Estimating Value-at-Risk for financial time series: an approach combining self-exciting processes and extreme value theory

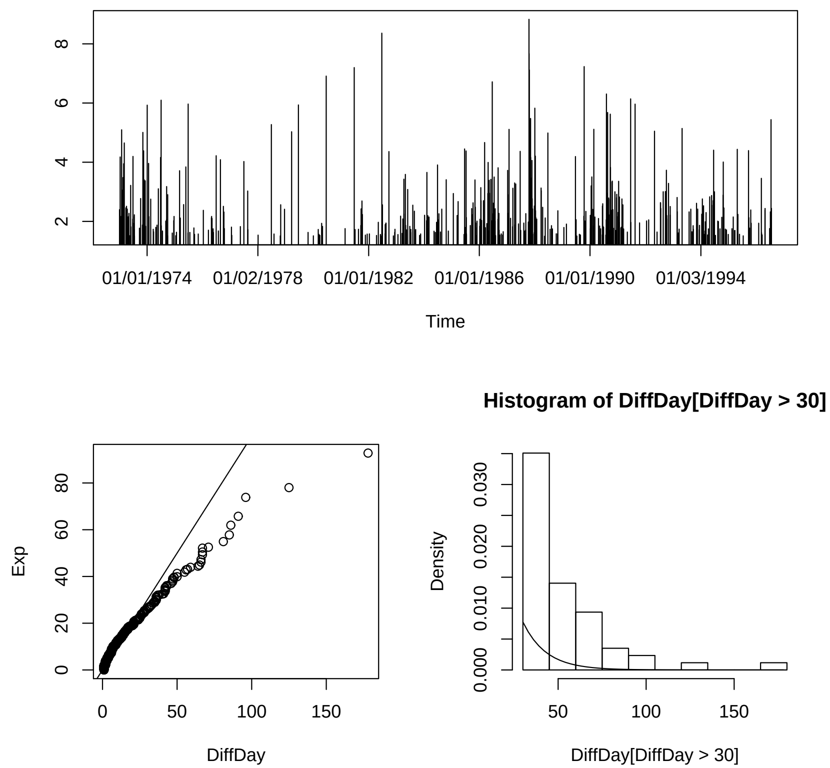
V. Chavez-Demoulin*, A. C. Davison† and A. J. McNeil*

ETHZ, Zürich*, EPFL, Lausanne†

valerie@math.ethz.ch

Riskday, 18th October 2002, Zürich

Times and marks of the negative daily log returns ($\times 100$) of Bayer shares over a threshold; data ranges from 2/1/1973 to 23/7/1996



Marked Point Process for (negative) log returns

- Times: Threshold exceedance times for some high threshold
- Marks: Excess returns over the threshold

- Time lengths between two successive events are clearly far from exponential
- Poisson process model for the occurrence of the events as in the classical POT (Peaks Over Threshold) method may be inappropriate

- Serial dependence causes clustering of large values
- An ad hoc declustering method is often applied: Fit the point process model to cluster maxima
- Practical problem is identification of clusters
- No attempt made to model within cluster behavior

We aim to model the behaviour of the data within a cluster

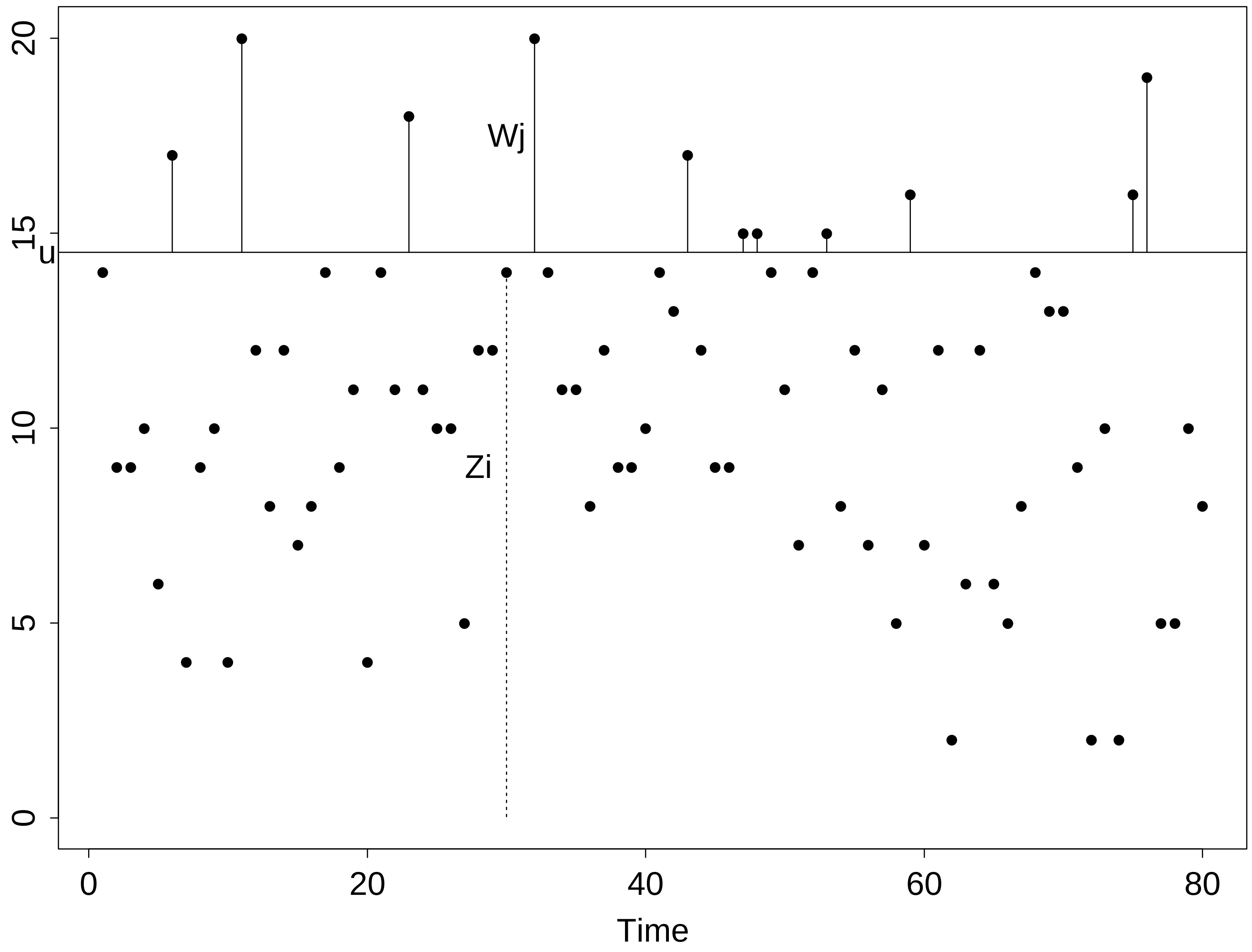
- We introduce a marked point process combining a self-exciting process for the threshold exceedance times with a time dependent process for the threshold excesses
- The form of the process allows realistic models in which recent events affect the current intensity more than do distant ones but it also allows the intensity to depend on the marks
- We show how the model may be used to estimate Value-at-Risk (VaR)

Contents

Classical POT Method

- Assumption: $Z_1, \ldots, Z_n$ iid. (or only very weakly dependent) data from $F_Z(z)$ in a wide class of continuous distribution functions
- Model:
  - Poisson process with intensity λ for the occurrence of events
  - Independent GPD threshold excesses
- Use of Model:
  - Estimation of return levels
  - Estimation of tail quantiles of the threshold excess distribution

The Point Process of Exceedances (POT)



Given the threshold u and conditional on n exceedances, the excesses $W_j = Z_j - u$ are GPD

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

where $\sigma > 0$, and the support is $w \geq 0$ when $\xi \geq 0$ and $0 \leq w \leq -\sigma/\xi$ when $\xi < 0$

Using (asymptotic) independence of the frequency and sizes of the exceedances, the loglikelihood is

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

and estimation can be performed separately for the point process of exceedance times and for the excesses

New Method

- Assumption: data from some stationary process
- Model:
  - Self-exciting process for the occurrence of events (intensity depends on the entire past of process)
  - GPD distribution marks with mark-dependent parameters
- Use of model:
  - Estimation of conditional VaR

Self-Exciting Process

$t_1 < \cdots < t_n$: times; $m_1, \ldots, m_n$: the associated marks; $\mathcal{H}(t)$ denotes the history up to time t

Usually hard to specify a tractable (well-behaved) but realistic form for $\lambda_{\mathcal{H}}(t)$. A possible form is a self-exciting process in which

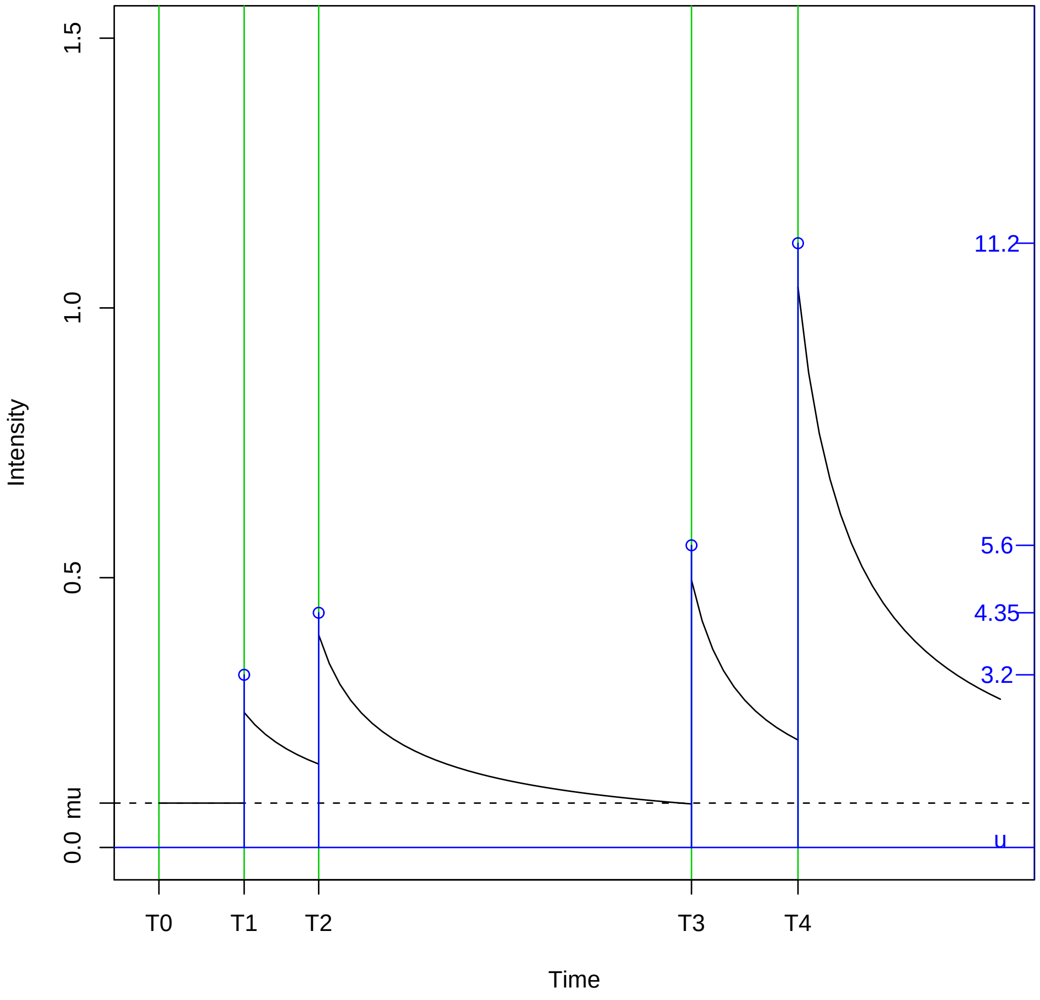
$$\lambda_{\mathcal{H}}(t) = \mu + \sum_{j:t_j<t} w(t - t_j).$$

where μ is a positive constant and $w(u)$ is monotonic decreasing, so recent events affect the current intensity more than do the distant ones.

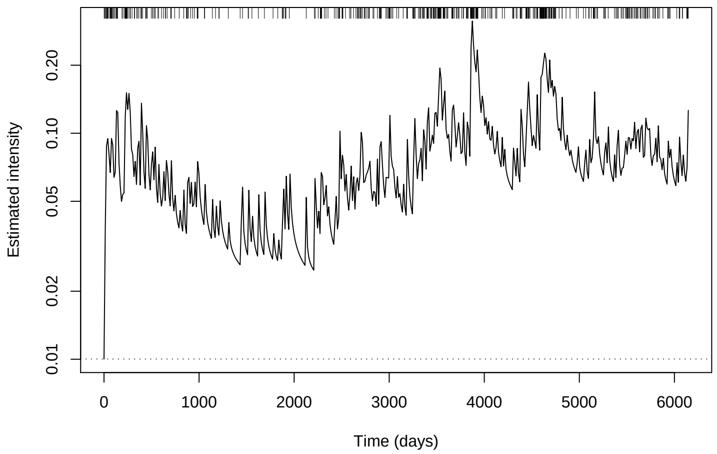
A quite general form:

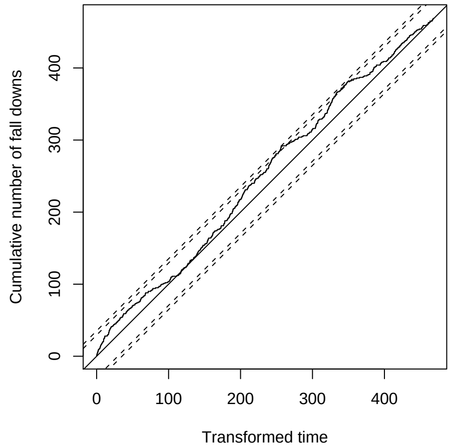
$$w(t - t_j; m_j) = \frac{\psi e^{\beta(m_j - U)}}{(t - t_j + \gamma)^{\rho}}, \qquad t > t_j$$

where $\rho, \gamma, \psi, \beta, \mu > 0$, and U is a threshold higher than u. 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



1.5
1.0
0.5
0.0
mu
Intensity
11.2
5.6
4.35
3.2
u
T0
T1
T2
T3
T4
Time





Fitted intensity $\hat{\lambda}_{\mathcal{H}}(t)$ and cumulative intensity for $\hat{\Lambda}_{\mathcal{H}}(t_j) = \int_0^{t_j} \lambda_{\mathcal{H}}(u)du$ which would be a straight line of unit gradient if the model fitted perfectly

Marks

Natural way to model the marks $(M_n)_{n \in \mathbb{N}}$ is to consider a GPD model with mark and/or time dependent parameters that we denoted by GPD(ξ_j, σ_j)

Several investigations (model comparisons based on likelihood ratio statistics) have shown that a reasonable model uses:

- $\xi_j = \xi$
- $\log \sigma_j = a + bm_{j-1}$

That is a first order Markov chain or autoregressive model of order 1

$$M_i \mid M_{i-1} = m_{i-1} \sim GPD_{\xi, a+bm_{i-1}}$$

with joint distribution

$$\prod_{i=1}^{n} P(M_i = m_i \mid M_{i-1} = m_{i-1}) P(M_1 = m_1)$$

Simulation Study for GPD

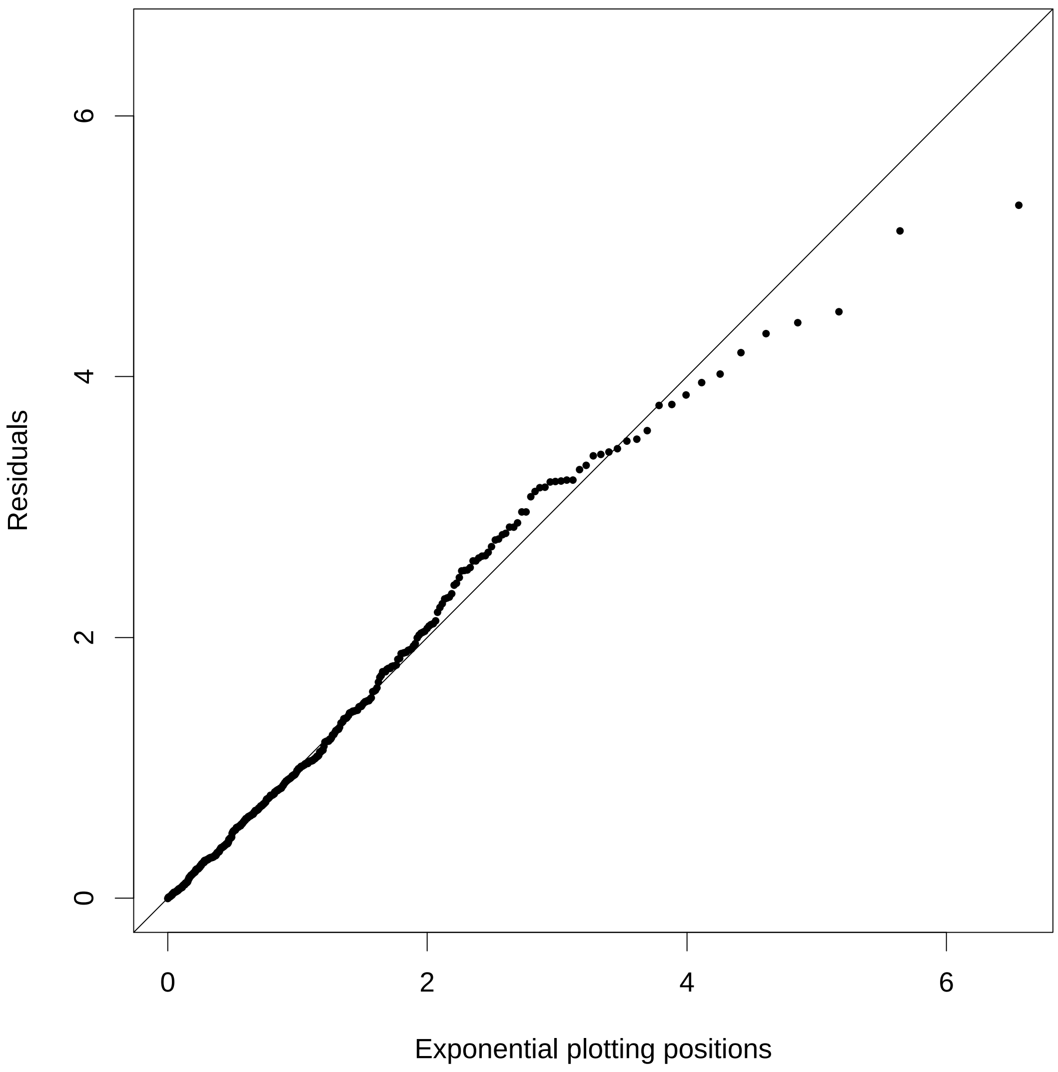
- 1000 samples of size $n = 200$ from GPD ($\xi = 0.2$, $\sigma = 0.8$)
- Null hypothesis H_0: $b = 0$
- rejected in 5% of the cases

Several applications to real data have shown that the model $\log \sigma_j = a + bm_{j-1}$ is better

To assess the model for the GPD parameters, a possible diagnostic is based on the result that when the model is correct, the residuals

$$R_j = \hat{\xi}^{-1} \log \left\{ 1 + \hat{\xi} m_j / \hat{\sigma}_j \right\}, \qquad j = 1, \ldots, n,$$

are approximately independent unit exponential variables



Residuals against exponential plotting positions using mark dependent GPD parameters

Value-at-Risk

Quantile of the predictive distribution for the return over the next day:

$$z_q^t(1) = \inf\left\{z \in \mathbb{R} \colon F_{Z_{t+1}|\mathcal{H}_t}(z) \geq q\right\}$$

We want to estimate this based on information up to time t (for q close to 1)

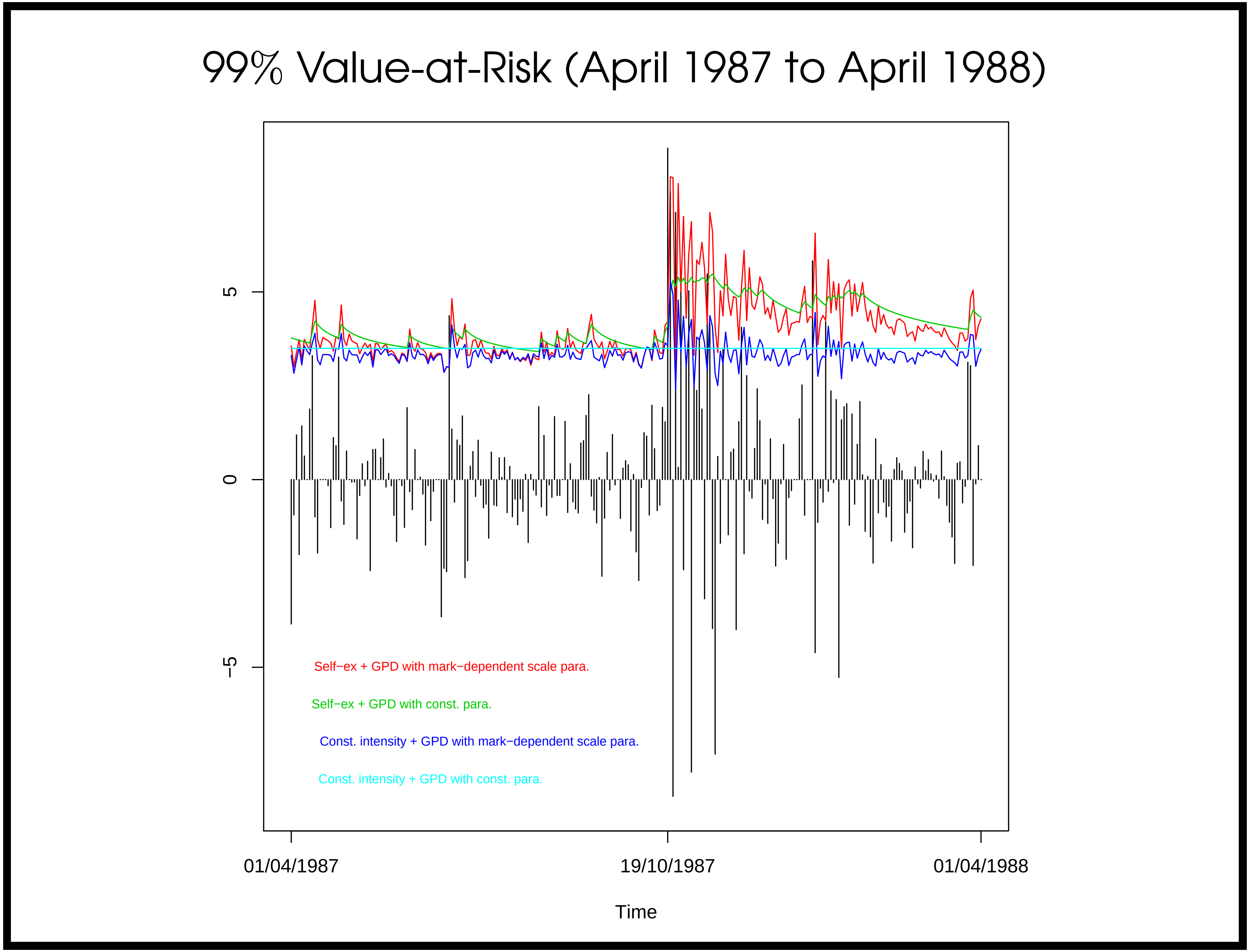
$$P\left(Z_{t+1} > z \mid \mathcal{H}_t\right) = P\left(Z_{t+1} - u > z \mid Z_{t+1} > u, \mathcal{H}_t\right)^{*} \times P\left(Z_{t+1} > u \mid \mathcal{H}_t\right)^{\dagger}$$

- † is calculated from the self-exciting process (conditional distribution of no event in $[t, t+1]$)
- * is calculated from the GPD with mark-dependent scale parameter

$$Z_{t+1} \mid \mathcal{H}_t \sim GPD \qquad \text{in } [u, \infty).$$

The quantile is the solution of

$$P\left(Z_{t+1} > z_q^t \mid \mathcal{H}_t\right) = 1 - q.$$


99% Value-at-Risk (April 1987 to April 1988)
5
0
−5
Self−ex + GPD with mark−dependent scale para.
Self−ex + GPD with const. para.
Const. intensity + GPD with mark−dependent scale para.
Const. intensity + GPD with const. para.
01/04/1987
19/10/1987
01/04/1988
Time

Backtesting

- For each day we compare the 99% estimated conditional VaR $\hat{z}^t_{0.99}$ with the observed value at time $t+1$, z_{t+1}. A violation is said to occur whenever $z_{t+1} > \hat{z}^t_{0.99}$
- Binomial test of the success of the marked point process estimation method based on the number of violations: H_0: The method correctly estimates the conditional quantiles

$$\sum_{t \in T} 1_{\{z_{t+1} > \hat{z}^t_q\}} \sim B(l, 1-q)$$

Backtesting

Sample size	Exp # of vio.	Nb of vio. (p-value)
Bayer(6146)	61	65 (0.61)
Djind(1303)	13	9 (0.26)
Dem(4274)	42	45 (0.65)

No rejection of the null hypothesis

Discussion

- The self-exciting process provides a more realistic model of clustering
- Possible improvement/generalisation by including smoothing methods to provide a flexible exploratory framework to model global changes (in time for instance) and trend of extremes