Smooth Extremal Models in Finance and Insurance

V. Chavez–Demoulin and P. Embrechts

Department of Mathematics
Swiss Federal Institute of Technology
CH–1015 Lausanne, Switzerland
`http://statwww.epfl.ch/people/chavez/`

and

Department of Mathematics
ETH-Zentrum, HG G37.1
CH–8092 Zürich, Switzerland
`http://www.math.ethz.ch/~embrechts/`

May 11, 2001

1 Introduction

Extreme Value Theory (EVT) has developed very rapidly over the past two decades both methodologically and with respect to applications. Whereas (non–life) actuaries have, at least implicitly, used EVT techniques for a long time, mainly through the emergence of quantitative Risk Management, EVT has entered only more recently the finance stage as a useful toolkit for describing non–standard (more precisely non–normal) price fluctuations. Econometricians for a long time were well aware of the so–called stylized facts of market data, which clearly showed that normal distribution based models (i.e. Brownian motion technology) are only a first step into the direction of finding more realistic models. What was perhaps not so clear were the next steps, i.e. how to use this "heavy–tailed" reality in pricing, hedging, portfolio management, risk management and even banking and insurance regulation. For the latter, despite the well supported evidence for non–normality, such tools like mean–variance optimization, Value–at–Risk (VaR), Sharpe–ratio etc. play a very dominant role. EVT offers a pair of glasses through which to look at these types of questions more realistically. Embrechts, Klüppelberg and Mikosch (1997) detail the mathematical theory of EVT and discuss its applications to financial and insurance risk management. Various updating material is to be found under the URLs http://www.math.ethz.ch/finance and http://www.risklab.ch. In the edited volume Embrechts (2000), various papers highlight the current state of the art on EVT modelling in Integrated Risk Management (IRM). See also Reiss and Thomas (2001) for a very readable discussion.

The traditional approach to EVT is based on extreme value limit distributions. Here, a model for extreme losses, say, is based on the possible parametric form of the limit distributions of maxima over independent, identically distributed (iid) (or weakly dependent) data; see for instance Embrechts, Klüppelberg and Mikosch (1997), p. 121. Whereas the original data may not be iid, perhaps by considering maxima over blocks of data within certain periods, one may hope to reduce the data to fairly uncorrelated maxima observations (for details of this so–called annual maxima method, see Embrechts, Klüppelberg and Mikosch (1997), p. 317). A more flexible model is based on a so–called point process characterization. The resulting Peaks Over Threshold (POT) method considers exceedances over a threshold u. For a pictorial presentation of the POT method, see Figure 1. In Figure 1, $Z_1, \ldots, Z_q$ denote the ground up losses

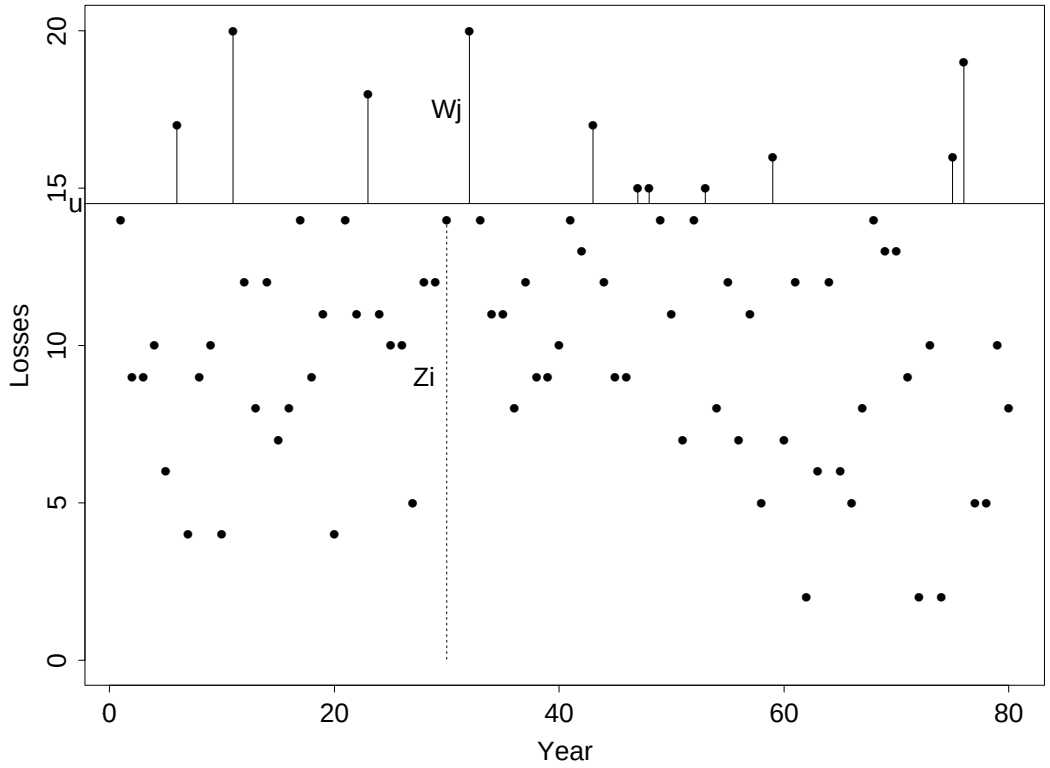

Figure 1: *The point process of exceedances (POT).*

(say), u a (typically high) threshold, n the number of exceedances by $Z_1, \ldots, Z_q$ of the level u, and $W_1, \ldots, W_n$ the corresponding excesses (loss exceeding u minus u). The level u may for instance correspond to the attachment point, or the lower level of an excess–of–loss reinsurance treaty; within finance, u could stand for a VaR number when managing market risk or a stress loss value within credit risk. A further example would correspond to larger operational losses $W_1, \ldots, W_n$ above a threshold u. Mathematical theory (see Leadbetter (1991)) supports the condition of a possibly inhomogeneous Poisson process with intensity λ for the number of exceedances combined with independent excesses W over the threshold. Given u, the excesses are treated as a random sample from the generalized Pareto distribution (GPD), with scale parameter σ and shape parameter κ; see (1) in Section 2 for the basic definitions.

An additional advantage of the threshold method over the method of annual maxima is that, since each exceedance is associated with a specific event, it is possible to let the scale and shape parameters depend on covariates. For instance, insurance losses can be of different types (so–called lines), credit loss data typically will be a function

of credit scores, business type, exogenous economic variables, time and other information. Operational losses will typically belong to various subclasses (fraud, system failures, backoffice errors, ...) and their occurrence no doubt shows a non–constant (often stochastic) intensity, possibly depending on such factors as business cycles, volume etc. Large losses may become more or less frequent over time, or indeed they may become more or less severe. It is also well–known that in general, insurance and financial losses show cyclic behaviour.

In this paper we discuss some of the more recent EVT methodology which may be useful in handling the presence of such covariates and the resulting modelling of extremal events. Several contributions in this context have been made: Smith and Shively (1995) assess the probability of high–level exceedances in the tropospheric ozone record as a function of meteorological information. Rootzén and Tajvidi (1997) use meteorological information in wind storm insurance and present a detailed analysis of Swedish wind storm claims. In particular, to obtain data as homogeneous as possible, they consider separately the storms from six Swedish meteorological and hydrological stations. They apply maximum likelihood theory to fit a GPD distribution with constant shape parameter κ and with $\sigma(t) = \exp\{\alpha + \beta t\}$, with t denoting time in years. They motivate this choice by the fact that previous experience of similar situations indicate that a constant κ should be reasonable and that the model for the scale parameter corresponds to a constant growth in cost, by $100 \times \left(e^{\beta} - 1\right)\%$ per year. Chavez-Demoulin (1999) finds that in an environmental context, when applied to extreme temperatures, a fully parametric form for the parameter κ often suffices.

The natural variability of the exceedances tends to mask any trends or other dependence on time. While variation due to the different covariates such as type of claims, of losses or geographical locations could be summarized parametrically, changes in time need not have a specific parametric form. In this paper we therefore propose to combine the point process for exceedances with smoothing methods to give a flexible exploratory approach to model changes in large values for insurance or financial data. In doing so, we closely rely on the methodology developed by Chavez-Demoulin (1999) and Chavez-Demoulin and Davison (2001).

A possible model might consist of an inhomogeneous Poisson process for the number of exceedances, with intensity of the form $\lambda(t) = \exp\left\{x^T\alpha + f(t)\right\}$, combined with the generalized Pareto distribution for the sizes of exceedances (the excesses) with

a parameterization of the form $\kappa(t) = x^T\beta + g(t)$ and $\log\sigma(t) = x^T\gamma + s(t)$ where α, β and γ are vectors of parameters and f, g and s are smooth functions (see Section 2 for the basic POT notation). The vector of covariates x can depend on time also, in particular taking into account possible discontinuities in λ, κ and σ, due for example to a worldwide crisis such as the one in the insurance industry of the early 1990s, or stockmarket crashes like 1987, 1998. Other reasons for discontinuous effects could be the announcement of economic policy decisions like interest rate changes.

Problems can arise when statistically identifying the functions g and s. These can be avoided by working with so–called orthogonal parameters. We might use either the reparameterization $\{\kappa, \nu(\kappa,\sigma)\}$ such that the parameters κ and ν are orthogonal with respect to the Fisher information metric or the reparameterization $\{\zeta(\kappa,\sigma),\sigma\}$ such that the parameters ζ and σ are orthogonal. As κ is hard to estimate and physically more stable than σ, we prefer to use the parameterization (κ,ν). Following the orthogonalization technique described in Cox and Reid (1987), we find the parameter $\nu = \sigma(1+\kappa)$ to be orthogonal to κ. With this parameterization we find that, as pointed out above, $g(t)$ is usually constant or at worst linear in t. Below, we shall use $\nu(t) = \exp\left\{x^T\eta + s(t)\right\}$. Whatever statistical estimation method we use, we are faced with a mixture of a finite dimensional problem (parameters α, β, η) and an infinite dimensional one (functions f, g, s). In order to handle the latter, some smoothness assumptions must typically be made. Estimation algorithms will carry a penalty component which is a function of the amount of smoothness we require for the functions f, g, s. Obviously, we could also restrict these functions to finitely parameterized classes of functions; we prefer however to let the data speak for themselves on this crucial time dependence and hence go for a fairly general and versatile model. The construction of such a model requires semiparametric techniques. Having observed $w_1,\ldots,w_n$, we might estimate α, β, η, f, g and s using maximum likelihood estimation based on penalized loglikelihood criteria. A motivation for the use of a procedure based on penalized loglikelihood is that it treats the entire dataset as a single entity. We use a Fisher scoring algorithm which has a clear justification through the penalized loglikelihood and furthermore allows for the incorporation of different smoothing methods.

The paper is organized as follows. In Section 2, we review some of the stochastic techniques underlying the threshold (POT) method. In Section 3, we propose a smoothing methodology as a new tool for practical extreme value exploration in

finance and insurance. In Section 4, the methods introduced will be applied to some idealized examples from insurance and finance.

2 The Threshold Method

The approach based on the threshold method considers a characterization of all observations which are extreme in the sense of having exceeded a high threshold u. Consider a sequence of independent and identically distributed random variables $Z_1, \ldots, Z_q$ from a distribution $F(z)$ in a wide class of continuous distribution functions. The number of exceedances over the level u has a Poisson distribution with mean λ and conditional on n exceedances, the excesses $W_j = Z_j - u$ are a random sample of size n from the generalized Pareto distribution (GPD)

$$G_{\kappa,\sigma}(w) = \begin{cases} 1 - (1 - \kappa w/\sigma)_+^{1/\kappa}, & \kappa \neq 0, \\ 1 - \exp(-w/\sigma), & \kappa = 0. \end{cases} \tag{1}$$

As $\kappa \to 0$, $G_{\kappa,\sigma}(w)$ tends to the exponential distribution with mean σ. Equation (1) can be used as the basis for a likelihood for σ and κ which is

$$l(\sigma, \kappa) \doteq -n \log \sigma - (1 - 1/\kappa) \sum_{j=1}^{n} \log (1 - \kappa w_j/\sigma)_+ ,$$

and the Poisson Process loglikelihood in term of λ, σ, κ is then

$$l(\lambda, \sigma, \kappa) \doteq n \log \lambda - \lambda - n \log \sigma - (1 - 1/\kappa) \sum_{j=1}^{n} \log(1 - \kappa w_j/\sigma)_+ . \tag{2}$$

In deriving (2), one obviously uses the (asymptotic) independence of the frequency and sizes of the losses over a high threshold u. Maximum likelihood estimation of the parameters κ and σ of a generalized Pareto random variable is non–regular in the sense that the score statistic is not asymptotically normal if $\kappa > 1/2$ (Davison (1984a, 1984b), Smith (1985)). For $\kappa > 1$ the GPD has infinite mean and so, whereas the usual Taylor expansions can be made, they do not yield a consistent estimator. Typically, in most applications the value of κ is close to zero and consistency and asymptotic efficiency of the maximum likelihood estimator hold. The generalized Pareto distributions yield a practical family for statistical estimation, provided that the threshold is taken sufficiently high.

The choice of the threshold is important. Smith (1987) proposes a graphical technique to get an aid for choosing the threshold and to assess the fit of the model; a "mean residual life plot" (see Yang (1978)) in which the mean excess over a threshold u is plotted against u, for a wide range of values u. See Davison and Smith (1990) and Embrechts, Klüppelberg and Mikosch (1997) for an extensive discussion of this approach. Matthys and Beirlant (2000) contains a nice review.

The level exceeded on average once in $1/p$ years (or any other relevant time period), called the $1/p$–year return level, is often a quantity of interest. Based on the threshold model its value is

$$y_{1-p} = u - \frac{\sigma}{\kappa}\left\{(\lambda/p)^{-\kappa} - 1\right\}, \tag{3}$$

which may be estimated by replacing σ, κ and λ by their maximum likelihood estimates. Interval estimates may be obtained by the delta method or by a reparameterization in terms of $(y_{1-p}, \lambda, \kappa)$, treating κ and λ as nuisance parameters, and solving (3) for σ. This method is also referred to as the profile likelihood approach.

Independence of widely separated extremes seems reasonable in most applications, but they almost always display short–range dependence in which clusters of extremes occur together. Serial dependence will typically imply clustering of large values: hot days tend to occur together, for example, and sea level maxima often occur during the strongest storm of the year. Likewise volatility bursts will produce clustered extremes in financial data, whereas environmental factors may result in clustering of catastrophic events like storms and floods. In these cases, it seems unrealistic to assume independence within each period, a year, say. In the threshold method, the usual solution is to fit the point process model to cluster maxima, as the use of the GPD for the peak excess in each cluster is justified. An important practical problem is the identification of clusters from data, provided that the cluster size is random and its distribution depends on the local correlation of the Z_i. The identification of clusters has been much influenced by earlier work like Leadbetter *et al.* (1983), and is a topic of much current research. Following Davison and Smith (1990), Robinson and Tawn (2000) propose a run approach for both the choice of a suitable high threshold and of a method to identify independent clusters. Suppose that a stationary series $Z_1, \ldots, Z_q$ has short range dependence, so extremes occur in clusters of mean size $1/\theta$, where $0 < \theta \leq 1$; θ is called the "extremal index" for the data and $\theta = 1$ corresponds to asymptotically independent clusters of size 1. For a fixed threshold u, we identify

different groups of exceedances over u as independent clusters only if there are at least v consecutive observations over u between them. Robinson and Tawn (2000) estimate the number of independent clusters for a sample of n observations by

$$C_n(u,v) = \sum_{j=1}^{n-v} X_j \left(1 - X_{j+1}\right) \cdots \left(1 - X_{j+v}\right) ,$$

where $X_j = I\left(Z_j > u\right)$. The estimate of the extremal index is then $\hat{\theta} = C_n(u,v)/N(u)$, where $N(u)$ is the number of exceedances over u. Two important remaining points are the choices of u and v. In practice, we recommend the use of sensible choices based on the specific situation. See Embrechts, Klüppelberg and Mikosch (1997, Section 8.1) for more details and alternative estimation procedures. An example showing how estimation can go wrong when θ is not taken into account in a financial risk management context, is to be found in Embrechts, Resnick and Samorodnitsky (1999).

3 The Description of the Methodology

In this section, we present the methodology that applies smoothing techniques to extreme values in an insurance or financial context. In these contexts, there is a need for exploratory data analysis to gain an understanding of the structure of the data. The goal of the method proposed is to offer some more advanced exploratory data analysis tools for financial and insurance risk management in the presence of extremal events. Theoretical results and details of the approach are described in Chavez-Demoulin (1999) and Chavez-Demoulin and Davison (2001).

As we see below, modelling of exceedance times in terms of a Poisson process combined with independent excesses over a threshold facilitates the procedure by permitting separate modelling of the number of exceedances and their sizes. Moreover, when modelling the two parameters of a GPD, it is straightforward to reparameterize the problem using orthogonal parameters. Chavez-Demoulin (1999) gives an example where a non–orthogonal parameterization leads to computational difficulties and non–convergence of the algorithm. This could arise when smoothing is performed directly on the three parameters of the generalized extreme value distribution (GEV) $H_\kappa\left(\frac{x-\theta_1}{\theta_2}\right)$ where

$$H_\kappa(x) = \begin{cases} \exp\left\{-(1-\kappa x)_+^{1/\kappa}\right\} , & \kappa \neq 0 , \\ \exp\{-\exp(-x)\} , & \kappa = 0 , \end{cases}$$

for which it is difficult to find an orthogonal reparameterization. For the link between the GPD and the GEV, see Embrechts, Klüppelberg and Mikosch (1997, Section 3.4).

Suppose a sequence of excesses results from the threshold method for which we recall the basic properties; see Leadbetter (1991):

i) The excesses over a high threshold u occur at the times of a Poisson process with intensity λ.

ii) The corresponding sizes over u are independent and have a GPD(κ, σ) distribution.

iii) Exceedance sizes and exceedance times are independent of each other.

For a precise formulation of the above properties, see the quoted reference of Leadbetter. The way we formulated i), ii), and iii) allows for an immediate application to the construction of likelihood functions. Hence the resulting process is of the so–called compound Poisson type. For this process, the overall loglikelihood in terms of the parameters λ, κ, σ is

$$l(\lambda, \kappa, \sigma) = \sum_{i=1}^{T} \left[n_i \log \lambda_i - \lambda_i - n_i \log \sigma_i - (1/\kappa_i + 1) \sum_{j=1}^{n_i} \log \{1 + \kappa_i w_j / \sigma_i\} \right] ,$$

where T is the observed number of time periods (the number of years, say), n_i is the number of exceedances during the ith year, and $w_k = z_k - u$, $k = 1, \ldots, N$, are the excesses, having observed a total number of $N = \sum n_i$ data $z_1, \ldots, z_N$ over a threshold u. Due to the following "cut" of the loglikelihood

$$l(\lambda, \kappa, \sigma) = l(\lambda) + l(\kappa, \sigma) ,$$

estimation can be performed separately for the point process of exceedance times and for the excesses. Based on Leadbetter's results, semiparametric models for exceedance times and excesses will be proposed.

Assuming an inhomogeneous Poisson process for the exceedance times, with intensity $\lambda(t) = \exp\left\{x^T \alpha + f(t)\right\}$, the point process part defines a semiparametric generalized linear model within the Poisson family. Following the approach of Green and Yandell (1985), we use the Fisher scoring algorithm to maximize the penalized loglikelihood

$$l(\lambda) - \frac{1}{2} \gamma_\lambda \int f''(t)^2 \, dt ,$$

where γ_λ is a smoothing parameter. Estimation and inference procedures are the usual ones for the exponential family (Green and Silverman (1994)).

Consider the semiparametric generalized Pareto model. We aim to fit $\kappa(t) = x^T\beta + g(t)$, $\nu(t) = \exp\{x^T\eta + s(t)\}$, and for that reason, attempt to maximize the penalized loglikelihood

$$l(\kappa,\nu) - \frac{1}{2}\gamma_\kappa \int_a^b g''(t)^2\,dt - \frac{1}{2}\gamma_\nu \int_a^b s''(t)^2\,dt\,, \tag{4}$$

where γ_κ and γ_ν control the degree of smoothing applied to the shape and scale parameters respectively. Here,

$$l(\kappa,\nu) = \sum_{i=1}^{T}\left[n_i\log(1+\kappa_i) + n_i\log\nu_i - (1/\kappa_i+1)\sum_{j=1}^{n_i}\log\left\{1+\kappa_i\left(1+\kappa_i\right)w_j\nu_i\right\}\right].$$

Details of the estimation procedure are given in Chavez-Demoulin (1999). A procedure can be implemented to estimate simultaneously the parameters κ and ν. Inference in semiparametric models for the GPD uses results for semiparametric generalized linear models or generalized additive models.

For the nonparametric component, we justify the use of simultaneous tests based on deviances for the two models by the orthogonality of the parameters κ and ν. That is, applications show that deviance test results for one model do not depend on the other parameter model form. Furthermore, for smoothing parameter selection we recommend the use of criteria such as AIC; see McQuarrie and Tsai (1998) for more details on such criteria. The behaviour of AIC curves for one parameter remains unchanged by a modification of the smooth function for the other parameter. Use of appropriate degrees of freedom in conjunction with deviances allows for the assessment of model adequacy. A possible general approach to assess uncertainty for semiparametric GPD estimates is to use bootstrap methods. A bootstrapping strategy is based on the result that, given the model is correct, the residuals

$$R_j = \hat\kappa_j^{-1}\log\left\{1+\hat\kappa_j W_j\left(1+\hat\kappa_j\right)\hat\nu_j\right\}, \qquad j=1,\ldots,N\,, \tag{5}$$

are distributed approximately as independent, unit exponential random variables. We define simulated responses by

$$W_j^* = \frac{1+\exp\left\{\hat\kappa_j\epsilon_j^*\right\}}{\hat\kappa_j\left(1+\hat\kappa_j\right)\hat\nu_j}, \qquad j=1,\ldots,N\,, \tag{6}$$

where $\epsilon_1^*, \ldots, \epsilon_N^*$ is a random sample from the residuals R_j defined in (5). This leads to basic bootstrap confidence intervals for $\hat{\kappa}$ and $\hat{\nu}$.

To summarize, the following points give general guidelines to take into account before proceeding:

a) Decide upon a model form for each parameter λ, κ and ν. In practice the Poisson process for the exceedance times is not necessarily homogeneous and hence a smoothing of $\lambda = \lambda(t)$ over t is appropriate. Additionally, the parameter κ which controls the weight of the tail of the extremal distribution is generally hard to estimate and a fully parametric form $\kappa = x^T\beta$ often suffices. The scale parameter σ (and hence ν) may vary much more with external explanatory variables.

b) The choice of the smoothing parameters depends on the aim of the analysis. If the aim is purely exploratory, the smoothers can be modified to suit the situation by changing the degrees of freedom. If an automatic procedure for selecting the smoothing parameters is required, we recommend the use of criteria such as AIC rather than cross–validation, which becomes computationally costly when the size of the data increases.

c) Informal inference based on deviance tests is also useful to assess models and their differences separately for each parameter. Uncertainty about the parameter estimates is assessed by constructing confidence intervals. For the nonparametric component, plots of pointwise confidence bands around the fitted curve yield useful visual tools. Procedures based on residuals are also useful to assess the goodness–of–fit of the model. A possible graphical diagnostic is based on the result that the residuals (5) are distributed approximately as independent unit exponential variables when the model is correct.

3.1 Applications

In the previous sections, we laid the foundation of an approach towards the analysis of extremes in a non–stationary environment, based on the exceedances of a high threshold. Nowadays, the threshold approach dominates most applications, being intuitively more appealing and flexible than the traditional GEV approach. We consider for illustrative purposes two examples; a first based on simulated data, and a second analysing some real data. These examples serve more the purpose of showing how the methodology can be turned into a real analysis, rather than giving an in-depth discussion of a particular application.

3.1.1 Simulated data

As already stated earlier in the paper, we concentrate on the exploratory power of the extreme value methodology presented. Therefore, we want to graphically present changing quantitative behaviour in extreme observations above some given threshold. A typical data set could be operational risk losses of k types, say, each recorded as an excess above some threshold value u_i, $i = 1, \ldots, k$. Another example could be credit losses over a given time period for k different business types or credit classes or indeed losses above given retention limits (thresholds) for k different lines of business in an excess–of–loss reinsurance example. For each of these cases, a risk measure may be given. We will concentrate on some high loss quantile also referred to earlier as the $1/p$–year return level in an insurance context or Value–at–Risk within finance. The latter was denoted earlier by y_{1-p}, see (3) in Section 2. So concretely, we want to model y_{1-p} as a function of time: is y_{1-p} constant or changing in time, and if the latter is the case, how does y_{1-p} change with time? Before we proceed, we would like to remark that other risk measures, like the conditional mean excess loss (also referred to as Conditional VaR) could have been taken.

As an example, Figure 2 represents the data (the losses) across $k = 4$ classes over a period of 50 years, say. After an extreme value analysis, we obtain the estimated $y_{1-0.05}$ quantile as a function of time for each of these classes. Below we explain the details underlying the simulated example.

The points in Figure 2 show four simulated datasets of excesses against time. To simulate the data, we proceeded as follows; for each time t_i, $i = 1, \ldots, 50$ we first simulated the number of exceedances from an inhomogeneous Poisson process with nonparametric intensity of the form $\lambda(t) = \exp\{f(t)\}$. The points in Figure 3 are the simulated numbers of exceedances from t_1 to t_{50} for four datasets and from a Poisson process with intensity shown by the line on each panel.

Given these numbers of exceedances, their sizes (points in Figure 2) are simulated from a GPD distribution with parameters $\kappa(t)$ and $\nu(t)$. As shown in Figure 4, we chose a fully parametric form for κ with $\kappa(t) = x^T\beta$, where x is any explanatory variable which differs for each dataset. Hence this parameterization assigns a possibly different value of κ to each of the data sets; in our case, the κ–values range from -0.4 to -0.55 corresponding to Fréchet–type tails for the underlying losses. In order to

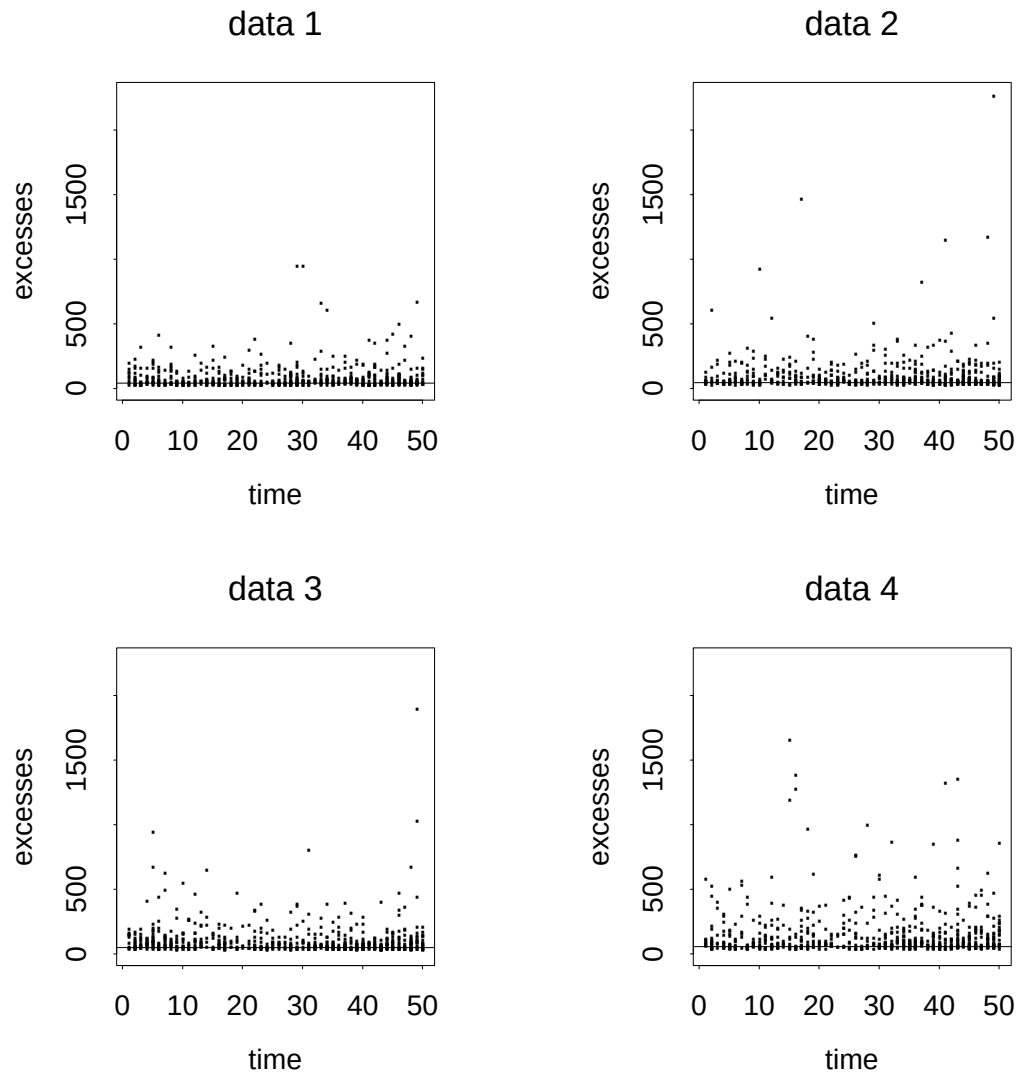

Figure 2: *Simulated data. Simulated datasets of excesses against time.*

interpret these values, note that with $\alpha = -1/\kappa$, the underlying ground up losses have a Pareto tail $x^{-\alpha}L(x)$, for some slowly varying function L and hence $\alpha + \epsilon$ divergent moments for all $\epsilon > 0$; see Embrechts, Klüppelberg and Mikosch (1997, Section 3.3 and 3.4). Note that in the latter reference, $-\kappa$ is denoted by ξ. Hence the parameter values chosen for κ correspond to heavy–tailed losses.

We chose a semiparametric form for ν, which is $\nu(t) = \exp\left\{x^T\gamma + s(t)\right\}$. The parameter ν, which depends on κ and on the scale parameter σ, is difficult to estimate and it is realistic to consider that in general it varies with time in a nonparametric way. The four curves of ν from which we simulated our datasets are shown in Figure 5.

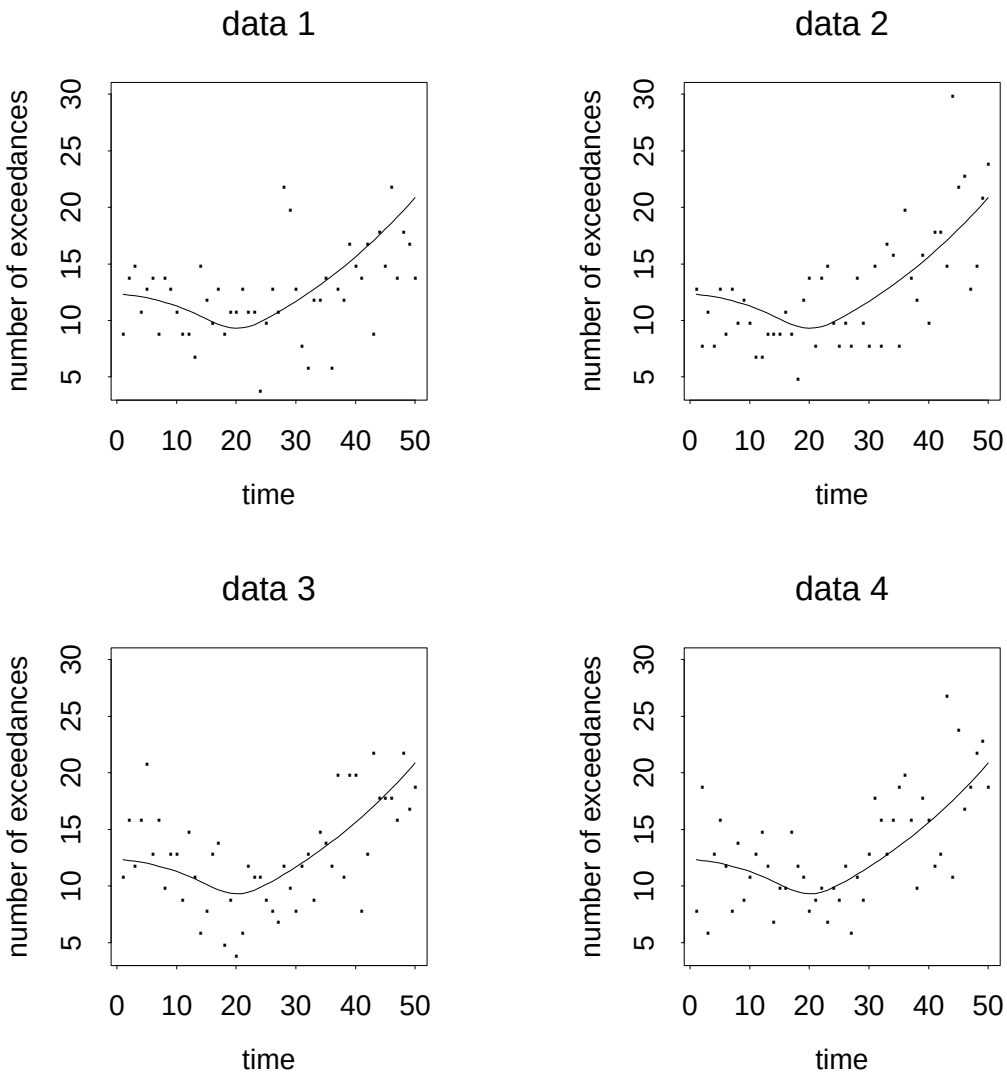

Figure 3: *Simulated data. The points are the numbers of exceedances for four datasets simulated from the Poisson process of intensity shown by the straight line in each panel.*

In this example the nonparametric part of the model for ν is the same for the four datasets.

Since we simulated the sizes of the exceedances and not the entire original data from which excesses are extracted, the choice of the threshold can be arbitrarily fixed. We chose a parametric form $u = x^T a$ with values (from the top left panel to the bottom right of Figure 2), $u_1 = 40$, $u_2 = 45$, $u_3 = 50$, $u_4 = 55$.

Following the methodology we outlined in the previous sections, we fit different models for λ, κ and ν and compare them using tests based on the likelihood ratio statistics.

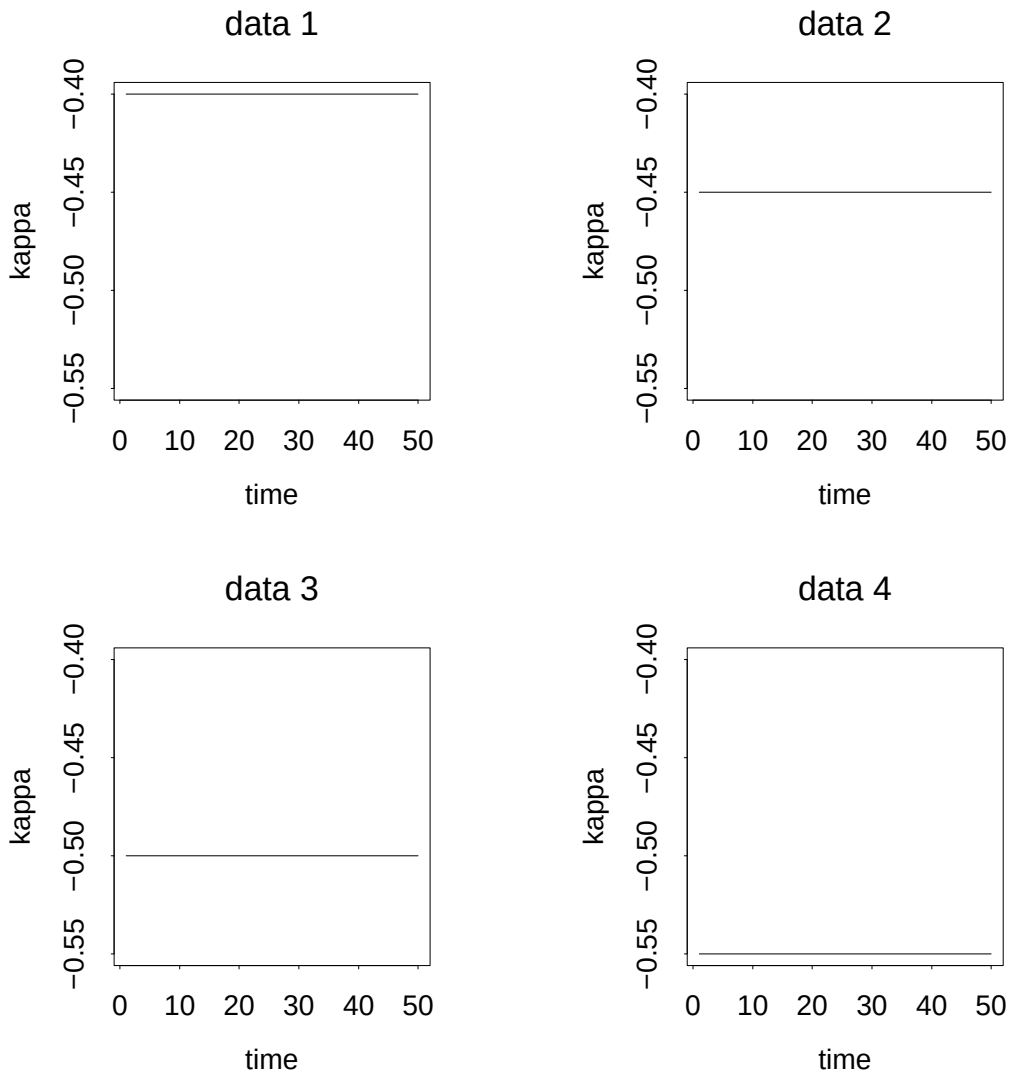

Figure 4: *Simulated data. The line of each panel represents $\kappa(t)$ of the GPD distribution.*

As mentioned in Section 2 and supposing that time t is measured in years, a quantity of interest is the $1/p$–year return level (3); that is the level crossed on average once in $1/p$ years. Figure 6 shows in dotted lines the estimated 20–year return level for each dataset. The solid lines of the figure are the 20–year return levels, based on the "true" values of λ, ν and κ. We refrain from discussing the accuracy of the method here as this example is only based on one simulation; likewise no confidence intervals will be given (for details on the accuracy of the method see Chavez–Demoulin (1999)). The curves show a slow decreasing trend from t_1 to t_{20} and an increasing trend from t_{20} for each dataset with more variation for the highest thresholds.

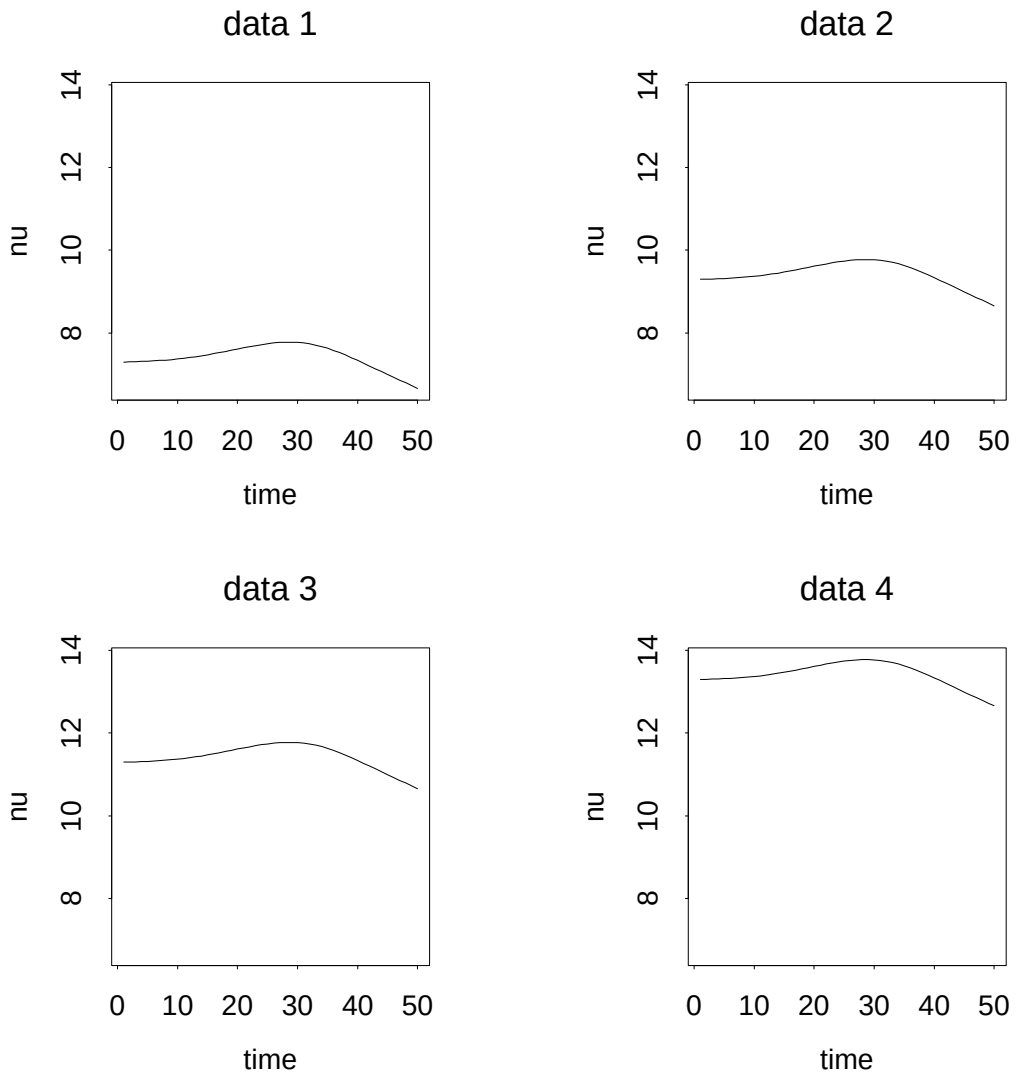

Figure 5: *Simulated data. The line of each panel represents $\nu(t)$ of the GPD distribution.*

Coming back to the discussion at the beginning of this section (keeping in mind that this is a simulated example), our analysis (Figure 6) shows that across the four rating classes there is a clear difference in the 95% Value–at–Risk (or quantile) curves. Concentrating for instance on class (data) 4, the 95% true Value–at–Risk decreases from about 2250 to a low of 1800 after 20 years, after that a marked increase over the next 30 years to a level around 3000 occurs. The estimated curve seems to catch some of the main features of this nonstationarity; of course, at this point, a more detailed analysis involving confidence intervals would be called for.

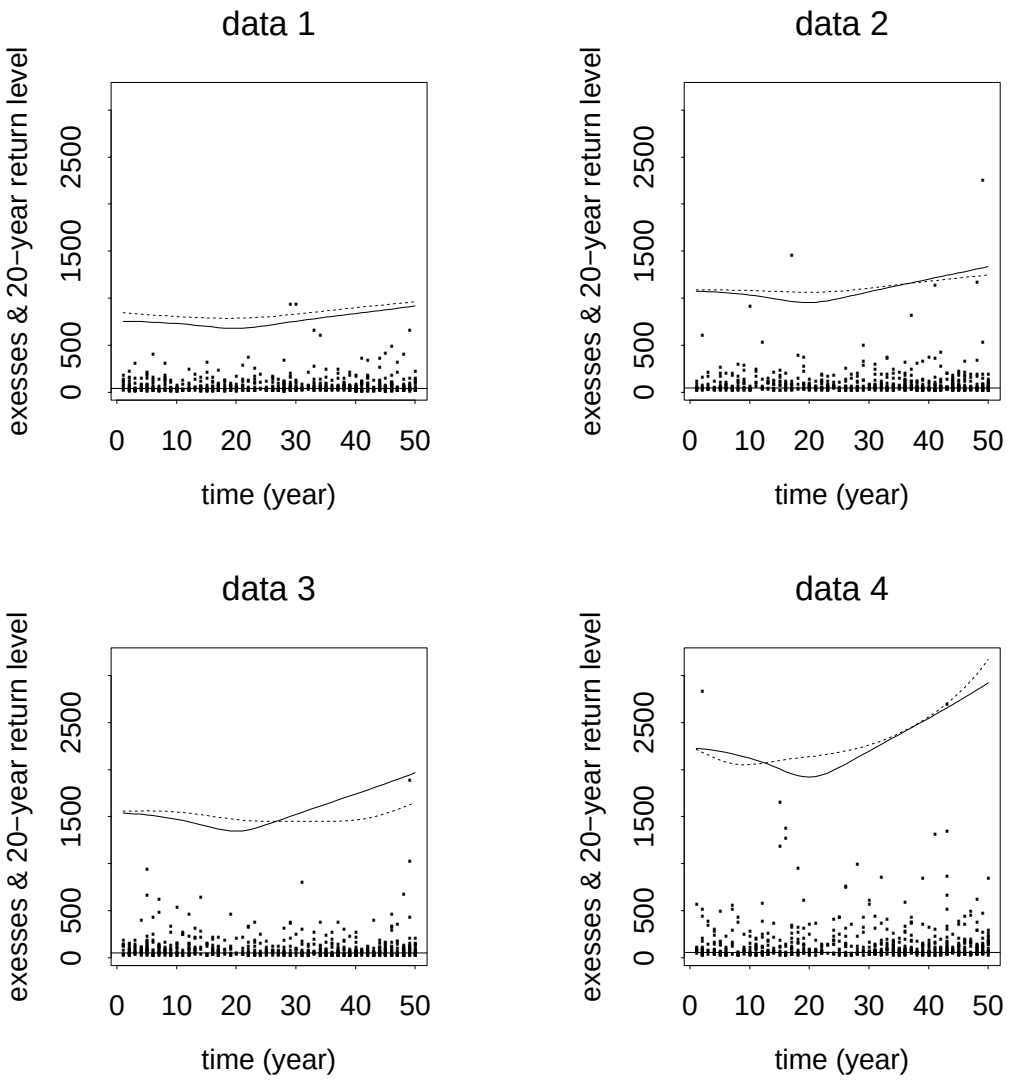

Figure 6: *Simulated data. The points are the simulated excesses against time (say year). The straight lines are the 20–year return levels and the dotted lines the estimated 20–year return levels.*

3.1.2 Coca Cola data

To briefly illustrate the methodology discussed on real data, we consider here the daily volumes of Coca Cola quotes, corrected for splits, from the 1st of January 1980 to the 31st of December 2000. The data come from the Web site `http://chart.yahoo.com/` with ticker symbol **ko**. The number of data per year is 252 and 253 for the leap year. The points of Figure 7 are the volumes $\times 10^{-5}$ from the 1st of January 1980 to the 31st of December 2000. For each year, we chose a threshold value such that about

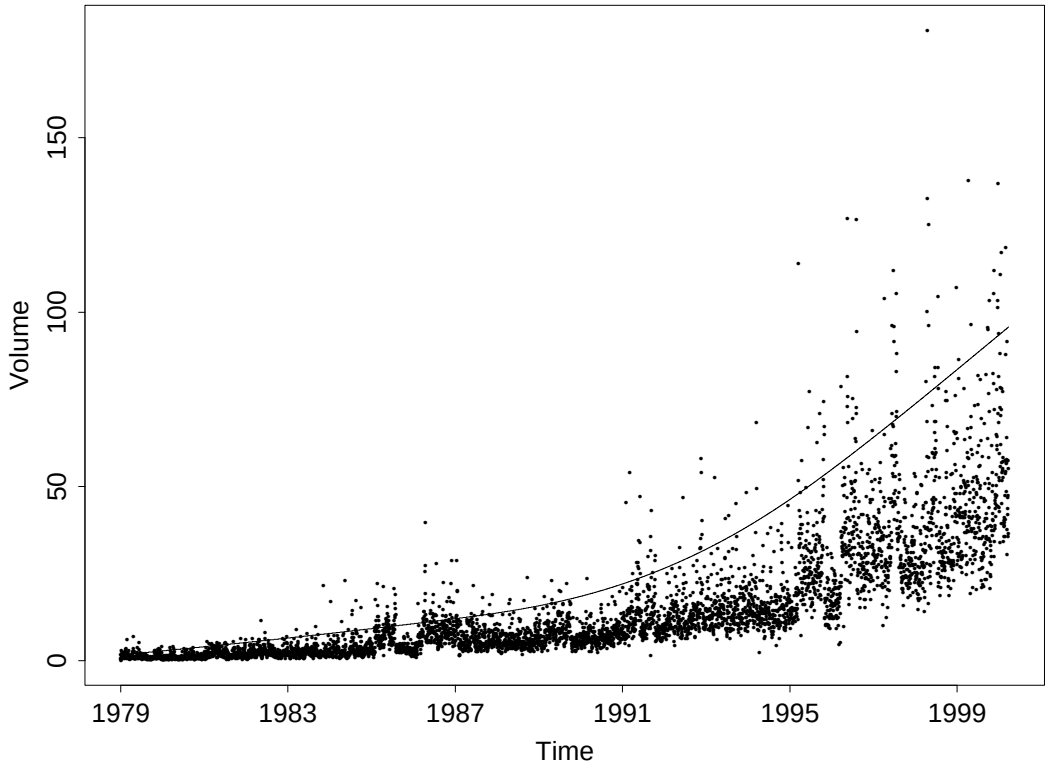

Figure 7: *Coca Cola data. Volumes* $\times 10^{-5}$ *from the 1st of January 1980 to the 31 of December 2000 (points) and smoothing threshold (line).*

10% of the data are excesses. The nonparametric threshold depicted by the solid line in Figure 7 is obtained by smoothing these yearly threshold values. The points of the left panel of Figure 8 show the jittered sizes of the exceedances over the threshold from 1980–2000 and could be compared and contrasted with any of the panels in Figure 2. The excesses against the days of year (points in the right panel of Figure 8) show that they seem to be greater at the beginning of the year but also that they are less important around June.

To account for both short and long range time dependence, we fit different non-parametric models for κ, ν and λ of the form

$$\kappa = r(d) + g(t)\,, \qquad \nu = \exp\{j(d) + s(t)\}\,, \qquad \log\lambda = z(d) + f(t)\,,$$

where d denotes the day and t the year. The points of Figure 9 are the yearly numbers of exceedances over the threshold. The line shows the fitted intensity Poisson process with

$$\log\hat{\lambda} = \hat{z}(d,2) + \hat{f}(t,2)\,,$$

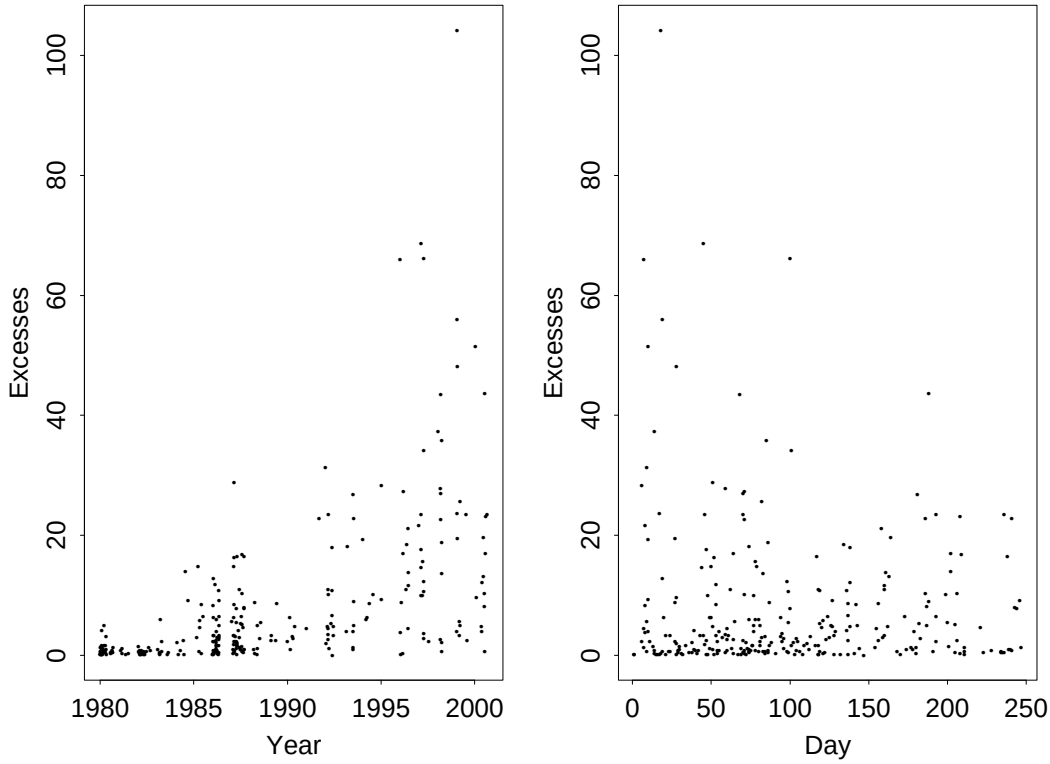

Figure 8: *Coca Cola data. Jittered excesses over threshold from 1980–2000 (left panel) and excesses against the days of year (right panel).*

where $z(d, Df)$ is the notation for the fitted spline with Df degrees of freedom. The estimated curve shows a decreasing trend from 1980 to the beginning of the 90s and a constant trend from 1994.

The models we selected for κ and ν are

$$\hat{\kappa} = \hat{r}(d, 4) + \hat{g}(t, 2)\,, \qquad \hat{\nu} = \exp\left\{\hat{j}(d, 4) + \hat{s}(t, 2)\right\}\,,$$

that is two fully nonparametric models on day and year. Figure 10 shows the volumes (points) against time (days from 1st of January 1980 to the 31st of December 2000) and the smoothed 5–year return level (line). The points of Figure 11 are the yearly volumes and the line is the 5–year return level. It shows an increasing trend amongst the years especially from the beginning of the 90s.

It seems that for the Coca Cola volumes, though the number of extremes seems to be constant since 1994, their values are highly increasing from the 90s. When dealing with such increasing volumes, it is useful to consider the logarithm of the data, since the return level is invariant under that transformation. The points of Figure 12 are

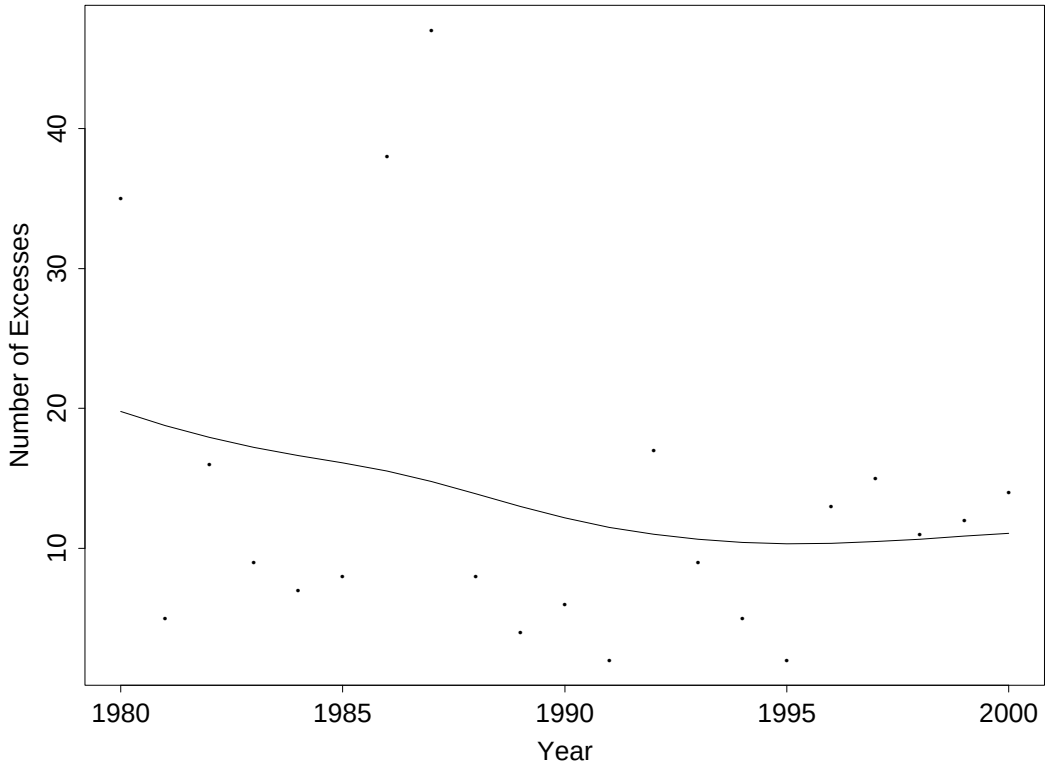

Figure 9: *Coca Cola data. Yearly numbers of exceedances (points) and estimated intensity Poisson process (line).*

the logarithm of the volumes. The figure better highlights the strange behaviour of the data around 1986–87 and again around 1996–97. Figure 13 shows the resulting 5–year return level from a smoothing modelization. If we compare with Figure 11, we see that the points which exceed the return level curve arise almost during the same years.

4 Comment

With the increasing interest on financial rare events like extreme losses or non–standard price fluctuations, there is a pressing need for flexible modelling of extremes. The smoothing extreme value method fitted by penalized loglikelihood provides a convenient, rapid and flexible data exploration technique. It also highlights features of the underlying distribution as the covariates change, and provides an objective tool to determine their relative importance.

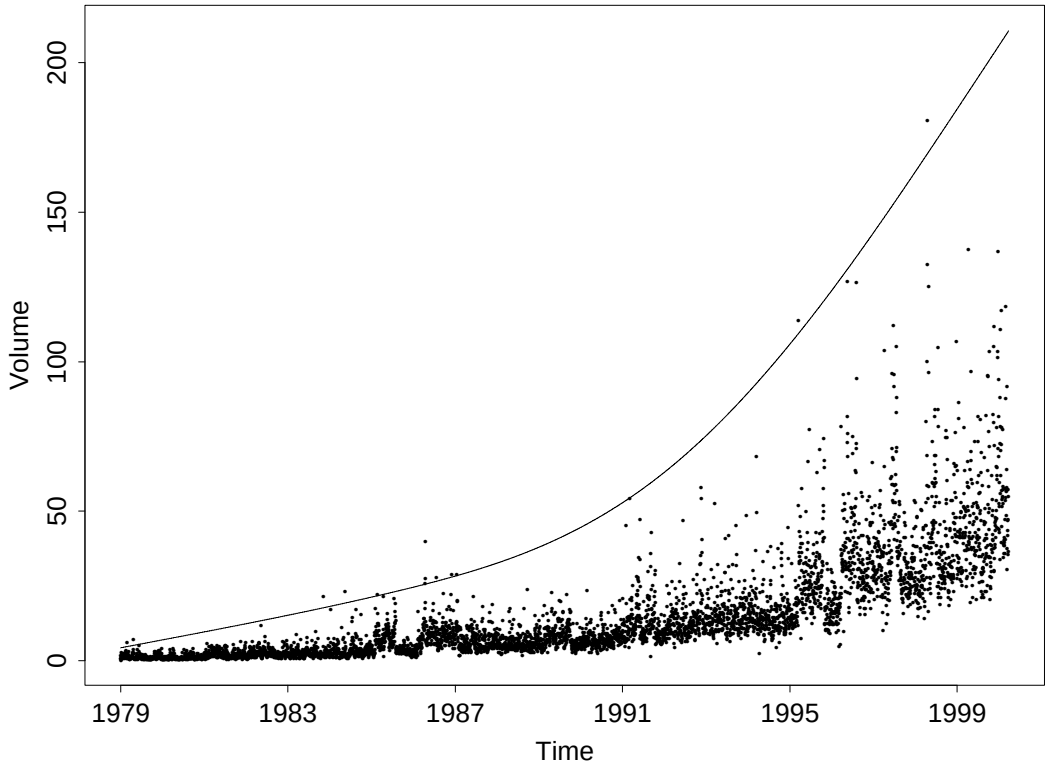

Figure 10: *Coca Cola data. Volumes* $\times 10^{-5}$ *(points) and estimated 5–year return level (line).*

Acknowlegement

This work was partly supported by the Swiss National Science Foundation. We thank Anthony Davison for helpful discussions and comments.

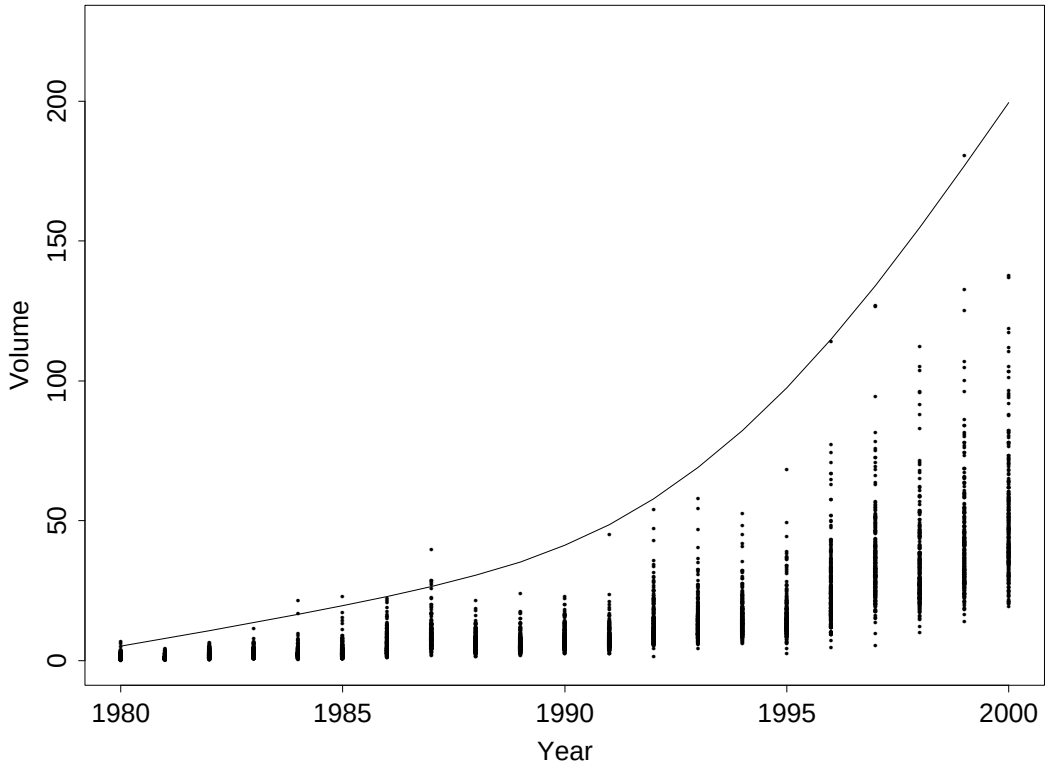

Figure 11: *Coca Cola data. Yearly volumes* $\times 10^{-5}$ *(points) and estimated 5–year return level (line).*

References

Chavez–Demoulin, V. (1999) *Two Problems in Environmental Statistics: Capture-Recapture Analysis and Smooth Extremal models.* Ph.D. thesis. Department of Mathematics, Swiss Federal Institute of Technology, Lausanne.

Chavez–Demoulin V. and Davison A.C. (2001) Generalized Additive Models for Sample Extremes. *Submitted to Applied Statistics.*

Cox, D. R. and Reid, N. (1987) Parameter orthogonality and approximate conditional inference (with Discussion), *Journal of the Royal Statistical Society, B,* **49**, 1–39.

Davison, A.C. (1984a) *A statistical model for contamination due to long–range atmospheric transport of radionuclides.* Ph.D. Thesis. Department of Mathematics, Imperial College of Science and Technology, London.

Davison, A.C. (1984b) Modelling excesses over high thresholds, with an applica-

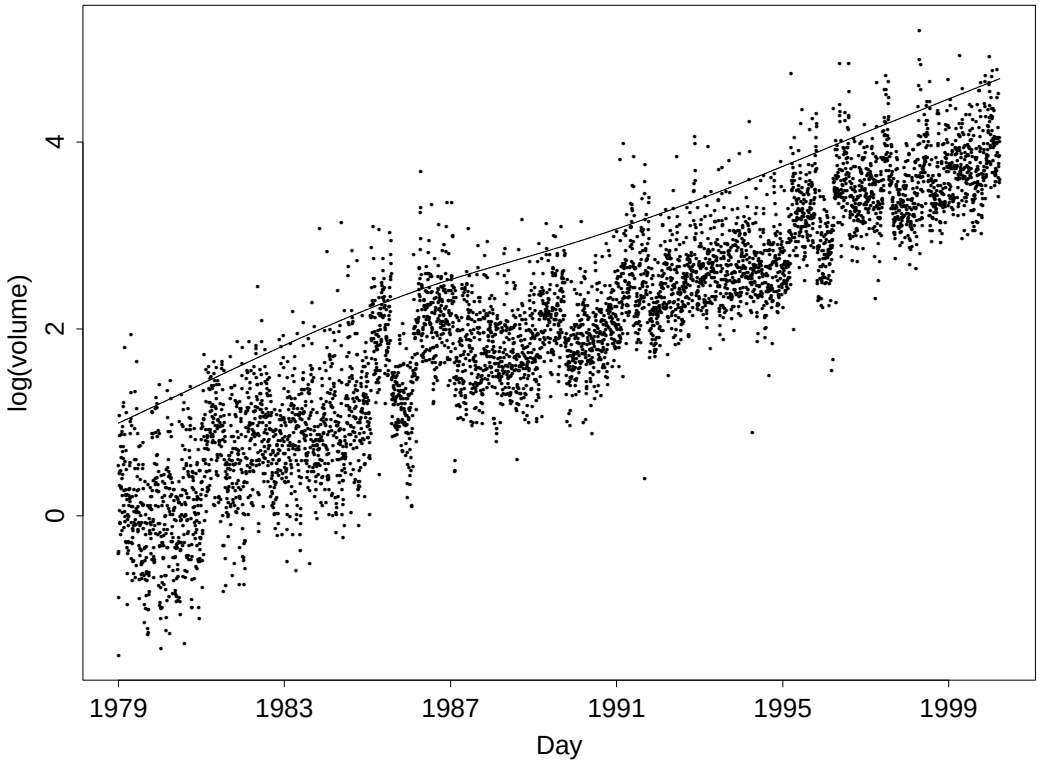

Figure 12: *Coca Cola data. Logarithm of the volumes* $\times 10^{-5}$ *(points) and smoothing threshold (line).*

tion. *Statistical Extremes and Applications.* Dordrecht: Reidel. Ed. J. Tiago de Oliveira, 461–482.

Davison, A. C. and Smith, R.L. (1990) Models for exceedances over high thresholds (with Discussion), *Journal of the Royal Statistical Society, B*, **52**, 393–442.

Embrechts, P. (Ed.) (2000) *Extremes and Integrated Risk Management.* London: Risk Waters Group.

Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997) *Modelling Extremal Events for Insurance and Finance.* Berlin: Springer.

Embrechts, P., Resnick, S.I. and Samorodnitsky, G. (1999) Extreme value theory as a risk management tool, *North American Actuarial J.* **3**, 30–41.

Green, P.J. and Silverman, B.W. (1994) *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach.* London: Chapman and Hall.

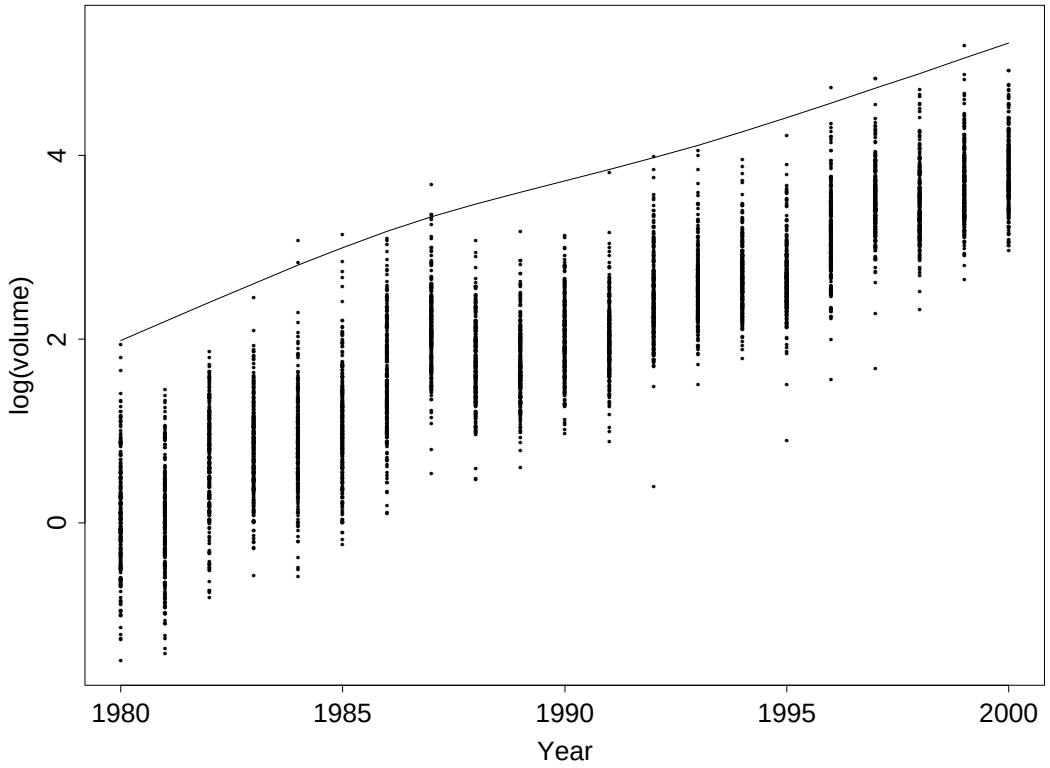

Figure 13: *Coca Cola data. Yearly logarithm of the volumes* $\times 10^{-5}$ *(points) and estimated 5–year return level (line).*

Green, P.J. and Yandell, B. (1985) Semi–parametric generalized linear models. *Proceedings 2nd International GLIM Conference.* Lancaster, Lecture Notes in Statistics, **32**, 44–55. New York: Springer.

Leadbetter, M.R. (1991) On a basis for 'peaks over threshold' modeling. *Statistics & Probability Letters* **12**, 357–362.

Leadbetter, M.R., Lindgren, G. and Rootzén, H. (1983) *Extremes and Related Properties of Random Sequences and Series.* New York: Springer.

Matthys, G. and Beirlant, J. (2000) Adaptive threshold selection in tail index estimation. In: *Extremes and Integrated Risk Management* (Ed. P. Embrechts). London: Risk Waters Group, 37–49.

McQuarrie, A.D.R. and Tsai, C.–L. (1998) *Regression and Time Series Model Selection.* Singapore: World Scientific.

Reiss, R.D. and Thomas, M. (2001) *Statistical Analysis of Extreme Values.* Basel: Birkhäuser.

Robinson, M.E. and Tawn, J.A. (2000) Extremal analysis of processes sampled at different frequencies, *Journal of the Royal Statistical Society, B*, **62**, 117–136.

Rootzén, H. and Tajvidi, N. (1997) Extreme value statistics and wind storm losses: a case study. *Scandinavian Actuarial Journal*, **??**, 70–94.

Smith, R.L. (1985) Maximum likelihood estimation in a class of non–regular cases, *Biometrika,* **72**, 67–90.

Smith, R.L. (1987) Estimating tails of probability distributions, *Annals of Statistics*, **15**, 1174–1207.

Smith, R.L. and Shively, T.S. (1995) A point process approach to modeling trends in tropospheric ozone, *Atmospheric Environment*, **29**, 3489–3499.

Yang, G.L. (1978) Estimation of a biometric function, *Annals of Statistics*, **6**, 112–116.