COMPUTATIONAL METHODS IN SCIENCE AND TECHNOLOGY **15(2)**, 195-201 (2009)

The Three-folded Skewness Test, when a Sample Size is Small

Piotr Sulewski

The Pomeranian Academy, ul. Arciszewskiego 22, 76-200 Słupsk, Poland
e-mail: informpiotr@interia.eu

(Received: 5 February 2009; revised: 16 July 2009; accepted: 16 July 2009; published online: 1 September 2009)

Abstract: The aim of this publication is to present a new goodness-of-fit test oriented toward normal distribution. The test uses three different versions of skewness measures: classic, median and Bowley's skewness. Critical values for these skewness measures at significance level α were determined. The power of the proposed test obtained on the basis of a numerical experiment was compared with the power of the Kolmogorov-Smirnov goodness-of-fit test (K-S test).

Key words: skewness test, empirical skewness, power test function, Visual Basic for Applications

I. INTRODUCTION

Statistical literature gives four different definitions of the skewness. Besides a very popular classic skewness, also Pearson's, median, and Bowley's skewnesses occur sporadically. In paper [6] the ability of these skewness measures toexpress asymmetry was compared, and the accuracy of their estimation from normal distribution was assessed. Testing for normality on the base classic skewness was presented in [5].

The presented article is intended to test for normality on the basis of the three-folded skewness test (TFS test). Classic, median, and Bowley's skewnesses are jointly employed to test whether a particular distribution comes from the normal population. In other words, the null hypothesis says that the actual distribution is a Normal one.

Execution of this type of investigations is possible by means of a computer in relation with a large number of calculations. A very popular Excel spreadsheet was chosen to develop the experimental environment. All calculations were performed by means of Visual Basic for Applications (VBA) procedures [7].

For the null hypothesis to be accepted, each of skewness estimates have to lie within its corresponding critical interval. These critical values were determined at significance level $\alpha \in \{0.01; 0.05; 0.1\}$, and for a sample size $n \in \{5,\ 10,\ 15,\ 20,\ 25,\ 30\}$.

Next, the power of the TFS test was calculated. The power test (PT) function shows how probability of rejecting the null hypothesis increases accordingly to the untruthfulness of the hypothesis. A role of untruth distribution is given to the lognormal distribution. Obtained results were compared with power of the K-S test.

II. THE LOGNORMAL DISTRIBUTION

The lognormal distribution belongs to distributions derived from the Normal distribution. The density function is given by [3]

$$f(x) = \frac{1}{s \cdot x \cdot \sqrt{2\pi}} \cdot \exp\left[-\frac{1}{2} \cdot \left(\frac{\ln(x) - m}{s}\right)^2\right] \quad (x > 0) \quad (1)$$

and cumulative distribution function

$$F(x) = \Phi\left[\frac{\ln(x) - m}{s}\right], \quad (2)$$

where $\Phi(.)$ is a normal cumulative distribution function. Formal similarity (1) to the density function of normal distribution is large; however, parameters m, s should not be interpreted as scale parameter and shape parameter.

The classic skewness of lognormal distribution takes a form

$$\gamma_K = \exp\left(s^2 + 1\right) \cdot \sqrt{\exp\left(s^2 - 1\right)}. \quad (3)$$

The density function of lognormal distribution for combinations of parameter values presented in table 1 was exemplified in Fig. 1. Calculations were performed by means of user function LogNormal, which was created in VBA

```
Function LogNormal(x As Double, m As Double, s As Double) As Double
   Dim pii As Double, mian As Double, wyk As Double
   Let pii = Application.pi()
   Let mian = x * s * sqr(2 * pii)
   Let wyk = (Log(x) - m) / s
   Let LogNormal = (1 / mian) * Exp(-0.5 * wyk ^ 2)
End Function
```

Table 1. Combinations of parameter values and values of the classic skewness

Combination	m	s	γ_K
I	0	0.641	2.5
II	0	0.920	5
III	0	1.165	10
IV	0	1.289	15
V	0	1.369	20

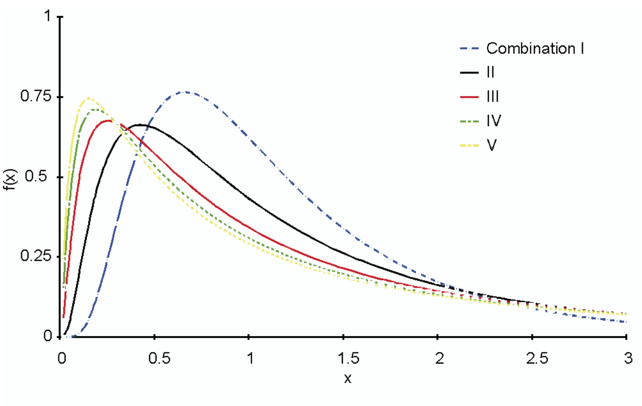

Fig. 1. Density function of lognormal distribution for combinations of parameter values presented in Table 1

For simulations to be performed, the use of random numbers is necessary. The generator of lognormal random numbers (function LogNormLos) arose from the generator of normal random numbers (function NormLos).

```
Function LogNormLos(m As Double, s As Double) As Double
   Dim wynik As Double
   Let wynik = NormLos(m, s)
   Let LogNormLos = Exp(wynik * s + m)
End Function
Function NormLos(m As Double, s As Double) As Double
   Dim i As Integer, wynik As Double, sum As Double
   Let sum = 0
   For i = 1 To 12
      Let sum = sum + Rnd
   Next i
   Let NormLos = s * (sum - 6) + m
End Function
```

III. THE EMPIRICAL CLASSIC SKEWNESS

The empirical classic skewness is calculated as [1]

$$\gamma_C^* = \frac{n \cdot \sqrt{n-1}}{n-2} \cdot \frac{\sum_{i=1}^{n}\left(x_i^* - \hat{\alpha}_1\right)^3}{\left[\sum_{i=1}^{n}\left(x_i^* - \hat{\alpha}_1\right)^2\right]^{3/2}}, \tag{4}$$

where $x_i^* \left(i = \overline{1,n}\right)$ are values of the random variable lognormal distributed as well as $\hat{\alpha}_1$ is a sample mean value

$$\hat{\alpha}_1 = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i^* . \tag{5}$$

The computer implementation of estimation of the classic skewness written in VBA was presented below. Comments were placed after apostrophes.

```
Sub ClassicSkewness()
'declaration of tables
Dim x() As Double
'declaration of variables
Dim mc2 As Double, mc3 As Double
Dim m As Double, s As Double
Dim i As Long, sum As Double
Dim c As Double, n As Long
Dim classic As Double, ave As Double

Randomize Timer
Worksheets("estimate").Select        'selecting worksheets "estimate"
'introduction of cells to variables
Let m = Cells(2, 1).Value            'parameter of lognormal distribution
Let s = Cells(2, 2).Value            'parameter of lognormal distribution
Let n = Cells(2, 3).Value            sample size
c = (n * Sqr(n - 1)) / (n - 2)
ReDim x(n)

Let sum = 0
  For i = 1 To n
    Let x(i) = LogNormLos(m, s)
    Let sum = sum + x(i)
  Next i
Let ave = sum / n

Let mc2 = 0
Let mc3 = 0
  For i = 1 To n
    Let mc2 = mc2 + (x(i) - ave) ^ 2
    Let mc3 = mc3 + (x(i) - ave) ^ 3
  Next i
Let classic = (c * mc3) / mc2 ^ (1.5)       'classic skewness
'introduction of result to cell
Let Cells(6, 1) = classic
End Sub
```

IV. THE EMPIRICAL MEDIAN SKEWNESS

The empirical median skewness is given by [6]

$$\gamma_M^* = \frac{\sqrt{n-1} \cdot \left(\hat{\alpha}_1 - x_{0.5}^*\right)}{\sqrt{\sum_{i=1}^{n}\left(x_i^* - \hat{\alpha}_1\right)^2}}, \quad (6)$$

where $x_{0.5}^*$ is a sample median. Unknown values of quantiles of k-th order were replaced by appropriate order statistics [2]

$$number = \text{int}\left[n * k\right] + 1 . \quad (7)$$

The computer implementation of estimation of the median skewness written in VBA was presented below.

```
Sub MedianSkewness()
Dim x() As Double
Dim mc2 As Double
Dim m As Double, s As Double
Dim i As Long, sum As Double
Dim n As Long, q2 As Double
Dim median As Double, ave As Double
Dim h As Byte, b As Double

Randomize Timer
Worksheets("estimate").Select
Let m = Cells(2, 1).Value
Let s = Cells(2, 2).Value
Let n = Cells(2, 3).Value
ReDim x(n)

Let sum = 0
  For i = 1 To n
    Let x(i) = LogNormLos(m, s)
    Let sum = sum + x(i)
  Next i
Let ave = sum / n

'sorting
powrot:
Let h = 0
  For i = 1 To n - 1
    If (x(i) <= x(i + 1)) Then GoTo dalej
    Let b = x(i)
    Let x(i) = x(i + 1)
    Let x(i + 1) = b
    Let h = 1
dalej:
  Next i
If h = 1 Then GoTo powrot

'the sample median
Let q2 = x(Int(n * 0.5)) + 1

Let mc2 = 0
  For i = 1 To n
    Let mc2 = mc2 + (x(i) - ave) ^ 2
  Next i
Let median = Sqr(n - 1) * (ave - q2) / Sqr(mc2)          'median skewness
Let Cells(6, 3) = median
End Sub
```

V. THE EMPIRICAL BOWLEY'S SKEWNESS

The empirical Bowley's skewness is defined as [4]

$$\gamma_B^* = \frac{\left(x_{0.75}^* - x_{0.5}^*\right) - \left(x_{0.5}^* - x_{0.25}^*\right)}{\left(x_{0.75}^* - x_{0.25}^*\right)}, \quad (8)$$

where $x_k^* \left(0 < k < 1\right)$ are sample quantiles of the k-th order.

The computer implementation of estimation of the Bowley's skewness written in VBA was presented below.

```
Sub BowleySkewness()
Dim x() As Double
Dim m As Double, s As Double
Dim i As Long, n As Long
Dim q1 As Double, q2 As Double
Dim q3 As Double, bowley As Double
Dim h As Byte, b As Double

Randomize Timer
Worksheets("estimate").Select
Let m = Cells(2, 1).Value
Let s = Cells(2, 2).Value
Let n = Cells(2, 3).Value
ReDim x(n)

For i = 1 To n
  Let x(i) = LogNormLos(m, s)
Next i

'sorting
powrot:
Let h = 0
  For i = 1 To n - 1
    If (x(i) <= x(i + 1)) Then GoTo dalej
    Let b = x(i)
    Let x(i) = x(i + 1)
    Let x(i + 1) = b
    Let h = 1
dalej:
  Next i
If h = 1 Then GoTo powrot

'the sample quantiles
Let q1 = x(Int(n * 0.25)) + 1
Let q2 = x(Int(n * 0.5)) + 1
Let q3 = x(Int(n * 0.75)) + 1

Let bowley = ((q3 - q2) - (q2 - q1)) / (q3 - q1)          'Bowley's skewness
Let Cells(6, 4) = bowley
End Sub
```

VI. THE GOODNESS-OF-FIT TEST APPLICABLE TO THE NORMAL DISTRIBUTION

In this paper, three skewness measures described in the previous sections are jointly employed to test whether a particular distribution comes from the Normal population.

In other words, the null hypothesis says that the actual distribution is a Normal one.

The essence of this joint employment is as follows. Three measures of skewness are estimated from the sample subjected to test. For the null hypothesis to be accepted, each of the skewness estimates have to lie within its corresponding critical interval $\left(X_i, X_j\right)$.

The first Monte-Carlo study on which this paper is based was intended to determine these critical values. The study consisted in generating samples coming from the Normal distribution. Such a great number of samples as 10 240 enabled precise determination of critical values X_i and X_j at significance level α according to the following formulas [2]

$$i = \text{int}\left[10\,240 \cdot \frac{\beta}{2}\right] + 1,$$

$$j = \text{int}\left[10\,240 \cdot \left(1 - \frac{\beta}{2}\right)\right] + 1, \qquad (9)$$

where $\beta = 1 - \left(1 - \alpha\right)^{1/3}$. Obtained results at significance level α are presented in Tables 2-4.

Table 2. Critical values at significance level $\alpha = 0.01$

Simple size	Classic skewness		Median skewness		Bowley's skewness	
n	X_i	X_j	X_i	X_j	X_i	X_j
5	-2.219	2.120	-0.646	0.710	-0.997	0.999
10	-2.134	2.010	-0.727	0.477	-0.967	0.803
15	-1.886	1.704	-0.570	0.470	-0.880	0.808
20	-1.728	1.482	-0.530	0.362	-0.729	0.734
25	-1.414	1.376	-0.408	0.438	-0.704	0.673
30	-1.497	1.374	-0.399	0.385	-0.664	0.595

Table 3. Critical values at significance level $\alpha = 0.05$

Simple size	Classic skewness		Median skewness		Bowley's skewness	
n	X_i	X_j	X_i	X_j	X_i	X_j
5	-2.026	2.018	-0.611	0.609	-0.984	0.980
10	-1.687	1.657	-0.625	0.395	-0.909	0.707
15	-1.405	1.448	-0.435	0.425	-0.725	0.706
20	-1.266	1.262	-0.448	0.323	-0.651	0.667
25	-1.084	1.137	-0.344	0.341	-0.606	0.589
30	-0.986	1.012	-0.360	0.281	-0.597	0.509

Table 4. Critical values at significance level $\alpha = 0.1$

Simple size	Classic skewness		Median skewness		Bowley's skewness	
n	X_i	X_j	X_i	X_j	X_i	X_j
5	-1.822	1.843	-0.561	0.579	-0.959	0.974
10	-1.564	1.593	-0.586	0.356	-0.888	0.672
15	-1.186	1.326	-0.385	0.385	-0.669	0.630
20	-1.002	1.125	-0.389	0.271	-0.553	0.623
25	-0.956	0.971	-0.302	0.329	-0.546	0.544
30	-0.870	0.933	-0.309	0.228	-0.545	0.409

Having critical values, one can start performing the second Monte-Carlo study intended to determine the power of the test. A substance PT function can be verbally expressed by saying that it is the function which shows how the probability of rejecting the null hypothesis increases accordingly to the untruthfulness of the hypothesis. A role of untruth distribution is given to the lognormal distribution. The computer procedure of the TFS test written in VBA was introduced below.

```
Sub TFS_test()
Dim x() As Double
Dim skw() As Double
Dim wk(6) As Double
Dim mc2 As Double, mc3 As Double
Dim m As Double, s As Double
Dim i As Long, sum As Double
Dim c As Double, n As Long, ave As Double
Dim h As Byte, b As Double, k As Long
Dim q1 As Double, q2 As Double, q3 As Double
Dim u As Long, power As Double, wiersz As Byte

Randomize Timer
Worksheets("estimate").Select
Let m = Cells(2, 1).Value
Let s = Cells(2, 2).Value
Let n = Cells(2, 3).Value
ReDim x(n)
ReDim skw(10240, 3)
Let c = (n * Sqr(n - 1)) / (n - 2)
For k = 1 To 10240
   'generating of pseudo random numbers
  For i = 1 To n
   Let x(i) = LogNormLos(m, s)
   Next i
   'sorting
powrot:
  Let h = 0
  For i = 1 To n - 1
   If (x(i) <= x(i + 1)) Then GoTo dalej
            Let b= x(i)
            Let x(i) = x(i + 1)
            Let x(i + 1) = b
            Let h = 1
```

```
dalej:
   Next i
   If h = 1 Then GoTo powrot
   'classic skewness
   Let sum = 0
      For i = 1 To n
         Let sum = sum + x(i)
      Next i
   Let ave = sum / n
   Let mc2 = 0
    Let mc3 = 0
   For i = 1 To n
       Let mc2 = mc2 + (x(i) - ave) ^ 2
       Let mc3 = mc3 + (x(i) - ave) ^ 3
   Next i
   Let skw(k, 1) = (c * mc3) / mc2 ^ (1.5)

    'median skewness
   Let q2 = x(Int(n * 0.5) + 1)
   Let skw(k, 2) = Sqr(n - 1) * (ave - q2) / Sqr(mc2)
  'Bowley's skewness
   Let q1 = x(Int(n * 0.25) + 1)
   Let q3 = x(Int(n * 0.75) + 1)
   Let skw(k, 3) = ((q3 - q2) - (q2 - q1)) / (q3 - q1)
Next k
Select Case n
   Case 5
      Let wiersz = 6
   Case 10
      Let wiersz = 7
   Case 15
      Let wiersz = 8
   Case 20
      Let wiersz = 9
   Case 25
      Let wiersz = 10
   Case 30
      Let wiersz = 11
End Select
'critical values written in table
Let wk(1) = Cells(wiersz, 2)
Let wk(2) = Cells(wiersz, 3)
Let wk(3) = Cells(wiersz, 4)
Let wk(4) = Cells(wiersz, 5)
Let wk(5) = Cells(wiersz, 6)
Let wk(6) = Cells(wiersz, 7)
u = 0
For k = 1 To 10240
   If (skw(k, 1) > wk(1) And skw(k, 1) < wk(2)) And _
      (skw(k, 2) > wk(3) And skw(k, 2) < wk(4)) And _
      (skw(k, 3) > wk(5) And skw(k, 3) < wk(6)) Then
      Let u = u + 1
   End If
Next k
'power of test
Let power = (10240 - u) / 10240
MsgBox "Power of the three-folded skewness test is " & power
End Sub
```

The classic skewness (3) was chosen as an argument of the PT function shown in Figures 2-4.

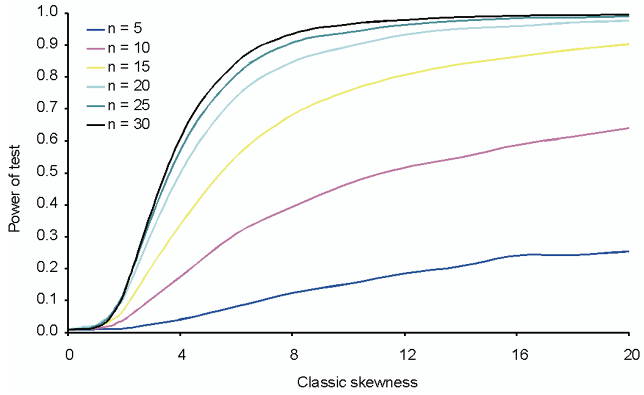

Fig. 2. Power of TFS test for a different sample size, at significance level $\alpha = 0.01$

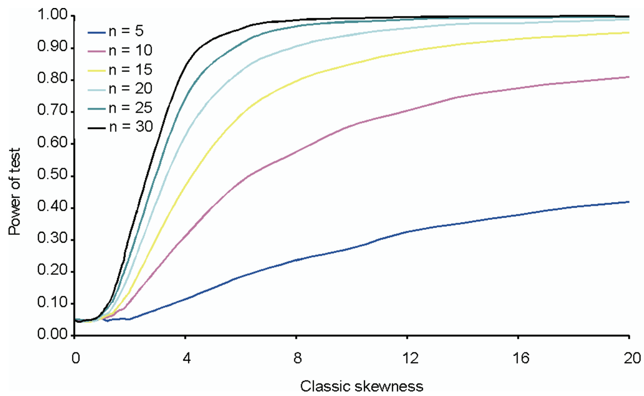

Fig. 3. Power of TFS test for a different sample size, at significance level $\alpha = 0.05$

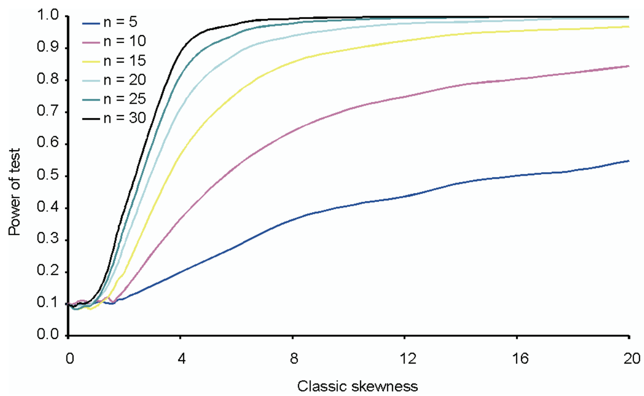

Fig. 4. Power of TFS test for a different sample size, at significance level $\alpha = 0.1$

The TFS test was compared to the K-S test [8] for which an appropriate Monte Carlo study was carried out in parallel. This test is well-known and therefore only its results were shown. Better performance of the TFS test is readily seen (Figures 5-10).

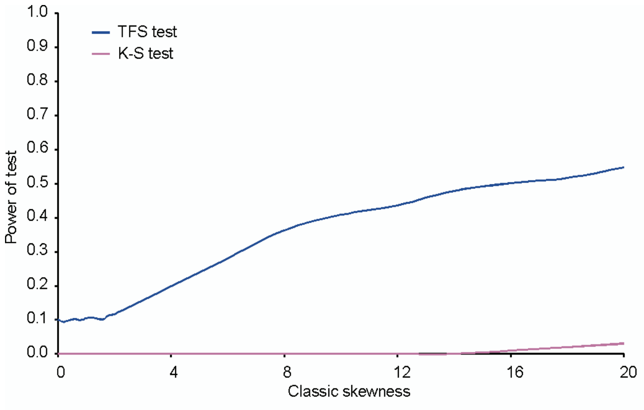

Fig. 5. Power of tests for a sample size $n = 5$, at significance level $\alpha = 0.1$

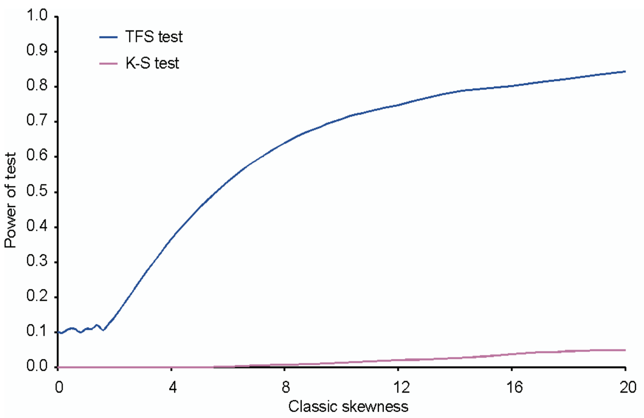

Fig. 6. Power of tests for a sample size $n = 10$, at significance level $\alpha = 0.1$

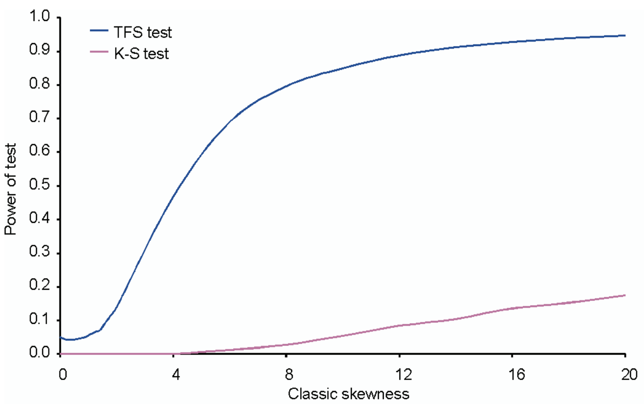

Fig. 7. Power of tests for a sample size $n = 15$, at significance level $\alpha = 0.05$

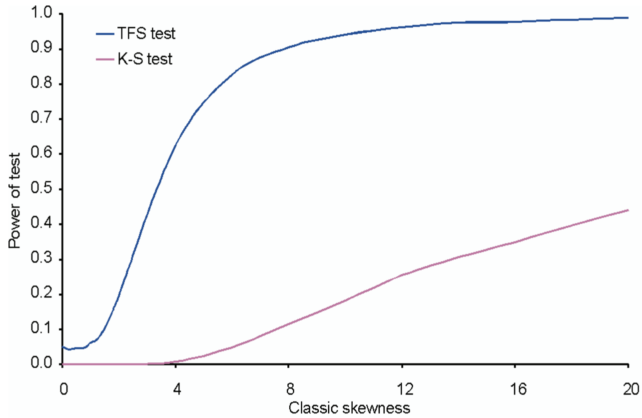

Fig. 8. Power of tests for a sample size $n = 20$, at significance level $\alpha = 0.05$

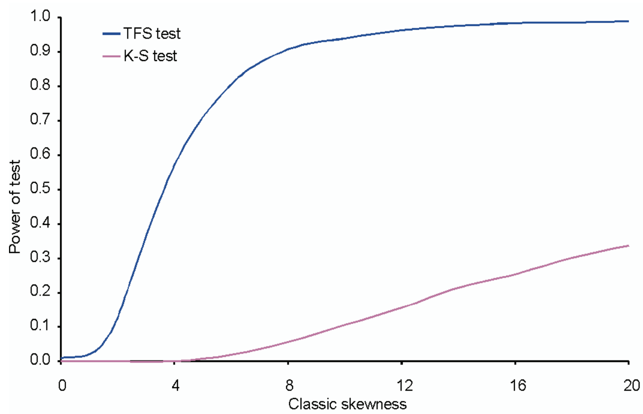

Fig. 9. Power of tests for a sample size $n = 25$, at significance level $\alpha = 0.01$

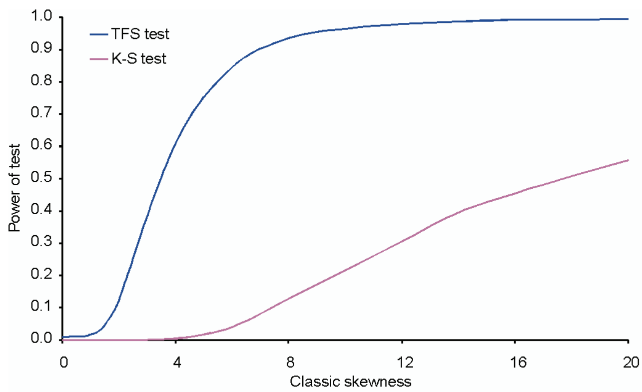

Fig. 10. Power of tests for a sample size $n = 30$, at significance level $\alpha = 0.01$

References

[1] H. Cramer, *Mathematical methods in statistics*, Warsaw 1958.
[2] H.A. David, *Order statistics*, Wiley, New York 1970.
[3] A. Drapella, *Statistical inference on the base skewness and kurtosis*, Pomeranian Pedagogical Academy, Słupsk 2004.
[4] J.F. Kenney, E.S. Keeping, *Mathematics of statistics*, Pt. 1, 3rd ed. Princeton, NJ: Van Nostrand 101-102 (1962).
[5] P. Sulewski, *Testing for normality on the base skewness and kurtosis*, Słupskie Prace Matematyczno – Fizyczne nr 3, 45-59 (2005).
[6] P. Sulewski, *On differently defined skewness*, Computational Methods in Science and Technology 14(1), 39-46 (2008).
[7] J. Walkenbach, *Excel 2007 Power Programming with VBA*, John Wiley & Sons 2001.
[8] W. Zieliński, *Statistical tables*, SGGW, Warsaw 2000.

Piotr Sulewski graduated in Mathematics in 1996. Since then he has been working at the Institute of Mathematics at Pomeranian Academy in Słupsk. He received the PhD in reliability theory in 2001 from the Systems Research Institute of Polish Academy of Sciences in Warsaw. His research interests concern reliability mathematics and computational methods in statistics.