Modelling Extreme values of the PX Index returns

Ján Gogola [1]

Abstract

In this contribution we focused on the daily log returns of investment in the Prague stock exchange index, PX-Index. We analysed data from January 1st, 1995 to June 30th, 2018. We can see that the data has fatter left and right-hand tails than the normal distribution. Conclusions of our basic analysis are that the daily log returns are leptokurtic and heavy tailed. They are not i.i.d. and volatility varies over time. Further we investigated extreme values of daily log returns. The focus is on how Extreme Value Theory fares in contrast to the assumption of normally distributed losses. Also we can say that extreme daily log returns appear in clusters.

Key words

PX index, daily log returns, extreme value theory, generalized extreme value distribution, risk measures

JEL Classification: C13, C40, C60

1. Introduction

Risk managements are motivated to search methodologies able to cope with rare events that may have serious consequences for them. The question arises: "If things go wrong, how wrong can they be? Extreme value theory (EVT) provides a firm theory on which we can build statistical models describing extreme events. In many fields of modern science, engineering, hydrology and insurance (see e.g. Embrechts (1999), Jindrová (2015), Pacáková (2009)), extreme value theory is well suited. Our contribution deals with the behaviour of the daily returns of the PX Index. The PX Index is the official index of major stocks that trade on the Prague Stock Exchange.

Before we turn to the topic of modelling extreme values of PX Index returns, it is worthwhile to consider and review typical characteristics of financial market data. Typical characteristics of financial market data are summarized in the literature (see e.g. McNeil et. al. (2005)) as "stylized facts". These stylized facts are:

- time series data of returns, in particular daily return series, are in general not independent and identically distributed (i.i.d.);
- the volatility of return processes is not constant with respect to time;
- the absolute or squared returns are highly autocorrelated;
- the distribution of financial market returns is leptokurtic. The occurrence of extreme events is more likely compared to the normal distribution;
- extreme returns are observed closely in time (volatility clustering);
- the empirical distribution of returns is skewed to the left, negative returns are more likely to occur than positive.

[1] Ján Gogola, RNDr. Ph.D., Fakulta ekonomicko-správní, Univerzita Pardubice, jan.gogola@upce.cz

First we will check whether these stylizes facts are valid to the returns of the PX Index. Then we continue with the EVT. Two approaches to modelling extreme values are presented: the block maxima method and the peaks-over-threshold method. More treatment of the EVT we can find in Pfaff (2013). Our results are accomplished by means of the R language (a free statistical software) and its currently available packages.

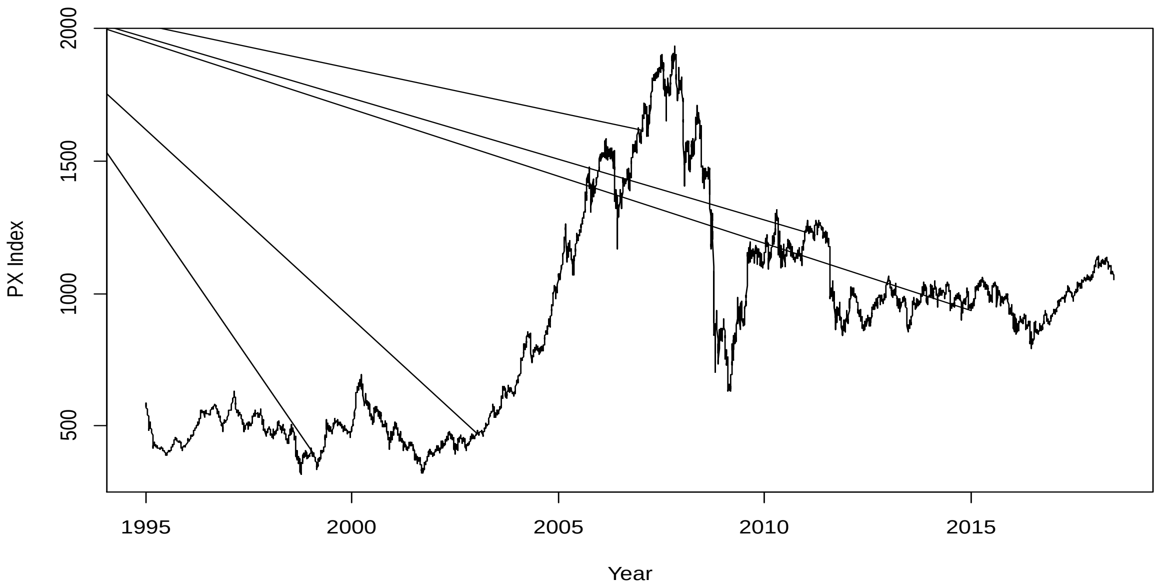
2. Data and methodology

The PX Index is the official index of major stocks that trade on the Prague Stock Exchange. The index was calculated for the first time on March 20, 2006 when it replaced the PX 50 and PX-D indices. The index took over the historical values of the PX50. The starting day of PX 50 was April 5, 1994 and its opening value was fixed at 1 000 points. At this time the index included 50 companies on the Prague Stock Exchange.

Since data in the year 1994 are irregular, we decided to analyse data from from January 1st 1995 to June 30th, 2018. Figure 1. shows the development of the PX Index. From the beginning of 1995 to about 2004 we can see something that looks like business cycles. Business cycles of this type might exist but the cycles are all of different lengths, the timing of the peaks and the lows are difficult to predict. The PX Index reaches its top on October 29, 2007 with 1936 points. As result of financial crisis reached 700 points on October 27, 2008 losing almost 50% of its value in two months.

In our contribution we are going to focus on the daily log returns (Figure 2.) and analyze these returns.

Figure 1: PX Index (in CZK) from 1. 1. 1995 to 30. 6. 2018. Source: www.pse.cz [6]



Particularly, the focus is on the use of extreme value theory to analyse losses (left tail) of the PX Index.

Our first approach considers the maximum the variable takes in successive periods. These selected observations constitute the extreme events, also called ***block maxima***. Assume that a sequence of random variables $X_1, X_2, \dots, X_m$ over a time span of *n* periods is given. The time span could be a calendar period such as a month, quarter, half year, or year. With respect to EVT, the question arises as to which distribution the maximum of these random variables, $M_n = \max\{X_1, X_2, \dots, X_k\}$, follows or, to put it more precisely, is asymptotically best approximated by. When modelling the maxima of a random variable, EVT plays the same fundamental role as the central limit theorem plays when modelling sums of random variables. In both cases, the theory tells us what the limiting distributions are. The limit law for the block maxima M_n, is given by the following theorem Pfaff (2013):

Theorem A: Let $\{X_m\}$ be a sequence of i.i.d. random variables. If there exist constants $c_n > 0$, $d_n \in R$ and some nondegenerate distribution function G such that

$$P\left(\frac{M_n - d_n}{c_n} \le z\right) \to G(z), \text{ for } n \to \infty,$$

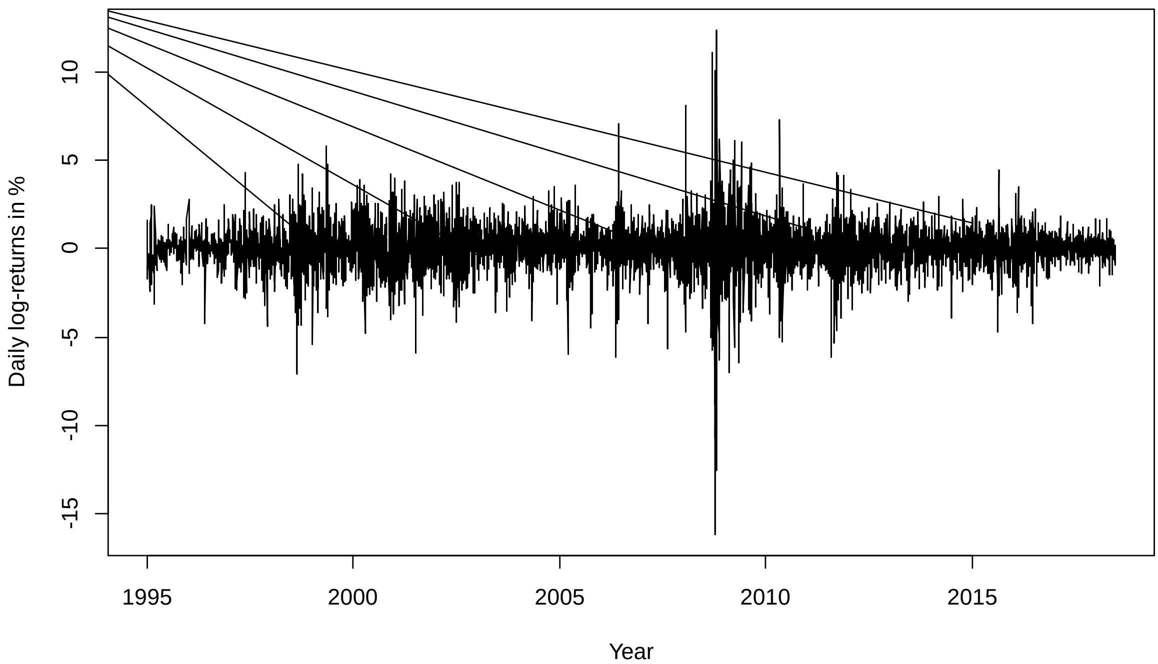
then G belongs to one of the following distributions: Gumbel, Fréchet or Weibull.

The three distributions can be subsumed into the ***generalized extreme value*** (GEV) distribution,

$$G(z) = \exp\left\{-\left[1 + \xi \cdot \left(\frac{z-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}\right\}. \tag{1}$$

The GEV is a three-parameter distribution where μ is the location, σ the scale and ξ the shape parameter. For the limit $\xi \to 0$ the Gumbel distribution is obtained, for $\xi > 0$ the Fréchet, and for $\xi < 0$ the Weibull. The Weibull has a finite right point. The density is exponential in the case of Gumbel and polynomial for the Fréchet distribution. Hence, the characteristics and properties of the GEV can be deduced from the value of the shape parameter. Fréchet distribution has a polynomially decaying tail and therefore suits well heavy tailed distributions.

Figure 2: Percentage daily return on the PX Index (5876 observations). Source: Own calculation



Our second approach focuses on the realizations exceeding a given (high) threshold u - the peaks-over-threshold method. This can be summarized for a given threshold u by the following probability expression:

$$P(X > u + y \mid X > u) = \frac{1 - F(u+y)}{1 - F(u)}, \; y > 0. \tag{2}$$

In practice the distribution $F(x)$ is generally unknown and hence, similarly to the derivation of the GEV, one needs an approximative distribution for sufficiently large threshold values.
The law is given by the following theorem Pfaff (2013):

Theorem B: For a large class of underlying distribution function F the conditional excess distribution function $F_u(y) = P(X - u \le y \mid X > u)$, for u large is well approximated by $F_u(y) \approx G_{\xi,\sigma}(y),\ u \to \infty$,
where

$$G_{\xi,\sigma}(y) = \begin{cases} 1-\left(1+\dfrac{\xi}{\sigma}y\right)^{-\frac{1}{\xi}}, \text{if } \xi \neq 0 \\ 1-e^{-\frac{y}{\sigma}}, \text{ if } \xi = 0 \end{cases}. \tag{3}$$

$G_{\xi,\sigma}$ is the so called ***generalized Pareto distribution*** (GPD).

The shape parameter ξ gives an indication of the heaviness of the tail, the larger ξ, the heavier the tail.
In practice, a difficulty arises when an adequate threshold has to be selected. An adequate threshold value can be determined graphically by means of a mean residual life (MRL) plot. This kind of plot is based on the expected value of the GPD: $E(Y) = \dfrac{\sigma}{1-\xi}$. For a given range of thresholds u the conditional expected values

$$E(X - u \mid X > u) = \frac{\sigma_u + \xi \cdot u}{1-\xi} \tag{4}$$

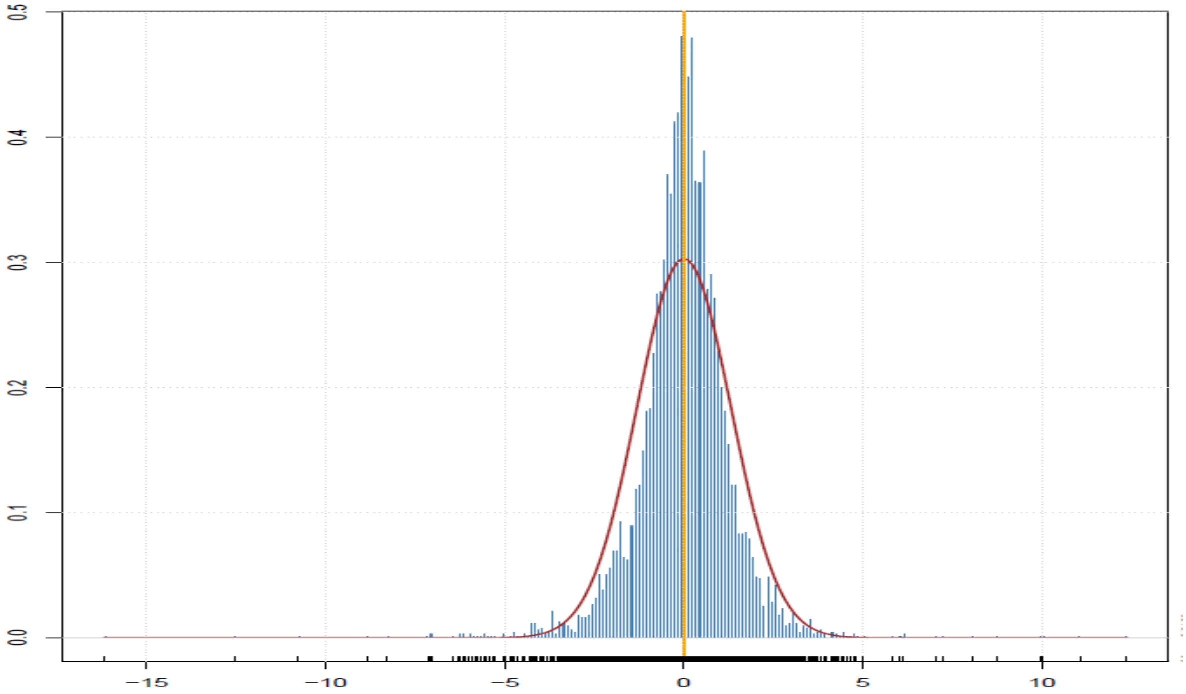
are plotted against u.
This equation is linear with respect to the threshold u. A suitable value for u is given when this line starts to become linear.

3. Results

First we will establish certain stylised facts about daily returns. The PX Index daily returns data has a skewness of $\sqrt{b} = -\ 0.455$ and a kurtosis of k = 12.55. These empirical coefficients look quite different from 0 and 3 respectively. The data is clearly non-normal from these analyses. We add to this by looking at the data using graphical techniques, such as histogram and Q-Q plot.

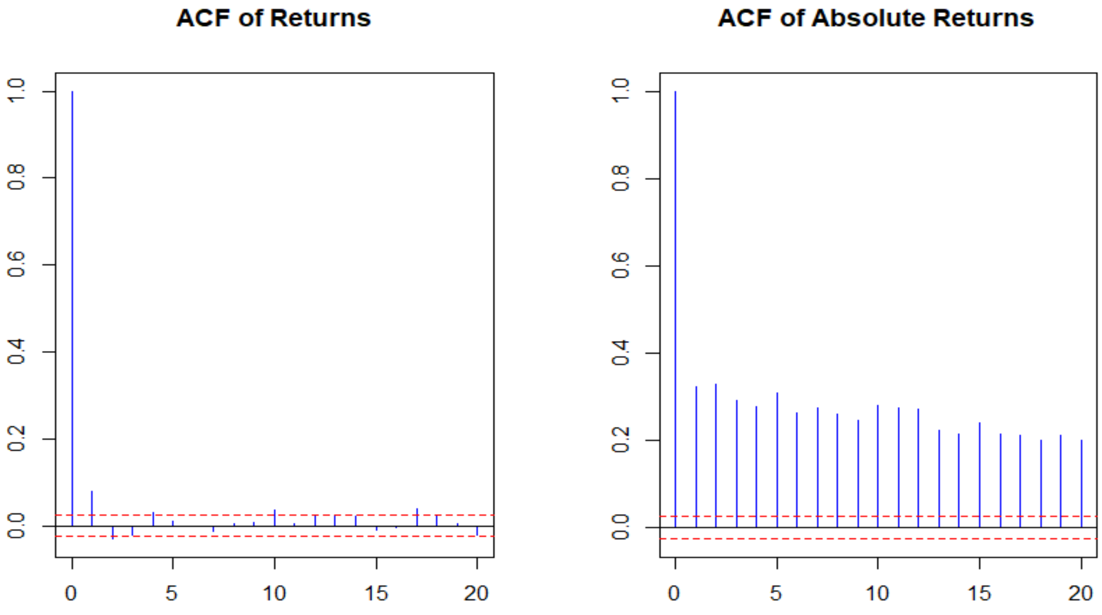
We have plotted in Figure 3. the histogram of the daily log returns on the PX Index. We have also drawn in the density function for the $N(\hat{\mu},\hat{\sigma}^2)$ distribution. We can easily see from this that the data exhibit a narrower peak than the *best-fitting* normal distribution. Less obviously, but certainly a feature of the data is, that it has a fatter left and right-hand tails than the *best-fitting* normal distribution. A Q-Q plot is a dot plot that plots the ordered sample against the corresponding quantiles of the distribution that we are considering to model the data. If the data were genuinely normally distributed then we would expect to see the 5876 points much more in a straight line. The fact that Figure 5. (left) actually exhibits an inverted "S" shape means that the data has fat left and right-hand tails. The downturn in the plot at the left-hand end means that the left-hand tail is fatter than the normal distribution: in other words we should expect rather more large losses over time than we would predict using the Normal distribution. This inverted "S" shape therefore points to the data being leptokurtic.

Figure 3: Histogram of the percentage daily returns on the PX Index. Source: Own calculation



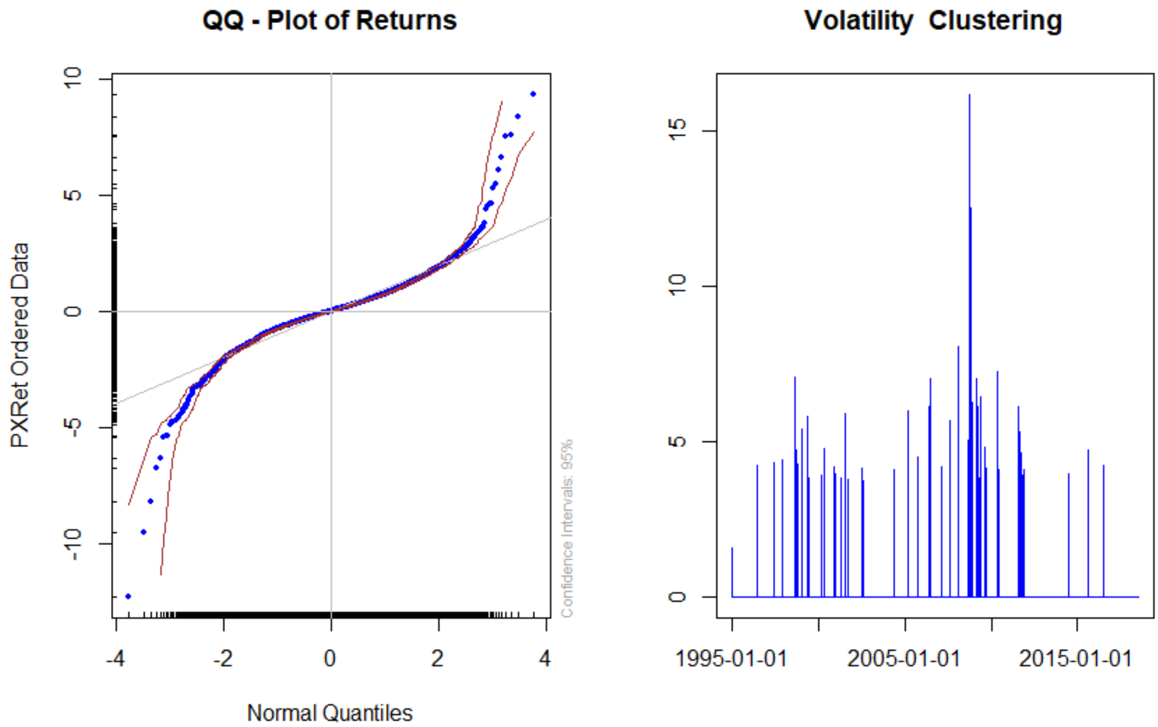
The less-formal graphical/diagnostic tests clearly indicate that the assumption that returns are normally distributed is not valid. Additionally Figure 5. (right) also suggests that the daily log returns are not i.i.d.. Instead, it looks like there are clear clusters of high and low volatility. The PX Index log returns have clusters of high volatility (e.g. in 2008) and low volatility (e.g. 2013).

Figure 4: Left: Sample autocorrelation function for the volatility-standardised residuals.
Right: Sample autocorrelation function for absolute value of volatility-standardised residuals.
Source: Own calculation



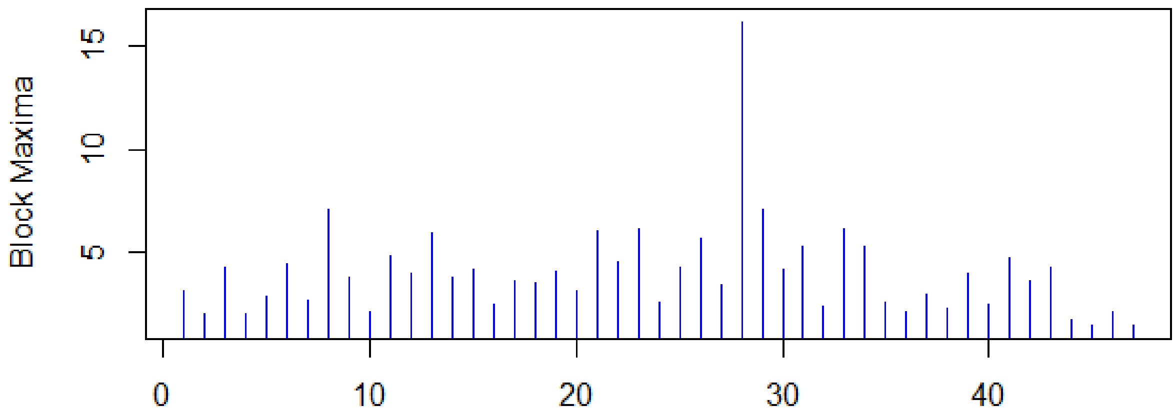
The autocorrelation function (ACF) for the daily returns (Figure 4. - left) hint at a slight autocorrelation of first order. The ACF of the absolute daily returns (Figure 4. - right) looks differently. There is a moderate, but nevertheless highly significant correlation between next days. Clearly, these are significantly different from zero and taper off only slowly. The significant autocorrelations in the absolute values of daily returns imply that the market goes through phases of high and low volatility.

Figure 5: Stylized facts for the PX Index. Source: Own calculation



Now we can apply the extreme value theory to the PX Index data. First the daily returns are converted to positive figures expressed as percentages. The application of the method of block maxima requires the following steps: divide the sample in *n* blocks of equal length, collect the maximum value in each block, fit the GEV distribution to the set of maxima. The key point of this method is the appropriate choice of the periods defining the blocks. We choose half-year periods to get 47 blocks (which are likely to be sufficiently large for *Theorem A* to hold).

Figure 6: Block maxima (left) for the PX Index. Source: Own calculation



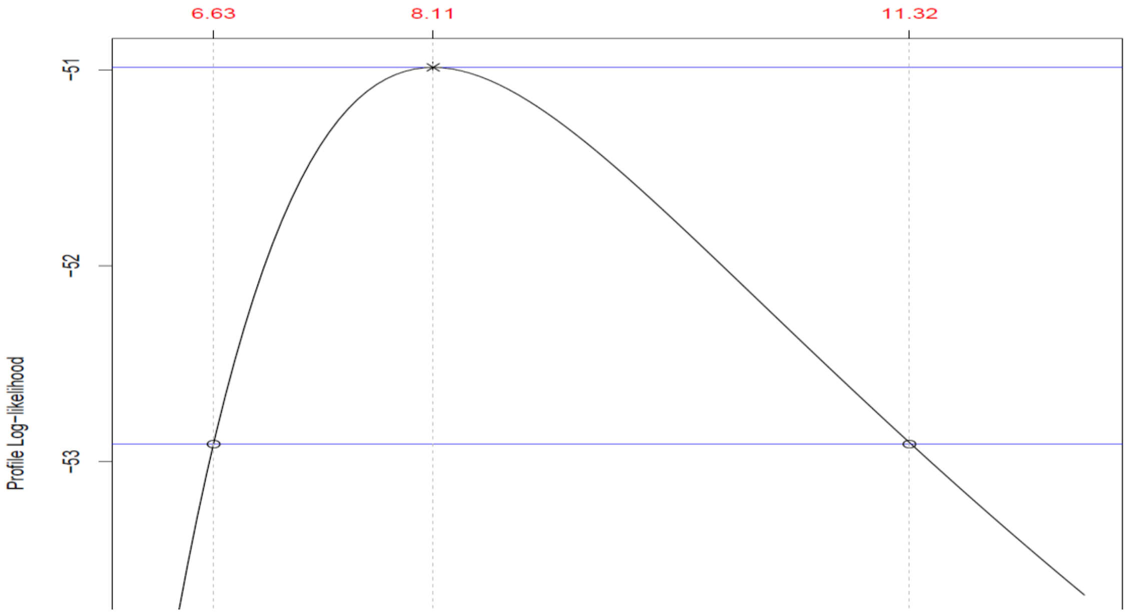
The maximum return in each of the blocks (Figure 6.) constitute the data points for the sample of maxima *M* which is used to estimate the GEV distribution. Point estimates for the parameters are given in Table 1. Positive value for ξ implies that the limiting distribution of maxima belongs to the Fréchet family.

Table 1: Fitted GEV to block maxima of PX Index

GEV	ξ	σ	μ
Estimate	0.178	1.263	3.010
Standard Error	0.119	0.165	0.210

Further inference from the model can be made using the profile log-likelihoods. Figure 7. shows those for a 10-year return level. A daily loss as high as 8.11 % would be observed once every 10 years. This point estimate falls within in a 95% confidence level ranging from 6.63 % to 11.32 %. As can be clearly seen, the confidence level (the horizontal line) is asymmetric.

Figure 7: Profile log-Likelihood for fitted GEV model



One way to better exploit information about extremes in the data sample is to use the peak-over-threshold (POT) method. The POT method requires the following steps:
select the threshold u, fit the GPD function to the exceedances over u.
The key point of POT method is the selection of the threshold u. Theory tells us that u should be high. But the higher the threshold the less observations are left for the estimation of the parameters of the tail distribution function. There is no clear algorithm for the selection of the threshold u. A graphical tool that is very helpful for the selection of the threshold u is the MRL plot. The sample mean excess function should be linear. This property can be used as a criterion for the selection of u. Figure 8. shows the MRL plot corresponding to the PX Index data. From a closer inspection of the plot we suggest the value u = 3 %. For the given threshold a total of 104 exceedances result. This data set corresponds roughly to the upper 98% quantiles of the empirical distribution function.

Table 2: Fitted GPD of PX Index, threshold = 3 %

GPD	ξ	σ
Estimate	0.153	1.300
Standard Error	0.099	0.180

We know that the distribution of the observation above the threshold in the tail should be a generalized Pareto distribution (GPD). We compute the distribution parameters (Table 2.) that maximize the log-likelihood function for the sample defined by the observations exceeding the threshold u.

Figure 8:MRL plot for PX Index losses

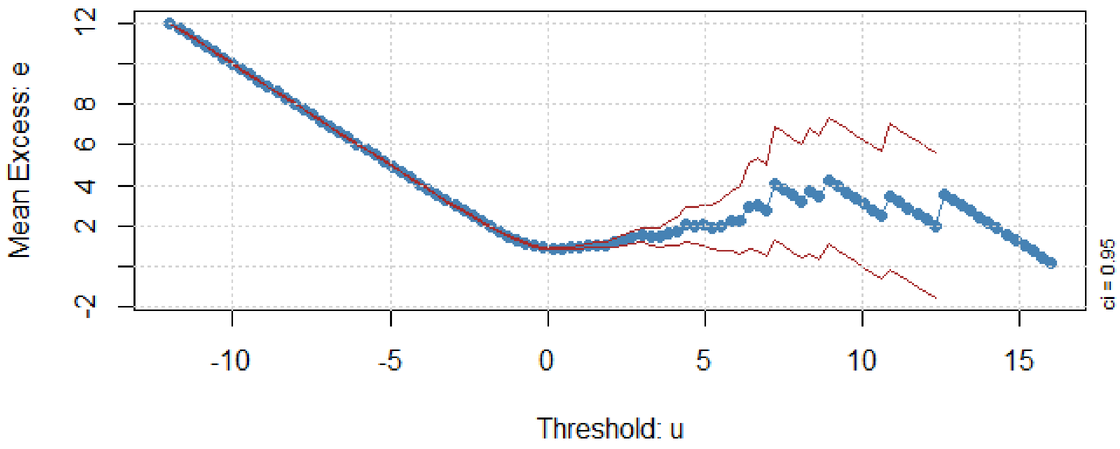


Table 4: Risk measures for PX Index

Confidence level	VaR	ES
95.0 %	1.752	3.061
99.0 %	3.775	5.451
99.5 %	4.813	6.676

4. Conclusion

We have illustrated how to analyse extreme values of daily returns of the PX Index. Our conclusion is that EVT can be useful for assessing the size of extreme events. The Value at Risk (VaR) and expected shortfall (ES) risk measures can be inferred directly from EVT. Point estimates for the VaR and ES risk measures (Table 4.) are computed for the 95%, 99% and 99.5% levels. These measures would qualify as unconditional risk assessment for the next business day.

References

[1] Embrechts, P., Klüppelberg, R. J., Mikosh, T. (1999). Modelling Extremal Events for Insurance and Finance, Application of Mathematics, Springer, New York, 2nd ed.

[2] Jindrová, P., Jakubínský, R. (2015). *Significance and possibilities of major accident insurance*. *E + M Ekonomie a Management*, 18(4), 121-131.

[3] McNeil A. J., Frey R., Embrechts P. (2005) *Quantitative Risk Management*. Princeton University Press, Princeton and Oxford, ISBN 13: 978-0-691-12255-7.

[4] Pacáková V., Linda B. (2009). *Simulation of Extreme Losses in Non-life insurance*. *E+M Ekonomie a management*, XII, no. 4., pp. 97-103, ISSN 1212-3609

[5] Pfaff, B. (2013). *Financial Risk Modelling and Portfolio Optimization with R*, John Wiley & Sons Ltd, West Sussex, UK, ISBN: 978-0-470-97870-2

[6] R Core Team. (2015). *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, http://www.R-project.org/ (accessed).

[7] www.pse.cz/indexy/hodnoty-indexu/historicka-data