A point process approach to Value-at-Risk estimation

Valerie Chavez-Demoulin
Anthony C. Davison
Alexander J. McNeil

First version: May 2002
Current version: September 2003

This research has been carried out within the NCCR FINRISK projects on “Mathematical Methods in Financial Risk Management”.
A point process approach to Value-at-Risk estimation

V. Chavez-Demoulin*, A. C. Davison† and A. J. McNeil*

*Department of Mathematics
ETH-Zentrum
8092 Zürich Switzerland
http://statwww.epfl.ch/people/chavez/
http://www.math.ethz.ch/~mcneil/

and

†Institute of Mathematics
Swiss Federal Institute of Technology
1015 Lausanne Switzerland
http://statwww.epfl.ch

September 11, 2003

Abstract

We consider the modelling of rare events in financial time series, and introduce a marked point process model for the exceedances of the time series over a high threshold that combines a self-exciting process for the exceedances with a mark (size) dependent process. This allows realistic models for rare events, in which recent events affect the current intensity for exceedances more than distant ones, but it also allows the intensity to depend on the marks of the events. Estimates of Value-at-Risk are derived for real datasets and backtested.

Keywords: Financial time series; Generalized Pareto distribution; Peaks over threshold; Self-exciting process; Statistics of extremes.
1 Introduction

The estimation of Value-at-Risk (VaR) for portfolios of traded assets over predefined holding periods has been a subject of considerable interest in recent years. According to the Capital Adequacy Directive of the Bank of International Settlements in Basle (the Basle Committee or BIS), the risk capital of a bank must be sufficient to cover losses on the bank’s trading portfolio over a 10-day holding period on 99% of occasions, and this is the value usually referred to as VaR. However for other applications different confidence levels and holding periods are considered; for purposes of internal control, for example, many financial firms use a holding period of one day and a confidence level of 95%. See Jorion (2001) or Crouhy, Galai, and Mark (2001).

In mathematical terms the VaR is a quantile of the profit-and-loss distribution of the portfolio over the holding period. In the academic literature a stylised version of the VaR estimation problem has frequently been considered whereby quantiles are estimated for univariate time series representing returns on single financial assets or sometimes portfolios thereof. In this paper we add to the methodology available for this problem.

A fundamental issue in applying VaR estimation methodology is deciding which quantile should be the focus of attention. Because financial asset and portfolio return series tend to show stochastic volatility (periods of clustered large values that contradict the simplest, purely random, model) we distinguish quantiles of the conditional return distribution of the asset over the holding period given current and past information from quantiles of the unconditional return distribution assuming a stationary model. We believe that the VaR calculation prescribed by the BIS should be interpreted as a problem of conditional quantile information, and this is the focus of this paper. It is certainly also useful to estimate unconditional quantiles, however; they provide additional information that is particularly relevant when we take a longer term view.

The existing literature on VaR estimation is vast and approaches of vary-
ing sophistication have been applied (Jorion 2001). Considering conditional quantile estimation methodology, a popular and influential method has been the method incorporated in RiskMetrics (Mina 2001) where returns over the holding period are assumed to be normally distributed and the volatility (the standard deviation of the conditional return distribution) is forecast using the exponential-weighted moving average method. In a similar spirit GARCH models with normal innovations have also been used.

In many applications of this approach (particularly to daily data) financial returns are neither unconditionally nor conditionally normally distributed, but rather their distribution is heavier-tailed; for example GARCH models with t innovations perform appreciably better than models with normal innovations. This has prompted the examination of methods from extreme value theory (EVT) (Embrechts, Klüppelberg, and Mikosch 1997, Coles 2001). The majority of papers on EVT for VaR estimation concern the estimation of unconditional quantiles; see for example Danielsson and de Vries (1997b), Danielsson and de Vries (1997c), and Danielsson and de Vries (1997a).

McNeil and Frey (2000) addressed the conditional quantile problem and suggested a method for applying EVT to the conditional return distribution by using a two-stage method combining GARCH models for forecasting volatility and EVT techniques applied to the residuals from the GARCH analysis. While this works quite well in practice it has the drawback of being a two-stage procedure, so the results of the EVT analysis will be sensitive to the fitting of a GARCH model to the entire dataset in the first stage.

A fundamental tenet of extreme value statistics is that the most extreme data should be allowed “to speak for themselves”. Our goal in this paper is to develop a model that is true to the spirit of EVT analysis and where the EVT models are applied directly to the raw data without the intermediate fitting of an econometric time series model. To this end we adapt a classical EVT model, the POT or peaks-over-thresholds model, which describes the
appearance of extremes in independent and identically distributed (iid) data, to obtain a new model which takes the serial dependence in financial data into account and allows the estimation of conditional quantiles and other risk measures of interest.

As motivation consider Figure 1, whose top panel shows the times and sizes of the negative daily percentage log returns of Bayer shares exceeding a threshold $u = 1.5\%$ between 2 January 1973 and 23 July 1996. In this paper these data will be treated as a realisation of a marked point process: in addition to the event times there is a mark (the negative log return) attached to each event. To clarify the marked point process terms used in this paper, we will refer to exceedances of the threshold $u$ as extreme events, to be described both by their times and their mark sizes, the latter being the size of the excess (negative log) percentage return over the threshold.

The classical POT model of EVT, which is motivated by the asymptotic behaviour of threshold exceedances for iid data, assumes that if $u$ has been chosen high enough then the events may be described by a particular marked Poisson process: events occur in time according to a homogeneous Poisson process, and the mark sizes are independently and identically distributed according to the generalised Pareto distribution (GPD). However, Figure 1 contradicts the classical model. Under a homogeneous Poisson process the inter-event times would be independent exponential random variables, but the upper panel suggests clustering of events relative to this, and this is confirmed by the lower panel. The lower panel shows an exponential probability plot for the inter-event times. These are clearly far from exponential, giving evidence against a Poisson process of exceedances.

Independence of widely separated extremes seems reasonable in most applications, but extremes often cluster together and so display short-range dependence stemming from serial dependence of the underlying series. In a financial context, volatility bursts will produce clustered extremes, and it seems unrealistic to assume independent exponential inter-event times and independent mark sizes. These problems are often addressed by the
application of a declustering method, which fits the Poisson model to cluster maxima only. In a stationary series it turns out that the distribution of
cluster maxima is the same as the marginal distribution of an exceedance, so fitting the model to cluster maxima does not bias the GPD fit, but it raises the question of how to identify clusters from data. A recent contribution to this topic is by Ferro and Segers (2003); it has also been discussed by Davison and Smith (1990), Robinson and Tawn (2000), Embrechts, Resnick, and Samorodnitsky (1998), and in Section 8.1 of Embrechts, Klüppelberg, and Mikosch (1997). Although appropriate in some applications, declustering is undertaken specifically to avoid modelling short-term behaviour of extremes, a central focus of the discussion below.

This paper introduces a marked point process model for the behaviour of clusters of financial extremes, by combining a self-exciting process for the threshold exceedance times with a time dependent process for the threshold excesses. The form of the process allows realistic models in which the recent past affects the current intensity of arrival of events more than does the distant past, and it allows the intensity to depend on the marks. Related papers are Ogata (1988), who discusses several classes of stochastic models for the origin times and magnitudes of earthquakes, and Klüppelberg and Mikosch (1995), who introduce a shot noise model for ruin problems.

In Section 2 we outline relevant aspects of the classical peaks over threshold model, and then in Sections 3 and 4 we describe the new approach. Numerical illustrations are given in Section 5, and Section 6 contains a brief discussion.

2 Peaks Over Threshold Analysis

Consider a series of iid random variables $Z_1, \ldots, Z_l$ from a continuous distribution $F_Z(z)$ with upper endpoint $z_+$, which may be infinite. Let $N_u$ denote the (random) number of these that exceed a high threshold $u$, and assume that these occur at (discrete) times $T_i$ and cause threshold excesses (mark sizes) $W_j = Z_j - u$. Mathematical theory (Leadbetter 1991) under which $l \to \infty$ and $u \to z_+$ in such a way that $l\{1 - F_Z(u)\} \to \lambda > 0$ supports the modelling of the rescaled times of these events by a homogeneous Poisson
process with intensity $\lambda$. Conditional on $N_u = n$, the mark sizes $W_1, \ldots, W_n$ are an iid sample from the generalised Pareto distribution (GPD)

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

(1)

where $\sigma > 0$, and the support of the $W_j$ is $w \geq 0$ when $\xi \geq 0$ and $0 \leq w \leq -\sigma/\xi$ when $\xi < 0$. This particular distributional choice is motivated by a limit result in Extreme Value Theory (Pickands 1975, Davison and Smith 1990, Embrechts, Klüppelberg, and Mikosch 1997). The class of distributions $F_Z$ for which this result holds falls into three groups according to the value of the shape parameter $\xi$ in the limiting GPD approximation to the distribution of excesses. The case $\xi > 0$ corresponds to heavy-tailed distributions whose tails decay like power functions, such as the Pareto, Student’s $t$, Cauchy, Burr, loggamma and Fréchet distributions. The case $\xi = 0$ corresponds to distributions like the normal, exponential, gamma and lognormal, whose tails decay essentially exponentially. The final group of distributions ($\xi < 0$) are short tailed distributions with a finite right endpoint, such as the uniform and beta distributions.

Equation (1) provides a loglikelihood for estimation of $\sigma$ and $\xi$, namely

$$l(\sigma, \xi) = -n \log \sigma - \left(1 + 1/\xi\right) \sum_{j=1}^{n} \log (1 + \xi w_j/\sigma)_{+} .$$

Maximum likelihood estimation of $\xi$ and $\sigma$ is non-regular in the sense that the score statistic is not asymptotically normal if $\xi \leq -1/2$ (Davison 1984a, Davison 1984b, Smith 1985). For $\xi \leq -1$ the score statistic has infinite mean, so the usual Taylor expansions do not yield a consistent estimator. In financial applications heavy-tailed data, for which $\xi > 0$, are common, and maximum likelihood estimation has its usual properties of consistency and efficiency, so the generalized Pareto distribution can be used for statistical estimation, provided that the threshold is taken sufficiently high. As the loglikelihood has a global maximum of infinity, so the maximum likelihood estimators (MLEs) are taken to be the values which yield a local maximum, which is found numerically; Hosking and Wallis (1987) describe a procedure
based on Newton–Raphson iteration. Davison and Smith (1990) and Coles (2001) give extensive discussions of this model.

The choice of the threshold \( u \) can be important. Smith (1987) proposes a graphical technique to help in choosing the threshold and to assess the fit of the model; a “mean residual life plot” (Yang 1978) in which the mean excess over/under a threshold \( u \) is plotted as a function of \( u \).

Since number and mark size for the threshold exceedances are assumed independent the overall loglikelihood is

\[
l(\lambda, \sigma, \xi) = \log \left\{ P \left( N_u = n \right) \prod_{j=1}^{n} g_{\xi, \sigma}(w_j) \right\}
= n \log \lambda - \lambda - \log n! - n \log \sigma - (1 + 1/\xi) \sum_{j=1}^{n} \log(1 + \xi w_j/\sigma)_+,
\]

where \( g_{\xi, \sigma} \) denotes the probability density function of the GPD. The loglikelihood can written as \( l(\lambda, \xi, \sigma) = l_{N_u}(\lambda) + l_{W|N_u}(\xi, \sigma) \), so inference can be performed separately for the frequency of exceedances and their sizes.

One advantage of the POT method compared to the traditional approach based on maxima is that, as each exceedance is associated with a specific event, it is possible to allow the parameters \( \sigma \) and \( \xi \) to depend on covariates. Davison and Smith (1990) give an example of this in nuclear safety, and Smith and Shively (1995) assess the probability of high-level exceedances in the tropospheric ozone record as a function of meteorological information. Non-parametric trends have been used recently by Hall and Tajvidi (2000) and Davison and Ramesh (2000) in meteorological applications of extremes. Rootzén and Tajvidi (1997) use meteorological information in wind storm insurance and present a detailed analysis of Swedish wind storm claims. Chavez-Demoulin (1999) and Chavez-Demoulin and Davison (2001) combine the point process for exceedances with smoothing methods to give a flexible exploratory framework to model changes of extremes. Chavez-Demoulin and Embrechts (2003) apply these ideas in an insurance and financial context.
3 Marked Point Processes

The theory outlined in Section 2 assumes that the underlying series \( \{Z_t\} \) is independent, but this is unrealistic in most applications. Indeed, here \( \{Z_t\} \) represents the daily values of the negative log return on a financial asset, and these are likely to be dependent. Chapter 5 of Coles (2001) describes general results on modelling stationary dependent series. The point process paradigm under which the Poisson process representation may be applied to the largest value in a cluster remains applicable under weak conditions, but this leads to the practical difficulties of identifying clusters from data, and of modelling their properties. Dependence in stationary series can take many different forms, and it is impossible to develop a general characterization of the behaviour of extremes unless some constraints are imposed.

Below we modify the approach outlined in Section 2 by incorporating a model for dependence of the frequencies and sizes of the events over a high threshold \( u \). This is achieved by constructing a marked point process model for the exceedances, involving a self-exciting process for their times with mark-dependent parameters for the GPD fitted to the excesses.

Consider an event process \((T_i, W_i), i \in \mathbb{Z}\), which when observed over the period \((0, t_0]\) gives data \((T_1, W_1), \ldots, (T_n, W_n)\). Let \( \mathcal{H}_t \) denote the entire history of the process up to time \( t \), that is, \( \mathcal{H}_t \) is the sigma-algebra generated by the process up to time \( t \). To compute the likelihood of the observed data, we note that the joint density of \((T_1, W_1), \ldots, (T_n, W_n)\) given \( \mathcal{H}_0 \) can be written

\[
\prod_{i=1}^n f_{T_i, W_i | \mathcal{H}_{t_{i-1}}}(t_i, w_i | \mathcal{H}_{t_{i-1}}).
\]

We also have the information that \( T_{n+1} > t_0 \) and so the joint density of the data seen over \((0, t_0]\) is

\[
\prod_{i=1}^n f_{T_i, W_i | \mathcal{H}_{t_{i-1}}}(t_i, w_i | \mathcal{H}_{t_{i-1}})P(T_{n+1} > t_0 | \mathcal{H}_{t_n}). \tag{2}
\]

We make the assumption of conditional independence of times and marks
given information at the previous event, that is,

\[ f_{T_i,W_i|\mathcal{H}_{t_{i-1}}}(t_i, w_i \mid \mathcal{H}_{t_{i-1}}) = f_{T_i|\mathcal{H}_{t_{i-1}}}(t_i \mid \mathcal{H}_{t_{i-1}}) f_{W_i|\mathcal{H}_{t_{i-1}}}(w_i \mid \mathcal{H}_{t_{i-1}}), \]

in which case the loglikelihood based on (2) can be split into two parts:

\[ \ell = \left[ \sum_{i=1}^{n} \log f_{T_i|\mathcal{H}_{t_{i-1}}}(t_i) + \log P(T_{n+1} > t_0 \mid \mathcal{H}_{t_n}) \right]^{A} + \]

\[ \left[ \sum_{i=1}^{n} \log f_{W_i|\mathcal{H}_{t_{i-1}}}(w_i) \right]^{B}, \]

where the terms labelled A and B are respectively the contributions from the times and from the marks given the history of the process.

We first consider term A. Let \( N(t, s) \) be the number of events in \( (t, s] \), and let \( N \) be the corresponding counting process with intensity function

\[ \lambda_H(t) = \lim_{\delta t \to 0^+} (\delta t)^{-1} P \{ N(t, t + \delta t) > 0 \mid \mathcal{H}_t \}, \]

assumed to be well defined (Cox and Isham 1979). This represents the conditional probability of an event at time \( t \) given the history of the process to then. In terms of this the contribution to (3) made by term A may be written as

\[ \log \left\{ \prod_{i=1}^{n} \lambda_H(t_i) \exp \left( - \int_{0}^{t_0} \lambda_H(u) du \right) \right\}. \]

In practice, it is usually hard to specify a tractable but plausible form for \( \lambda_H(t) \). One possibility that seems to fit financial time series rather well is a self-exciting process, in which

\[ \lambda_H(t) = \mu + \sum_{j : t_j < t} \omega(t - t_j), \]

where \( \mu \) is a positive constant and \( \omega(u) \) is non-negative for \( u > 0 \) and otherwise zero. This intensity function leads to events occurring in clusters, when an initial event increases the rate of arrival of events above the background \( \mu \), and this rate is raised by the superposition of the \( \omega(t - t_j) \) for previous
events. Typically \( \omega(u) \) is taken to be monotonic decreasing, so that recent events affect the current intensity more than do distant ones.

In the financial context it is natural to let \( \omega \) depend on the marks \( w_j \) also. We use the quite general form (Ogata 1988)

\[
\omega(t - t_j; w_j) = \frac{\psi e^{\beta w_j}}{(t - t_j + \gamma)^\rho}, \quad t > t_j,
\]

(5)

where \( \rho, \gamma, \psi, \beta, \mu > 0 \). Under this formulation the increase in intensity depends not only on the time since an event but also on the size of a past event. In our applications to financial time series we have found that setting \( \rho = 1 \) does not change the likelihood significantly, and we shall use this simpler model.

When applied to the Bayer data of Figure 1, the maximised value of part A of (3) is \(-1613\), very significantly higher than the corresponding value \(-1673\) for the model of constant \( \lambda \). The parameter estimates, with standard errors in parentheses, are \( \hat{\mu} = 0.0100 \) (0.0062) events/day, \( \hat{\psi} = 0.071 \) (0.0181) events/day, \( \hat{\gamma} = 3.6269 \) (1.1342) days and \( \hat{\beta} = 0.2348 \) (0.0877). The number of events generated by the background is about \( 253\hat{\mu} = 2.53 \) events/year. One interpretation of this is the following: after a drop in daily returns of size \( w_j = 3\% \), \( \lambda_H(t) \) jumps by \( \hat{\psi} e^{\hat{\beta} \hat{\gamma}} \hat{\gamma} \approx 0.0395 \) events/day while a drop of size \( w_j = 7\% \) induces a jump of \( \hat{\psi} e^{7\hat{\beta} \hat{\gamma}} / \hat{\gamma} \approx 0.10 \) events/day.

The top panel of Figure 2 shows the fitted intensity \( \hat{\lambda}_H(t) \), which is initially low due to the lack of data before \( t = 0 \). This would not be problematic in practice, as the purpose of such a model would be to provide predictions for future days, and in practice presumably enough past data would be available. The lower left panel shows the cumulative intensity for the transformed process \( \hat{\lambda}_H(t_j) = \int_0^{t_j} \lambda_H(u)du \), which would be a straight line of unit gradient if the model fitted perfectly. The cumulative intensity lies within overall 95% confidence limits and give no evidence against the model.

Now consider part B of (3), which corresponds to the amounts by which the process exceeds the threshold \( u \), conditional on the exceedance times. One natural way to model dependence among successive exceedances is to
Figure 2: Bayer share price data. The upper panel shows the estimated intensity $\lambda_H(t)$ events/day. The tick marks at the top of the panel show the event times. The lower left panel shows the estimated cumulative number of events $\Lambda_H(t_j)$ (solid) and two-sided 95% and 99% overall confidence limits (solid diagonal), based on the Kolmogorov–Smirnov statistic; the straight line shows perfect fit of the model. The lower right panel shows an exponential probability plot of the residuals using mark dependent GPD parameters.
consider a GPD model with mark and/or time dependent parameters, which we denote by $\text{GPD}(\xi_j, \sigma_j)$. Our exploratory analyses of financial time series have suggested that it is typically reasonable to take $\xi_j = \xi$ and a first order autoregressive process for the scale parameter, that is, $\log \sigma_j = a + bw_{j-1}$. This defines a first order Markov chain

$$W_i \mid W_{i-1} = w_{i-1} \sim \text{GPD}_{\xi, \exp(a + bw_{i-1})},$$

whose parameters $a, b$, and $\xi$ can be estimated by maximising

$$\prod_{i=2}^{n} f_{W_i \mid W_{i-1}}(w_i \mid w_{i-1}) f_{W_1}(w_1),$$

whose logarithm is term $B$ of (3). Note that the marginal distribution of $W_i$ will not be GPD, and that the chain need not be stationary. These models for the GPD parameters were obtained by model comparisons based on the likelihood ratio statistic. Amongst the possible alternative models, we tried different semi-parametric models for $\sigma_j$. A possible diagnostic for the marginal distribution may be based on that fact that when the GPD parameter model is correct, the residuals

$$R_j = \hat{\xi}^{-1} \log \left(1 + \frac{\hat{\xi} w_j}{\hat{\sigma}}\right), \quad j = 1, \ldots, n,$$

are approximately independent unit exponential variables. The lower right panel of Figure 2 shows an exponential probability plot of these quantities for the Bayer data, for which $\hat{\xi} = 0.2145$ (0.0467) and $\log \hat{\sigma}_j = -0.3061$ (0.1181) + 0.0820 (0.0454)$w_{j-1}$. The model appears quite reasonable. To check that there is no further time series structure, Figure 3 shows the correlogram or sample autocorrelation function (ACF) for the residuals (left panel) and for their squares (right panel). Both correlograms are negligible at nearly all lags: the residuals seem be a white noise process.

4 Conditional Value-at-Risk

In applications the purpose of fitting our model would be to estimate the conditional quantile of the predictive distribution for the return over the
next day,

\[ z_q^t(1) = \inf \{ z \in \mathbb{R} : F_{Z_{t+1}|\mathcal{H}_t}(z) \geq q \} , \]

using data observed up to time \( t \). In practice we would take \( q = 0.95 \) or 0.99, thus defining a conditional VaR. Now

\[ P ( Z_{t+1} > z \mid \mathcal{H}_t ) = P ( Z_{t+1} - u > z \mid Z_{t+1} > u, \mathcal{H}_t ) \times P ( Z_{t+1} > u \mid \mathcal{H}_t ) , \]

and the second term on the right of this expression can be estimated using the self-exciting process. It represents the conditional probability of an event in \((t, t+1]\), that is,

\[ 1 - P \{ N(t, t+1) = 0 \mid \mathcal{H}_t \} = 1 - \exp \left( - \int_t^{t+1} \lambda(\mu) d\mu \right) . \]

The first term may be estimated using the autoregressive GPD model,

\[ Z_{t+1} - u \mid \mathcal{H}_t \sim GPD_{\xi, a+b(z_{t}-u)} . \]

The required quantile \( z_q^t(1) \) is then the solution of the equation

\[ P ( Z_{t+1} > z_q^t \mid \mathcal{H}_t ) = 1 - q , \]

with parameters replaced by their maximum likelihood estimates.
In the next section we outline how these calculations may be used to provide estimates of the conditional VaR for three different real data sets, and investigate their properties by simulation.

5 Numerical Illustrations

5.1 Bayer data

Figure 4 shows estimates of the 99% VaR under various models for the period from April 1987 to April 1988, including the market crash of the 19th of October 1987. The horizontal line corresponds to the unconditional VaR calculated from the basic peaks over threshold model with constant parameters \( \lambda, \xi \) and \( \sigma \). The red line is the conditional VaR estimated for each day successively by fitting the self-exciting process for the exceedance times and the autoregressive GPD model for the exceedance values, to the preceding data. The green curve is the estimated conditional VaR when the self-exciting process is fitted to the times, but the exceedances have a GPD model with constant parameters. The blue curve corresponds to the model with constant \( \lambda \) and the autoregressive GPD model.

The effect of applying a self-exciting process is highlighted, providing higher VaR levels after higher extreme events like crashes. It is possible to calculate confidence intervals for the conditional VaR estimator based on the asymptotic normality of the estimators of the self-exciting and GPD coefficients. Figure 5 shows upper and lower bounds of a 95% confidence interval based on 200 Monte Carlo simulations. As one would expect, the interval is wider when the estimated conditional 99% VaR is higher, i.e. in periods of crashes or high volatility.

5.2 Djind data

Consider the negative daily log returns values (\( \times 100 \)) of the Djind index from 29 July 1996 to 25 July 2001 (Figure 6). We fix a threshold \( u = 1.4 \) such that around 10% of the data are excesses (this corresponds to the tail of a distribution) and fit the self-exciting process given by (4) and
Figure 4: Bayer share price data from April 1987 to April 1988. The red line shows the 99% estimated conditional VaR from the self-exciting process combined with the POT method for the sizes with GPD parameters: $\hat{\xi}=\text{constant and } \log \hat{\sigma}_j = \hat{\zeta} + \hat{b} w_{j-1}$. The green curve is the estimated VaR with self-exciting process combined with MLEs for the GPD parameters. The blue line corresponds to the model with constant intensity $\lambda$ combined with a GPD model with mark dependent scale parameter; $\hat{\xi}=\text{constant and } \log \hat{\sigma}_j = \hat{\zeta} + \hat{b} w_{j-1}$. The horizontal line corresponds to the model with constant estimated parameters $\hat{\lambda}, \hat{\xi}$ and $\hat{\sigma}$.

(5). The parameter estimates (standard errors) are $\hat{\mu} = 0.0263$ (0.0141), $\hat{\psi} = 0.0884$ (0.0616), $\hat{\gamma} = 6.349$ (5.1584), and $\hat{\beta} = 0.2589$ (0.2551). Compar-
Figure 5: Bayer share price data from April 1987 to April 1988. The points are the negative daily log returns (×100). The solid lines shows the 99% estimated conditional VaR from the self-exciting process combined with the POT method for the sizes with GPD parameters: $\xi=$constant and $\log \sigma_j = \hat{a} + \hat{b}w_{j-1}$. The dotted lines are the 95% Monte Carlo confidence intervals.

Figure 6: Djind index data. Negative daily percentage log returns (×100) from 29 July 1996 to 25 July 2001.

Comparison with the model of constant $\lambda$ gives strong evidence for the self-exciting process, as the maximised value of part A of (3) is $-377$, significantly higher than the corresponding value $-384$ for the model of constant $\lambda$.

The top panel of Figure 7 shows the fitted intensity $\hat{\lambda}_H(t)$, which is
Figure 7: Djind index data. The upper panel shows the estimated intensity $\hat{\lambda}_H(t)$ (events/day). The tick marks at the top of the panel show the event times. The lower left panel shows the estimated cumulative number of events $\hat{N}_H(t_j)$ (solid) and two-sided 95% and 99% overall confidence limits (solid diagonal), based on the Kolmogorov–Smirnov statistic; the straight line shows perfect fit of the model. The lower right panel shows an exponential probability plot of the residuals obtained mark dependent GPD parameters.
again low initially due to the lack of data before \( t = 0 \). The lower left panel shows that the cumulative intensity for the transformed process \( \hat{\Lambda}_H(t_j) \) lies well within overall 95% confidence limits; this gives no evidence against the model.

We fit the autoregressive GPD model to the excesses over the threshold \( u \) and find the estimates (standard errors) to be \( \hat{\xi} = 0.1281 (0.0760), \log \sigma_j = -0.1666 (0.2127) - 0.1079 (0.1126)w_{t-1} \). This model is better at the 5% level than the simplest model, as the maximised value of part B of (3) is \(-82\), significantly higher at the 5% level than the corresponding value \(-84\) for the model of constant \( \sigma \). The lower right panel of Figure 7 shows an exponential probability plot of the residuals, which suggests reasonable fit of the model. Furthermore the correlogram of the residuals shows no remaining time series structure.

Figure 8 shows estimates of the 99% VaR under the various models mentioned above. The initial values are again low when a self-exciting process is applied (red and green curves) due to the initial lack of data. The graph highlights overall changes in the conditional VaR and especially jumps in the red and green estimates after major drops in the market, due to the form of \( \omega \). The estimated conditional VaR from the model with constant \( \lambda \) and autoregressive GPD varies only locally and by much less than the other two dynamic models, compared to which it seems to overestimate the VaR.

### 5.3 dem exchange rate data

We consider the US dollar/Deutschmark exchange rate (dem) from 3 January 1980 to 21 May 1996 (negative daily log returns \( \times 100 \) in Figure 9). The value of \( u \) is 0.6. The estimates and their standard errors are \( \hat{\mu} = 0.0113 (0.0073), \hat{\psi} = 0.0430 (0.0149), \hat{\gamma} = 2.648 (0.8858), \hat{\beta} = 1.0142 (0.1249) \), and the maximised value of part A of (3) is \(-1249\), very significantly higher than the corresponding value \(-1299\) for the model of constant \( \lambda \). The top panel of Figure 10 shows the fitted intensity \( \hat{\Lambda}_H(t) \), and the lower left panel shows that the estimated cumulative intensity for the transformed process
Figure 8: Djind index data. The red line shows the 99% estimated conditional VaR from the self-exciting process combined with the POT method for the sizes with GPD parameters: $\xi=\text{constant}$ and $\log(\sigma_j) = \hat{\alpha} + \hat{b}w_{j-1}$. The green curve is the estimated VaR with self-exciting process combined with MLEs for the GPD parameters. The blue line corresponds to the model with constant intensity $\lambda$ combined with a GPD model with mark dependent scale parameter; $\xi=\text{constant}$ and $\log(\sigma_j) = \hat{\alpha} + \hat{b}w_{j-1}$. The horizontal line corresponds to the model with constant estimate parameters $\hat{\lambda}, \hat{\xi}$ and $\hat{\sigma}$.

lies within overall 95% confidence limits. On fitting the GPD model with autoregressive scale parameter to the marks we obtain $\hat{\xi} = 0.1229$ (0.0428), $\log(\sigma_j) = -1.1864$ (0.1192) + 0.1125 (0.1440)$w_{t-1}$. The lower right panel of
Figure 9: Dem index data. Negative daily log returns ($\times 100$) from 3 January 1980 to 21 May 1996.

Figure 10 shows the residuals against the exponential plotting positions using the estimate model and suggests a good model. Again, correlograms of the residuals give no evidence of time series structure.

We focus on the period from March 1992 to March 1993, which includes the market crash of 17 September 1992. With the same colors as above, Figure 11 shows the different estimated VaRs. Figure 12 gives the 95% Monte Carlo confidence interval (dotted lines) of the 99% conditional VaR estimator. The upper bound rises to values between 5 and 8% in the period of crash and then fluctuates more than previously.

5.4 Backtesting and simulation study

To backtest the method on a historical series $z_1, \ldots, z_l$, we compare the 99% estimated conditional VaR, $\hat{z}_{0.99}^t(1)$, given knowledge of returns up to and including day $t$ in the set $T = \{1, \ldots, l\}$, with the observed value at time $t + 1$, $z_{t+1}$. The conditional VaR is estimated using our marked point process on the series $\{w_1, \ldots, w_n\}$ of exceedances of the series $z_1, \ldots, z_l$ over a certain threshold $u$. That is the self-exciting process is combined with the autoregressive GPD model (red curves in the graphs of Section 5). A
Figure 10: Dem exchange rate data. The upper panel shows the estimated intensity $\lambda_H(t)$ events/day. The tick marks at the top of the panel show the event times. The lower left panel shows the estimated cumulative number of events $\Lambda_H(t_j)$ (solid) and two-sided 95\% and 99\% overall confidence limits (solid diagonal), based on the Kolmogorov-Smirnov statistic; the straight line shows perfect fit of the model. The lower right panel shows the residuals against exponential plotting positions using mark dependent GPD parameters.
violation is said to occur whenever $z_{t+1} > \hat{z}_{0.99}^t(1)$. Following the method given in Section 3 of McNeil and Frey (2000), it is possible to develop a
Figure 12: Dem exchange rate from March 1992 to March 1993. The points are the negative daily log returns (×100). The solid lines shows the 99% estimated conditional VaR from the self-exciting process combined with the POT method for the sizes with GPD parameters: $\hat{\xi}=$constant and $\log \hat{\sigma}_j = \hat{a} + b\psi_{j-1}$. The dotted lines are the 95% Monte Carlo confidence intervals.

The binomial test of the success of marked point process estimation based on the number of violations. The indicator for a violation at time $t$ is the Bernoulli variable
\[ I_t = 1 \{ z_{t+1} > z_q \} \sim B(q, 1-q), \]
and if we assume that $I_t$ and $I_s$ are independent for $t \neq s$, the total number of violations is binomially distributed under the model:
\[ \sum_{t \in T} I_t \sim B(l, 1-q). \]

Here we do not suppose that $Z_{t+1}$ and $Z_{s+1}$ are independent but we assume that the dependence of $I_t$ and $I_s$ is weak. To assess this assumption, we perform a runs test based on counting the number of runs (strings) of consecutive zeros and ones in the indicator sequence. For the Bayer share price data, for instance, the sample size is 6146, the number of zeros (no violations) is 6081 and the number of ones (violations) is 65. The observed number of runs is 130 which is close to the expect number 129 under independence. So the runs test does not reject the independence hypothesis of the sample of data.
Table 1: Backtesting results for three log return series. Theoretically expected number of violations, number of violations observed using our approach, and p-values for the binomial test.

<table>
<thead>
<tr>
<th>Sample name (size)</th>
<th>Expected number of violations</th>
<th>Observed number of violations</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bayer (6146)</td>
<td>61</td>
<td>65</td>
<td>0.61</td>
</tr>
<tr>
<td>Djind (1303)</td>
<td>13</td>
<td>11</td>
<td>0.58</td>
</tr>
<tr>
<td>Dem (4274)</td>
<td>42</td>
<td>42</td>
<td>1</td>
</tr>
</tbody>
</table>

If our method correctly estimates the conditional quantiles, the empirical version of this statistic $\sum_{t \in T} 1_{\{z_{t+1} > \hat{z}_{q}\}}$ has approximately the binomial distribution $B(l, 1 - q)$, and we can perform a two-sided test of the null hypothesis against the alternative that the method has a systematic estimation error and hence gives too few or too many violations. A p-value less than or equal to 0.05 would be interpreted as evidence against the null hypothesis. Table 1 gives the backtesting results for the three log return series studied on which the new method was applied. On no occasion did our approach lead to rejection of the null hypothesis.

To assess our model we perform a simulation study based on a GARCH(1,1) model to the Bayer data, using Student t innovations of estimated degrees of freedom. We generated 1000 samples of length 1500, and for each we calculate $\hat{z}_{0.95}$ from our new method and compare it to the “exact” conditional quantile $z_{0.95}$ from the GARCH model using the mean squared error, which equals 0.629.

For comparison we also perform a second simulation study, generating 1000 samples of size 1500 from a GARCH(1,1) model with normal innovations applied to the Bayer data. We calculate the unconditional VaR using the basic approach that uses MLEs for the parameters $\lambda$, $\xi$ and $\sigma$, and obtain an estimated MSE of 1.28, more than twice the MSE obtained from the simulation study using our new method.

In light of these studies and our data analysis it seems that our marked point process approach for the excesses over high threshold provides a rather
realistic model for the extremal behaviour of the negative log returns of the
types studied here.

6 Discussion

This paper concerns tail estimation for financial time series, and in particular
estimation of value at risk for return series as a measure of market risk. We propose an approach that models within cluster behaviour, based on
an extension of the classical POT model involving a self-exciting process
for the exceedance times and a form of autoregressive GPD process for the
exceedances themselves. The form of the self-exciting process allows realistic
models in which recent events affect the current intensity more than do
distant ones, and which allows the intensity to depend on the event sizes.

In the market risk area, VaR estimation involves portfolios of more than
one asset. We are optimistic that our method can be extended to such mul-
tivariate series.

Conditional quantile estimation problem using EVT techniques is a cur-
rent and important topic. We are currently developing a fast interactive
tool that returns the different VaRs and related values (volatility, expected
shortfall and so forth) from the most recent existing methods (including
the method described in this paper and the approach from McNeil and
Frey (2000)), and also from existing standard methods (such as variance-
covariance and EWMA procedures). Riskometer consists of an online in-
strument that enable private individuals or financial institutions to enter
the portfolio for which the VaR is required and allows procedures to be
compared based on real data sets.

Acknowledgements

We thank Professor Paul Embrechts for helpful comments. This work was
stimulated by a workshop on managing uncertainty held at the Isaac Newton
Institute in Cambridge in July–August 2001, and partly supported by the
Swiss National Science Foundation through the NCCR FINRISK.
References


