ESTIMATION OF THE DISTRIBUTION OF α-STABLE RETURN RATES OF STOCK MARKET INDICES BASED ON THE CRITERION OF MINIMIZATION OF CHI-SQUARE STATISTICS

Kamila Bednarz-Okrzyńska

University of Szczecin, Faculty of Management and Economics of Services, POLAND
e-mail: kamila.bednarz@wzieu.pl

Received 10 December 2018
Accepted 28 December 2018

JEL classification D53, G11

Keywords return rates, alpha-stable distribution, chi-square test

Abstract

One of the most frequently considered problems related to the capital market is the appropriate modelling of the distributions of rates of return for specific financial instruments. The results of such modelling are often used as an element of a number of tools and methods used for analyzes, diagnoses and forecasts of specific phenomena occurring on financial markets. An adoption a priori of certain assumptions as to the density function of distribution of return rates, seems to be a highly risky approach. A significant deviation of the actual rates of return from the assumed ones may cause a number of negative consequences, including among others that it may be the basis for questioning the credibility and thus the applicability of a number of techniques, methods and models used for analyzes, diagnoses and forecasts of the capital market. The main objective of the study will be to determine the impact of the change in the optimization criterion when estimating the parameters of the stable distribution, on the probability of obtaining a distribution consistent with the theoretical. In addition, the potential impact on this probability of such factors as the adoption of a specific assumption regarding the method of construction of individual numerical intervals or the inclusion of a specific rate of return will also be examined.

Introduction

An element inseparably connected with research conducted on the capital market are issues related to proper modelling of specific financial variables, including in particular the returns on financial instruments. The simplest and at the same time the most commonly used approach assuming the normality of the rates of return, despite high practicality, is unacceptable from a theoretical point of view.

If the deviations from the normal distribution of return rates are large, they may call into question the utility of a number of tools developed within the framework of modern financial theory. For the above reasons, the question of the probability distribution of return rates from shares is an extremely important issue, although there is still no universal way to solve this problem. Therefore, for a long time it has been the subject of interest for both Polish researchers (Tomasik, Echaust, 2008, pp. 34–66; Bednarz, 2012, pp. 103–113; Czyżycki, 2013, pp. 1530–1535) as well as foreign (Piasecki, Tomasik, 2013 work on this subject). There are many schedules that can replace a normal distribution: t-distribution and skrew t-distribution, GED, Laplace distribution. One of the more frequently used is also the alpha-stable distribution.

Unfortunately, apart from three cases (normal distribution, Cauchy distribution, Levy distribution), there are not known explicit forms of distribution of the density function of stable distributions, which is the basic problem in the use of this class of statistical distributions. In the literature on the subject there are a number of proposals related to the estimation of stable distribution parameters, including estimation of parameters by moments method (Press, 1972, pp. 842–846; Fielitz, Rozelle, 1981, pp. 303–320; Kuruoglu, 2001, pp. 2192–2201), by the quantile method (Fama, Roll, 1971, pp. 331–338; McCulloch, 1986, pp. 1109–1136), using the regression function (Koutrouvelis, 1980, pp. 918–928; Koutrouvelis, 1981, pp. 17–28), or on the basis of the Maximum Likelihood Method (Paulson, Holcomb, Leitch, 1975, pp. 163–170). The conducted simulation tests indicate that the worst approximation properties characterize the method of moments, then the regression and quantile method, while the most accurate estimation results are obtained using the method of the Maximum Likelihood Method (Borak, Härdle, Weron, 2005, pp. 21–44). The estimation of parameters of alpha-stable distributions by the Maximum Likelihood Method is not significantly different from the estimation by this method of the distribution parameters of other classes. Having the given observation vector $x = \{x1, x2, ..., xn\}$, the estimation of the parameter vector θ = (α, β, μ, σ) of the alpha-stable distribution by the Maximum Likelihood Method is obtained by maximizing the logarithm of the likelihood function: estimation of the parameter vector θ = (α, β, μ, σ) of the alpha-stable distribution by the Maximum Likelihood Method is obtained by maximizing the logarithm of the likelihood function:

$$\Phi(t) = \begin{cases} \exp\left[i\mu \times t - \sigma^{\alpha} |t|^{\alpha} \times \left(1 - i\beta \times \mathrm{sgn}(t) \times \mathrm{tg}\left(\frac{\alpha \cdot \pi}{2} \right) \right) \right] & \text{if} \quad \alpha \neq 1 \\ \exp\left[i\mu \times t - \sigma |t| \times \left(1 + i\beta \times \mathrm{sgn}(t) \times \frac{\pi}{2} \times \ln(t) \right) \right] & \text{if} \quad \alpha = 1 \end{cases},$$

where: $\alpha \in \left(0, 2\right\rangle$ – an index of stability (tail index, tail exponent of characteristic exponent); $\beta \in \left\langle -1, 1 \right\rangle$ – a skewness parameter; $\sigma > 0$ – a scale parameter; $\mu \in R$ – a location parameter; $i = \sqrt{-1}$; $\mathrm{sgn}(t) = \frac{1}{|t|}$.

The rate of return on a given financial instrument can be defined either as an arithmetic rate of return or as a logarithmic rate of return. In the literature on modelling rates of return we can meet both forms; moreover, studies conducted so far do not indicate any particular advantage in modelling the distribution of any of these returns (Czyżycki, 2016, pp. 19–29; Bednarz-Okrzyńska, 2014, pp. 11–25).

As previously indicated, the most commonly used test to assess the fit of the theoretical distribution to the empirical distribution of the rate of return is the chi-square test, defined as:

$$\chi^2 = \sum_{j=1}^{k} \frac{\left(n_j - \hat{n}_j\right)^2}{\hat{n}_j},$$

where: n_j – empirical size in j^{th} interval; $\hat{n}_j$ – theoretical size in j^{th} interval; k – number of numerical intervals, wherein the most common assumption is that the width of all ranges is the same. It is possible, however, to assume the existence of the same empirical size or the same theoretical size in particular intervals. Irrespective of the assumptions regarding the method of constructing numerical intervals, the concordance of the theoretical distribution with the empirical distribution is the greater, the smaller the differences between the empirical and theoretical sizes in individual intervals, i.e. the lower the value of the chi-square statistics.

In order to verify such hypotheses, modelling of the following returns was carried out: a daily simple rate of return formed during the period of 25 quotations (monthly estimation period) (Rt _25); a daily logarithmic rate of return formed during the period of 25 quotations (R*t _25); a daily simple rate of return formed over a period of 125 quotations (half-year estimation period) (Rt _ 125); a daily logarithmic rate of return formed during the period of 125 quotations (R*t _ 125); a daily simple rate of return formed during the period of 250 quotations (annual estimation period) (Rt _250); a daily logarithmic rate of return shaped over a period of 250 quotations (R*t _250); a daily simple rate of return formed during the 1,250 quotations (five-year estimation period) (Rt _1250); a daily logarithmic rate of return formed during the period of 1,250 quotations (R*t _1250).

The above rates of return will be modelled from the first quotation in which such modelling is possible (e.g. for a daily return rate covering a period of 250 quotations, the first model was obtained for the 251st quotation of a given index, the second model for 252nd quotation included return rates from quotations from 2 to 251 etc.) until the quotation taking place on June 28, 2018. This means that for the S & P500 index, for each of the analyzed distributions, 12,208 models were obtained for the monthly estimation period, 12,108 for the half-year estimation period, 11,983 models for the annual estimation period and 10,983 models for the five-year estimation period. In the case of the WIG index, the number of models obtained was at 5,913, 5,813, 5,688 and 4,688, respectively.

Findings

In order to verify the hypothesis concerning the impact of the adopted optimization criterion on the probability of obtaining a distribution consistent with the considered stable distribution, the parameters for the previously indicated all variants were estimated. In the next step, based on the Chi-square compatibility test, it was checked for how many of the n distributions obtained there are no grounds to reject the hypothesis that they are consistent with the considered theoretical distribution (p-value = 0.05). Then, based on the test for two structure indices, the p-value level was determined, at which it can be assumed that there is a statistically significant difference in the frequency (probability) of obtaining a distribution consistent with the assumed, taking into account a function of density and Chi-square statistics as a function of likelihood. Based on the results of the above tests contained in tables 1–3, the following conclusions can be drawn: (1) the probability of obtaining a distribution consistent with the assumed one depends on the adopted method of construction of the interval series - this is particularly evident in the case of a short estimation period. The lowest probability occurs when considering the equal width of the numerical intervals, the largest – in the case of assuming the same empirical size in each range; (2) irrespective of the adopted assumption regarding the method of building a numerical series, the probability of obtaining a distribution consistent with the assumed significantly increases in the case of adopting in the estimation of the parameters of the stable distribution the criterion of minimization of chi-square statistics, instead of the likelihood function described in the formula 1 which is conventionally used in this role. An increase of this probability is inversely proportional to the length of the observation vector (of the length of the estimation period).

Table 1. Results of modelling of selected rates of return using the Maximum Likelihood Method, adopting in the function of likelihood the function of density of stable distribution (ML (1)) and chi-square statistics (ML (2)) and the equal range of the intervals of the numerical series

Distribution	n	Number of consistent distributions		p-value for test for two structure indicators
		ML(1)	MNW(2)	
SP500_Rt_25	12,208	8,735	9,795	0.000E+00
SP500_Rt*_25	12,208	8,760	9,778	0.000E+00
WIG _Rt_25	5,913	4,032	4,567	0.000E+00
WIG _Rt*_25	5,913	4,033	4,566	0.000E+00
SP500_Rt_125	12,108	10,253	10,454	2.431E-04
SP500_Rt*_125	12,108	10,243	10,473	2.630E-05
WIG _Rt_125	5,813	5,009	5,116	3.082E-03
WIG _Rt*_125	5,813	5,022	5,126	3.786E-03
SP500_Rt_250	11,983	10,086	10,192	5.775E-02
SP500_Rt*_250	11,983	10,306	10,400	7.653E-02
WIG _Rt_250	5,688	5,164	5,218	7.299E-02
WIG _Rt*_250	5,688	5,147	5,213	3.002E-02
SP500_Rt_1250	10,983	8,279	8,334	3.874E-01
SP500_Rt*_1250	10,983	8,299	8,357	3.607E-01
WIG _Rt_1250	4,688	3,670	3,722	1.886E-01
WIG _Rt*_1250	4,688	3,669	3,721	1.887E-01

Source: author's own study.

Table 2. Results of modelling selected rates of return using the Maximum Likelihood Method, taking into account in the function of likelihood the function of density of stable distribution (ML (1)) and chi-square statistics (ML (2)) and equal empirical size in particular intervals of the numerical series

Distribution	n	Number of consistent distributions		p-value for test for two structure indicators
		ML(1)	ML(2)	
1	2	3	4	5
SP500_Rt_25	12,208	11,290	11,995	0.000E+00
SP500_Rt*_25	12,208	11,281	11,992	0.000E+00
WIG _Rt_25	5,913	5,426	5,779	0.000E+00
WIG _Rt*_25	5,913	5,430	5,779	0.000E+00
SP500_Rt_125	12,108	10,847	11,002	7.964E-04
SP500_Rt*_125	12,108	10,295	10,466	1.678E-03
WIG _Rt_125	5,813	5,035	5,112	3.210E-02
WIG _Rt*_125	5,813	5,027	5,353	0.000E+00
SP500_Rt_250	11,983	10,305	10,394	9.384E-02
SP500_Rt*_250	11,983	10,306	10,400	7.653E-02

1	2	3	4	5
WIG _Rt_250	5,688	5,161	5,218	5.877E-02
WIG _Rt*_250	5,688	5,165	5,214	1.042E-01
SP500_Rt_1250	10,983	8,104	8,294	3.208E-03
SP500_Rt*_1250	10,983	8,159	8,323	1.057E-02
WIG _Rt_1250	4,688	3,670	3,722	1.886E-01
WIG _Rt*_1250	4,688	3,669	3,721	1.887E-01

Source: author's own study.

Table 3. Results of modelling selected rates of return using the Maximum Likelihood Method, taking into account in the function of likelihood the function of density of stable distribution (ML (1)) and chi-square statistics (ML (2)) and equal theoretical size in particular intervals of the numerical series

Distribution	n	Number of consistent distributions		p-value for test for two structure indicators
		ML(1)	ML(2)	
SP500_Rt_25	12,208	9,064	9,790	0.000E+00
SP500_Rt*_25	12,208	8,946	10,317	0.000E+00
WIG _Rt_25	5,913	4,104	4,761	0.000E+00
WIG _Rt*_25	5,913	4,423	4,584	5.116E-04
SP500_Rt_125	12,108	10,444	10,626	5.039E-04
SP500_Rt*_125	12,108	10,467	10,660	2.010E-04
WIG _Rt_125	5,813	5,323	5,439	4.100E-05
WIG _Rt*_125	5,813	5,034	5,115	2.408E-02
SP500_Rt_250	11,983	10,383	10,521	7.578E-03
SP500_Rt*_250	11,983	10,306	10,400	7.653E-02
WIG _Rt_250	5,688	5,161	5,218	5.877E-02
WIG _Rt*_250	5,688	5,165	5,214	1.042E-01
SP500_Rt_1250	10,983	7,814	8,067	1.365E-04
SP500_Rt*_1250	10,983	8,299	8,357	3.607E-01
WIG _Rt_1250	4,688	3,670	3,722	1.886E-01
WIG _Rt*_1250	4,688	3,669	3,721	1.887E-01

Source: author's own study.

Based on the information contained in figures 1–3, which present graphically the differences in the frequency of obtained distributions consistent with the assumed (positive value means the advantage of the distributions obtained assuming minimization of chi-square statistics as a function of likelihood, while negative value – the advantage of distributions obtained assuming maximizing the function of likelihood, taking into account the function of density of the stable distribution), the hypothesis on a higher probability of obtaining a distribution consistent with the stable distribution can be additionally confirmed with the assumption of minimizing chi-square statistics.

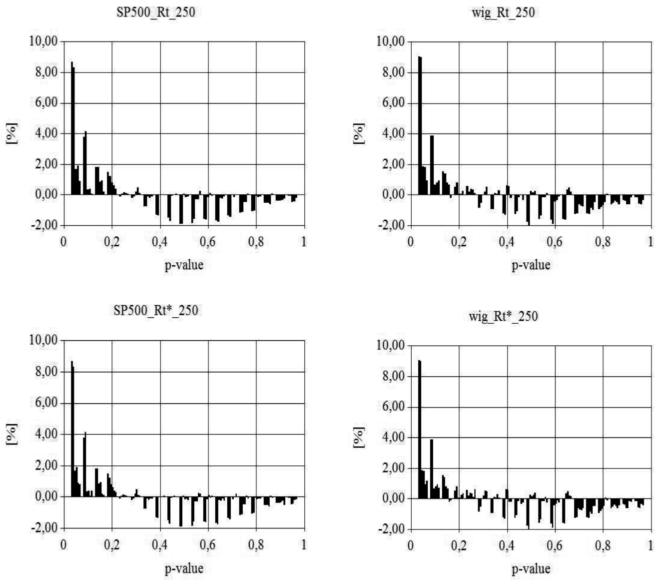

Figure 1. Differences in the frequency of obtained distributions consistent with the stable distribution with a differently defined function of likelihood in the Maximum Likelihood Method, the estimation period of 250 quotations and the equal range of the intervals of the numerical series

Source: author's own study.

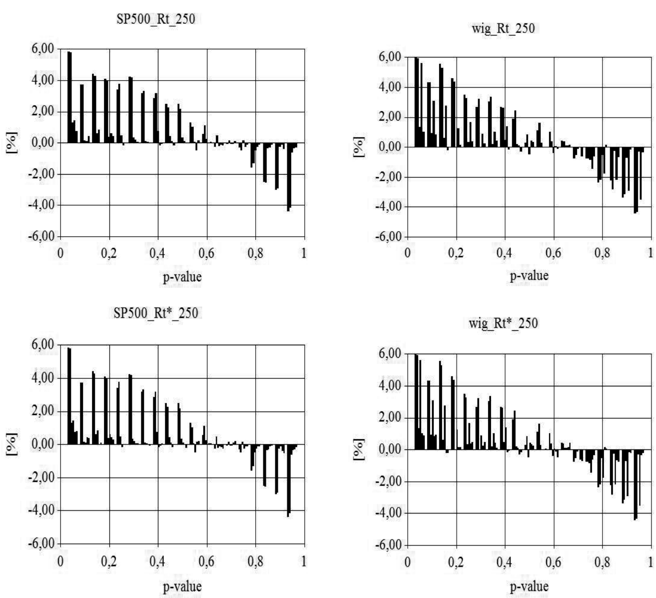

Figure 2. Differences in the frequency of obtained distributions consistent with the stable distribution with a differently defined function of likelihood in the Maximum Likelihood Method, the estimation period of 250 quotations and equal empirical size in particular intervals of the numerical series

Source: author's own study.

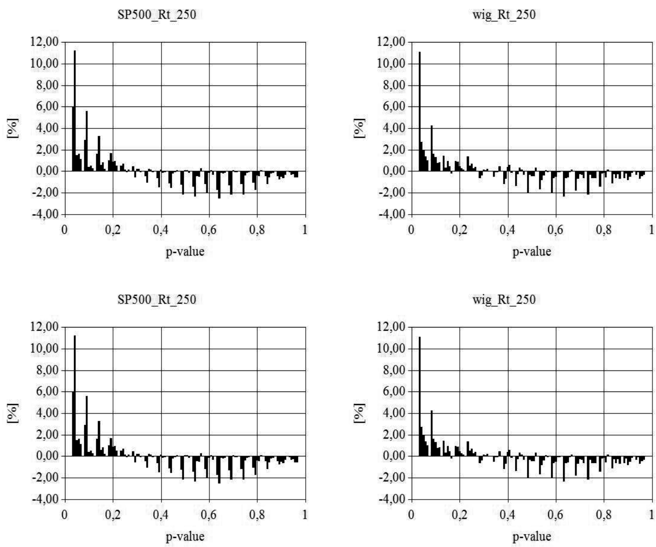

Figure 3. Differences in the frequency of obtained distributions consistent with the stable distribution with a differently defined function of likelihood in the Maximum Likelihood Method, the estimation period of 250 quotations and equal theoretical size in particular intervals of the numerical series

Source: author's own study.

In order to verify the additional hypothesis (H2), which states that the probability of obtaining a distribution consistent with the assumed one is the same in developed and developing capital markets, the results of modelling these rates for the SP500 index and WIG index were compared. However, the results obtained in this respect are ambiguous. In the case of a fixed spread of numerical intervals, the probability of obtaining a distribution consistent with the assumed one was different in a statistically significant way for both markets irrespective of the optimization criterion adopted. In the case of assuming a constant empirical size in individual intervals, the lack of such a difference can be seen in logarithmic modelling of the rate of return for the monthly estimation period (n = 25), while with the same theoretical sizes in particular intervals of the distribution series, the lack of differences in the probability of obtaining a distribution consistent with the assumed one occurs in the case of a logarithmic rate of return and half-year estimation period (n = 125), taking into account the classical likelihood function and the arithmetic rate of return for the monthly estimation period and the logarithmic rate of return and half-year estimation period in the case of chi-square statistics as a function of likelihood. Detailed information on the results of the conducted study is presented in tables 4–6.

Table 4. Values of p for the test for two structure indices, examining the frequency of occurrence of distributions consistent with the assumed one for the rates of return from the WIG index and SP500 assuming an equal spread of the intervals of the numerical series (based on data from Table 1)

Distribution	p-value for test for two structure indicators	
	ML(1)	ML(2)
Rt_25	3.290E-06	3.069E-06
Rt*_25	8.713E-07	7.902E-06
Rt_125	8.645E-03	1.935E-03
Rt*_125	1.539E-03	1.663E-03
Rt_250	0.000E+00	0.000E+00
Rt*_250	0.000E+00	0.000E+00
Rt_1250	9.128E-05	1.750E-06
Rt*_1250	2.673E-04	7.606E-06

Source: author's own study.

Table 5. Values of p for the test for two structure indices, examining the frequency of occurrence of distributions consistent with the assumed one for the rates of return from the WIG index and SP500 assuming an equal empirical size in particular intervals of the numerical series (based on data from Table 2)

Distribution	p-value for test for two structure indicators	
	ML(1)	ML(2)
Rt_25	9.088E-02	1.633E-02
Rt*_25	1.754E-01	2.269E-02
Rt_125	4.630E-09	1.149E-09
Rt*_125	9.753E-03	0.000E+00
Rt_250	0.000E+00	0.000E+00
Rt*_250	0.000E+00	0.000E+00
Rt_1250	2.445E-09	1.475E-07
Rt*_1250	1.174E-07	1.050E-06

Source: author's own study.

Table 6. Values of p for the test for two structure indices, examining the frequency of occurrence of distributions consistent with the assumed one for the rates of return from the WIG index and SP500 assuming an equal theoretical size in particular intervals of the numerical series (based on data from Table 3)

Distribution	p-value for test for two structure indicators	
	ML(1)	ML(2)
Rt_25	7.162E-12	6.069E-01
Rt*_25	2.902E-02	0.000E+00
Rt_125	0.000E+00	0.000E+00
Rt*_125	7.804E-01	9.254E-01
Rt_250	6.217E-15	4.663E-15
Rt*_250	0.000E+00	0.000E+00
Rt_1250	0.000E+00	2.887E-15
Rt*_1250	2.673E-04	7.606E-06

Source: author's own study.

Conclusions

On the basis of the conducted considerations, it seems justified to propose the following conclusions:

1. In the case of estimating the parameters of the stable distribution with the Method of Maximum Likelihood, using the minimization of chi-square statistics instead of the usual maximization of the likelihood function based on the density function of the stable distribution significantly increases the probability of obtaining a distribution consistent with the assumed one.
2. In the case of chi-square compatibility test, an adopted assumption related to the method of determining empirical and theoretical sizes in particular intervals of the distribution series has a significant effect on the probability of obtaining a distribution consistent with the assumed one. The lowest probability occurs in the case of accepting the equal width of the numerical intervals, the largest - in the case of assuming the same empirical size in each interval.
3. The probability of obtaining a distribution of the return rate, consistent with the theoretical one, does not depend significantly on the rate of return adopted for modelling. Both in the case of the simple and the logarithmic rate of return, the chance of obtaining a distribution consistent with the theoretical is the same regardless of the length of the assumed horizon of estimation.
4. The degree of development of the capital market is significant. In the case of emerging markets, which is undoubtedly the Polish capital market, the probability of obtaining a distribution consistent with the modelled one is significantly smaller compared to the modelling of indices of developed stock markets (SP500).

References

Bednarz, K. (2012). Goodness of fit tests in modeling the distribution of the daily rate of return of the WIG20 companies. *Folia Oeconomica Stetinensia*, *10* (18), 103–113.

Bednarz-Okrzyńska, K. (2014). Modelowanie empirycznych rozkładów stóp zwrotu z akcji notowanych na Giełdzie Papierów Wartościowych w Warszawie za pomocą logarytmicznej i klasycznej stopy zwrotu. *Studia i Prace WNEiZ US*, *36* (2), 11–25.

Borak, S., Härdle, W., Weron, R. (2005). Stable distributions. Statistical tools for finance and insurance. *Springer*, 21–44.

Czyżycki, R. (2013). Using GED (Generalized Error Distribution) for modeling distribution of the rates of return. Hradec Kralove, The Czech Republic: International Masaryk Conference for Ph.D. Students and Young Researchers, 1530–1535.

Czyżycki, R. (2016). Arithmetic or logarithmic rate of return? The impact of the choice made on the distribution modelling results. In: A. Sokół, I. Figurska, K. Drela (eds.), *Contemporary socio-economic issues and problems*. Bratislava: KARTPRINT.

Fama, E.F., Roll, R. (1971). Parameter estimates for symmetric stable distributions. *Journal of the American Statistical Association*, *66* (334), 331–338.

Fielitz, B.D., Rozelle, J.P. (1981). Method-of-moments estimators of stable distribution parameters. *Applied Mathematics and Computation*, *8* (4), 303–320.

Koutrouvelis, I.A. (1980). Regression-type estimation of the parameters of stable laws. *Journal of the American statistical association*, *75* (372), 918–928.

Koutrouvelis, I.A. (1981). An iterative procedure for the estimation of the parameters of stable laws: An iterative procedure for the estimation. *Communications in Statistics-Simulation and Computation*, *10* (1), 17–28.

Kuruoglu, E.E. (2001). Density parameter estimation of skewed alpha-stable distributions. *IEEE Transactions on signal processing*, *49* (10), 2192–2201.

McCulloch, J.H. (1986). Simple consistent estimators of stable distribution parameters. *Communications in Statistics-Simulation and Computation*, *15* (4), 1109–1136.

Paulson, A.S., Holcomb, E.W., Leitch, R.A. (1975). The estimation of the parameters of the stable laws. *Biometrika*, *62* (1), 163–170.

Piasecki, K., Tomasik, E. (2013). *Rozkłady stop zwrotu z instrumentów polskiego rynku kapitałowego.* Kraków–Warszawa: edu-Libri.

Press, S.J. (1972). Estimation in univariate and multivariate stable distributions. *Journal of the American Statistical Association*, *67* (340), 842–846.

Tomasik, E., Echaust, K. (2008). Wybrane rozkłady prawdopodobieństwa w modelowaniu empirycznych stóp zwrotu akcji notowanych na GPW w Warszawie. In: W. Sikora (ed.), *Z prac Katedry Badań Operacyjnych*. Poznań: Wydawnictwo Akademii Ekonomicznej.

Cite this article as: Bednarz-Okrzyńska, K. (2018). Estimation of the distribution of α-stable return rates of stock market indices based on the criterion of minimization of chi-square statistics. *European Journal of Service Management*, *4* (28/2), 55–64. DOI: 10.18276/ejsm.2018.28/2-06.