

Bootstrap Estimation Methods of Value at Risk

Author: Dorota Pekasiewicz

Abstract

Interval bootstrap methods can be used to estimate Value at Risk, defined as a quantile of fixed order of random variable being the value of losses from investments. These methods are applied when there is no information about the distribution class of the variable considered, which is the advantage of bootstrap methods compared with parametric methods. Semiparametric estimation procedures are of particular importance. They can be used in the estimation of high-order quantiles. They guarantee the occurrence of large values in the generated bootstrap samples. The paper presents nonparametric and semiparametric bootstrap estimation methods and the results of simulation studies for higher-order quantiles of a heavy-tailed distribution. The application of the methods analysed provide confidence intervals with greater accuracy compared to the nonparametric classical method. The procedures under discussion are used in *VaR* estimation of daily returns of selected shares at Warsaw Stock Exchange.

Keywords: bootstrap estimation, accuracy of estimation, Value at Risk
JEL: C13, C14, C15

History: Otrzymano 2015-11-21, poprawiono 2016-06-11, zaakceptowano 2016-07-05

Introduction

In investment activity we are faced with a variety of risks, including market risk (currency rate, interest rate, stock prices, prices of goods), as well as credit, operational, liquidity, legal or event risk. Recent years have seen a very rapid development of methods involved in measuring those risks. Viewing risk as a random variable with a continuous distribution, we can

divide risk measurements into differentiated groups (K. Jajuga, 2007, pp. 39-48). One of them are quantiles of distribution, amongst which there is *VaR* – *Value at Risk*, a very popular risk measurement in recent years. In classical approach, in order to estimate market risk using the *VaR*, it is necessary to know the price distribution or the rate of return of the instruments investigated. The present assumptions on the normality of the

distribution of variables considered frequently fail to be correct. In many cases we come across heavy-tailed distributions, which results in greater probabilities that substantial losses will occur compared to the situations where variables are characterized by normal distributions. This leads to significant estimation errors. Lack of information on the distribution class of the price of instruments or their rate of return, and the application of the quantile of the empirical distribution for point estimation of value at risk is also affected by a large error. However, the solution to this problem may lie in providing interval evaluation instead of point evaluation and in applying the non-classical methods.

The paper demonstrates the application proposals regarding bootstrap interval estimation procedures for estimating value at risk, defined on the basis of the concept of the appropriate quantile-order when the distribution of the variable under study is unknown. Bootstrap confidence intervals for quantile Q_p of order p of the variable X distribution represent an alternative for the "exact" nonparametric confidence interval form (Zieliński, 2008):

$$P(X_{(r)}^{(n)} \leq Q_p \leq X_{(s)}^{(n)}) = \sum_{i=r}^{s-1} \binom{n}{i} p^i (1-p)^{n-i} = 1 - \alpha \tag{1}$$

where $X_{(r)}^{(n)}, X_{(s)}^{(n)}$ are order statistics with ranks, r and s , respectively, determined based on a random variable X_1, X_2, \dots, X_n , whereas $1 - \alpha$ is a fixed likelihood coefficient.

The accuracy of the quantile Q_p estimation obtained using formula (1), measured by the length of the confidence interval obtained, for very small and very large orders p may prove to be unsatisfactory, particularly in the case of heavy-tailed distributions. The bootstrap methods allow

obtaining a more accurate estimations of quantiles.

Among bootstrap methods, one can distinguish parametric, nonparametric and semiparametric methods. For estimating higher-order quantiles, semiparametric methods are of considerable relevance, since they use the information on the distribution class of the right tail of the variable with which the size of losses is identified, and further, they guarantee the occurrence of large values of the random variable in the bootstrap samples generated. As a counterpart, if the estimation refers to lower-order quantiles, it is important that small values of the left tail of the distribution of the variable tested occur in bootstrap samples. The simulation analysis, whose results have been presented in the paper, allow for certain conclusions concerning the effectiveness of bootstrap procedures, especially of the semiparametric bootstrap method.

Risk Measurement Using VaR Method

The value at risk measure is defined as such loss in the market value of a financial instrument, portfolio or investment that the probability of reaching or exceeding it, in a certain time interval, is equal to the pre-determined tolerance level. Thus, if W_t, W_{t+1} denote the price of the instrument under discussion or the investment value, respectively, at the moment t and $t+1$, the VaR can be expressed as follows:

$$P(W_{t+1} \leq W_t - VaR) \leq p, \tag{2}$$

where $p \in (0,1)$ is the fixed tolerance level.

If the loss distribution, that is the distribution of the variable $X_t = W_{t+1} - W_t$, has been determined using a known distribution function F , then the Value at Risk can be determined based on the formula below:

$$VaR = -F^{-1}(p). \tag{3}$$

This approach stems from the static risk management, where only unconditional distribution of the random variable is analysed. The methods of value at risk estimation thus defined have been shown, e.g. in works of: T. Bałamut, 2002; C. Domański, 2011, pp. 142-152.

In literature there are more complex methods of VaR estimation illustrated using information on the properties of the distribution of rates of return on shares, i.e. leptokurtosis, the left tail of the distribution is heavier than the right tail (which implies that the probability of having large, atypical changes of the rates is bigger than if the rates of return were to come from a normal distribution), skewness of the rates of return distribution (the distributions which are most frequently observed are right skewed), volatility aggregation effect (which means that both small and large shifts in rates occur in series), or the effect of a negative correlation between the level of rates and the volatility level, the so called leverage effect. Owing to the specificity of the methods examined in the paper, that kind of properties were not taken into account in the analysis.

Bootstrap Methods of Quantile Estimation

Bootstrap methods are applied in the estimation of parameters of a random variable’s distribution, when there is no information about the distribution class and when it is not possible to use the asymptotic properties of estimators of the parameters which are being estimated.

Let X be a random variable of an unknown distribution F with which we identify the population under study, and let X_1, X_2, \dots, X_n be a simple sample drawn from this population. Moreover, let $1 - \alpha$ be the likelihood coefficient determined.

One of the nonparametric bootstrap methods of estimating order p quantile of the random variable X is the percentile method (C. Domański, K. Pruska, 2000, pp. 262-264). The method involves generating, on the basis of the random sample X_1, X_2, \dots, X_n , N bootstrap samples $X_1^*, X_2^*, \dots, X_n^*$, according to the bootstrap distribution given by:

$$P(X^* = x_i) = \frac{1}{n}, \quad \text{for } i = 1, 2, \dots, n, \quad (4)$$

where x_1, x_2, \dots, x_n are elements of the drawn sample.

Following that, on the basis of each bootstrap sample, the quantile $X_{p,k}^*$ of order p is determined, where $k = 1, 2, \dots, N$. The sequence of sorted quantiles $X_{p,1}^*, \dots, X_{p,N}^*$ thus obtained allows approximating the sample p quantile distribution. The quantiles of $\frac{\alpha}{2}$ order and $1 - \frac{\alpha}{2}$ that is statistics

$$X_{\frac{\alpha}{2};N}^* \quad \text{and} \quad X_{1-\frac{\alpha}{2};N}^*$$

are used to construct the confidence interval for Q_p quantile, given by:

$$P\left(X_{\frac{\alpha}{2};N}^* < Q < X_{1-\frac{\alpha}{2};N}^*\right) \approx 1 - \alpha. \quad (5)$$

The number of replications N is selected so that $\frac{N\alpha}{2}$ may be an integer, then

the quantile $X_{\frac{\alpha}{2};N}^*$ is the value of the $X_{p,1}^*, \dots, X_{p,N}^*$, whereas quantile $X_{1-\frac{\alpha}{2};N}^*$ is

the value of the order statistic of rank $N - \frac{N\alpha}{2}$.

This way of estimation may not be effective when estimating Q_p quantiles for large and small values of p because the bootstrap samples which are generated may not include observations of large and small values, respectively, which will not allow obtaining a “good” approximation of the distribution tail, and, by inference, a “good” estimation of the distribution of the

quantile considered. In those situations, the semiparametric approach can be of help.

The semiparametric estimation method is distinguished by double bootstrap simulations. While estimating high-order quantiles, $n-k$ bootstrap sample values below the fixed threshold u are generated from the empirical distribution F_n , whereas k values, which are over the threshold, from the distribution that takes into account the asymptotic properties of the tail distribution (M. D. Pandey et al., 2001, 2003).

In this case the distribution function of the bootstrap distribution has the following form:

$$F^*(x|u) = \begin{cases} (1 - F_n(u))F_0(x) + F_n(u), & \text{if } x > u, \\ F_n(x), & \text{if } x \leq u, \end{cases} \quad (6)$$

where F_n denotes the empirical distribution function, while F_0 denotes the distribution function of the generalized Pareto distribution $GPD(\xi, \beta)$.

One usually assumes as the threshold value u , the value of at least 0.9 quantile.

Since the distribution function of the generalized Pareto distribution is expressed by the following formula:

$$F_0(x) = \begin{cases} 1 - \left(1 + \frac{\xi x}{\beta}\right)^{-\frac{1}{\xi}} & \text{for } \xi \neq 0 \\ 1 - \exp\left(-\frac{x}{\sigma}\right) & \text{for } \xi = 0 \end{cases} \quad (7)$$

then the estimated distribution function (6) is given by:

- for $\xi \neq 0$

$$\hat{F}^*(x|u) = 1 - \frac{k}{n} \left(1 + \frac{\hat{\xi}(x-u)}{\hat{\beta}}\right)^{-\frac{1}{\hat{\xi}}} \text{ for } x > u, \quad (8)$$

- for $\xi = 0$:

$$\hat{F}^*(x|u) = 1 - \frac{k}{n} \exp\left(-\frac{x-u}{\hat{\beta}}\right), \quad (9)$$

where k denotes the number of elements over the fixed threshold value in the sample consisting of n -elements,

with $\hat{\xi}, \hat{\beta}$ being the estimators of parameters ξ, β respectively.

Therefore p quantiles have the following form:

$$X_{p;n} = \begin{cases} u + \frac{\hat{\beta}}{\hat{\xi}} \left(\left(\frac{n}{k}(1-p) \right)^{-\hat{\xi}} - 1 \right) & \text{for } \hat{\xi} \neq 0, \\ u - \hat{\beta} \ln \left(\frac{n}{k}(1-p) \right) & \text{for } \hat{\xi} = 0. \end{cases} \quad (10)$$

The values of parameters ξ, β are estimated using the moment method or probability-weighted moment method, conditional on the information on the existence of the first and second order- moments (D. Pekasiewicz, 2015, pp. 75–89). When $\xi = 0$, parameter β can be estimated using the maximum likelihood method.

Comparative Analysis of the Estimation Accuracy of the Quantiles Obtained Using Bootstrap Methods

The aim of the studies conducted is to compare the length of confidence intervals generated using the bootstrap methods and the classical nonparametric estimation method (formula (1)), and to compare the estimated probabilities of covering the theoretical quantile value by the intervals obtained.

The at least 0.95 quantiles were estimated for random variables characterized by heavy-tailed distributions. For the analyses, the following distributions have been chosen:

- Student's t $S(k)$, where k is the number of degrees of freedom,
- Pareto $Pa(\theta, a)$, where $\theta, a > 0$,
- Log-gamma $LG(p, \lambda, \mu)$, where $p, \lambda > 0, \mu \in \mathbb{R}$

The choice of the distribution classes is connected with the various properties of the distributions – Student's t is a symmetric distribution, whereas Pareto and Log-gamma distribution are characterized by their asymmetry. Moreover, in financial analyses, for

example, for the approximation of the distributions of the rates of return on shares, Student's t distribution and a skew- Student's t distribution are applied (cf. Piontek, 2005). Thus, the results of the analyses focusing on the estimation of the quantiles of the Student's t distribution used to determine VaR are significant. Also, the results obtained for the asymmetric distribution quantiles are also relevant as the empirical distributions of logarithmic rates of return have precisely this kind of feature.

The samples comprising at least 1000 elements were included in the analyses. The size of samples must be large because it is necessary to estimate the parameters of the generalized Pareto distribution. For ξ, β parameters estimation, the probability-weighted moment method and moment method were used (J.M. Landwehr et al., 1979).

The construction of the nonparametric and semiparametric bootstrap confidence intervals has the effect that these methods can be examined and compared only through simulation analyses; therefore, the confidence interval length and the probability γ that the true value of quantile is within the interval determined, were estimated by having the estimation procedure repeated a 1000 times. This probability should be roughly equal to the fixed likelihood coefficient $1 - \alpha$. For the classic nonparametric estimation procedure, the confidence interval length was determined by computing the expected values of the order statistics $X_{(985)}^{(1000)}, X_{(998)}^{(1000)}$, being the lower and upper end of the confidence interval, respectively. The following formula was used:

where $g_{(i)}^{(n)}(x)$ is a density function of the order statistic $X_{(i)}^{(n)}$, whereas f, F are, respectively, the density function and distribution function of the variable X .

The statistics ranks were determined such that the estimation likelihood was equal 0.9495.

Table 1 contains the results of the studies carried out with a view to evaluate the efficiency of the bootstrap estimation method for 0.99 quantiles of the selected distributions, with the confidence coefficient equal to 0.95. The size of the sample was 1000 elements. The parameters were selected in such a way as to include distributions which have and those which do not have the expected value. For the Pareto and Log-gamma distributions, the expected value equals 3, while for the Student's t distribution 0, if, of course, it exists.

The findings of the studies show that for the distributions which were analysed the semiparametric estimation methods proved to provide the most accurate estimations of 0.99 quantiles. The confidence intervals obtained using the nonparametric percentile bootstrap method were characterized by having a slightly bigger span. Thus, for estimating 0.99 quantiles of the Pareto, Log-gamma or Student's t distributions we may expect a much greater precision in estimations when applying bootstrap methods compared to the nonparametric method based on order statistics determined on the basis of the value of the sample drawn. Moreover, value $\gamma(II)$ proved to be slightly closer to the determined confidence coefficient than $\gamma(I)$. Value $\gamma(III)$ was not distinguished in Table 1, for it equals exactly 0.9495, which results from the confidence interval construction.

$$E\left(X_{(i)}^{(n)}\right) = \int_{-\infty}^{\infty} x g_{(i)}^{(n)}(x) dx = \frac{n!}{(i-1)!(n-i)!} \int_{-\infty}^{\infty} x [F(x)]^{i-1} [1-F(x)]^{n-i} f(x) dx, i = 985, 998, n = 1000 \quad (11)$$

Table 1: The confidence interval length and estimation likelihood of 0.99 quantiles

Distribution	Nonparametric bootstrap percentile method		Semiparametric bootstrap method		Classic method
	$d(I)$	$\gamma(I)$	$d(II)$	$\gamma(II)$	$d(I)$
$S(3)$	2.1148	0.940	1.9098	0.955	3.7491
$S(3)$	4.5764	0.910	4.0748	0.960	9.2162
$S(3)$	40.6448	0.931	29.0684	0.892	137.9481
$Pa(2; 3)$	3.8876	0.937	3.4622	0.949	6.9966
$Pa(1.5, 2)$	9.4868	0.924	8.1446	0.935	19.3847
$Pa(1; 1.5)$	19.0926	0.929	15.0672	0.944	41.2171
$LG(1; 0.5; 2)$	6.4478	0.931	5.5168	0.952	12.9231
$LG(2; 0.5; 0)$	19.8290	0.921	17.3168	0.939	42.8796
$LG(1; 0.75; 0)$	31.6229	0.924	23.6356	0.902	77.5623

Source: Author's own study.

The problem, however, constitute super-heavy tailed distributions of the $LG(1; 0.75; 0)$, $S(1)$, or more generally $S(k)$ for $k \leq 1$, for which the application of bootstrap methods increases the estimation accuracy, yet the estimation likelihood decreases. This type of distributions need further investigation and modification of the estimation procedures.

Bootstrap Method Application in VaR Estimation of the Rate of Returns of Selected WSE Indices

The bootstrap methods under discussion were applied to estimate value at risk for logarithmic daily rates of return on shares of Clothing Company "Bytom" (Zakłady Odzieżowe Bytom) and Computer Works ELZAB (Zakłady Urządzeń Komputerowych ELZAB S.A.). Variable $R_t = \ln \frac{P_t}{P_{t-1}}$ was analysed,

where P_t, P_{t-1} denote share prices at the time t and $t-1$, respectively.

Fig. 1 and 2 illustrate changes in daily logarithmic rates of return on shares between 29 March 2011 and 16 October 2015. The data used were obtained from <http://stooq.pl/q/?s=btm> oraz <http://stooq.pl/q/?s=elz>.

In estimating value at risk for daily logarithmic rates of return on shares traded on the WSE, the approaches chosen were constrained to the following: nonparametric and semiparametric, building on the logarithmic unconditional distributions of the rates of return on shares.

According to formula (3), for estimating value at risk low-order quantiles (order p) should be used, that is, those from the left tail of the empirical distribution of the logarithmic rates of return. In order to be able to employ the semiparametric bootstrap method, a variable $Y_t = -R_t$ was introduced and $(1-p)$ th quantiles were estimated,

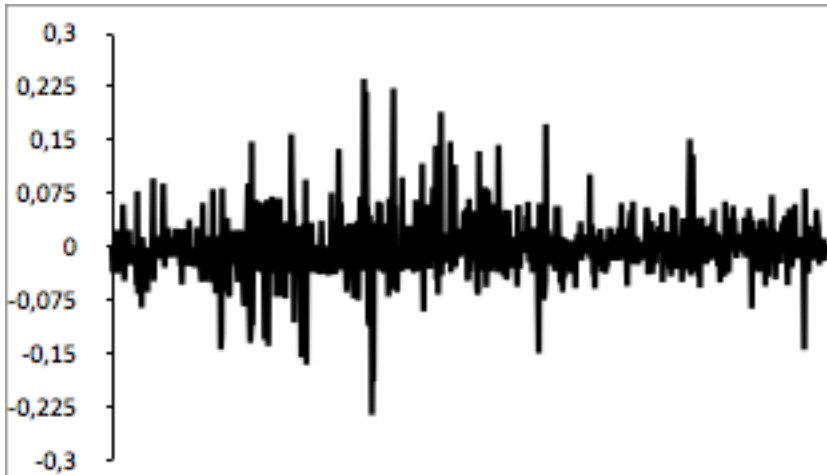


Figure 1: Daily logarithmic rates of return on shares of Clothing Company "Bytom" between 29.03.2011 – 16.10.2015

Source: Author's own study.

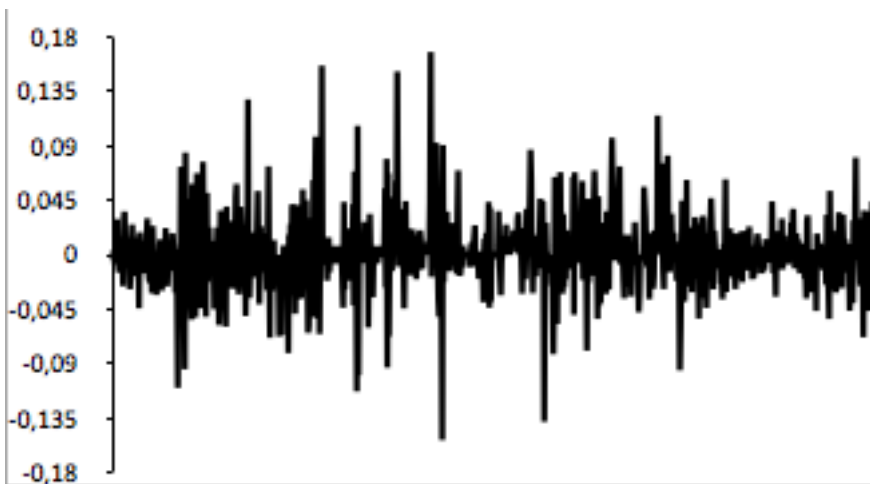


Figure 2: Daily logarithmic rates of return on shares of Computer Works ELZAB between 29.03.2011 – 16.10.2015

Source: Author's own study.

that is, the quantiles from the right tail of the empirical distribution of the random variable Y_t . Introducing the variable Y_t allowed for using the generalised Pareto distribution for the

estimation of the right tail distribution (over 0,9 quantile).

In the application of a semiparametric bootstrap estimation method, it is necessary to decide as to the form of

the distribution function used for generating the bootstrap samples. Choosing either (7) or (8) formula has to be based on the information about the distribution Y_i (whether it is a heavy-tailed distribution, i.e. $\hat{\xi} > 0$, or a thin-tailed distribution, i.e. $\hat{\xi} = 0$) or it has to be based on the extreme index estimated value $\hat{\xi}$.

For the preliminary estimation of parameter ξ a moment estimator was applied in the form:

$$\hat{\xi}_{m,n}^M = M_m^{(1)} + 1 - \frac{1}{2} \left(1 - \frac{(M_m^{(1)})^2}{M_m^{(2)}} \right)^{-1}, \quad (12)$$

where $M_m^{(l)} = \frac{1}{m} \sum_{i=1}^m \left(\ln X_{(n-i+1)}^{(n)} - \ln X_{(n-m)}^{(n)} \right)^l$

for $l = 1, 2$, whereas order statistics $X_{(n-i)}^{(n)}$ were determined based on the n -elements simple sample X_1, X_2, \dots, X_n . For the variable Y_i determined on the basis of the daily logarithmic rates of return on shares of Clothing Company "Bytom", $\hat{\xi}=0.1832$ was obtained, while for the variable associated with the rates of return of shares of Computer Works ELZAB, it was $\hat{\xi}=0.2639$. Since $\hat{\xi} < 0.5$, for the estimation of the parameters of the generalized Pareto distribution the moment method was used.

The confidence intervals for value at risk of investing in the shares of the companies under discussion at the tolerance level of 0.01, and the interval lengths obtained have been included in Table 2.

Conclusion

The application of bootstrap methods allows obtaining the interval estimation of quantiles which are used to determine value at risk. The simulation analysis refer to the situation when the random variables specifying the amount of losses are characterized by distributions which belong to the distribution group with heavy tails. In the majority of cases analysed, the bootstrap estimation methods are more effective compared to the classic nonparametric method, for the confidence intervals obtained have much smaller span while having similar estimation likelihood. The application of the semiparametric bootstrap method proved to be the most effective, for the confidence intervals which were obtained had the smallest span. Moreover, the probabilities with which the intervals, obtained through the semiparametric method, covered

Table 2. Confidence intervals for VaR of the selected WSE shares

Type of shares	Estimation method	Confidence interval	Confidence interval length
Clothing Company „Bytom”	Bootstrap percentile	(0.0984; 0.1591)	0.0607
	semiparametric	(0.1064; 0.1515)	0.0451
	classic	(0.1161; 0.2183)	0.1022
Computer Works ELZAB S.A.	Bootstrap percentile	(0.0733; 0.1068)	0.0335
	semiparametric	(0.0785; 0.1058)	0.0273
	classic	(0.0995; 0.1517)	0.0722

Source: Author’s own study

the estimated value of the quantile were closer to the likelihood coefficient determined than the equivalent probabilities for the percentile method. The only thing that is needed is to choose properly the threshold value for a given sample size. The number of elements in excess of the threshold cannot be too small for then large errors occur, associated with the estimation of parameters of the generalized Pareto distribution, which is used to estimate the distribution tail of the variable analysed. In the cases demonstrated, this number was 100, since the threshold was determined at the level of 0.9 quantile, proving to be sufficient in size.

In some situations the increase in the estimation accuracy as the result of the bootstrap method application involved the decrease in the estimation likelihood, e.g. for the analyses of

random variables with very heavy distributions, e.g. Cauchy ($S(1)$), or other Student's t distributions with the number of freedom degrees being not more than 1, and thus extreme index ξ being bigger than 1.

The results of the empirical tests are consistent with the results of the simulation analyses. The accuracy of the interval estimation of value at risk for daily logarithmic rates of return on shares of Clothing Company "Bytom" and Computer Works ELZAB is the greatest if the semiparametric bootstrap method is applied. For other stocks quoted on the Warsaw Stock Exchange the distributions of their daily logarithmic rates of return are not characterized by high values of index ξ , thus, it appears advisable to apply the bootstrap methods outlined in the paper for the estimation of value at risk.

Bibliography

Bałamut T. (2002), *Metody estymacji Value at Risk*, „Materiały i Studia”, 147, Narodowy Bank Polski, Warszawa.

Domański C. (ed.) (2011), *Nieklasyczne metody oceny efektywności ryzyka*, Polskie Wydawnictwo Ekonomiczne, Warszawa.

Domański C., Pruska K. (2000), *Nieklasyczne metody statystyczne*, Polskie Wydawnictwo Ekonomiczne, Warszawa.

Jajuga K. (2007), *Zarządzanie ryzykiem*, Wydawnictwo Naukowe PWN, Warszawa.

Landwehr, J. M., Matalas, N. C., Wallis J. R. (1979) *Probability Weighted Moments Compared with Some Traditional Techniques in Estimating Gumbel Parameters and Quantiles*,

Water Resources Research 15(5), pp. 1055–1064.

Pandey M. D., Van Gelder P. H. A. J. M., Vrijling J. K. (2001), *The Estimation of Extreme Quantiles of Wind Velocity Using Safety*”, 23, pp. 179–192.

Pandey M. D., Van Gelder P. H. A. J. M., Vrijling J. K. (2003), *Bootstrap Simulations for Evaluating the Uncertainty Associated with Peaks-over-Threshold Estimates of Extreme Wind Velocity*, *Environmetrics*”, 14, pp. 27–43.

Piontek, K., (2005), *Modelowanie własności szeregów stóp zwrotu – skośność rozkładów*, *Ekonometria*, Prace Naukowe Akademii Ekonomicznej we Wrocławiu, Wrocław.

Pekasiewicz D. (2015), *Statystyki pozycyjne w procedurach estymacji i ich*

zastosowania w badaniach społeczno-ekonomicznych, Wydawnictwo Uniwersytetu Łódzkiego, Łódź.

Zieliński W. (2008), *Przykład zastosowania dokładnego nieparametrycznego przedziału ufności dla VaR*, „Metody Ilościowe w Badaniach Ekonomicznych”, 9, pp. 239–244.

Bootstrapowe metody estymacji wartości zagrożonej ryzykiem

Abstrakt

Przedziałowe metody bootstrapowe mogą być stosowane do szacowania wartości zagrożonej ryzykiem, definiowanej jako kwantyl ustalonego rzędu rozkładu zmiennej losowej określającej wielkość strat z inwestycji. Zaletą tych metod jest możliwość ich zastosowania w przypadku braku informacji o klasie rozkładu badanej zmiennej. Szczególne znaczenie ma semiparametryczna procedura estymacji wykorzystywana do estymacji kwantyli o wysokich rzędach. Gwarantuje ona pojawienie się w generowanych próbach bootstrapowych dużych wartości. W artykule zaprezentowane zostały nieparametryczne i semiparametryczne bootstrapowe metody estymacji wraz z wynikami badań symulacyjnych dotyczących szacowania kwantyli wyższych rzędów rozkładów o grubych ogonach. Analizowane metody prowadzą do uzyskania oszacowań o większej dokładności w porównaniu z klasyczną nieparametryczną procedurą estymacji. Rozważane procedury zostały wykorzystane do szacowania *VaR* dziennych stóp zwrotu wybranych akcji notowanych na Giełdzie Papierów Wartościowych w Warszawie.

Słowa kluczowe: estymacja bootstrapowa, dokładność oszacowania, wartość zagrożona ryzykiem