Ann. Inst. Statist. Math.
Vol. 45, No. 2, 265–278 (1993)

STATISTICAL PROCEDURES BASED ON SIGNED RANKS IN k SAMPLES WITH UNEQUAL VARIANCES

TAKA-AKI SHIRAISHI

Department of Mathematics, Yokohama City University, Yokohama 236, Japan

(Received November 2, 1991; revised April 16, 1992)

Abstract. In k samples with unequal variances, test procedures based on signed ranks for the homogeneity of k location parameters are proposed. The asymptotic χ^2-distribution of the test statistics is shown. It is found that the asymptotic relative efficiency of the rank tests relative to Welch's test (1951, *Biometrika*, **38**, 330–336) under local alternatives agrees with that of the one-sample signed rank tests relative to the t-test. A simulation study for the goodness of the χ^2-approximate of significance points is done. Then, surprisingly it can be seen that the χ^2-approximate for the critical points of the proposed tests is better than that of Kruskal-Wallis test and the Welch-type test. Next R-estimators and weighted least squares estimators for common mean of k samples under the homogeneity of k location parameters are compared in the same way as the test case. Furthermore, positive-part shrinkage versions of R-estimators for the k location parameters are considered along with a modified James-Stein estimation rule. The asymptotic distributional risks of the usual R-estimators, the positive-part shrinkage R-estimators (PSRE's), and the preliminary test and shrinkage R-versions under an arbitrary quadratic loss are derived. Under Mahalanobis loss, it is shown that the PSRE's dominate the other R-estimators for $k \geq 4$. A simulation study leads strong support to the claims that the PSRE's dominate the other type R-estimators and they are robust about outliers.

Key words and phrases: Hypothesis-testing, parameter estimation, Behrens-Fisher's problem, asymptotic relative efficiency, asymptotic distributional risk, modified James-Stein rule, simulation study.

1. Introduction

The problem of testing the equality of location parameters and of estimating the parameters for $k \geq 2$ populations occupies a fundamental role in the statistical literature. Making robust procedures for inference of the parameters is the most important because the procedures are applicable for more relaxed models than parametric models and are not sensitive to contaminated models, for instance, models including outliers. Kruskal and Wallis (1952) proposed a rank test of Wilcoxon type. Using Chernoff and Savage's technique (1958), Puri (1964) derived the asymptotic local power of rank test based on general scores and showed that the Pitman asymptotic relative efficiency (ARE) of the rank tests with respect to the F-test agrees with that of the two-sample rank tests with respect to the t-test. As discussed in Hodges and Lehmann (1956) and Chernoff and Savage (1958), the rank test procedures are more efficient than the parametric test in many cases, and they are nearly efficient even when the underlying distribution is normal. Especially the ARE of the normal scores test with respect to the parametric test is larger than or equal to 1, and the equality is attained only when the underlying distribution is normal. The present paper deals with k samples with unequal variances ($k \geq 2$).

Consider the k samples so that the j-th observation X_{ij} in the i-th level has continuous distribution function $F((x-\mu_i)/\sigma_i)$ for $j = 1, \ldots, n_i$, $i = 1, \ldots, k$, with σ_i's nuisance parameters. It is assumed that X_{ij}'s are independent, the density $f(x) = F'(x)$ is symmetric about zero, i.e., $f(-x) = f(x)$ for all x, and the second moment of $F(x)$ is finite. $\int_{-\infty}^{\infty} x^2 dF(x) = 1$ are assumed for convenience. Then μ_i and σ_i^2 become, respectively, a mean and variance of the i-th population. We want to test the null hypothesis $H_0; \mu_1 = \cdots = \mu_k$ versus the alternative $A; \mu_i \neq \mu_j$ for some $i \neq j$.

When $F(x)$ is normal and $k = 2$, it is well-known that this is the Behrens-Fisher problem, and Welch (1949) had considered a test based on a statistic which has asymptotically a standard normal distribution under H_0. Furthermore, Welch (1951) proposed a test statistic having approximately F-distribution for $k \geq 3$. Pratt (1964) showed that the asymptotic level depends on σ_1/σ_2 and that it is not stable. We propose test procedures based on signed ranks and show the asymptotic χ^2-distributions of the test statistics as all the sample sizes tend to infinity. Then it is found that the asymptotic relative efficiency of the rank tests relative to the Welch test under local alternatives agrees with that of two-sample rank tests relative to the t-test. Furthermore, when $\sigma_i = \sigma$ (unknown) for $i = 1, \ldots, k$, the proposed tests are asymptotically power-equivalent to Kruskal-Wallis type tests, which are used for testing the homogeneity of the means in k samples with an equal variance. As a result of a simulation study of 10,000 repetitions, surprisingly it is verified that the χ^2-approximate for the null distribution of the proposed test statistics is better than that of the Kruskal-Wallis test statistics and the Welch test.

Next when H_0 holds, we discuss R-estimators (REC's) for the common mean $\nu = \mu_i$ ($i = 1, \ldots, k$) by the asymptotic theory and a simulation study. Furthermore, we may estimate $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_k)'$ by utilizing a k dimensional statistic, whose i-th element is Hodges-Lehmann's one-sample R-estimator (1963) made of the observations $X_{i1}, \ldots, X_{in_i}$ in the i-th sample. We refer to the statistic as usual R-estimator (URE) of $\boldsymbol{\mu}$. So by the URE, we construct a positive-part shrinkage R-estimator (PSRE), based on the modified James-Stein rule stated in Stein (1966). We also consider a shrinkage R-estimator (SRE), based on the usual James-Stein rule (1961), and a preliminary test R-estimator (PTRE) taking the REC derived under the restriction of H_0 or the unrestrained URE of $\boldsymbol{\mu}$ according as the rank test leads to the acceptance or rejection of H_0. We derive asymptotic risks, based on an arbitrary quadratic loss, of the URE, PSRE, SRE and PTRE under a contiguous sequence of local alternatives. Under a Mahalanobis loss, we may show that the PSRE dominates the other three type R-estimators for $k \geq 4$. A simulation study leads strong support to the claim that the PSRE dominates the other type R-estimators and least squares estimators for any underlying distributions and that the PSRE is robust relative to outliers.

2. Statistical procedures

For real θ, let $X_{ij}(\theta) = X_{ij} - \theta$ and let $R_{ij}^+(\theta)$ be the rank of $|X_{ij}(\theta)|$ among $\{|X_{ij}(\theta)| : j = 1, \dots, n_i\}$. Then let us put aligned signed rank statistic

$$S_i(\theta) = \sum_{j=1}^{n_i} \mathrm{sign}\{X_{ij}(\theta)\} a_{n_i}^+(R_{ij}^+(\theta))/\sqrt{N}, \tag{2.1}$$

where $\mathrm{sign}(t) = 1$ for $t > 0; = 0$ for $t = 0; = -1$ otherwise, $a_n^+(m) = E\{\psi(U_n^{(m)})\}$ or $\psi(m/(n+1))$, $\psi(\cdot)$ is a nonnegative nondecreasing function defined on (0, 1), and $N = \sum_{i=1}^k n_i$. $\psi(u) = u$, $\Phi^{-1}(1/2 + u/2)$ are respectively corresponding to $a_n^+(m) = m/(n+1)$ (Wilcoxon score), $\Phi^{-1}(1/2 + m/\{2(n+1)\})$ (normal score), where $\Phi(x)$ is a standard normal distribution function. Then we propose to reject H_0 when the following ST is too large.

$$ST = N \sum_{i=1}^k [\{T_i(\tilde{\nu}_N)\}^2 / \{n_i \hat{\sigma}_i^2(\psi)\}], \tag{2.2}$$

where $T_i(\theta) = S_i(\theta) - \{n_i/(N\tilde{\sigma}_i)\} \sum_{j=1}^k \{S_j(\theta)/\tilde{\sigma}_j\} / \sum_{j=1}^k \{n_j/(N\tilde{\sigma}_j^2)\}$, $\tilde{\nu}_N$ is a consistent estimator of common mean ν such that $\tilde{\nu}_N - \nu = O_p(1/\sqrt{N})$ under H_0, $\tilde{\sigma}_i = \sqrt{\tilde{\sigma}_i^2}$, $\tilde{\sigma}_i^2$ is a consistent estimator of σ_i^2 such that $\tilde{\sigma}_i^2 - c_0\sigma_i^2 = O_p(1/\sqrt{n_i})$ and $\hat{\sigma}_i^2(\psi) = \sum_{l=1}^{n_i} \{a_{n_i}^+(l)\}^2 / n_i$, where c_0 is a constant not depending on i's. We may take as robust choices of $\tilde{\nu}_N$ and $\tilde{\sigma}_i$, respectively, an R-estimator $\hat{\nu}_N$ of the common mean (REC) defined afterwards and the sample mean deviation of the i-th sample, i.e., $\hat{\sigma}_i = \sum_{j=1}^{n_i} |X_{ij} - \bar{X}_{i\cdot}|/n_i$. A one-sample R-estimator for μ_i due to Hodges and Lehmann (1963) is given by

$$\hat{\mu}_i = [\inf\{\theta : S_i(\theta) \leq 0\} + \sup\{\theta : S_i(\theta) \geq 0\}]/2 \qquad (i = 1, \dots, k).$$

By using $\hat{\mu}_i$'s and $\hat{\sigma}_i^2$'s, we propose as an R-estimator of the common mean ν

$$\hat{\nu}_N = \sum_{i=1}^k (n_i \hat{\mu}_i / \hat{\sigma}_i^2) \bigg/ \sum_{i=1}^k (n_i / \hat{\sigma}_i^2). \tag{2.3}$$

Although we consider $\hat{\boldsymbol{\mu}}_N = (\hat{\mu}_1, \dots, \hat{\mu}_k)'$ for an R-estimator of $\boldsymbol{\mu} = (\mu_1, \dots, \mu_k)'$, as a procedure improved on $\hat{\boldsymbol{\mu}}_N$, by using $\hat{\boldsymbol{\mu}}_N$, $\hat{\nu}_N$ and ST, we propose a positive-part shrinkage R-estimator:

$$\hat{\boldsymbol{\mu}}_N^{PS} = \hat{\nu}_N \cdot \mathbf{1}_k + \max\{1 - c/ST, 0\} \cdot \hat{\boldsymbol{\alpha}}_N, \tag{2.4}$$

along with a modified James-Stein rule, where

$$\hat{\boldsymbol{\alpha}}_N = \hat{\boldsymbol{\mu}}_N - \hat{\nu}_N \cdot \mathbf{1}_k. \tag{2.5}$$

3. Tests

3.1 *Asymptotics*

We impose the following conditions to discuss the asymptotic theory.

(C.1) $F(x)$ has a positive and finite Fisher information with respect to location, i.e.,

$$0 < \int_{-\infty}^{\infty} \{-f'(x)/f(x)\}^2 dF(x) < +\infty.$$

(C.2) $\lim_{N\to\infty}(n_i/N) = \lambda_i > 0$ for $i = 1, \ldots, k$.

Then the densities $\{\prod_{i=1}^{k}\prod_{j=1}^{n_i}[(1/\sigma_i)f((x_{ij} - \nu - \Delta_i/\sqrt{N})/\sigma_i)]\}$ are contiguous to the densities $\{p_N(\boldsymbol{x}) = \prod_{i=1}^{k}\prod_{j=1}^{n_i}[(1/\sigma_i)f((x_{ij} - \nu)/\sigma_i)]\}$ as N tends to infinity. By the contiguity, we can show the following asymptotic linearity

$$\lim_{N\to\infty} P\left\{ \sup_{|\zeta|<C_1, |\Delta_i|<C_2} |S_i(\nu + \zeta/\sqrt{N} + \Delta_i/\sqrt{N}) - S_i(\nu) \right. \qquad \tag{3.1}$$
$$\left. + \lambda_i\eta(\psi, \phi^+)(\zeta + \Delta_i)/\sigma_i| > \epsilon \right\} = 0$$

for $\epsilon > 0$ and $C_1, C_2 > 0$, where $(X_{11}, \ldots, X_{kn_k})$ has the joint density $p_N(\boldsymbol{x})$, $\eta(\psi, \phi^+) = -\int_{-\infty}^{\infty}\psi(u)\phi^+(u,f)du$, $\phi^+(u,f) = -f'(F^{-1}((u+1)/2))/f(F^{-1}((u+1)/2))$ and $F^{-1}(u) = \inf\{x : u \le F(x)\}$.

THEOREM 3.1. *Assume that* (C.1) *and* (C.2) *are satisfied. Then under the null hypothesis* H_0, ST *as* $N \to \infty$ *has asymptotically a* χ^2*-distribution with* $k-1$ *degrees of freedom.*

PROOF. From (3.1), we get, under $\{p_N(\boldsymbol{x})\}$,

$$T_i(\tilde{\nu}_N) = T_i(\nu) + o_p(1). \tag{3.2}$$

Let us put $\boldsymbol{T}(\theta) = (T_1(\theta), \ldots, T_k(\theta))'$. Then using the central limit theorem and Cramer-Wold technique, we find that, under H_0, $\boldsymbol{T}(\nu) \xrightarrow{L} N_k(\boldsymbol{0}, \sigma^2(\psi)\Lambda)$, where $\xrightarrow{L}$ denotes convergence in law, $N_k(\boldsymbol{\mu}^*, \Sigma^*)$ denotes the k-variate normal distribution with mean $\boldsymbol{\mu}^*$ and variance-covariance matrix Σ^*, $\sigma^2(\psi) = \int_0^1 \psi^2(u)du$, $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_k) - \boldsymbol{\beta}\boldsymbol{\beta}'/b$, $\boldsymbol{\beta} = (\lambda_1/\sigma_1, \ldots, \lambda_k/\sigma_k)'$, and

$$b = \sum_{i=1}^{k}\{\lambda_k/\sigma_k^2\}. \tag{3.3}$$

From (3.2), it follows that $\boldsymbol{T}(\tilde{\nu}_N) \xrightarrow{L} N_k(\boldsymbol{0}, \sigma^2(\psi)\Lambda)$. $\hat{\Lambda}^- = \mathrm{diag}(N/n_1, \ldots, N/n_k)$ converges to a generalized inverse of Λ in probability and $\hat{\sigma}_i^2(\psi) \xrightarrow{P} \sigma^2(\psi)$, where $\xrightarrow{P}$ denotes convergence in probability. Hence from Theorem 1.4.1 of Muirhead (1982), $ST = \{\boldsymbol{T}(\tilde{\nu}_N)\}' \mathrm{diag}(1/\hat{\sigma}_1^2(\psi), \ldots, 1/\hat{\sigma}_k^2(\psi))\hat{\Lambda}^- \boldsymbol{T}(\tilde{\nu}_N)$ has asymptotically a χ^2-distribution. Furthermore, the degree is equal to the rank of Λ which is $k-1$. Therefore the conclusion is shown. □

Next we consider the local alternatives

$$A_N : \mu_i = \nu + \Delta_i/\sqrt{N}. \tag{3.4}$$

THEOREM 3.2. *Under the assumptions of Theorem* 3.1 *and under* A_N, ST *as* $N \to \infty$ *has asymptotically a noncentral* χ^2*-distribution with* $k-1$ *degrees of freedom and a noncentrality parameter* δ^2, *where*

$$\delta^2 = \xi^2(\psi, \phi^+) \sum_{i=1}^{k} \{\lambda_i(\Delta_i - \bar{\bar{\Delta}})^2/\sigma_i^2\}, \tag{3.5}$$

$\xi^2(\psi, \phi^+) = \{\eta(\psi, \phi^+)\}^2/\sigma^2(\psi)$ *and* $\bar{\bar{\Delta}} = \sum_{i=1}^{k}(\lambda_i\Delta_i/\sigma_i^2)/\sum_{i=1}^{k}(\lambda_i/\sigma_i^2)$.

PROOF. Let

$$TL_i(\tilde{\nu}_N, \boldsymbol{\Delta}) = S_i(\tilde{\nu}_N + \Delta_i/\sqrt{N}) - \{n_i/(N\tilde{\sigma}_i)\} \sum_{j=1}^{k} \{S_j(\tilde{\nu}_N + \Delta_j/\sqrt{N})/\tilde{\sigma}_j\} \Big/ \sum_{j=1}^{k} \{n_j/(N\tilde{\sigma}_j^2)\},$$

where $\boldsymbol{\Delta} = (\Delta_1, \ldots, \Delta_k)$. Then since $TL_i(\tilde{\nu}_N, \mathbf{0}) = T_i(\tilde{\nu}_N)$, we note that $TL_i(\tilde{\nu}_N, \boldsymbol{\Delta})$ is an extension of $T_i(\tilde{\nu}_N)$. From (3.1), we get, under $\{p_N(\boldsymbol{x})\}$,

$$TL_i(\tilde{\nu}_N, \boldsymbol{\Delta}) = T_i(\nu) - \lambda_i\eta(\psi, \phi^+)(\Delta_i - \bar{\bar{\Delta}})/\sigma_i + o_p(1). \tag{3.6}$$

It follows that, under A_N, $\boldsymbol{T}(\tilde{\nu}_N) \xrightarrow{L} N_k(\boldsymbol{\mu}_0, \sigma^2(\psi)\Lambda)$, where $\boldsymbol{\mu}_0 = \eta(\psi, \phi^+) \cdot (\lambda_1(\Delta_1 - \bar{\bar{\Delta}})/\sigma_1, \ldots, \lambda_k(\Delta_k - \bar{\bar{\Delta}})/\sigma_k)'$. The rest of the proof is parallel to the proof of Theorem 3.1 and the noncentrality parameter is given by $\boldsymbol{\mu}_0'\Lambda^-\boldsymbol{\mu}_0/\sigma^2(\psi) = \delta^2$. □

If $F(x)$ is normal, we may reject H_0 when the following statistic having an asymptotic χ^2-distribution is too large.

$$NT = \sum_{i=1}^{k} \{n_i(\bar{X}_{i\cdot} - \bar{\bar{X}}_{\cdot\cdot})^2/\tilde{\sigma}_{0i}^2\},$$

where

$$\bar{\bar{X}}_{\cdot\cdot} = \sum_{i=1}^{k}(n_i\bar{X}_{i\cdot}/\tilde{\sigma}_{0i}^2) \Big/ \sum_{i=1}^{k}(n_i/\tilde{\sigma}_{0i}^2) \tag{3.7}$$

Table 1. The upper 5% points of the χ^2-distribution and the estimated critical points of level 0.05 for the Kruskal-Wallis test.

k	2	3	5
Upper 5% points of the χ^2-distribution with $(k-1)$ degrees of freedom	3.84	5.99	9.49
Critical points of the Kruskal-Wallis test	3.98	5.80	9.08

Table 2. The estimated critical points for the ST-test of level 0.05 based on Wilcoxon score and Welch-type test.

k	Variances	$F(x)$	Critical points	
			ST-test	Welch-type test
2	$\sigma_1^2 = 1.0$, $\sigma_2^2 = 5.0$	Normal	3.88	4.92
		Logistic	3.91	4.77
		Double exponential	3.87	4.60
		Contaminated normal	3.88	4.94
3	$\sigma_1^2 = 1.0$, $\sigma_2^2 = 3.0$,	Normal	6.02	8.23
	$\sigma_3^2 = 5.0$	Logistic	5.94	7.69
		Double exponential	5.86	7.28
		Contaminated normal	5.97	8.24
5	$\sigma_1^2 = 1.0$, $\sigma_2^2 = 2.0$,	Normal	9.40	13.45
	$\sigma_3^2 = 3.0$, $\sigma_4^2 = 4.0$,	Logistic	9.31	13.14
	$\sigma_5^2 = 5.0$	Double exponential	9.25	12.13
		Contaminated normal	9.41	13.25

and $\tilde{\sigma}_{0i}^2 = \sum_{j=1}^{n_i}(X_{ij} - \bar{X}_{i\cdot})^2/(n_i - 1)$. When $k = 2$, the statistic is reduced to $NT = (\bar{X}_{1\cdot} - \bar{X}_{2\cdot})^2/(\tilde{\sigma}_{01}^2/n_1 + \tilde{\sigma}_{02}^2/n_2)$. The test based on NT for $k = 2$ is equivalent to Welch's test (1949). Welch (1951) also proposed a test statistic having approximately F-distribution under H_0. However, as the degree of freedom of the F-distribution depends on the data, the critical point is not stable. It is found that the asymptotic relative efficiency (ARE) of the rank test based on ST relative to the Welch-type test based on NT under A_N is given by $\xi^2(\psi, \phi^+)$. It is equal to the classical ARE-result of two-sample rank tests relative to the t-test. When $\sigma_i^2 = \sigma^2$ (unknown) for $i = 1, \ldots, k$ can be assumed, Kruskal-Wallis type tests based on general scores are utilized. The asymptotic theory of the test statistics was done by Puri (1964) and was reviewed by Hájek and Šidák (1967). From Theorem 3.2 and Puri (1964), under the local alternatives $A_N : \mu_i = \nu + \Delta_i/\sqrt{N}$ and $\sigma_i^2 = \sigma^2$ (unknown) for $i = 1, \ldots, k$, the ST-tests are asymptotically power-equivalent to the Kruskal-Wallis type tests.

3.2 *Simulation results*

In this section, the results of a simulation study for the goodness of χ^2-approximate of significance points for the proposed test based on Wilcoxon score ($a_n^+(m) = m/(n+1)$) and for the Welch-type test are presented. We limited attention to $k = 2$, 3, 5. Table 1 presents the upper 5 percent points for the χ^2-distribution and significance points of Kruskal-Wallis test for level 0.05 when all the sample sizes are equal to 8, i.e., $n_1 = \cdots = n_k = 8$, and $\sigma_i^2 = \sigma^2$ (unknown) for $i = 1, \dots, k$. The critical points were estimated by a simulation study of 10,000 replicates. On the other hand, the critical points of the proposed tests and the Welch-type test were given in Table 2 for the four error distributions $F =$ normal, logistic, double exponential, contaminated normal distribution: $0.9\Phi(x) + 0.1\Phi(x/3)$. $\hat{\nu}_N$ defined by (2.3), based on Wilcoxon score, was used as $\tilde{\nu}_N$ in the statistic (2.2). The sample mean deviation $\hat{\sigma}_i^2$ was used as $\tilde{\sigma}_i^2$. Although the statistics of the proposed test, Kruskal-Wallis test and Welch-type test have asymptotically the same χ^2-distribution, from Tables 1 and 2, it can be seen that the χ^2-approximate for the critical points of the proposed test is better than that of the Kruskal-Wallis test and the Welch-type test.

4. *R*-estimation for common mean

When $F(x)$ is normal and $k = 2$, Graybill and Deal (1959) considered the following weighted least squares estimator $\bar{\bar{X}}_{\cdot\cdot}$ defined by (3.7) for ν. Graybill and Deal (1959) showed that $\bar{\bar{X}}_{\cdot\cdot}$ improves on the i-th sample mean $\bar{X}_{i\cdot}$ in the light of a mean squared error under a normal underlying distribution. However, we may conjecture that $\bar{\bar{X}}_{\cdot\cdot}$ is not robust. As a matter of fact, the ARE of the proposed estimator $\hat{\nu}_N$ based on Wilcoxon score relative to $\bar{\bar{X}}_{\cdot\cdot}$ agrees with the ARE stated in the previous section. We define the relative risk efficiency of $\hat{\nu}_N$ relative to $\bar{\bar{X}}_{\cdot\cdot}$ by $E(\bar{\bar{X}}_{\cdot\cdot} - \nu)^2/E(\hat{\nu}_N - \nu)^2$. Then as n_i tends to infinity, the relative risk efficiency converges to the ARE under the suitable condition. We give the estimated values of the relative risk efficiency due to the simulation study of 10,000 replicates in Table 3. We limit sample sizes $n_1 = \cdots = n_k = 5, 10$ and the four error distributions $F =$ normal, logistic, double exponential, contaminated normal distribution combined with outlier: $0.9\Phi(x) + 0.1I_{[3,\infty)}(x)$. Then Table 3 shows that the proposed R-estimator is better than the weighted least squares estimator except for the underlying normal distribution.

5. *R*-estimation for means

When it is doubtful whether H_0 is true, except $\hat{\boldsymbol{\mu}}_N^{PS}$, we may consider a shrinkage R-estimator along with a James-Stein rule:

$$\hat{\boldsymbol{\mu}}_N^S = \hat{\nu}_N \cdot \mathbf{1}_k + (1 - c/ST) \cdot \hat{\boldsymbol{\alpha}}_N,$$

and a preliminary test R-estimator taken as $\hat{\nu}_N \cdot \mathbf{1}_k$ or $\hat{\boldsymbol{\mu}}_N$ according to the acceptance or rejection of H_0 after the preliminary ST test:

$$\hat{\boldsymbol{\mu}}_N^{PT} = \hat{\nu}_N \cdot \mathbf{1}_k + I(ST > \chi^2_{k-1,\gamma}) \cdot \hat{\boldsymbol{\alpha}}_N,$$

Table 3. The relative risk efficiency of R-estimator with respect to weighted least squares estimator for a common mean.

k	Variances	$F(x)$	Relative efficiency: Sample size 5	10
2	$\sigma_1^2 = 1.0$, $\sigma_2^2 = 5.0$	Normal	0.94	0.93
		Logistic	0.99	1.04
		Double exponential	1.11	1.28
		Contamination	1.07	1.20
3	$\sigma_1^2 = 1.0$, $\sigma_2^2 = 1.5$, $\sigma_3^2 = 2.0$	Normal	0.95	0.94
		Logistic	0.98	1.02
		Double exponential	1.08	1.26
		Contamination	1.10	1.30
5	$\sigma_1^2 = 1.0$, $\sigma_2^2 = 2.0$, $\sigma_3^2 = 3.0$, $\sigma_4^2 = 4.0$, $\sigma_5^2 = 5.0$	Normal	0.96	0.93
		Logistic	0.98	1.02
		Double exponential	1.07	1.25
		Contamination	1.11	1.40

where $\hat{\boldsymbol{\alpha}}_N$ is defined by (2.5), $I(B)$ denotes the indicator function of the set B and $\chi^2_{k-1,\gamma}$ is the upper 100γ-percent point of the χ^2-distribution function with $k-1$ degrees of freedom.

In the following two sections, the risks of the positive-part shrinkage R-estimator (PSRE), shrinkage R-estimator (SRE), preliminary test R-estimator (PTRE), and usual R-estimator are compared due to the asymptotic theory and a simulation study.

5.1 *Asymptotic theory*

Assume that, for an estimator $\hat{\boldsymbol{\mu}}_N^{\#}$ of the k dimensional vector $\boldsymbol{\mu}$, $F^{\#}(\boldsymbol{x}) = \lim_{N\to\infty} P\{\sqrt{N}(\hat{\boldsymbol{\mu}}_N^{\#} - \boldsymbol{\mu}) \le \boldsymbol{x} \mid A_N\}$. Then for a suitable (p.d.) matrix Q associated with the quadratic loss function $L(\hat{\boldsymbol{\mu}}_N^{\#}; \boldsymbol{\mu}) = N(\hat{\boldsymbol{\mu}}_N^{\#} - \boldsymbol{\mu})'Q(\hat{\boldsymbol{\mu}}_N^{\#} - \boldsymbol{\mu})$, the asymptotic distributional quadratic risk (ADQR) of $\hat{\boldsymbol{\mu}}_N^{\#}$ is defined as

$$\begin{aligned}(5.1)\qquad \mathcal{R}(\hat{\boldsymbol{\mu}}_N^{\#}, \boldsymbol{\mu} : Q) &= \lim_{b\to\infty}\lim_{N\to\infty} E[\min\{L(\hat{\boldsymbol{\mu}}_N^{\#}; \boldsymbol{\mu}), b\}] \\ &= \int_{R^k} \boldsymbol{x}'Q\boldsymbol{x}\,dF^{\#}(\boldsymbol{x}) = \operatorname{tr}(\Sigma^{\#}Q),\end{aligned}$$

where $\Sigma^{\#} = \int_{R^k} \boldsymbol{x}\boldsymbol{x}'dF^{\#}(\boldsymbol{x})$. Let us put $\xi^2 = \xi^2(\psi, f)$ for the simplicity of the notation throughout this section. Since $\hat{\mu}_i$ is an admissible point of $\{\theta : S_i(\theta) = \text{minimum}\}$, we may verify

$$(5.2)\qquad \sqrt{N}(\hat{\mu}_i - \mu_i)/\{\lambda_i\eta(\psi, \phi^+)\} \sim S_i(\mu_i),$$

where $V \sim W$ denotes that $V - W$ converges to 0 in probability. It follows that

$$\sqrt{N}(\hat{\boldsymbol{\mu}}_N - \boldsymbol{\mu}) \xrightarrow{L} N_k(\mathbf{0}, \Sigma_0/\xi^2), \tag{5.3}$$

where

$$\Sigma_0 = \mathrm{diag}(\sigma_1^2/\lambda_1, \ldots, \sigma_k^2/\lambda_k). \tag{5.4}$$

Then from (5.1), it can be verified that under A_N including H_0,

$$\begin{aligned} &\sqrt{N}(\hat{\nu}_N - \nu^*) \sim \mathcal{L}(\overline{\overline{Y}}), \qquad \sqrt{N}\hat{\boldsymbol{\alpha}}_N \sim \mathcal{L}(\boldsymbol{Y} - \overline{\overline{Y}}\,\mathbf{1}_k) \quad \text{and} \\ &ST \sim \mathcal{L}\left(\xi^2 \sum_{i=1}^{k} \{\lambda_i (Y_i - \overline{\overline{Y}})^2/\sigma_i^2\}\right), \end{aligned} \tag{5.5}$$

where $\nu^* = \sum_{i=1}^k (n_i \mu_i/\tilde{\sigma}_i^2)/\sum_{i=1}^k (n_i/\tilde{\sigma}_i^2)$, $V_N \sim \mathcal{L}(W)$ denotes that V_N has asymptotically a distribution of the random variable W, $\boldsymbol{Y} = (Y_1, \ldots, Y_k)'$ is defined by the random variable having $N_k(\boldsymbol{\Delta}_0, \Sigma_0/\xi^2)$, $\boldsymbol{\Delta}_0 = \boldsymbol{\Delta} - \overline{\overline{\Delta}}\,\mathbf{1}_k$, Σ_0 is defined by (5.4), $\overline{\overline{Y}} = \sum_{i=1}^k \{\lambda_i Y_i/(b\sigma_i^2)\}$ and b is defined by (3.3). Let us put $\hat{\boldsymbol{\mu}}_N^h = \hat{\nu}_N \mathbf{1}_k + \{1 - h(ST)\}\hat{\boldsymbol{\alpha}}_N$ for a real valued function $h(\cdot)$ defined on $[0, \infty)$. From (5.5), we have

$$\begin{aligned} \mathcal{R}(\hat{\boldsymbol{\mu}}_N^h, \boldsymbol{\mu} : Q) = {} & \mathrm{tr}(Q\Sigma_0)/\xi^2 - 2E\{h(U)(\boldsymbol{Y} - \overline{\overline{Y}}\,\mathbf{1}_k)' Q (\boldsymbol{Y} - \overline{\overline{Y}}\,\mathbf{1}_k - \boldsymbol{\Delta}_0)\} \\ & + E\{h^2(U)(\boldsymbol{Y} - \overline{\overline{Y}}\,\mathbf{1}_k)' Q (\boldsymbol{Y} - \overline{\overline{Y}}\,\mathbf{1}_k)\}. \end{aligned}$$

Furthermore, by using the fact that $\overline{\overline{Y}}$ is independent of $\boldsymbol{Y} - \overline{\overline{Y}}\,\mathbf{1}_k$ and U, it follows that

$$\begin{aligned} \mathcal{R}(\hat{\boldsymbol{\mu}}_N^h, \boldsymbol{\mu} : Q) = {} & \mathrm{tr}(Q\Sigma_0)/\xi^2 - 2\,\mathrm{tr}\, E[\{h(U) Q (\boldsymbol{Y} - \overline{\overline{Y}}\,\mathbf{1}_k - \boldsymbol{\Delta}_0)\boldsymbol{Y}'\}] \\ & + \mathrm{tr}[QE\{h^2(U)(\boldsymbol{Y} - \overline{\overline{Y}}\,\mathbf{1}_k)\boldsymbol{Y}'\}]. \end{aligned}$$

Since U has a noncentral χ^2-distribution with $k-1$ degrees of freedom and a noncentrality parameter δ^2,

$$\begin{aligned} E\{h(U)\} = {} & \sum_{m=0}^{\infty} \{e^{-\delta^2/2}(\delta^2/2)^m/(m!)\} \\ & \cdot \int_0^\infty [h(t) t^{(k-1)/2+m-1} e^{-t/2}/\{2^{(k-1)/2+m}\Gamma((k-1)/2+m)\}]dt. \end{aligned}$$

By taking the differential operator $\boldsymbol{d}/\boldsymbol{d\Delta}_0 = (\partial/\partial\Delta_{01}, \ldots, \partial/\partial\Delta_{0k})'$ to the above both sides, we get

$$E\{h(U)\boldsymbol{Y}\} = \boldsymbol{\Delta}_0 E\{h(\chi_{k+1}^2(\delta^2))\}. \tag{5.6}$$

Furthermore, by taking $(\boldsymbol{d}/\boldsymbol{d\Delta}_0)'$, we find

$$E\{h(U)\boldsymbol{YY}'\} = \Sigma_0 E\{h(\chi^2_{k+1}(\delta^2))\}/\xi^2 + \boldsymbol{\Delta}_0\boldsymbol{\Delta}_0' E\{h(\chi^2_{k+3}(\delta^2))\}. \tag{5.7}$$

Using the relations (5.6) and (5.7), after routine computation, we get

$$\begin{aligned}\mathcal{R}(\hat{\boldsymbol{\mu}}^h_N, \boldsymbol{\mu} : Q) = {} & \mathrm{tr}(Q\Sigma_0)/\xi^2 \\ & - \mathrm{tr}(Q\Sigma)[2E\{h(\chi^2_{k+1}(\delta^2))\} - E\{h^2(\chi^2_{k+1}(\delta^2))\}]/\xi^2 \\ & + \mathrm{tr}(Q\Delta_0\Delta_0')[2E\{h(\chi^2_{k+1}(\delta^2))\} \\ & \qquad - 2E\{h(\chi^2_{k+3}(\delta^2))\} + E\{h^2(\chi^2_{k+3}(\delta^2))\}].\end{aligned} \tag{5.8}$$

By setting $h(U) = 1 - \max\{1 - c \cdot U^{-1}, 0\}$, c/U, and $1 - I(U \geq \chi^2_{k-1,\gamma})$, we may get, respectively, $\mathcal{R}(\hat{\boldsymbol{\mu}}^{PS}_N, \boldsymbol{\mu} : Q)$, $\mathcal{R}(\hat{\boldsymbol{\mu}}^{S}_N, \boldsymbol{\mu} : Q)$, and $\mathcal{R}(\hat{\boldsymbol{\mu}}^{PT}_N, \boldsymbol{\mu} : Q)$. Then in a way similar to Shiraishi (1991), under general Q and under A_N, we may show the asymptotic dominance of the SRE over the URE. Also under general Q and under H_0, we may show the asymptotic dominance of the PSRE over the SRE, PTRE and URE. However, using the same discussion as in Theorem 4.3 of Sen and Saleh (1987), we can verify that neither of the SRE nor PTRE dominate each other. We may not show the dominance of the PSRE over the other R-estimators for arbitrary Q. Hence we consider the loss $L(\hat{\boldsymbol{\mu}}^{\#}_N; \boldsymbol{\mu}) = \xi^2 N(\hat{\boldsymbol{\mu}}^{\#}_N - \boldsymbol{\mu})'\Sigma_0^{-1}(\hat{\boldsymbol{\mu}}^{\#}_N - \boldsymbol{\mu})$, i.e., we choose $Q = \xi^2\Sigma_0^{-1}$. Then (5.8) becomes

$$\begin{aligned}\mathcal{R}(\hat{\boldsymbol{\mu}}^h_N, \boldsymbol{\mu} : \xi^2\Sigma_0^{-1}) = {} & k - (k-1)[2E\{h(\chi^2_{k+1}(\delta^2))\} - E\{h^2(\chi^2_{k+1}(\delta^2))\}] \\ & + \delta^2[2E\{h(\chi^2_{k+1}(\delta^2))\} \\ & \qquad - 2E\{h(\chi^2_{k+3}(\delta^2))\} + E\{h^2(\chi^2_{k+3}(\delta^2))\}].\end{aligned}$$

Hence in a way similar to the proof of Theorem 5.4 of Shiraishi (1991), we get

THEOREM 5.1. *Suppose that $k \geq 4$.*

(i) *If $c = \chi^2_{k-1,\gamma}$ is chosen as a shrinkage constant and $0 < c \leq 2(k-3)$ is satisfied, $\mathcal{R}(\hat{\boldsymbol{\mu}}^{PS}_n, \boldsymbol{\mu} : \xi^2\Sigma_0^{-1}) < \mathcal{R}(\hat{\boldsymbol{\mu}}^{PT}_n, \boldsymbol{\mu} : \xi^2\Sigma_0^{-1})$ for all $\delta^2 \geq 0$.*

(ii) *$\mathcal{R}(\hat{\boldsymbol{\mu}}^{PS}_n, \boldsymbol{\mu} : \xi^2\Sigma_0^{-1}) < \mathcal{R}(\hat{\boldsymbol{\mu}}^{S}_n, \boldsymbol{\mu} : \xi^2\Sigma_0^{-1})$ for all $\delta^2 \geq 0$ and $c > 0$.*

5.2 *Relative risks*

The least squares estimator for $\boldsymbol{\mu}$ is given by $\tilde{\boldsymbol{\mu}}_N = (\bar{X}_{1\cdot}, \ldots, \bar{X}_{k\cdot})'$. The covariance matrix of $\tilde{\boldsymbol{\mu}}_N$ is $\Sigma_N = \mathrm{diag}(\sigma_1^2/n_1, \ldots, \sigma_k^2/n_k)$. Hence it is plausible that $(\hat{\boldsymbol{\mu}}^{\#}_N - \boldsymbol{\mu})'\Sigma_N^{-1}(\hat{\boldsymbol{\mu}}^{\#}_N - \boldsymbol{\mu})$ is taken as a loss for an estimator $\hat{\boldsymbol{\mu}}^{\#}_N$ of $\boldsymbol{\mu}$. Then the risk of $\hat{\boldsymbol{\mu}}^{\#}_N$ is given by $\mathcal{R}_N(\hat{\boldsymbol{\mu}}^{\#}_N, \boldsymbol{\mu}) = E\{(\hat{\boldsymbol{\mu}}^{\#}_N - \boldsymbol{\mu})'\Sigma_N^{-1}(\hat{\boldsymbol{\mu}}^{\#}_N - \boldsymbol{\mu})\}$. Under a suitable condition, we may find

$$\lim_{N\to\infty} \mathcal{R}_N(\hat{\boldsymbol{\mu}}^{\#}_N, \boldsymbol{\mu}) = \mathcal{R}(\hat{\boldsymbol{\mu}}^{\#}_N, \boldsymbol{\mu} : \Sigma_0^{-1}). \tag{5.9}$$

We define the relative risk efficiency of $\hat{\boldsymbol{\mu}}^{\#}_N$ with respect to the usual R-estimator $\hat{\boldsymbol{\mu}}_N$ by $\mathcal{R}_N(\hat{\boldsymbol{\mu}}_N, \boldsymbol{\mu})/\mathcal{R}_N(\hat{\boldsymbol{\mu}}^{\#}_N, \boldsymbol{\mu})$, which is denoted by $\mathrm{RRE}(\hat{\boldsymbol{\mu}}^{\#}_N, \hat{\boldsymbol{\mu}}_N)$. If $\mathrm{RRE}(\hat{\boldsymbol{\mu}}^{\#}_N,$

$\hat{\boldsymbol{\mu}}_N) > 1$ (< 1), $\hat{\boldsymbol{\mu}}_N^{\#}$ is better (worse) than $\hat{\boldsymbol{\mu}}_N$. From (5.9), it is natural that the asymptotic relative risk efficiency of $\hat{\boldsymbol{\mu}}_N^{\#}$ with respect to $\hat{\boldsymbol{\mu}}_N$ is defined by $\text{ARRE}(\hat{\boldsymbol{\mu}}_N^{\#}, \hat{\boldsymbol{\mu}}_N) = \mathcal{R}(\hat{\boldsymbol{\mu}}_N, \boldsymbol{\mu} : \Sigma_0^{-1})/\mathcal{R}(\hat{\boldsymbol{\mu}}_N^{\#}, \boldsymbol{\mu} : \Sigma_0^{-1})$. We consider the four-type R-estimators based on Wilcoxon score. The values of the risks for $\hat{\boldsymbol{\mu}}_N^{PS}$, $\hat{\boldsymbol{\mu}}_N^{S}$, $\hat{\boldsymbol{\mu}}_N^{PT}$, $\hat{\boldsymbol{\mu}}_N$, and $\tilde{\boldsymbol{\mu}}_N$ were estimated due to the Monte Carlo simulation. We limited attention to $k = 4$, $n_1 = n_2 = n_3 = n_4 = 5$, 10, 15, $(\sigma_1^2, \sigma_2^2, \sigma_3^2, \sigma_4^2) = (1.0, 2.0, 3.0, 5.0)$, and $(\mu_1, \mu_2, \mu_3, \mu_4) = \sqrt{5/n_1} \cdot (\sigma_1^2\Delta, \sigma_2^2\Delta, \sigma_3^2\Delta, \sigma_4^2\Delta)$ such that $\Delta = 0.0(0.04)2.0$. Also from Theorem 5.1, $c = \chi^2_{k-1,\gamma} = k - 3 = 1$ is taken as a shrinkage constant and a critical point of the preliminary test. The underlying distribution chosen here are normal, logistic, double exponential, and mixture of the normal and outlier: $0.9\Phi(x) + 0.1I_{[3,\infty)}(x)$. For each setting, 10,000 repetitions are used. By the way, since $ST \fallingdotseq 0$ and $\hat{\boldsymbol{\alpha}}_N \neq \mathbf{0}$ often happened in the values of statistics computed from pseudo random numbers, the loss of $\hat{\boldsymbol{\mu}}_N^{S}$, $(\hat{\boldsymbol{\mu}}_N^{S} - \boldsymbol{\mu})'\Sigma_N^{-1}(\hat{\boldsymbol{\mu}}_N^{S} - \boldsymbol{\mu})$, often became too large. Hence, we used $\hat{\nu}_N \cdot \mathbf{1}_k + \max\{(1 - c/ST), -70\} \cdot \hat{\boldsymbol{\alpha}}_N$ as $\hat{\boldsymbol{\mu}}_N^{S}$ instead of $\hat{\nu}_N \cdot \mathbf{1}_k + (1 - c/ST) \cdot \hat{\boldsymbol{\alpha}}_N$. Based on the estimated values of the risks, we can calculate the values for the RRE's among the estimators. Especially $\text{RRE}(\hat{\boldsymbol{\mu}}_N, \tilde{\boldsymbol{\mu}}_N)$ does not depend on $\boldsymbol{\Delta}$. Also $\text{ARRE}(\hat{\boldsymbol{\mu}}_N, \tilde{\boldsymbol{\mu}}_N)$ agrees with the ARE(ST, NT) stated in Subsection 3.1 when the underlying distribution is symmetric. Hence $\text{RRE}(\hat{\boldsymbol{\mu}}_N, \tilde{\boldsymbol{\mu}}_N)$ was given in Table 4. From Table 4, it can be seen that (i) $\hat{\boldsymbol{\mu}}_N$ is slightly worse than $\tilde{\boldsymbol{\mu}}_N$ for a normal underlying distribution, (ii) the former is better than the latter for the other underlying distributions, and (iii) $\text{RRE}(\hat{\boldsymbol{\mu}}_N, \tilde{\boldsymbol{\mu}}_N)$ increases in the sample size.

Table 4. The relative risk efficiency of the R-estimator with respect to the least squares estimator.

Distribution	Sample size			
	5	10	15	∞
Normal	0.93	0.93	0.94	0.95
Logistic	1.01	1.05	1.07	1.10
Double exponential	1.15	1.29	1.34	1.50
Mixture of the normal and outlier	1.02	1.09	1.17	—

In Figs. 1.1–1.4, we drew the graphical pictures of $\text{RRE}(\hat{\boldsymbol{\mu}}_N^{PS}, \hat{\boldsymbol{\mu}}_N)$, $\text{RRE}(\hat{\boldsymbol{\mu}}_N^{S}, \hat{\boldsymbol{\mu}}_N)$, $\text{RRE}(\hat{\boldsymbol{\mu}}_N^{PT}, \hat{\boldsymbol{\mu}}_N)$, and $\text{RRE}(\tilde{\boldsymbol{\mu}}_N, \hat{\boldsymbol{\mu}}_N)$ for $n_1 = 10$ and the respective distributions. The graphical pictures for $n_1 = 5$, 15 are similar to Figs. 1.1–1.4. In Fig. 2, we drew the picture of $\text{ARRE}(\hat{\boldsymbol{\mu}}_N^{PS}, \hat{\boldsymbol{\mu}}_N)$, $\text{ARRE}(\hat{\boldsymbol{\mu}}_N^{S}, \hat{\boldsymbol{\mu}}_N)$, and $\text{ARRE}(\hat{\boldsymbol{\mu}}_N^{PT}, \hat{\boldsymbol{\mu}}_N)$. These ARRE's depend on underlying distribution only through the noncentral parameter δ^2. The following conclusions are drawn from Figs. 1.1–1.4 and 2:

1. The PSRE dominates the other R-estimators even for small sample sizes.
2. The dominance of the PSRE over the other R-estimators decreases in Δ.
3. Except the case that the underlying distribution is a mixture of the normal and outlier, the RRE's approach to 1 as Δ tends to 2.

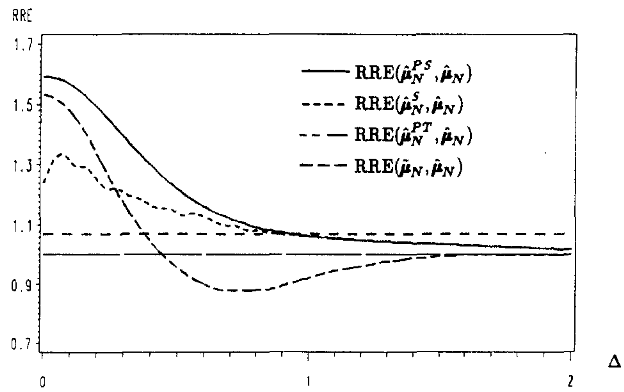

Fig. 1.1. The relative risk efficiency of $\hat{\mu}_N^{PS}$, $\hat{\mu}_N^{S}$, $\hat{\mu}_N^{PT}$, and $\tilde{\mu}_N$ with respect to $\hat{\mu}_N$ for F = Normal.

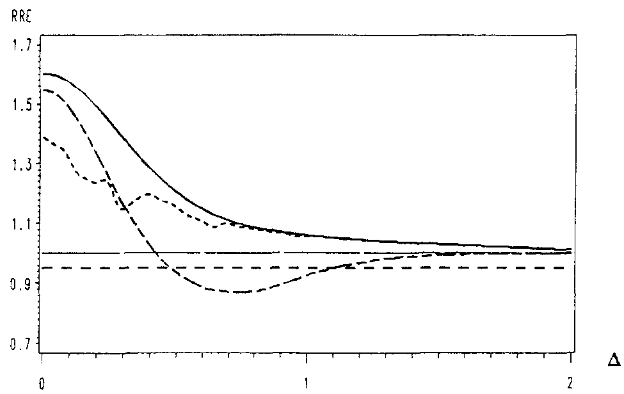

Fig. 1.2. The relative risk efficiency of $\hat{\mu}_N^{PS}$, $\hat{\mu}_N^{S}$, $\hat{\mu}_N^{PT}$, and $\tilde{\mu}_N$ with respect to $\hat{\mu}_N$ for F = Logistic.

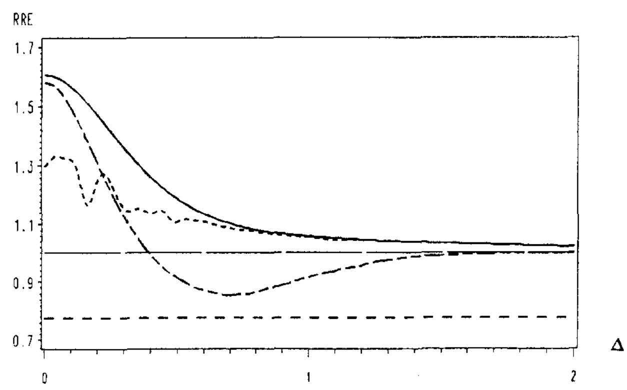

Fig. 1.3. The relative risk efficiency of $\hat{\mu}_N^{PS}$, $\hat{\mu}_N^{S}$, $\hat{\mu}_N^{PT}$, and $\tilde{\mu}_N$ with respect to $\hat{\mu}_N$ for F = Double exponential.

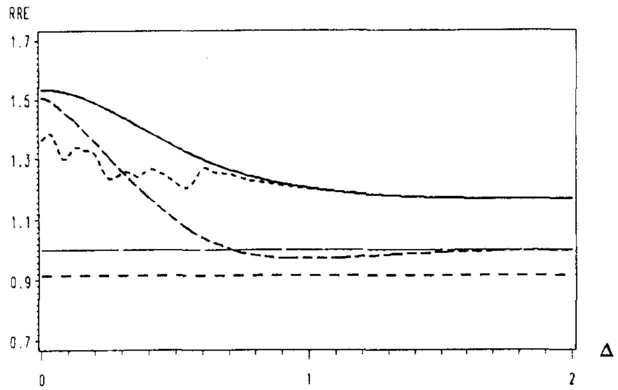

Fig. 1.4. The relative risk efficiency of $\hat{\mu}_N^{PS}$, $\hat{\mu}_N^{S}$, $\hat{\mu}_N^{PT}$, and $\tilde{\mu}_N$ with respect to $\hat{\mu}_N$ for F = Mixture of the normal and outlier.

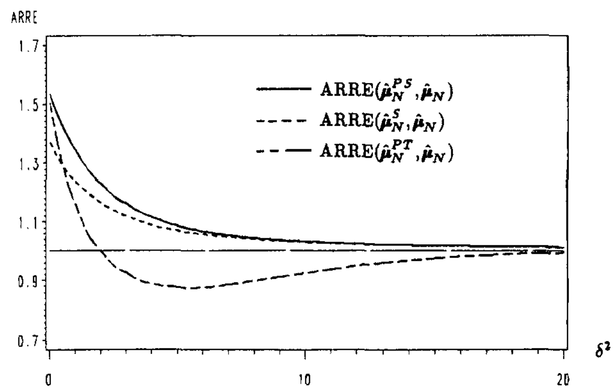

Fig. 2. The asymptotic relative risk efficiency of $\hat{\mu}_N^{PS}$, $\hat{\mu}_N^{S}$, and $\hat{\mu}_N^{PT}$ with respect to $\hat{\mu}_N$.

4. When the underlying distribution is a mixture of the normal and outlier, since the values of $\mathrm{RRE}(\hat{\mu}_N^{PS}, \hat{\mu}_N)$ are larger than 1.17, the PSRE is greatly effective for the outliers.

5. The SRE is not stable since the test statistic $ST \fallingdotseq 0$ often happens.

6. $\mathrm{RRE}(\hat{\mu}_N^{PT}, \hat{\mu}_N)$ is larger than 1 for small Δ, while it is smaller than 1 for $\Delta \geq 0.72$.

7. Since $\mathrm{RRE}(\hat{\mu}_N^{PS}, \hat{\mu}_N)$ decreases in the sample size n_i, the PSRE is more effective for small sample sizes.

Acknowledgements

The author would like to thank the referees for their valuable comments.

References

Chernoff, H. and Savage, I. R. (1958). Asymptotic normality and efficiency of certain nonparametric test statistics, *Ann. Math. Statist.*, **29**, 972–994.

Graybill, F. A. and Deal, R. B. (1959). Combining unbiased estimators, *Biometrics*, **15**, 543–550.

Hájek, J. and Šidák, Z. (1967). *Theory of Rank Tests*, Wiley, New York.

Hodges, J. L., Jr. and Lehmann, E. L. (1956). The efficiency of some nonparametric competitors of the t-test, *Ann. Math. Statist.*, **27**, 324–335.

Hodges, J. L., Jr. and Lehmann, E. L. (1963). Estimates of location based on rank tests, *Ann. Math. Statist.*, **34**, 598–611.

James, W. and Stein, C. (1961). Estimation with quadratic loss, *Proc. 4th Berkeley Symp. on Math. Statist. Prob.*, Vol. 1, 361–379, University of California Press, Berkeley.

Kruskal, W. H. and Wallis, W. A. (1952). Use of ranks in one criterion variance analysis, *J. Amer. Statist. Assoc.*, **57**, 583–621.

Muirhead, R. J. (1982). *Aspects of Multivariate Statistical Theory*, Wiley, New York.

Pratt, J. W. (1964). Robustness of some procedures for the two-sample location problem, *J. Amer. Statist. Assoc.*, **59**, 665–680.

Puri, M. L. (1964). Asymptotic efficiency of a class of c-sample tests, *Ann. Math. Statist.*, **35**, 102–121.

Sclove, S. L., Morris, C. and Radhakrishnan, R. (1972). Non-optimality of preliminary test estimators for the mean of a multi-variate normal distribution *Ann. Math. Statist.*, **43**, 1481–1490.

Sen, P. K. and Saleh, K. Md. E. (1987). On preliminary test and shrinkage M-estimation in linear models, *Ann. Statist.*, **15**, 1580–1592.

Shiraishi, T. (1991). Hypotheses testing and parameter estimation based on M-statistics in k samples with unequal variances, *Metrika*, **38**, 163–178.

Stein, C. (1966). An approach to the recovery of interblock information in balanced incomplete block designs, *Research Papers in Statistics: Festschrift for J. Neyman*, (ed. F. N. David), 351–366, Wiley, New York.

Welch, B. L. (1949). Further notes on Mrs. Aspin's tables, *Biometrika*, **36**, 243–246.

Welch, B. L. (1951). On the comparison of several mean values: an alternative approach, *Biometrika*, **38**, 330–336.