Asymptotic Efficiency of Two Nonparametric Competitors of Wilcoxon's Two Sample Test

Myles Hollander

February 1, 1967

Technical Report Number 12
NASA Grant Number NGR-10-004-029

Department of Statistics
Florida State University
Tallahassee, Florida

ASYMPTOTIC EFFICIENCY OF TWO NONPARAMETRIC COMPETITORS OF WILCOXON'S TWO SAMPLE TEST[1]

Myles Hollander[2]
The Florida State University

ABSTRACT

Wilcoxon's signed rank test and a test based on the uniform minimum variance unbiased estimator of $P(X_1+ X_2 < Y_1+ Y_2)$ are considered as competitors of the Mann-Whitney-Wilcoxon (U) test. The criteria used to compare the tests are Bahadur and Pitman efficiency. For pure translation alternatives U is superior, but both tests compare favorably with respect to U for certain contamination alternatives.

1. INTRODUCTION

Let $X_1,\ldots,X_m$ be independent and identically distributed according to F_1 and $Y_1,\ldots,Y_n$ be independent and identically distributed according to F_2 where F_1,F_2 are assumed continuous. The excellent properties of the Mann-Whitney-Wilcoxon U statistic ([12], [7]) for testing H_0: $F_1= F_2= F$ against translation alternatives (3.1) are well known (e.g. see [3], [4]). These properties may be attributed, in part, to the fact that U/mn (2.1) is the uniform minimum variance unbiased estimator of $P(X_1 < Y_1)$, a result given by Lehmann [5]. This suggests the investigation of tests based on statistics which, when suitably scaled, are consistent estimators of the related parameter $P(X_1+ X_2 < Y_1+ Y_2)$.

---

[1] Respectfully dedicated to the memory of Frank Wilcoxon.

[2] This research was supported by the National Institutes of Health under Grant 17-946 and by NASA Grant No. NGR-10-004-029.

In this paper we concern ourselves with two statistics having this property.

The W statistic (2.3) is Wilcoxon's signed rank statistic [12] applied to a random pairing of the X's with the Y's. The V statistic (2.5), which is the uniform minimum variance unbiased estimator of $P(X_1 + X_2 < Y_1 + Y_2)$, is defined to be the proportion of the $\binom{m}{2}\binom{n}{2}$ quadruples $(X_i, X_j; Y_k, Y_\ell)$ with $i < j$ and $k < \ell$ satisfying the inequality $X_i + X_j < Y_k + Y_\ell$.

In section 2 we show that V is not distribution-free under H_0, but an asymptotically distribution-free procedure based on V is defined.

Section 3 is devoted to efficiency comparisons of U,V, and W for translation and contamination alternatives. Although V is slightly more Pitman efficient than U for normal translation, the calculations of this section illustrate the general superiority of U to V and W for translations both near and away from H_0. However, when we consider contamination alternatives of the form $F_2(x) = (1-p) F_1(x) + pF_1(x-\theta)$, for p close to 0 and θ large our efficiency calculations favor V and W. In this section (and section 2) we also discuss the relationship of W to V.

Section 4 contrasts the use of the random-paired signed rank test W in place of U with normal theory practice where the random-paired t-test is sometimes preferred to the usual two sample t-test.

2. DEFINITIONS AND BASIC FACTS

The Mann-Whitney form of Wilcoxon's statistic is

$$U = \sum_{i=1}^{m} \sum_{j=1}^{n} \phi(X_i, Y_j) \tag{2.1}$$

where

$$\phi(a,b) = \begin{cases} 1 & \text{if } a < b \\ 0 & \text{otherwise.} \end{cases} \tag{2.2}$$

To define the W statistic, let us first take m=n and assume for simplicity (and without loss of generality) that the random pairing of the X's with the Y's results in the pairs (X_i,Y_i), $i=1,2,\ldots,n$. Let $D_i = |X_i - Y_i|$ and R_i = rank of D_i in the joint ranking from least to greatest of $[D_\alpha]_{\alpha=1}^n$. Then Wilcoxon's signed rank statistic is

$$W = \sum_{i=1}^{n} R_i \phi(X_i,Y_i). \tag{2.3}$$

If $m \neq n$ we define the W test as the one obtained by computing (2.3) after we have randomly discarded observations from the larger sample to equalize the sample sizes. In this case the "n" of (2.3) is replaced by $n^* = \min[n,m]$.

Using a representation due to Tukey [11] we may write W as

$$W = \sum_{i<j}^{n} \phi(X_i + X_j,\ Y_i + Y_j) + \sum_{i=1}^{n} \phi(X_i,Y_i). \tag{2.4}$$

Letting W' denote the first term on the right of (2.4), it is easily seen that W and W' are asymptotically equivalent test statistics and that $2W'/n(n-1)$ is an unbiased and consistent estimator of $P(X_1 + X_2 < Y_1 + Y_2)$.

The uniform minimum variance unbiased estimator of $P(X_1 + X_2 < Y_1 + Y_2)$ is

$$V = \binom{m}{2}^{-1}\binom{n}{2}^{-1} \sum_{\substack{i<j \\ k<\ell}} \phi(X_i + X_j,\ Y_k + Y_\ell). \tag{2.5}$$

This follows from a direct application of a lemma due to Lehmann and Scheffe (Lemma 3.2 of [5]). We remark that the statistic V, even when m=n, is based on more "information" than W' as the indicator function $\phi(X_i + X_j,\ Y_k + Y_\ell)$ is computed for $n^2(n-1)^2/4$ quadruples in the case of V versus $n(n-1)/2$ for W'.

Unlike U and W, V is not distribution-free under H_0. To see this we first apply Lehmann's generalized U-statistic theorem [5] to obtain

Theorem 1: If $0 < P(X_1 < Y_1) < 1$, $m=sN$, $n=(1-s)N$ with $0 < s < 1$, then $N^{\frac{1}{2}}(V-P(X_1+X_2 < Y_1+Y_2))$ has a limiting normal distribution with mean 0 and asymptotic variance $4(s^{-1}\delta_{10}+(1-s)^{-1}\delta_{01})$ where

$$\delta_{10} = E[\phi(X_1+X_2, Y_1+Y_2)\phi(X_1+X_3, Y_3+Y_4)] - E^2\phi(X_1+X_2, Y_1+Y_2) \tag{2.6}$$

and

$$\delta_{01} = E[\phi(X_1+X_2, Y_1+Y_2)\phi(X_3+X_4, Y_1+Y_3)] - E^2\phi(X_1+X_2, Y_1+Y_2). \tag{2.7}$$

Under H_0,

$$\delta_{10} = \delta_{01} = \lambda(F) - 1/4 \tag{2.8}$$

where

$$\lambda(F) = P(X_1 < X_2+X_3-X_4;\ X_1 < X_5+X_6-X_7) \tag{2.9}$$

when $X_1, X_2, \ldots, X_7$ are independent and identically distributed according to F. Lehmann [6] has obtained different values of $\lambda(F)$ for various F and thus the null distribution of V will depend on F.

In the remainder of this paper the phrase "the V test" will mean the asymptotically distribution-free procedure which treats $(V-(1/2))/\hat{\sigma}_A(V)$ as a unit normal random variable under H_0 where

$$\sigma_A^2(V) = (4\lambda(F)-1)(m^{-1}+n^{-1}) \tag{2.10}$$

and $\hat{\sigma}_A^2(V)$ is defined by replacing $\lambda(F)$ with a consistent estimate in (2.10). One such estimate, similar to one proposed by Lehmann [6] in another context, is the following. Let $Z_1, Z_2, \ldots, Z_N$ denote the combined sequence of X's and Y's and define $\hat{\lambda}(F)$ to be the relative frequency of the event $(Z_{\alpha_1} < Z_{\alpha_2} + Z_{\alpha_3} - Z_{\alpha_4};$

$Z_{\alpha_1} < Z_{\alpha_5} + Z_{\alpha_6} - Z_{\alpha_7}$). This estimate is tedious to compute and in practice only a small proportion of the total number of such simultaneous inequalities should be checked.

3. EFFICIENCIES FOR TRANSLATION AND CONTAMINATION ALTERNATIVES

We first consider translation alternatives

$$H_1: F_1(x) = F(x),\ F_2(x) = F(x-\theta),\ \theta > 0, \tag{3.1}$$

and utilize Bahadur efficiency ([1], [2]) to obtain a measure of asymptotic performance for each fixed θ.

For the efficiency calculations of this section we lose no generality in assuming $m \geq n$ and thus we write $m=sN$, $n=(1-s)N$, with $1/2 \leq s < 1$. We define

$$T_U^{(N)} = \frac{U-E_0(U)}{\sigma_0(U)} = \frac{U-(mn/2)}{(mn(m+n+1)/12)^{\frac{1}{2}}}, \tag{3.2}$$

$$T_V^{(N)} = \frac{V-E_0(V)}{\sigma_A(V)} = \frac{V-(1/2)}{[(4\lambda(F)-1)(m^{-1}+n^{-1})]^{\frac{1}{2}}}, \tag{3.3}$$

$$T_W^{(N)} = \frac{W-E_0(W)}{\sigma_0(W)} = \frac{W-(n(n+1)/4)}{(n(n+1)(2n+1)/24)^{\frac{1}{2}}}, \tag{3.4}$$

where the subscript 0 denotes that the moment is computed under H_0 and $\lambda(F)$ and $\sigma_A(V)$ are defined by (2.9) and (2.10). By using Chebychev's inequality it follows that

$$b_U(\theta) = \text{p-lim}\,\frac{T_U^{(N)}}{N^{\frac{1}{2}}} = (12s(1-s))^{\frac{1}{2}}[\textstyle\int F(x+\theta)\,dF(x)-(1/2)], \tag{3.5}$$

$$b_V(\theta) = \text{p-lim}\,\frac{T_V^{(N)}}{N^{\frac{1}{2}}} = \frac{(s(1-s))^{\frac{1}{2}}\,[\int G(x+2\theta)\,dG(x)-(1/2)]}{4\lambda(F)-1}\,, \tag{3.6}$$

$$b_W(\theta) = \text{p-lim}\,\frac{T_W^{(N)}}{N^{\frac{1}{2}}} = (3(1-s))^{\frac{1}{2}}\,[\int G(x+2\theta)\,dG(x)-(1/2)]\,, \tag{3.7}$$

where G is the distribution function of $X_1 - X_2$ when X_1, X_2 are independent and identically distributed according to F. In equations (3.5) - (3.7), "p-lim" denotes the probability limit of the random variable computed under the H_1 alternatives.

Conditions I, II and III of Bahadur ([1], p.276) are immediately verified and we may state

Theorem 2: For the H_1 alternatives (3.1), the Bahadur efficiencies are

$$B_\theta(W,U) = (b_W(\theta)/b_U(\theta))^2 = \frac{[\int G(x+2\theta)\,dG(x)-(1/2)]^2}{4s[\int F(x+\theta)\,dF(x)-(1/2)]^2}\,, \tag{3.8}$$

$$B_\theta(V,U) = (b_V(\theta)/b_U(\theta))^2 = s(12\lambda(F)-3)^{-1}B_\theta(W,U)\,. \tag{3.9}$$

The quantities $b_i^2(\theta)$, $i=U,V,W$ are, in the terminology of Bahadur, the asymptotic slopes of the tests based on $T_U^{(N)}$, $T_V^{(N)}$, and $T_W^{(N)}$ respectively.

Part of Bahadur's motivation (specialized to our statistics) of this efficiency measure is the following. The approximate levels attained by the statistics $T_i^{(N)}$ which reject for large values are $1 - \Phi(T_i^{(N)})$, $i = U,V,W$, where Φ is the unit normal cumulative distribution function. The word approximate relates to the fact that the limiting null distribution Φ of $T_i^{(N)}$, rather than the exact null distribution, is used. Suppose now that H_1 is true. For a given outcome of the N X's and Y's it makes sense to say the $T_W^{(N)}$ procedure, for example, is better than the $T_U^{(N)}$ procedure if $1-\Phi(T_W^{(N)}) < 1-\Phi(T_U^{(N)})$, or equivalently if $K_U^{(N)} < K_W^{(N)}$ where[3] $K_i^{(N)} = -2\log[1-\Phi(T_i^{(N)})]$. The efficiency measure $B_\theta(W,U)$ has the property that the random variable $K_W^{(N)}/K_U^{(N)}$ converges in probability to $B_\theta(W,U)$ as $N \to \infty$. Paraphrasing Bahadur, with probability tending to 1, $T_W^{(N)}$ is less successful (for the particular θ) than $T_U^{(N)}$ if $B_\theta(W,U) < 1$, more successful if $B_\theta(W,U) > 1$, and equally successful to this degree of approximation if $B_\theta(W,U) = 1$. Of course this is not the only motivation given by Bahadur, and the reader interested in other interpretations, advantages, and pitfalls of this efficiency measure should refer to the papers of Bahadur and Gleser.

Table 1 gives values of the Bahadur efficiencies when F is normal with variance σ^2. In all the Tables of this paper the entries involving W are only valid for $s = 1/2$, but the W efficiencies for $s > 1/2$ are obtained simply by dividing the tabular values by 2s.

TABLE 1: BAHADUR EFFICIENCIES FOR NORMAL TRANSLATION

θ/σ	.25	.5	1	2	3
$B_\theta(W,U)$	.990	.960	.860	.641	.533
$B_\theta(V,U)$	1.025	.995	.891	.665	.552

From (3.8) and (3.9) we see that $B_\theta(V,U)$ does not depend on s. Also,

$$B_\theta(W,V) = (B_\theta(W,U)/B_\theta(V,U)) = s^{-1}(12\lambda(F)-3), \qquad (3.10)$$

---

[3] The $K_i^{(N)}$ random variable is introduced by Bahadur for mathematical convenience.

and this expression is independent of θ. Furthermore, letting $\theta \to \infty$ in (3.8) and noting that

$$\lim_{\theta \to \infty} [\int G(x+2\theta) dG(x) - (1/2)] = 1/2 = \lim_{\theta \to \infty} [\int F(x+\theta) dF(x) - (1/2)], \qquad (3.11)$$

we have

Corollary 1: $\lim_{\theta \to \infty} B_\theta(W,U) = (4s)^{-1}$. $\lim_{\theta \to \infty} B_\theta(V,U) = (48\lambda(F)-12)^{-1}$.

Even in the most favorable case for W, namely s = 1/2, this limiting Bahadur efficiency is only .5. Also, the values of $\lim_{\theta \to \infty} B_\theta(V,U)$ are .518, .510, and .529 for the normal, uniform, and exponential distributions, respectively. Thus for pure translation alternatives which are far from H_0 we cannot, using Bahadur efficiency as a criterion, recommend either W or V as a satisfactory competitor for U.

The standing of W and V, as competitors of U, is improved only slightly for translation alternatives near H_0. By letting $\theta \to 0$ in (3.8) and applying a result of Bahadur ([1], Appendix 2) we may state

Corollary 2: The Pitman efficiencies for the sequence of alternatives $F_2^{(N)}(x) = F_1(x-(c/N^{\frac{1}{2}}))$ are

$$E(W,U) = (s)^{-1}(\int g^2/\int f^2)^2, \qquad (3.12)$$

$$E(V,U) = (12\lambda(F)-3)^{-1}(\int g^2/\int f^2)^2, \qquad (3.13)$$

where f,g are the densities (now assumed to exist) corresponding to F,G.

The Pitman efficiencies are also easily derived by a direct application of Pitman's formula [8]. In fact, equation (3.12) should not be regarded as new as it is implicit in the work of Pitman [9] where the efficacies of both the Wilcoxon signed rank test and the Wilcoxon-Mann-Whitney rank sum test are given.

For $s = 1/2$, $E(W,U)$ equals 2/3 of the Pitman efficiency of the signed rank test with respect to the sign test for a single sample from the distribution G. Some values are given in Table 2.

TABLE 2: PITMAN EFFICIENCIES OF THE RANDOM-PAIRED SIGNED RANK TEST W WITH RESPECT TO U FOR TRANSLATION ALTERNATIVES

Distribution	Density	E(W,U)
1. Normal	$f(x) = (2\pi)^{-1/2} e^{-x^2/2}, \quad -\infty < x < \infty.$	1.00
2. Uniform	$f(x) = 1,\ 0 \leq x \leq 1;\quad 0$ otherwise.	.889
3. $\Gamma(2)$	$f(x) = xe^{-x},\ x > 0;\quad 0$ otherwise.	.781
4. Cosine	$f(x) = (1-\cos(x))/\pi x^2, \quad -\infty < x < \infty$.	.720
5. Exponential	$f(x) = e^{-x},\ x > 0;\quad 0$ otherwise.	.500
6. Cauchy	$f(x) = (\pi(1+x^2))^{-1}, \quad -\infty < x < \infty$.	.500

Although there is no loss in Pitman efficiency for normal translation, the values in Table 2 favor U over W. The status of V for these alternatives is substantially the same as that of W. Using (3.13) we find the values of $E(V,U)$ for densities 1,2, and 5 of Table 2 are, respectively, 1.036, .906, and .529.

There is some independent theoretical interest in the relationship of W to V. The statistic V is tedious to compute and not distribution-free under H_0. The random-paired signed rank test removes these difficulties but has the disadvantage of utilizing an irrelevant randomization. How much efficiency is lost by using W in place of V? The $E(W,V)$ expression is given by (3.10) and in Table 3 we list some values.

TABLE 3: EFFICIENCY OF W WITH RESPECT TO V

Distribution	Normal	Uniform	Exponential
E(W,V)	.965	.981	.944

The losses in efficiency, when using W in place of V in the equal sample size case, are only 1.9, 3.5, and 5.6 per cent respectively, for the uniform, normal, and exponential distributions. (Lehmann [6] has proved $\lambda(F) \leq 7/24$ which implies $E(W,V) \leq 1$ for all F.)

We next consider the contamination alternatives

$$H_2\text{: } F_1(x) = F(x),\ F_2(x) = (1-p)F(x) + pH(x),\ H(x) \leq F(x). \tag{3.14}$$

Hodges and Lehmann [3] have used these alternatives to compare the U test with the normal theory t-test. From (2.1), (2.4), and (2.5) we have

$$E_p(U/mn) = \int F_1 dF_2 = ((1 + p)/2) - p\int H dF, \tag{3.15}$$

$$E_p(2W'/n(n-1)) = E_p(V) = \int (F_1 {}^* F_1) d(F_2 {}^* F_2)$$

$$= 1-\int [(1-p)^2 F^* F + 2p(1-p)F^* H + p^2 H^* H] d(F^* F), \tag{3.16}$$

where "*" denotes convolution and the subscript p indicates the moment is computed under H_2. From Chebychev's inequality we obtain

$$c_U(p) = \text{p-lim}\, \frac{T_U^{(N)}}{N^{\frac{1}{2}}} = (s(1-s)12)^{\frac{1}{2}}[(p/2) - p\int H dF], \tag{3.17}$$

$$c_V(p) = \text{p-lim}\, \frac{T_V^{(N)}}{N^{\frac{1}{2}}} = (s(1-s))^{\frac{1}{2}}\, L(p,F,H)/(4\lambda(F)-1)^{\frac{1}{2}}, \tag{3.18}$$

$$c_W(p) = \text{p-lim} \frac{T_W^{(N)}}{N^{\frac{1}{2}}} = (3(1-s))^{\frac{1}{2}} L(p,F,H), \qquad (3.19)$$

where

$$L(p,F,H) = [p-(p^2/2) - 2p(1-p)\int(F^*H)d(F^*F) - p^2\int(H^*H)d(F^*F)], \qquad (3.20)$$

and the symbol "p-lim" in equations (3.17) - (3.19) now denotes the probability limit of the random variable computed under the H_2 alternatives. We then have

Theorem 3: For the H_2 alternatives (3.14), the Bahadur efficiencies are

$$B_p(W,U) = (c_W(p)/c_U(p))^2 = \frac{L^2(p,F,H)}{4s[(p/2)-p\int HdF]^2} \qquad (3.21)$$

$$B_p(V,U) = (c_V(p)/c_U(p))^2 = s(12\lambda(F)-3)^{-1}B_p(W,U). \qquad (3.22)$$

The efficiency $B_p(W,V)$ is independent of p and H and is again given by (3.10). Thus Table 3 is also applicable for the contamination alternatives.

When $H(x) = F(x-\theta)$,

$$\lim_{\theta \to \infty} L(p,F,H) = p - \frac{p^2}{2}, \qquad (3.23)$$

and we then obtain

Corollary 3: For the H_2 alternatives with $H(x) = F(x-\theta)$,

$$\lim_{\theta \to \infty} B_p(W,U) = s^{-1}(1-(p/2))^2. \quad \lim_{\theta \to \infty} B_p(V,U) = (12\lambda(F)-3)^{-1}(1-(p/2))^2. \qquad (3.24)$$

The limiting Bahadur efficiencies of Corollary 3 are decreasing functions of p but remain above 1 for fairly large p. For example, for s = 1/2, $\lim_{\theta \to \infty} B_p(W,U)$ is greater than 1 as long as p is less than (approximately).536. The indication is that for contamination with a large translation, V and W rate as serious competitors of U, especially for small p. This impression is further justified by the Pitman efficiencies. Letting $p \to 0$ in (3.21) yields

Corollary 4: For the H_2 alternatives, the Pitman efficiencies ($p \to 0$) are

$$E(W,U) = \frac{s^{-1}[\int(F^*F)d(F^*H)-(1/2)]^2}{[\int FdH - (1/2)]^2} , \tag{3.25}$$

$$E(V,U) = \frac{[\int(F^*F)d(F^*H)-(1/2)]^2}{(12\lambda(F)-3)[\int FdH-(1/2)]^2} . \tag{3.26}$$

The entries in Table 4 are selected values of the Pitman efficiencies when $H(x) = F(x-\theta)$ and F is normal with variance σ^2.

TABLE 4: PITMAN EFFICIENCIES FOR

CONTAMINATION BY A NORMAL SHIFT

θ/σ	.25	.5	1	2	3
E(W,U)	1.005	1.021	1.082	1.313	1.608
E(V,U)	1.041	1.058	1.122	1.360	1.667

Corollary 5: For the contamination alternatives with $H(x) = F(x-\theta)$, we have

a. $\lim_{\theta \to \infty} E(W,U) = s^{-1}$. $\lim_{\theta \to \infty} E(V,U) = (12\lambda(F)-3)^{-1}$. (3.27)

b. For all F, $\lim_{\theta \to \infty} E(V,U) \geq 2$.

Part a. of Corollary 5 follows directly from (3.25) and b. follows from a. and Lehmann's upper bound of 7/24 for $\lambda(F)$. We also note that (3.27) agrees with the results one obtains by letting $p \to 0$ in (3.24). In other words, for the B_p efficiencies of Theorem 3, with $H(x) = F(x-\theta)$ we have $\lim_{\theta \to \infty} \lim_{p \to 0} B_p = \lim_{p \to 0} \lim_{\theta \to \infty} B_p$.

4. SOME COMPARISONS WITH THE PAIRED t-TEST

It is interesting to compare the relationship of W to U with the relationship of the paired t-test to the unpaired t-test for the case $m = n$.

When $F_1 = N(\mu_1,\sigma_1^2)$ and $F_2 = N(\mu_2,\sigma_2^2)$ with $\sigma_1^2 = \sigma_2^2$ an exact test of $H_0\colon \mu_1 = \mu_2$ can be based on

$$t_1 = \frac{(n(n-1))^{\frac{1}{2}}(\overline{Y}-\overline{X})}{\left[\sum_{i=1}^{n} \left((X_i-\overline{X})^2 + (Y_i-\overline{Y})^2\right)\right]^{\frac{1}{2}}}, \tag{4.1}$$

which, under H_0, has the Student t-distribution on 2n-2 degrees of freedom. If $\sigma_1^2 \neq \sigma_2^2$ the t_1 test will not be exact. By randomly pairing the X's with the Y's an exact test of H_0 based on

$$t_2 = \frac{(n(n-1))^{\frac{1}{2}}\,\overline{Z}}{\left(\sum_{i=1}^{n} (Z_i - \overline{Z})^2\right)^{\frac{1}{2}}} \tag{4.2}$$

is obtained. (Here $Z_i = Y_i - X_i$ and we have again assumed, for simplicity of notation, that the random pairing results in the pairs (X_i, Y_i).) Under H_0, t_2 has the Student t-distribution on n-1 degrees of freedom even when the populations have different variances.

The situation in the nonparametric case is partially analogous. If F_1 and F_2 differ by a scale parameter the U test will not be exact (e.g. see [10]). On the other hand, W will preserve its null distribution when F_1 and F_2 are symmetric about the same point. In particular, if $F_1(x) = H(x-\theta)$, $F_2(x) = H(c(x-\theta))$ with $c \neq 1$ and H symmetric about 0, the W test will be exact but the U test will not be exact. If considerations of exact size are important to the user, this would represent an advantage of the W test.

There may be other reasons to pair - not at random. For example, we might want to eliminate the nuisance parameters in a model corresponding to $E(X_i) = \mu_1 + b_i$, $E(Y_i) = \mu_2 + b_i$, $i=1,2,\ldots n$. At any rate suppose we pair when it is not really necessary. (The phrase "not really necessary" could refer to $\sigma_1^2 = \sigma_2^2$ when we are wary of unequal variances, or all the b_i's being equal in the model just mentioned.) How much efficiency is lost by pairing? With t-testing, asymptotically we lose nothing as the Pitman efficiency of t_2 with respect to t_1 is 1. However, we can lose asymptotic efficiency by using W in place of U. The efficiency loss for various distributions can be obtained from Table 2, section 3.

One important dissimilarity between the t_1- t_2 and U-W correspondences is the following. By using t_2 in place of t_1, we retain the same consistency parameter. The two-sided t_1 and t_2 tests will be consistent if $E(Y-X) \neq 0$ (assuming finite variances). But the set of alternatives for which the U test is consistent is different than the set of alternatives for which the W test is consistent. The two-sided test based on U is consistent if and only if $P(X_1 < Y_1) \neq 1/2$ while the two-sided tests based on V and W are consistent if and only if $P(X_1 + X_2 < Y_1 + Y_2) \neq 1/2$. If we consider the densities $f_1(x) = 1$ if $4 \leq x \leq 5$, and 0 otherwise, and $f_2(x) = a$ if $1 \leq x \leq 2$, b if $10 \leq x \leq 11$, and 0 otherwise, a simple calculation shows that for $a = b = 1/2$ the two-sided W and V tests are consistent but the two-sided U test is not consistent, whereas for $a = 1/\sqrt{2}$, $b = 1-(1/\sqrt{2})$ we get the opposite conclusion. Similar examples are easily constructed for the one-sided tests.

5. CONCLUSION

For pure translation alternatives, neither W nor V proved to be a worthy competitor of the U test. On the basis of this work the author recommends W and V for consideration in situations where protection against alternatives of the form $F_2(x) = (1-p)F_1(x) + pF_1(x-\theta)$ is desirable. For example, we may suspect that the treatment (with translation θ) will be active on a fraction p of the subjects who receive it. Suppose we have little information about the value of p. For p close to 1, these alternatives resemble the pure translation alternatives H_1(3.1). For p close to 0, Table 4 and Corollary 5 are relevant. If good sensitivity to these alternatives with p near 0 is important, V or W may be preferred to U.

Confronted with a choice between W and V, the decision will depend on the vagaries of the user. Many people will immediately dismiss W due to its dependence on randomization. On the other hand, V is not distribution-free and is more difficult to compute than W. The efficiency loss incurred by using W in place of V is small when s is approximately 1/2 but otherwise becomes intolerable (divide the entries of Table 3 by 2s).

6. ACKNOWLEDGMENTS

I am grateful to the late Professor Frank Wilcoxon for suggesting this investigation and for many interesting discussions pertaining to the signed rank and rank sum tests. The suggestions and criticisms of Professor John Pratt, Professor I. Richard Savage, and the referees have improved the manuscript considerably. I thank these people for their generous aid without implying they agree with either the material presented, or the point of view I take toward it.

REFERENCES

[1] Bahadur, R. R., "Stochastic comparison of tests," <u>Annals of Mathematical Statistics</u>, 31 (1960), 276-295.

[2] Gleser, L. J., "On a measure of test efficiency proposed by R. R. Bahadur," <u>Annals of Mathematical Statistics</u>, 35 (1964), 1537-1544.

[3] Hodges, J. L. and Lehmann, E. L., "The efficiency of some nonparametric competitors of the t-test," <u>Annals of Mathematical Statistics</u>, 27 (1956), 324-335.

[4] Hodges, J. L. and Lehmann, E. L., "Comparison of the normal scores and Wilcoxon tests," <u>Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability</u>, Berkeley and Los Angeles: University of California Press. Vol. 1. 307-317. 1961.

[5] Lehmann, E. L., "Consistency and unbiasedness of certain nonparametric tests," <u>Annals of Mathematical Statistics</u>, 22 (1951), 165-179.

[6] Lehmann, E. L., "Asymptotically nonparametric inference in some linear models with one observation per cell," <u>Annals of Mathematical Statistics</u>, 35 (1964), 726-734.

[7] Mann, H. B., and Whitney, D. R., "On a test of whether one of two random variables is stochastically larger than the other," <u>Annals of Mathematical Statistics</u>, 18 (1947), 50-60.

[8] Noether, G. E., "On a theorem of Pitman," <u>Annals of Mathematical Statistics</u>, 26 (1955), 64-68.

[9] Pitman, E. J. G., <u>Lecture Notes on Non-parametric Statistics</u>, New York: Columbia University, (1948).

[10] Pratt, J. W., "Robustness of some procedures for the two-sample location problem," Journal of the American Statistical Association, 59 (1964), 665-680.

[11] Tukey, J. W., "The simplest signed-rank tests," Princeton University Statistical Research Group Memo Report, No. 17 (1949).

[12] Wilcoxon, F., "Individual comparisons by ranking methods," Biometrics 1 (1945), 80-83.