

*Wojciech Zieliński\**

## ROBUST TEST FOR COMPARISON OF TWO VARIANCES

**Abstract.** The size of the F test for equality of variances is not robust against nonnormality. In the paper there is proposed a test the size of which is more robust to nonnormality than the size of the F test.

**Key words:** F test for variances, robust test

### I. INTRODUCTION

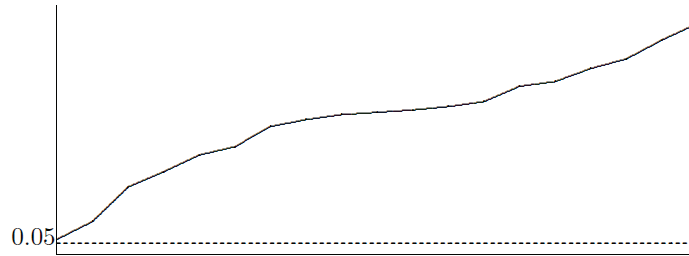
Consider a problem of testing hypothesis  $H_0: \sigma_1^2 = \sigma_2^2$  vs  $H_1: \sigma_1^2 > \sigma_2^2$  on the basis of two independent samples  $X_{11}, \dots, X_{1n_1}$  and  $X_{21}, \dots, X_{2n_2}$ . Under assumption of normality of both samples the most powerful test is based on the statistics  $F = S_1^2 / S_2^2$ , where  $S_1^2$  and  $S_2^2$  are sample variances. It appears that the size of that test strongly depends on the kurtosis of the underlying distribution. In the Figure 1 the size of the test for  $n_1 = n_2 = 5$  and  $\alpha = 0.05$  is shown in dependence on the kurtosis (x axis) of the distribution of the first sample. The shown results are obtained by simulation, but similar (asymptotical) results may be found in Scheffé (1959).

The problem is to find a test a size of which is more robust to nonnormality than the one of the F test.

In literature the problem was considered in a more general way, namely as a special case of the problem of testing equality of  $k \geq 2$  variances. The problem has a very long history (Pearson and Atyanthaya 1929, Pearson 1932). The first test the size of which is not influenced by the kurtosis may be found in Box and Andersen (1955) who introduced the term "robustness" into statistics. Levene (1960) proposed a test which is more robust than the Bartlett test. This test was slightly modified by Brown and Forsythe (1974). Other tests, more robust than the F test, were proposed by Miller (1968), Layard (1973), Flinger and Killen (1976) and Tiku and Balakrishnan (1984). The most recent results on the subject may be found in Tsou (2003, 2006).

---

\*Profesor, Department of Econometrics and Statistics University of Life Sciences.

Figure 1. The size of the  $F$  test

All mentioned tests are asymptotical and for small samples they don't work well. We are interested in the small sample approach.

In what follows we consider a situation in which the first sample is nonnormally distributed while the second one is normally distributed. Such a situation may occur in Goldfeld-Quandt test (Greene 2000). This test is devoted to investigate homoscedasticity in a classical linear regression problem  $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ ,  $i=1, \dots, n$  ( $x_1 < \dots < x_n$ ). It is assumed that random errors  $\varepsilon_i$ 's are independent random variables distributed as  $N(0, \sigma^2)$ . In the test, the sample is divided into two disjoint subsamples of size  $n_1$  and  $n_2$  respectively, so two models are considered  $Y_i = \beta_0^{(1)} + \beta_1^{(1)} x_i + \varepsilon_i^{(1)}$ ,  $i=1, \dots, n_1$  and  $Y_i^{(2)} = \beta_0^{(2)} + \beta_1^{(2)} x_i + \varepsilon_i^{(2)}$ ,  $i=n_2, \dots, n$  with the assumption:  $\varepsilon^{(1)} \sim N(0, \sigma_1^2)$  and  $\varepsilon^{(2)} \sim N(0, \sigma_2^2)$ . Verified hypothesis is  $H_0: \sigma_1^2 = \sigma_2^2$ .

The Goldfeld-Quandt test statistic is  $S_1^2/S_2^2$ , where  $S_1^2$  and  $S_2^2$  are residual variances in the first and the second model respectively.

In chapter 2 the new test is proposed. In chapter 3 properties of the test are investigated in Monte-Carlo experiment. The test is compared with the classical  $F$  test and the Levene test. The last one was chosen among others because this test is widely applied in statistical software. Concluding remarks and propositions of generalization may be found in chapter 4

## II. ROBUST TEST

Let  $X_{11}, \dots, X_{1n_1}$  be i.i.d.  $N(\mu_1, \sigma_1^2)$  and  $X_{21}, \dots, X_{2n_2}$  be i.i.d.  $N(\mu_2, \sigma_2^2)$ . The problem is to verify  $H_0: \sigma_1^2 = \sigma_2^2$  vs.  $H_1: \sigma_1^2 \neq \sigma_2^2$ .

The classical test is based on the  $F$  statistic:

$$F = S_1^2/S_2^2,$$

where

$$S^2_i = \frac{1}{n_i} \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2, \quad i = 1, 2$$

Hypothesis is rejected at a level  $\alpha$ , if  $F > F(\alpha; n_1-1, n_2-1)$ , where  $F(\alpha; v_1, v_2)$  is the  $\alpha$  critical value of the F distribution with  $v_1$  and  $v_2$  d.f.

Now suppose that  $X_{11}, \dots, X_{1n_1}$  are i.i.d. with nonnormal distribution and let

$$\gamma_2 = \frac{E(\xi_t - E\xi_t)^4}{(E(\xi_t - E\xi_t)^2)^2} - 3$$

be the kurtosis of the distribution. The kurtosis of a normal distribution is zero. The size of the classical F test strongly depends on the kurtosis. It seems, that the one of the reasons of such behavior is the dependence of the variance of the F statistic on the kurtosis.

Let  $\mathbf{x}_1 = (X_{11}, \dots, X_{1n_1})'$  and  $\mathbf{x}_2 = (X_{21}, \dots, X_{2n_2})'$ . Then

$$(n_i - 1)S^2_i = \mathbf{x}_i' \left( \mathbf{I} - \frac{1}{n_i} \mathbf{1}\mathbf{1}' \right) \mathbf{x}_i, \quad i = 1, 2$$

where  $\mathbf{I}$  denotes the identity matrix and  $\mathbf{1}$  denotes the vector of ones. Hence, F statistic is the quotient of two quadratic forms.

Let  $\mathbf{y} = (Y_1, \dots, Y_n)'$  be a vector of i.i.d. r.v. and let  $\mathbf{A}$  be a symmetric  $n \times n$  matrix. Assume that  $\mathbf{y}'\mathbf{A}\mathbf{y}$  is shift-invariant, i.e.  $\mathbf{A}\mathbf{1} = \mathbf{0}$ . It is well known (Atiqullah 1962) that

$$\begin{aligned} E[\mathbf{y}'\mathbf{A}\mathbf{y}] &= \sigma^2 \text{tr} \mathbf{A} \\ D^2[\mathbf{y}'\mathbf{A}\mathbf{y}] &= \sigma^4 (\gamma_2 \mathbf{a}'\mathbf{a} + 2\text{tr} \mathbf{A}^2), \end{aligned}$$

where  $\mathbf{a}$  is the vector of diagonal elements of  $\mathbf{A}$ . Hence

$$D^2 S^2_i = \frac{2\sigma^4}{n_i} ((n_i - 1)\gamma_2 + 1)$$

and

$$D^2 F = \frac{2\sigma^4}{n_1} ((n_1 - 1)\gamma_2 + 1) D^2 S^{-2}_2$$

It is seen that the variance of the F test linearly depends on the kurtosis of the underlying distribution. To avoid this dependence it is proposed to construct a test with a test statistic

$$R = \frac{x_1' A x_1}{S_2^2}$$

such that diagonal elements of  $A$  equal zero, i.e.  $\mathbf{a}=\mathbf{0}$ . If  $\mathbf{x} \sim N_n(\mu, \sigma^2 \mathbf{I})$ , then  $\mathbf{x}' A \mathbf{x}$  is distributed as  $\sigma^2 \sum_{j=1}^n \lambda_j \chi_j^2(1)$ , where coefficients  $\lambda$ 's are negative as well as positive and  $\chi_j^2(1)$ ,  $j=1, \dots, n$  are independent r.v.'s distributed as chi-square with one d.f. Hypothesis  $H_0: \sigma_1^2 = \sigma_2^2$  is rejected for small negative or big positive values of  $R$ .

It is expected that the size  $\alpha_R(\gamma_2)$  should be more stable than the size  $\alpha_F(\gamma_2)$  of the F test. Those sizes were estimated on the basis of the Monte Carlo experiment.

Levene (1960) proposed a robust test for comparison of many variances of normal distribution. In considered problem this test takes on the form

$$W = \frac{\bar{Z}_1 - \bar{Z}_2}{\sqrt{\sum_{j=1}^{n_1} (Z_{1j} - \bar{Z}_1)^2 + \sum_{j=1}^{n_2} (Z_{2j} - \bar{Z}_2)^2}} \cdot \frac{1}{\sqrt{\frac{1}{n_1 + n_2 - 2} \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

where  $Z_{ij} = |X_{ij} - X_i|$ ,  $i=1, 2$ . The null distribution of the test is (asymptotically) the t distribution with  $n_1 + n_2 - 2$  d.f. The hypothesis is rejected at a level  $\alpha$ , if  $W > t(\alpha; n_1 + n_2 - 2)$ .

In simulation studies the size  $\alpha_L(\gamma_2)$  of the Levene test was also estimated.

### III. MONTE CARLO EXPERIMENT

In Monte Carlo experiment there were considered two cases:  $n_1 = n_2 = 4$  and  $n_1 = n_2 = 5$ . It was taken  $\sigma_1^2 = \sigma_2^2 = 1$ .

For  $n_1 = 4$  the nominator of the robust test was the quadratic form with

$$A = \begin{bmatrix} 0 & 0 & 1 & -1 \\ 0 & 0 & -1 & 1 \\ 1 & -1 & 0 & 0 \\ -1 & 1 & 0 & 0 \end{bmatrix}$$

If  $\mathbf{x}_1$  is distributed as  $N_4(\mu, \sigma_2 \mathbf{I})$ , then  $\mathbf{x}_1' \mathbf{A} \mathbf{x}_1$  is distributed as  $\sigma^2(\chi^2(1) - \chi^2(1))$ . The null distribution of the test statistic R in normal model is given by the cumulative distribution function

$$HR(x) = \begin{cases} \int_0^\infty \int_{-xz}^\infty \int_0^{t+xz} g_1(u)g_1(t)g_3(z)dudtdz, & \text{for } x < 0, \\ \int_0^\infty \int_0^\infty \int_0^{t+xz} g_1(u)g_1(t)g_3(z)dudtdz, & \text{for } x \geq 0, \end{cases}$$

$$= \begin{cases} \frac{1}{2\pi} \int_0^\infty \int_{-xz}^\infty \sqrt{\frac{z}{t}} \exp\left(-\frac{1}{2}(t+z)\right) \text{Erf}\left(\sqrt{\frac{t+xz}{2}}\right) dt dz, & \text{for } x < 0, \\ \frac{1}{2\pi} \int_0^\infty \int_0^\infty \sqrt{\frac{z}{t}} \exp\left(-\frac{1}{2}(t+z)\right) \text{Erf}\left(\sqrt{\frac{t+xz}{2}}\right) dt dz, & \text{for } x \geq 0, \end{cases}$$

where  $g_v(\cdot)$  is a pdf of chi-square distribution with  $v$  d.f. and  $\text{Erf}(x) = \frac{1}{2\pi} \int_0^x e^{-t^2} dt$ .

For  $n_1=5$  the nominator of the robust test was the quadratic form with

$$A = \begin{bmatrix} 0 & 1 & 1 & -1 & -1 \\ 1 & 0 & -1 & 1 & -1 \\ 1 & -1 & 0 & -1 & 1 \\ -1 & 1 & -1 & 0 & 1 \\ -1 & -1 & 1 & 1 & 0 \end{bmatrix}.$$

If  $\mathbf{x}_1$  is distributed as  $N_5(\mu, \sigma_2 \mathbf{I})$ , then  $\mathbf{x}_1' \mathbf{A} \mathbf{x}_1$  is distributed as  $\sigma^2(\chi^2(2) - \chi^2(2))$ . The null distribution of the test statistic R in normal model is given by the cumulative distribution function

$$HR(x) = \begin{cases} \int_0^\infty \int_{-xz}^\infty \int_0^{t+xz} g_2(u)g_2(t)g_4(z)dudtdz = \frac{1}{2(1-x)^2}, & \text{for } x < 0, \\ \int_0^\infty \int_0^\infty \int_0^{t+xz} g_2(u)g_2(t)g_4(z)dudtdz = \frac{1+4x-2x^2}{2(1+x)^2}, & \text{for } x \geq 0, \end{cases}$$

Critical values for considered robust test are given in the Table 1.

Table 1. Critical values

$\alpha$	0.01	0.05	0.10
$n_1=n_2=4$	13.75547	4.17208	2.33421
$n_1=n_2=5$	9.00000	3.47213	2.16228

As a family of nonnormal distributions the Tukey contamination was taken:

$$(1-\varepsilon)N(0,\tau_1^2)+\varepsilon N(0,\tau_2^2), \varepsilon \in [0,1].$$

Parameters were chosen in such a way the variance is one and kurtosis  $\gamma_2$  of the distribution is a given number, i.e. parameters are a solution of the equations:

$$(1-\varepsilon)\tau_1^2+\varepsilon\tau_2^2=1$$

$$3\{(1-\varepsilon)\tau_1^4+\varepsilon\tau_2^4-1\}=\gamma_2$$

Parameters of distributions involved in simulations are given in the Table 2.

Table 2. Parameters  $\tau_1$ ,  $\tau_2$ ,  $\gamma_2$ 

$\varepsilon=0.05$			$\varepsilon=0.10$			$\varepsilon=0.20$		
$\tau_1$	$\tau_2$	$\gamma_2$	$\tau_1$	$\tau_2$	$\gamma_2$	$\tau_1$	$\tau_2$	$\gamma_2$
0.7706	5.3589	3	0.6667	4.0000	3.0	0.5000	3.0000	3.0
0.6756	7.1644	6	0.5918	4.6742	4.5	0.4599	3.1602	3.5
0.6026	8.5498	9	0.5286	5.2426	6.0	0.4226	3.3094	4.0
0.5412	9.7178	12	0.4730	5.7434	7.5	0.3876	3.4495	4.5
0.4870	10.7468	15	0.4226	6.1962	9.0	0.3545	3.5820	5.0
0.4380	11.6771	18	0.3764	6.6125	10.5	0.3230	3.7080	5.5
0.3930	12.5326	21	0.3333	7.0000	12.0	0.2929	3.8284	6.0
0.3511	13.3288	24	0.2929	7.3640	13.5	0.2640	3.9439	6.5
0.3118	14.0767	27	0.2546	7.7082	15.0	0.2362	4.0551	7.0
0.2745	14.7840	30	0.2183	8.0356	16.5	0.2094	4.1623	7.5
0.2391	15.4568	33	0.1835	8.3485	18.0	0.1835	4.2660	8.0
0.2053	16.0997	36	0.1502	8.6485	19.5	0.1584	4.3665	8.5
0.1728	16.7162	39	0.1181	8.9373	21.0	0.1340	4.4641	9.0
0.1416	17.3095	42	0.0871	9.2158	22.5	0.1102	4.5590	9.5
0.1115	17.8819	45	0.0572	9.4853	24.0	0.0871	4.6515	10.0
0.0823	18.4356	48				0.0646	4.7417	10.5
0.0541	18.9722	51				0.0426	4.8297	11.0
0.0267	19.4932	54						

Results of simulations are shown in the Figures 2 and 3. On x--axis there is kurtosis of the underlying distribution while on the y--axis the simulated sizes  $\alpha_F(\gamma_2)$ ,  $\alpha_R(\gamma_2)$  and  $\alpha_L(\gamma_2)$  of the F, robust and Levene test respectively. The assumed significance level in normal model was 0.05.

One observes that the size of the new test is more robust than the size of the standard one and the Levene test. Note that, the size of the Levene test in normal model ( $\gamma_2=0$ ) is not exactly assumed 0.05 because this test is the asymptotic one.

For  $\varepsilon=0.1$  and  $\varepsilon=0.2$  results are similar, so they are not presented here.

In Figures 4 and 5 the powers  $\beta_F(\sigma_1^2/\sigma_2^2)$ ,  $\beta_R(\sigma_1^2/\sigma_2^2)$  and  $\beta_L(\sigma_1^2/\sigma_2^2)$  of the investigated tests are shown for  $\sigma_1^2/\sigma_2^2 \geq 1$ . Of course, the power of the robust test is less than the power of the standard one, but it is the prize for the stabilization of the size. It is interesting, that the power of the Levene test is comparable with the power of the robust test.

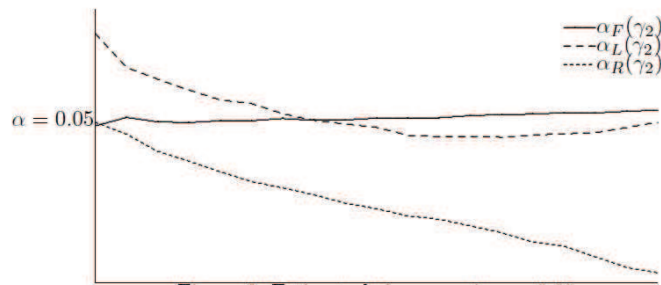


Figure 2. Estimated sizes  $n = 4$ ,  $\varepsilon = 0.05$

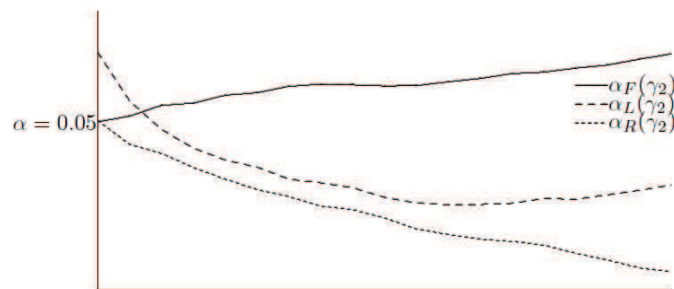
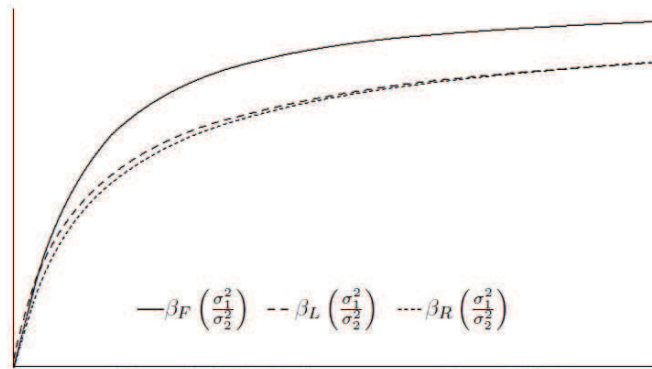
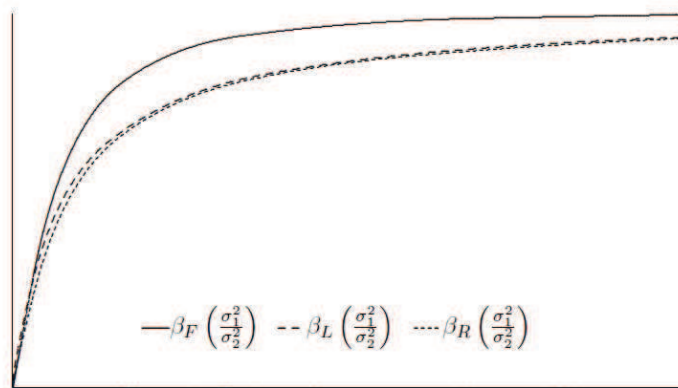


Figure 3. Estimated sizes  $n = 5$ ,  $\varepsilon = 0.05$

Figure 4. Estimated powers  $n = 4$ ,  $\alpha = 0.05$ Figure 5. Estimated powers  $n = 5$ ,  $\alpha = 0.05$ 

#### IV. CONCLUDING REMARKS

It seems that for any  $n$  the statistic of the robust test may be chosen such that the nominator is a quadratic form distributed in the normal model as  $\chi^2\left(\frac{n-1}{2}\right) - \chi^2\left(\frac{n-1}{2}\right)$  for odd  $n$  and  $\chi^2\left(\frac{n}{2}-1\right) - \chi^2\left(\frac{n}{2}-1\right)$  for even  $n$ . Such a quadratic form may be found for a given  $n$  with the aid of an appropriate computer program. The matrix of such a quadratic form may be found in the following manner (Zieliński 1992). Let the matrix  $\mathbf{A}$  of the robust test have eigenvalues  $\lambda_1 = \dots = \lambda_{(n-1)/2} = 1$ ,  $\lambda_{(n+1)/2} = \dots = \lambda_{n-1} = -1$ ,  $\lambda_n = 0$  in case of odd  $n$ , and  $\lambda_1 = \dots = \lambda_{n/2-1} = 1$ ,  $\lambda_{n/2+1} = \dots = \lambda_{n-2} = -1$ ,  $\lambda_{n-1} = \lambda_n = 0$  for  $n$  even. The vectors



$\mathbf{z}_i'=(z_{1i},\dots,z_{ni})$  ( $i=1,\dots,n$ ) in order to be eigenvectors of matrix  $\mathbf{A}$  must satisfy the following system of equations:

n odd:

$$\left\{ \begin{array}{l} \sum_{i=1}^{(n-1)/2} z_{ki}^2 = \sum_{i=(n+1)/2}^{n-1} z_{ki}^2, \quad k=1,\dots,n, \\ \sum_{i=1}^n z_{ki}^2 = 1, \quad k=1,\dots,n, \\ \sum_{i=1}^n z_{ik}z_{il} = 0, \quad k,l=1,\dots,n, \quad k > l, \\ \sum_{i=1}^n z_{ki} = 0, \quad k=1,\dots,n-1, \\ z_{n1} = \dots = z_{nn} = 1/\sqrt{n}; \end{array} \right.$$

n even:

$$\left\{ \begin{array}{l} \sum_{i=1}^{n/2-1} z_{ki}^2 = \sum_{i=n/2+1}^{n-2} z_{ki}^2, \quad k=1,\dots,n, \\ \sum_{i=1}^n z_{ki}^2 = 1, \quad k=1,\dots,n, \\ \sum_{i=1}^n z_{ik}z_{il} = 0, \quad k,l=1,\dots,n, \quad k > l, \\ \sum_{i=1}^n z_{ki} = 0, \quad k=1,\dots,n-1, \\ z_{n-1} + z_n = 1. \end{array} \right.$$

Then  $\mathbf{A}=\mathbf{Z}'\mathbf{\Lambda}\mathbf{Z}$ , where  $\mathbf{Z}'=(\mathbf{z}_1,\dots,\mathbf{z}_n)$  and  $\mathbf{\Lambda}$  is a diagonal matrix with diagonal elements  $\lambda_1,\dots,\lambda_n$ . The quadratic form  $\mathbf{x}'\mathbf{A}\mathbf{x}$  in normal model is distributed as a difference of two independent chi-square random variables.

In the paper there were considered a case of nonnormality of the first sample while the second one remains normal. Of course, the opposite situation is

possible, i.e. the first sample is normal and the second one not. The question is, what is the robust test in such a case. The most general case is in which both samples are not normally distributed. Investigations on the both cases are in progress and will be presented in a separate paper.

#### BIBLIOGRAPHY

- Atiqullah M. (1962), The estimation of residual variance in quadratically balanced least squares problems and the robustness of the F-test, *Biometrika* 49, 83–91.
- Box G. E. P., Andersen S. L. (1955), Permutation theory in the derivation of robust criteria and the study of departures from assumptions, *J. Roy. Statist. Society, Ser. B.* 17, 1–34.
- Brown M. B., Forsythe A. B. (1974), Robust tests for the equality of variances. *Journal of the American Statistical Association*, 69, 264–267.
- Flinger M. A., Killen T. J. (1976), Distribution-free two-sample tests for scale, *J. Amer. Statist. Assoc.*, 71, 210–213.
- Greene W. H. (2000) *Econometric Analysis*, Prentice Hall Inc., New Jersey (ch. 12).
- Layard M. W. J. (1973), Robust large-sample tests for homogeneity of variance, *J. Amer. Statist. Assoc.*, 68, 195–198.
- Levene H. (1960), In *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*, I. Olkin et al., eds. Stanford University Press, Stanford, Calif., 278–292.
- Miller R. G. (1968), Jackknifing variances, *Ann. Math. Statist.*, 39, 567–582.
- Pearson E. S. (1929), The analysis of variance in cases of nonnormal variation, *Biometrika*, 23, 114–133.
- Pearson E. S., Adyanthaya N. K. (1929), The distribution of frequency constants in small samples from non-normal symmetrical and skew populations, *Biometrika*, 21, 259–286.
- Scheffé H. (1959) *The Analysis of Variance*, Wiley, New York.
- Tiku M. L., Balakrishnan N. (1984) Testing equality of population variances the robust way, *Commun. Statist.-Theor. Meth.*, 13, 2143–2159.
- Tsou Tsung-Shan (2003,) Comparing two populations means and variances – a parametric robust way, *Commun. Statist.-Theor. Meth.*, 32, 2013–2029.
- Tsou Tsung-Shan (2006), Parametric robust test for several variances with unknown underlying distribution, *Metrika*, 64, 333–349.
- Zieliński W. (1992), A robust test for variance, *Applicaciones Mathematicae* 21, 441–448

*Wojciech Zieliński*

#### ODPORNY TEST PORÓWNIANIA DWÓCH WARIANCJI

Rozmiar testu F porównania dwóch wariancji jest nieodporny na zaburzenie normalności rozkładu. W pracy zaproponowano test, którego rozmiar jest bardziej odporny niespełnienie założenia o normalności niż testu F.