© *The Journal of Risk and Insurance*, 2004, Vol. 71, No. 2, 183-199

Smooth Extremal Models in Finance and Insurance

V. Chavez-Demoulin
P. Embrechts

Abstract

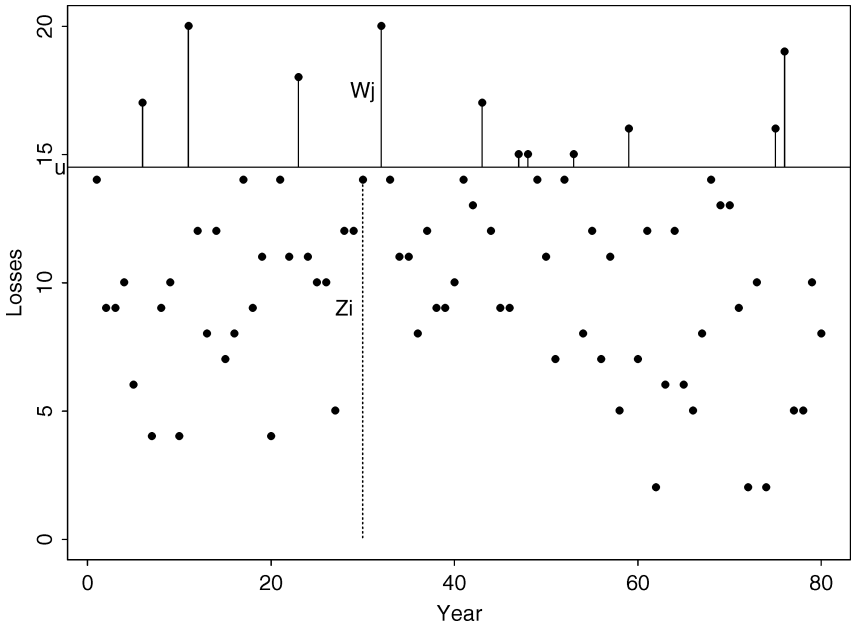
This article describes smooth nonstationary generalized additive modeling for sample extremes, in which spline smoothers are incorporated into models for exceedances over high thresholds. We summarize the smoothing methodology as a new tool for practical extreme value exploration in finance and insurance.

Introduction

Extreme value theory (EVT) has developed very rapidly over the past two decades both methodologically and with respect to applications. Whereas (nonlife) actuaries have, at least implicitly, used EVT techniques for a long time, mainly through the emergence of quantitative risk management, EVT has entered the finance stage more recently as a useful toolkit for describing nonstandard (more precisely nonnormal) 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 nonnormality, 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 at http://www.math.ethz.ch/finance and http://www.risklab.ch. In Embrechts (2000), various papers highlight the current state-of-the-art on EVT modeling in Integrated Risk Management (also see Reiss and Thomas, 2001, and Coles, 2001, for very readable discussions).

Chavez-Demoulin and Embrechts are from the Department of Mathematics, ETH-Zentrum, 8092 Zürich, Switzerland. This work was partly supported by the NCCR FINRISK Swiss research program. We thank Anthony Davison for helpful discussions and comments. We also thank two anonymous referees for several useful comments on the first version of this article.

FIGURE 1
The Point Process of Exceedances (POT)



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 and identically distributed (i.i.d.) (or weakly dependent) data (see, for instance, Embrechts, Klüppelberg, and Mikosch, 1997, p. 121). For a complete discussion concerning possible dependencies on the data we refer to Coles (2001, Chapter 5). Another model is based on a so-called point process characterization. The resulting Peaks Over Threshold (POT) method appears more flexible and it 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 (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. For the latter, see Cruz (2002, Chapter 4). 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 "The Description of the Methodology" 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 so forth) and their occurrence no doubt shows a nonconstant (often stochastic) intensity, possibly depending on such factors as business cycles, transaction intensity, 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 behavior.

In this article, we discuss some of the more recent EVT methodology, which may be useful in handling the presence of such covariates and the resulting modeling 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. Nonparametric trends have been used recently by Hall and Tajvidi (2000) in other meteorological applications. 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 separately consider 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 $\sigma(t) = \exp\{\alpha + \beta t\}$, where t denotes time in years. They motivate this choice by the fact that previous experience of similar situations indicates that a constant κ should be reasonable and that the model for the scale parameter corresponds to a constant growth in cost, by $100 \times (e^{\beta} - 1)$ percent 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 article, we therefore propose to summarize and illustrate the methodology proposed by Chavez-Demoulin (1999) and Chavez-Demoulin and Davison (2004) to an insurance and financial context. Their approach combines the point process for exceedances with smoothing methods to give a flexible exploratory framework to model changes in large values for insurance or financial data.

A possible model might consist of an inhomogeneous Poisson process for the number of exceedances, with intensity of the form $\lambda(t) = \exp\{x^T\alpha + f(t)\}$, combined with the GPD 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 "The Description of the Methodology" 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. These can be summarized parametrically introducing in the model factors with different levels. In our context, they can be 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. Seasonal effects can be explored by the nonparametric part of the model by increasing the degrees of freedom (for more details, see Chavez-Demoulin, 1999).

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. We will use $\nu(t) = \exp\{x^T\eta + s(t)\}$ below. 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 that 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 data set 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 article is organized as follows. In "The Description of the Methodology," we review some of the stochastic techniques underlying the threshold (POT) method and summarize the smoothing methodology as a new tool for practical extreme value exploration in finance and insurance. In "Comment," the methods will be illustrated on some idealized examples from insurance and finance.

THE DESCRIPTION OF THE METHODOLOGY

The approach based on the threshold method considers a characterization of all observations that are extreme in the sense of having exceeded a high threshold u. Consider a sequence of i.i.d. 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 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 Equation (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 nonregular 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 GPDs 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 and also see Matthys and Beirlant (2000) which 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

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

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 Equation (3) for σ. This method is also referred to as the profile likelihood approach and has been programmed into the freeware EVIS available at http://www.math.ethz.ch/~mcneil.

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. Similarly, 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 an earlier work by Leadbetter, Lindgren, and Rootzén (1983),

and is a topic of much current research. 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. 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).

In the context of insurance or finance, there is a need for exploratory data analysis to gain an understanding of the structure of the data. The goal of applying a smoothing method 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 (2004). To summarize, the following points give general guidelines to take into account before proceeding:

1. 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.
2. 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 (Akaike, 1974) rather than cross-validation, which becomes computationally costly when the size of the data increases.
3. 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. In particular, bootstrapping strategies are presented in Chavez-Demoulin (1999). 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.

Applications

In the previous sections, we briefly laid the foundation of an approach toward the analysis of extremes in a nonstationary 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 two examples for illustrative purposes: (1) based on simulated data and (2) analyzing some real data. These examples serve 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.

Simulated Data. As already stated earlier in the article, we concentrate on the exploratory power of the extreme value methodology presented. Therefore, we want

to graphically present changing quantitative behavior 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 VaR within finance. The latter was denoted earlier by y_{1-p}, see (3) in "The Description of the Methodology." 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. See Artzner et al. (1999) for the pros and cons of these and other financial risk measures.

As an example, Figure 2 represents the data (the losses) across $k = 4$ classes over a period of say, 50 years. After an extreme value analysis, we obtain the estimated

FIGURE 2
Simulated Data. Simulated Data Sets of Excesses Against Time

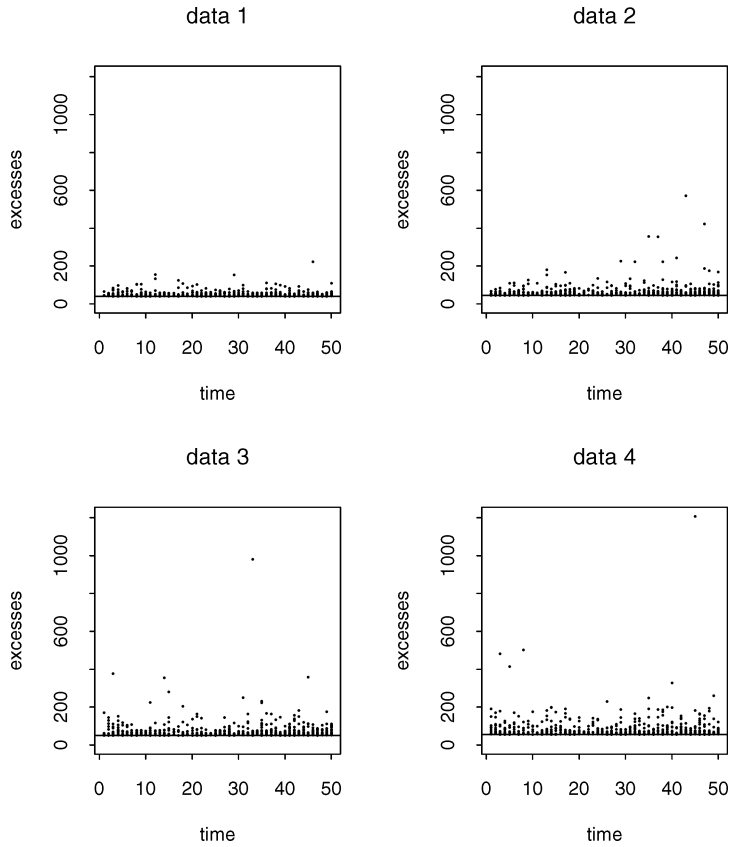
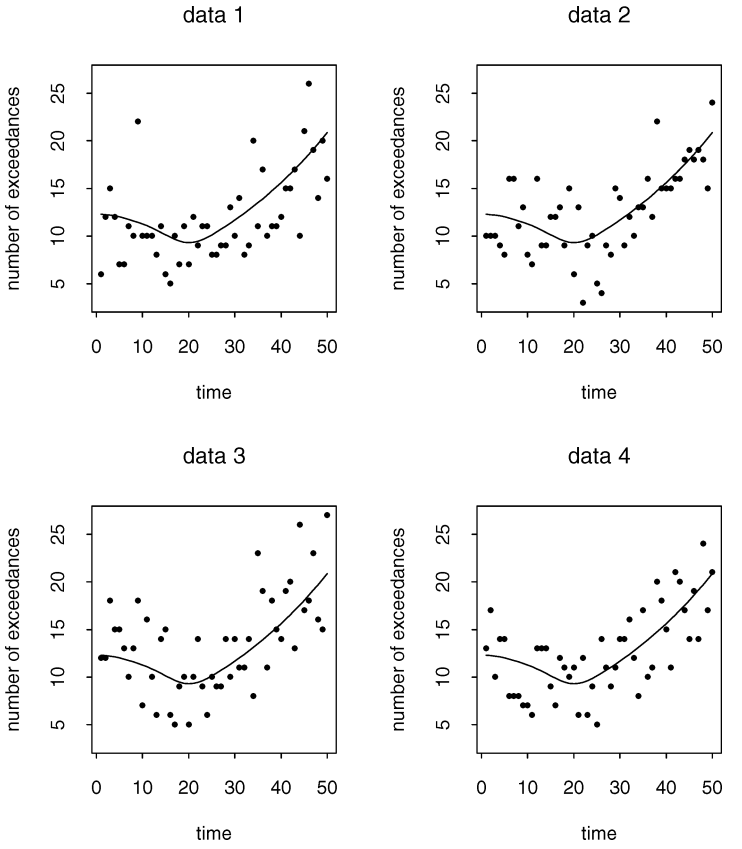


FIGURE 3
Simulated Data. The Points Are the Numbers of Exceedances for Four Data Sets Simulated From the Poisson Process of Intensity Shown by the Solid Line in Each Panel

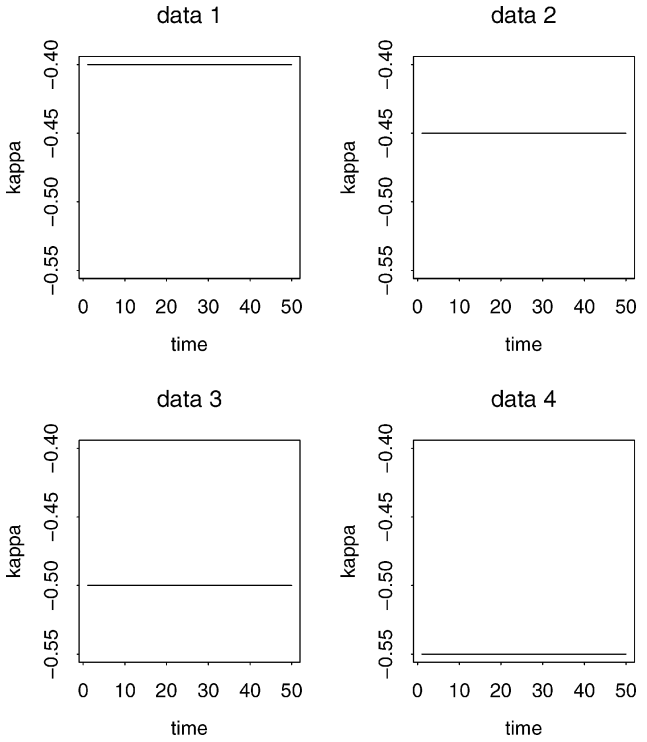


$y_{1-0.05}$ quantile as a function of time for each of these classes. We explain the details underlying the simulated example below.

The points in Figure 2 show four simulated data sets 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 data sets 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 that differs for each data set. 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 (i.e., power-like) tails for the underlying losses. In order to interpret these values, note that with $\alpha = -1/\kappa$, the underlying ground

FIGURE 4
Simulated Data. The Line of Each Panel Represents $\kappa(t)$ of the GPD Distribution



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, Sections 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.

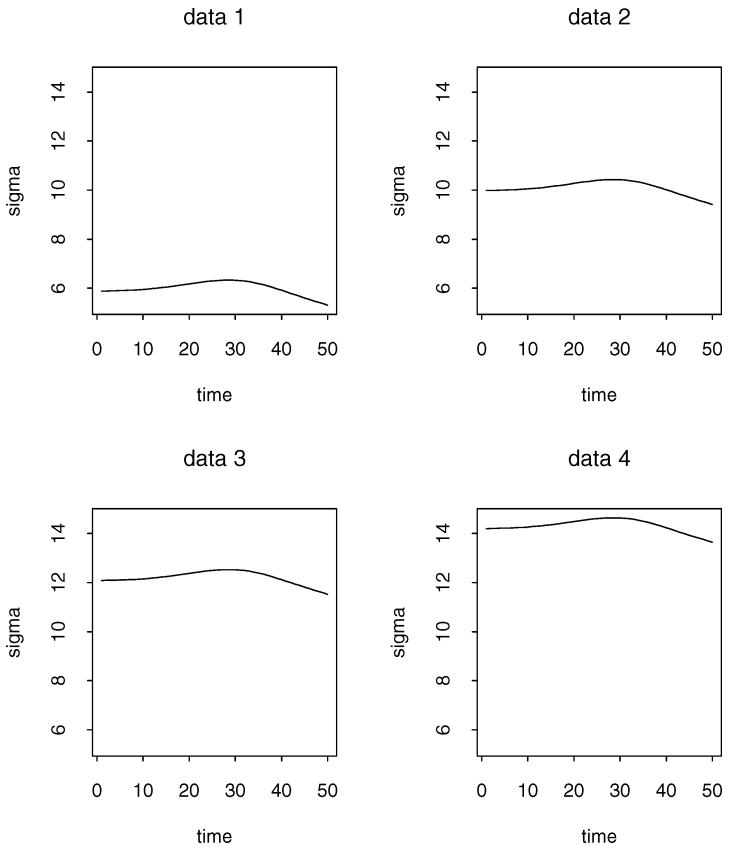
We chose a semiparametric form for ν, which is $\nu(t) = \exp\{x^T\gamma + s(t)\}$. The parameter ν, which depends on κ and on the scale parameter σ, is difficult to estimate and it is realistic to consider that in general it varies with time in a nonparametric way. The four curves of σ from which we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

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

As mentioned in "The Description of the Methodology" and supposing that time t is measured in years, a quantity of interest is the $1/p$-year return level (3), that is, the

FIGURE 5
Simulated Data. The Line of Each Panel Represents $\sigma(t)$ of the GPD Distribution

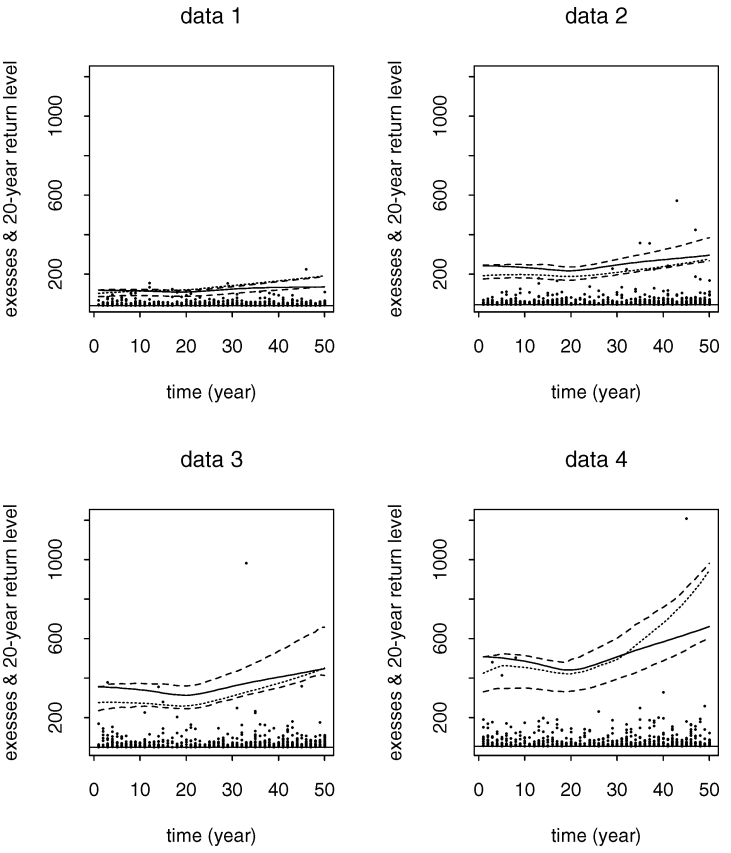


level crossed on average once in $1/p$ years. Figure 6 shows in solid lines the 20-year return level, based on the "true" values of λ, ν, and κ for each data set. The dotted lines are the estimated 20-year return levels corresponding to the simulated data. The dashed lines are the 95 percent pointwise bootstrap confidence intervals based on 99 simulations (for more details on the accuracy of the method and construction of confidence intervals, see Chavez-Demoulin, 1999). The curves show a low decreasing trend from t_1 to t_{20} and an increasing trend from t_{20} for each data set with more variation for the highest thresholds.

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 percent VaR (or quantile) curves. The estimated curves seem to catch some of the main features of this nonstationarity.

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 January 1, 1980 to December 31, 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.

FIGURE 6
Simulated Data. The Points Are the Simulated Excesses Against Time (Say, Year). The Solid Lines Are the 20-Year Return Levels and the Dotted Lines Are the Estimated 20-Year Return Levels. The Dashed Lines Are the 95 Percent Pointwise Bootstrap Confidence Intervals Based on 99 Simulations



The points in Figure 7 are the volumes $\times 10^{-5}$ from January 1, 1980 to December 31, 2000. For each year, we chose a threshold value such that about 10 percent 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 to 2000 and could be compared and contrasted with any of the panels in Figure 2. The excesses against the days of the 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 nonparametric models for κ, ν, and λ of the form

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

where d denotes the day and t the year.

FIGURE 7
Coca Cola Data. Volumes $\times 10^{-5}$ From January 1, 1980 to December 31, 2000 (Points) and Smoothing Threshold (Line)

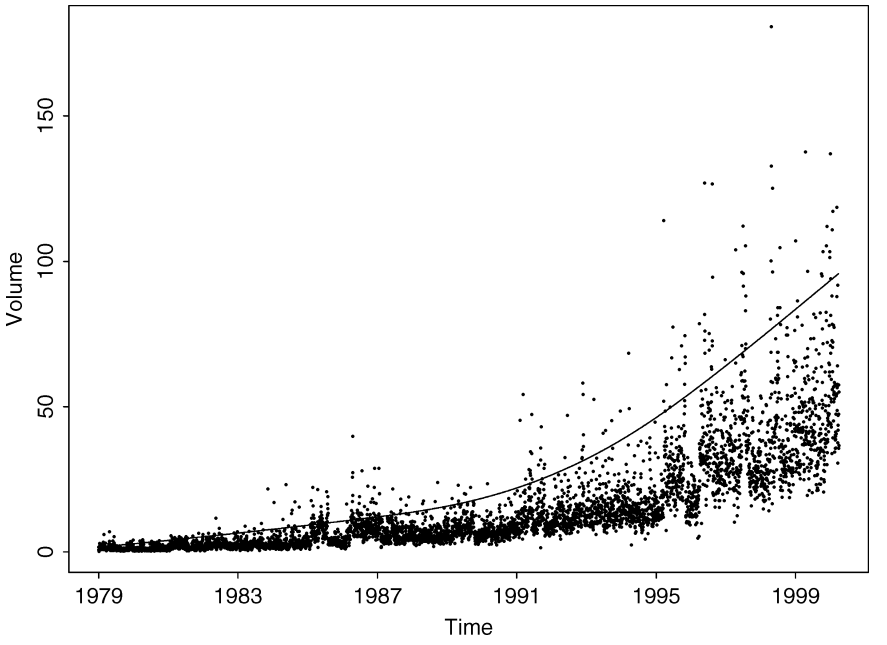


FIGURE 8
Coca Cola Data. Jittered Excesses Over Threshold From 1980 to 2000 (Left Panel) and Excesses Against the Days of Year (Right Panel)

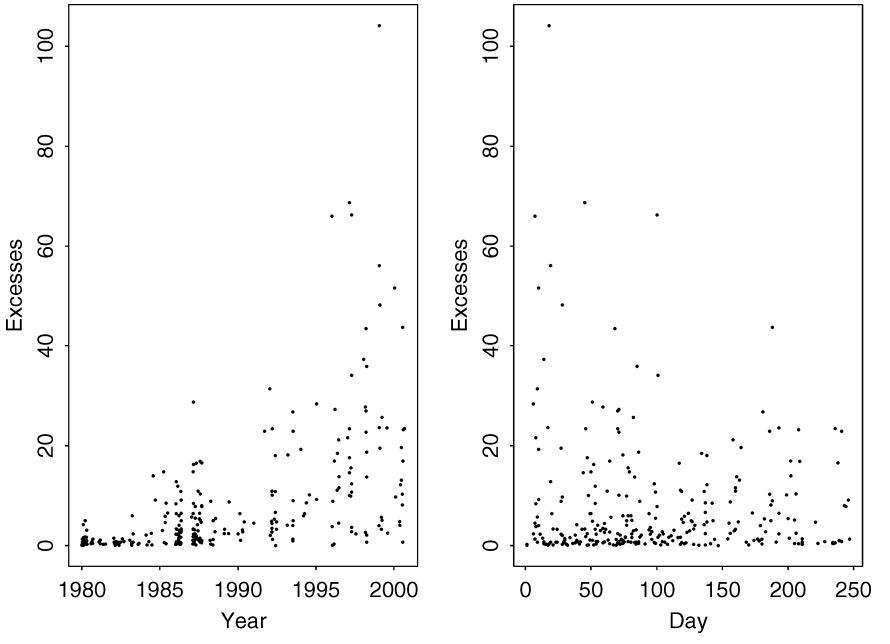
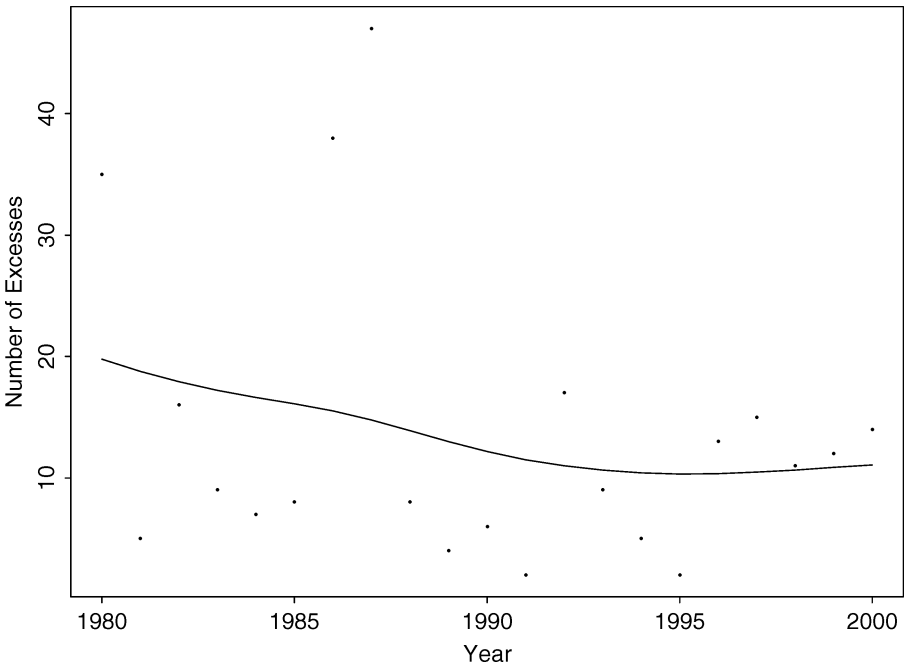


FIGURE 9
Coca Cola Data. Yearly Numbers of Exceedances (Points) and Estimated Intensity Poisson Process (Line)



The points of Figure 9 are the yearly numbers of exceedances over the threshold. The line shows the fitted intensity Poisson process with

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

where $z(d, Df)$ is the notation for the fitted spline with Df, the degrees of freedom. The estimated curve shows a decreasing trend from 1980 to the beginning of the 1990s and a constant trend from 1994.

The models we selected for κ and ν are

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

that is two fully nonparametric models on day and year. Figure 10 shows the volumes (points) against time (days from January 1, 1980 to December 31, 2000) and the smoothed 5-year return level (line). It shows an increasing trend among the years especially from the beginning of the 1990s.

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 1990s. 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 in Figure 11 are the logarithm of the volumes and the line is the resulting 5-year return level from a smoothing-model Ansatz.

FIGURE 10
Coca Cola Data. Volumes $\times 10^{-5}$ (Points) and Estimated 5-Year Return Level (Line)

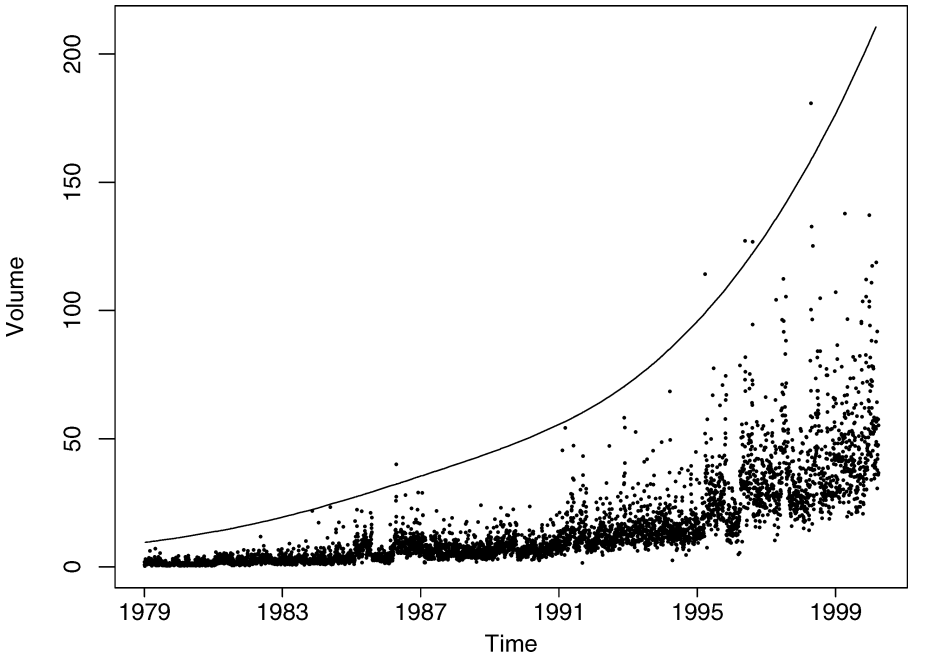


FIGURE 11
Coca Cola Data. Logarithm of the Volumes $\times 10^{-5}$ (Points) and Estimated 5-Year Return Level (Line)

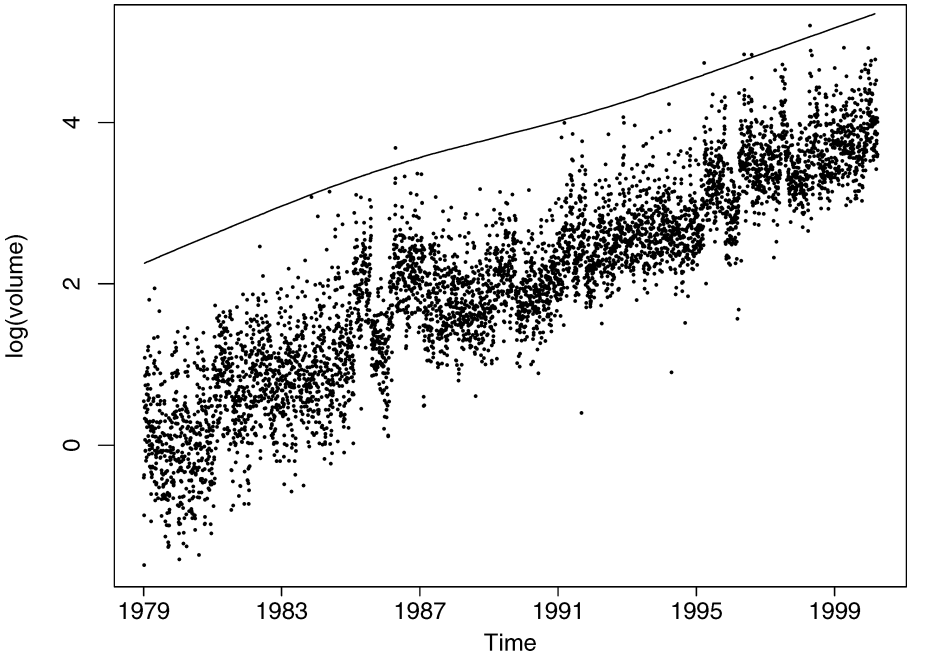
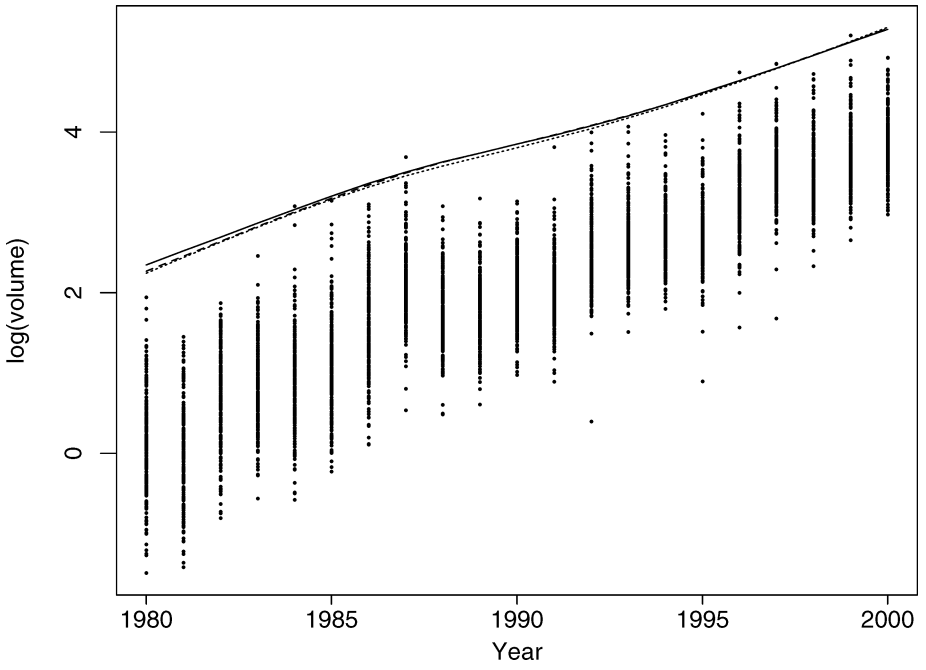


FIGURE 12
Coca Cola Data. Yearly Logarithm of the Volumes $\times 10^{-5}$ (Points) and Estimated 5-Year Return Level (Lines) From a Smoothing Model With Three Different Thresholds



A matter of concern is the need for letting the threshold depend on time in order to avoid biased results. We are grateful to a referee for suggesting a sensitivity analysis of the results with respect to the choice of the threshold. The solid line in Figure 12 shows the resulting 5-year return level (solid line) from a smoothing model for which in each year the top 10 percent of the data are used. The dotted line is the result for which in each year the top 8 percent of the data are used and the dashed line the result for which in each year the top 12 percent of the data are used. We see that the three estimated 5-year return levels are rather close together. Small variations in the value of the threshold have nearly no impact.

COMMENT

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

REFERENCES

Akaike, H., 1974, A New Look at the Statistical Model Identification, *IEEE Transactions on Automatic Control*, 19: 716-723.

Artzner, P., F. Delbaen, J. M. Eber, and D. Heath, 1999, Coherent Measures of Risk, *Mathematical Finance*, 9: 203-228.

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

Chavez-Demoulin, V., and A. C. Davison, 2004, Generalized Additive Models for Sample Extremes, *Applied Statistics*, forthcoming.

Coles, S., 2001, *An Introduction to Statistical Modeling of Extreme Values* (London: Springer).

Cox, D. R., and N. Reid, 1987, Parameter Orthogonality and Approximate Conditional Inference (with Discussion), *Journal of the Royal Statistical Society B*, 49: 1-39.

Cruz, M. G., 2002, *Modeling, Measuring and Hedging Operational Risk* (Chichester: Wiley).

Davison, A. C., 1984a, *A Statistical Model for Contamination Due to Long-Range Atmospheric Transport of Radionuclides*. Ph.D. thesis, Imperial College of Science and Technology, London.

Davison, A. C., 1984b, Modeling Excesses Over High Thresholds, With an Application, in: J. Tiago de Oliveira, ed., *Statistical Extremes and Applications* (Dordrecht: D. Reidel Publishing Company), pp. 461-482.

Davison, A. C., and R. L. Smith, 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 Books, Risk Waters Group).

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

Embrechts, P., I. S. Resnick, and G. Samorodnitsky, 1999, Extreme Value Theory as a Risk Management Tool, *North American Actuarial Journal*, 3: 30-41.

Hall, P., and N. Tajvidi, 2000, Nonparametric Analysis of Temporal Trend When Fitting Parametric Models to Extreme-Value Data, *Statistical Science*, 15(2), 153-167.

Leadbetter, M. R., 1991, On a Basis for 'Peaks Over Threshold' Modeling, *Statistical Probability Letters*, 12: 357-362.

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

Matthys, G., and J. Beirlant, 2000, Adaptive Threshold Selection in Tail Index Estimation, in: P. Embrechts, ed., *Extremes and Integrated Risk Management* (London: Risk Waters Group), pp. 37-49.

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

Robinson, M. E., and J. A. Tawn, 2000, Extremal Analysis of Processes Sampled at Different Frequencies, *Journal of the Royal Statistical Society B*, 62(1): 117-136.

Rootzén, H., and N. Tajvidi, 1997, Extreme Value Statistics and Wind Storm Losses: A Case Study, *Scandinavian Actuarial Journal*, 97: 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 T. S. Shively, 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.