Analysis of occurrence of extremes in a time series with a trend

Petr Volf[1]

Abstract. We consider a random series of values and are interested in the analysis and modeling the occurrence of extremes. There exist several possible approaches. One of them is the analysis of sequence of block maxima. As we assume that the series has a trend, we first select a proper regression model for the block maxima development. From it, a Markov chain of the sequence of extremes is derived. As the transition probabilities of the chain are not tractable analytically, we use the Monte Carlo generation of the chain behavior. Then, from the sample representing the series of block maxima development we obtain a representation of corresponding predictive distribution. Finally, we shall apply such a method to real data.

Keywords: extreme, regression model, random walk, prediction.

JEL classification: C41, J64
AMS classification: 62N02, 62P25

1 Introduction

Let us consider a sequence of maximal values in a series of random variables $X_1, X_2, \ldots$. New maximum (record) is established when $X_{t+1} > max\{X_1, ..., X_t\}$. The case of i.i.d. (independent, identically distributed) random variables X_t has been analyzed by many authors (cf. Embrechts et al, 1997). It has been shown that the probability of new record at time t is proportional to $1/t$, and that the sequence new record values $R_1 < R_2 < \ldots$ behaves as a random point process with intensity $h_x(r)$ equal to the intensity of distribution of random variables X_t.

However, quite frequently the assumption of i.i.d. variables is not adequate. Especially in cases when the series of variables is dependent and changes along certain trend. This contradiction between reality and the i.i.d. scheme led to the construction of models describing the sequence of extremes (i. e. values, increments, times) with the aid of convenient functional models for intensity, regression, or time-series (though we shall speak mostly on maxima, the same concerns the minimal values).

In the paper we are interested in the following questions: In Section 2 we deal with the problem of trend model fitted to the data. We use the approach of block maxima, i.e. the data are reduced to maximal values over certain periods. It has an advantage that the dependence in a series is reduced, on the other hand some local extremes are lost. The statistical tools for the model fit diagnostics are recalled, too. Then the method of the prediction of further series development is proposed. It is based on the simulation, as the numerical computations are hardly tractable in this case. The simulation of future trajectories enables us to estimate prediction bands, i.e. curves which are crossed with given probability (see "Peaks over threshold", POT method, for instance in Beirlant et al, 2004). Notice that when we deal with block maxima, the POT approach yields a piece-vise constant threshold curve, which is also one of possible choices of thresholds of POT analysis for the whole data series.

In Section 3 we recall the attempts to model the occurrence of extremes as a random point process. We actually link up both approaches, by the formulation of random walk process of new extremes based on the analysis of trend of block maxima. Finally, the last part deals with a brief illustration of presented methods.

[1]Institute of Information Theory and Automation, Prague 8, e-mail: volf@utia.cas.cz

2 A model of trend

Let $X(t)$ be a series of maximal values in periods t, so that t is the discrete time, periods $t = 1, 2, \dots$ given by our selection of data. We shall consider the following model form:

$$X(t) = m(t) + r(t),$$

where $m(t)$ stands for a trend function, $r(t)$ are the residuals, random errors. We allow for their autoregressive structure, so that

$$r(t) = \sum_{j=1}^{K} a_j \cdot r(t-j) + \sigma \cdot \varepsilon_t,$$

where ε_t are already i.i.d. $N(0, 1)$ (standard normal) variables, $\sigma > 0$ is constant. Naturally, during the analysis those assumptions have to be checked. An alternative can consider time-dependent variance, either given by a function of time (as in Volf, 2011) or for instance by an ARCH model. As regards the selection of trend function, we consider two possibilities:

1. A parametrized function corresponding to the shape of trend. Frequent choices are S-shape Gompertz function or exponential decay curve (see Kuper and Sterken, 2003).
2. Trend constructed from a linear combination of functional units (eg. polynomials, goniometric functions, polynomial splines or others). Optimal selection of units and a degree of model is achieved with the aid of tests of parameters significance, the choice may be supported by the use of some penalized criterion, for instance the BIC. However, as we are interested in a global trend and also in the possibility to extrapolate (predict) it, the choice of basic functions is rather limited.

Estimation procedure is then based on the method of least squares, even in the case of autoregression among errors $r(t)$. It can seem that, because we deal with values of block maxima, the use of GEV (generalized extremal values) distribution should be preferred to normal errors. In such a case, method of maximal likelihood should be employed. However, the difference is negligible, as we show also in our example.

Diagnostics of model fit: The goodness-of-fit of selected model, namely the correspondence of errors ε_t to the standard normal distribution, can be tested both graphically (by the Q-Q plot) and numerically, e.g. with the aid of the Kolmogorov-Smirnov test. The selection of degree of eventual autoregression is standardly based on the maximum likelihood estimation (i.e. the mean squares in the Gauss distribution case) of autoregression parameters, on the tests of their significance, and also may be supported by the BIC criterion.

The constantness of σ, i.e. the homoskedasticity of remaining term, can be tested by the White test (White, 1980). It is based on the coefficient of determination in the linear regression of squared residuals (i.e. estimated $\sigma \cdot \varepsilon_t$) on regressors contained in $m(t)$. The test statistics is the coefficient of determination multiplied by the sample size, its critical value is given by corresponding chi-square quantile with $p - 1$ degrees of freedom, where p is the number of parameters in $m(t)$.

Finally, the independence of ε_t can be tested for instance by simple nonparametric tests ("series above and below median", "series up and down").

Generally, the models with smooth trend do not consider any change of conditions, though such changes are quite frequent in practice. Then, the analysis can be amended by a method searching for potential changes, as well as by the detection of outlied values.

3 Models of random point process

Such models are based on the notion of intensity of new extreme occurrence (cf. Embrechts at al, 1997, Beirlant et al, 2004, and references there). The methodology is borrowed from the survival analysis. Models also allow to incorporate the dependence of intensity on influencing factors, for instance in the framework of Cox's regression model. In order to enlarge the point process scheme to the description of both new extremes times and values, we can use a model of compound point process. Compound process means the process of random increments at random times, formally

$$C(t) = \int_0^t Z(s)\,\mathrm{d}N(s),$$

where $Z(s)$ are (nonnegative) random variables and $N(s)$ is a counting process. If $N(s)$ has intensity $\lambda(s)$ and the mean and variance of $Z(s)$ are $\mu(s)$, $\sigma^2(s)$, respectively, then the mean development of $C(t)$ is given as

$$EC(t) = \int_0^t \lambda(s)\,\mu(s)\,\mathrm{d}s \quad \text{and} \quad \operatorname{var} C(t) = \int_0^t \lambda(s)\,\left(\mu^2(s) + \sigma^2(s)\right)\mathrm{d}s.$$

Now, both components of compound process can depend on explaining factors, via conveniently selected regression model (as is for instance the model presented in Volf, 2005). In the discrete-time case, i.e. also in the block maxima approach, we register just whether the new maximum was achieved or not in certain period. Then the compound process changes to a random walk model. It is described by probabilities $p(t)$ of new extreme occurrence in period t and random variables $Z(t)$ of its increase.

3.1 New maximum occurrence and value

Let us assume that up to time 0 the maximal value of the series was R. Further, let the block maxima in following periods be described as (continuous type) random variables $X(t)$, $t = 1, 2, \ldots$, with probability densities, distribution functions, survival functions $f_t,\ F_t,\ , S_t = 1 - F_t$, respectively. Then the probability that a new maximum will occur in period k is

$$p(k, R) = P\left\{X(j) < R, j = 1, 2, .., k-1,\ X(k) > R\right\} = \tag{1}$$

$$= \left\{\prod_{j=1}^{k-1} P(X(j) < R)\right\} \cdot P(X(k) > R) = \left\{\prod_{j=1}^{k-1} F_j(R)\right\} \cdot S_k(R)$$

when $X(t)$ are independent (conditionally, given the trend function). Further, the new maximum value is given by the density

$$g_k(r, R) = P(X(k) = r | X(k) > R) = \frac{f_k(r)}{S_k(R)}, \quad \text{for } r > R, \tag{2}$$

provided k is the period of new maximum occurrence.

Therefore, when the joint distribution of $X(t)$ is known (estimated, in the present context), the distributions of random variables T_R - the period of new record occurrence, and Z_R of new maximum improvement, can be derived easily. Namely,

$$P(T_R = k) = p(k, R), \quad k = 1, 2, \ldots$$

and the distribution of Z_R has the density

$$g(z, R) = \sum_{k=1}^{\infty} p(k, R) \cdot g_k(R + z, R). \quad \text{for } z > 0. \tag{3}$$

In most instances, including the case of normal distribution, these formulas can be evaluated just numerically.

3.2 A Markov chain of extremes, its prediction

The process of growth of the maximal value can also be treated as a Markov chain, with discrete time and continuous state space. Again, assume that at time t the actual maximum value is R_t. Then, at time $t+1$, no change will occur with probability $P(R_{t+1} = R_t) = P(X(t+1) \le R_t) = F_{t+1}(R_t)$, while the transition probability to a higher value $r > R_t$ is given as in (2), namely by density $g_{t+1}(r, R_t) = f_{t+1}(r)/S_{t+1}(R_t)$.

Such a Markov scheme is convenient for random generation of future process paths. Namely, assume that the data up to period T, $X(1), \ldots, X(T)$, are available and that the parameters of trend model are estimated from them. The objective is to predict the process for next periods. First, the trend of $X(t)$ is extrapolated to $t > T$. On this basis, we generate random trajectories of Markov process of extremes described above, starting from value R_T at T. From a large number of such random trajectories, the development of certain sample characteristics of future process course can be computed. For instance the mean values, variances as well as different quantiles.

3.3 Prediction bands and POT view

Random generation of future trajectories of analyzed process yields a sample of them, say $f_m(t), t \in (T, T_1), m = 1, .., M$. Then the point-wise (at each t) 'prediction' intervals for $X(t)$ are obtained immediately from sample quantiles of f_m at t. Methods for construction of prediction bands on the whole interval (T, T_1) could be, theoretically, connected with the concept of 'depth of data' (see for instance Zuo and Serfling, 2000). Practically, the approach corresponds to the construction of multivariate quantiles, for instance in the following way: Let us consider a sample of functions $f_m(t)$ given empirically by values at the same set of points $t_j; j = 1, .., J$. For each $k < M/2$, point-wise k/M or $(M-k)/M$ sample quantiles (i.e. at each t_j) can be constructed. If we join them to a band, we can try to find such k that, approximately, a given proportion (95%, say) of functions lies below it. As an additional finer criterion we can compare numbers of points at which the quantiles are crossed. In other words, in such a way we construct an empirical version of the threshold which is crossed (on the whole interval) just with probability 5%.

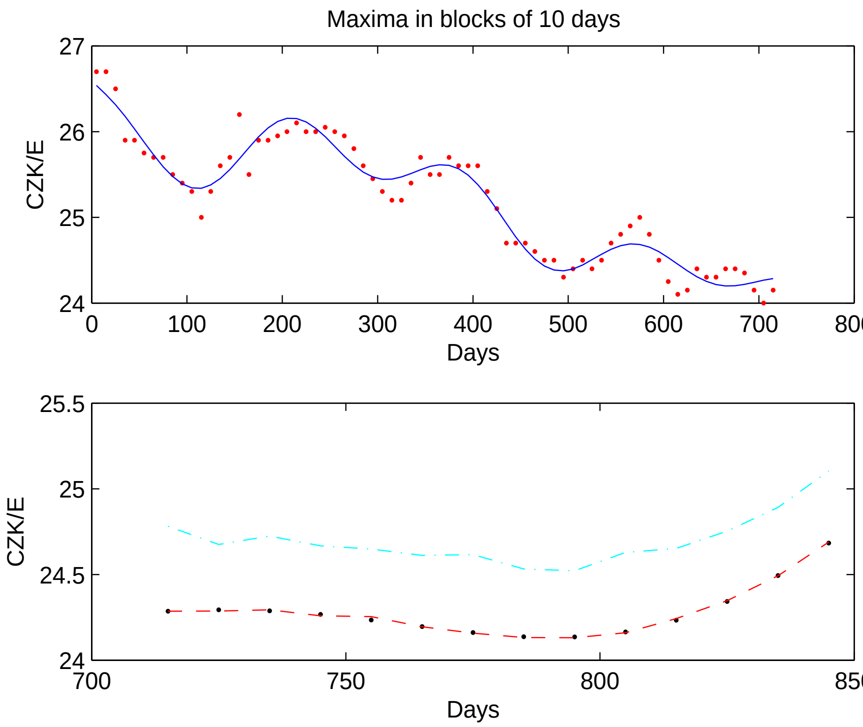

Figure 1: Block maxima and trend function (above), extrapolation of trend and 90% prediction band

4 Application

Figure 1, in its upper subplot, displays, by points, the development of exchange rate of CZK to Euro, namely 10 days maxima during approximately last 2 years (May 2009 to April 2011). As we know, the trend was decreasing, moreover, the data show certain seasonal components and also other non-regular disturbances. The plot contains also fitted trend curve. It was created as a linear combination of certain basic function, their selection was optimized as described in Part 2, following the approach 2. Namely,

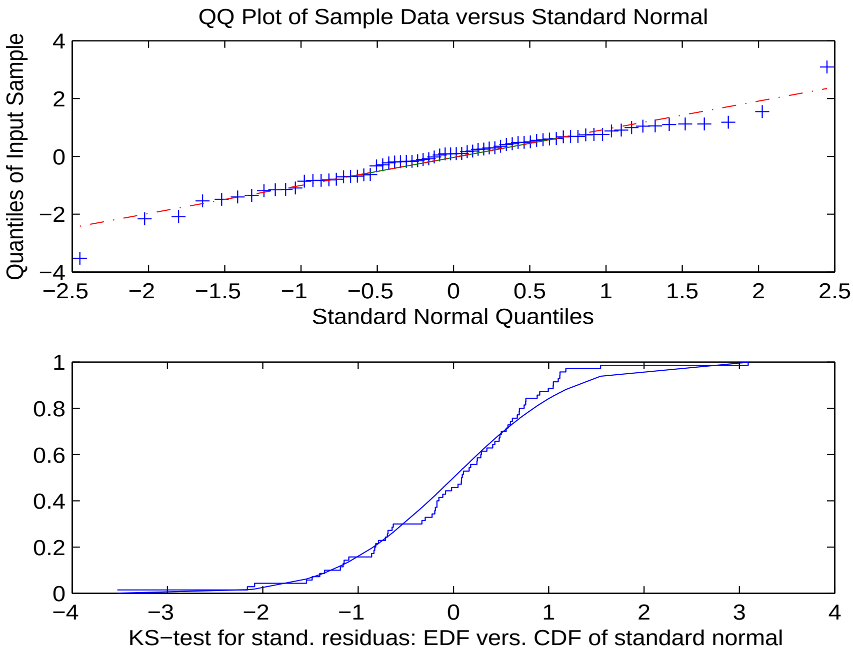

Figure 2: Q-Q plot (above) and K-S test (below) comparing empirical distribution of standardized errors ε_t with standard normal distribution

we obtained

$$m(t) = b_1 + b_2 \cdot t + b_3 \cdot t^2 + b4 \cdot cos(2\pi/T) + b_5 \cdot sin(4\pi/T) + b_6 \cdot sin(6\pi/T) + b_7 \cdot sin(8\pi/T) + b_8 \cdot cos(8\pi/T). \quad (4)$$

As T is actually the range of times t, $sin(2k\pi/T)$ is a function with k periods during T. Hence, the trend function has also several periodic components, with a period of 1 year and also with another period half-year long.

Further, it was detected that the residual values behaved as AR(1) series,

$$r(t) = a \cdot r(t-1) + \sigma \cdot \varepsilon_t.$$

Parameters were estimated by the least squares method, with results (half–widths of 95% confidence intervals are in parentheses):

$b_1 = 27.05412\ (18.01561)$, $b_2 = -0.00896\ (0.00271)$,

$b_3 = 0.000008\ (0.000004)$, $b_4 = -0.58707\ (0.20364)$,

$b_5 = -0.25745\ (0.06260)$, $b_6 = -0.12884\ (0.05968)$,

$b_7 = 0.21106\ (0.05863)$, $b_8 = 0.11985\ (0.05860)$,

$a = 0.42870\ (0.21395)$.

Residual standard deviation σ was estimated as $s = 0.15344$.

Figure 2 shows the Q-Q plot (above) and graphical version of the Kolmogorov-Smirnov test (below), both comparing empirical distribution of realizations of ε_t with $N(0,1)$ distribution. Neither graph contradict to the good fit. Numerically, in K-S test the maximal departure of empirical and hypothetical distribution function was 0.0903 while the critical value for $n = 72$ is larger, 0.1601.

The homoskedasticity of residual term was tested with the White test (described in Part 2). The test statistics yielded 12.27, while the 95% quantile of the chi-square distribution with 7 degrees of freedom was 14.07, so that the hypothesis of constant σ was not rejected.

The assumption of independence of errors ε_t was checked, too. Two nonparametric tests ("series above and below median", "series up and down") were employed. The independence was not formally rejected by any test, P-values were 0.0576 and 0.1406, respectively (though these P-values are quite close to their critical border).

Finally, we also tried to predict the behavior of series $X(t)$ for next 13 periods 10 days long. The result is displayed in Figure 1, lower subplot. The points correspond to extrapolated trend curve $m(t)$, at $t = 715, 725, ..., 845$. We then generated 1000 realizations of series of future block maxima, following the approach described in subsection 3.2. Dashed curve close to $m(t)$ is connecting point-wise sample medians obtained from them. Remaining curve, dot-dashed, is the threshold which was crossed just by 100 (i.e 10%) of trajectories, in other words, it is an empirical (so that approximate) 90% prediction band for maxima in the series of exchange rate development.

5 Conclusion

We have proposed a model for the occurrence of extremal values, taking into account the series of block maxima. Its development is represented by a nonlinear regression (trend) model. From it, the Markov chain of new extremes occurrence and values has been derived. While, explicitly, the model depends just on time, an implicit dependence of new maximum on the past maximum duration and value is involved, too. An application to real data has shown usefulness and good performance of the model. A future improvement should concentrate to the problems of detection (and prediction) of changes in analyzed time series. Except the use of statistical methods for changes and outliers detection, one can think also on selection of informative factors indicating the changes of conditions, and on methods of pattern recognition for the analysis of those factors.

Acknowledgements

The research is supported by the project of GA CR No 402/10/0956.

References

[1] Beirlant, J., Goegebeur, Y., Segers, J., and Teugels, J.: *Statistics of Extremes: Theory and Applications.* Wiley, New York, 2004.

[2] Embrechts, P., Kluppelberg, K., and Mikosch, T.: *Modelling Extremal Events.* Springer, Berlin, 1997.

[3] Kuper, G.H. and E. Sterken (2003): Endurance in speed skating: the development of world records, *European Journal of Operational Research* **148** (2003), 293–301.

[4] Volf, P.: Hazard rate model and statistical analysis of a compound point process, *Kybernetika* **41** (2005), 773–786.

[5] Volf, P.: A stochastic model of progression of athletic records, *IMA Journal of Management Mathematics* **22** (2011), 157–169.

[6] White, H.: A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity, *Econometrica* **48** (1980), 817-838.

[7] Zuo, Y. and Serfling, R.: General notion of statistical depth function. *Annals Statist.* **28** (2000), 461–482.