1976), "Distribution-Free Two-Sample Tests for Scale," *Journal of the American Statistical Association*, 71, 210–213.

Good, P., and Chernick, M. (1993), "Testing the Equality of Variances of Two Populations" unpublished manuscript.

Hajek, J., and Sidak, Z. (1967), *Theory of Rank Tests*, New York: Academic Press.

Hald, A. (1967), *Theory With Engineering Applications* (7th printing), New York: Wiley.

Hall, P., and Padmanabhan, A. R. (1990), "Adaptive Inference for the Two-Sample Scale Problem," Research Report, Monash University, Dept. of Mathematics.

Hogg, R. V., Randles, R. H., and Fisher, D. M. (1975), "A Two-Sample Adaptive Distribution-Free Test," *Journal of the American Statistical Association*, 70, 656–661.

Hollander, M., and Wolfe, D. A. (1973), *Nonparametric Statistics Methods*, New York: Wiley.

Mendenhall, W., Wackerly, D. D., and Schaeffer, R. L. (1990), *Mathematical Statistics With Applications* (4th ed.), Boston: PWS-Kent.

Miller, R. G. (1986), *Beyond ANOVA: Basics of Applied Statistics*, New York: Wiley.

Moberg, T. F., Randles, R. H., and Ramberg, J. S. (1980), "An Adaptive Multiple Regression Procedure Based on M-Estimators," *Technometrics*, 22, 213–224.

Nair, V. N. (1984), "On the Behavior of Some Estimators From Probability Plots," *Journal of the American Statistical Association*, 79, 823–830.

Qureishi, A. S. (1964), "The Discrimination Between Two Weibull Processes," *Technometrics*, 6, 57–75.

Schafer, R. E., and Sheffield, T. S. (1976), "On Procedures for Comparing Two Weibull Populations," *Technometrics*, 18, 231–235.

Shorack, G. (1969), "Testing and Estimating Ratios of Scale Parameters," *Journal of the American Statistical Association*, 64, 999–1013.

Simpson, J. (1972), "Use of the Gamma Distribution in Single-Cloud Rainfall Analysis," *Monthly Weather Review*, 100, 309–312.

Woinsky, M. N. (1972), "A Composite Nonparametric Test for A Scale Slippage Alternative," *The Annals of Mathematical Statistics*, 43, 65–73.

Yuh, L., and Hogg, R. V. (1988), "On Adaptive M-Regression," *Biometrics*, 44, 433–435.