ELSEVIER
Physica A 287 (2000) 269–278
PHYSICA A
www.elsevier.com/locate/physa

Property insurance loss distributions

Krzysztof Burnecki[a], Grzegorz Kukla[a], Rafał Weron[b,*,1]

[a] *Institute of Mathematics, Wrocław University of Technology, 50-370 Wrocław, Poland*

[b] *Hugo Steinhaus Center for Stochastic Methods, Wrocław University of Technology, 50-370 Wrocław, Poland*

Received 12 July 2000

Abstract

Property claim services (PCS) provides indices for losses resulting from catastrophic events in the US. In this paper, we study these indices and take a closer look at distributions underlying insurance claims. Surprisingly, the lognormal distribution seems to give a better fit than the Paretian one. Moreover, lagged autocorrelation study reveals a mean-reverting structure of indices returns.

PACS: 02.50.Rj; 05.45.Tp; 89.90.+n

Keywords: Econophysics; Property insurance; Loss distribution; PCS index

1. Introduction

The property insurance industry has paid out over \$75 billion in losses in the last five years due to increasingly severe catastrophes. In 1999, insured losses from natural catastrophes and man-made disasters amounted to \$28.6 billion, of which 85% was caused by natural disasters and 15% by man-made ones. This amounted to the second-heaviest claims burden ever for insurers, behind 1992, the year of hurricane Andrew (\$32.4 billion of insured losses of which \$19 billion were due to the hurricane alone; all losses in 1999 prices) [1]. The main cause of the heavy loss burden resulted from storms Anatol, Lothar (\$4.5 billion) and Martine, which ravaged much of western Europe [2]. Some industry experts believe that even larger catastrophes are coming because population and building development continue to increase. Moreover,

* Corresponding author. Fax: +48-71-320-26-54.

E-mail address: rweron@im.pwr.wroc.pl (R. Weron).

[1] Research supported by KBN Grant No. 8 T10B 034 17.

PII: S0378-4371(00)00453-2

there are reasons to believe that such catastrophes will occur more frequently and with greater force in the future due to changes in the earth's atmosphere [1,2].

At present, however, the insurance and reinsurance industries simply do not have the resources needed to support a major catastrophe. For example, the primary insurance and reinsurance industry in the US has capital to cover only 1% of about $30 trillion of national property. So a $50 billion disaster would probably wipe out a large number of insurance companies from business [3]. In order for the needed capacity to be achieved, non-traditional forms of capital, such as hedge and pension funds, need to be engaged. However, this cannot be done without standardization and commoditization of insurance risks into tradable securities. These securities must also be efficient hedging mechanisms and risk management tools that the insurance industry is willing to buy. The Chicago Board of Trade's catastrophe options – based on PCS indices – were among the first products which could potentially meet these needs, and thus bridge the gap between the capital and insurance markets [4].

The purpose of this paper is to see what type of distributions fit PCS indices data and whether there is significant autocorrelation of indices returns [5,6]. These questions have to be answered before more sophisticated methods are used for pricing of structured derivatives based on PCS indices.

This paper is structured so that the next section describes the PCS indices. After that, we present the four distributions and their mixtures fitted later on to the quarterly PCS National index values from the period 1950–99. In Section 4 we briefly discuss the well- and not-so-well-known non-parametric tests often used for judging which distribution fits the empirical financial data best [7]. Finally, in the last section we present results of our statistical analysis. Surprisingly, contrary to earlier reports [8,9], the lognormal distribution seems to give a better fit than the Paretian one.

2. PCS indices

Property claim services (PCS) is recognized around the world as the property/casualty insurance industry's authority on insured catastrophic events. Since the inception of the catastrophe serial number system in 1949, PCS has been responsible for estimating insured property damage resulting from catastrophes affecting the US.

PCS provides indices for losses resulting from catastrophic events on a daily basis. By definition, a catastrophe is an event that causes over $5 million of insured property damage and affects a significant number of policy holders and insurance companies. PCS compiles its estimates of insured property damage using a combination of procedures, including a general survey of insurers, its National Insurance Risk Profile (NIRP), and, where appropriate, its own on-the-ground survey [4].

A survey of companies, agents, and adjusters is one part of the estimating process. PCS conducts confidential surveys of at least 70% of the market based on premium-written market share. PCS then develops a composite of individual loss and claim estimates reported by these sources. Using both actual and projected claim

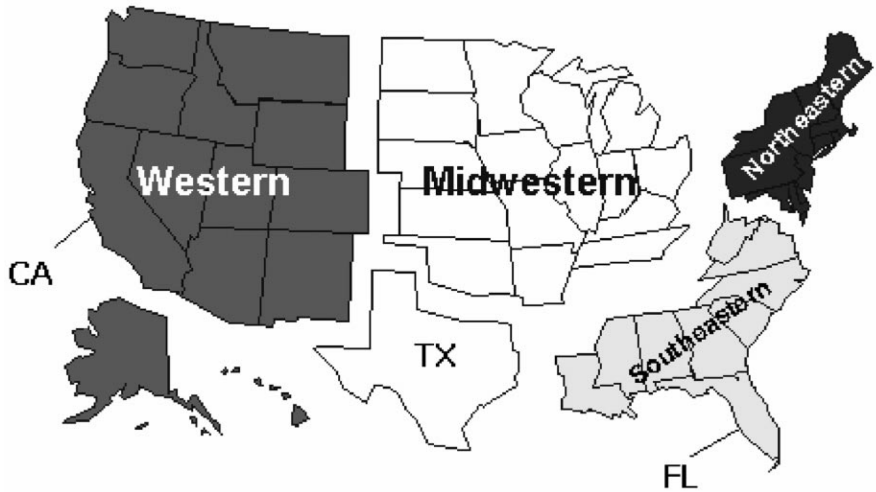

Fig. 1. Regions covered by the PCS indices traded on CBOT (CA stands for California, TX for Texas, and FL for Florida).

figures, PCS extrapolates to a total industry estimate by comparing this information to the market share data.

The PCS indices include direct and indirect insurance losses, i.e.,

- real property (buildings, detached garages, sheds, pool cabanas, etc.);
- contents of the building;
- living expenses (in the case of homeowners' insurance);
- extra or business interruption expenses (in the case of commercial properties);
- personal boats (not ocean liners).

PCS indices track insured catastrophic loss estimates on a national, regional, and state basis from information obtained by PCS. Nine PCS indices are listed for trading on the Chicago Board of Trade, see Figs. 1 and 2:

- a National index covering all insured losses in the US;
- Eastern – consisting of Northeastern (storms) and Southeastern (hurricanes), Midwestern (floods, snow storms), and Western (earthquakes, tsunami waves) regional indices;
- state indices for most exposed regions: Florida (hurricanes), Texas (tornadoes), and California (earthquakes).

Only options (and combinations of options – spreads) on these indices are available for trading.

3. Distributions

The derivation of claim size distributions from the claim data could be considered to be a separate discipline in its own right, applying the methods of mathematical statistics

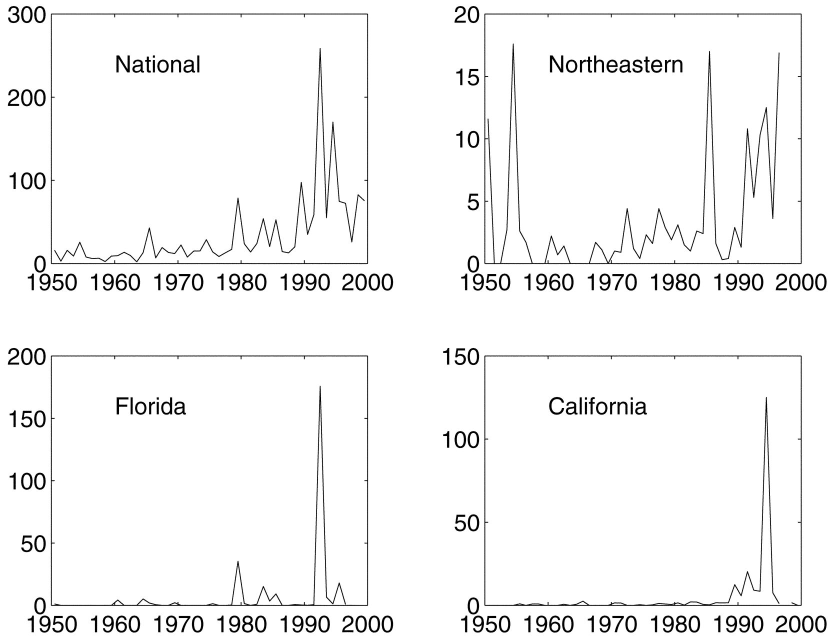

Fig. 2. Annual PCS indices since 1950: National, Northeastern, Florida and California. All values represent index points (1 pt = \$100 mln).

[8]. The objective is to find a distribution function (d.f.) F which fits the observed data in a satisfactory manner. The models used for the d.f. F can be classified into the following three basic types:

(1) F is expressed in an analytic form which is fitted to the observed data;

(2) F is derived directly from statistical data in a tabular, parameter-free discrete form;

(3) F is not specified explicitly, but the lowest main characteristics, in particular the mean, standard deviation and skewness, are derived from the data.

The analytical form is the approach most frequently adopted in the actuarial literature. The problem is to find a suitable analytic expression which fits the observed data well and which is easy to handle. The lognormal, Pareto, Burr and gamma distributions, to be dealt in the sequel, are typical candidates to be considered for applications.

Consider a random variable X which has the normal distribution. Let $Y = \exp X$. The distribution of Y is called a lognormal distribution. The d.f. is given by

$$F(x) = \Phi\left(\frac{\ln x - \mu}{\sigma}\right) = \int_0^x \frac{1}{\sqrt{2\pi}\sigma y} \mathrm{e}^{-(1/2)((\ln y - \mu)/\sigma)^2} \, \mathrm{d}y, \quad x, \sigma > 0, \ \mu \in R,$$

where $\Phi(x)$ is the standard normal (with mean 0 and variance 1) d.f. The lognormal distribution is very useful in modeling of claim costs. It has a thick right tail and fits many situations well. For small σ it resembles a normal distribution, although this is not always desirable.

One of the most frequently used analytic claim size distributions is the Pareto d.f. which is defined by the formula

$$F(x) = 1 - \left(\frac{\lambda}{\lambda + x} \right)^{\alpha}, \quad x, \alpha, \lambda > 0 \, .$$

The first parameter controls how heavy a tail distribution has: the smaller the α, the heavier the tail. Experience has shown that the Pareto formula is often an appropriate model for the claim size distribution, particularly where exceptionally large claims may occur [8,9]. However, there is a need to find heavy-tailed distributions which offer greater flexibility than the Pareto d.f. Such flexibility is provided by the Burr distribution

$$F(x) = 1 - \left(\frac{\lambda}{\lambda + x^{\tau}} \right)^{\alpha}, \quad x, \alpha, \lambda, \tau > 0 \, ,$$

which is just a generalization of the Pareto distribution.

All the three presented distributions suffer from some mathematical drawbacks (e.g. lack of a closed form representation for the Laplace transform or non-existence of the moment generating function). On the other hand, the gamma distribution

$$F(x) = \int_0^x \frac{1}{\Gamma(\alpha)\beta^{\alpha}} y^{\alpha-1} \mathrm{e}^{-y/\beta} \, \mathrm{d}y, \quad x, \alpha, \beta > 0 \, ,$$

does not have these drawbacks. It is one of the most important distributions for modeling (not only insurance claims) because it has very tractable mathematical properties and is related to other distributions [10].

Up to this point, we have assumed that all observations were positive valued. In our case, like in other insurance contexts, PCS indices can have zero values as well. However, such situations present no difficulties in the calculation of the total d.f., namely we can write

$$F(x) = P(X = 0) + P(X > 0)F_{+}(x), \quad x \geqslant 0 \, ,$$

where F_{+} is a d.f. related to positive values of the PCS index and is given by one of the above analytic forms. This means that the spike at zero can be easily removed. When this is done, the model is reduced to one of the positive claims only.

4. Non-parametric tests

Once the distribution is selected, we must obtain parameter estimates. In what follows we use moment (for the Pareto distribution only) and maximum likelihood estimation. The next step is to test whether the fit is adequate. This is usually done by comparing the fitted and empirical d.f.'s, more precisely, by checking whether values of the fitted d.f. at sample points form a uniform distribution [11]. We applied the well- and not-so-well-known non-parametric tests verifying the hypothesis of uniformity. The critical values C_{α} of the tests, given a significance level α (e.g. $\alpha = 0.05$), can be easily found in the literature [11,12].

A very natural and well-known is the χ^2 statistics

$$\chi_k^2 = k \sum_{i=1}^{k} \frac{(n_i - n/k)^2}{n} ,$$

where n is the overall number of observations and n_i is the number of observations which fall into the interval $[(i-1)/k, i/k]$. χ_k^2 has an approximate chi-squared distribution with $k-1$ degrees of freedom. In general, the better the fit, the smaller the χ_k^2. This test becomes more discriminating as the sample size becomes larger.

Another classical measure of fit is the Kolmogorov–Smirnov statistics

$$D_n = \sup_{x \in R} |\hat{F}(x) - F_+(x)| ,$$

where the empirical d.f. is defined as $\hat{F}(x) = (1/n)\sum_{i=1}^{n} 1_{\{x_i \leqslant x\}}$. Statistics D_n measures the distance between the empirical and fitted d.f. in the supremum norm.

The two other tests we apply are the Cramer–von Mises and Anderson–Darling tests [11,5]. The former uses

$$C_n = n \int_{-\infty}^{+\infty} (\hat{F}(x) - F_+(x))^2 \, \mathrm{d}F_+(x)$$

statistics while the latter (considered to be the best within the class of tests based on empirical d.f.) uses

$$AD = n \int_{-\infty}^{+\infty} \frac{(\hat{F}(x) - F_+(x))^2}{F_+(x)(1 - F_+(x))} \, \mathrm{d}F_+(x) .$$

In order to interpret the results of the tests we compare with the corresponding critical values C_α (for the same significance level α). When the value of the test is less than the corresponding value C_α we accept the fit as adequate (there is no reason to reject the null hypothesis). A problem arises when there is more then one distribution that fits the sample data. We should somehow distinguish them. For this reason we introduce the following test function [13]:

$$Z = \sum_{i=1}^{K} \frac{T_i}{C_{\alpha,i}} , \qquad (1)$$

where K denotes the number of tests under consideration, T_i their values (e.g. T_1 may denote the value of the χ^2 test, T_2 of the AD test etc.) and $C_{\alpha,i}$ – the corresponding critical values. The smallest value of such a function indicates the best fit (in the sense of the l_1 norm).

In the case when none of the distributions fits the sample data or when we want to improve the fit, we can take into consideration a mixture of distributions. A mixture of two distributions can be written as

$$G(x) = aF_1(x) + (1-a)F_2(x), \quad a \in (0,1) .$$

In this case testing the fit of the model requires estimating not only the parameters of F_1 and F_2 but the coefficient a as well.

5. Empirical analysis

Empirical studies were conducted for the quarterly National PCS index. The National index was chosen since it covers losses in the whole United States and thus has the most non-zero values. Of the two available data sets – annual and quarterly– the latter series was selected because of its length, see Fig. 3. Note, however, that the annual index value does not necessarily equal the sum of the quarterly index values. This fact is due to the rounding of the raw data in formulating the indices and to later revisions of loss estimates performed by PCS.

At the beginning distributions were fitted using moments and maximum likelihood estimation. The results of parameter estimation and test statistics are presented in Table 1. Only three distributions passed all tests and it is hard to judge which one is the best. For this reason, in the next step we have conducted parameter estimation via

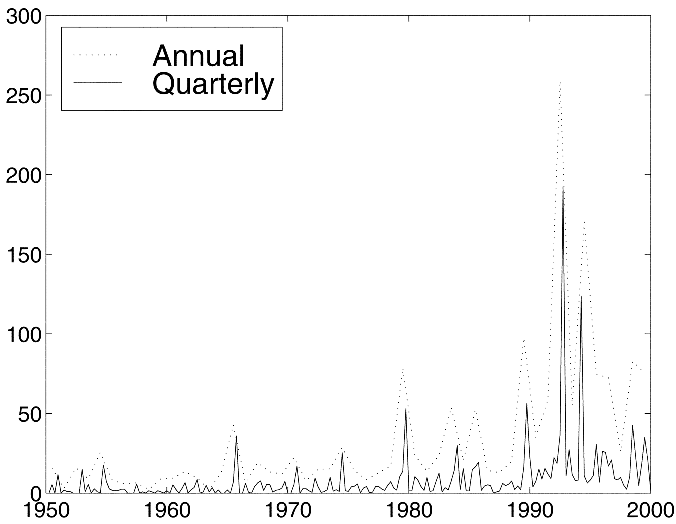

Fig. 3. Annual and quarterly National PCS indices since 1950.

Table 1
Parameter estimates and test statistics for the quarterly National PCS index. Parameter estimates were obtained through moments (Pareto distribution only) or maximum likelihood estimation

Distributions	Lognormal	Pareto	Burr	Gamma
Parameters	$\mu = 1.5240$	$\alpha = 2.7163$	$\alpha = 1.0552$	$\alpha = 0.78156$
	$\sigma = 1.2018$	$\lambda = 16.8759$	$\lambda = 9.1614$	$\beta = 12.5811$
			$\tau = 1.4085$	
Test values (*in parentheses*: *critical values for* $\alpha = 0.05$)				
χ^2 (24.9958)	17.48235	14.47059	17.29412	19.55294
D_n (0.103138)	0.05916	0.06700	0.06873	0.09921
C_n (0.460636)	0.05028	0.09670	0.09017	0.58648
AD (2.49200)	0.30930	0.98915	0.57751	3.43060

Table 2
Parameter estimates and test statistics for the quarterly National PCS index. Parameter estimates were obtained through minimization of the test function Z (see Eq. (1))

Distributions	Lognormal	Pareto	Burr	Gamma
Parameters	$\mu = 1.4856$ $\sigma = 1.2377$	$\alpha = 2.798$ $\lambda = 17.1001$	$\alpha = 1.1718$ $\lambda = 8.3050$ $\tau = 1.2817$	$\alpha = 0.90443$ $\beta = 8.4846$
Test values (*in parentheses*: *critical values for* $\alpha = 0.05$)				
χ^2 (24.9958)	3.74118	8.07059	5.24706	14.6588
D_n (0.103138)	0.04448	0.06845	0.04511	0.06792
C_n (0.460636)	0.04691	0.08632	0.07226	0.14116
AD (2.49200)	0.31697	0.95679	0.59704	1.91120

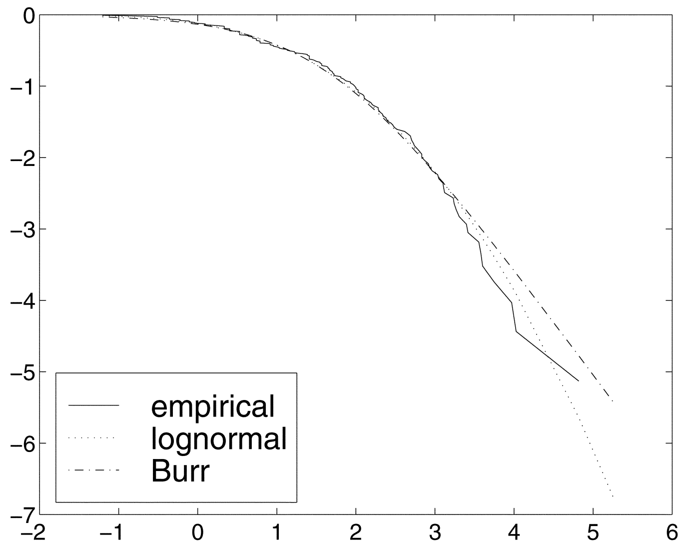

Fig. 4. Right tails of the quarterly National PCS index and two best approximating distributions on a double-logarithmic paper: $\ln(1 - F(t))$ is plotted against $\ln(t)$.

minimization of the test function Z, see Eq. (1) and Table 2. This time the lognormal distribution came out as the winner (as it had the smallest values in all tests) with the Burr distribution following closely. Both distributions are plotted in Fig. 4 on a double-logarithmic paper.

Finally, after fitting distributions themselves we have estimated parameters of mixtures and tested the results using the four presented tests. As we can see in Table 3, we have not been able to obtain a much better fit. The lognormal–lognormal mixture is the best, but the test values are almost identical to the ones for the lognormal distribution itself!

Table 3
Parameter estimates of mixtures and test statistics for the quarterly National PCS index. Again parameter estimates were obtained through minimization of the test function Z (see Eq. (1)). Ln means lognormal, Pa – Pareto, Bu – Burr, Ga – Gamma

Mixtures	Ln+Ln	Ln+Pa	Ln+Bu	Ln+Ga	Pa+Bu
Parameters	$\mu_1 = 1.6988$	$\mu = 1.4193$	$\mu = 1.2657$	$\mu = 1.4105$	$\alpha_1 = 2.5842$
	$\sigma_1 = 1.2332$	$\sigma = 1.1756$	$\sigma = 1.1650$	$\sigma = 1.1560$	$\lambda_1 = 17.0442$
	$\mu_2 = 1.2888$	$\alpha = 2.7891$	$\alpha = 1.2065$	$\alpha = 0.77057$	$\alpha_2 = 1.1487$
	$\sigma_2 = 1.2075$	$\lambda = 17.8431$	$\lambda = 11.0105$	$\beta = 12.5486$	$\lambda_2 = 8.5787$
			$\tau = 1.2185$		$\tau = 1.3829$
	$a = 0.48599$	$a = 0.52536$	$a = 0.55139$	$a = 0.5472$	$a = 0.48223$
Test values (*in parentheses*: *critical values for* $\alpha = 0.05$)					
χ^2 (24.9958)	3.74118	8.07059	3.74118	9.38824	12.2118
D_n (0.103138)	0.04439	0.04473	0.04663	0.04898	0.04992
C_n (0.460636)	0.04660	0.04906	0.05744	0.05145	0.05448
AD (2.49200)	0.31168	0.48146	0.43008	0.57705	0.55939

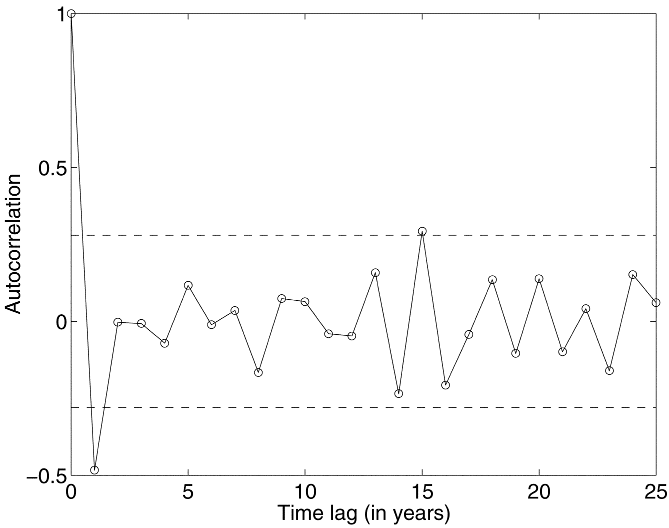

Fig. 5. Lagged autocorrelation function for log-returns of the quarterly National PCS index. Dashed horizontal lines represent the 95% confidence interval of a Gaussian random walk.

Seasonality of a time series of returns r_t (logarithmic changes of the index) can be demonstrated by plotting the autocorrelation function [14]

$$acf(r,k) = \frac{\sum_{t=k+1}^{N}(r_t - \bar{r})(r_{t-k} - \bar{r})}{\sum_{t=1}^{N}(r_t - \bar{r})^2},$$

where N is the sample length and $\bar{r} = (1/N)\sum_{t=1}^{N} r_t$, for different time lags k as in Fig. 5. For the annual National PCS index log-returns there is a strong anticorrelation for lag $k = 1$ year (quarterly index values contained zeros making it impossible to conduct the autocorrelation analysis). For almost all other lags the correlation falls into the confidence interval for the Brownian motion, indicating no dependence. This result is similar to that for electricity spot price returns [15,16], i.e., both processes are mean-reverting, and unlike that for most financial data [17,18].

References

[1] Swiss Reinsurance Co., Sigma no. 2/2000, www.swissre.com, 2000.
[2] Guy Carpenter & Co., Natural Hazards – Review of the Year 1999, www.guycarp.com/publications/white/whitep.html, 2000.
[3] Guy Carpenter & Co., The World Catastrophe Reinsurance Market 1999, www.guycarp.com/publications/white/whitep.html, 1999.
[4] Chicago Board of Trade, PCS Catastrophe Insurance Options: A User's Guide, CBOT, Chicago, 1995.
[5] A. Weron, R. Weron, Financial Engineering: Derivatives Pricing, Computer Simulations, Market Statistics, WNT, Warsaw, 1998, (in Polish).
[6] R.N. Mantegna, H.E. Stanley, An Introduction to Econophysics: Correlations and Complexity in Finance, Cambridge University Press, Cambridge, 1999.
[7] L. Gillemot, J. Toyli, J. Kertesz, K. Kaski, Physica A 282 (2000) 304.
[8] C.D. Daykin, T. Pentikainen, M. Pesonen, Practical Risk Theory for Actuaries, Chapman & Hall, London, 1994.
[9] Yu. Baryshnikov, A. Mayo, D.R. Taylor, Working Paper, www.cam.wits.ac.za/mfinance/pages/research.html, 1998.
[10] H.H. Panjer, G.E. Willmot, Insurance Risk Models, Society of Actuaries, Schaumburg, 1992.
[11] R.B. D'Agostino, M.A. Stephens, Goodness-of-Fit Techniques, Marcel Dekker, New York, 1986.
[12] M.A. Stephens, JASA 69 (1974) 730.
[13] G. Kukla, Insurance Risk Derivatives, M.Sc. Thesis, Wrocław University of Technology, 2000.
[14] J. Beran, Statistics for Long-Memory Processes, Chapman & Hall, New York, 1994.
[15] A. Weron, R. Weron, Power Exchange: Risk Management Strategies, CIRE, Wrocław, 2000 (in Polish).
[16] R. Weron, Physica A 285 (2000) 127.
[17] J.P. Bouchaud, M. Potters, Theory of Financial Risk, Alea-Saclay, Eyrolles, Paris, 1997 (in French).
[18] V. Plerou, P. Gopikrishnan, B. Rosenow, L.A.N. Amaral, H.E. Stanley, Physica A 279 (2000) 443.