A C T A   U N I V E R S I T A T I S   L O D Z I E N S I S
FOLIA OECONOMICA 132, 1993

*Czesław Domański**, *Andrzej S. Tomaszewicz***

TESTS BASED ON RUN LENGTH FOR TWO SAMPLES

Abstract. In the practice of statistical research, tests based on the number of runs are often applied. It concerns the tests for one, two or more samples. Seldom the length of runs is applied as a test statistic.

In the paper, we present a test for two samples based on the run length. Its power is compared with the t-Student parametric test, non-parametric Wilcoxon test and the Wald-Wolfowitz test based on the number of runs.

Key words: Two sample tests, runs tests, Wald-Wolfowitz test, non-parametric Wilcoxon tests.

1. THE PROBLEM

Let

$$X_1, X_2, \ldots, X_m \quad \text{and} \quad Y_1, Y_2, \ldots, Y_n \tag{1}$$

be two independent samples drawn from a population with continuous distribution functions F and G respectively. We verify the hypothesis

$$H_0 : F = G. \tag{2}$$

It is well known that if F and G are normal distribution functions with the same variance, the statistic

* Professor at the Institute of Econometrics and Statistics of the University of Łódź

** Late Professor at the Institute of Ekonometrics and Statistics of the University of Łódź.

$$t = \frac{\bar{Y} - \bar{X}}{\sqrt{\left(\frac{1}{m} + \frac{1}{n}\right) \frac{(m - 1)s_x^2 + (n - 1)s_y^2}{n + m - 2}}} \tag{3}$$

has the Student distribution with $m + n - 2$ degrees of freedom. In this case

$$G(x) = F(x + \Delta), \tag{4}$$

so (2) is equivalent to

$$H_0 : \Delta = 0 \tag{5}$$

and the test based on (2) is the most powerful. Usually

$$H_1 : \Delta > 0 \tag{6}$$

is accepted as the alternative hypothesis.

When F and G differ from the normal distribution function or, because of not equal variances, (3) does not hold, the t test does not have optimal properties and it is reasonable to search for alternative solutions.

In this paper, we compare the test based on the run length with the Wald-Wolfowitz test [W a l d-W o l f o w i t z (1940)], t test (3) and the Wilcoxon test [W i l c o x o n (1945)]. Generally speaking, the last one is supposed to be the most powerful among non-parametric two-sample tests [see e.g. M i l t o n (1970)].

The test procedure based on run length is as follows.

(a) All values (1) are ordered in the increasing sequence

$$z_1 < z_2 < \ldots < z_{m+n}. \tag{7}$$

(b) The maximum length of runs L in (7) is counted.

(c) H_0 is rejected if $L \geq l_\alpha$ where l_α is the critical point

$$l_\alpha = \min \{l : P_0(L \geq l) \leq \alpha\} \tag{8}$$

corresponding to a chosen significance level α and P_0 denotes probabilities when H_0 is true. (A small length of runs testifies to H_0, so the critical region is right-hand sided).

The Wald-Wolfowitz test statistic is the number of runs in (7).

The Wilcoxon test against the alternative (6) can be defined as follows:

(a) All values (1) are ordered in the sequence (7). Let's assume that w_i $(i = 1, \ldots, m)$ is the rank of element Y_i, that means $Y_i = Z_{w_i}$.

(b) The test statistic

$$W = \sum_{i=1}^{n} w_i$$

is computed.

(c) H_0 is rejected if $W \geq w_\alpha$ where w_α is the critical point

$$w_\alpha = \min \{w : P_0(W \geq w) \leq \alpha\}. \quad (9)$$

Both the Wald-Wolfowitz statistic and the Wilcoxon statistic W are discrete. Thus, in order to compare the power we apply randomized tests.

According to the randomized length-of-run test

- H_0 is rejected if $L \geq l_\alpha$,
- H_0 is rejected with the probability

$$p_\alpha^L = \frac{\alpha - P_0(L \geq l_\alpha)}{P_0(L = l_\alpha - 1)} \quad (10)$$

when $L = l_\alpha - 1$,

- H_0 is accepted if $L < l_\alpha - 1$.

The size of this test is obviously equal to α:

$$P_0(L \geq l_\alpha) + p_\alpha P_0(K = l_\alpha - 1) = \alpha,$$

and its power:

$$1 - \beta^L = P_1(L \geq l_\alpha) + p_\alpha^L P_1(L = l_\alpha - 1),$$

where P_1 denotes probabilities when H_1 is true.

Analogously we define the randomized Wilcoxon test, the randomized Wald-Wolfowitz test and their powers

$$1 - \beta^W = P_1(W \geq w_\alpha) + p_\alpha^W P_1(W = w_\alpha - 1),$$

$$1 - \beta^R = P_1(R \leq r_\alpha) + p_\alpha^R P_1(R = r_\alpha + 1).$$

2. MONTE CARLO EXPERIMENT

We attempted to evaluate the length-of-runs two-sample test power using the Monte Carlo methods[1]. For

- 2 values of $c = 1, 3$,
- 7 values of $\Delta = 0, 0.5, 1, 1.5, 2, 2.5, 3$,
- 5 sample sizes $m + n = 12, 24, 36, 48, 60$,
- 2 values of the quotient $m/n = 1, 2$,
- 3 types of probability distribution F: normal, exponential, double exponential

the amount of $q = 1000$ $(m + n)$ - observation samples was drawn; each sample consisted of

- m observations from the distribution $F(x)$,
- n observations from the distribution $G(x) = F(cx + \Delta)$.

The results are shown on the graphs (Fig. 1-12). In each graph the horizontal axis contains values of Δ in the range 0.0-3.0. On the vertical one the values of empirical test power (Fig. 1-3) or the power difference (Fig. 4-12) are expressed. Two curves are drawn in each graph. The continuous one shows the dependence between the test power and the location parameter Δ when $c = 1$. The dashed line depicts the case when the scale parameter is different in the two samples $c = 3$. Ten graphs on each figure correspond to the chosen pairs of (m, n).

3. CONCLUSIONS

1. For distributions with equal variances, the t-Student test is much more powerful than the test based on the length of run (see Fig. 4-6).

2. In the case when distributions have different variances $(c = 3)$ and the difference between scale parameters is not large $(\Delta \leq 1)$, the test based on the run length is more powerful than t-Student test (Fig. 4-6).

---

[1] We took into account, among others, the experience of [R a n d l e s and W o l f e (1979)] and [D o m a ń s k i and T o m a s z e w i c z (1989)].

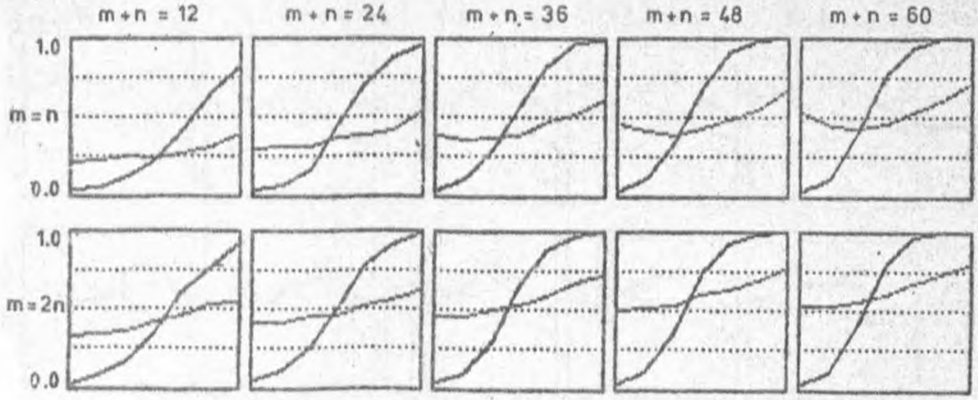

Fig. 1. Length-of-run test power - normal

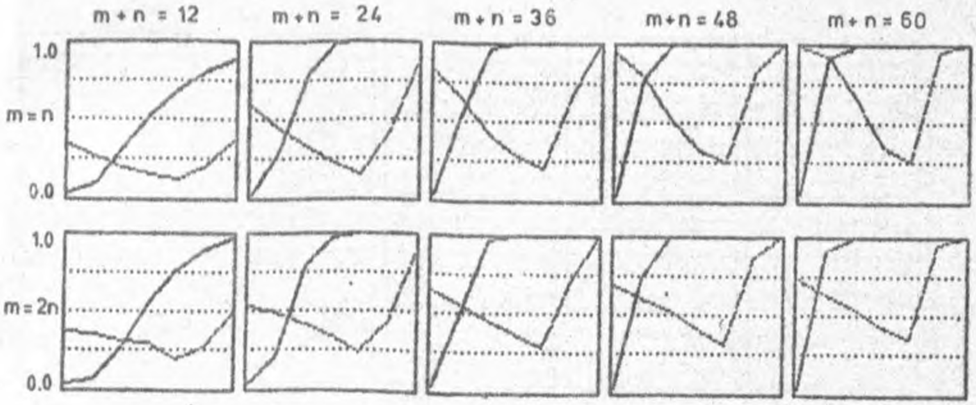

Fig. 2. Length-of-run test power - exponential

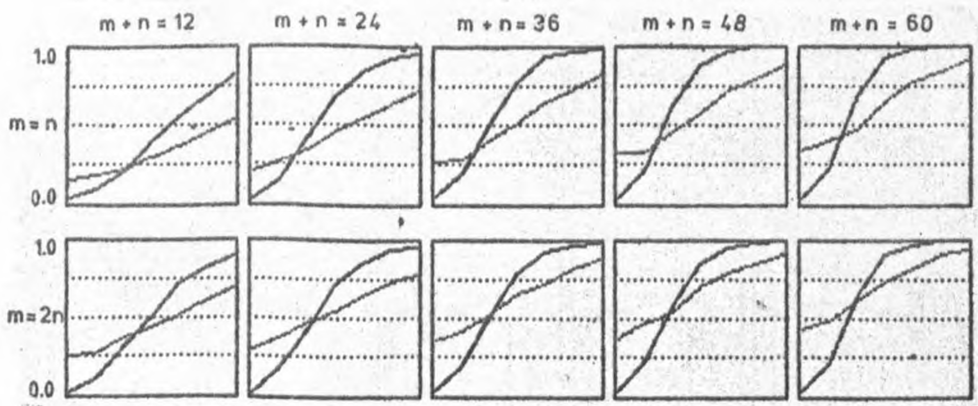

Fig. 3. Length-of-run test power - double exp.

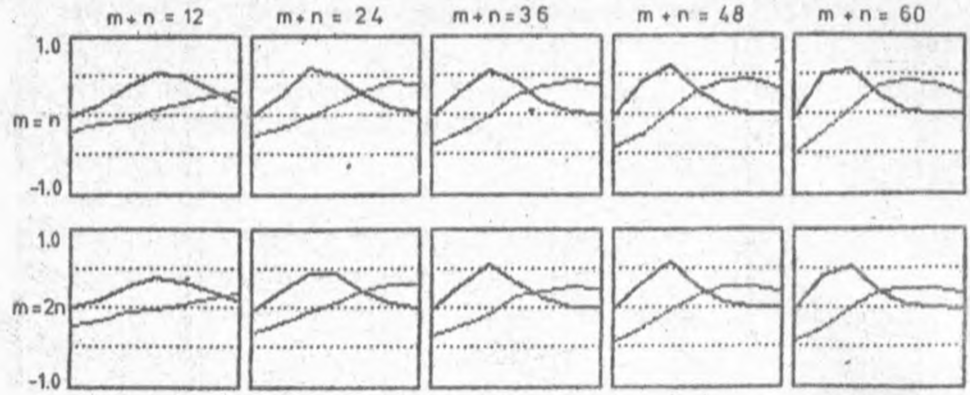

Fig. 4. t-test versus length-of-run test power - normal

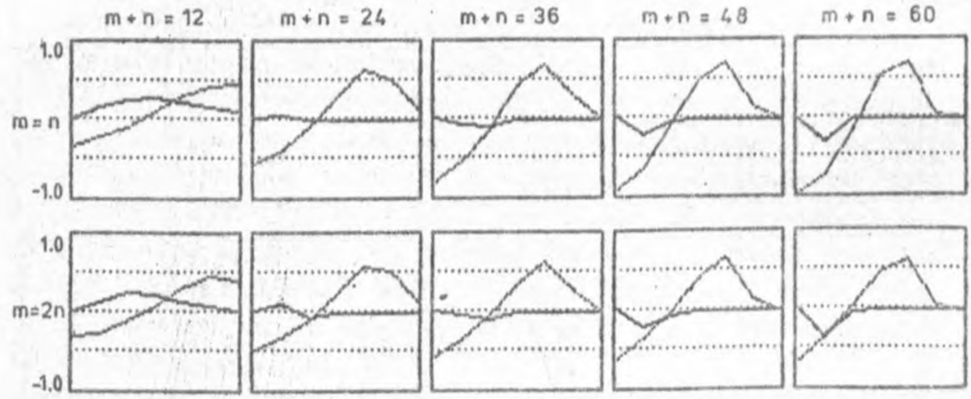

Fig. 5. t-test versus length-of-run test power - exponential

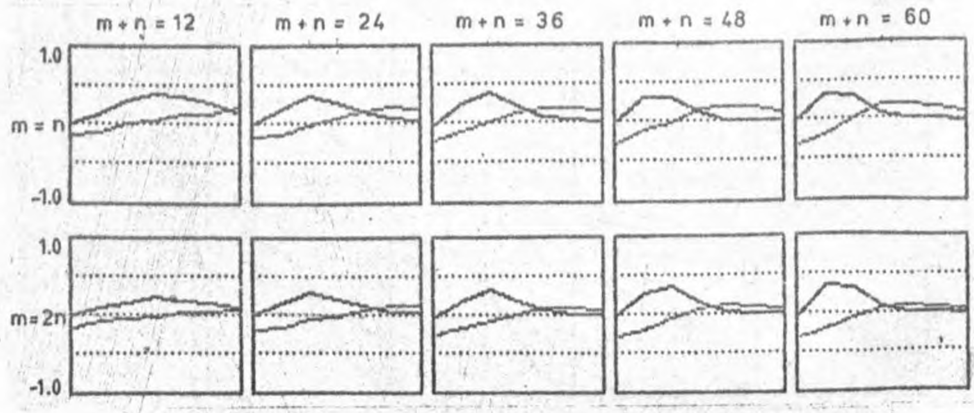

Fig. 6. t-test versus length-of-run test power - double exp.

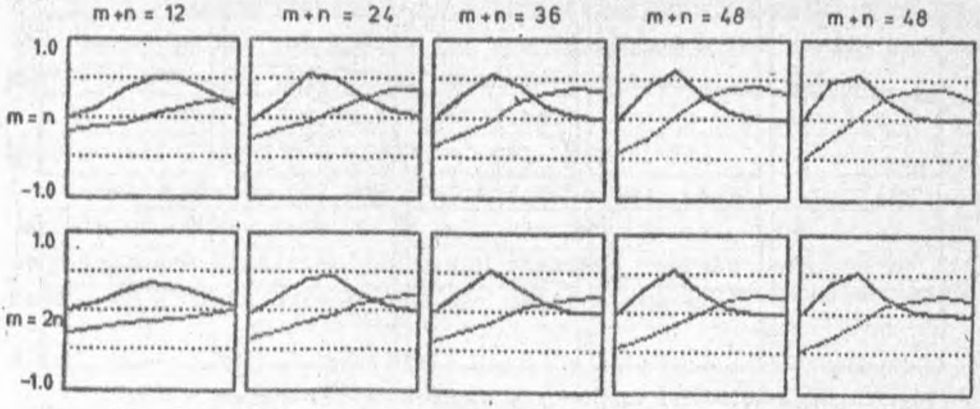

Fig. 7. Wilcoxon test versus length-of-run test power - normal

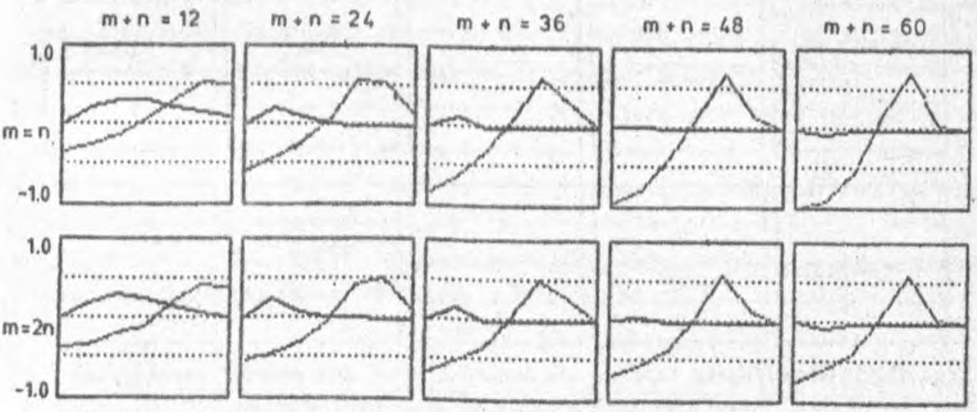

Fig. 8. Wilcoxon test versus length-of-run test power - exponential

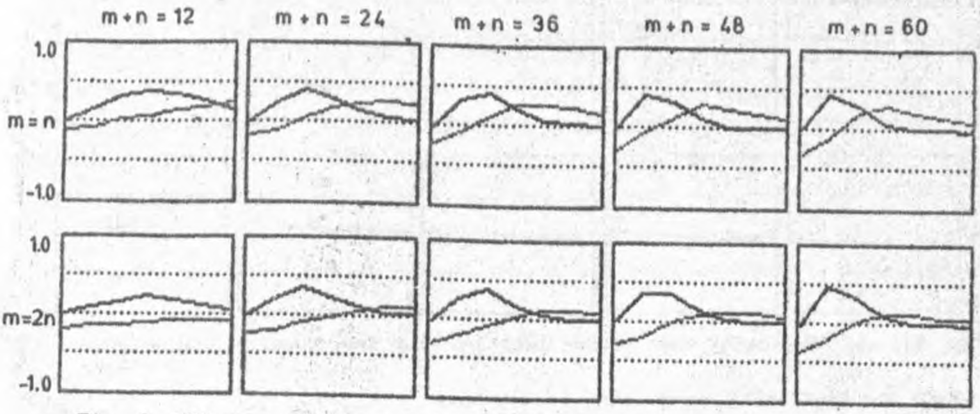

Fig. 9. Wilcoxon test versus length-of-run test power - double exp.

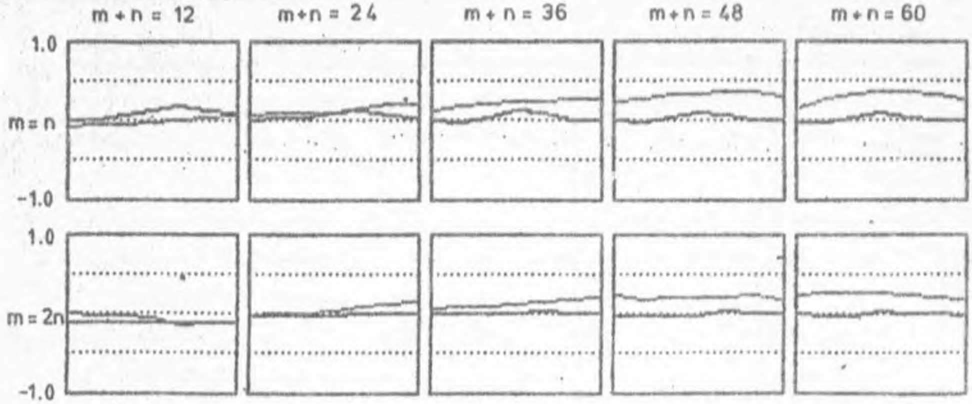

Fig. 10. Wald-Wolfowitz test versus length-of-run test power - normal

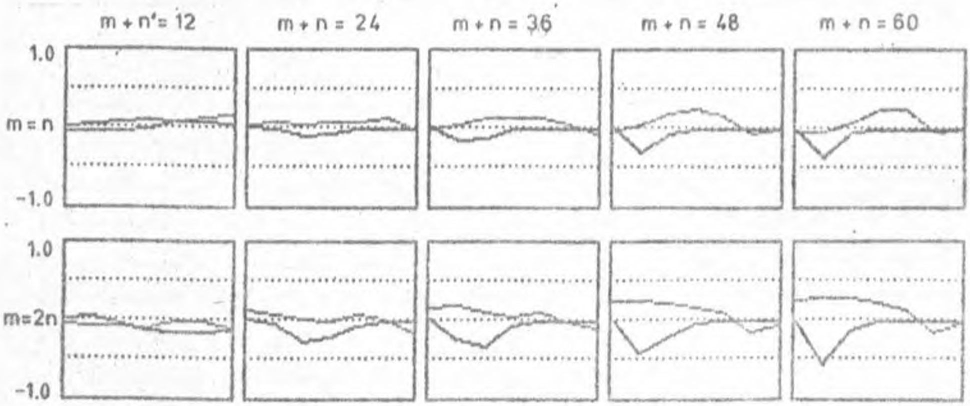

Fig. 11. Wald-Wolfowitz test versus length-of-run test power - exponential

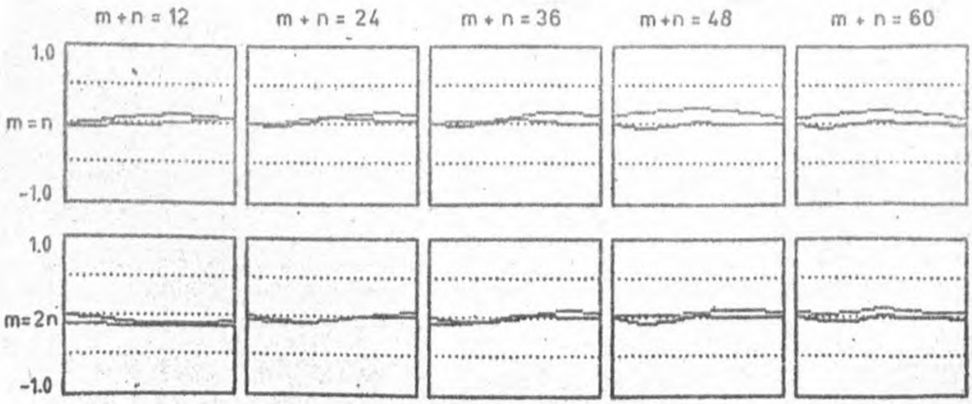

Fig. 12. Wald-Wolfowitz test versus length-of-run test power - double exp.

Note for Figs 1-12: —— c = 1 —— c = 3

3. The Wald-Wolfowitz test is generally more powerful than the test based on the run length for $n \geq 24$, although, for small samples the length-of-run test shows some advantage (Fig. 10-12).

4. The test based on the run length for $c = 3$ and $\Delta \leq 1$ is a more powerful then the Wilcoxon test (Fig. 7-9).

Summing up, we can say that the obtained results confirm the usefulness of the test based on the run length, especially when we expect the variations in the examined populations to be different to a great extent and the samples are small.

REFERENCES

Domański C., Tomaszewicz A. S. (1989): *Evaluation of Wald-Wolfowitz test power*, XIII SOR Ulm (to be published in the Proceedings).

Milton R. (1970); *Rank order probabilities*, Wiley, New York.

Randles H., Wolfe D. A. (1979): *Introduction to the theory of nonparametric statistics*, Wiley, New York.

Wald A., Wolfowitz J. (1940): *On a test whether two samples are from the same population*, AMS, 11, p. 147-162.

Wilcoxon F. (1945): *Individual comparisons by ranking methods*, "Biometrics Bulletin" (later "Biometrics"), 1, p. 80-87.

Czesław Domański, *Andrzej S. Tomaszewicz*

TESTY OPARTE NA DŁUGOŚCI SERII DWÓCH PRÓB

W artykule przedstawiamy pewien test hipotezy o równości dystrybuant dla przypadku dwóch prób losowych. Test ten oparty jest na długości serii. Moc tego testu została porównana z empiryczną mocą parametrycznego testu t-Studenta, testu Wilcoxona, oraz mocą testu Walda-Wilcoxona opartego na liczbie serii. Załączono 12 wykresów mocy empirycznej ww. testów.