Black swans or dragon kings? A simple test for deviations from the power law

Janczura, Joanna and Weron, Rafal

Wrocław University of Technology, Poland

2 May 2011

Black swans or dragon kings? A simple test for deviations from the power law

Joanna Janczura[1,a] and Rafał Weron[2,b]

[1] Hugo Steinhaus Center for Stochastic Methods, Wrocław University of Technology, Poland
[2] Institute of Organization and Management, Wrocław University of Technology, Poland

Abstract. We develop a simple test for deviations from power law tails, which is based on the asymptotic properties of the empirical distribution function. We use this test to answer the question whether great natural disasters, financial crashes or electricity price spikes should be classified as dragon kings or 'only' as black swans.

1 Introduction

In a recent article Didier Sornette [18] presents a number of data sets exhibiting power law or other heavy-tailed behavior with extreme outliers. They range from French city sizes with Paris as the extreme outlier to financial drawdowns (see the discussion in Section 3.3), turbulent velocity fluctuations and earthquake magnitudes. Sornette calls these extreme events *dragon kings* and says that they are the result of positive feedback mechanisms that make them much larger than their peers. Being outliers to heavy-tailed behavior, these dragon kings are unaccounted for by power laws, for which Nassim Taleb coined the nowadays popular term *black swans* [17].

Interestingly, Taleb [19] does not define black swans as power law events. Rather he describes them in a more literary way as events which: (i) are unpredictable outliers, beyond the realm of regular expectations, (ii) carry an extreme impact and (iii) human nature makes them explainable and predictable ... after they have happened. Taleb regards almost all major scientific discoveries (like the personal computer, the Internet) and historical events (like the market crash of 1987, the 9/11 attack) as black swans. He also makes a distinction between the 'totally intractable' black swans and – what he calls – the *Mandelbrotian gray swans*, which are 'tractable scientifically' ... by means of power laws. But the gray swan terminology has not picked up and black swans prevail. Why? Partly due to the fact that power law events pretty well fit Taleb's definition anyway. Recall, that power law distributions are scale invariant. If extreme events are described by a power law distribution there is no way to predict them because nothing can distinguish them from the smaller events: their extreme impacts are ... beyond the realm of normal expectations.

According to Sornette, extreme events are significantly more likely to happen than power laws suggest. Are dragon kings then 'to blame' for experiencing two or three once-in-a-millennium floods or financial crises in the last couple of decades? What is a

---

[a] e-mail: `joanna.janczura@pwr.wroc.pl`
[b] e-mail: `rafal.weron@pwr.wroc.pl`

dragon king anyway? And how do we know that a given observation is a dragon king and not simply a 'normal' deviation to be expected of any random sample? Definitely these issues are intriguing and worth an investigation.

Having the latter question in focus, in Section 2 we develop a simple test for deviations from power law tails. Actually, from the tails of any distribution. The test is based on the asymptotic properties of the empirical distribution function (edf) and – as we show in Section 3 – is a universal tool that can be used in many practical situations.

2 Testing for dragon kings

The simple test we propose here is based on the properties of the empirical distribution function (edf):

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}_{\{x_i < x\}}, \tag{1}$$

which is defined for a sample of observations $(x_1, x_2, ..., x_n)$, with $\mathbb{I}$ being the indicator function. To construct the test first note, that the number Y of observations x_i smaller than x is binomially distributed with parameters n and $r = F(x)$, i.e.

$$Y = \sum_{i=1}^{n} \mathbb{I}_{\{x_i \leq x\}} \sim B(n, F(x)), \tag{2}$$

where $F(x)$ is the true cumulative distribution function (cdf). Recall, that the first parameter of the binomial law represents the number of trials (here: sample size n) and the second – the success probability in each trial (here: the value of the cdf at point x). Hence, we can write:

$$P\left(q_{\frac{\alpha}{2}}[n, F(x)] < Y \leq q_{1-\frac{\alpha}{2}}[n, F(x)]\right) = 1 - \alpha, \tag{3}$$

where $q_\eta[n, r]$ is the η-quantile of the binomial distribution $B(n, r)$. Now, using definition (1) and property (3), we can construct confidence intervals (CI) for the edf and an arbitrarily chosen confidence level $(1 - \alpha)$. Indeed, we have:

$$P\left(\frac{1}{n}q_{\frac{\alpha}{2}}[n, F(x)] < F_n(x) \leq \frac{1}{n}q_{1-\frac{\alpha}{2}}[n, F(x)]\right) = 1 - \alpha. \tag{4}$$

Analogously, for the right tail we have:

$$P\left(\frac{1}{n}q_{\frac{\alpha}{2}}[n, 1 - F(x)] < 1 - F_n(x) \leq \frac{1}{n}q_{1-\frac{\alpha}{2}}[n, 1 - F(x)]\right) = 1 - \alpha. \tag{5}$$

Note, that these are pointwise intervals, meaning that for each specified value of x, we are $(1 - \alpha) \times 100\%$ confident of observing edf(x) within those limits. This is not the same as constructing the so-called confidence bands which guarantee, with a given confidence level, that the edf falls within the band for all x's (in some interval); these bands are wider than the curves one obtains by using pointwise CI [2,13].

At this point assume that the true distribution F has power law tails, i.e.

$$F(x) \approx b_1|x|^{p_1} \quad \text{for} \quad x \to -\infty, \tag{6}$$
$$1 - F(x) \approx b_2 x^{p_2} \quad \text{for} \quad x \to \infty. \tag{7}$$

From (4), the left tail of the edf should lie in the interval

$$\left(\frac{1}{n} q_{\frac{\alpha}{2}}[n, b_1|x|^{p_1}], \frac{1}{n} q_{1-\frac{\alpha}{2}}[n, b_1|x|^{p_1}] \right] \tag{8}$$

with probability $1 - \alpha$. Similarly, from (5), the right tail should lie in the interval

$$\left(\frac{1}{n} q_{\frac{\alpha}{2}}[n, b_2 x^{p_2}], \frac{1}{n} q_{1-\frac{\alpha}{2}}[n, b_2 x^{p_2}] \right] \tag{9}$$

with probability $1 - \alpha$. Now, it suffices to fit a power law to the left or right tail of the edf built from the analyzed sample and plot the respective intervals (8) or (9). Observations lying outside the curves spanned by the CI are likely to be (i.e. with probability $1 - \alpha$) dragon kings.

Although formulas (8) and (9) seem simple, the computation of the CI is not straightforward – there are no closed form formulas (not involving special functions) for the inverse of the binomial cdf. Naturally, the quantiles of the binomial distribution $B(n, r)$ can be obtained by inverting the binomial cdf, but this can be time consuming for a large number of trials (here: the sample size). Alternatively the quantile $q_\eta[n, r]$ can be obtained as a solution of

$$\eta = \beta_{1-r}(n - q_\eta[n, r], 1 + q_\eta[n, r]), \tag{10}$$

where β is the incomplete beta function [16]. But again this can be burdensome as β is a special function and has to be approximated numerically. In particular, the Matlab implementation of β is inefficient and leads to inaccurate CI. Moreover, for the sample sizes considered in this paper the computational times are at least twice longer than for the algorithm utilizing binomial inversion.

As a remedy to these numerical issues we suggest to use an approximation based on asymptotic properties of the edf. Recall that by the Central Limit Theorem, $F_n(x)$ is asymptotically (i.e. as $n \to \infty$) normally distributed [1]. Precisely,

$$\frac{\sqrt{n}[F_n(x) - F(x)]}{\sqrt{F(x)[1 - F(x)]}} \xrightarrow{d} N(0, 1), \tag{11}$$

where $\xrightarrow{d}$ denotes convergence in distribution. Letting $z_{\frac{\alpha}{2}}$ and $z_{1-\frac{\alpha}{2}}$ denote the $(\frac{\alpha}{2})$ and $(1 - \frac{\alpha}{2})$-quantiles of the standard normal distribution, respectively, we have:

$$P\left(z_{\frac{\alpha}{2}} < \frac{\sqrt{n}[F_n(x) - F(x)]}{\sqrt{F(x)[1 - F(x)]}} < z_{1-\frac{\alpha}{2}} \right) \approx 1 - \alpha, \tag{12}$$

provided that n is large enough. Note, that since the standard Gaussian law $N(0, 1)$ is symmetric around 0, $z_{1-\frac{\alpha}{2}} = -z_{\frac{\alpha}{2}}$. Formula (12) implies that:

$$P\left(F(x) + \sqrt{\frac{F(x)[1 - F(x)]}{n}} z_{\frac{\alpha}{2}} < F_n(x) < F(x) + \sqrt{\frac{F(x)[1 - F(x)]}{n}} z_{1-\frac{\alpha}{2}} \right) \approx 1 - \alpha. \tag{13}$$

Analogously, for the right tail we have:

$$P\left(1 - F(x) + \sqrt{\frac{F(x)[1 - F(x)]}{n}} z_{\frac{\alpha}{2}} < 1 - F_n(x) < \right.$$

$$\left. < 1 - F(x) + \sqrt{\frac{F(x)[1 - F(x)]}{n}} z_{1-\frac{\alpha}{2}} \right) \approx 1 - \alpha. \tag{14}$$

Now, assuming power law tails, the left tail of the edf should lie in the interval

$$\left(b_1|x|^{p_1} + \sqrt{\frac{b_1|x|^{p_1}(1 - b_1|x|^{p_1})}{n}} z_{\frac{\alpha}{2}}, \ b_1|x|^{p_1} + \sqrt{\frac{b_1|x|^{p_1}(1 - b_1|x|^{p_1})}{n}} z_{1-\frac{\alpha}{2}} \right) \quad (15)$$

with probability $1 - \alpha$. Similarly, from (14) the right tail should lie in the interval

$$\left(b_2 x^{p_2} + \sqrt{\frac{b_2 x^{p_2}(1 - b_2 x^{p_2})}{n}} z_{\frac{\alpha}{2}}, \ b_2 x^{p_2} + \sqrt{\frac{b_2 x^{p_2}(1 - b_2 x^{p_2})}{n}} z_{1-\frac{\alpha}{2}} \right) \quad (16)$$

with probability $1 - \alpha$. Formulas (15)-(16) are numerically much more efficient than their exact, binomial-law based counterparts (8)-(9). The quantiles of the standard normal distribution are easily computable since they can be expressed in terms of the inverse error function, for which fast and accurate numerical algorithms exist. Moreover, to construct the CLT-based CI only one value of the standard normal quantiles, i.e. $z_{\frac{\alpha}{2}} = -z_{1-\frac{\alpha}{2}}$, is needed no matter what is the range of the CI, while in the case of the exact CI the quantiles of the binomial distribution, $q_{\frac{\alpha}{2}}[n, b|x|^p]$ and $q_{1-\frac{\alpha}{2}}[n, b|x|^p]$, have to be found for each point x. The speedup is considerable. For a sample size of 1000 observations the CLT-based CI for a power law fitted to the largest 10%-1% values can be computed ca. 60 times faster than the exact CI, but for a sample size of 5000 observations over 220 times faster (0.003 vs. 0.665 seconds on an Intel Core i7-820QM processor running Matlab 7.9). Computational times under one second should not be worrying, however, for larger samples or fits to a wider range of values (say, top 25% observations), they can become substantial for the exact CI.

Being aware of the extreme speedup in calculations, what remains to be tested is the accuracy of the approximate CLT-based CI. We will deal with this issue in Section 3.1. But before we do that let us only remark, that the presented approach is general. The true distribution can be arbitrary, say, stretched exponential (also known as Weibull). In that case the confidence intervals would be computed from relations (4)-(5) or (13)-(14) using the stretched exponential cdf, see Figure 1.

3 Empirical examples

3.1 Simulated data

We start the empirical analysis with a simulation study to check the effectiveness of the test for dragon kings. The results summarized in Tables 1 and 2 concern random samples from two heavy-tailed distributions with power law decay in the tail(s):

- Cauchy with cdf $F(x; \mu, \sigma) = \frac{1}{\pi} \arctan\left(\frac{x-\mu}{\sigma}\right) + \frac{1}{2}$ where $x, \mu \in \mathbb{R}$ and $\sigma > 0$,
- Pareto with cdf $F(x; \lambda, \alpha) = 1 - \lambda^\alpha (x + \lambda)^{-\alpha}$ where $x \geq 0$ and $\lambda, \alpha > 0$,

and two lighter tailed laws:

- symmetric Hyperbolic with the cdf obtained by numerically integrating the probability density function (pdf) $f(x; \alpha, \delta) = \exp(-\alpha\sqrt{\delta^2 + x^2})/\{2\delta K_1(\delta\alpha)\}$, where K_1 is the modified Bessel function of the third kind, $x \in \mathbb{R}$ and $\alpha, \delta > 0$,
- stretched exponential (or Weibull) with cdf $F(x; \beta, \tau) = 1 - e^{-\beta x^\tau}$ where $x \geq 0$ and $\beta, \tau > 0$.

For simulation and estimation issues, as well as, sample applications of these distributions in finance and insurance see e.g. [8,14,15].
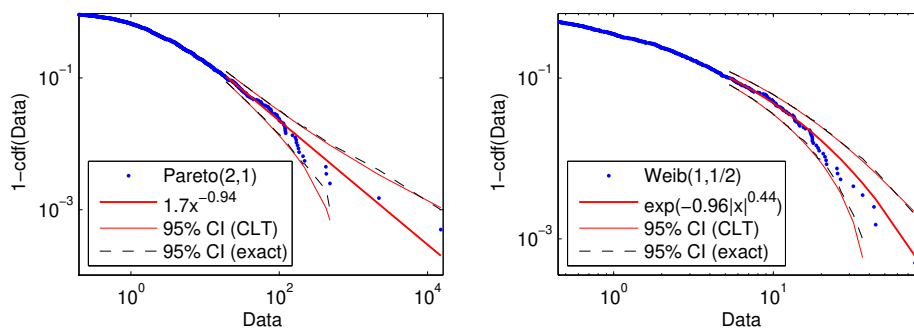
Fig. 1. Comparison of exact, binomial law-based CI and approximate, CLT-implied CI for simulated samples of 1000 observations of Pareto(2,1)-distributed (*left panel*) and Weibull(1,0.5)-distributed random numbers (*right panel*). The power law and Weibull tails were fitted to the 10%-1% largest observations. Apparently for samples sizes of 1000 or more observations the differences between the CI are negligible, even in the very tails of the distribution.

Table 1. Percentage of pointwise outliers with respect to the exact, binomial law-based CI given by formula (5) and $F(x)$ being a power law fitted to the 10%-1% or 25%-2.5% largest observations. For Cauchy and Pareto distributions, outliers with respect to the true power law (implied by the cdf) are also provided. The number of simulated samples is equal to 10^4.

CI	Sample size	Cauchy(0,1)		Pareto(2,1)		Hyp(2,1)	Weib(1,$\frac{1}{2}$)
		True	Fitted	True	Fitted	Fitted	Fitted
		Power law fitted to 10%-1% largest observations					
90%	1000	10.0%	4.5%	9.8%	4.5%	24.7%	16.6%
95%	1000	5.2%	2.2%	4.6%	2.1%	10.7%	6.8%
99%	1000	1.1%	0.3%	1.0%	0.3%	0.7%	0.2%
		Power law fitted to 25%-2.5% largest observations					
90%	1000	"	12.4%	"	11.1%	100.0%	96.9%
95%	1000	"	6.1%	"	5.4%	99.9%	92.2%
99%	1000	"	1.2%	"	0.8%	98.2%	70.6%
		Power law fitted to 10%-1% largest observations					
90%	5000	9.9%	10.1%	10.0%	11.0%	99.8%	98.5%
95%	5000	5.1%	5.2%	4.8%	5.4%	99.3%	96.7%
99%	5000	1.1%	0.7%	1.0%	1.1%	96.3%	84.8%

The validation procedure is the following. We calculate the edf for 10^4 simulated samples from each of the four distributions and two sample sizes. Next, using least squares regression we fit a power law to the 10%-1% (or 25%-2.5%) largest observations. The upper 1% (or 2.5%) of values are not used for calibration due to the sensitivity of the least squares fit to outlying observations. Then, we compute the pointwise CI given by the exact, binomial law-based formula (9), as well as, the CLT-based approximation (16), see the left panel in Figure 1. Finally, we check whether for a particular value of x the edf lies within the CI. Motivated by the fact that dragon kings are to be expected in the very tails of the distribution, we arbitrarily set x to be the fourth largest observation, i.e. $x_{(4)}$. However, we have checked that qualitatively identical results are obtained for $x_{(8)}$ and $x_{(12)}$.

Table 2. Percentage of pointwise outliers with respect to the approximate, CLT-based CI given by formula (14) and $F(x)$ being a power law fitted to the 10%-1% or 25%-2.5% largest observations. For Cauchy and Pareto distributions, outliers with respect to the true power law (implied by the cdf) are also provided. The number of simulated samples is equal to 10^4.

CI	Sample size	Cauchy(0,1) True	Cauchy(0,1) Fitted	Pareto(2,1) True	Pareto(2,1) Fitted	Hyp(2,1) Fitted	Weib(1,$\frac{1}{2}$) Fitted
Power law fitted to 10%-1% largest observations							
90%	1000	9.6%	4.2%	9.1%	4.3%	29.2%	20.4%
95%	1000	4.8%	2.1%	4.1%	1.7%	10.1%	6.1%
99%	1000	1.1%	0.6%	0.9%	0.4%	0.1%	<0.1%
Power law fitted to 25%-2.5% largest observations							
90%	1000	"	10.7%	"	11.2%	99.1%	90.9%
95%	1000	"	4.4%	"	4.7%	99.8%	78.1%
99%	1000	"	0.7%	"	0.5%	74.3%	35.0%
Power law fitted to 10%-1% largest observations							
90%	5000	10.0%	9.9%	9.8%	11.3%	99.8%	98.9%
95%	5000	4.6%	4.3%	4.7%	5.1%	99.5%	96.5%
99%	5000	1.0%	1.2%	1.0%	0.6%	93.3%	78.9%

Table 3. Percentage of pointwise outliers with respect to the exact, binomial law-based CI given by formula (5) and $F(x)$ being a stretched exponential (or Weibull) law fitted to the 10%-1% or 25%-2.5% largest observations. For the Weibull distribution, outliers with respect to the tail of the true cdf are also provided. The number of simulated samples is equal to 10^4.

CI	Sample size	Cauchy(0,1) Fitted	Pareto(2,1) Fitted	Hyp(2,1) Fitted	Weib(1,$\frac{1}{2}$) True	Weib(1,$\frac{1}{2}$) Fitted
Weibull tail fitted to 10%-1% largest observations						
90%	1000	66.1%	63.3%	10.6%	10.7%	11.2%
95%	1000	57.5%	53.5%	5.5%	4.8%	6.6%
99%	1000	41.5%	38.6%	0.8%	1.2%	1.9%
Weibull tail fitted to 25%-2.5% largest observations						
90%	1000	97.4%	97.9%	59.3%	"	14.6%
95%	1000	95.8%	97.0%	47.6%	"	8.0%
99%	1000	90.6%	92.7%	25.8%	"	2.4%

The percentages of pointwise outliers at $x = x_{(4)}$ with respect to the exact, as well as, approximate CI are given in Tables 1 and 2, respectively. As we can observe, the results are highly dependent on the subsample used to calibrate the power law. For the larger sample size (5000 observations), the 10%-1% range leads to accurate coverage rates for the Cauchy and Pareto laws, while for the lighter tailed distributions $x_{(4)}$ is classified as an outlier nearly in all samples. For the smaller sample size (1000 observations), the power law fitted to the 10%-1% range yields overly conservative CI for the Cauchy and Pareto laws – the rejection rates are roughly twice lower. In such a case, the 25%-2.5% range has to be used to obtain a reliable coverage by the CI. However, the percentages of pointwise outliers with respect to the true power law (implied by the parameters of cdf) are accurate. The latter indicates that the dragon king test works well – even for small sample sizes – provided that the power law is

Table 4. Percentage of pointwise outliers with respect to the approximate, CLT-based CI given by formula (14) and $F(x)$ being a stretched exponential (or Weibull) law fitted to the 10%-1% or 25%-2.5% largest observations. For the Weibull distribution, outliers with respect to the tail of the true cdf are also provided. The number of simulated samples is equal to 10^4.

| CI	Sample size	Cauchy(0,1) Fitted	Pareto(2,1) Fitted	Hyp(2,1) Fitted	Weib$(1,\frac{1}{2})$ True	Fitted
		Weibull tail fitted to 10%-1% largest observations				
90%	1000	62.1%	58.3%	11.5%	9.6%	10.6%
95%	1000	56.4%	52.4%	4.9%	4.8%	5.8%
99%	1000	44.8%	41.3%	0.6%	1.1%	2.4%
		Weibull tail fitted to 25%-2.5% largest observations				
90%	1000	96.7%	97.5%	63.2%	"	14.0%
95%	1000	95.3%	96.6%	47.7%	"	8.0%
99%	1000	91.8%	94.1%	19.0%	"	2.6%

correctly estimated. Otherwise it may lead to a lower rejection rate and, hence, a lower number of observations classified as outliers.

As we have said previously, the test of Section 2 is universal in the sense that the true distribution can be arbitrary. We thus repeat the simulation study with another popular heavy-tailed, but not power-law tailed, distribution – the stretched exponential. Note, that in actuarial sciences and statistics it is more commonly known as the Weibull law [8,14]. Now the intervals (8)-(9), or (15)-(16), are computed from relations (4)-(5), or (13)-(14), using the Weibull cdf, see the right panel in Figure 1. To save space, the results summarized in Tables 3 and 4 concern only the smaller sample size (1000 obs.), which better reflects the sizes of the datasets considered in Sections 3.2-3.4 (they range from just under 500 to just over 1500). This time the situation is quite the opposite to the one in Tables 1 and 2. The 10%-1% range leads to accurate coverage rates for the Weibull law, while the rejection rate tends to be too high if the 25%-2.5% range is used (due to a poor fit in the tail of the distribution). For the power law type distributions, the percentage of observations identified as outliers significantly exceeds the expected rejection rates. Only the tails of the hyperbolic law, with their exponential decay, are hard to distinguish from the stretched exponential tails (note, that for $\tau = 1$, the Weibull law reduces to the exponential distribution).

Summing up, the 10%-1% range of largest observations used to fit the power law or the stretched exponential tail seems to be a good choice. Only in the former case, the resulting CI can be overly conservative yielding a lower number of observations classified as outliers. Thus, once an observation is classified as a dragon king it is very likely to be so. On the other hand, if the power law fit is not accurate, an observation inside but close to the edge of the CI may be a dragon king as well.

Finally, comparing the exact, binomial law-based CI and approximate, CLT-implied CI we can conclude that for the considered sample sizes the differences between the CI are negligible. This can be also seen in Figure 1, where both types of CI are plotted for two samples of Pareto(2,1)-distributed and Weibull(1,0.5)-distributed random numbers. Interestingly, in a related context of computing CI for the edf (not for the power law or Weibull fits as in this paper) of seismic data, Pisarenko and Rodkin [16] conclude that the CLT approximation is good enough even for much smaller sample sizes of $n > 20$. We are not as optimistic in our conclusions, but can certainly declare that for sample sizes exceeding a few hundred observations the CLT-based CI are a very good approximation. Hence, taking into account the huge differences
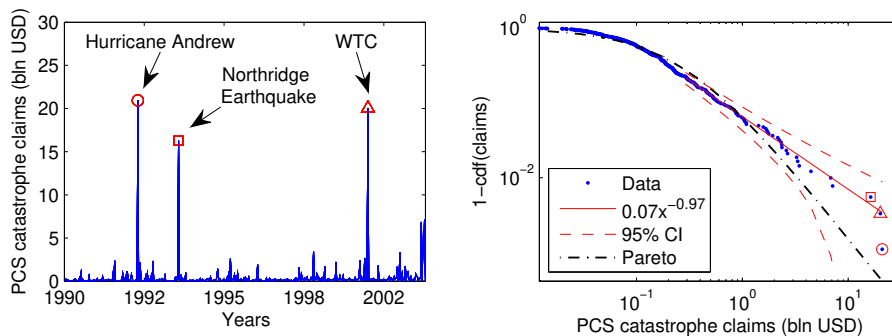
Fig. 2. *Left panel:* PCS catastrophe loss data, 1990-2004. The three largest losses in this period were caused by Hurricane Andrew (24 August 1992), the Northridge Earthquake (17 January 1994) and the terrorist attack on WTC (11 September 2001). *Right panel:* Right tail of the empirical distribution of claim sizes. The three largest losses do not deviate significantly from the fitted power law. However, two out of three seem to be outliers with respect to the Pareto law fit.

in computational efficiency (see the discussion at the end of Section 2), we suggest to use the simpler CLT-based CI. We will also use them in the empirical examples of the following sections.

3.2 Catastrophe claims

Severities of catastrophic events are known to exhibit a heavy-tailed behavior, even after excluding the most extreme outliers [7,10]. We will now check whether these outliers can be considered as dragon kings. To this end, we study the Property Claim Services (PCS) dataset, which covers losses resulting from catastrophic events in the U.S. The data include 1990-2004 market loss amounts in USD, adjusted for inflation using the monthly values of the Consumer Price Index (CPI). Only natural events – except for the terrorist attack on World Trade Center, WTC, and the August 2003 blackout – that caused damage exceeding five million dollars (nominal value, i.e. unadjusted for inflation) were taken into account. The three largest losses in this period were caused by Hurricane Andrew (24 August 1992), the Northridge Earthquake (17 January 1994) and the terrorist attack on WTC (11 September 2001), see the left panel in Figure 2.

An earlier study [4] of the PCS dataset spanning a five-year shorter time period (1990-1999) revealed that the two largest losses of the 1990s (Hurricane Andrew and the Northridge Earthquake) were outliers. At least such a conclusion could be drawn from the Pareto probability plot, see Figure 7.5 in [4]. Using the technique developed in Section 2, we want to check whether this assertion is justified. To this end, we plot the right tail of the empirical distribution of claim sizes, see the right panel in Figure 2, and fit a power law to the 10%-1% largest observations.

The estimated power law exponent of $p = -0.97$ indicates a very heavy tailed distribution (with a decay in the tail comparable to the Cauchy law). However, none of the three largest claim sizes exceed the 95% CI allowing us to conclude that they are not dragon kings, 'merely' black swans. Apparently, while the Pareto law exhibits a power law tail, the fit to the whole PCS dataset is far from perfect. Indeed, the reported maximum likelihood estimate of the Pareto law exponent for the 1990-1999 dataset was twice higher, i.e. the slope of the tail was much steeper [4]. With such a
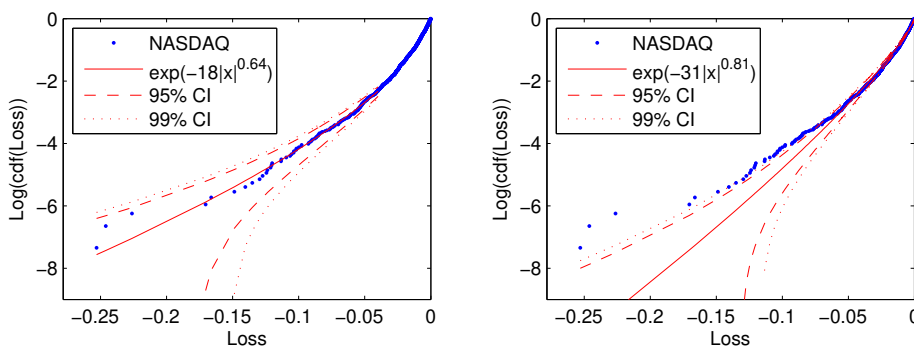
Fig. 3. *Left panel:* Stretched exponential (or Weibull) fit to the 10%-1% largest NASDAQ drawdowns from the period Feb. 5, 1971 – May 31, 2000. Even the most extreme drawdowns cannot be classified as outliers. *Right panel:* Weibull fit to all NASDAQ drawdowns from the same period. The estimated exponent of 0.81 is roughly the same as obtained in [12]. Nearly all of the largest 10% of drawdowns deviate significantly from the Weibull law, indicating that perhaps a heavier-tailed distribution would yield a better fit.

fast decay, the two largest claims in the 1990s would seem to be outliers. However, performing the test introduced in Section 2 after fitting a power law only to the tail of the distribution (as in the right panel of Figure 2) results in the rejection of the dragon king hypothesis.

3.3 Financial drawdowns

Johansen and Sornette [12] claim that dragon kings are common in the distributions of financial drawdowns. In this Section we test whether this is really the case. Following [12], we define a drawdown as the loss between the local price maximum and the following local minimum. It is calculated as the percentage difference between the lowest price and the highest price in a decline period, i.e. the time series $1, 2, 5, 4, 3, 3, 1, 3, 4, 3, 2, 3$ would result in two drawdowns, namely $\frac{1-5}{5}$ and $\frac{2-4}{4}$.

The dataset analyzed here comprises 7661 NASDAQ index closing values from the period Feb. 5, 1971 – May 31, 2000. The data were sampled at daily frequency and obtained from the Reuters EcoWin database. Although the time period studied seems to be the same as the one in [12], denoted there by [1971.1:2000.5], most likely it is not identical. The number of drawdowns we study is 1543, compared to 1495 in [12]. Perhaps the precision of the index closing values is not the same (two decimal places in our dataset) or the ways of computing the drawdowns are slightly different.

In the left panel of Figure 3 we plot the stretched exponential (or Weibull) fit to the 10%-1% largest NASDAQ drawdowns. Note that, in fact, we fit an inverse of the Weibull distribution with the cdf $F(x) = 1 - e^{-\beta|x|^{\tau}}$ defined for negative values of x. Apparently, even the most extreme drawdowns cannot be classified as outliers in this case. However, if we follow the same approach as Johansen and Sornette [12] and using maximum likelihood calibrate the Weibull distribution to all NASDAQ drawdowns we obtain a nearly identical fit: $\beta \approx 31$ and $\tau \approx 0.81$; compare with the values in Table 2 in [12]. This time almost all of the largest 10% of drawdowns deviate significantly from the Weibull law, see the right panel in Fig. 3 and compare it with Fig. 15 in [18]. In our opinion this is not evidence for dragon kings, but rather indicates that a heavier-tailed distribution would better describe the data. At the same time, this simple exercise shows that dragon kings are a model dependent feature. Depending
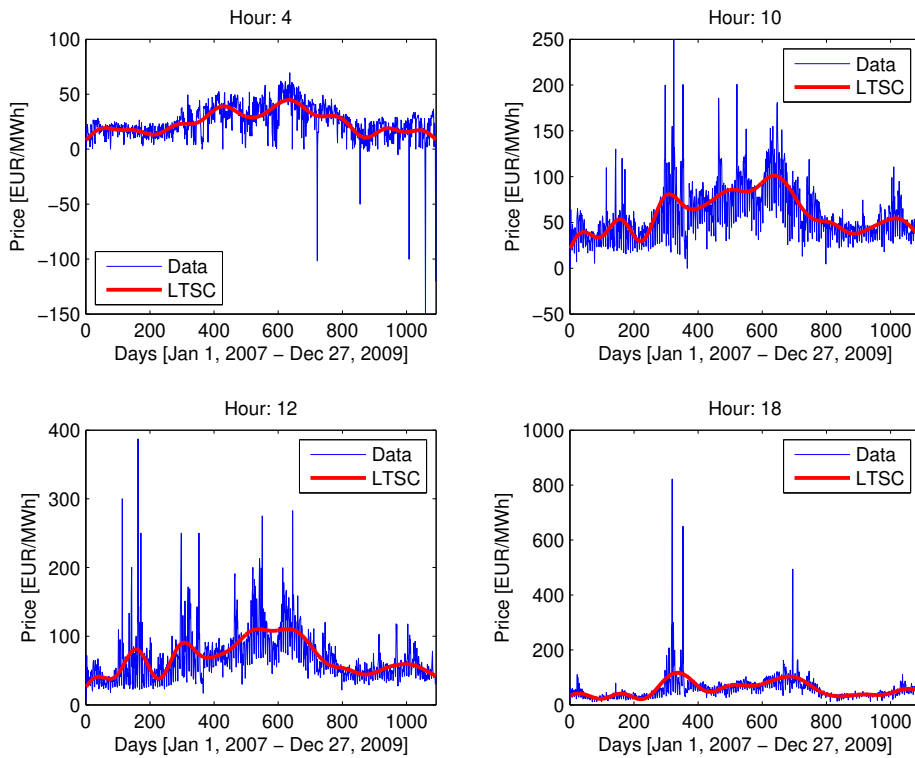
Fig. 4. Electricity spot prices and the estimated long term seasonal components (LTSC) in the German EEX market for four sample hours and 1092 days in the period Jan. 1, 2007 – Dec. 27, 2009. Note, the different y axis scales in the four panels.

on the model (stretched exponential distribution vs. tail) the same observations can or cannot be classified as outliers.

3.4 Electricity spot prices

3.4.1 Why do they spike?

Electricity is a unique commodity and the power markets exhibit behavior like no other financial or commodity markets. Severe weather conditions, often in combination with the execution of market power by some players, have led in the recent past to spectacular price fluctuations – ranging even two orders of magnitude within a matter of hours, see Figures 4 and 5. These abrupt and short-lived price changes are known as spikes and are one of the most profound features of electricity spot prices [3,5,21]. But why do electricity spot prices spike in the first place?

The answer lies in the way these prices are determined. The spot price is typically set in a one- or two-sided uniform-price auction for each hour or a half-hourly time interval of the next day. It is determined as the intersection of the supply curve constructed from aggregated supply bids and the demand curve constructed from aggregated demand bids (for two-sided auctions) or system operator's demand forecast (for one-sided auctions). On the supply side of the market, all generation units of a utility or of a set of utilities in a given region are ranked (and form the so-called supply stack). This ranking is based on many factors, such as the marginal cost of
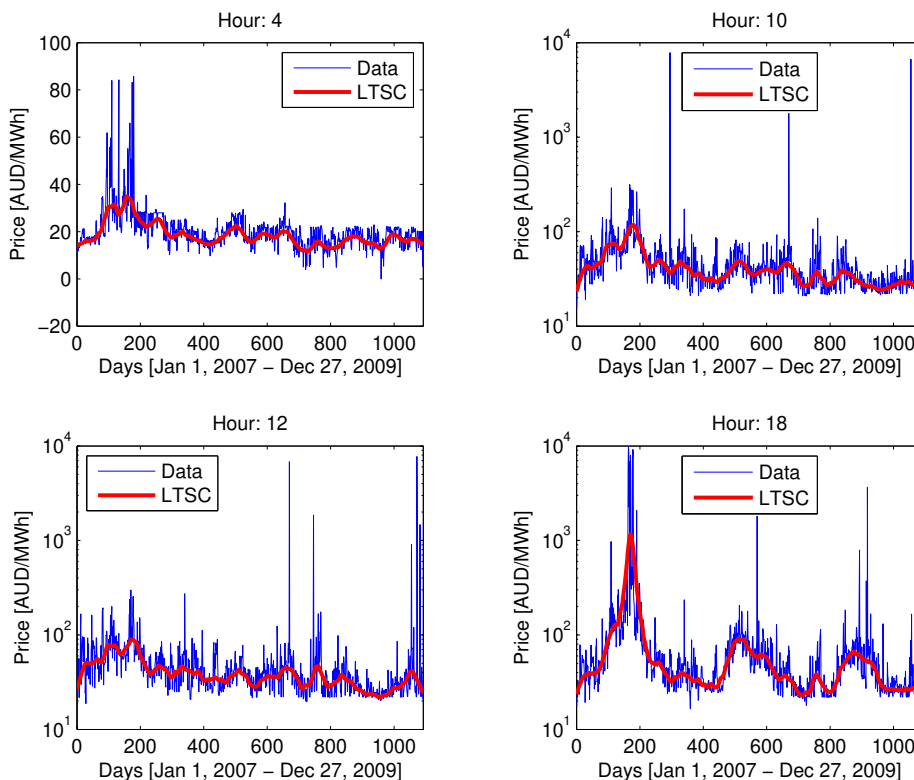
Fig. 5. Electricity spot prices and the estimated LTSC in the Australian NSW market for four sample hours (in fact, half-hourly intervals starting at full hours) and 1092 days in the period Jan. 1, 2007 – Dec. 27, 2009. Note, the semilogarithmic scale for hours 10, 12 and 18. The few extremely spiky prices would render the 'normal' prices invisible on a linear scale.

production and the response time. The utility will typically first dispatch nuclear and hydro units, if available, followed by coal units. These types of plants are generally used to cover the so-called base load, whereas oil-, gas-fired and hydro-storage plants are used to meet peak demand.

Demand, on the other hand, exhibits seasonal fluctuations reflecting variable business activities and changing climate conditions. In Northern Europe and Canada the demand peaks normally in the winter because of excessive heating. In Australia or mid-western U.S. it peaks in the summer months due to air-conditioning. Unexpected weather conditions can cause sudden shocks with demand typically falling back to its normal level as soon as the underlying weather phenomenon is over.

The spot price is not very sensitive to demand shifts when the demand is low, since the supply stack is flat in the low-demand region. However, when demand is high and a larger fraction of power comes from 'expensive' sources, even a small increase in consumption can force the prices to rise substantially. Then, when the demand drops, the price can rapidly decrease to the normal level. Likewise, if the consumption stays almost constant, price spikes can still appear when a considerable amount of 'cheap' generation is withdrawn from the market (due to outages, maintenance, etc.).

While the supply-demand equilibrium explains price volatility it does not, however, justify the extreme severity of the spikes. It is not simply a matter of higher marginal costs. Rather, the spikes are a result of the bidding strategies of the market

12

participants. Since electricity is a non-storable commodity, some agents are willing to pay almost any price to secure a sufficient and continuous supply. On a regular basis they place bids at the maximum level allowed [21]. The risk of having to pay the maximum price is relatively low, because in uniform-price auctions the spot price is what a buyer has to pay for each unit of power irrespective of what he or she did bid initially, as long as the bid was not less than the spot price.

### 3.4.2 Dragon kings or black swans?

We will now try to answer the question whether electricity price spikes should be classified as dragon kings or 'only' as black swans. Or perhaps they are so common that they cannot be even regarded as black swans. To this end, we analyze the hourly (day-ahead) spot prices from two major power markets: the European Energy Exchange (EEX; Germany) and the New South Wales region of the National Electricity Market (NSW; Australia), see Figures 4 and 5. For each market the sample totals 1096 daily observations for each hour of the day ($h = 1, 2, ..., 24$) and covers the 3-year period Jan. 1, 2007 – Dec. 27, 2009.

It is well known that electricity spot prices exhibit several characteristic features, which have to be taken into account when analyzing or modeling such processes [9, 21]. These include seasonality on the annual, weekly and daily level. To cope with it we use the standard time series decomposition approach and let the electricity spot price P_t for a particular hour h be represented by a sum of two independent parts: a predictable (seasonal) component f_t and a stochastic component X_t, i.e. $P_t = f_t + X_t$. Following [22] the deseasonalization is conducted in three steps. First, the long term seasonal component (LTSC) T_t is estimated from the spot price P_t using a wavelet filter-smoother of order 6 (for the EEX prices; for details see [20]) or a Gaussian kernel smoother with a bandwidth of 2^6 (for the NSW prices). In the latter case the kernel smoother is used instead of the wavelet one due to it's lower sensitivity to extreme observations. A single non-parametric LTSC is used here to represent the long-term non-periodic fuel price levels, the changing climate/consumption conditions throughout the years and strategic bidding practices. As shown by Janczura and Weron [11], the wavelet-estimated LTSC pretty well reflects the 'average' fuel price level, understood as a combination of natural gas, crude oil and coal prices.

The price series without the LTSC is obtained by subtracting the T_t approximation from P_t. Next, weekly periodicity s_t is removed by subtracting the 'average week' calculated as the arithmetic mean (for EEX data) or median (for NSW data; again for the sake of robustness) of prices corresponding to each day of the week. Additionally for the EEX dataset, the German national holidays are treated as the eight day of the week. Finally, the deseasonalized prices, i.e. $X_t = P_t - T_t - s_t$, are shifted so that the minimum of the new process X_t is the same as the minimum of P_t.

In Figures 6 and 7 we present the dragon king test results for the right tails of the distributions of deseasonalized price changes, i.e. of ΔX_t, for four selected hours of the day. The more common returns (or changes of the log-prices) cannot be analyzed due to the existence of negative spot prices, which is another peculiarity of electricity markets. Two interesting conclusions can be drawn from these two figures.

First, looking at the price trajectories in Figures 4-5 and the estimated power law exponents in Figures 6-7 it is apparent that Australian prices are more volatile and exhibit heavier-tails than their German counterparts. This can be explained by the fact that Australian markets operate as 'energy only' markets, meaning that the wholesale electricity price should provide compensation to investors for both variable and fixed costs. Indeed it does. For instance, in South Australia the installed capacity increased by nearly 50% in the period 1998-2003, almost half of it being open cycle gas turbines (OCGT) for peaking purposes [21].
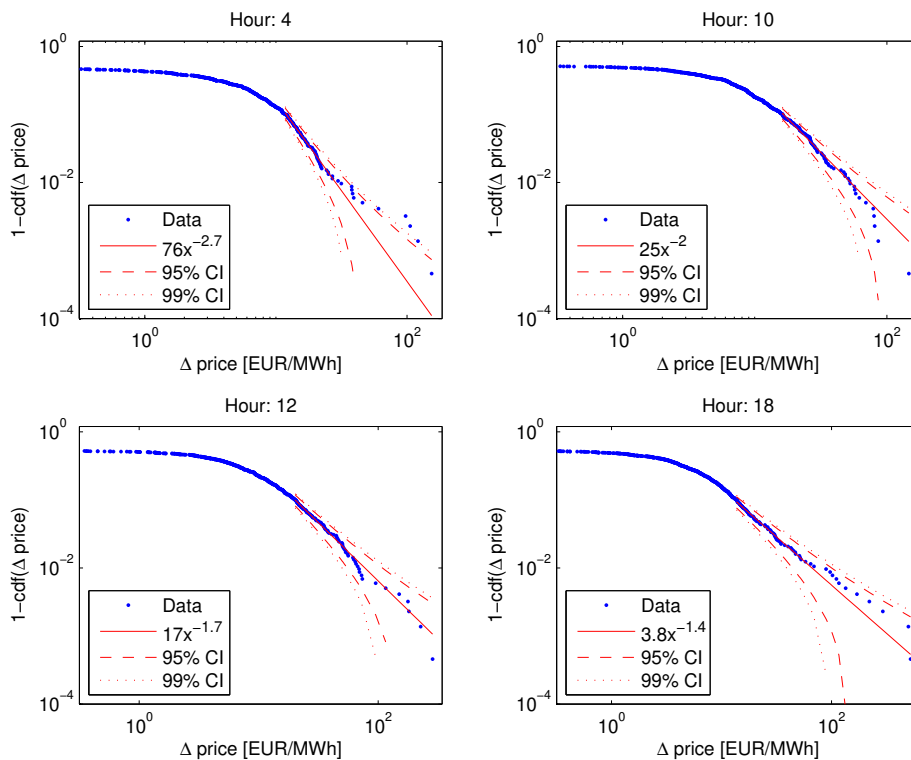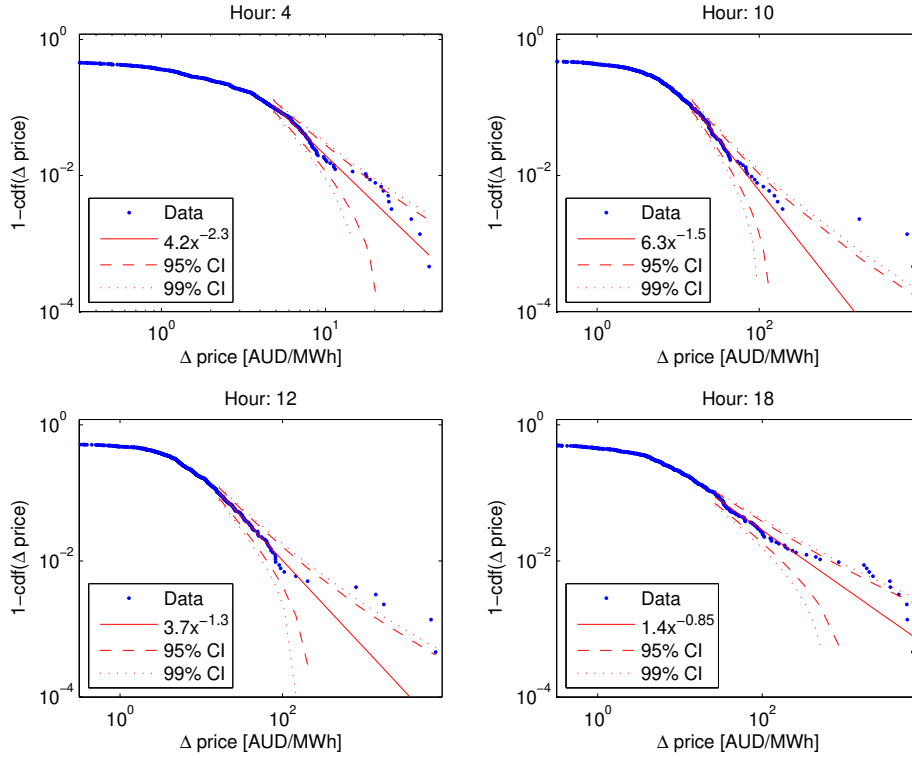
Fig. 6. Right tails of the empirical distribution of electricity spot price changes, computed for EEX prices depicted in Figure 4. The solid lines represent the fitted power law tails (to the lowest or highest 1%-10% of observations). The dashed and dotted curves indicate the 95% and 99% CI, respectively.

Second, in the EEX market only night hours (in particular 4 a.m.) yield price changes that can be classified as outliers, see the upper left panel in Figure 6. This is due to extreme negative prices on some days, possibly resulting from relatively large wind farm generation and very limited demand at this time of night. Somewhat surprisingly, the most spiky afternoon peak hours (like 6 p.m.) do not lead to dragon kings. The price changes on some days are extreme, but they do not deviate significantly from the fitted power law, see the lower right panel in Figure 6. On the other hand, the Australian NSW market is abundant in dragon kings. Nearly all daytime hours (including the three depicted in Figure 7) yield outliers, deviating significantly from the power law tails. Even some night hours (like 4 a.m.) exhibit extreme price changes that are nearly dragon kings.

4 Conclusions

In this paper we have developed a simple test for deviations from power law tails, which is based on the asymptotic properties of the empirical distribution function. We have used it to test whether great natural disasters, financial crashes or electricity price spikes should be classified as dragon kings or 'only' as black swans. While not every observation deviating from a power law (or stretched exponential) fit can be called a dragon king, the bottom line is that outliers to power law tails exist.

Fig. 7. Right tails of the empirical distribution of electricity spot price changes, computed for NSW prices depicted in Figure 5. The solid lines represent the fitted power law tails (to the lowest or highest 1%-10% of observations). The dashed and dotted curves indicate the 95% and 99% CI, respectively.

Following Sornette [18] we can call them dragon kings. We have to stress, however, that the qualification is (strongly) model dependent. As we have shown in Section 3.3, depending on the model (stretched exponential distribution vs. tail) the same observations can or cannot be classified as outliers.

Interestingly, Sornette actually goes one step further and argues that dragon kings may have properties that make them predictable. Now, this is controversial. It's one thing to identify dragon kings but quite another to spot the event that triggers a crash or a price spike. The assertion that dragon kings are more easily predictable than other events requires proofs, which go well beyond the scope of this paper.

Let us just mention one fact that might justify this line of thought. Recently Cartea et al. [6] have shown in the context of power market modeling that predicting the timing of electricity price spikes is possible, at least to some extent, when forward looking information on capacity constraints is taken into account. Unfortunately, in most power markets the availability (to every power market participant) of the reserve margin data is limited.

Acknowledgments

References

1. Bickel, P.J., Doksum, K.A., *Mathematical Statistics, 2nd ed.* (Prentice Hall, 2001).
2. Borgan, Ø., Kaplan-Meier estimator, in *Encyclopedia of Biostatistics*, eds. P. Armitage, T. Colton (Wiley, 1998).
3. Bottazzi, G., Sapio, S., Secchi, A., Physica A **355** (2005) 54-61.
4. Burnecki, K., Weron, R., Visualization Tools for Insurance Risk Processes. In *Handbook of Data Visualization*, eds. Ch. Chen, W. Härdle, A. Unwin (Springer, 2008) 899-920.
5. Byström, H.N.E., International Review of Economics and Finance **14** (2005) 41-55.
6. Cartea, A., Figueroa, M., Geman, H., Applied Mathematical Finance **16(2)** (2009) 103-122.
7. Chernobai, A., Burnecki, K., Rachev, S.T., Trück, S., Weron, R., Computational Statistics **21** (2006) 537-555.
8. Čížek, P., Härdle, W., Weron, R. (eds.), *Statistical Tools for Finance and Insurance, 2nd ed.* (Springer, 2011).
9. Eydeland, A., Wolyniec, K., *Energy and Power Risk Management* (Wiley, 2003).
10. Härdle, W., Lopez Cabrera, B., Journal of Risk and Insurance **77** (2010) 625-650.
11. Janczura, J., Weron, R., Energy Economics **32(5)** (2010) 1059-1073.
12. Johansen, A., Sornette, D., Journal of Risk **4(2)** (2001) 69-110.
13. Klein, J.P., Moeschberger, M.L., *Survival Analysis* (Springer, 2003).
14. Klugman, S.A., Panjer, H.H., Willmot, G.E., *Loss Models: From Data to Decisions, 3rd ed.* (Wiley, 2008).
15. Paolella, M.S., *Intermediate Probability: A Computational Approach* (Wiley, 2007).
16. Pisarenko, V.F., Rodkin, M.V., Izvestia, Physics of the Solid Earth **44(1)** (2008) 1-8.
17. Shaywitz, D.A., *Shattering the Bell Curve*, The Wall Street Journal, April 24 (2007) D8.
18. Sornette, D., International Journal of Terraspace Science and Engineering **2** (2009) 1-18.
19. Taleb, N.N., *The Black Swan: The Impact of the Highly Improbable* (Random House, 2007).
20. Trück, S., Weron, R., Wolff, R., Bulletin of the International Statistical Institute **62**, 1524. Available at MPRA: http://mpra.ub.uni-muenchen.de/4711/.
21. Weron, R., *Modeling and Forecasting Electricity Loads and Prices: A Statistical Approach* (Wiley, 2006).
22. Weron, R., Mathematical Methods of Operations Research **69(3)** (2009) 457-473.