J. Korean Math. Soc.
Vol. 14, No. 2, 1978.

USE OF THE WILCOXON TEST FOR THE PARALLELISM OF TWO REGRESSION LINES

BY MOON SUP SONG

1. Introduction.

Consider the regression model

$$Y_{ij}=\alpha_i+\beta_i X_{ij}+e_{ij},\ j=1,2,\cdots,N_i,\ i=1,2$$

where the X's are known constants, the α's are nuissance parameters, and the β's are regression parameters. The Y's are observable while the e's are mutually independent unobservable random variables. For line i, the e_i's are identically distributed according to the unknown continuous cdf $F_i(e)$.

We wish to test the null hypothesis that two regression lines are parallel, i.e., we want to test

$$H_0:\beta_1=\beta_2.$$

When the error terms e_1's and e_2's are normally distributed with the same variance, the Student's t-test may be applied. However, there are many good and simple nonparametric tests for the regression problems which do not require any assumptions other than independent samples from identically distributed continuous populations. This paper presents a test which is valid for two arbitrary distributions of two sets of error terms.

Sen (1969) studied a class of more general models including the case of non-identical distributions. Hollander (1970) proposed a distribution-free signed rank test for testing H_0. But his method is applicable only for the case of equal numbers of observations from the two lines. Potthoff (1974) introduced a conservative nonparametric test which resembles the Wilcoxon two-sample test. According to the Monte Carlo values in Hollander (1970), the Pottoff's test is too conservative for potential users. Potthoff's test also requires the assumption that no X_1's are equal and no X_2's are equal.

Trying to avoid the disadvantages of the Hollander's and Potthoff's tests, we propose a method to utilize the ordinary Wilcoxon two-sample statistic applied to the slope estimators $(Y_{1i}-Y_{1i'})/(X_{1i}-X_{1i'})$ and $(Y_{2j}-Y_{2j'})/(X_{2j}-X_{2j'})$. This method is nonparametric in the sense that continuity is the

Received by the editors Oct. 20, 1977.

only assumption made concerning the error distributions F_1 and F_2, but it is not distribution-free. We present some Monte Carlo results which suggest the usage of the ordinary Wilcoxon two-sample statistic even when $F_1 \neq F_2$.

2. Procedures and statement of results.

We assume that $N_1=2m$ and $N_2=2n$. To use the slope estimators proposed by Hollander (1970), for line $i=1$ pair the X_{1j}'s to form m groups of the form $(X_1, X_{1'})$, and for line $i=2$ pair X_{2j}'s to form n groups of the form $(X_2, X_{2'})$, each containing two unequal X's. For each group compute a slope estimator of β_i of the form $(Y-Y')/(X-X')$ where Y, Y' are observations corresponding to the paired X, X' points. The grouping depends only on the $\{X_{ij}\}$ configurations. Random pairing may be suggested as a safeguard against the introduction of biases.

Then, the Wilcoxon (1945) two-sample test, or equivalently, the Mann–Whitney (1947) two-sample test can be applied to these slope estimators. Thus our test of H_0 will be based on the statistic

$$W=(1/mn)\sum_{i=1}^{m}\sum_{j=1}^{n}U(V_{ij}) \tag{2.1}$$

where $U(V)$ equals 0 if $V \leq 0$ and 1 if $V>0$, and where

$$V_{ij}=\frac{Y_{2j}-Y_{2j'}}{X_{2j}-X_{2j'}}-\frac{Y_{1i}-Y_{1i'}}{X_{1i}-X_{1i'}}.$$

Note that the V's are, in general, not identically distributed, but they are independent and symmetric.

The one-sided test of H_0 based on (2.1) against alternatives $\beta_2>\beta_1$ rejects for large values of W. Our proposed test W will use the exact critical values of the ordinary Wilcoxon test statistic or will treat

$$(W-1/2)/[(m+n+1)/12mn]^{1/2} \tag{2.2}$$

as an approximate $N(0,1)$ variate under H_0.

To justify our proposal we note the following three points:

(i) Regardless of what F_1 and F_2 are, under H_0 the distribution of $[W-E(W)]/[\mathrm{Var}(W)]^{1/2}$ is asymptotically $N(0,1)$ under mild restrictions. For proof and statement of conditions, refer to Hoeffding (1948) or the section 5 of Potthoff (1974). The proof is similar to that of Potthoff to warrant ommision.

(ii) Under H_0, no matter what F_1 and F_2 are,

$$E(W)=1/2$$

which follows immediately from the symmetry of V about the origin. If H_0 is not true, $E(W)\neq 1/2$. For the case $\beta_2>\beta_1$, we have $E(W)>1/2$, and for the case $\beta_2<\beta_1$, we have $E(W)<1/2$. (See section 3 of Potthoff (1974).)

(iii) For the equally spaced design, which is explained in Section 3, the upper bound of the variance is given by

$$\mathrm{Var}(W)\leq 1/[4\ \min(m,n)] \tag{2.3}$$

which follows from Birnbaum and Klose (1957). When $F_1=F_2$ and H_0 is true, W has the usual null distribution of Wilcoxon-Mann-Whitney two-sample test statistics. In this case W has variance $(m+n+1)/12mn$. For the case $m=n$, we find from the equation (2.6) of Birnbaum and Klose (1957), that

$$\mathrm{Var}(W)\geq (2n+1)/12n^2, \tag{2.4}$$

using the symmetry of V about 0.

REMARKS: To apply a conservative test of H_0 we may use the Wilcoxon statistics, with $(m+n+1)/12mn$ replaced by $1/[4\ \min(m,n)]$. That is, to test H_0 against alternatives $\beta_2>\beta_1$, if we use the critical region

$$C=2[\min(m,n)]^{\frac{1}{2}}(W-1/2)>z_{1-\alpha} \tag{2.5}$$

where $z_{1-\alpha}$ is the $1-\alpha$ percentile point of $N(0,1)$, then the test based on (2.5) will be approximately of size α. The C test based on (2.5) was developed by Potthoff (1963) concerning the use of the Wilcoxon statistic for a generalized Behrens-Fisher problem.

3. Monte Carlo study with equally spaced design.

Hollander (1970) has proposed a test of H_0 based on

$$H=\sum_{i=1}^{n}r_iU(V_i)=\sum_{i\leq j}^{n}U(V_i+V_j) \tag{3.1}$$

where $V_i=(Y_{2i}-Y_{2i'})/(X_{2i}-X_{2i'})-(Y_{1i}-Y_{1i'})/(X_{1i}-X_{1i'})$, $i=1,2,\cdots,n$, and r_i is the rank of $|V_i|$ in the joint ranking from the least to greatest of $|V_1|, |V_2|, \cdots, |V_n|$. Under H_0, the H test statistics has the null distribution of Wilcoxon's signed rank statistics.

The Student's t-test of H_0 can be based on

$$t=(b_2-b_1)/sZ,$$

where

$$b_i=\sum_{j=1}^{N_i}(Y_{ij}-Y_{i\cdot})(X_{ij}-X_{i\cdot})/\sum_{j=1}^{N_i}(X_{ij}-X_{i\cdot})^2,$$

$$s^2=(N_1+N_2-4)^{-1}\sum_{i=1}^{2}\sum_{j=1}^{N_i}(Y_{ij}-a_i-b_iX_{ij})^2,$$

$$a_i=Y_{i\cdot}-b_iX_{i\cdot},\quad Z^2=\sum_{i=1}^{2}\left[1/\sum_{j=1}^{N_i}(X_{ij}-X_{i\cdot})^2\right].$$

When F_1 and F_2 are normally distributed with the same variance and H_0 is true, t has the Student's t-distribution with (N_1+N_2-4) degrees of freedom.

Potthoff (1974) has suggested a test of H_0 based on

$$P=\left[\binom{N_1}{2}\binom{N_2}{2}\right]^{-1}\sum_{i<i'}\sum_{j<j'}U\Big[(Y_{2j}-Y_{2j'})/(X_{2j}-X_{2j'}) - (Y_{1i}-Y_{1i'})/(X_{1i}-X_{1i'})\Big],$$

and the one-sided test against $\beta_2>\beta_1$ would have as critical region

$$(3.2)\qquad (P-1/2)[(2M+5)/18M(M-1)]^{-\frac{1}{2}}>z_{1-\alpha}$$

where $M=\min(N_1, N_2)$, $z_{1-\alpha}$ is the $1-\alpha$ percentile point of a $N(0,1)$ distribution.

Monte Carlo sampling with the equally spaced design, which was suggested by Hollander (1970) for the H test, was used to compare the powers of the H and W tests. Setting $N_1=N_2=N=2kn'$, we have n' observations at each of $2k$ points

$$(3.3)\qquad C_i+2c_ij,\quad j=0,\cdots,2k-1,$$

where C_i, c_i are arbitrary constants with $c_i>0$, $i=1,2$. Take the k intervals

$$[C_i+2c_ij,\ C_i+2c_i\ (j+k)],\quad j=0,\cdots,k-1$$

and for each interval obtain n' independent slope estimators of β_i, $i=1,2$.

For the design (3.3) with $k=10$, $n'=1$, $c_1=c_2=1/2$, 500 samples with each sample consisting of 20 random values from F_1 and F_2 are used to estimate powers. Parameters are chosen to compare the results with those of Hollander (1970, Table 3), and the Monte Carlo results are summarized in Table 1. The one-sided t, C (based on (2.5)), W, and H tests are used to test H_0 against alternatives $\beta_2>\beta_1$ for various (F_1, F_2) pairs and several values of

$$\Delta=(\beta_2-\beta_1)/\left\{\left[(\sigma_1^2+\sigma_2^2)/2\right]\sum_{i=1}^{2}\left[1/\sum_{j=1}^{2k'}(X_{ij}-X_{i\cdot})^2\right]\right\}^{\frac{1}{2}}.$$

The letters N, U, and E represent, respectively, the distributions with density functions $f(x)=(2\pi)^{-\frac{1}{2}}\exp(-x^2/2)$ for $-\infty<x<\infty$; $f(x)=1$ for $0\leq x\leq 1$ and 0 otherwise; $f(x)=\exp(-x)$ for $x\geq 0$ and 0 otherwise. The four $\triangle$ values were chosen so that the actual powers of the t test for $(F_1, F_2)=(N, N)$ would be .0098, .3, .7, and .9. The powers of the the W test are interpolated using the exact table.

Table 1. Estimated Power for the Equally Space Design

(F_1, F_2)	$\triangle=0.00$				$\triangle=1.90$			
	t	C	W	H	t	C	W	H
(N, N)	.006	.002	.012	.010	.298	.086	.241	.224
(U, U)	.006	.000	.012	.008	.296	.066	.200	.204
(E, E)	.002	.000	.007	.008	.352	.122	.313	.264
(U, N)	.012	.004	.022	.014	.322	.102	.237	.228
(E, N)	.012	.004	.014	.004	.350	.112	.265	.242
(U, E)	.004	.000	.013	.004	.342	.174	.356	.288

(F_1, F_2)	$\triangle=3.01$				$\triangle=3.79$			
	t	C	W	H	t	C	W	H
(N, N)	.726	.250	.552	.516	.910	.492	.783	.718
(U, U)	.748	.226	.493	.464	.890	.464	.775	.754
(E, E)	.764	.358	.619	.520	.878	.586	.840	.710
(U, N)	.722	.280	.543	.524	.886	.456	.743	.712
(E, N)	.674	.310	.625	.558	.896	.532	.769	.722
(U, E)	.728	.420	.665	.552	.872	.626	.851	.726

4. Conclusions.

The Monte Carlo results show that the C test based on (2.5) is very consertive. The powers of the C test are almost same as those of the Potthoff's P-test based on (3.2) (refer to Table 3, Hollander (1970)) for the equally spaced design with the same sample sizes. The W test has greater power than the Hollander's H-test, but the probability of Type I error of the W test exceeds the significance level $\alpha=.0098$ when $F_1\neq F_2$.

In the equation (2.4), the equality sign holds if and only if $F_1=F_2$. Thus, in the case $m=n$, we are using a lower bound for the variance of W

and the probability of Type I error will exceed the intended value α, unless $F_1=F_2$. But, the Monte Carlo calculations find that it is not so serious.

If equal numbers of observations for the two lines are available, then the Hollander's distribution-free H-test is suggested. The P test based on (3.2) or the C test based on (2.5) will have superior power in situations where the variance of P or C under H_0 is sufficiently close to the upper limit.

References

[1] Birnbaum, Z. W. and Klose, O. M. (1957). *Bounds for the variance of the Mann-Whitney statistics*, Ann. Math. Statist. **28**, 933–945.
[2] Hoeffding, W. (1948). *A class of statistics with asymptotically normal distribution*, Ann. Math. Statist. **19**, 293–325.
[3] Hollander, M. (1970). *A distribution-free test for parallelism*, J. Amer. Statist. Assoc. **65**, 387–394.
[4] Mann, H. B. and Whitney, D. R. (1947). *On a test of whether one of two random variables is stochastically larger than the other*, Ann. Math. Statist. **18**, 50–60.
[5] Potthoff, F. (1963). *Use of the Wilcoxon statistics for a generalized Behrens-Fisher problem*, Ann. Math. Statist. **34**, 1596–1599.
[6] Potthoff, F. (1974). *A nonparametric test of whether two simple regression lines are parallel*, Ann. Statist. **2**, 295–310.
[7] Sen, P. K. (1969). *On a class of rank order tests for the parallelism of several regression lines*, Ann, Math. Statist. **40**, 1668–1683.
[8] Wilcoxon, F. (1945). *Individual comparisons by ranking methods*, Biometrics Bull. **1**, 80–83.

Seoul National University