Modelling Temporal Dependencies of Extreme Events via Point Processes

Andreas Fuest

Modelling Temporal Dependencies of Extreme Events via Point Processes

Master Thesis

presented by

Andreas Fuest

to

Prof. Stefan Mittnik, PhD
Seminar für Finanzökonometrie

Institut für Statistik
Ludwig–Maximilians–Universität München

June 25, 2009

Contents

List of Figures

List of Tables

1 Introduction

This thesis deals with models for extreme events in financial time series data. As it is written in the middle of the largest crisis of the global financial system for the last 80 years, the virulence of the topic is almost all–too obvious.

The current crisis is not new insofar as it again shows the habit of extreme returns to occur in clusters. Market risk measures which are not able to account for the clustering phenomenon are not appropriate since their inability results in the clustering of violations. In presence of an appropriate risk measure, such violations should occur non–systematically, i.e. randomly. Figure 1.1 offers one striking example from the dusk of the crisis for the inadequacy of VaR estimates when they are confronted with the clustering phenomenon. The model does obviously not have the ability to estimate a reliable VaR from September 2008 on at latest.[1]

Clusters of extreme events can also be seen as an implication of the already well–known pattern of volatility clustering in financial returns. In times of high volatility extreme events occur very often, in times of low volatility we observe none of them.

Unlike the well established volatility models — with GARCH and its derivatives being the most prominent — the main goal of this work is to better understand the temporal aspect, i.e. the durations between extreme events. Particularly it should be analyzed to which extent extreme events can be forecast or anticipated.[2] This might help to construct tools for risk managing purposes that perform better in view of volatile times.

To this end, the thesis is organized as follows. Chapter 2 reviews the typical properties (stylized facts) of financial returns on the basis of Dow Jones Industrial Average index data, particularly with durations in view.

The main section is divided into two parts. Part I builds the core of the thesis. It deals with self–exciting peaks–over–thresholds (POT) models. After a short review of some prerequisites (extreme value theory, point processes, risk measures), some point process models which aim to capture the change between tranquil and stormy periods are proposed (chapter 3). Applications of these models to univariate daily return series including goodness–of–fit testing and the in–sample and out–of–sample performance of Value–at–Risk estimations based on these models follow in chapter 4.

Part II introduces one further approach to model clustering of extreme values: duration

[1] An event that might be associated with a possible structural break is the fall of Lehman Brothers on September 15, 2008, when the firm filed for bankruptcy protection.

[2] The possibility of forecasts based on historical data contradicts the postulate of efficient capital markets, i.e. the martingale property of security prices as stated by Fama (1970).

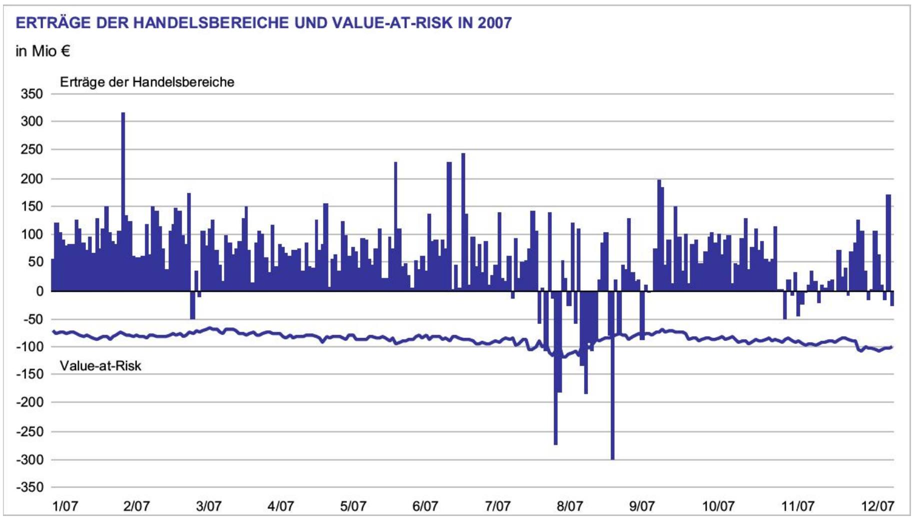


Figure 1.1.: Aggregated 0.99 Value–at–Risk for all trading divisions of Deutsche Bank in 2008. Most loss exceedances over the predicted VaR occur within relatively short intervals of a few days. Moreover, these smaller violation clusters can be combined to a larger cluster in the fourth quarter of 2008. Source: `http://geschaeftsbericht.deutsche-bank.de/2008/gb/serviceseiten/downloads/files/dbfy2008_risikobericht.pdf`

models (chapter 5) which originally stem from the econometric literature on intraday transaction data.

The final chapter 6 provides a concluding comparison of generalized POT and ACD models, some economic remarks and an outlook on further research to be done.

2

Statistical properties of financial data with durations in view

In this chapter, we give an overview of some typical properties of daily financial return data. In terms of these stylized facts, many financial (return) data from different areas as e.g. stocks and stock indices, commodity prices and exchange rates of currencies exhibit surprisingly similar properties. For instance, ARCH–models for volatility clustering have originally been applied to UK inflation data (Engle (1982)), but have meanwhile become a standard technique for portfolios consisting of a variety of asset classes. Thus, in the following we use continuous returns of the Dow Jones Industrial Average (DJIA) index data as representative for financial data. Some of its properties continue to hold even if we look at longer (weeks, months) or shorter (intraday) time intervals.

Since stock prices are mostly non–stationary (I(1)), it is common to model relative changes of prices, the return series. Major stylized facts of financial returns are

- Little serial dependence of returns, but strong correlation of squared or absolute returns,
- hence time-dependent volatility and volatility clustering.
- As a consequence of clustered volatility extreme events (losses as well as gains) appearing in clusters.
- Heavy-tailedness / leptokurtosis, i.e. the assumption of normally distributed returns (that is made e.g. within the Black–Scholes model) is not justifiable.

The plots of continuous or log returns of the DJIA index and its autocorrelations shown in figures 2.2 and 2.3 underline the first two stylized facts. Returns seem to fluctuate almost symmetrically around zero, and there are no periods that can exclusively be characterized as either boom or bust stages, i.e. the returns itself show little serial dependence. In contrast, the return series exhibits volatility clusters which can be inferred from both eye inspection of the return series as well as the ACF plot presented in the right panel of figure 2.3.

An immediate consequence of volatility clustering is clustering of extreme returns — positive as well as negative.

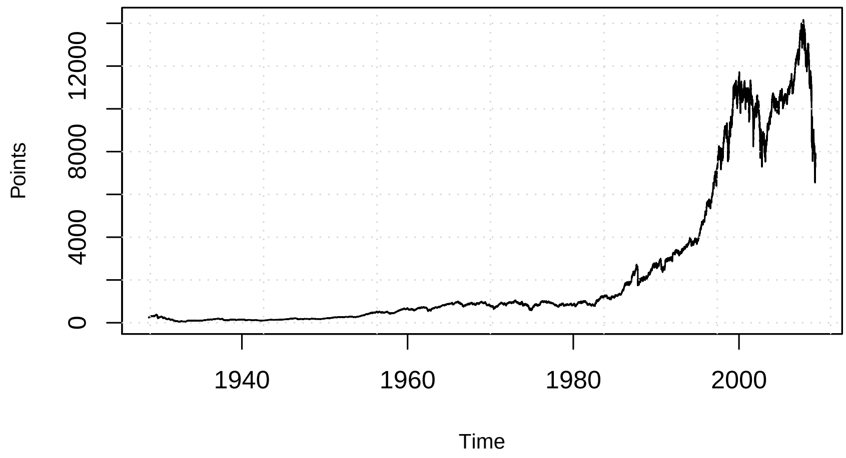


Figure 2.1.: Dow Jones Industrial Average prices between 1928-10-02 and 2009-03-27

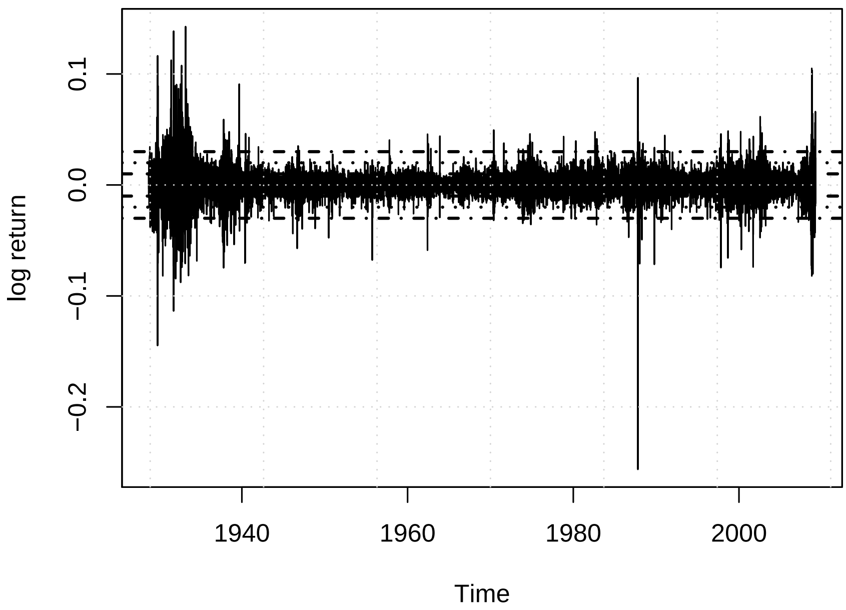


Figure 2.2.: Continuous returns of the Dow Jones Industrial Average (DJIA) index between 1928-10-02 and 2009-03-27. The horizontal lines indicate losses and gains larger than 1, 2 and 3 %, respectively.

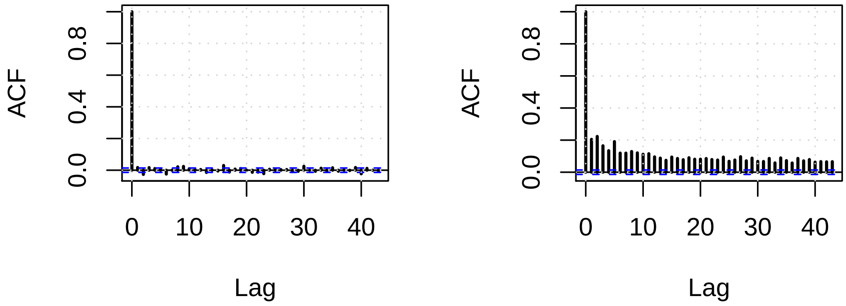


Figure 2.3.: Autocorrelations of the Dow Jones Industrial Average returns between 1928-10-02 and 2009-03-27 (left panel) and the absolute returns (right panel).

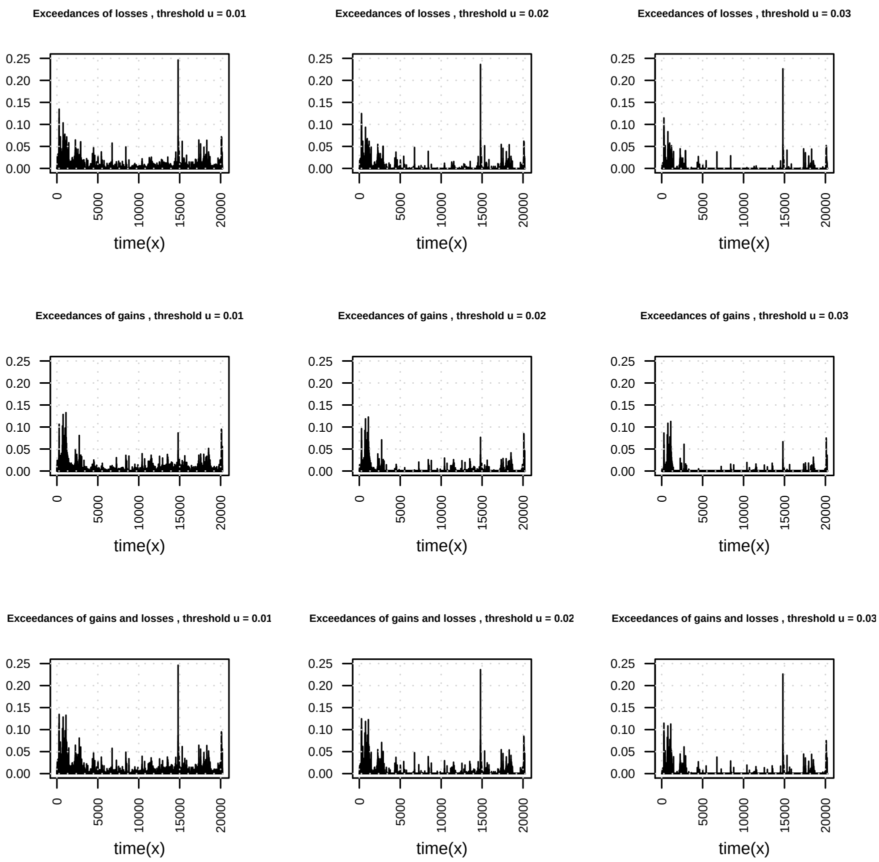


Figure 2.4.: Threshold exceedances of continuous DJIA returns (losses, gains, and both) between 1928-10-02 and 2009-03-27 with thresholds of 1, 2 and 3 %.

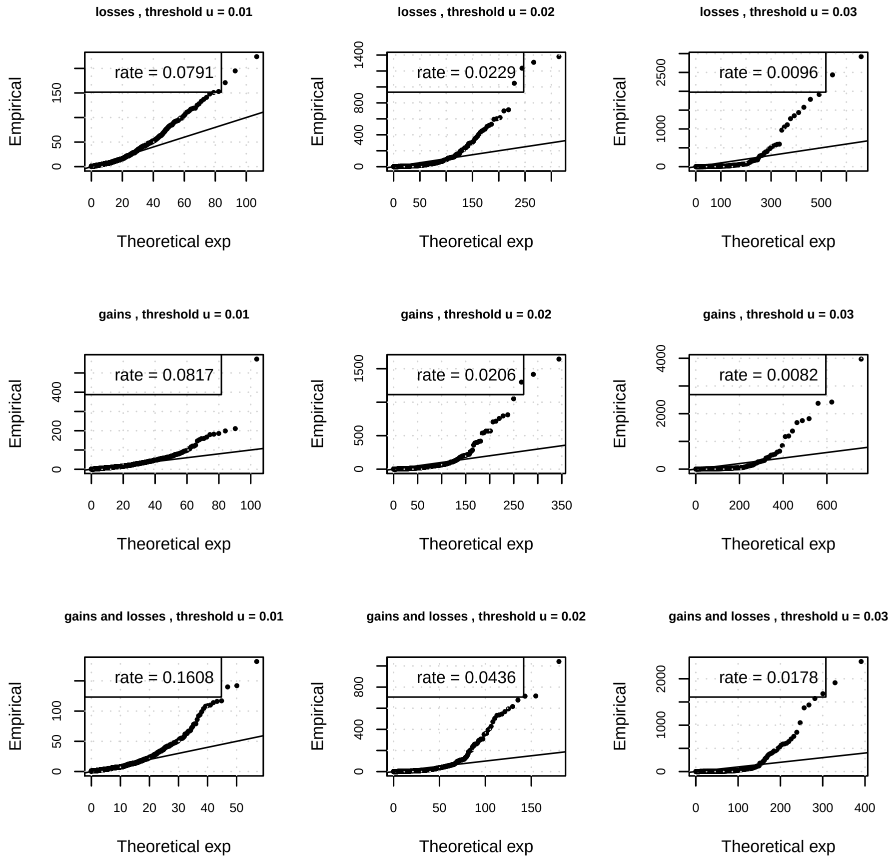


Figure 2.5.: Emprirical vs. theoretical quantiles of durations between extreme DJIA returns (losses, gains, and both), 1928-10-02 to 2009-03-27, with thresholds of 1, 2 and 3 %. The rate of the exponential distribution is estimated via ML, i.e. the reciprocal of the mean duration between events. In all cases most of the observations (70 to 90 %) lie below the line with slope 1, whereas some outliers can be found very high above. While the latter represent periods of tranquility, the former can be interpreted as durations within clusters of extreme events.

Laws of seismicity and financial returns

The models to be discussed in chapters 3 and 4 originally stem from seismological applications of point process models. Interestingly, parallels between seismological and financial earthquakes do not only exist in metaphorical terms: magnitudes and temporal structure of earthquakes and extreme returns comprise similar patterns.

To show this, we assume exponentially distributed magnitudes of earthquakes or extreme returns. The probability of an exceedance of threshold u is thus given by

$$\begin{aligned} P(\text{exceedance}|u) &= \int_u^\infty \lambda e^{-\lambda s} ds \\ &= \left[-e^{-\lambda s}\right]_u^\infty \\ &= e^{-\lambda u} \end{aligned}$$

For the exponential distribution — in point process language: for a homogeneous Poisson process N — this probability times the width of the interval under observation, a constant n, yields the expected number of events $E(N)$. Taking logarithms we obtain

$$E(\log N) = \log n - \lambda u.$$

Therefore, under the assumption of an exponential distribution the logarithmized number of exceedances should be linear in the threshold u. Figure 2.6 illustrates that this connection known from earthquake magnitudes and called the *Gutenberg–Richter law* in the seismological literature (Gutenberg and Richter (1954)) is fulfilled surprisingly well by our Dow Jones data.

Likewise, the decaying probability of further occurrences of shocks after a mainshock which is responsible for the temporal clustering of events is known in the seismological literature as the *Omori law* (Omori (1894)). Its modified version says that the decay of the aftershock frequency (or "rate") R can be described by

$$R(t) = \frac{k}{(c+t)^{-(\rho+1)}},$$

where k, c and ρ are constant parameters. As we will see, this "rate" R is not precisely equivalent to the intensity or hazard rate $\lambda(t)$ of our point process models, but it is one major ingredient.

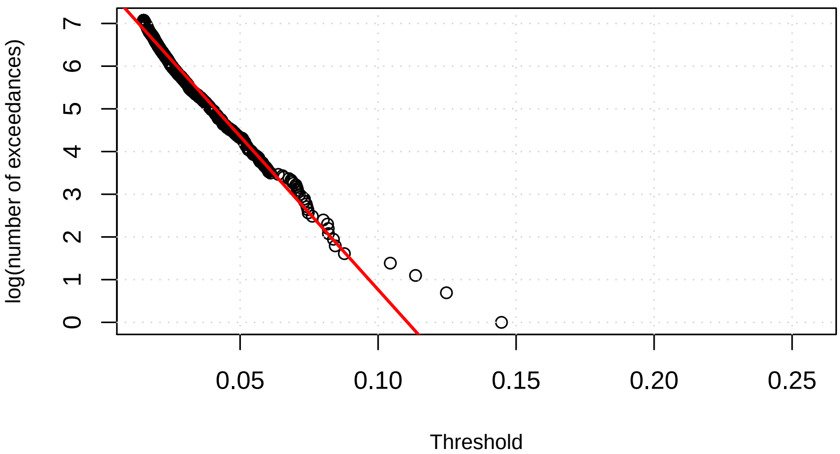


Figure 2.6.: The Gutenberg–Richter law says that if we plot the (originally: base 10, here: natural) logarithm of the number of events with magnitude greater than a threshold u (vertical axis) against u (horizontal axis), there should be a straight line. This is equivalent to magnitudes having an exponential distribution.

Part I.

Point process models

3

Methodological background

The two constituting components of the kind of processes we are interested in are times and "marks", i.e. exceedances. The temporal part is basically modelled using point processes. Moreover, *marked* point processes are also able to capture potential effects of marks on times. The distribution of the marks itself is modelled using results from *extreme value theory* (EVT).

Therefore, in the next section basic ideas from EVT are introduced. In the succeeding section we outline and discuss *self–exciting marked* point processes on the line and their usage within a generalized class of POT models.

3.1. Extreme value theory

EVT deals with the distribution of largest– and smallest–order statistics, i.e. EVT focuses on the tail of a distribution. This section contains to some extent a "customized" distillate of the seminal treatment of EVT in Embrechts, Klüppelberg, and Mikosch (1997) — not much unlike many other depictions to be found in the literature. There are two types of model for extreme values that can be distinguished, *block maxima* models and *threshold exceedance* models. Although we confine ourselves to the latter type, we introduce some concepts from the analysis of maxima first.

The central limit theorem (CLT) says that appropriately standardized sums $S_n = X_1 + \cdots + X_n$ of iid random variables X_i converge in distribution to the standard normal distribution for large n. The generalized extreme value distribution (GEV) plays a similar role within the analysis of extremes. According to EVT, normalized maxima $M_n = \max(X_1, \ldots, X_n)$ of "blocks" with size n converge in distribution to distributions of the GEV family.

Definition 1 (The generalized extreme value (GEV) distribution)**.** *The distribution function of the standardardized GEV distribution is given by*

$$H_\xi(x) = \begin{cases} exp(-(1+\xi x)^{-1/\xi}), & \xi \neq 0, \\ exp(-e^{-x}), & \xi = 0, \end{cases}$$

where $1 + \xi x > 0$. A three–parameter family with location $\mu \in \mathbb{R}$ and scale $\sigma > 0$ can be constructed as follows:

$$H_{\xi,\mu,\sigma}(x) := H_\xi((x-\mu)/\sigma).$$

The parameter ξ is the shape parameter of the GEV. Moreover, GEV embraces Fréchet ($\xi > 0$), Gumbel ($\xi = 0$), and Weibull distributions ($\xi < 0$) as special cases. The Weibull distribution has a finite right endpoint (it is "short-tailed"), whereas the Fréchet distribution is "fat-tailed".

We assume there exist sequences of real constants (d_n) and (c_n) with $c_n > 0 \quad \forall \quad n$ such that

$$\lim_{n\to\infty} P((M_n - d_n)/c_n \leq x) = \lim_{n\to\infty} F^n(c_n x + d_n) = H(x) \tag{3.1}$$

for non–degenerate d.f. $H(x)$. The central role of the GEV distribution is based on the following theorem.

Definition 2 (Maximum domain of attraction). *If (3.1) holds for some non–degenerate d.f. H, F is said to be in the* maximum domain of attraction *of H; we write $F \in MDA(H)$.*

Theorem 1 (Fisher-Tippett, Gnedenko). *If $F \in MDA(H)$ for some non–degenerate d.f. H, then H has to be a distribution from the GEV family.*

Block maxima methods do not deal parsimoniously with the data since there is a trade–off between the size of the blocks and the number of blocks to be constructed from a given dataset. A different characterization of "extremeness" can be achieved by considering only data which exceed a certain "high" threshold. The pivotal distribution for exceedances over thresholds is the generalized Pareto distribution (GPD).

Definition 3 (Generalized Pareto distribution (GPD)). *The distribution function of the GPD is given by*

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

where $\beta > 0$ and $x \geq 0$ when $\xi \geq 0$ and $0 \leq x \leq -\beta/\xi$ when $\xi < 0$. The parameters ξ and β are responsible for shape *and* scale *of the distribution, respectively.*

It is possible to add a location parameter, say ν, by replacing x by $(x - \nu)$. Special cases of the GPD include the ordinary Pareto distribution (with $\alpha = 1/\xi$ and $\kappa = \beta/\xi$) and the exponential distribution when $\xi = 0$; when $\xi < 0$, the distribution is short–tailed, i.e. it has a finite right endpoint (for examples see figure 3.1). Moreover, $G_{\xi,\beta} \in \text{MDA}(H_\xi)$ holds for any $\xi \in \mathbb{R}$.

Because it plays a crucial role within the model likelihoods in later sections, we take the first derivative of the cdf including the location parameter ν in order to obtain the corresponding density function.

$$g_{\xi,\nu,\beta}(x) = \begin{cases} -\frac{1}{\beta}\left(1 + \xi\frac{x-\nu}{\beta}\right)^{-1/\xi-1}, & \xi \neq 0, \\ \frac{1}{\beta}\exp\left(-\frac{x-\nu}{\beta}\right), & \xi = 0, \end{cases}$$

The GPD is employed to model the *excess distribution* over high thresholds. Therefore we present two more definitions that are useful in this context.

Definition 4 (Excess Distribution over Threshold u). *Let X be a rv with d.f. F. The distribution of excesses over the threshold u is given by its d.f.*

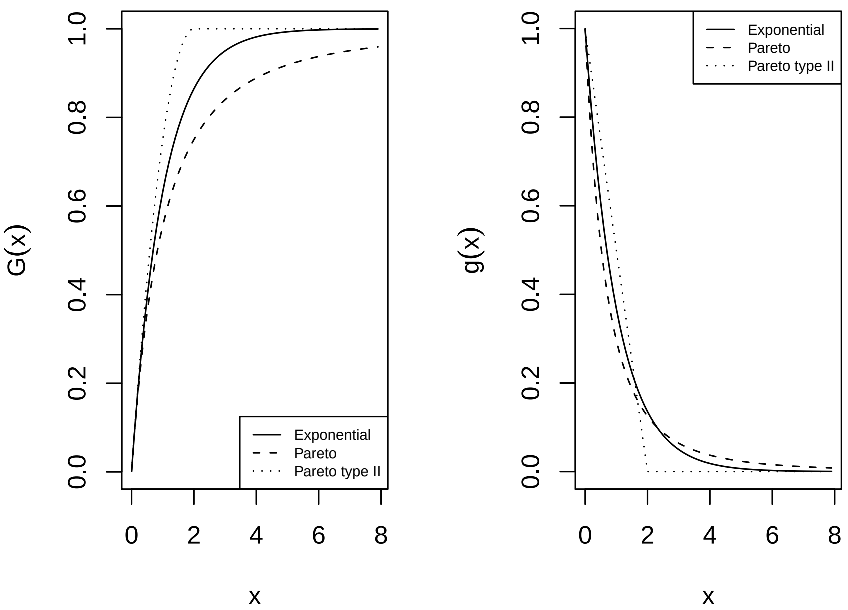


Figure 3.1.: Distribution functions (left) and densities (right) for three special cases of the GPD

$$F_u(x) = P(X - u \leq x| > u) = \frac{F(x+u) - F(u)}{1 - F(u)} \tag{3.2}$$

where $0 \leq x < x_F - u$ *and* $x_F \leq \infty$ *is the right endpoint of* F*.*

Definition 5 (Mean excess function)**.** *The mean excess function of a rv* X *with finite mean is given by*

$$e(u) = E(X - u|X > u). \tag{3.3}$$

Using equation 3.2, it is easy to obtain the excess distributions for the exponential and GPD distribution, respectively. For the former we have

$$\begin{aligned} F_u(x) &= \frac{1 - e^{-\lambda(x+u)} - (1 - e^{-\lambda u})}{1 - 1 + e^{-\lambda u}} \\ &= 1 - e^{-\lambda x} \\ &= F(x), \end{aligned}$$

whereas the latter yields ($\xi \neq 0$)

$$\begin{aligned} F_u(x) &= \frac{1 - (1 + \xi(x+u)/\beta)^{-1/\xi} - 1 - (1 + \xi x/\beta)^{-1/\xi}}{1 - 1 - (1 + \xi u/\beta)^{-1/\xi}} \\ &= 1 - \left(\frac{1 + \xi x/\beta + \xi u/\beta}{1 + \xi u/\beta}\right)^{-1/\xi}, \end{aligned}$$

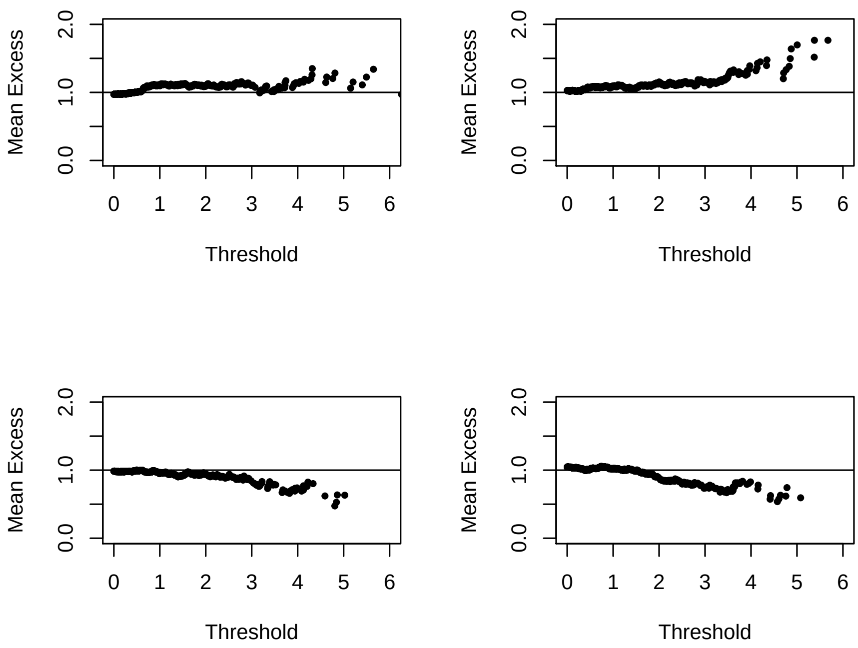


Figure 3.2.: Mean excess plots for 4 samples (size $n = 1000$) drawn from the standard exponential distribution. As the exceedance sizes are independent of the threshold u, mean excesses fluctuate randomly around the mean (unity).

and by expanding the fraction we finally obtain

$$
\begin{aligned}
&= 1 - \left(1 + \frac{\xi x}{\beta + \xi u}\right)^{-1/\xi} \\
&= G_{\xi,\beta(u)}(x),
\end{aligned}
$$

where $\beta(u) := \beta + \xi u$, $0 \leq x < \infty$ if $\xi > 0$, and $0 \leq x \leq -(\beta/\xi) - u$ if $\xi < 0$.

For the exponential distribution, the excess distribution is independent of the threshold u. This result is illustrated in figure 3.2.

The mean of the GPD exists if $\xi < 1$ holds. It is given by

$$E(X) = \frac{\beta}{1 - \xi}.$$

As we have shown, the mean excess distribution of the GPD is also a GPD with a modified scaling that grows linearly with the threshold u. Hence we have the following mean excess function for the GPD:

$$e(u) = \frac{\beta + \xi u}{1 - \xi}.$$

This linearity property is illustrated in figure 3.3. It is crucial for checking the appropriateness of the GPD model for threshold exceedances. The GPD model should only be applied to data exceeding a threshold u for which the mean excess grows linearly.

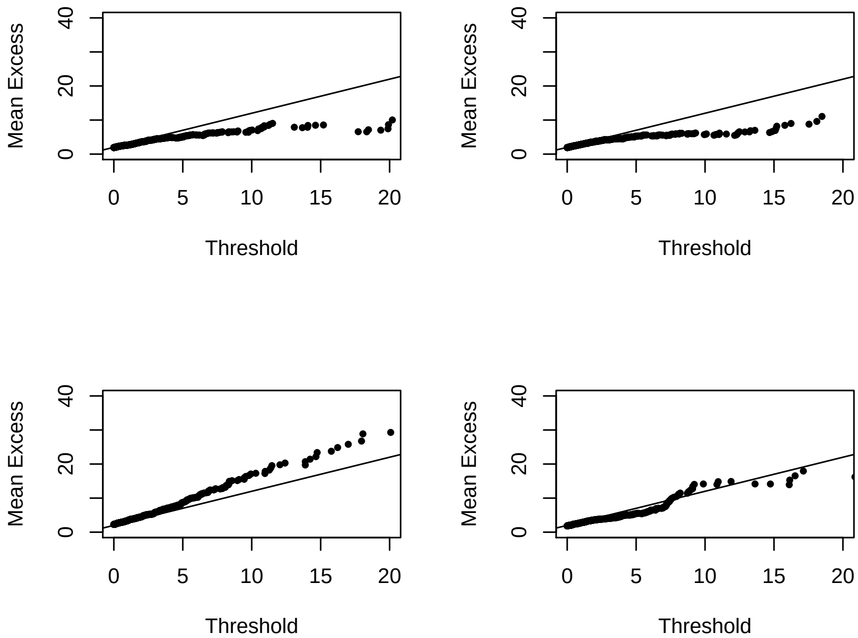


Figure 3.3.: Mean excess plots for 4 samples (size $n = 1000$) drawn from the GPD with $\xi = 0.5$ and $\beta = 1$. The mean excesses fluctuate randomly around the straight line with intercept $\beta/(1-\xi) = 2$ and slope $\xi/(1-\xi) = 1$.

3.2. Point processes

A point process is a random measure N that specifies the number of points $N(A)$ located in a compact set A which is part of the state space $\mathcal{X}$ of the process.[1] Therefore it is a method to randomly allocate points to intervals on the real line ($\mathbb{R}$) or hyper–rectangles in d–dimensional Euclidean space ($\mathbb{R}^d$).

According to Daley and Vere-Jones (2003),

> [...] a point process is completely defined if the joint probability distributions are known for the number of events in all finite families of disjoint intervals (or rectangles etc.).

In case of the one–dimensional point process, the real line is often interpreted as the time axis. Hence a point process provides the opportunity to model the occurrences of events in time. Before we display these point process models we present some basic definitions from survival analysis.

3.2.1. Basic definitions and relations from survival analysis

Definition 6 (Survivor function)**.** *The* survivor function *is the (unconditional) probability of surviving time t, i.e. the probability that no event occurs up to t.*

$$S(t) = P(T > t) \tag{3.4}$$

When T is a continuous random variable, the survivor function is the complement of the cumulative distribution function $F(t)$, hence

$$S(t) = 1 - F(t) = \int_t^\infty f(x)dx, \tag{3.5}$$

where $f(t)$ denotes the density corresponding to $F(t)$. Thus

$$\frac{dS(t)}{dt} = -f(t). \tag{3.6}$$

Definition 7 (Hazard function)**.** *The hazard function is the given by probability of an event ("death") within the interval $[t, t + \Delta t]$ given survival up to t, normalized by the width Δt of the interval:*

$$h(t) = \frac{P(t \leq T \leq t + \Delta t \mid T \geq t)}{\Delta t}. \tag{3.7}$$

Note that the hazard or intensity function is *not* a probability and can take values above 1. The expression $\lim_{\Delta t \to \infty} h(t)$ is called hazard or intensity as well. Particularly in presence of a constant hazard, it is often represented by the letter λ.

Definition 8 (Cumulative hazard)**.** *Let $h(t)$ be some hazard function. The corresponding* cumulative hazard *is given by*

[1] For comprehensive treatments see Karr (1991) and the two volumes Daley and Vere-Jones (2003) and Daley and Vere-Jones (2008) which are the major references for the following presentation. Vere–Jones also contributed prominent articles on seismological applications of marked point process models (Vere-Jones (1970), Vere-Jones and Ozaki (1982)).

$$H(t) = \int_0^t h(x)dx. \tag{3.8}$$

An important relation of these quantities is given by

$$h(t) = \frac{f(t)}{S(t)}. \tag{3.9}$$

3.2.2. One–dimensional point processes

A point process on the line is a natural model for occurrences of events in time. It can equivalently be described by a non–decreasing integer–valued step function, by the sequence of its points $\{t_i\} = t_0, t_1, t_2, \ldots$, or by the sequence of intervals (durations) between successive events $\{\tau_i\} = \tau_1, \tau_2, \ldots$, where $\tau_i = t_i - t_{i-1}$.

Let A be an interval on the real line and $N(A)$ the number of events within A. Writing A_i for indexed partitions of the state space which are mutually disjoint we have

$$N\left(\bigcup_{i=1}^{r} A_i\right) = \sum_{i=1}^{r} N(A_i). \tag{3.10}$$

Assuming that the process starts in $t = 0$, we use the following equivalent notations.

$$N(t) = N(0, t] = N((0, t]), \qquad (0 < t \leq \infty).$$

If the distribution of the number of points lying in an interval does not depend on the interval's location on the time axis but only on its length, we call the point process *stationary*. Similarly, we can establish such a stationarity property for the operational time series of durations between events τ_i also.

Formally we can give the following definition.

Definition 9 (Stationarity and duration stationarity)**.** *A point process is* stationary *when for every* $r = 1, 2, \ldots$ *and all bounded Borel subsets* $A_1, \ldots, A_r$ *of the real line, the joint distribution of*

$$\{N(A_1 + t), \ldots, N(A_r + t)\}$$

does not depend on t $(-\infty < t < \infty)$*.*

It is duration stationary *when for every* $r = 1, 2, \ldots$ *and all integers* $i_i, \ldots, i_r$ *the joint distribution of* $\{\tau_{i_1+k}, \ldots, \tau_{i_r+k}\}$ *does not depend on* k $(k = 0, \pm 1, \ldots)$*.*

According to *Khinchin's Existence Theorem*, the hazard rate for a stationary process exists. When it is finite, we can rewrite the definition of the hazard (3.7) as

$$\begin{aligned} P(N(t, t+h] > 0) &= P(\text{there occurs at least one event in (t,t+h]}) \\ &= \lambda h + o(h) \qquad \text{when } h \downarrow 0. \end{aligned}$$

Moreover, a point process is called *simple* if

$$P(N(\{t\}) = 0 \text{ or } 1 \; \forall \; t),$$

i.e. if events do not occur simultaneously. It is *regular* if — roughly speaking — the probability of events is a continuous function of the size of the evaluated interval.

A generalization to higher dimensions is straightforwardly achieved. Therefore, the Poisson process is assumed to take its values in a complete separable metric space (c.s.m.s.) $\mathcal{X}$ — usually d–dimensional Euclidean space. Instead of points in the interval (one–dimensional case), the probability of $N(A) = n$ points falling into the bounded set A is modelled via the parameter measure $\Lambda(\cdot)$ of the process.

In principle, the exceedances could enter the model as a second dimension of the point process, but the following analysis centers around the concept of *marked point processes* (MPP) where the actual point process only consists of event times. The marks (exceedances) can be seen as properties of the events, determining the "family" the events are from.

3.2.3. Marked processes and cluster processes

To account for possible interconnections between durations and exceedances, the latter can be included as marks of a marked point process. Times and marks in principle constitute processes in their own right. The locations $\{x_i\}$ where (in our special case: $\{t_i\}$ and "when") the events occur is also called the *ground process*, denoted by N_g. The formal definition of a marked point process can be written as follows.

Definition 10 (Marked point process).
A marked point process (MPP) *with locations in the c.s.m.s.* $\mathcal{X}$ *and marks in the c.s.m.s.* $\mathcal{K}$ *is a point process* $\{(x_i, \kappa_i)\}$ *on* $\mathcal{X} \times \mathcal{K}$ *with the additional property that the ground process* $N_g(\cdot)$ *is itself a point process; i.e. for bounded* $A \in \mathcal{B}_{\mathcal{X}}$ *we have* $N(A \times \mathcal{K}) < \infty$.

The concepts of simplicity (no more than one event per point in time) and stationarity can also be defined for MPPs.

Definition 11 (Simplicity and stationarity of MPPs).

(a) The MPP N *is* simple *if the ground process* N_g *is simple.*

(b) The MPP N *on* $\mathcal{X}$ *is* stationary *(or* homogeneous*) if the probability structure of the process is invariant under shifts in* $\mathcal{X}$.

Definition 12 (Independent marks and unpredictable marks). *Let the MPP* $N = \{(x_i, \kappa_i)\}$ *on the product space* $\mathcal{X} \times \mathcal{K}$ *be given.*

(a) N *has* independent marks *if, given* N_g*, the marks are mutually independent random variables such that the distribution of* κ_i *depends only on the corresponding location* x_i.

(b) For $\mathcal{X} = \mathbb{R}$, N *has* unpredictable marks *if the distribution of the mark at* x_i *is independent of locations and marks* $\{(x_j, \kappa_j)\}$ *for which* $x_j < x_i$.

Hence, in our case independent marks would be present if the size of exceedances was independent of the size of adjacent exceedances, which is a questionable supposition. Because we can observe a clustering effect of returns exceeding a certain threshold, the sizes are likely to be clustered, i.e. highly autocorrelated, as well.

Unpredictable marks imply in our context that the history of times and sizes of exceedances does not make any contribution to explain later realizations of exceedances.

The notion of cluster processes provides a different interpretation of MPPs: Points of cluster processes can be used to describe the locations of individuals from consecutive generations of a branching process.

This offers in deviation from efficient market theory (EMT) an economic interpretation as well: The market's reaction to exogenous shocks ("news arrival") may give birth to further reactions of the market. Market participants might not be completely aware of the reactions of the remainder of the market, consequently the market might exhibit some momentum without news arrival, a phenomenon one might call "endogeneous news".

Such cluster processes can be best understood by the separation of two components: the locations of clusters on the one hand and the locations of cluster members within its particular cluster on the other hand. The superposition of the latter constitutes the observed process, while the intrinsic dissociation of the process into centre process and "within–cluster processes" can not be observed. The cluster elements constitute point processes $N(\cdot \mid y_i)$ indexed by the cluster centres $\{y_i\}$. The cluster locations itselves can be described by the process N_c of cluster centres whose realization consists of the points $\{y_i\} \subset \mathcal{Y}$. Usually both processes are defined on the same state space, i.e. $\mathcal{Y} = \mathcal{X}$. Within the branching process interpretation, the centres might be seen as germs or ancestors for the clusters they generate. However, the observed process does not include information which events might be ancestors and which might be offspring.

The formal definition of a cluster process is the following.

Definition 13 (Cluster process). *N is a* cluster process *on the completely separable metric space $\mathcal{X}$ with centre process N_c on the c.s.m.s. $\mathcal{Y}$ and component processes $N(\cdot \mid y) : y \in \mathcal{Y}$ associated with one of the centres, when for every bounded $A \in \mathcal{B}_{\mathcal{X}}$*

$$N(A) = \int_{\mathcal{Y}} N(A \mid y) N_c(dy) = \sum_{y_i \in N_c(\cdot)} N(A \mid y_i) < \infty \text{ almost surely.} \tag{3.11}$$

One special class of cluster processes is the *Poisson cluster process.* We speak of a Poisson cluster process when the cluster centres are the points of a Poisson process.

Note that the suggestive word "cluster" should not be taken too literally: let mark space and centre space be indentical, points belonging to the same family within the mark space $\mathcal{K} = \mathcal{Y}$, i.e. belonging to the same cluster centres, are not necessarily neighbours within the primary state space $\mathcal{X}$ (the time axis). However, if such a connection between locations within $\mathcal{K}$ and $\mathcal{X}$ exists, it can be modelled in the MPP framework.

The Hawkes process

The Hawkes process is a so–called *self–exciting process* or *infectivity model.* The points $\{x_i\}$ of a Hawkes process are of two types: "immigrants" without extant parents in the process which can be seen as triggering events in the finance context, and "offspring" that are "breeded" by existing points. The arrival of immigrants occurs at (constant) rate τ, while the offspring arise as elements of a finite Poisson process that is associated with some point already constructed. This interpretation is obviously closely linked to the Poisson special case of definition 13 above. Any point of the process has the potential to give birth to further points whose locations are those of a (finite) Poisson process.

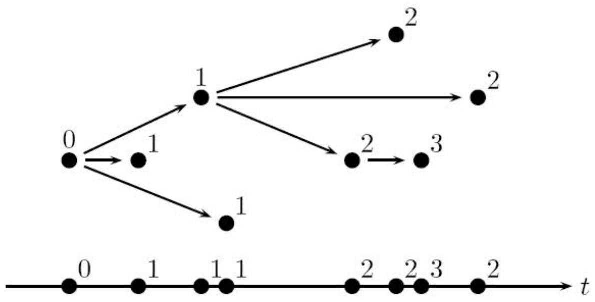


Figure 3.4.: The branching structure of the Hawkes process (top) and the observed process on the time axis (bottom). Source: Møller and Rasmussen (2005).

The processes are assumed to be mutually independent, points within one cluster (i.e. given the triggering event) to be identically distributed and independent of the immigrant process as well. Hence, each immigrant has the potential to produce descendants whose numbers in successive generations constitute a so–called *Galton–Watson branching process* with Poisson offspring distribution. When the branching process is of finite total size, it is also called *sub–critical*. Hence, clusters are defined as immigrants (= ancestors) plus their descendants. A Hawkes process is constituted by the entirety of these clusters.

Besides its cluster process representation, the process can also be defined by its intensity function that is given by

$$\lambda^*(t) = \tau + \sum_{0<t_i<t} \eta(t - t_i) = \tau + \int_0^t \eta(t-u)N(du).$$

This representation is very illustrative as it can be divided into an "immigrant part" τ and the time–varying "offspring part" in the sum. Usually, the function η is descending $(\partial\eta(t-t_i)/\partial(t-t_i) < 0)$ in order to capture the clustering phenomenon mentioned above — "young" events should be more likely to produce further events than "old" ones. The precise functional forms chosen for our applications are discussed in later sections.

The marked Hawkes process

We now consider a marked Hawkes process with unpredictable marks. According to the branching process interpretation of Hawkes processes, events can be characterized as either "immigrants" or "descendants" of these immigrants. Within the *marked* point process framework, the marks κ_i associated with the times x_i can be interpreted as the "type" of the individual event, i.e. in our case its degree of extremeness.

In the context of cluster processes, we denoted the sequence of cluster centres by $\{y_i\}$. In Hawkes processes, the cluster centres are the "immigrants" who also act as ancestors of further members of their family, the family being characterized by their marks. They are assumed to arrive according to a compound homogeneous Poisson process $N(dy \times d\kappa)$ with rate μ_c and fixed mark distribution $F(d\kappa)$. Immigrants as well as descendants have the potential to act as ancestors of first–generation offspring; they constitute an ordinary Poisson process with rate depending on the ancestor's mark and time elapsed since the ancestor event. The marks of the offspring form an i.i.d. sequence with distribution function $F(d\kappa)$, hence the family members share this d.f. as "surname".

The ETAS model

Ogata (1988) introduced the ETAS (Epidemic Type After–Shock) model to earthquake modelling. It can be interpreted as a special case of the marked Hawkes process with occurrence times $x_i = t_i$ and magnitudes of the shocks κ_i. The intensity is a decreasing function of time elapsed since the last event and increasing with respect to the mark size or "magnitude". More precisely, due to seismological considerations the intensity follows the Gutenberg–Richter law of magnitudes and the empirical Omori law of event clustering mentioned in the introduction.

The *Gutenberg–Richter (GR) Law* implies exponentially distributed magnitudes, an assumption that is close to the newer GPD model for threshold exceedances. The Omori law ceteris paribus postulates an hyperbolic decay with respect to time elapsed.

3.2.4. Conditional intensities & likelihoods

General relations

To construct a likelihood in presence of a realization $(x_1, \ldots, x_n)$ of points (or events), we require the joint probability density of the x_i . Such a likelihood is considered as a function of the parameters defining the joint density, whereas the points x_i and its number n are assumed to be given.

The most important building blocks of point process likelihoods are the so–called *Janossy densities*.[2] Roughly speaking, these densities are used to describe the probability of specific allocations of events in the state space $\mathcal{X}$ (usually $\mathcal{X} = \mathbb{R}^d$), or possibly in some subset A of the state space.

Let $j_n(x_1, \ldots, x_n)$ be the density of the Janossy measure $J_n(\cdot)$ with $x_i \neq x_j$ for $i \neq j$ (only one point of the process per location). The probability of one specific allocation is given by

$$j_n(x_1, \ldots, x_n) dx_1 \cdots dx_n = P \begin{pmatrix} \text{there are exactly } n \text{ points in the} \\ \text{process, one in each of the } n \text{ distinct} \\ \text{infinitesimal small regions } (x_i, x_i + dx_i) \end{pmatrix}.$$

When and if interest centers around a bounded subset A of the entire state space $\mathcal{X}$, e.g. a certain time interval, the following definition is particularly useful.

Definition 14 (Local Janossy measures and densities)**.** *Given any bounded Borel set A, the* Janossy measures *localized to A are the measures $J_n(\cdot \mid A)$, $n = 1, 2, \ldots$, with locations $x_i \in A$, $i = 1, \ldots, n$, for which*

$$J_n(dx_1 \times \cdots \times dx_n) = P \begin{pmatrix} \textit{there are exactly } n \textit{ points in } A \\ \textit{at locations } dx_1, \ldots, dx_n \end{pmatrix}. \quad (3.12)$$

The densities of these measures are the local Janossy densities*.*

Under some regularity conditions fulfilled by the point process N, the likelihood of its realizations can be defined.

[2]The Janossy measure is named in Daley and Vere-Jones (2003) after Hungarian physicist Lajos Janossy who introduced it to model particle showers. For mathematical details see Daley and Vere-Jones (2003), chapter 5 (Janossy measures) and chapter 7 (conditional intensities and likelihoods).

Definition 15 (Point process likelihood)**.** *The* likelihood *of a realization $x_1, \ldots, x_n$ of a (regular) point process N on a bounded Borel set $A \subseteq \mathbb{R}^d$ where $N(A) = n$ is the local Janossy density*

$$L_A(x_1, \ldots, x_n) = j_n(x_1, \ldots, x_n \mid A). \tag{3.13}$$

If the neighbourhood A is identical with the entire state space $\mathcal{X}$, the notation can be further simplified by dropping the subscript A.

The one–dimensional case

According to Daley and Vere-Jones (2003), the evaluation of such likelihoods on general state spaces is difficult. Since we are interested in the particular state space $\mathcal{X} = \mathbb{R}_+$ — the time axis — there is an alternative approach to obtain tractable likelihood functions: the use of *conditional* intensity functions. The period under observation can be denoted by $A = [0, T]$, therefore we have a point process on state space $\mathbb{R}_+$. The ordered set of occurrence times of extreme events is denoted by $\{t_1, \ldots, t_{N(T)}\}$, and the t_i as well as the durations $\tau_i = t_i - t_{i-1}$ between events are assumed to be well–defined rvs. The (Janossy) density for the location of n events on the interval $(0, w)$ is denoted by $j_n(t_1, \ldots, t_n \mid w)$, the corresponding local Janossy df is denoted by $J_0(w) = J_0((0, w))$.

Furthermore we introduce *conditional survivor functions* by defining

$$S_k(w \mid t_1, \ldots, t_{k-1}) := P(\tau_k > w \mid t_1, \ldots, t_{k-1}), \tag{3.14}$$

i.e. the probability that the kth duration exceeds w given all former event times ($\in A$). The densities corresponding to the survivor functions $S_k(\cdot \mid \cdot)$ are denoted by $p_k(\cdot \mid \cdot)$.

Proposition 1 (Conditional survivor function and conditional densities)**.** *Under some regularity conditions, for a point process on $\mathcal{X} = \mathbb{R}_+$ there exists a uniquely determined family of conditional density functions $p_n(t \mid t_1, \ldots, t_{n-1})$ and corresponding survivor functions*

$$S_n(t \mid t_1, \ldots, t_{j-1}) = 1 - \int_{t_{n-1}}^{t} p(u \mid t_1, \ldots, t_{n-1}) du \qquad (t > t_{n-1}) \tag{3.15}$$

defined on $0 < t_1 < \cdots < t_{n-1} < t$ such that each $p_n(\cdot \mid t_1, \ldots, t_{n-1})$ has support carried by (t_{n-1}, ∞).

For $n \geq 1$ and finite $[0, T]$ with $T > 0$, the Janossy densities for the locations of $t_1, \ldots, t_{n-1}$ can be obtained recursively:

$$J_0(T) = S_1(T) \tag{3.16}$$

$$\begin{aligned} j_n(t_1, \ldots, t_n \mid T) &\equiv j_n(t_1, \ldots, t_n \mid (0, T)) \\ &= p_1(t_1) p_2(t_2 \mid t_1) \cdots p_n(t_n \mid t_1, \ldots, t_{n-1}) \times S_{n+1}(T \mid t_1, \ldots, t_n). \end{aligned} \tag{3.17}$$

The *conditional hazard function* can be obtained using (3.9) and (3.15).

$$h_n(t \mid t_1, \ldots, t_{n-1}) = \frac{p_n(t \mid t_1, \ldots, t_{n-1})}{S_n(t \mid t_1, \ldots, t_{n-1})}. \tag{3.18}$$

As $S_n(t \mid \cdot) = \exp(-H_n(t \mid \cdot))$, we can write

$$p_n(t \mid t_1, \ldots, t_{n-1}) = h_n(t \mid t_1, \ldots, t_{n-1}) \exp\left(-\int_{t_{n-1}}^{t} h(u \mid t_1, \ldots, t_{n-1}) du\right). \quad (3.19)$$

On the basis of (3.18), we are able to define a conditional intensity function given some sequence $\{t_i\}$ with $0 < t_1 < \cdots < t_n < \cdots$. Usually the terms "hazard" and "intensity" are used synonymously, but within the following definition the *conditional intensity function* is a hybrid consisting of the unconditional hazard in $(0, t_1]$ and the conditional hazard known from (3.18) for $(t_1, \infty]$.

Definition 16 (Conditional intensity function)**.** *The conditional intensity function* $\lambda^*(\cdot)$ *for a (regular) point process on* $\mathcal{X} = \mathbb{R}_+ = [0, \infty)$ *is defined by*

$$\lambda^*(t) = \begin{cases} h_1(t) & \text{for } 0 < t \leq t_1, \\ h_n(t \mid t_1, \ldots, t_{n-1}) & \text{for } t_{n-1} < t \leq t_n,\ n \geq 2. \end{cases} \quad (3.20)$$

Note that in this definition conditioning is based on the past of the process. More formally speaking: its sequence of σ–algebras (its filtration/ information set) $\mathcal{H}_{t-}$ consists only of the occurrence times $\{t_1, t_2, \ldots, t_{n-1}\}$ up to time t (but not including t). In principle the filtration can also include covariate information. We are going to exploit this fact in the next chapter by taking exceedances as covariates.

Proposition 2 (Likelihood of a self–exciting point process)**.** *Let* N *be a (regular) point process on* $[0, T]$ *for some finite positive* T*, and let* $t_1, \ldots, t_{N(T)}$ *be a realization of that process in* $[0, T]$*. Then, the likelihood* L *of the process can be expressed by*

$$L = \left[\prod_{i=1}^{N(T)} \lambda^*(t_i)\right] \exp\left(-\int_0^T \lambda^*(u) du\right). \quad (3.21)$$

Proof. To see this, we use (3.17), (3.19), and (3.25). We can write the likelihood as

$$L = j_{N(T)}(t_1, \ldots, t_{N(T)} \mid (0, T),\ N(T)).$$

Since we have a realization of known size $N(T)$, the last term of (3.17) disappears via conditioning and we obtain

$$\begin{aligned} L &= p_1(t_1) p_2(t_2 \mid t_1) \cdots p_{N(t)}(t_{N(T)}, \ldots, t_1) \\ &= h_1(t_1) \exp\left(-\int_0^{t_1} h(u) du\right) \cdot h_2(t_2 \mid t_1) \exp\left(-\int_{t_1}^{t_2} h(u \mid t_1) du\right) \\ &\quad \cdots \times h_{N(T)}(t_{N(T)})) \mid t_{N(T)-1}, \ldots, t_1) \exp\left(-\int_{t_{N(T)-1}}^{t_{N(T)}} h(u \mid t_{N(T)-1}, \ldots, t_1) du\right) \\ &= \left[h_1(t_1) h_2(t_2 | t_1) \cdots h_{N(T)}(t_{N(T)} | t_{N(T)-1}, \ldots, t_1)\right] \exp\left(-\sum_{i=1}^{N(T)} \int_{t_{i-1}}^{t_i} \lambda^*(u) du\right) \\ &= \left[\prod_{i=1}^{N(T)} \lambda^*(t_i)\right] \exp\left(-\int_0^T \lambda^*(u) du\right), \end{aligned}$$

where $t_{N(T)} = T$ and $t_0 = 0$.

$\square$

Using this result, the expansion to MPPs is straightforward. The state space of MPPs on the line is given by $[0, \infty) \times \mathcal{K}$, the corresponding Janossy density can be written

$$j_n(x_1, \dots, x_n, \kappa_1, \dots, \kappa_n \mid A \times \mathcal{K}). \tag{3.22}$$

The recursion of (3.17) can be expanded and we obtain

$$\begin{aligned} J_0(T) &= S_1(T) && (3.23) \\ j_1(t_1, \kappa_1 \mid T) &= p_1(t_1, \kappa_1) = p_1(t_1) f_1(\kappa_1 \mid t_1) \\ j_2(t_1, t_2, \kappa_1, \kappa_2 \mid T) &= p_1(t_1) f_1(\kappa_1 \mid t_1) p_2(t_2 \mid (t_1, \kappa_1)) f_2(\kappa_2 \mid (t_1, \kappa_1), t_2)) \\ &\vdots && (3.24) \end{aligned}$$

The history $\mathcal{H}$ of the MPP consists of pairs of event locations and the associated marks. The expanded conditional intensity function can be defined as follows.

Definition 17 (Conditional intensity function for MPPs)**.** *The conditional intensity function $\lambda^*(\cdot)$ for a (regular) marked point process on $\mathcal{X} = \mathbb{R}_+ = [0, \infty) \times \mathcal{K}$ is defined by*

$$\lambda^*(t) = \begin{cases} h_1(t) f_1(\kappa \mid t) & \textit{for } 0 < t \le t_1, \\ \quad \vdots & \\ h_n(t \mid (t_1, \kappa_1), \dots, (t_{n-1}, \kappa_{n-1})) & \\ \times f_n(\kappa \mid (t_1, \kappa_1), \dots, (t_{n-1}, \kappa_{n-1}), t) & \textit{for } t_{n-1} < t \le t_n,\ n \ge 2 \\ \quad \vdots & \end{cases} \tag{3.25}$$

We furthermore introduce the more convenient notation

$$\lambda^*(t, \kappa) = \lambda_g^*(t) f^*(\kappa | t), \tag{3.26}$$

where the right–hand side abbreviates the product of the hazard *conditional* (when $n \ge 2$) on the history of the MPP, and the mark density *conditional* on that same history and t. The first term is also called the intensity of the *ground process.*

Marked point process likelihood

Proposition 3 (Likelihood of a self–exciting marked point process)**.** *Let N be a (regular) MPP on $[0, T] \times \mathcal{K}$ for some finite positive T, and let $t_1, \dots, t_{N(T)}$ be a realization of that process in $[0, T]$. Then, the likelihood L of the process can be expressed by*

$$L = \left[\prod_{i=1}^{N_g(T)} \lambda_g^*(t_i) \right] \left[\prod_{i=1}^{N_g(T)} f^*(\kappa_i \mid t_i) \right] \exp\left(- \int_0^T \lambda_g^*(u) du \right). \tag{3.27}$$

Proof. Marks are conditionally independent of the associated ground process. Therefore, the product of mark densities simply has to be multiplied with the likelihood of the ground process that has already been derived (proof of proposition 2). □

3.3. Generalizing the POT model

Time series data of threshold exceedances comprise two components, times and sizes of events. As we have shown, in a marked point process model with one–dimensional point process the components can be represented by (conditionally) independent terms in the log–likelihood.

3.3.1. The basic POT model

Within the basic peaks over thresholds (POT) model, events follow a homogeneous Poisson process, hence the intensity is a constant and durations between events are iid. The excess sizes above the threshold are iid generalized Pareto, particularly they are independent of their locations in time.

In this simple case, the associated likelihood can be obtained even by a two–dimensional point process approach (McNeil, Frey, and Embrechts (2005)). However, to be consistent we present the MPP version of the likelihood for all models.

Using the general MPP likelihood from (3.27) and the conditional intensity from (3.25) and (3.26), we can write the log–likelihood of the basic POT model as

$$l(\lambda, \beta, \xi) = \underbrace{\sum_{i=1}^{N_T} \log(\lambda_g^*(t_i)) - \int_0^T \lambda_g^*(u)\ du}_{\text{"times"}} + \underbrace{\sum_{i=1}^{N_T} \log f(\kappa_i\ |t_i)}_{\text{"marks"}}$$

and using the (mutual and time–) independence of both event times and marks we obtain

$$= \sum_{i=1}^{N_T} \log \tau - \int_0^T \tau\ du + \sum_{i=1}^{N_T} \log f(\kappa_i),$$

where $f(\kappa_i)$ is the density of the GPD. Hence

$$= n \log \tau - T\tau - n \log \beta - (1 + 1/\xi) \sum_{i=1}^{N_T} \log(1 + \xi\kappa_i/\beta). \qquad (3.28)$$

3.3.2. The self–exciting POT model

We derive the likelihood of several versions of the self–exciting POT model, namely

- model without and with influence of excess sizes on the intensity,
- model without and with time–varying GPD scale parameter β.

The core of all these models is the Hawkes–type conditional intensity function of the ground process given by

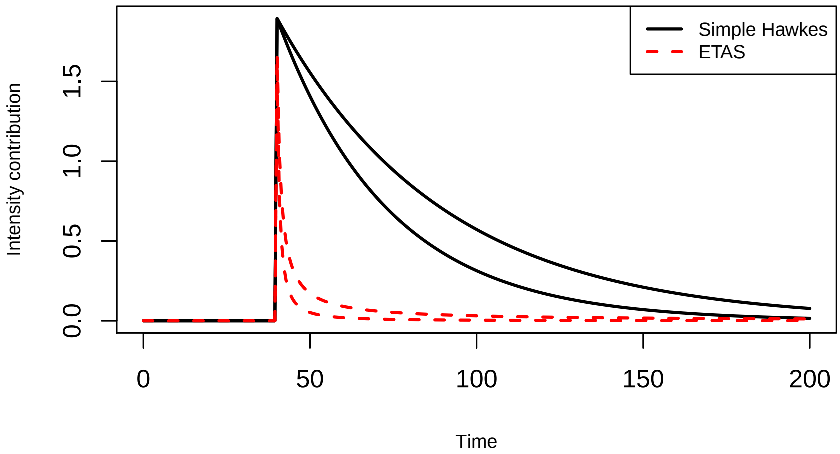


Figure 3.5.: Contribution of a single event at $t = 40$ to the intensity for the simple Hawkes (straight lines) and the ETAS model (dashed lines) and several choices of the parameters. Note that the parameter values are not comparable due to the different functional forms of the models (exponential vs. hyperbolic decay).

$$\begin{aligned}\lambda_g^*(t) &= \tau + \psi \sum_{j:0<t_j<t} \eta(t - t_j, \kappa_j) \\ &= \tau + \psi v^*(t) \end{aligned} \tag{3.29}$$

where $v^*(t)$ is a convenient notation of $v^*(t \mid (t_1, \kappa_1), \ldots)$. We use two versions of $\eta(\cdot)$,

- $\eta(s, \kappa) = (1 + \delta\kappa) \exp(-\gamma s)$ where s denotes the time elapsed since an event. We refer to this version as the "simple Hawkes model". Furthermore we have
- $\eta(s, \kappa) = (1 + \delta\kappa)(1 + \frac{s}{\gamma})^{-(1+\rho)}$. This version is referred to as the "ETAS model" (Epidemic Type After–Shock model).

The function $\eta(\cdot)$ is decreasing in the elapsed time since the last event $s = t - t_j$. Ceteris paribus older events have less weight than younger ones. The effect of the excess sizes $\kappa_j = x_j - u$ on $\eta(\cdot)$ and hence on the intensity remains subject of the empirical evaluation, i.e. we do not constrain the responsible parameter(s) (we do not put a priori information into the analysis). Moreover, it is possible to test for the presence of an effect of mark sizes on the intensity ("mark influence") using likelihood ratio tests.

However, the specific functional forms of $\eta(\cdot)$ will be further discussed in the next chapter on estimation and empirical results. It is largely dictated by the feasability of numerical optimization routines.[3]

[3] Problems with the numerical optimization of likelihoods even based on simple versions of Hawkes–type conditional intensities are well known, see Daley and Vere-Jones (2003), p. 234. See also appendix A.

Model with unpredictable marks

In order to obtain likelihood functions for our models, we start by considering the generalized since self–exciting times model while still assuming iid GP distributed marks. We can write the logarithmized version of (3.27) as

$$l(\lambda, \beta, \xi, \ldots) = \underbrace{\sum_{i=1}^{N_T} \log(\lambda_g^*(t_i)) - \int_0^T \lambda_g^*(u)\ du}_{\text{"times"}} + \underbrace{\sum_{i=1}^{N_T} \log f(\kappa_i\ |t_i)}_{\text{"marks"}},$$

hence

$$= \sum_{i=1}^{N_T} \log\left(\tau + \psi \sum_{j:0<t_j<t_i} \eta(t_i - t_j, \kappa_j)\right) - \int_0^T \lambda_g^*(u)\ du + \sum_{i=1}^{N_T} \log f(\kappa_i\ |\mathcal{H}_{t_i-})$$

Note that *all* events preceding each t_i have some influence on $\lambda_g^*(t_i)$, i.e. the intensity process has infinite memory. This property is therefore reproduced in the exceedance process.

The second term — the log survivor function at T equalling the area under the intensity — can be obtained analytically for the simple Hawkes model and our version of the ETAS model.

$$\begin{aligned}\int_0^T \lambda_g^*(s)\ ds &= \int_0^T \tau + \psi \sum_{i=1}^{n} \eta(s - t_i, \kappa_i)\ ds \\ &= \tau T + \psi \sum_{i=1}^{n} \int_0^T \eta(s - t_i, \kappa_i)\ ds.\end{aligned}$$

For one single arbitrary event $(t_i, \kappa_i),\ s > t_i$ with $\eta(\cdot)$ of *simple Hawkes type* we have

$$\begin{aligned}\int_0^T (1 + \delta\kappa_i) \exp(-\gamma(s - t_i))\ ds &= (1 + \delta\kappa_i) \int_0^T \exp(-\gamma(s - t_i))\ ds \\ &= (1 + \delta\kappa_i) \left[\frac{1 - \exp(-\gamma(s - t_i))}{\gamma}\right]_0^T \\ &= (1 + \delta\kappa_i) \frac{1 - \exp(-\gamma(T - t_i))}{\gamma},\end{aligned}$$

where the constant of integration is chosen such that the log survivor function for $t = 0$ equals 0. Similarly we obtain for this single arbitrary event and $\eta(\cdot)$ taken from the *ETAS model*

$$\begin{aligned}(1 + \delta\kappa_i) \int_0^T \left(1 + \frac{s - t_i}{\gamma}\right)^{-(\rho+1)} ds &= (1 + \delta\kappa_i) \left[\frac{\gamma(1 - (1 + (s - t_i)/\gamma)^{-\rho})}{\rho}\right]_0^T \\ &= (1 + \delta\kappa_i) \frac{\gamma\left(1 - \left(1 + \frac{T - t_i}{\gamma}\right)^{-\rho}\right)}{\rho}\end{aligned}$$

as can be easily verified.

Altogether, the second term of the log–likelihood can be written

$$\int_0^T \lambda_g^*(s)\; ds = \begin{cases} \tau T + \psi \sum_i (1+\delta\kappa_i)\frac{1-\exp(-\gamma(T-t_i))}{\gamma} & \text{for the simple Hawkes model,} \\ \tau T + \psi \sum_i (1+\delta\kappa_i)\frac{\gamma\left(1-\left(1+\frac{T-t_i}{\gamma}\right)^{-\rho}\right)}{\rho} & \text{for the ETAS model.} \end{cases} \tag{3.30}$$

The history of the process $\mathcal{H}_{t_i-}$ up to but not including t_i in the last term is irrelevant when we assume iid GP distributed marks.

Model with predictable marks

We now turn to the case of *predictable marks* where the mark distribution depends on the history of the intensity process also. Recalling the conditional intensity function from (3.26), the difference to the model with unpredictable marks lies only in the mark distribution — the ground model described by $\lambda_g^*(t)$ remains the same.

Precisely, the scale parameter of the GPD is enriched by a time–dependent term. In times of high intensity, the scale parameter β is thus able to grow or shrink. Thereto, it is triggered by some additional parameter α leading to the conditional intensity

$$\begin{aligned} \lambda^*(t,\kappa) &= \lambda_g^*(t) f^*(\kappa|t) \\ &= \underbrace{(\tau + \psi v^*(t))}_{\text{ground model intensity}} \underbrace{\frac{1}{\beta + \alpha v^*(t)}\left(1+\xi\frac{x-u}{\beta+\alpha v^*(t)}\right)^{-1/\xi-1}}_{\text{mark density}}. \end{aligned} \tag{3.31}$$

3.4. Measures of risk

3.4.1. Market risk

The risk that the value of a financial instrument will fluctuate as a result of changes in market prices is called *market risk*. These changes are either caused by factors specific to the individual instrument or its issuer, or by factors affecting all securities traded in the market. Depending on the specific asset class market risk may include e.g. equity risk, commodity risk, interest rate risk, or currency risk. In many situations quantile based measures of the loss distribution are used as risk measures, such as Value-at-Risk (VaR) or expected shortfall (ES).

It has been recalled by the current financial crisis that these statistic measures can only predict what history has shown already. They usually lose their value for risk that arises from single events such as takeovers, sudden changes of market conditions (regime shifts) or liquidity supply.

3.4.2. Value–at–Risk

Value–at–Risk (VaR) is likely to be the most common risk measure, since it is crucial within the Basel II capital–adequacy framework. It is usually easy to calculate, and it can

model name	modelling of times	modelling of marks	parameters	intensity
basic peaks over threshold (POT)	homogeneous Poisson process, i.e. exponentially distributed durations, "lack of memory"	iid GPD	τ, β, ξ	$\lambda^*(t,\kappa) = \lambda^*(\kappa) = \frac{\tau}{\beta}\left(1 + \xi\frac{\kappa}{\beta}\right)^{-1/\xi-1}$
basic self-exciting model	intensity depends on elapsed time since last event *and* mark sizes	not modelled	$\tau, \phi, \gamma, \delta, \rho$	$\lambda^*(t) = \tau + \phi v^*(t)$
self-exciting POT model (with unpredictable marks)	intensity depends on elapsed time since last event *and* (potentially) mark sizes	iid GPD	$\tau, \phi, \gamma, \delta, \rho, \beta, \xi$	$\lambda^*(t,\kappa) = \frac{\tau + \psi v^*(t)}{\beta}\left(1 + \xi\frac{\kappa}{\beta}\right)^{-1/\xi-1}$
self-exciting POT model (with predictable marks)	intensity depends on elapsed time since last event *and* (potentially) mark sizes	conditional GPD, dependent on t	$\tau, \phi, \gamma, \delta, \rho, \beta, \xi, \alpha$	$\lambda^*(t,\kappa) = \frac{\tau + \psi v^*(t)}{\beta + \alpha v^*(t)}\left(1 + \xi\frac{\kappa}{\beta + \alpha v^*(t)}\right)^{-1/\xi-1}$

Table 3.1.: Overview of point process models. The time–dependence of the intensity is determined by $v^*(t) = \sum_{j:0<T_j<t} \eta(t - t_j, \kappa)$ which is a function of the time elapsed since the event in t_j and the excess size κ, while $\eta(\cdot)$ includes the parameters γ, ρ, δ. The parameter ρ is only present if $\eta(\cdot)$ is of ETAS type, δ triggers the potential influence of exceedance sizes ("marks") on the intensity function. A possible autoregressive structure of the scale of the GPD is captured by α. Hence, the "times" model exhibits up to five parameters $(\tau, \phi, \gamma, (\delta), (\rho))$, the "marks" model two or three $(\beta, \xi, (\alpha))$.

be interpreted easily, making it accessible to corporate executives without much statistical training.

Definition 18 (Value–at–Risk (VaR))**.** *Given some confidence level* $\alpha \in (0,1)$*, the* $1-\alpha$ Value–at–Risk *of an asset (or portfolio of assets) S is the* $1-\alpha$ *quantile of the corresponding loss distribution, formally*

$$VaR_{\alpha} = \inf\{l \in \mathbb{R} : P(L > l) \leq 1-\alpha\} = \inf\{l \in \mathbb{R} : F_L(l) \geq \alpha\},$$

where the rv L denotes the losses.

Losses can be defined either in terms of absolute values or in terms of changes ("returns") of the value of a portfolio. Since we use daily returns over long time horizons, we confine ourselves to distributions of (positive and negative) log returns, i.e. *relative losses.*

VaR is not a coherent risk measure. The definition of coherence includes several axiomatic requirements postulated in Artzner, Delbain, Eber, and Heath (1999) that a reasonable risk measure should fulfil. VaR is not coherent because it is not subadditive which means that for two financial instruments S_1 and S_2 it is possible that $\text{VaR}(S_1 + S_2) > \text{VaR}(S_1) + \text{VaR}(S_2)$ holds, i.e. merging two portfolios can create extra risk.

A favourable since coherent alternative is the Expected Shortfall (ES) which is the expected loss given that the loss is exceeding a certain level. However, ES is mostly difficult to obtain because the entire tail of the loss distribution has to be known (or estimated at least). In case of the self–exciting POT model, the loss distribution comprehends the intensity of the exceedance process as well as the excess distribution. Thus, already the computation of the conditional VaR is precarious, obtaining the ES is even more challenging.

3.4.3. Conditional Value–at–Risk

The conditional VaR is a quantile of the predicted loss distribution. According to the models discussed above, there are three types of predicted distributions that are relevant for VaR estimation, namely

- basic POT: events follow a homogeneous Poisson process with iid generalized Pareto distributed exceedances;
- self–exciting POT with unpredictable marks: the probability of events is time–dependent, whereas the exceedances retain the iid GPD assumption;
- self–exciting POT with predictable marks: event times as well as exceedance sizes are time dependent.

Let κ_t denote an exceedance at time t and $G_{\kappa_{t+\Delta t}|\mathcal{H}_t}$ the predictive df using information up to t to predict the loss distribution Δt days ahead.[4] The q–quantile of this distribution and hence VaR_q is then given by

$$\kappa_t^q = \inf\{\kappa \in \mathbb{R} : \; G_{\kappa_{t+\Delta t}|\mathcal{H}_t} \geq q\}. \tag{3.32}$$

[4]Note that κ denotes the rv as well as its realization in order to retain the notation from chapter 3.

Because of the conditional independence of times and marks, the probability of an exceedance higher than a certain level and Δt days ahead can be written

$$P(\kappa_{t+\Delta t} > \kappa \mid \mathcal{H}_t) = P(\kappa_{t+\Delta t} - u > \kappa - u \mid \kappa_{t+\Delta t} - u, \mathcal{H}_t) \tag{3.33}$$
$$\times P(\kappa_{t+\Delta t} > u \mid \mathcal{H}_t) \tag{3.34}$$

The second term on the right–hand side is determined by the conditional intensity — event or no event —, while the first term comprises the size of the possible exceedance which is assumed to follow a GPD.

Let us now assume $\Delta t = 1$. The second term can only be approximated since the continuous time intensity model in contrast to the daily data allows for more than one event per day. Neglecting the probabilities for $2, 3, \ldots$ events we write

$$\begin{aligned} P(N(t, t+1) = 1 \mid \mathcal{H}_t) &\approx 1 - P(N(t, t+1) = 0 \mid \mathcal{H}_t) \\ &= 1 - S(t+1 \mid \mathcal{H}_t) \\ &= 1 - \exp\left(-\int_t^{t+1} \lambda^*_{\mathcal{H}_t}(s) ds\right). \end{aligned} \tag{3.35}$$

Thus, it is possible to estimate the Value–at–Risk for the self–exciting POT models, which is done in chapter 4.

4 Empirical results

4.1. The data

In this thesis, we consider point process models for *univariate* financial time series only. However, in order to obtain meaningful results, we perform our analyses using several datasets of sufficient length. We consider daily closing prices of the DJIA index and five stocks that have been traded at the NYSE over the whole period under investigation: General Electric (GE), IBM, Coca Cola (CO), Boeing (BO), and General Motors (GM), with data ranging from 1962-01-03 to 2009-03-27.[1] Although the DJIA dataset is actually longer, estimation results presented in this chapter are based on the same period. Some descriptive statistics of the data are shown in table 4.1. The Ljung–Box statistics for raw and absolute returns and 10 lags indicate that all series exhibit the autocorrelation structure typical for financial returns as described above.

Series	Mean	Std.Dev.	Skewness	Exc.Kurt.	Ljung-Box(10)	Ljung-Box(10), abs. ret.
DJIA	0.0002	0.0103	-1.3327	37.2374	47.9990	6397.6999
GE	-0.0002	0.0237	-18.5071	681.5657	8.9214	707.6842
IBM	-0.0002	0.0239	-22.8697	1121.0437	7.8137	331.9966
CO	-0.0001	0.0239	-19.4231	676.0397	15.1179	292.4246
BO	-0.0000	0.0249	-6.6211	161.0556	25.5071	537.5002
GM	-0.0002	0.0222	-2.9076	103.8712	80.7144	7699.4933

Table 4.1.: Descriptive statistics of univariate return series

4.2. Estimation and model choice

Figure 4.1 shows mean excess plots for the six datasets under investigation. At least in the cases of GE, Coca Cola, and Boeing, the linearity property and hence the GPD assumption appears to be reasonable for thresholds above 0.05, if at all.

In this section, estimation results and model selection are summarized, whereas detailed tables with all estimates, standard errors, maximum log likelihoods etc. can be found in the appendices B and C. Generally, estimation of self–exciting POT models is difficult and

[1] All datasets are available from `http://finance.yahoo.com`.

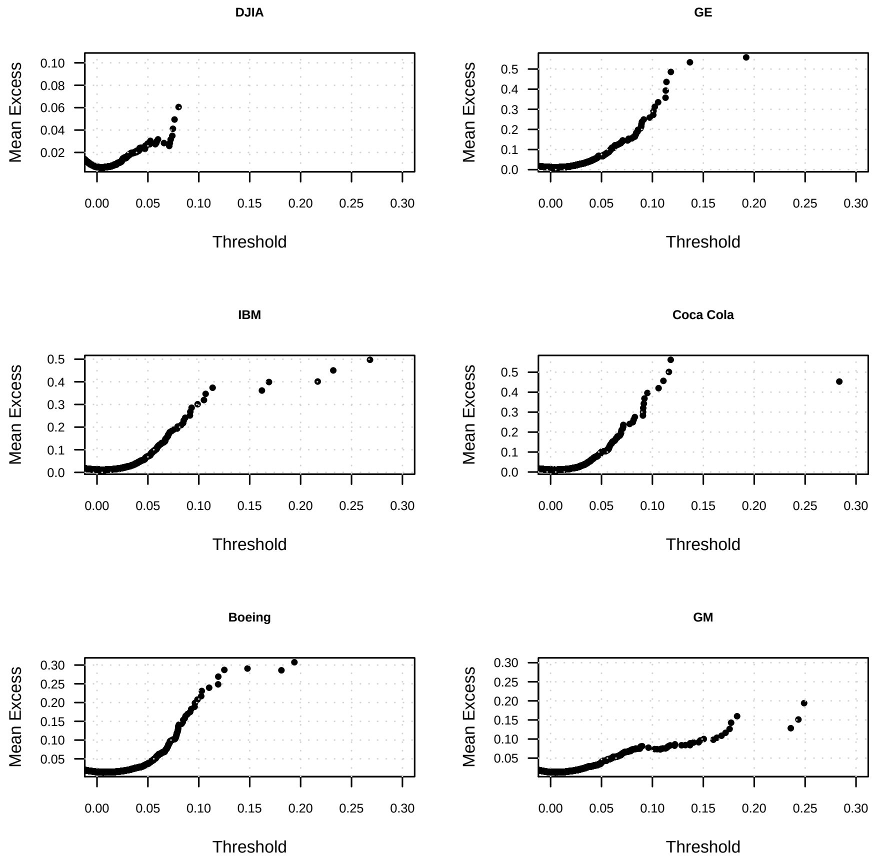


Figure 4.1.: Mean excess plots for the six datasets. In case of the DJIA data, the plot appears to be linear above a relatively low threshold of about 0.01. The fact that the index exhibits less volatility is well–known from portfolio theory. GM and GE seem to be piecewise linear. For Boeing, GM and Coca Cola, the mean excesses start to be linear at a relatively high threshold of roughly 0.05 which questions the applicability of the GPD model for lower thresholds in these cases.

often averted by numerical problems (for details see appendix A on optimization routines). Thus, results presented here are based on those models which could be estimated. It is possible that some ”feasibility bias” caused by the numerical problems arises, i.e. estimation results may not be ”missing at random”. However, since there are no obvious indiosyncratic properties of the underlying data which are responsible for the estimation problems, it still may be legitimate to assume that the obtained results are passably representative for stock returns in general.

For the six univariate datasets described above the model constellations summarized in table 4.2 have been estimated. All versions of the ETAS model (hyperbolic intensity decay) could only be estimated for the Dow Jones data, possibly due to the fact that an index might exhibit some idiosyncrasies that are favourable in this context, e.g. lower volatility because of diversification effects. Therefore results and conclusions presented are almost entirely based on the simple Hawkes–type models (with exponential decay of the conditional intensity).

self–excitement type	mark influence	predictability	# parameters
Hawkes	no	no	3 (times) + 2 (marks) = 5
Hawkes	yes	no	4 (times) + 2 (marks) = 6
Hawkes	no	yes	3 (times) + 3 (marks) = 6
Hawkes	yes	yes	4 (times) + 3 (marks) = 7
ETAS	no	no	4 (times) + 2 (marks) = 6
ETAS	yes	no	5 (times) + 2 (marks) = 7
ETAS	no	yes	4 (times) + 3 (marks) = 7
ETAS	yes	yes	5 (times) + 3 (marks) = 8

Table 4.2.: Parametrizations of self–exciting POT models.

The models are estimated for both loss and gain processes, i.e. for the point processes generated by threshold exceedances of both negative and positive log returns. We consider three thresholds which are determined by the data (0.90, 0.95 and 0.99 quantile of losses / gains of each dataset) plus three arbitrary thresholds (returns of 0.01, 0.02, 0.03) to achieve two different kinds of comparability.

The choice of an appropriate model implies the answers to the following questions:

1. **Temporal homogeneity of extreme events:** Do the conditional intensity models perform significantly better than the homogeneous Poisson model?

2. **Influence of marks on the intensity:** Does the size of the exceedances have an effect on the probability of further exceedances in the near future?

3. **Predictability of marks:** Are exceedances time–dependent? Does the conditional GPD model perform significantly better than the iid GPD model?

4. **ETAS**: Does the ETAS model outperform the simpler parameterization of the intensity function?

In order to answer these questions, we use the Akaike information criterion given by

$$AIC = 2 \times \#\text{ parameters} - 2 \times \text{ maximum log likelihood},$$

which penalizes the addition of parameters and rewards improvement of the maximum of the likelihood and has to be minimized by the model of choice.

Furthermore, we back these results by employing likelihood ratio (LR) tests. The LR statistic is given by

$$\mathrm{LR} = 2(l(\hat{\theta}) - l(\tilde{\theta})) = 2 \log \frac{L(\hat{\theta})}{L(\tilde{\theta})}, \tag{4.1}$$

where $\tilde{\theta}$ denotes the ML estimator under H_0 (restricted model), and $\hat{\theta}$ signifies the MLE under the alternative. A convenient result says that as the sample size n approaches ∞, the test statistic LR for a nested model will be asymptotically χ^2 distributed with degrees of freedom equal to the difference in dimensionality of the models under the null and alternative hypotheses, respectively. We use this result in order to test for model appropriateness.

The times model

The model for the temporal part of the point process can be chosen separately because it is independent of the associated mark distribution. In contrary the scale of the time–dependent marks model is dependent on the times model. If and when there is an influence of exceedance sizes (the only covariate in our model) or possibly further covariates on the probability of events, this might in turn cause a feedback to the mark distribution and so on. The consequence of this property in terms of model choice is that one has to choose the times model first to be able to choose the associated marks model. More precisely: If mark sizes affect the intensity of the process, this might also affect the mark sizes, which is a feature typical for complex systems and also attractive in terms of economic interpretations.

Tables 4.3 and 4.4 depict the AICs for loss and gain processes, in each case without (upper panel) and with (centre panel) mark influence.[2] The lower panel shows the difference between the two. Empty cells indicate that the respective model could not be estimated due to numerical difficulties. Since the AIC should be minimized in presence of a ”good model”, negative differences indicate that exceedances seem not to influence the intensity, positive differences suggest that they do. In case of the loss exceedance processes, evidence in favour of the presence of a mark influence is strong for low thresholds for all datasets, while exceedance sizes over high thresholds do not appear to have such an influence. For DJIA index returns, the 1 percent most extreme events (implying a threshold of less than 0.03 in contrast to the other five series) do have an influence on the intensity, marks exceeding 0.03 do not. For the other series, the 0.99 quantile implies a much higher threshold in absolute terms, and no mark influence is diagnosed. This may lead to the conclusion that the presence of some mark influence is up to the *absolute* level of thresholds, or even more roughly: extremeness might be attributed in absolute terms. On the other hand we may conclude that, given a certain level of extremeness, the exact amount of extremeness is no more relevant for further events to occur.

In remarkable contrast, the degree of extremeness of gains affects the probability of further extreme gains *irrespective of the level of the threshold.* Hence we seem to have

[2]For the corresponding likelihood ratio tests — which essentially yield the same results — see appendix C.

uncovered an asymmetry between gains and losses. The estimate of the parameter δ determining the mark influence is positive in all cases: the higher the excess sizes, the higher the probability of further events in the near future (parameter estimates can be found in appendix B).

dataset	0.01	0.02	0.03	0.90 quantile	0.95 quantile	0.99 quantile
DJIA	8926.79	2405.85	808.13	8283.49	4769.76	1194.63
GE	14340.15	6463.76	2824.22	8343.82	4763.15	1174.47
IBM	14606.17	6716.08	3098.61	8399.86		
CO	13846.49	6112.96	2702.73	8427.35	4889.09	1349.05
BO	17152.91	9993.26	5435.93	8482.76	4939.32	
GM	15798.67	8297.28	4151.75	8297.28	4628.18	1109.64
DJIA	8924.04	2404.57	810.10	8281.30	4767.06	1193.98
GE	14323.46	6460.96	2826.22	8341.78	4762.40	1176.47
IBM	14597.63	6716.60	3100.50	8398.51		
CO	13835.59	6112.85	2704.57	8421.98	4889.47	1351.05
BO	17144.25	9989.96	5435.96	8479.06	4940.99	
GM	15792.06	8292.69	4152.71	8292.69	4628.64	1111.64
DJIA	2.76	1.28	-1.97	2.19	2.70	0.65
GE	16.70	2.79	-2.00	2.04	0.76	-2.00
IBM	8.53	-0.52	-1.89	1.36		
CO	10.90	0.11	-1.84	5.37	-0.37	-2.00
BO	8.66	3.30	-0.03	3.70	-1.67	
GM	6.61	4.60	-0.96	4.60	-0.46	-2.00

Table 4.3.: AIC for the times models without (upper panel) and with (central panel) mark influence and difference between the two for negative log returns and several thresholds.

The marks model

Since exceedance sizes seem to affect the conditional intensity, we compare the performance of marks models based on a mark–influenced version of the times model. The competing marks or exceedance models are the iid GPD model vs. the GPD model with time–dependent scale parameter β. Recall from (3.31) that the conditional mark density $f^*(\kappa|t)$ enters the conditional intensity of the MPP such that

$$\begin{aligned}\lambda^*(t,\kappa) &= \lambda_g^*(t) f^*(\kappa|t)\\ &= \underbrace{(\tau+\psi v^*(t))}_{\text{ground model intensity}} \underbrace{\frac{1}{\beta+\alpha v^*(t)}\left(1+\xi\frac{x-u}{\beta+\alpha v^*(t)}\right)^{-1/\xi-1}}_{\text{mark density}}.\end{aligned}$$

The expression $\beta + \alpha v^*(t)$ represents the time–dependent scale of the GPD, where α in practice is always positive though not restricted, and $v^*(t)$ fulfilling $v^*(t) > 0 \; \forall \; t \in \mathcal{X}$ comprises the sum of all past influences on the present intensity. As we have chosen the

dataset	0.01	0.02	0.03	0.90 quantile	0.95 quantile	0.99 quantile
DJIA	9329.74	2681.31	947.46	8398.32	4857.17	1223.26
GE		7205.54	3341.12	8437.45	4879.75	
IBM				8457.86	4915.47	
CO		6998.37	3275.32	8554.76	4996.92	1311.09
BO				8580.67	5018.83	1372.80
GM	15653.83	8671.16	4560.46	8335.58	4873.38	1163.17
DJIA	9314.85	2655.47	941.84	8382.65	4836.46	1213.74
GE		7186.78	3326.66	8414.72	4860.92	
IBM				8422.98	4894.78	
CO		6979.20	3252.19	8533.14	4973.43	1307.43
BO				8568.73	5007.85	1372.66
GM	15621.83	8661.31	4551.61	8325.57	4861.55	1154.78
DJIA	14.89	25.85	5.62	15.67	20.71	9.53
GE		18.76	14.46	22.73	18.83	
IBM				34.88	20.68	
CO		19.17	23.14	21.62	23.49	3.65
BO				11.95	10.98	0.15
GM	32.01	9.85	8.85	10.01	11.83	8.40

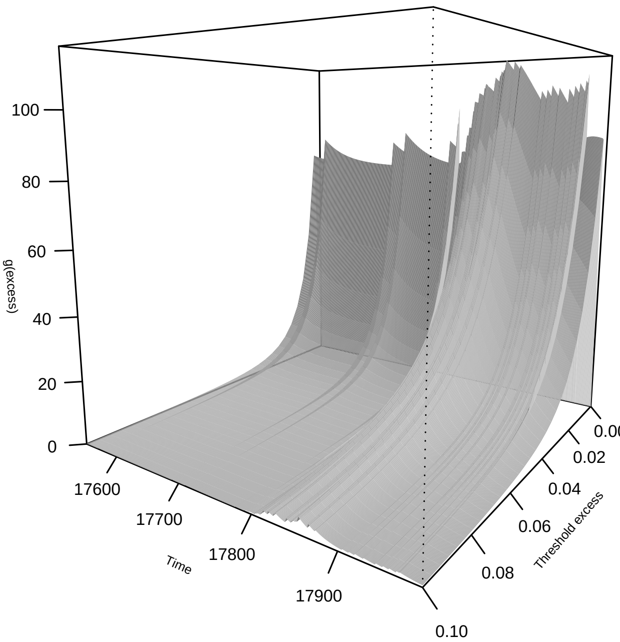
Table 4.4.: AIC for the times models without (upper panel) and with (central panel) mark influence and difference between the two for positive log returns and several thresholds.

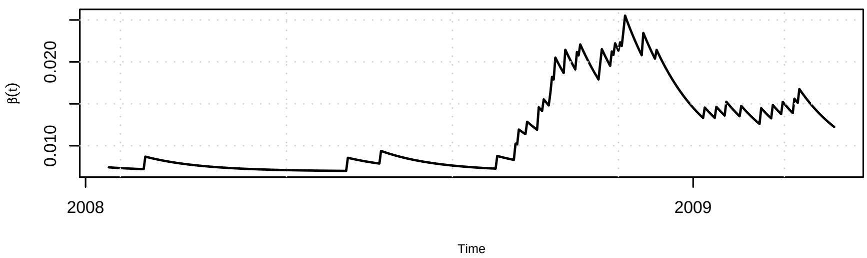
times model *with* mark influence we retract the notational simplification introduced above to write

$$v^*(t) = v^*(t, \kappa),$$

emphasizing the mark dependence of the conditional GPD model. Within the self–exciting POT model with predictable marks (conditional GPD) extreme events cause the occurrence of further events via an increasing intensity, which in turn makes the events likely to be more extreme by expanding the scale of the mark distribution. This connection is illustrated in figure 4.2 for the DJIA data.

We again examine the necessity of such time dependence using Akaike's information criterion. The results can be found in tables 4.5 for loss exceedances and 4.6 for gain exceedances. The AIC differences appearing in the lower panels of the tables are mostly large, unambiguously indicating that the mark distribution is time dependent for each of the considered thresholds.





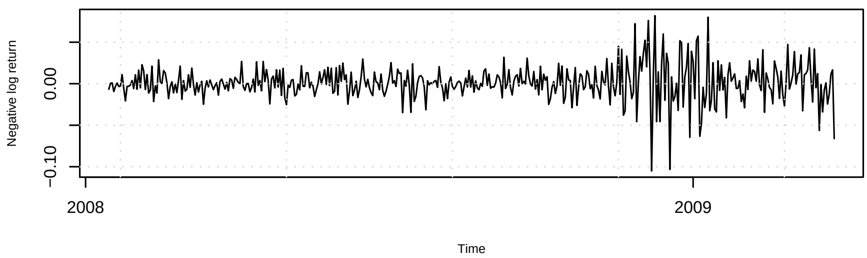


Figure 4.2.: Implied exceedance distributions from 2008–01–01 to 2009–03–27 according to the simple Hawkes model with mark influence for DJIA data (upper panel). The GPD's scaling (centre panel) evidently reacts strongly to the volatility rise in the fourth quarter of 2008 (lower panel).

dataset	0.01	0.02	0.03	0.90 quantile	0.95 quantile	0.99 quantile
DJIA	-10374.67	-1988.37	-476.13	-9422.67	-4609.12	-782.79
GE	-17479.73	-5786.13	-1925.75	-8134.36	-3814.33	-588.23
IBM	-17971.32	-6022.26	-2124.59	-8161.57		
CO	-16829.64	-5391.87	-1727.63	-8318.18	-3993.22	-604.68
BO	-21154.90	-9623.35	-4188.11	-7649.20	-3661.10	
GM	-19217.69	-7642.45	-3085.99	-7642.06	-3604.93	-536.69
DJIA	-10579.37	-2014.02	-485.09	-9598.37	-4714.41	-792.76
GE	-17771.74	-5911.13	-1954.78	-8299.20	-3887.48	-592.00
IBM	-17989.75	-6095.23	-2131.73	-8254.87		
CO	-16842.28	-5407.31	-1730.04	-8364.82	-4000.96	-604.62
BO	-21339.02	-9702.10	-4194.57	-7702.27	-3665.44	
GM	-19713.84	-7902.42	-3198.14	-7902.01	-3724.72	-556.22
DJIA	204.70	25.65	8.96	175.71	105.29	9.98
GE	292.01	125.00	29.04	164.84	73.15	3.77
IBM	18.44	72.96	7.15	93.30		
CO	12.64	15.45	2.41	46.64	7.74	-0.06
BO	184.12	78.74	6.45	53.06	4.34	
GM	496.15	259.97	112.14	259.95	119.79	19.53

Table 4.5.: AIC for the marks models without (upper panel) and with (centre panel) predictable mark distribution and difference between the two (lower panel) for negative log returns and several thresholds.

dataset	0.01	0.02	0.03	0.90 quantile	0.95 quantile	0.99 quantile
DJIA	-10946.63	-2192.75	-604.06	-9463.72	-4602.40	-828.58
GE		-6741.48	-2503.73	-8346.62	-4054.99	
IBM				-8239.97	-4003.25	
CO		-6398.92	-2412.98	-8362.74	-4128.82	-778.19
BO				-7675.01	-3745.52	-725.28
GM	-18602.83	-8158.55	-3342.91	-7705.59	-3631.30	-583.33
DJIA	-11085.12	-2208.11	-606.19	-9593.81	-4665.72	-829.07
GE		-6856.08	-2544.74	-8471.30	-4116.02	
IBM				-8308.03	-4025.24	
CO		-6457.55	-2422.53	-8441.40	-4157.03	-778.38
BO				-7715.32	-3751.66	-726.89
GM	-18629.24	-8336.16	-3476.25	-7876.38	-3767.74	-612.30
DJIA	138.49	15.36	2.13	130.08	63.31	0.49
GE		114.59	41.01	124.68	61.03	
IBM				68.06	21.98	
CO		58.63	9.55	78.66	28.21	0.19
BO				40.31	6.14	1.61
GM	26.41	177.61	133.34	170.79	136.44	28.97

Table 4.6.: AIC for the marks models without (upper panel) and with (centre panel) predictable mark distribution and difference between the two (lower panel) for positive log returns and several thresholds.

4.3. Goodness of fit

The model of choice identified by likelihood ratio tests and AIC is the simple Hawkes model with mark influence and time–dependent mark distribution. It should be underlined that this model may be seen as the best among the existing models because it shows the best global fit. However, this does not mean that no better model is possible.

Besides the relative comparison of models we therefore have to assess the absolute goodness of fit of the chosen model. To be able to appraise the quality of the times component of our model of choice, we employ residual analysis methods proposed by Ogata (1988) which are very similar to the Cox–Snell residuals of Cox's proportional hazards regression model.

Generally speaking, in presence of a "good fit" residuals should not comprise any deterministic pattern, because this pattern should have been identified and extracted by the model. However, while in regression models the construction of appropriate residuals is straightforward in general — difference between response and fit at the observed covariate values —, it is not that obvious in terms of point processes. The model output is essentially an estimated intensity process, i.e. a realization of a stochastic process in continuous time, that implies probabilities of event occurrences within certain time intervals. The data (of the times component) consist of discrete locations of events on the time axis only.

Ogata (1988) proposes the construction of yet another point process, the *residual process*. Its aim is to compare the estimated conditional intensity function of the ground process $\hat{\lambda}_g^*(\cdot)$ to its true intensity process $\lambda_g^*(\cdot)$ whose outcome are the locations of events in time. Once again we consider the sequence of observed event times $t_1, t_2, \ldots, t_{N(T)}$ generated by this true intensity process on $[0, T]$. According to Papangelou (1972), the points of the cumulative intensity process

$$\Lambda(t_i) = \int_0^{t_i} \lambda_g^*(s) ds \tag{4.2}$$

for $i = 1, \ldots, N(T)$ constitute a homogeneous Poisson process of rate 1 on an interval $[0, N(T)]$ which hence is part of a *transformed time axis*. This is intuitive since — roughly speaking — there should occur one event per unit of area under the intensity function on average, i.e. one event per unit of time on the transformed time axis. The resulting property of exponentially distributed durations enables us to test for the presence of a homogeneous Poisson process via a Kolmogorov–Smirnov test.

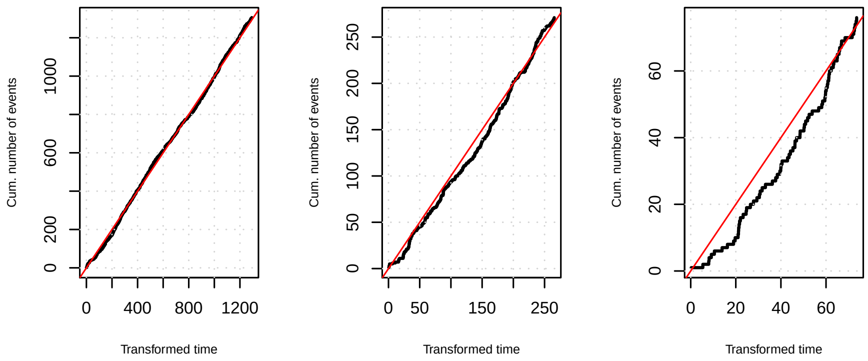


Figure 4.3.: Plots of the residual processes for DJIA's self–exciting POT models with the thresholds 0.01, 0.02, and 0.03. The p–values of Kolmogorov–Smirnov tests for the presence of a homogeneous Poisson process (with iid exponentially distributed inter–event times) are 0.007, 0.7114, and 0.8767, respectively. Hence, for the two higher thresholds the self–exciting POT model appears to be appropriate, for the lowest threshold it is clearly not since the homogeneous Poisson hypothesis is rejected.

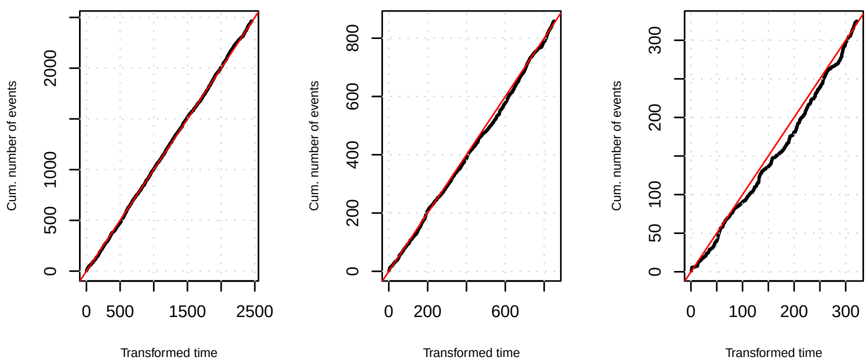


Figure 4.4.: Plots of the residual processes for GE's self–exciting POT models with the thresholds 0.01, 0.02, and 0.03. The p–values of Kolmogorov–Smirnov tests for the presence of a homogeneous Poisson process (with iid exponentially distributed inter–event times) are 0, 0.0538, and 0.8699, respectively. Hence, for the highest threshold the self–exciting POT model appears to be appropriate, for the lowest threshold it is clearly not since the homogeneous Poisson hypothesis is rejected. Evidence for $u = 0.02$ is not clear.

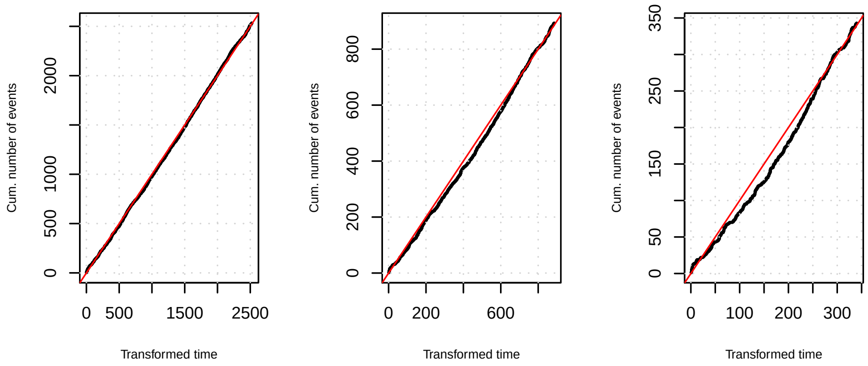


Figure 4.5.: Plots of the residual processes for IBM's self–exciting POT models with the thresholds 0.01, 0.02, and 0.03. The p–values of Kolmogorov–Smirnov tests for the presence of a homogeneous Poisson process (with iid exponentially distributed inter–event times) are 0, 0.1017, and 0.8803, respectively. Hence, for the highest threshold the self–exciting POT model appears to be appropriate, for both lower thresholds it is not since the homogeneous Poisson hypothesis is rejected.

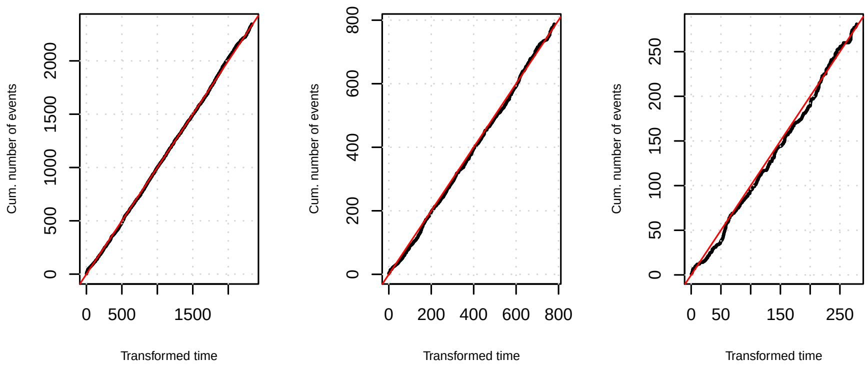


Figure 4.6.: Plots of the residual processes for Coca Cola's self–exciting POT models with the thresholds 0.01, 0.02, and 0.03. The p–values of Kolmogorov–Smirnov tests for the presence of a homogeneous Poisson process (with iid exponentially distributed inter–event times) are 0, 0.6372, and 0.7921, respectively. Hence, for the highest threshold the self–exciting POT model appears to be appropriate, for the lower thresholds it is clearly not since the homogeneous Poisson hypothesis is rejected.

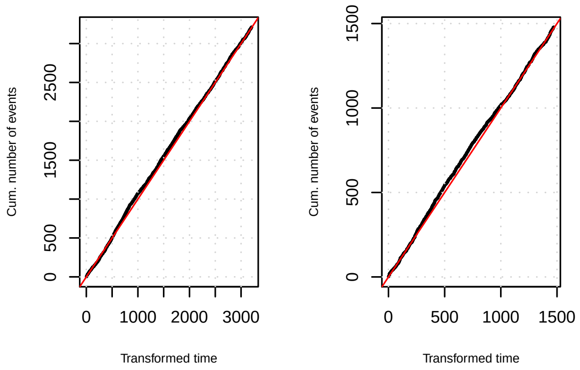


Figure 4.7.: Plots of the residual processes for Boeing's self–exciting POT models with the thresholds 0.01, 0.02, and 0.03. The p–values of Kolmogorov–Smirnov tests for the presence of a homogeneous Poisson process (with iid exponentially distributed inter–event times) are 0, 8e-04, and 0.5371, respectively. Hence, for the two higher thresholds the self–exciting POT model appears to be appropriate, for the lowest threshold it is clearly not since the homogeneous Poisson hypothesis is rejected.

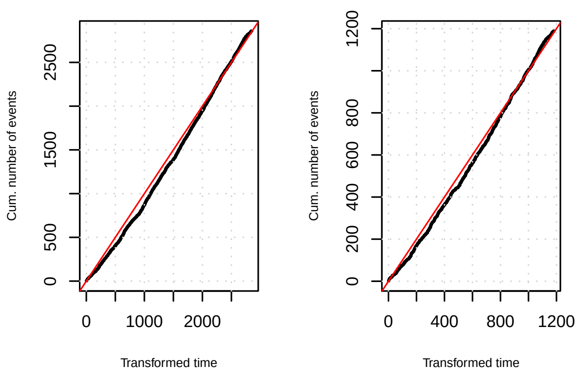


Figure 4.8.: Plots of the residual processes for GM's self–exciting POT models with the thresholds 0.01, 0.02, and 0.03. The p–values of Kolmogorov–Smirnov tests for the presence of a homogeneous Poisson process (with iid exponentially distributed inter–event times) are 0, 0.0148, and 0.6743, respectively. Hence, for the two higher thresholds the self–exciting POT model appears to be appropriate, for the lowest threshold it is clearly not since the homogeneous Poisson hypothesis is rejected.

4.4. Estimation of risk measures

The empirical examination of the conditional Value–at–Risk proposed by Chavez-Demoulin, Davison, and McNeil (2005) and explained in section 3.4 requires just a few words. Negative returns and corresponding in–sample 0.99 VaR one–step–ahead predictions based on negative returns exceeding 1% are depicted in figure 4.9 for the six datasets. Visual inspection suggests that the conditional excess distribution's faster reaction to increasing and decreasing volatility may lead to a better performance compared to the unconditional marks model. This would also be in accordance with the suggestions in section 4.2. However, both models perform very poorly, in case of the DJIA data the conditional GPD model is even worse than the unconditional model because the VaR is underestimated when volatility is declining. The percentage of violations which is meant to be about one percent ranges between about 4 up to about 12 (!) percent.

The assessment of the model performance in terms of interval forecasts as proposed in Christoffersen (1998) is neither possible for small nor for large (up to eight years) moving windows due to the numerical problems already mentioned.

Since even the in–sample performance is rather poor, the forecasting ability can not be expected to be better anyway. Estimation of the models using data until 2008–06 only, i.e. before the last substantial rise of volatility which can be observed for all time series considered, does not change the parameter estimates and VaR forecasts heavily.

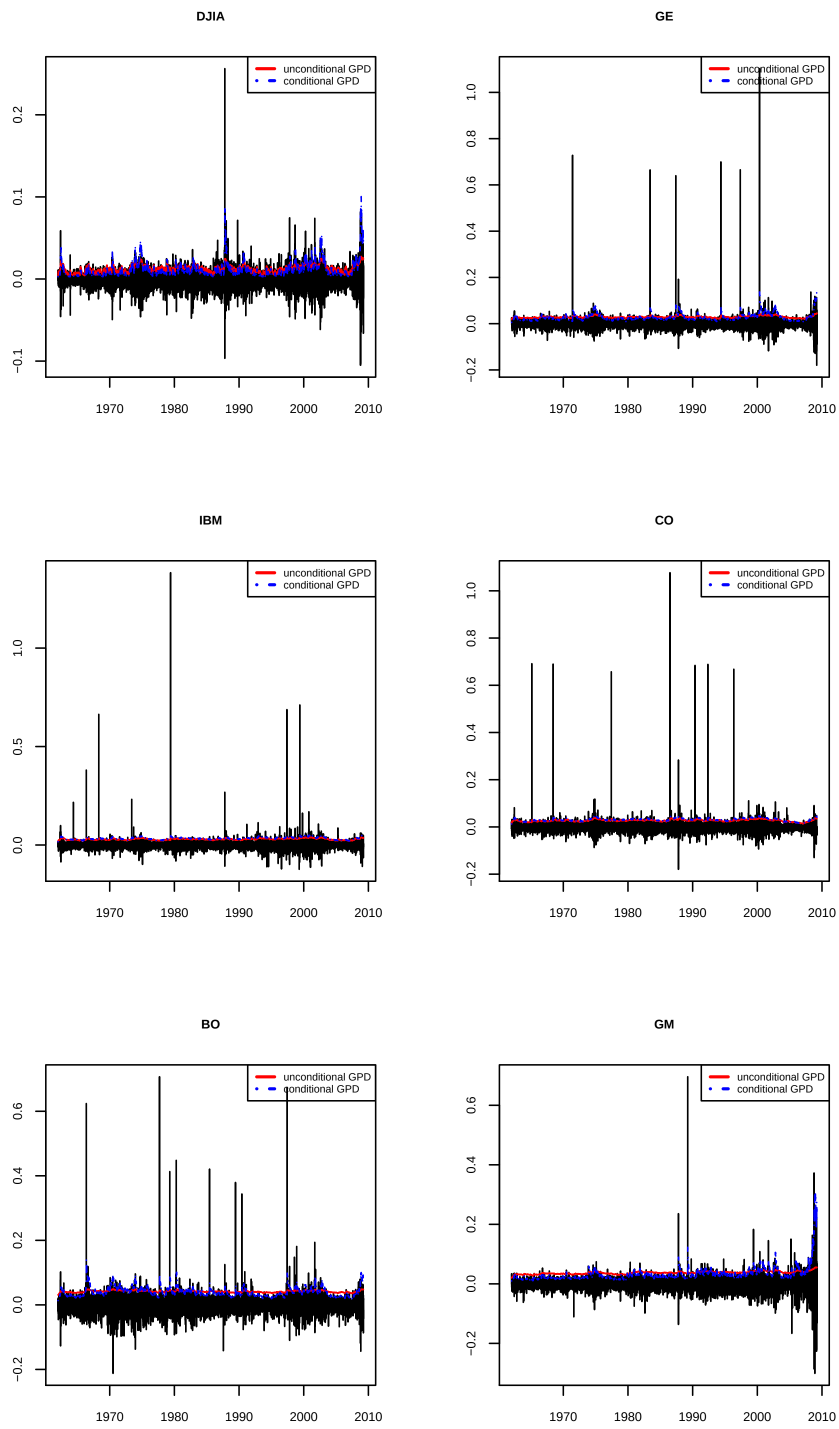


Figure 4.9.: Plots of negative log returns and estimated 0.99 Value–at–Risk for the six datasets (in–sample).

Part II.

Duration models

5 Duration models

5.1. ACD models

ACD models (Engle and Russell (1998)) were originally designed to understand the processing of exogeneous information in financial markets, leading to applications to durations between trades, arrivals of orders, updates of quotes, or price changes. These applications became possible due to the availability of so–called tick–by–tick or high frequency financial data. According to classical financial theory, trading activities are only necessary when new information is arriving at the market. However, duration models aim at analyzing possible market frictions and the presence of arbitrage opportunities. All these data have in common that they exhibit non–equally spaced durations. The transfer to our problem at hand is obvious: time intervals between threshold exceedances are not equally spaced, either. Hence, from a methodological point of view it does not matter whether we model intraday transaction data with stochastic trading times or daily excess data with stochastic occurrences.

Engle and Russell (1998) propose a framework similar to Engle (1982)'s ARCH models for volatility. ARCH models are concerned with conditional volatilities, whereas the so–called ACD (autoregressive conditional duration) models deal with conditional durations.

The durations between event times $\tau_i = t_i - t_{i-1}$ constitute an operational time series whose properties are modeled directly within the ACD framework — in contrast to the models treated in chapters 3 and 4, where the conditional intensity is directly estimated resulting in an occurrence probability for each day and in principle for each point in continuous time. However, the self–exciting ground model (without mark influence) and the ACD model correspond as they both employ the same information about the point process $N(0,T]$. Although the ACD model on the one hand and conditional intensity models on the other represent alternative strategies of event modelling, it is possible to characterize ACD model in terms of their *implied conditional intensity*.

In the ACD(q,p) model it is assumed that the expected duration $E(\tau_i|\mathcal{H}_{i-}) = \psi_i$ — where $\mathcal{H}_{i-}$ denotes the filtration or information set up to but not including the i-th event — depends on its own past, the past durations and possibly further possibly time–dependent covariates:

$$\psi_i = \omega + \sum_{j=1}^{q} \alpha_j \tau_{i-j} + \sum_{j=1}^{p} \beta_j \psi_{i-j} + \zeta z_i,$$

where durations which are standardized by some appropriate function $f(\cdot)$ satisfy $\tau_i/f(\psi_i) \stackrel{iid}{\sim} G$. Thus, introducing an error term ε_i, the duration can be written in terms of the expected duration as

$$\tau_i = f(\psi_i)\varepsilon_i. \tag{5.1}$$

The distribution G is not time–dependent since the dependence structure is entirely captured by the sequence of conditional durations $\{\psi_i\}$. To secure that ψ_i is always positive, involving the exponential function to obtain the log ACD model can be useful (Bauwens and Giot (2001)).

Similar to GARCH models, the difference process $\delta_i = \tau_i - f(\psi_i)$ is a martingale difference sequence, i.e. $E(\delta_i|\mathcal{H}_{i-}) = 0$. This fact motivates the following ARMA representation of the duration process:

$$\tau_i = \omega + \sum_{j=1}^{\max(p,q)} (\alpha_j + \beta_j)\tau_{i-j} - \sum_{j=1}^{q} \beta_j\delta_{i-j} + \varepsilon_j.$$

The two distributional assumptions considered within the original ACD model (Engle and Russell (1998)) are the exponential and the Weibull distribution. In case of the exponential distribution, when standardization can be achieved very easily ($f(\psi_i) = \psi_i$), the density function $g(\cdot)$ corresponding to G can be expressed as

$$g\left(\frac{\tau_i}{\psi_i}\,|\mathcal{H}_{i-}, \theta_g\right) = \exp\left(-\frac{\tau_i}{\psi_i}\right),$$

where θ_g is the rate of the exponential distribution. When we know the information included in $\mathcal{H}_{i-}$ we also know ψ_i and can write

$$g\left(\tau_i\,|\mathcal{H}_{i-}, \theta_g\right) = \frac{1}{\psi_i}\exp\left(-\frac{\tau_i}{\psi_i}\right).$$

It can easily be seen that the parameter of the exponential distribution equals the reciprocal of the conditional duration. The conditional intensity implied by the EACD model right before the i-th event is hence given by

$$\lambda_g^*(t_i) = \psi_i^{-1}.$$

The implied conditional intensity remains constant until the subsequent event, then the intensity takes another constant value and so on. The process can be called somewhat contradictory a ”piecewise homogeneous” Poisson process. Although exponentially distributed, durations only seemingly have a lack of memory since the rate of the process depends on the conditional duration which is determined by the past according to the order of the model — the larger the order, the longer the memory. Moreover, the fact that the process is only piecewise homogeneneous in principle facilitates to model the clustering phenomenon.

For the Weibull case (WACD model) and with standardization function $f(\psi_i) = \psi_i[\Gamma(1+1/\gamma)]$ the following conditional density and conditional intensity functions can be obtained:

$$g(\tau_i \mid \mathcal{H}_{i-}, \theta_g) = \frac{\gamma}{\tau_i}\left(\frac{\tau_i}{f(\psi_i)}\right)^{\gamma}\exp\left[-\left(\frac{\tau_i}{f(\psi_i)}\right)^{\gamma}\right],$$

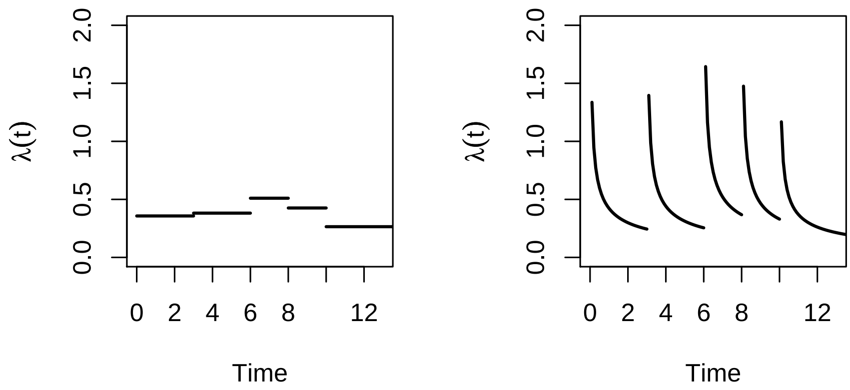


Figure 5.1.: Implied conditional intensities of EACD(1,1) (left panel) and WACD(1,1) (right panel) models based on five artificial data points. The resulting conditional durations are very similar, the implicit intensities are not.

and

$$\lambda_g^*(t_i) = f(\psi_i)^{-\gamma}\tau_i^{\gamma-1}\gamma.$$

The conditional intensity decreases ($0 < \gamma < 1$) or increases ($\gamma > 0$) monotonicly with respect to τ_i, and it is reduced to the EACD model for $\gamma = 0$. It is hence not possible to model a non–monotonic course of the conditional duration. Besides the functional form, the remarks about the memory of the process that could be stated for the EACD model also hold for the WACD case: the memory depends on the order of the model.

An important difference of the implied WACD intensity to the explicitely modelled conditional intensity function of part I lies in its behaviour in the neighbourhood of the latest preceding event. In the latter case, neglecting possible mark influences, the contribution of the last event to the conditional intensity is finite (about 1) in the simple Hawkes as well as in the ETAS model. The WACD intensity decays much more rapidly, the probability of further events in the close neighbourhood is very high (figure 5.1). Since we actually analyze discrete data with minimum duration of one day, this feature might not be a desirable one.

Generalizations of these distributional assumptions have foremostly been introduced in order to account for possible non–monotonicity of durations. Such distributional assumptions include the generalized Gamma distribution (GACD model, Zhang, Russell, and Tsay (2001)) and the Burr distribution (Burr–ACD model, Grammig and Maurer (2000)). While durations between trades can indeed be assumed to have a monotonic implied hazard, price durations — the time span required for the price change to exceed a certain threshold — and volume durations — time span to achieve a certain cumulative trading volume — can not. However, these kinds of generalizations may match the needs of some tick–by–tick transaction data, e.g. price and volume durations.

To inspect the appropriateness of the Weibull model for daily exceedance durations we use the kernel density estimations and empirical ACFs depicted in figures 5.3 and 5.4, respectively. Taking into account the lowest possible duration for daily data, the

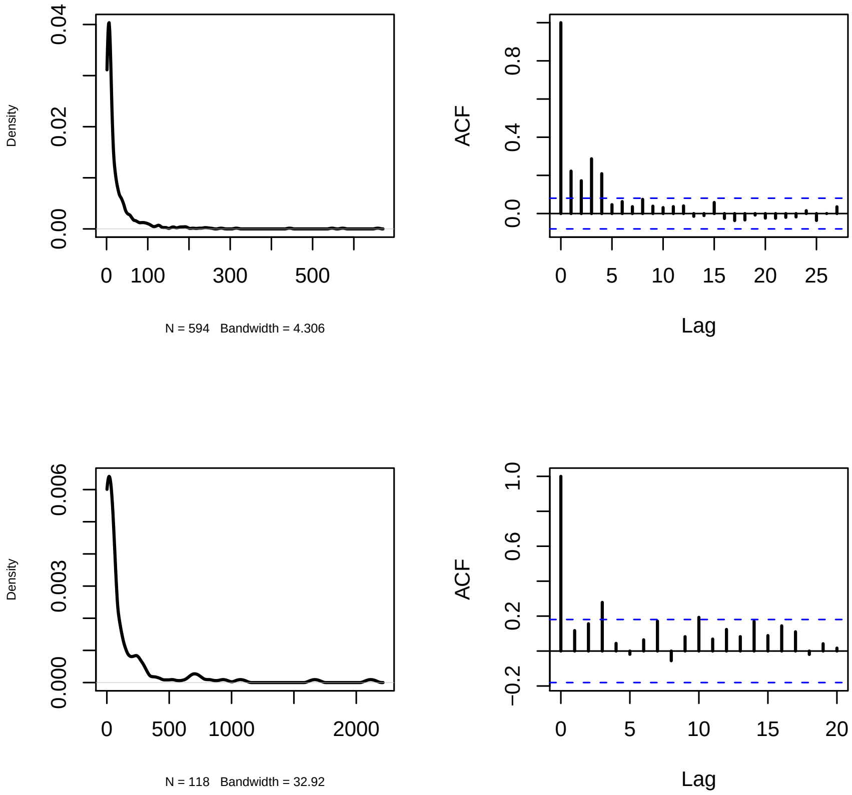


Figure 5.2.: Density estimate using a biweight kernel and empirical ACF of exceedances of IBM stock's negative returns over the thresholds 0.0243 (upper panel) and 0.0431 (lower panel), i.e. the 0.95 and 0.99 quantiles.

density estimation is based on the continuous support $[1, \infty)$. Figure 5.3 indicates that the Weibull model might be sufficient insofar as the duration density appears to decay monotonicly. The autocorrelograms in figure 5.4 suggest that the duration process exhibits significant autocorrelations even for higher lags. However, it may again be underlined that the meaning of long memory of the duration series and memory of the exceedance process on the original time axis have very different implications. Due to the clustering phenomenon, tranquil periods are represented by much fewer observations τ_i than volatile periods, hence significant autocorrelations for higher lags may nonetheless imply short memory on the original time axis if they are based on many, but clustered events. The results shown for the 0.95 quantile thresholds hold for higher thresholds as well.

Maximum likelihood estimation

The likelihood function of an ACD(q,p) model of durations $\tau_2, \ldots, \tau_N$ given the vector of parameters θ is[1]

[1] Because of the definition of τ_i the first duration is τ_2. For convenience and because of the large sample size we can ignore the first (left–censored) and last (right–censored) observation and obtain $N-1$ durations between N events.

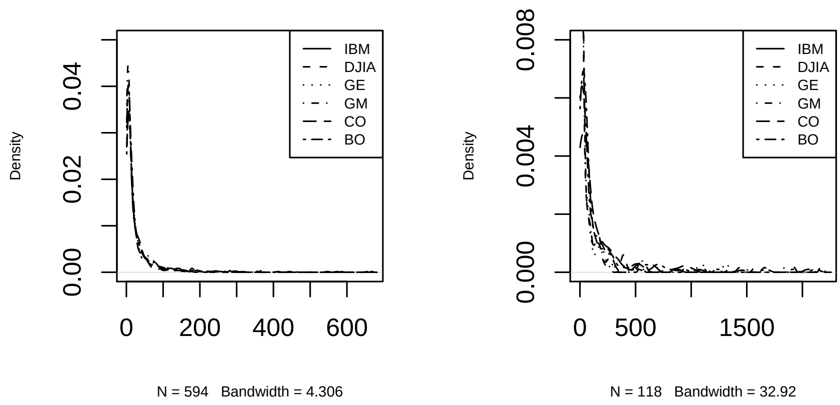


Figure 5.3.: Density estimate using a biweight kernel for exceedance durations of the six negative return series over the thresholds determined by the 0.95 and 0.99 quantiles.

$$L(\tau_2, \ldots, \tau_N \mid \theta) = \prod_{i=\max p,q+2}^{N} g(\tau_i \mid \mathcal{H}_{i-}, \theta) \times g(\tau_{\max(p,q)} \mid \theta).$$

It is acceptable to ignore the marginal density function $g(\cdot|\theta)$ for sufficiently large N and compute parameter estimates based on the conditional part. In the following, we consider the exponential (EACD), Weibull (WACD), and generalized Gamma (GACD) models. The conditional log likelihood of the latter can be written as

$$\begin{aligned} l(\tau_2, \ldots, \tau_N \mid \gamma, \kappa, t_1) = \sum_{i=\max(p,q)+2}^{N} & \log\left(\frac{\gamma}{\Gamma(\kappa)}\right) + (\kappa\gamma - 1)\log(\tau_i) \\ & - \kappa\gamma \log(\Gamma^{-1}(1 + 1/\gamma)\psi_i) - \left(\frac{\tau_i}{\Gamma^{-1}(1 + 1/\gamma)\psi_i}\right)^{\gamma}, \end{aligned}$$

including the special cases of the WACD model (for $\kappa = 1$) and the EACD model (for $\kappa = \gamma = 1$).

5.2. Empirical results

Not only does the inspection of the estimated duration densities suggest a monotonic decay and hence a monotonic decay of the implied intensity. Estimation results for the generalized Gamma ACD model indicate that the model is ”too general” as well, since the estimation of its shape parameter κ is highly dependent on the initial value within the numerical optimization. Imposing the restriction $\kappa = 1$ leads to the WACD model without much impact on the value of the maximum log likelihood. This result strongly supports the conjecture of a monotonic conditional intensity function.

Therefore, in the following only EACD and WACD models are considered. In both cases model selection somewhat surprisingly (because of the ”long memory” suggested by the

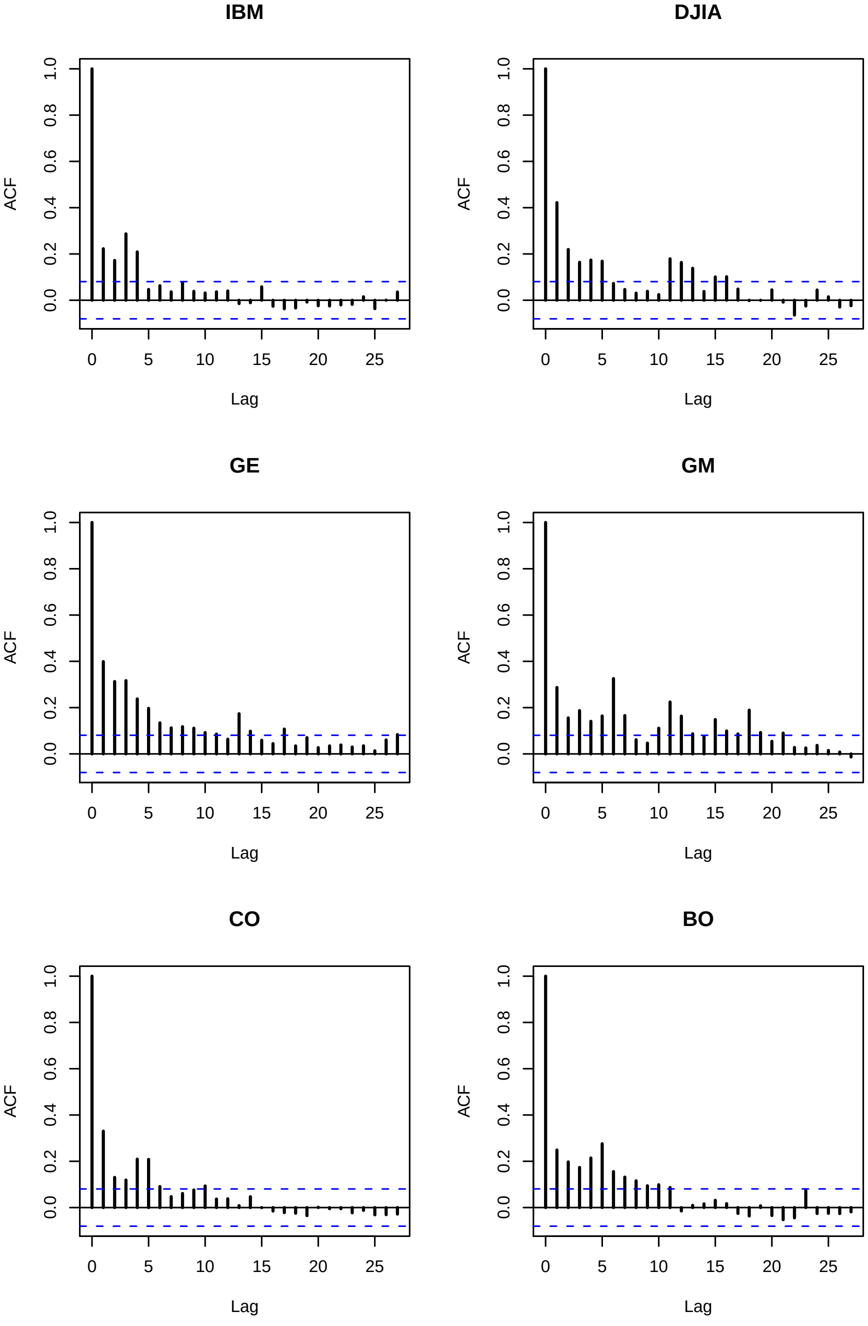


Figure 5.4.: Empirical ACF of exceedance durations of the six negative return series over the thresholds determined by the 0.95 quantiles.

empirical ACFs) displays the most parsimonious versions — EACD(1,1) and WACD(1,1) — as the best ones in terms of likelihood ratio tests and AIC.

Moreover, we can examine whether the WACD model performs significantly better than the EACD model because of their nestedness.

	IBM	DJIA	GE	GM	CO	BO
90%	0.0172	0.0106	0.0167	0.0200	0.0163	0.0227
95%	0.0243	0.0149	0.0233	0.0285	0.0227	0.0315
99%	0.0431	0.0254	0.0441	0.0531	0.0395	0.0552

Table 5.1.: Quantiles of negative returns that are used as thresholds to construct the point processes of exceedances under investigation.

The tables 5.2 and 5.3 show maximum log likelihoods and the associated AIC values for three stocks. The latter leads us to the following conclusions:

- For the low threshold, the more general Weibull model does not provide much improvement compared to the EACD model (if any), irrespective of the order of the model (i.e. the memory). The result holds for even higher orders as well.
- For both higher thresholds, the WACD models outperform the EACD models.
- However, evidence for the appropriate order of the WACD models is ambiguous. For the IBM and GM stocks, the WACD(1,1) model appears to be of sufficient order. For the GE stock and especially the 0.99 quantile threshold, the inclusion of durations with lag 2 leads to a substantial improvement of the fit. However, inclusion of yet more lags does not provide any further improvement.

The first result is intuitive in principle since the lower the threshold is chosen, the more the resulting point process has to resemble a homogeneous Poisson process because of the very nature of the data, though the characterization of the 0.9 quantile as ”low” is not that evident indeed. The stylized fact of a decaying intensity discussed earlier is captured by the WACD model only and appears to be more appropriate at least for the higher thresholds. The memory of the duration process seems not to be very long since the order (p=2, q=1) is favourable according to ACD models' AICs. In order to assess the reliability of these results we perform the same residual analysis as for the self–exciting POT models.

model	data	.90	.95	.99
EACD(1,1)	IBM	-4210.8191	-2465.6422	-679.9352
WACD(1,1)	IBM	-4210.1167	-2436.0797	-643.1026
EACD(2,1)	IBM	-4209.9843	-2465.6856	-677.7949
WACD(2,1)	IBM	-4209.3232	-2436.1437	-642.3329
EACD(1,1)	GE	-4187.6044	-2383.3849	-651.0213
WACD(1,1)	GE	-4186.2518	-2381.0740	-603.0535
EACD(2,1)	GE	-4187.5809	-2381.7651	-633.9026
WACD(2,1)	GE	-4186.2288	-2379.6575	-592.4614
EACD(1,1)	GM	-4148.9296	-2346.2637	-552.7331
WACD(1,1)	GM	-4148.7254	-2334.4298	-543.3283
EACD(2,1)	GM	-4147.5675	-2347.6047	-549.5884
WACD(2,1)	GM	-4147.3414	-2335.2118	-541.6241

Table 5.2.: Log likelihood values at the ML estimate for the thresholds marked by the 0.9 / 0.95 / 0.99 quantiles of negative log returns of IBM, GE and GM.

model	data	.90	.95	.99
EACD(1,1)	IBM	8427.6383	4937.2843	1365.8704
WACD(1,1)	IBM	8428.2333	4880.1593	1294.2052
EACD(2,1)	IBM	8427.9686	4939.3712	1363.5899
WACD(2,1)	IBM	8428.6464	4882.2874	1294.6659
EACD(1,1)	GE	8381.2089	4772.7698	1308.0426
WACD(1,1)	GE	8380.5036	4770.1481	1214.1071
EACD(2,1)	GE	8383.1618	4771.5302	1275.8052
WACD(2,1)	GE	8382.4576	4769.3151	1194.9227
EACD(1,1)	GM	8303.8591	4698.5275	1111.4661
WACD(1,1)	GM	8305.4507	4676.8597	1094.6567
EACD(2,1)	GM	8303.1350	4703.2093	1107.1767
WACD(2,1)	GM	8304.6828	4680.4236	1093.2482

Table 5.3.: AIC values for the thresholds marked by the 0.9 / 0.95 / 0.99 quantiles of negative log returns of IBM, GE and GM.

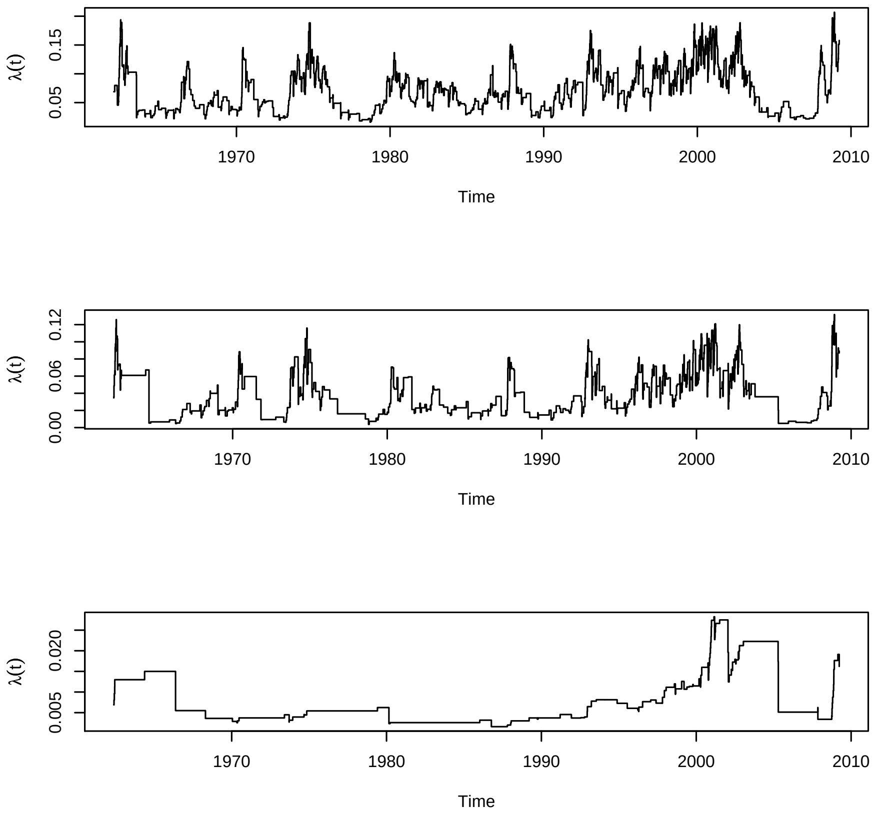


Figure 5.5.: Implied conditional intensities for the EACD(1,1) model and events being the 10 (5, 1) percent largest daily losses of the IBM stock.

5.3. Goodness of fit

In order to examine the goodness of fit of the models, i.e. their ability to reproduce the temporal structure of occurrence times, again the strategy of Ogata (1988) is chosen. This means that we do not examine the residuals of the actual ACD models but the residual process constituted by the implied intensities instead.

We recall from (4.2) that the points of the cumulative intensity process

$$\Lambda(t_i) = \int_0^{t_i} \lambda_g^*(s)ds$$

for $i = 1, \ldots, N(T)$ constitute a homogeneous Poisson process of rate 1 on an interval $[0, N(T)]$ being a subset of a transformed time axis. The resulting property of exponentially distributed durations establishes the applicability of the Kolmogorov–Smirnov test. Evidence is quite unambiguous this time: all p–values indicate that the null hypothesis is rejected, neither the EACD nor the WACD yield a good fit to the data.

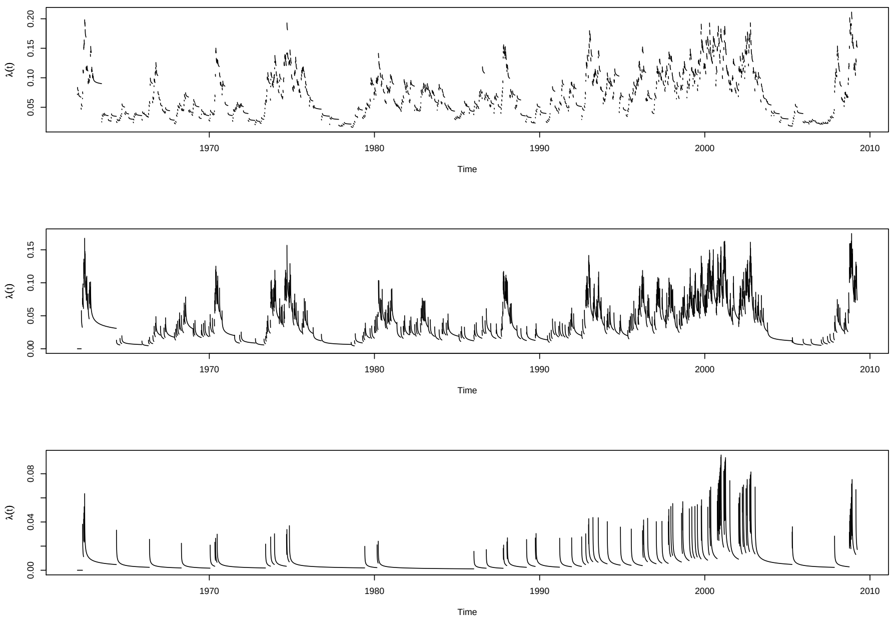


Figure 5.6.: Implied conditional intensities for the WACD(1,1) model and events being the 10 (5, 1) percent largest daily losses of the IBM stock.

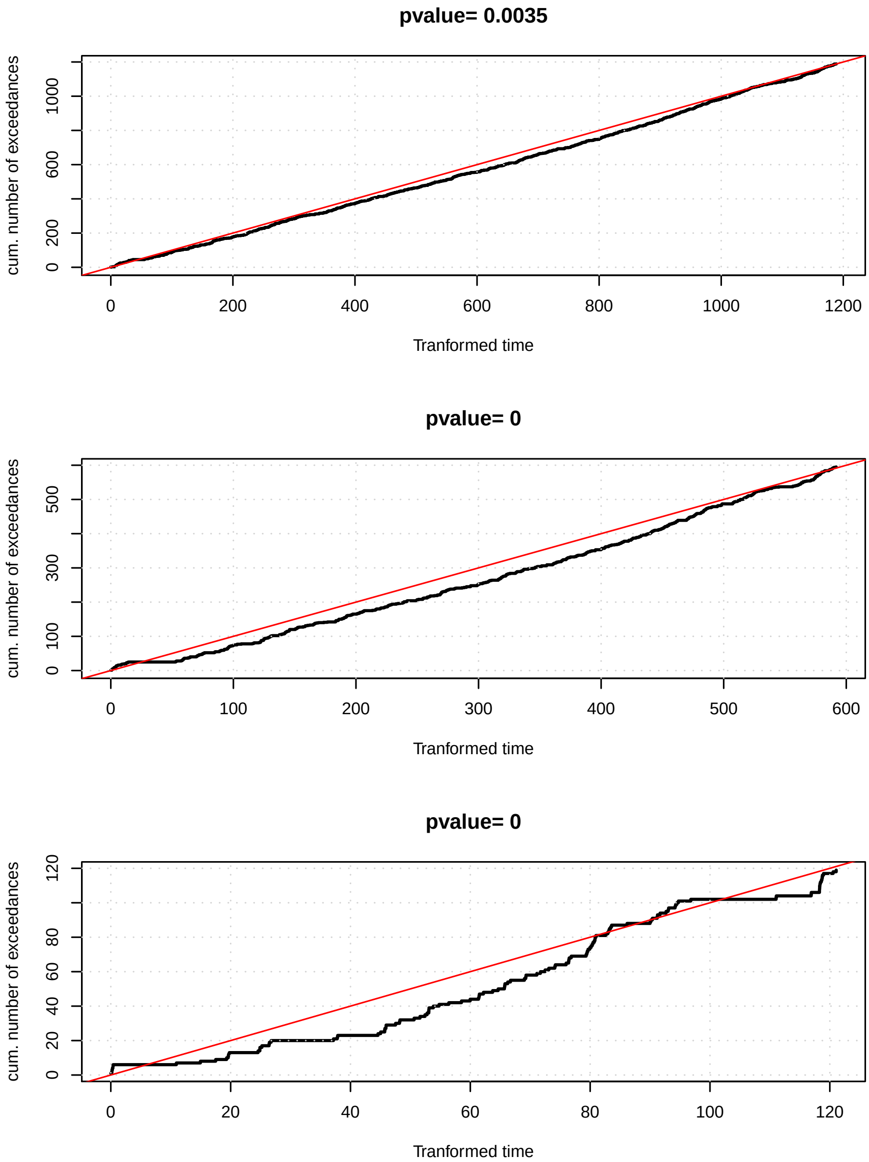


Figure 5.7.: Residual processes for the EACD(1,1) model and events being the 10 (5, 1) percent largest daily losses of the IBM stock.

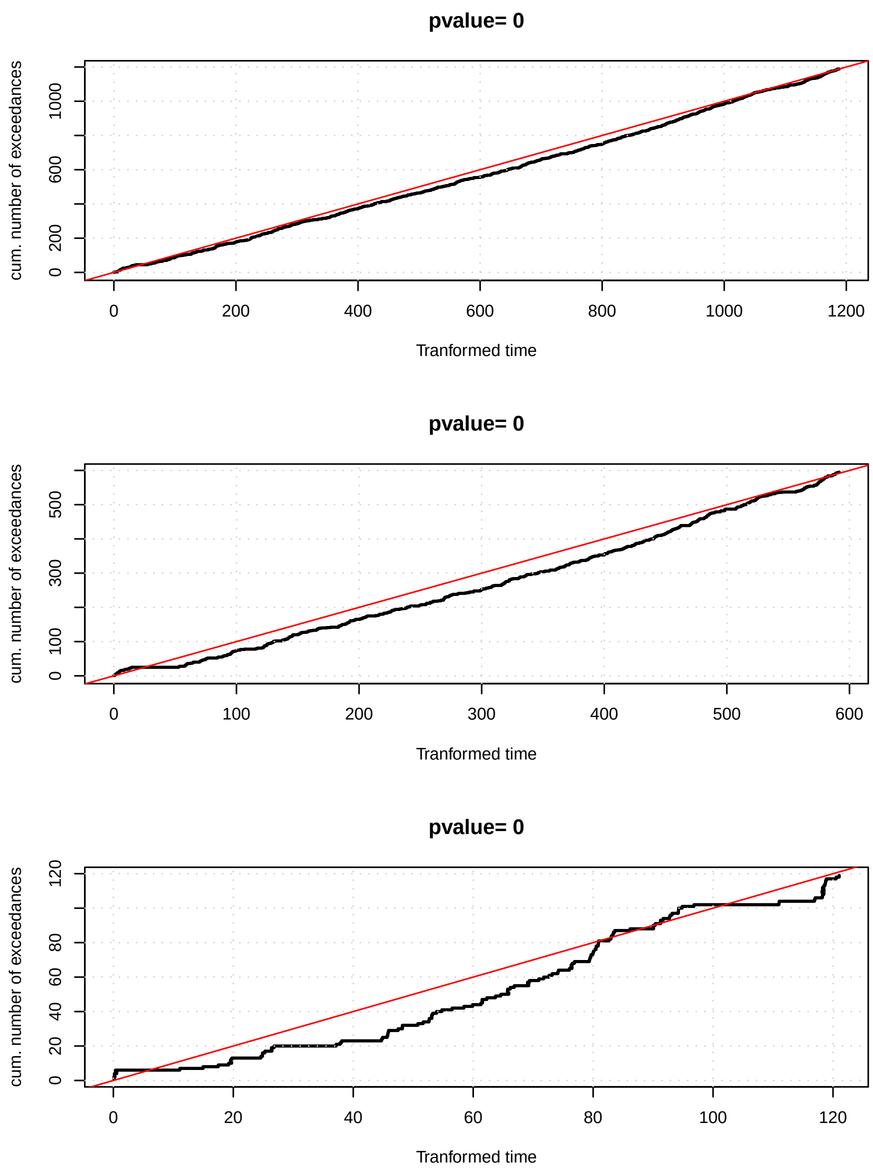


Figure 5.8.: Residual processes for the WACD(1,1) model and events being the 10 (5, 1) percent largest daily losses of the IBM stock.

6 Summary, conclusions and perspectives

In this section, major results are summarized, economic implications are discussed, and perspectives of further research are outlined.

Table 6.1 contrasts central features of self–exciting POT and ACD models that are discussed in detail within the chapters 3 to 5.

The goodness–of–fit assessment using the residual process of Ogata (1988) strongly supports the self–exciting models which are appropriate at least for higher thresholds. None of the ACD models is able to perform similarly well, the null hypothesis of a Poisson residual process with rate 1 is rejected for all ACD models assessed and all thresholds. However, the endeavour of estimating a conditional Value–at–Risk via self–exciting POT models can be cautiously described as unsuccessful, either — its predictions are outperformed even by an unconditional Gaussian model. This result might partly be caused by the inappropriateness of the GPD model: mean excess plots suggest that even a threshold of 0.05 might not be sufficiently high for three of the datasets. Nevertheless, the main reason for the bad performance is the inability of extreme events to speak for all returns. The results become much worse using yet higher thresholds, although for the extreme events itselves the GPD model as well as the point process model are successful according to the goodness–of–fit tests.

Generalizations of ACD models mostly concern the monotonicity assumption of the intensity, but this constraint appears to be appropriate for daily exceedance data. Within transaction data applications the phenomenon of a delay in information processing might be common leading to a local maximum of the implied hazard apart from zero. In case of daily data it can be assumed that this delay is not very important, the stylized facts can be captured by a model with monotonic intensity.

From an economist's point of view, the possibility of separate analyses of gains and losses is very appealing. We have shown that the conditional intensity is differently affected by exceedances of positive and negative returns. Extreme positive returns increase the probability of further extreme positive returns in the near future according to their size — even very high gains affect the probability of further very high gains—, while negative returns lose this predictive property above some level. Moreover, this level appears to be determined by an *absolute* threshold size. Although we do not have any information on an individual market participant's level, this might imply that market participants' loss aversion is inelastic from a certain threshold on. This phenomenon is indeed known from

Aspect	SEPP models	ACD models
Main focus	estimation of monotonously decaying conditional intensity function	estimation of conditional durations
Role of the intensity	explicitely modelled	implicitely modelled
Role of durations	implicitely modelled	explicitely modelled
Covariate information	time elapsed + mark size, potentially further influences (marks) possible	former durations and conditional durations (autoregression), further influences possible
Memory	infinite	usually short, depending on the order of the model

Table 6.1.: Comparison self–exciting PP vs. ACD models

behavioral economics and, e.g., incorporated within *prospect theory*.[1] However, it should not be concealed that such diagnoses based on one data point per day are somewhat speculative.

The self–exciting point process models have large potential to be expanded and refined. More general functional forms of the conditional intensity are in principle possible, e.g. via time dependent parameters. Therefore, the stability of parameter estimates has to be examined which can be done using the model scores (Zeileis and Hornik (2007)). On the other hand, restrictions of the memory of the process might simplify the estimation without loss of predictive power. Further covariate influences such as market liquidity could enter the model as marks.

In order to use self–exciting POT models for risk managment purposes, combinations of estimated intensities for several thresholds might improve VaR estimates towards competitiveness with other models.

[1] See Barberis and Thaler (2002) for an overview and Benartzi and Thaler (1995) for an application of Tversky's prospect theory to the equity premium puzzle.

Part III.

Appendix

A Implementation of self–exciting models

The self–exciting models introduced in chapter ?? were estimated using functions based on the R package `QRMlib`, which is the R implementation of Alexander McNeils's S-Plus library accompanying McNeil, Frey, and Embrechts (2005).

All self–exciting point processes considered have something in common: their conditional intensity $\lambda^*(t) = \mu + \psi v^*(\cdot)$ is of Hawkes type, hence it has the main ingredient

$$v^*(t) = \sum_{j:0<T_j<t} \eta(t - T_j; \tilde{X}_j - u),$$

whereas the choice of $\eta(s,x)$ determines the specific model, e.g. "Hawkes" or "ETAS" in McNeil's terminology.

As was pointed out in chapter 3, Hawkes–type models and ETAS models are defined and distinguished somewhat differently from the earthquake literature they were adopted from. Moreover, there are some differences between the model formulae presented in Chavez-Demoulin, Davison, and McNeil (2005) and the book accompanied by the package (McNeil, Frey, and Embrechts (2005), see p. 306 therein) on the one hand and McNeil's implementation on the other. The formulae from both the book and the C subroutines used within `QRMlib` are shown in table A.1.

Remarkably, model estimation with the R implementations of the respective functions $\eta(s,x)$ presented in McNeil, Frey, and Embrechts (2005) make the model estimation unfeasible (singular hessian within the numerical optimization routine) in even more cases than before, presumably because of the parameter(s) within the exponential function. In general such difficulties with numerical optimization of Hawkes–type likelihoods are a well–known fact according to Daley and Vere-Jones (2003).

The sensitivity of estimation practicability was inspected with several directions in view. Smaller intervals might be preferable in presence of structural breaks that are not captured by — e.g — time dependent parameters. Larger intervals might be necessary if the method

model name	QRM book	`QRMlib` implementation
Hawkes	$\eta(s,x) = \exp(\delta x - \gamma s)$	$\eta(s,x) = (1 + \delta x)\exp(-\gamma s)$
ETAS	$\eta(s,x) = \exp(\delta x)(s + \gamma)^{-(\rho+1)}$	$\eta(s,x) = (1 + \delta x)(1 + \frac{s}{\gamma})^{-(\rho+1)}$

Table A.1.: Model formulae in McNeil, Frey, and Embrechts (2005) vs. implementation within the `QRMlib` package in R is very wasteful of data. However, the feasibility of estimation does not appear to depend on the width of the interval.

Moreover, it does not appear to depend larglely on the optimization method chosen. The numerical problems could not be remediated by using alternate optimization algorithms, either. The `QRMlib` implementation uses optimization using `PORT` routines (R function `nlminb` within the `stats` package, a quasi Newton optimizer).[1] The Nelder–Mead algorithm essentially yields the same results — they especially suffer from the same numerical difficulties for the same models as the `PORT` optimization. Several further optimizers — including BFGS that was preferable within ACD model estimation — do converge even less often. The sensitivity to starting values for the parameters is not substantial.

[1] The PORT documentation is available at `http://netlib.bell-labs.com/cm/cs/cstr/153.pdf`.

dataset	threshold	POT	se1	se2	se3	se4	sePOT1a	sePOT2a	sePOT3a	sePOT4a	sePOT1b	sePOT2b	sePOT3b	sePOT4b
DJIA	0.01	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
DJIA	0.02	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
DJIA	0.03	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
BA	0.01	yes	yes	yes	no	no	yes	yes	no	no	yes	yes	no	no
BA	0.02	yes	yes	yes	no	no	yes	yes	no	no	yes	yes	no	no
BA	0.03	yes	yes	yes	no	no	yes	yes	no	no	yes	yes	no	no
GE	0.01	yes	yes	yes	no	no	yes	yes	no	no	yes	yes	no	no
GE	0.02	yes	yes	yes	no	no	yes	yes	no	no	yes	yes	no	no
GE	0.03	yes	yes	yes	no	no	yes	yes	no	no	yes	yes	no	no
GM	0.01	yes	yes	yes	no	no	yes	yes	no	no	yes	yes	no	no
GM	0.02	yes	yes	yes	no	no	yes	yes	no	no	yes	yes	no	no
GM	0.03	yes	yes	yes	no	no	yes	yes	no	no	yes	yes	no	no
IBM	0.01	yes	yes	yes	no	no	yes	yes	no	no	yes	yes	no	no
IBM	0.02	yes	yes	yes	no	no	yes	yes	no	no	yes	yes	no	no
IBM	0.03	yes	yes	yes	no	no	yes	yes	no	no	yes	yes	no	no
KO	0.01	yes	yes	yes	no	no	yes	yes	no	no	yes	yes	no	no
KO	0.02	yes	yes	yes	no	no	yes	yes	no	no	yes	yes	no	no
KO	0.03	yes	yes	yes	no	no	yes	yes	no	no	yes	yes	no	no

Table A.2.: Overview of all possible models with the six datasets under investigation for three different thresholds. The entries indicate whether the model has been estimated or could not be estimated due to difficulties within the numerical optimization routine. The numbers 1 and 2 stand for the simple Hawkes model without and with mark influence, 3 and 4 likewise for the ETAS model. The letter ”a” signifies models with iid GP distributed marks, ”b” stands for conditional GP distributed (”predictable”) marks.

B Model tables

B.1. Losses, thresholds 0.01, 0.02, 0.03

dataset	threshold	$\hat{\lambda}$	$\hat{\xi}$	$\hat{\beta}$	$se_{\hat{\lambda}}$	$se_{\hat{\xi}}$	$se_{\hat{\beta}}$	ll_{max}	$ll_{max,times}$	AIC_{times}	$ll_{max,marks}$	AIC_{marks}
DJIA	0.0100	0.0756	0.1938	0.0058	0.0021	0.0301	0.0002	515.2600	-4674.0755	9350.1511	5189.3355	-10374.6711
DJIA	0.0200	0.0157	0.3656	0.0065	0.0010	0.0821	0.0006	-400.4131	-1396.6003	2795.2006	996.1872	-1988.3743
DJIA	0.0300	0.0044	0.3163	0.0114	0.0005	0.1494	0.0021	-248.2285	-488.2918	978.5836	240.0633	-476.1266
GE	0.0100	0.1429	0.2566	0.0082	0.0029	0.0211	0.0002	1480.7702	-7261.0970	14524.1940	8741.8672	-17479.7345
GE	0.0200	0.0498	0.4506	0.0081	0.0017	0.0464	0.0005	-540.8100	-3435.8767	6873.7534	2895.0667	-5786.1334
GE	0.0300	0.0188	0.5340	0.0111	0.0010	0.0753	0.0010	-650.9623	-1615.8351	3233.6701	964.8727	-1925.7455
IBM	0.0100	0.1468	0.2375	0.0083	0.0029	0.0206	0.0002	1595.1853	-7392.4737	14786.9474	8987.6590	-17971.3180
IBM	0.0200	0.0518	0.5054	0.0081	0.0017	0.0561	0.0004	-524.0764	-3537.2077	7076.4153	3013.1312	-6022.2625
IBM	0.0300	0.0199	0.5220	0.0098	0.0011	0.0713	0.0008	-622.5454	-1686.8380	3375.6759	1064.2925	-2124.5851
KO	0.0100	0.1360	0.2517	0.0079	0.0028	0.0209	0.0002	1390.1794	-7026.6415	14055.2830	8416.8210	-16829.6419
KO	0.0200	0.0457	0.4334	0.0078	0.0016	0.0445	0.0004	-521.9336	-3219.8683	6441.7366	2697.9347	-5391.8694
KO	0.0300	0.0163	0.6697	0.0087	0.0010	0.0900	0.0009	-572.1381	-1437.9531	2877.9061	865.8150	-1727.6299
BA	0.0100	0.1862	0.1671	0.0116	0.0033	0.0162	0.0003	1968.1552	-8611.2959	17224.5919	10579.4511	-21154.9022
BA	0.0200	0.0857	0.2720	0.0108	0.0022	0.0285	0.0004	-298.4924	-5112.1693	10226.3386	4813.6769	-9623.3538
BA	0.0300	0.0390	0.3739	0.0112	0.0015	0.0465	0.0007	-756.8309	-2852.8867	5707.7733	2096.0558	-4188.1116
GM	0.0100	0.1657	0.2251	0.0102	0.0031	0.0203	0.0003	1613.0829	-7997.7615	15997.5229	9610.8444	-19217.6887
GM	0.0200	0.0689	0.3306	0.0106	0.0020	0.0356	0.0005	-546.0628	-4369.2867	8740.5733	3823.2238	-7642.4477
GM	0.0300	0.0301	0.4331	0.0122	0.0013	0.0607	0.0009	-795.9374	-2340.9342	4683.8684	1544.9968	-3085.9936

Table B.1.: POT parameter estimates, standard errors, and log–likelihood evaluated at the MLEs for the three thresholds and six datasets. Note the somewhat peculiar fact of positive log–likelihoods for some of the models. A positive log-likelihood is acceptable in a model with continuous response because the probability density function can be greater than one, which implies that the log–likelihood can be positive.

dataset	threshold	$\hat{\mu}$	$\hat{\phi}$	$\hat{\gamma}$	$se_{\hat{\mu}}$	$se_{\hat{\phi}}$	$se_{\hat{\gamma}}$	ll_{max}	AIC
dji	0.0100	0.0173	0.0156	0.0201	0.0029	0.0021	0.0031	-4460.3963	8926.7927
dji	0.0200	0.0033	0.0200	0.0250	0.0006	0.0033	0.0044	-1199.9248	2405.8497
dji	0.0300	0.0015	0.0249	0.0375	0.0003	0.0067	0.0096	-401.0666	808.1332
ge	0.0100	0.0411	0.0079	0.0110	0.0076	0.0012	0.0017	-7167.0766	14340.1532
ge	0.0200	0.0104	0.0115	0.0143	0.0020	0.0016	0.0022	-3228.8776	6463.7552
ge	0.0300	0.0041	0.0145	0.0178	0.0008	0.0022	0.0029	-1409.1085	2824.2169
ibm	0.0100	0.0428	0.0080	0.0112	0.0077	0.0012	0.0018	-7300.0832	14606.1664
ibm	0.0200	0.0120	0.0142	0.0184	0.0020	0.0020	0.0029	-3355.0389	6716.0777
ibm	0.0300	0.0058	0.0176	0.0246	0.0009	0.0029	0.0043	-1546.3070	3098.6140
coco	0.0100	0.0313	0.0074	0.0094	0.0068	0.0010	0.0014	-6920.2431	13846.4862
coco	0.0200	0.0110	0.0126	0.0165	0.0020	0.0020	0.0029	-3053.4808	6112.9617
coco	0.0300	0.0062	0.0134	0.0216	0.0011	0.0029	0.0055	-1348.3667	2702.7334
bo	0.0100	0.0764	0.0050	0.0084	0.0127	0.0011	0.0018	-8573.4560	17152.9119
bo	0.0200	0.0219	0.0102	0.0135	0.0042	0.0016	0.0023	-4993.6284	9993.2568
bo	0.0300	0.0098	0.0135	0.0179	0.0018	0.0021	0.0032	-2714.9655	5435.9310
gm	0.0100	0.0376	0.0049	0.0063	0.0084	0.0009	0.0012	-7896.3325	15798.6650
gm	0.0200	0.0147	0.0128	0.0160	0.0028	0.0019	0.0027	-4145.6424	8297.2849
gm	0.0300	0.0052	0.0145	0.0170	0.0011	0.0023	0.0030	-2072.8726	4151.7452

Table B.2.: Parameter estimates, standard errors, and maximum log–likelihood for the self–exciting ground process without mark influence.

dataset	threshold	$\hat{\mu}$	$\hat{\phi}$	$\hat{\gamma}$	$\hat{\delta}$	$se_{\hat{\mu}}$	$se_{\hat{\phi}}$	$se_{\hat{\gamma}}$	$se_{\hat{\delta}}$	ll_{max}	AIC
dji	0.0100	0.0195	0.0138	0.0209	19.0897	0.0032	0.0022	0.0034	10.7647	-4460.3963	8924.0360
dji	0.0200	0.0034	0.0169	0.0249	17.2331	0.0007	0.0034	0.0044	11.9241	-1199.9248	2404.5672
dji	0.0300	0.0015	0.0242	0.0376	1.8112	0.0003	0.0076	0.0096	10.6988	-401.0666	810.1018
ge	0.0100	0.0561	0.0048	0.0102	26.4636	0.0086	0.0012	0.0017	10.8905	-7167.0766	14323.4562
ge	0.0200	0.0109	0.0101	0.0139	5.6019	0.0021	0.0016	0.0022	3.1647	-3228.8776	6460.9640
ge	0.0300	0.0041	0.0145	0.0178	0.0000	0.0008	0.0022	0.0029		-1409.1085	2826.2169
ibm	0.0100	0.0505	0.0057	0.0105	17.9170	0.0082	0.0013	0.0018	8.3075	-7300.0832	14597.6342
ibm	0.0200	0.0120	0.0134	0.0179	2.5965	0.0021	0.0021	0.0029	2.5207	-3355.0389	6716.5969
ibm	0.0300	0.0058	0.0174	0.0245	0.3653	0.0009	0.0030	0.0043	1.1774	-1546.3070	3100.5016
coco	0.0100	0.0381	0.0058	0.0097	16.9811	0.0073	0.0011	0.0014	6.7028	-6920.2431	13835.5904
coco	0.0200	0.0109	0.0119	0.0162	2.7089	0.0020	0.0020	0.0029	2.1518	-3053.4808	6112.8511
coco	0.0300	0.0062	0.0133	0.0217	0.5243	0.0011	0.0029	0.0055	1.3772	-1348.3667	2704.5701
bo	0.0100	0.0941	0.0028	0.0084	34.7151	0.0141	0.0011	0.0022	20.8557	-8573.4560	17144.2523
bo	0.0200	0.0239	0.0087	0.0137	9.6313	0.0044	0.0016	0.0024	5.3567	-4993.6284	9989.9584
bo	0.0300	0.0098	0.0125	0.0177	3.7117	0.0018	0.0021	0.0031	3.1176	-2714.9655	5435.9560
gm	0.0100	0.0510	0.0035	0.0058	15.1437	0.0101	0.0010	0.0013	7.7965	-7896.3325	15792.0560
gm	0.0200	0.0178	0.0116	0.0175	9.0201	0.0031	0.0019	0.0031	4.4793	-4145.6424	8292.6864
gm	0.0300	0.0054	0.0137	0.0172	2.8703	0.0011	0.0024	0.0031	3.1698	-2072.8726	4152.7080

Table B.3.: Parameter estimates, standard errors, and maximum log–likelihood for the self–exciting ground process with mark influence.

dataset	threshold	$\hat{\mu}$	$\hat{\phi}$	$\hat{\gamma}$	$\hat{\xi}$	$\hat{\beta}$	$se_{\hat{\mu}}$	$se_{\hat{\phi}}$	$se_{\hat{\gamma}}$	$se_{\hat{\xi}}$	$se_{\hat{\beta}}$	ll_{max}	$ll_{max,times}$	AIC_{times}	$ll_{max,marks}$	AIC_{times}
dji	0.0100	0.0173	0.0156	0.0201	0.1938	0.0058	0.0029	0.0021	0.0031	0.0301	0.0002	728.9392	-4460.3963	8926.7927	5189.3355	-10374.6711
dji	0.0200	0.0033	0.0200	0.0250	0.3656	0.0065	0.0006	0.0033	0.0044	0.0821	0.0006	-203.7377	-1199.9248	2405.8497	996.1872	-1988.3743
dji	0.0300	0.0015	0.0249	0.0375	0.3163	0.0114	0.0003	0.0067	0.0096	0.1494	0.0021	-161.0033	-401.0666	808.1332	240.0633	-476.1266
ge	0.0100	0.0411	0.0079	0.0110	0.2566	0.0082	0.0076	0.0012	0.0017	0.0211	0.0002	1574.7907	-7167.0766	14340.1532	8741.8672	-17479.7345
ge	0.0200	0.0104	0.0115	0.0143	0.4506	0.0081	0.0020	0.0016	0.0022	0.0464	0.0005	-333.8109	-3228.8776	6463.7552	2895.0667	-5786.1334
ge	0.0300	0.0041	0.0145	0.0178	0.5340	0.0111	0.0008	0.0022	0.0029	0.0753	0.0010	-444.2357	-1409.1085	2824.2169	964.8727	-1925.7455
ibm	0.0100	0.0428	0.0080	0.0112	0.2375	0.0083	0.0077	0.0012	0.0018	0.0206	0.0002	1687.5758	-7300.0832	14606.1664	8987.6590	-17971.3180
ibm	0.0200	0.0120	0.0142	0.0184	0.5054	0.0081	0.0020	0.0020	0.0029	0.0561	0.0004	-341.9076	-3355.0389	6716.0777	3013.1312	-6022.2625
ibm	0.0300	0.0058	0.0176	0.0246	0.5220	0.0098	0.0009	0.0029	0.0043	0.0713	0.0008	-482.0145	-1546.3070	3098.6140	1064.2925	-2124.5851
coco	0.0100	0.0313	0.0074	0.0094	0.2517	0.0079	0.0068	0.0010	0.0014	0.0209	0.0002	1496.5779	-6920.2431	13846.4862	8416.8210	-16829.6419
coco	0.0200	0.0110	0.0126	0.0165	0.4334	0.0078	0.0020	0.0020	0.0029	0.0445	0.0004	-355.5461	-3053.4808	6112.9617	2697.9347	-5391.8694
coco	0.0300	0.0062	0.0134	0.0216	0.6697	0.0087	0.0011	0.0029	0.0055	0.0900	0.0009	-482.5517	-1348.3667	2702.7334	865.8150	-1727.6299
bo	0.0100	0.0764	0.0050	0.0084	0.1671	0.0116	0.0127	0.0011	0.0018	0.0162	0.0003	2005.9951	-8573.4560	17152.9119	10579.4511	-21154.9022
bo	0.0200	0.0219	0.0102	0.0135	0.2720	0.0108	0.0042	0.0016	0.0023	0.0285	0.0004	-179.9515	-4993.6284	9993.2568	4813.6769	-9623.3538
bo	0.0300	0.0098	0.0135	0.0179	0.3739	0.0112	0.0018	0.0021	0.0032	0.0465	0.0007	-618.9097	-2714.9655	5435.9310	2096.0558	-4188.1116
gm	0.0100	0.0376	0.0049	0.0063	0.2251	0.0102	0.0084	0.0009	0.0012	0.0203	0.0003	1714.5118	-7896.3325	15798.6650	9610.8444	-19217.6887
gm	0.0200	0.0147	0.0128	0.0160	0.3306	0.0106	0.0028	0.0019	0.0027	0.0356	0.0005	-322.4186	-4145.6424	8297.2849	3823.2238	-7642.4477
gm	0.0300	0.0052	0.0145	0.0170	0.4331	0.0122	0.0011	0.0023	0.0030	0.0607	0.0009	-527.8758	-2072.8726	4151.7452	1544.9968	-3085.9936

Table B.4.: Parameter estimates, standard errors, and maximum log–likelihoods for the self–exciting POT model. The ground process exhibits no mark influence, the marks are assumed to be unpredictable (iid GPD).

dataset	threshold	$\hat{\mu}$	$\hat{\phi}$	$\hat{\gamma}$	$\hat{\delta}$	$\hat{\xi}$	$\hat{\beta}$	$se_{\hat{\mu}}$	$se_{\hat{\phi}}$	$se_{\hat{\gamma}}$	$se_{\hat{\delta}}$	$se_{\hat{\xi}}$	$se_{\hat{\beta}}$	ll_{max}	$ll_{max,times}$	AIC_{times}	$ll_{max,marks}$	AIC_{marks}
dji	0.0100	0.0195	0.0138	0.0209	19.0897	0.1938	0.0058	0.0032	0.0022	0.0034	10.7647	0.0301	0.0002	731.3176	-4458.0180	8924.0360	5189.3355	-10374.6711
dji	0.0200	0.0034	0.0169	0.0249	17.2331	0.3656	0.0065	0.0007	0.0034	0.0044	11.9241	0.0821	0.0006	-202.0964	-1198.2836	2404.5672	996.1872	-1988.3743
dji	0.0300	0.0015	0.0242	0.0376	1.8112	0.3163	0.0114	0.0003	0.0076	0.0096	10.6988	0.1494	0.0021	-160.9876	-401.0509	810.1018	240.0633	-476.1266
ge	0.0100	0.0561	0.0048	0.0102	26.4636	0.2566	0.0082	0.0086	0.0012	0.0017	10.8905	0.0211	0.0002	1584.1391	-7157.7281	14323.4562	8741.8672	-17479.7345
ge	0.0200	0.0109	0.0101	0.0139	5.6019	0.4506	0.0081	0.0021	0.0016	0.0022	3.1647	0.0464	0.0005	-331.4153	-3226.4820	6460.9640	2895.0667	-5786.1334
ge	0.0300	0.0041	0.0145	0.0178	0.0000	0.5340	0.0111	0.0008	0.0022	0.0029		0.0753	0.0010	-444.2357	-1409.1085	2826.2169	964.8727	-1925.7455
ibm	0.0100	0.0505	0.0057	0.0105	17.9170	0.2375	0.0083	0.0082	0.0013	0.0018	8.3075	0.0206	0.0002	1692.8419	-7294.8171	14597.6342	8987.6590	-17971.3180
ibm	0.0200	0.0120	0.0134	0.0179	2.5965	0.5054	0.0081	0.0021	0.0021	0.0029	2.5207	0.0561	0.0004	-341.1672	-3354.2985	6716.5969	3013.1312	-6022.2625
ibm	0.0300	0.0058	0.0174	0.0245	0.3653	0.5220	0.0098	0.0009	0.0030	0.0043	1.1774	0.0713	0.0008	-481.9583	-1546.2508	3100.5016	1064.2925	-2124.5851
coco	0.0100	0.0381	0.0058	0.0097	16.9811	0.2517	0.0079	0.0073	0.0011	0.0014	6.7028	0.0209	0.0002	1503.0258	-6913.7952	13835.5904	8416.8210	-16829.6419
coco	0.0200	0.0109	0.0119	0.0162	2.7089	0.4334	0.0078	0.0020	0.0020	0.0029	2.1518	0.0445	0.0004	-354.4909	-3052.4256	6112.8511	2697.9347	-5391.8694
coco	0.0300	0.0062	0.0133	0.0217	0.5243	0.6697	0.0087	0.0011	0.0029	0.0055	1.3772	0.0900	0.0009	-482.4701	-1348.2851	2704.5701	865.8150	-1727.6299
bo	0.0100	0.0941	0.0028	0.0084	34.7151	0.1671	0.0116	0.0141	0.0011	0.0022	20.8557	0.0162	0.0003	2011.3250	-8568.1261	17144.2523	10579.4511	-21154.9022
bo	0.0200	0.0239	0.0087	0.0137	9.6313	0.2720	0.0108	0.0044	0.0016	0.0024	5.3567	0.0285	0.0004	-177.3023	-4990.9792	9989.9584	4813.6769	-9623.3538
bo	0.0300	0.0098	0.0125	0.0177	3.7117	0.3739	0.0112	0.0018	0.0021	0.0031	3.1176	0.0465	0.0007	-617.9222	-2713.9780	5435.9560	2096.0558	-4188.1116
gm	0.0100	0.0510	0.0035	0.0058	15.1437	0.2251	0.0102	0.0101	0.0010	0.0013	7.7965	0.0203	0.0003	1718.8164	-7892.0280	15792.0560	9610.8444	-19217.6887
gm	0.0200	0.0178	0.0116	0.0175	9.0201	0.3306	0.0106	0.0031	0.0019	0.0031	4.4793	0.0356	0.0005	-319.1194	-4142.3432	8292.6864	3823.2238	-7642.4477
gm	0.0300	0.0054	0.0137	0.0172	2.8703	0.4331	0.0122	0.0011	0.0024	0.0031	3.1698	0.0607	0.0009	-527.3572	-2072.3540	4152.7080	1544.9968	-3085.9936

Table B.5.: Parameter estimates, standard errors, and maximum log–likelihoods for the self–exciting POT model. The ground process accounts for possible mark influence, the marks are assumed to be unpredictable (iid GPD).

dataset	threshold	$\hat{\mu}$	$\hat{\phi}$	$\hat{\gamma}$	$\hat{\xi}$	$\hat{\beta}$	$\hat{\alpha}$	$se_{\hat{\mu}}$	$se_{\hat{\phi}}$	$se_{\hat{\gamma}}$	$se_{\hat{\xi}}$	$se_{\hat{\beta}}$	$se_{\hat{\alpha}}$	ll_{max}	$ll_{max,times}$	AIC_{times}	$ll_{max,marks}$	AIC_{marks}
dji	0.0100	0.0191	0.0182	0.0240	0.1082	0.0026	0.0008	0.0028	0.0020	0.0029	0.0250	0.0002	0.0001	814.6177	-4460.3963	8926.7927	5275.0140	-10544.0280
dji	0.0200	0.0034	0.0206	0.0259	0.2862	0.0044	0.0011	0.0006	0.0034	0.0044	0.0735	0.0006	0.0003	-192.6409	-1199.9248	2405.8497	1007.2839	-2008.5678
dji	0.0300	0.0015	0.0250	0.0375	0.2795	0.0070	0.0036	0.0003	0.0067	0.0095	0.1392	0.0018	0.0015	-155.5071	-401.0666	808.1332	245.5595	-485.1190
ge	0.0100	0.0458	0.0097	0.0141	0.1947	0.0013	0.0007	0.0074	0.0012	0.0017	0.0179	0.0004	0.0001	1704.8436	-7167.0766	14340.1532	8871.9202	-17737.8404
ge	0.0200	0.0112	0.0128	0.0163	0.3673	0.0032	0.0011	0.0020	0.0016	0.0023	0.0395	0.0004	0.0002	-271.1383	-3228.8776	6463.7552	2957.7393	-5909.4785
ge	0.0300	0.0043	0.0155	0.0193	0.4911	0.0064	0.0016	0.0008	0.0024	0.0031	0.0673	0.0009	0.0004	-428.7171	-1409.1085	2824.2169	980.3914	-1954.7827
ibm	0.0100	0.0499	0.0107	0.0160	0.2071	0.0019	0.0007	0.0076	0.0013	0.0020	0.0186	0.0004	0.0001	1778.7137	-7300.0832	14606.1664	9078.7970	-18151.5939
ibm	0.0200	0.0129	0.0161	0.0213	0.3514	0.0043	0.0012	0.0020	0.0020	0.0028	0.0369	0.0005	0.0002	-304.0184	-3355.0389	6716.0777	3051.0205	-6096.0409
ibm	0.0300	0.0059	0.0187	0.0264	0.5367	0.0072	0.0014	0.0009	0.0031	0.0046	0.0710	0.0010	0.0005	-477.3931	-1546.3070	3098.6140	1068.9139	-2131.8278
coco	0.0100	0.0330	0.0082	0.0107	0.2376	0.0024	0.0004	0.0068	0.0010	0.0014	0.0197	0.0004	0.0001	1560.3163	-6920.2431	13846.4862	8480.5594	-16955.1187
coco	0.0200	0.0117	0.0139	0.0185	0.4278	0.0057	0.0005	0.0020	0.0021	0.0032	0.0436	0.0006	0.0002	-347.0579	-3053.4808	6112.9617	2706.4229	-5406.8458
coco	0.0300	0.0063	0.0136	0.0220	0.6649	0.0071	0.0009	0.0011	0.0028	0.0055	0.0885	0.0010	0.0005	-480.3185	-1348.3667	2702.7334	868.0482	-1730.0963
bo	0.0100	0.0880	0.0071	0.0133	0.1514	0.0028	0.0006	0.0125	0.0011	0.0019	0.0150	0.0006	0.0001	2071.3433	-8573.4560	17152.9119	10644.7992	-21283.5985
bo	0.0200	0.0231	0.0110	0.0150	0.2598	0.0047	0.0009	0.0042	0.0015	0.0022	0.0268	0.0007	0.0001	-140.9137	-4993.6284	9993.2568	4852.7147	-9699.4295
bo	0.0300	0.0100	0.0140	0.0186	0.3755	0.0088	0.0007	0.0018	0.0022	0.0033	0.0459	0.0010	0.0003	-614.5693	-2714.9655	5435.9310	2100.3962	-4194.7923
gm	0.0100	0.0480	0.0084	0.0117	0.1196	0.0013	0.0006	0.0087	0.0010	0.0014	0.0164	0.0003	0.0001	1883.4816	-7896.3325	15798.6650	9779.8142	-19553.6283
gm	0.0200	0.0141	0.0120	0.0148	0.1656	0.0034	0.0013	0.0027	0.0016	0.0021	0.0269	0.0005	0.0002	-203.6431	-4145.6424	8297.2849	3941.9993	-7877.9987
gm	0.0300	0.0052	0.0143	0.0167	0.2556	0.0057	0.0019	0.0010	0.0022	0.0028	0.0466	0.0007	0.0004	-473.0173	-2072.8726	4151.7452	1599.8553	-3193.7107

Table B.6.: Parameter estimates, standard errors, and maximum log–likelihoods for the self–exciting POT model. The ground process does not account for possible mark influence, the marks are assumed to be predictable (GPD with conditional scale parameter).

dataset	threshold	$\hat{\mu}$	$\hat{\phi}$	$\hat{\gamma}$	$\hat{\delta}$	$\hat{\xi}$	$\hat{\beta}$	$\hat{\alpha}$	$se_{\hat{\mu}}$	$se_{\hat{\phi}}$	$se_{\hat{\gamma}}$	$se_{\hat{\delta}}$	$se_{\hat{\xi}}$	$se_{\hat{\beta}}$	$se_{\hat{\alpha}}$	ll_{max}	$ll_{max,times}$	AIC_{times}	$ll_{max,marks}$	AIC_{marks}
dji	0.01	0.03	0.01	0.03	62.80	0.08	0.00	0.00	0.00	0.00	0.00	17.62	0.02	0.00	0.00	834.67	-4458.02	8924.04	5292.68	-10579.37
dji	0.02	0.00	0.02	0.03	29.83	0.27	0.00	0.00	0.00	0.00	0.00	14.35	0.07	0.00	0.00	-188.27	-1198.28	2404.57	1010.01	-2014.02
dji	0.03	0.00	0.02	0.04	0.12	0.28	0.01	0.00	0.00	0.01	0.01	9.16	0.14	0.00	0.00	-155.51	-401.05	810.10	245.54	-485.09
ge	0.01	0.07	0.01	0.01	35.83	0.19	0.00	0.00	0.01	0.00	0.00	10.16	0.02	0.00	0.00	1731.14	-7157.73	14323.46	8888.87	-17771.74
ge	0.02	0.01	0.01	0.02	5.84	0.37	0.00	0.00	0.00	0.00	0.00	2.90	0.04	0.00	0.00	-267.91	-3226.48	6460.96	2958.57	-5911.13
ge	0.03	0.00	0.02	0.02	0.00	0.49	0.01	0.00	0.00	0.00	0.00	0.00	0.07	0.00	0.00	-428.72	-1409.11	2826.22	980.39	-1954.78
ibm	0.01	0.05	0.01	0.01	17.92	0.24	0.01	0.00	0.01			7.76	0.02			1703.06	-7294.82	14597.63	8997.88	-17989.75
ibm	0.02	0.01	0.02	0.02	1.38	0.35	0.00	0.00	0.00	0.00	0.00	1.90	0.04	0.00	0.00	-303.69	-3354.30	6716.60	3050.61	-6095.23
ibm	0.03	0.01	0.02	0.03	0.12	0.54	0.01	0.00	0.00	0.00	0.00	1.01	0.07	0.00	0.00	-477.39	-1546.25	3100.50	1068.87	-2131.73
coco	0.01	0.04	0.01	0.01	16.98	0.25	0.01	0.00	0.01	0.00	0.00	6.53	0.02	0.00	0.00	1510.35	-6913.80	13835.59	8424.14	-16842.28
coco	0.02	0.01	0.01	0.02	2.88	0.43	0.01	0.00	0.00	0.00	0.00	2.09	0.04	0.00	0.00	-345.77	-3052.43	6112.85	2706.66	-5407.31
coco	0.03	0.01	0.01	0.02	0.40	0.66	0.01	0.00	0.00	0.00	0.01	1.31	0.09	0.00	0.00	-480.27	-1348.29	2704.57	868.02	-1730.04
bo	0.01	0.12	0.00	0.01	116.00	0.15	0.00	0.00	0.01	0.00	0.00	61.82	0.01	0.00	0.00	2104.38	-8568.13	17144.25	10672.51	-21339.02
bo	0.02	0.02	0.01	0.01	10.07	0.26	0.00	0.00	0.00	0.00	0.00	4.66	0.03	0.00	0.00	-136.93	-4990.98	9989.96	4854.05	-9702.10
bo	0.03	0.01	0.01	0.02	3.38	0.38	0.01	0.00	0.00	0.00	0.00	2.99	0.05	0.00	0.00	-613.70	-2713.98	5435.96	2100.28	-4194.57
gm	0.01	0.11	0.00	0.02	114.43	0.08	0.00	0.00	0.01	0.00	0.00	41.54	0.01	0.00	0.00	1967.89	-7892.03	15792.06	9859.92	-19713.84
gm	0.02	0.02	0.01	0.02	19.34	0.15	0.00	0.00	0.00	0.00	0.00	5.57	0.03	0.00	0.00	-188.13	-4142.34	8292.69	3954.21	-7902.42
gm	0.03	0.01	0.01	0.02	6.64	0.25	0.01	0.00	0.00	0.00	0.00	3.54	0.05	0.00	0.00	-470.29	-2072.35	4152.71	1602.07	-3198.14

Table B.7.: Parameter estimates, standard errors, and maximum log–likelihoods for the self–exciting POT model. The ground process accounts for possible mark influence, the marks are assumed to be predictable (GPD with conditional scale parameter).

B.2. Losses, thresholds equalling 0.90 (0.95, 0.99) quantile

dataset	threshold	$\hat{\tau}$	$\hat{\xi}$	$\hat{\beta}$	$se_{\hat{\tau}}$	$se_{\hat{\xi}}$	$se_{\hat{\beta}}$	ll_{max}	$ll_{max,times}$	AIC_{times}	$ll_{max,marks}$	AIC_{marks}
DJIA	0.0106	0.0689	0.1996	0.0058	0.0020	0.0316	0.0002	344.0475	-4369.2867	8740.5733	4713.3342	-9422.6684
DJIA	0.0149	0.0345	0.3012	0.0057	0.0014	0.0519	0.0004	-291.8444	-2598.4031	5198.8061	2306.5586	-4609.1173
DJIA	0.0254	0.0069	0.3052	0.0100	0.0006	0.1164	0.0015	-317.8107	-711.2037	1424.4075	393.3931	-782.7861
GE	0.0167	0.0689	0.4890	0.0078	0.0020	0.0473	0.0004	-300.1047	-4369.2867	8740.5733	4069.1820	-8134.3639
GE	0.0233	0.0344	0.4717	0.0092	0.0014	0.0554	0.0006	-685.8716	-2595.0352	5192.0703	1909.1635	-3814.3271
GE	0.0441	0.0069	0.7890	0.0139	0.0006	0.1538	0.0023	-415.0895	-711.2037	1424.4075	296.1143	-588.2285
IBM	0.0172	0.0689	0.3265	0.0086	0.0020	0.0330	0.0004	-286.5033	-4369.2867	8740.5733	4082.7833	-8161.5667
IBM	0.0243											
IBM	0.0431											
CO	0.0163	0.0688	0.3522	0.0078	0.0020	0.0329	0.0003	-202.8435	-4363.9355	8729.8709	4161.0919	-8318.1839
CO	0.0227	0.0345	0.4871	0.0079	0.0014	0.0533	0.0005	-599.7934	-2598.4031	5198.8061	1998.6097	-3993.2193
CO	0.0395	0.0069	0.9238	0.0113	0.0006	0.1678	0.0019	-406.8636	-711.2037	1424.4075	304.3401	-604.6802
BO	0.0227	0.0689	0.2927	0.0110	0.0020	0.0322	0.0005	-542.6851	-4369.2867	8740.5733	3826.6016	-7649.2032
BO	0.0315	0.0344	0.3860	0.0115	0.0014	0.0496	0.0007	-762.4833	-2595.0352	5192.0703	1832.5519	-3661.1037
BO	0.0552											
GM	0.0200	0.0689	0.3305	0.0106	0.0020	0.0356	0.0005	-546.2576	-4369.2867	8740.5733	3823.0290	-7642.0581
GM	0.0285	0.0345	0.4356	0.0115	0.0014	0.0573	0.0008	-793.9387	-2598.4031	5198.8061	1804.4644	-3604.9288
GM	0.0531	0.0069	0.5110	0.0228	0.0006	0.1460	0.0038	-440.8576	-711.2037	1424.4075	270.3461	-536.6922

Table B.8.: POT parameter estimates, standard errors, and log–likelihood evaluated at the MLEs for the three thresholds and six datasets. Note the somewhat peculiar fact of positive log–likelihoods for some of the models. A positive log-likelihood is acceptable in a model with continuous response because the probability density function can be greater than one, which implies that the log–likelihood can be positive.

dataset	threshold	$\hat{\tau}$	$\hat{\phi}$	$\hat{\gamma}$	$\hat{\xi}$	$\hat{\beta}$	$se_{\hat{\tau}}$	$se_{\hat{\phi}}$	$se_{\hat{\gamma}}$	$se_{\hat{\xi}}$	$se_{\hat{\beta}}$	ll_{max}	$ll_{max,times}$	AIC_{times}	$ll_{max,marks}$	AIC_{times}
DJIA	0.0106	0.0153	0.0163	0.0207	0.1996	0.0058	0.0025	0.0021	0.0030	0.0316	0.0002	574.5890	-4138.7452	8283.4904	4713.3342	-9422.6684
DJIA	0.0149	0.0074	0.0170	0.0213	0.3012	0.0057	0.0013	0.0024	0.0033	0.0519	0.0004	-75.3198	-2381.8785	4769.7570	2306.5586	-4609.1173
DJIA	0.0254	0.0019	0.0195	0.0259	0.3052	0.0100	0.0004	0.0043	0.0060	0.1164	0.0015	-200.9221	-594.3151	1194.6303	393.3931	-782.7861
GE	0.0167	0.0139	0.0115	0.0142	0.4890	0.0078	0.0029	0.0017	0.0023	0.0473	0.0004	-99.7300	-4168.9120	8343.8239	4069.1820	-8134.3639
GE	0.0233	0.0074	0.0121	0.0151	0.4717	0.0092	0.0013	0.0017	0.0022	0.0554	0.0006	-469.4123	-2378.5759	4763.1517	1909.1635	-3814.3271
GE	0.0441	0.0015	0.0142	0.0165	0.7890	0.0139	0.0004	0.0032	0.0041	0.1538	0.0023	-288.1183	-584.2326	1174.4652	296.1143	-588.2285
IBM	0.0172	0.0171	0.0133	0.0176	0.3265	0.0086	0.0028	0.0018	0.0026	0.0330	0.0004	-114.1476	-4196.9309	8399.8619	4082.7833	-8161.5667
IBM	0.0243															
IBM	0.0431															
CO	0.0163	0.0136	0.0092	0.0114	0.3522	0.0078	0.0030	0.0013	0.0017	0.0329	0.0003	-49.5837	-4210.6757	8427.3514	4161.0919	-8318.1839
CO	0.0227	0.0087	0.0127	0.0168	0.4871	0.0079	0.0016	0.0021	0.0031	0.0533	0.0005	-442.9373	-2441.5470	4889.0940	1998.6097	-3993.2193
CO	0.0395	0.0037	0.0135	0.0289	0.9238	0.0113	0.0006	0.0040	0.0098	0.1678	0.0019	-367.1841	-671.5242	1349.0485	304.3401	-604.6802
BO	0.0227	0.0160	0.0102	0.0131	0.2927	0.0110	0.0034	0.0017	0.0024	0.0322	0.0005	-411.7776	-4238.3792	8482.7583	3826.6016	-7649.2032
BO	0.0315	0.0091	0.0128	0.0172	0.3860	0.0115	0.0017	0.0021	0.0032	0.0496	0.0007	-634.1081	-2466.6600	4939.3200	1832.5519	-3661.1037
BO	0.0552															
GM	0.0200	0.0147	0.0128	0.0160	0.3305	0.0106	0.0028	0.0019	0.0027	0.0356	0.0005	-322.6134	-4145.6424	8297.2849	3823.0290	-7642.0581
GM	0.0285	0.0058	0.0153	0.0179	0.4356	0.0115	0.0011	0.0023	0.0030	0.0573	0.0008	-506.6245	-2311.0890	4628.1779	1804.4644	-3604.9288
GM	0.0531	0.0007	0.0073	0.0063	0.5110	0.0228	0.0003	0.0018	0.0020	0.1460	0.0038	-281.4757	-551.8218	1109.6436	270.3461	-536.6922

Table B.9.: Parameter estimates, standard errors, and maximum log–likelihoods for the self–exciting POT model. The ground process exhibits no mark influence, the marks are assumed to be unpredictable (iid GPD).

dataset	threshold	$\hat{\tau}$	$\hat{\phi}$	$\hat{\gamma}$	$\hat{\delta}$	$\hat{\xi}$	$\hat{\beta}$	$se_{\hat{\tau}}$	$se_{\hat{\phi}}$	$se_{\hat{\gamma}}$	$se_{\hat{\delta}}$	$se_{\hat{\xi}}$	$se_{\hat{\beta}}$	ll_{max}	$ll_{max,times}$	AIC_{times}	$ll_{max,marks}$	AIC_{marks}
DJIA	0.0106	0.0170	0.0144	0.0214	17.7443	0.1996	0.0058	0.0027	0.0022	0.0032	10.4937	0.0316	0.0002	576.6845	-4136.6497	8281.2993	4713.3342	-9422.6684
DJIA	0.0149	0.0080	0.0148	0.0220	19.9063	0.3012	0.0057	0.0013	0.0024	0.0034	11.5645	0.0519	0.0004	-72.9700	-2379.5286	4767.0572	2306.5586	-4609.1173
DJIA	0.0254	0.0019	0.0150	0.0260	20.2554	0.3052	0.0100	0.0004	0.0044	0.0059	17.2907	0.1164	0.0015	-199.5956	-592.9887	1193.9773	393.3931	-782.7861
GE	0.0167	0.0149	0.0103	0.0139	5.5473	0.4890	0.0078	0.0029	0.0017	0.0023	3.3071	0.0473	0.0004	-97.7090	-4166.8910	8341.7820	4069.1820	-8134.3639
GE	0.0233	0.0074	0.0108	0.0145	3.8618	0.4717	0.0092	0.0013	0.0017	0.0022	2.8100	0.0554	0.0006	-468.0346	-2377.1981	4762.3962	1909.1635	-3814.3271
GE	0.0441	0.0015	0.0142	0.0165	0.0000	0.7890	0.0139	0.0004	0.0032	0.0037		0.1538	0.0023	-288.1183	-584.2326	1176.4652	296.1143	-588.2285
IBM	0.0172	0.0172	0.0119	0.0169	4.8743	0.3265	0.0086	0.0028	0.0019	0.0026	3.3180	0.0330	0.0004	-112.4695	-4195.2529	8398.5057	4082.7833	-8161.5667
IBM	0.0243																	
IBM	0.0431																	
CO	0.0163	0.0140	0.0079	0.0108	7.8084	0.3522	0.0078	0.0031	0.0013	0.0016	3.6746	0.0329	0.0003	-45.8988	-4206.9907	8421.9814	4161.0919	-8318.1839
CO	0.0227	0.0086	0.0120	0.0165	2.0899	0.4871	0.0079	0.0016	0.0021	0.0030	1.8791	0.0533	0.0005	-442.1247	-2440.7343	4889.4687	1998.6097	-3993.2193
CO	0.0395	0.0037	0.0135	0.0289	0.0000	0.9238	0.0113	0.0006	0.0040	0.0098	0.0000	0.1678	0.0019	-367.1841	-671.5242	1351.0485	304.3401	-604.6802
BO	0.0227	0.0168	0.0086	0.0130	8.9288	0.2927	0.0110	0.0034	0.0017	0.0024	4.8394	0.0322	0.0005	-408.9273	-4235.5289	8479.0578	3826.6016	-7649.2032
BO	0.0315	0.0090	0.0124	0.0170	1.4778	0.3860	0.0115	0.0017	0.0022	0.0032	2.7590	0.0496	0.0007	-633.9428	-2466.4947	4940.9893	1832.5519	-3661.1037
BO	0.0552																	
GM	0.0200	0.0178	0.0116	0.0175	9.0203	0.3305	0.0106	0.0031	0.0019	0.0031	4.4794	0.0356	0.0005	-319.3142	-4142.3432	8292.6864	3823.0290	-7642.0581
GM	0.0285	0.0061	0.0143	0.0180	3.5102	0.4356	0.0115	0.0012	0.0024	0.0030	3.2259	0.0573	0.0008	-505.8549	-2310.3193	4628.6386	1804.4644	-3604.9288
GM	0.0531	0.0007	0.0073	0.0063	0.0000	0.5110	0.0228	0.0003	0.0018	0.0020	0.0001	0.1460	0.0038	-281.4757	-551.8218	1111.6436	270.3461	-536.6922

Table B.10.: Parameter estimates, standard errors, and maximum log–likelihoods for the self–exciting POT model. The ground process accounts for possible mark influence, the marks are assumed to be unpredictable (iid GPD).

dataset	threshold	$\hat{\tau}$	$\hat{\phi}$	$\hat{\gamma}$	$\hat{\xi}$	$\hat{\beta}$	$\hat{\alpha}$	$se_{\hat{\tau}}$	$se_{\hat{\phi}}$	$se_{\hat{\gamma}}$	$se_{\hat{\xi}}$	$se_{\hat{\beta}}$	$se_{\hat{\alpha}}$	ll_{max}	$ll_{max,times}$	AIC_{times}	$ll_{max,marks}$	AIC_{marks}
DJIA	0.0106	0.0166	0.0185	0.0241	0.1135	0.0028	0.0008	0.0025	0.0020	0.0029	0.0262	0.0002	0.0001	649.4257	-4138.7452	8283.4904	4788.1709	-9570.3418
DJIA	0.0149	0.0079	0.0189	0.0241	0.1813	0.0028	0.0012	0.0013	0.0024	0.0033	0.0412	0.0003	0.0002	-27.2756	-2381.8785	4769.7570	2354.6029	-4703.2059
DJIA	0.0254	0.0019	0.0217	0.0294	0.2544	0.0069	0.0021	0.0004	0.0048	0.0068	0.1043	0.0013	0.0009	-195.3080	-594.3151	1194.6303	399.0071	-792.0142
GE	0.0167	0.0147	0.0124	0.0155	0.2852	0.0032	0.0009	0.0028	0.0015	0.0021	0.0292	0.0004	0.0001	-19.1177	-4168.9120	8343.8239	4149.7942	-8293.5885
GE	0.0233	0.0080	0.0137	0.0174	0.4026	0.0044	0.0013	0.0013	0.0018	0.0024	0.0477	0.0006	0.0002	-431.2937	-2378.5759	4763.1517	1947.2822	-3888.5644
GE	0.0441	0.0015	0.0148	0.0174	0.8198	0.0092	0.0016	0.0004	0.0034	0.0045	0.1521	0.0023	0.0008	-285.2336	-584.2326	1174.4652	298.9990	-591.9980
IBM	0.0172	0.0182	0.0149	0.0201	0.2995	0.0039	0.0010	0.0027	0.0018	0.0026	0.0298	0.0004	0.0002	-66.1636	-4196.9309	8399.8619	4130.7674	-8255.5348
IBM	0.0243																	
IBM	0.0431																	
CO	0.0163	0.0152	0.0108	0.0137	0.3424	0.0043	0.0005	0.0031	0.0014	0.0020	0.0317	0.0005	0.0001	-24.7921	-4210.6757	8427.3514	4185.8836	-8365.7673
CO	0.0227	0.0091	0.0137	0.0184	0.4841	0.0061	0.0005	0.0016	0.0022	0.0033	0.0526	0.0007	0.0002	-437.9670	-2441.5470	4889.0940	2003.5800	-4001.1601
CO	0.0395	0.0036	0.0132	0.0279	0.9390	0.0093	0.0023	0.0006	0.0039	0.0094	0.1689	0.0022	0.0020	-366.2134	-671.5242	1349.0485	305.3109	-604.6218
BO	0.0227	0.0176	0.0117	0.0155	0.2858	0.0057	0.0010	0.0033	0.0016	0.0024	0.0306	0.0007	0.0002	-383.9930	-4238.3792	8482.7583	3854.3862	-7702.7723
BO	0.0315	0.0093	0.0132	0.0179	0.3886	0.0092	0.0007	0.0017	0.0022	0.0033	0.0490	0.0010	0.0003	-630.9368	-2466.6600	4939.3200	1835.7232	-3665.4464
BO	0.0552																	
GM	0.0200	0.0141	0.0120	0.0148	0.1655	0.0034	0.0013	0.0027	0.0016	0.0021	0.0269	0.0005	0.0002	-203.8478	-4145.6424	8297.2849	3941.7946	-7877.5892
GM	0.0285	0.0058	0.0153	0.0180	0.2455	0.0055	0.0018	0.0011	0.0022	0.0028	0.0442	0.0007	0.0003	-448.3793	-2311.0890	4628.1779	1862.7097	-3719.4193
GM	0.0531	0.0008	0.0081	0.0073	0.3790	0.0128	0.0019	0.0003	0.0021	0.0023	0.1103	0.0027	0.0006	-270.7118	-551.8218	1109.6436	281.1100	-556.2200

Table B.11.: Parameter estimates, standard errors, and maximum log–likelihoods for the self–exciting POT model. The ground process does not account for possible mark influence, the marks are assumed to be predictable (GPD with conditional scale parameter).

dataset	threshold	$\hat{\tau}$	$\hat{\phi}$	$\hat{\gamma}$	$\hat{\delta}$	$\hat{\xi}$	$\hat{\beta}$	$\hat{\alpha}$	$se_{\hat{\tau}}$	$se_{\hat{\phi}}$	$se_{\hat{\gamma}}$	$se_{\hat{\delta}}$	$se_{\hat{\xi}}$	$se_{\hat{\beta}}$	$se_{\hat{\alpha}}$	ll_{max}	$ll_{max,times}$	AIC_{times}	$ll_{max,marks}$	AIC_{marks}
DJIA	0.01	0.02	0.01	0.03	54.05	0.09	0.00	0.00	0.00	0.00	0.00	15.76	0.02	0.00	0.00	665.54	-4136.65	8281.30	4802.19	-9598.37
DJIA	0.01	0.01	0.01	0.03	34.56	0.17	0.00	0.00	0.00	0.00	0.00	12.77	0.04	0.00	0.00	-19.32	-2379.53	4767.06	2360.20	-4714.41
DJIA	0.03	0.00	0.02	0.03	22.61	0.25	0.01	0.00	0.00	0.00	0.01	17.75	0.10	0.00	0.00	-193.61	-592.99	1193.98	399.38	-792.76
GE	0.02	0.02	0.01	0.01	7.97	0.28	0.00	0.00	0.00	0.00	0.00	3.32	0.03	0.00	0.00	-14.29	-4166.89	8341.78	4152.60	-8299.20
GE	0.02	0.01	0.01	0.02	2.58	0.40	0.00	0.00	0.00	0.00	0.00	2.31	0.05	0.00	0.00	-430.46	-2377.20	4762.40	1946.74	-3887.48
GE	0.04	0.00	0.01	0.02	0.00	0.82	0.01	0.00	0.00	0.00	0.00		0.15	0.00	0.00	-285.23	-584.23	1176.47	299.00	-592.00
IBM	0.02	0.02	0.01	0.02	3.42	0.30	0.00	0.00	0.00	0.00	0.00	2.51	0.03	0.00	0.00	-64.82	-4195.25	8398.51	4130.44	-8254.87
IBM	0.02																			
IBM	0.04																			
CO	0.02	0.02	0.01	0.01	6.05	0.34	0.00	0.00	0.00	0.00	0.00	2.95	0.03	0.00	0.00	-21.58	-4206.99	8421.98	4185.41	-8364.82
CO	0.02	0.01	0.01	0.02	1.86	0.48	0.01	0.00	0.00	0.00	0.00	1.77	0.05	0.00	0.00	-437.25	-2440.73	4889.47	2003.48	-4000.96
CO	0.04	0.00	0.01	0.03	0.00	0.94	0.01	0.00	0.00	0.00	0.01	0.00	0.17	0.00	0.00	-366.21	-671.52	1351.05	305.31	-604.62
BO	0.02	0.02	0.01	0.02	7.21	0.29	0.01	0.00	0.00	0.00	0.00	3.96	0.03	0.00	0.00	-381.40	-4235.53	8479.06	3854.13	-7702.27
BO	0.03	0.01	0.01	0.02	1.44	0.39	0.01	0.00	0.00	0.00	0.00	2.71	0.05	0.00	0.00	-630.77	-2466.49	4940.99	1835.72	-3665.44
BO	0.06																			
GM	0.02	0.02	0.01	0.02	19.34	0.15	0.00	0.00	0.00	0.00	0.00	5.57	0.03	0.00	0.00	-188.34	-4142.34	8292.69	3954.00	-7902.01
GM	0.03	0.01	0.01	0.02	7.40	0.24	0.01	0.00	0.00	0.00	0.00	3.57	0.04	0.00	0.00	-444.96	-2310.32	4628.64	1865.36	-3724.72
GM	0.05	0.00	0.01	0.01	0.00	0.38	0.01	0.00	0.00	0.00	0.00		0.11	0.00	0.00	-270.71	-551.82	1111.64	281.11	-556.22

Table B.12.: Parameter estimates, standard errors, and maximum log–likelihoods for the self–exciting POT model. The ground process accounts for possible mark influence, the marks are assumed to be predictable (GPD with conditional scale parameter).

B.3. Gains, thresholds 0.01, 0.02, 0.03

dataset	threshold	$\hat{\tau}$	$\hat{\xi}$	$\hat{\beta}$	$se_{\hat{\tau}}$	$se_{\hat{\xi}}$	$se_{\hat{\beta}}$	ll_{max}	$ll_{max,times}$	AIC_{times}	$ll_{max,marks}$	AIC_{marks}
DJIA	0.0100	0.0794	0.1726	0.0057	0.0021	0.0312	0.0002	637.5587	-4837.7579	9677.5158	5475.3166	-10946.6331
DJIA	0.0200	0.0170	0.2279	0.0069	0.0010	0.0743	0.0006	-388.7315	-1487.1077	2976.2153	1098.3762	-2192.7524
DJIA	0.0300	0.0050	0.1914	0.0093	0.0005	0.1250	0.0015	-243.1773	-547.2054	1096.4108	304.0281	-604.0562
GE	0.0100											
GE	0.0200	0.0560	0.1878	0.0093	0.0018	0.0373	0.0005	-377.7171	-3750.4579	7502.9157	3372.7407	-6741.4815
GE	0.0300	0.0217	0.2480	0.0100	0.0011	0.0647	0.0008	-553.0677	-1806.9322	3615.8645	1253.8645	-2503.7291
IBM	0.0100											
IBM	0.0200											
IBM	0.0300											
CO	0.0100											
CO	0.0200	0.0529	0.1058	0.0100	0.0018	0.0339	0.0005	-394.7469	-3596.2061	7194.4122	3201.4593	-6398.9185
CO	0.0300	0.0205	0.1472	0.0104	0.0011	0.0570	0.0008	-517.3805	-1725.8725	3453.7451	1208.4920	-2412.9840
BO	0.0100											
BO	0.0200											
BO	0.0300											
GM	0.0100	0.1619	0.1825	0.0110	0.0031	0.0201	0.0003	1425.0510	-7878.3654	15758.7309	9303.4165	-18602.8329
GM	0.0200	0.0726	0.3020	0.0105	0.0021	0.0359	0.0005	-454.8814	-4536.1561	9074.3121	4081.2746	-8158.5492
GM	0.0300	0.0319	0.3498	0.0124	0.0014	0.0561	0.0009	-771.6858	-2445.1389	4892.2778	1673.4531	-3342.9062

Table B.13.: POT parameter estimates, standard errors, and log–likelihood evaluated at the MLEs for the three thresholds and six datasets. Note the somewhat peculiar fact of positive log–likelihoods for some of the models. A positive log-likelihood is acceptable in a model with continuous response because the probability density function can be greater than one, which implies that the log–likelihood can be positive.

dataset	threshold	$\hat{\tau}$	$\hat{\phi}$	$\hat{\gamma}$	$\hat{\xi}$	$\hat{\beta}$	$se_{\hat{\tau}}$	$se_{\hat{\phi}}$	$se_{\hat{\gamma}}$	$se_{\hat{\xi}}$	$se_{\hat{\beta}}$	ll_{max}	$ll_{max,times}$	AIC_{times}	$ll_{max,marks}$	AIC_{times}
DJIA	0.0100	0.0103	0.0085	0.0096	0.1726	0.0057	0.0030	0.0011	0.0013	0.0312	0.0002	813.4475	-4661.8691	9329.7382	5475.3166	-10946.6331
DJIA	0.0200	0.0033	0.0128	0.0154	0.2279	0.0069	0.0007	0.0020	0.0025	0.0743	0.0006	-239.2811	-1337.6573	2681.3146	1098.3762	-2192.7524
DJIA	0.0300	0.0017	0.0193	0.0278	0.1914	0.0093	0.0004	0.0051	0.0072	0.1250	0.0015	-166.6998	-470.7279	947.4559	304.0281	-604.0562
GE	0.0100															
GE	0.0200	0.0105	0.0088	0.0106	0.1878	0.0093	0.0024	0.0013	0.0017	0.0373	0.0005	-227.0291	-3599.7699	7205.5397	3372.7407	-6741.4815
GE	0.0300	0.0045	0.0104	0.0127	0.2480	0.0100	0.0010	0.0016	0.0021	0.0647	0.0008	-413.6962	-1667.5607	3341.1215	1253.8645	-2503.7291
IBM	0.0100															
IBM	0.0200															
IBM	0.0300															
CO	0.0100															
CO	0.0200	0.0121	0.0077	0.0098	0.1058	0.0100	0.0028	0.0012	0.0016	0.0339	0.0005	-294.7251	-3496.1843	6998.3687	3201.4593	-6398.9185
CO	0.0300	0.0061	0.0094	0.0132	0.1472	0.0104	0.0012	0.0018	0.0028	0.0570	0.0008	-426.1703	-1634.6623	3275.3246	1208.4920	-2412.9840
BO	0.0100															
BO	0.0200															
BO	0.0300															
GM	0.0100	0.0495	0.0033	0.0046	0.1825	0.0110	0.0099	0.0007	0.0010	0.0201	0.0003	1479.5005	-7823.9160	15653.8319	9303.4165	-18602.8329
GM	0.0200	0.0097	0.0088	0.0100	0.3020	0.0105	0.0025	0.0013	0.0015	0.0359	0.0005	-251.3053	-4332.5799	8671.1598	4081.2746	-8158.5492
GM	0.0300	0.0043	0.0087	0.0097	0.3498	0.0124	0.0012	0.0014	0.0017	0.0561	0.0009	-603.7771	-2277.2302	4560.4604	1673.4531	-3342.9062

Table B.14.: Parameter estimates, standard errors, and maximum log–likelihoods for the self–exciting POT model. The ground process exhibits no mark influence, the marks are assumed to be unpredictable (iid GPD).

dataset	threshold	$\hat{\tau}$	$\hat{\phi}$	$\hat{\gamma}$	$\hat{\delta}$	$\hat{\xi}$	$\hat{\beta}$	$se_{\hat{\tau}}$	$se_{\hat{\phi}}$	$se_{\hat{\gamma}}$	$se_{\hat{\delta}}$	$se_{\hat{\xi}}$	$se_{\hat{\beta}}$	ll_{max}	$ll_{max,times}$	AIC_{times}	$ll_{max,marks}$	AIC_{marks}
DJIA	0.0100	0.0159	0.0059	0.0103	63.1322	0.1726	0.0057	0.0036	0.0012	0.0014	25.2460	0.0312	0.0002	821.8937	-4653.4229	9314.8458	5475.3166	-10946.6331
DJIA	0.0200	0.0042	0.0051	0.0161	168.5121	0.2279	0.0069	0.0009	0.0019	0.0027	89.1934	0.0743	0.0006	-225.3578	-1323.7340	2655.4679	1098.3762	-2192.7524
DJIA	0.0300	0.0017	0.0088	0.0273	99.4008	0.1914	0.0093	0.0004	0.0046	0.0072	80.7478	0.1250	0.0015	-162.8903	-466.9184	941.8368	304.0281	-604.0562
GE	0.0100																	
GE	0.0200	0.0156	0.0049	0.0114	65.6265	0.1878	0.0093	0.0029	0.0013	0.0019	29.8879	0.0373	0.0005	-216.6488	-3589.3896	7186.7792	3372.7407	-6741.4815
GE	0.0300	0.0060	0.0055	0.0141	72.5354	0.2480	0.0100	0.0012	0.0017	0.0025	38.3607	0.0647	0.0008	-405.4656	-1659.3301	3326.6602	1253.8645	-2503.7291
IBM	0.0100																	
IBM	0.0200																	
IBM	0.0300																	
CO	0.0100																	
CO	0.0200	0.0196	0.0033	0.0121	117.0703	0.1058	0.0100	0.0034	0.0014	0.0021	73.2872	0.0339	0.0005	-284.1394	-3485.5987	6979.1974	3201.4593	-6398.9185
CO	0.0300	0.0079	0.0022	0.0144	252.2115	0.1472	0.0104	0.0013	0.0018	0.0029	252.1499	0.0570	0.0008	-413.6017	-1622.0937	3252.1875	1208.4920	-2412.9840
BO	0.0100																	
BO	0.0200																	
BO	0.0300																	
GM	0.0100	0.0940	0.0003	0.0039	356.8210	0.1825	0.0110	0.0138	0.0006	0.0010	715.3534	0.0201	0.0003	1496.5036	-7806.9128	15621.8256	9303.4165	-18602.8329
GM	0.0200	0.0122	0.0069	0.0100	17.2434	0.3020	0.0105	0.0028	0.0013	0.0016	7.2095	0.0359	0.0005	-245.3783	-4326.6529	8661.3059	4081.2746	-8158.5492
GM	0.0300	0.0052	0.0066	0.0095	16.0490	0.3498	0.0124	0.0014	0.0014	0.0018	7.1218	0.0561	0.0009	-598.3533	-2271.8064	4551.6128	1673.4531	-3342.9062

Table B.15.: Parameter estimates, standard errors, and maximum log–likelihoods for the self–exciting POT model. The ground process accounts for possible mark influence, the marks are assumed to be unpredictable (iid GPD).

dataset	threshold	$\hat{\tau}$	$\hat{\phi}$	$\hat{\gamma}$	$\hat{\xi}$	$\hat{\beta}$	$\hat{\alpha}$	$se_{\hat{\tau}}$	$se_{\hat{\phi}}$	$se_{\hat{\gamma}}$	$se_{\hat{\xi}}$	$se_{\hat{\beta}}$	$se_{\hat{\alpha}}$	ll_{max}	$ll_{max,times}$	AIC_{times}	$ll_{max,marks}$	AIC_{marks}
DJIA	0.0100	0.0112	0.0096	0.0110	0.1207	0.0024	0.0004	0.0030	0.0011	0.0013	0.0282	0.0003	0.0001	860.1197	-4661.8691	9329.7382	5521.9888	-11037.9776
DJIA	0.0200	0.0034	0.0133	0.0160	0.2031	0.0051	0.0007	0.0007	0.0021	0.0026	0.0728	0.0007	0.0003	-233.9059	-1337.6573	2681.3146	1103.7514	-2201.5027
DJIA	0.0300	0.0017	0.0208	0.0304	0.1106	0.0074	0.0025	0.0004	0.0055	0.0080	0.1235	0.0015	0.0015	-164.6094	-470.7279	947.4559	306.1185	-606.2370
GE	0.0100																	
GE	0.0200	0.0116	0.0103	0.0127	0.1049	0.0045	0.0009	0.0024	0.0014	0.0018	0.0354	0.0005	0.0002	-184.4347	-3599.7699	7205.5397	3415.3352	-6824.6703
GE	0.0300	0.0046	0.0107	0.0132	0.1251	0.0057	0.0017	0.0010	0.0016	0.0021	0.0623	0.0008	0.0004	-394.4266	-1667.5607	3341.1215	1273.1341	-2540.2683
IBM	0.0100																	
IBM	0.0200																	
IBM	0.0300																	
CO	0.0100																	
CO	0.0200	0.0126	0.0082	0.0107	0.0757	0.0053	0.0008	0.0028	0.0012	0.0016	0.0319	0.0007	0.0002	-270.7530	-3496.1843	6998.3687	3225.4314	-6444.8627
CO	0.0300	0.0061	0.0096	0.0135	0.1316	0.0076	0.0010	0.0012	0.0018	0.0028	0.0538	0.0010	0.0004	-420.6121	-1634.6623	3275.3246	1214.0502	-2422.1004
BO	0.0100																	
BO	0.0200																	
BO	0.0300																	
GM	0.0100	0.0561	0.0053	0.0080	0.1185	0.0009	0.0005	0.0105	0.0008	0.0011	0.0190	0.0004	0.0001	1584.5934	-7823.9160	15653.8319	9408.5094	-18811.0187
GM	0.0200	0.0096	0.0088	0.0099	0.1798	0.0043	0.0008	0.0025	0.0011	0.0013	0.0324	0.0005	0.0001	-191.9582	-4332.5799	8671.1598	4140.6217	-8275.2435
GM	0.0300	0.0042	0.0083	0.0092	0.1710	0.0051	0.0016	0.0012	0.0012	0.0014	0.0447	0.0008	0.0003	-550.4587	-2277.2302	4560.4604	1726.7715	-3447.5429

Table B.16.: Parameter estimates, standard errors, and maximum log–likelihoods for the self–exciting POT model. The ground process does not account for possible mark influence, the marks are assumed to be predictable (GPD with conditional scale parameter).

dataset	threshold	$\hat{\tau}$	$\hat{\phi}$	$\hat{\gamma}$	$\hat{\delta}$	$\hat{\xi}$	$\hat{\beta}$	$\hat{\alpha}$	$se_{\hat{\tau}}$	$se_{\hat{\phi}}$	$se_{\hat{\gamma}}$	$se_{\hat{\delta}}$	$se_{\hat{\xi}}$	$se_{\hat{\beta}}$	$se_{\hat{\alpha}}$	ll_{max}	$ll_{max,times}$	AIC_{times}	$ll_{max,marks}$	AIC_{marks}
DJIA	0.01	0.02	0.00	0.01	153.94	0.09	0.00	0.00	0.00	0.00	0.00	51.73	0.03	0.00	0.00	892.14	-4653.42	9314.85	5545.56	-11085.12
DJIA	0.02	0.00	0.00	0.02	199.56	0.18	0.01	0.00	0.00	0.00	0.00	107.69	0.07	0.00	0.00	-216.68	-1323.73	2655.47	1107.06	-2208.11
DJIA	0.03	0.00	0.01	0.03	91.26	0.11	0.01	0.00	0.00	0.01	0.01	73.09	0.12	0.00	0.00	-160.82	-466.92	941.84	306.09	-606.19
GE	0.01																			
GE	0.02	0.02	0.00	0.01	116.99	0.07	0.01	0.00	0.00	0.00	0.00	49.47	0.03	0.00	0.00	-158.35	-3589.39	7186.78	3431.04	-6856.08
GE	0.03	0.01	0.01	0.01	75.81	0.11	0.01	0.00	0.00	0.00	0.00	37.67	0.06	0.00	0.00	-383.96	-1659.33	3326.66	1275.37	-2544.74
IBM	0.01																			
IBM	0.02																			
IBM	0.03																			
CO	0.01																			
CO	0.02	0.02	0.00	0.01	147.94	0.07	0.01	0.00	0.00	0.00	0.00	89.76	0.03	0.00	0.00	-253.82	-3485.60	6979.20	3231.78	-6457.55
CO	0.03	0.01	0.00	0.01	222.28	0.13	0.01	0.00	0.00	0.00	0.00	202.55	0.05	0.00	0.00	-407.83	-1622.09	3252.19	1214.27	-2422.53
BO	0.01																			
BO	0.02																			
BO	0.03																			
GM	0.01	0.09	0.00	0.00	356.82	0.18	0.01	0.00	0.01	0.00	0.00	59.05	0.04	0.00	0.00	1510.71	-7806.91	15621.83	9317.62	-18629.24
GM	0.02	0.02	0.00	0.01	53.55	0.12	0.01	0.00	0.00	0.00	0.00	15.23	0.03	0.00	0.00	-155.57	-4326.65	8661.31	4171.08	-8336.16
GM	0.03	0.01	0.00	0.01	33.40	0.12	0.01	0.00	0.00	0.00	0.00	10.70	0.04	0.00	0.00	-530.68	-2271.81	4551.61	1741.12	-3476.25

Table B.17.: Parameter estimates, standard errors, and maximum log–likelihoods for the self–exciting POT model. The ground process accounts for possible mark influence, the marks are assumed to be predictable (GPD with conditional scale parameter).

B.4. Gains, thresholds equalling 0.90 (0.95, 0.99) quantile

dataset	threshold	$\hat{\tau}$	$\hat{\xi}$	$\hat{\beta}$	$se_{\hat{\tau}}$	$se_{\hat{\xi}}$	$se_{\hat{\beta}}$	ll_{max}	$ll_{max,times}$	AIC_{times}	$ll_{max,marks}$	AIC_{marks}
DJIA	0.0108	0.0689	0.1809	0.0057	0.0020	0.0340	0.0003	364.5756	-4369.2867	8740.5733	4733.8622	-9463.7244
DJIA	0.0152	0.0345	0.1903	0.0063	0.0014	0.0478	0.0004	-295.2013	-2598.4031	5198.8061	2303.2017	-4602.4035
DJIA	0.0266	0.0069	0.1455	0.0096	0.0006	0.0988	0.0013	-294.9143	-711.2037	1424.4075	416.2894	-828.5788
GE	0.0179	0.0689	0.1654	0.0093	0.0020	0.0322	0.0004	-193.9791	-4369.2867	8740.5733	4175.3076	-8346.6151
GE	0.0248	0.0345	0.2072	0.0099	0.0014	0.0488	0.0006	-568.9059	-2598.4031	5198.8061	2029.4972	-4054.9944
GE	0.0418											
IBM	0.0180	0.0689	0.1607	0.0098	0.0020	0.0330	0.0004	-247.3036	-4369.2867	8740.5733	4121.9831	-8239.9662
IBM	0.0253	0.0345	0.1884	0.0105	0.0014	0.0491	0.0007	-594.7770	-2598.4031	5198.8061	2003.6260	-4003.2521
IBM	0.0451											
CO	0.0173	0.0689	0.0867	0.0100	0.0020	0.0284	0.0004	-183.2432	-4366.6115	8735.2230	4183.3682	-8362.7365
CO	0.0245	0.0345	0.1190	0.0102	0.0014	0.0421	0.0006	-531.9913	-2598.4031	5198.8061	2066.4118	-4128.8236
CO	0.0424	0.0069	0.2118	0.0111	0.0006	0.1061	0.0015	-320.1066	-711.2037	1424.4075	391.0972	-778.1943
BO	0.0243	0.0689	0.0910	0.0133	0.0020	0.0313	0.0006	-529.7806	-4369.2867	8740.5733	3839.5061	-7675.0121
BO	0.0336	0.0345	0.0567	0.0149	0.0014	0.0384	0.0008	-723.6426	-2598.4031	5198.8061	1874.7604	-3745.5209
BO	0.0584	0.0069	0.1727	0.0144	0.0006	0.1069	0.0020	-346.5632	-711.2037	1424.4075	364.6406	-725.2811
GM	0.0205	0.0689	0.2989	0.0107	0.0020	0.0366	0.0005	-514.4901	-4369.2867	8740.5733	3854.7965	-7705.5930
GM	0.0288	0.0345	0.3300	0.0125	0.0014	0.0523	0.0008	-780.7553	-2598.4031	5198.8061	1817.6478	-3631.2956
GM	0.0534	0.0069	0.4807	0.0193	0.0006	0.1621	0.0035	-417.5364	-711.2037	1424.4075	293.6674	-583.3348

Table B.18.: POT parameter estimates, standard errors, and log–likelihood evaluated at the MLEs for the three thresholds and six datasets. Note the somewhat peculiar fact of positive log–likelihoods for some of the models. A positive log-likelihood is acceptable in a model with continuous response because the probability density function can be greater than one, which implies that the log–likelihood can be positive.

dataset	threshold	$\hat{\tau}$	$\hat{\phi}$	$\hat{\gamma}$	$\hat{\xi}$	$\hat{\beta}$	$se_{\hat{\tau}}$	$se_{\hat{\phi}}$	$se_{\hat{\gamma}}$	$se_{\hat{\xi}}$	$se_{\hat{\beta}}$	ll_{max}	$ll_{max,times}$	AIC_{times}	$ll_{max,marks}$	AIC_{times}
DJIA	0.0108	0.0097	0.0092	0.0106	0.1809	0.0057	0.0026	0.0012	0.0015	0.0340	0.0003	537.7010	-4196.1612	8398.3224	4733.8622	-9463.7244
DJIA	0.0152	0.0062	0.0111	0.0133	0.1903	0.0063	0.0013	0.0017	0.0021	0.0478	0.0004	-122.3837	-2425.5854	4857.1708	2303.2017	-4602.4035
DJIA	0.0266	0.0021	0.0215	0.0299	0.1455	0.0096	0.0004	0.0053	0.0077	0.0988	0.0013	-192.3429	-608.6323	1223.2646	416.2894	-828.5788
GE	0.0179	0.0136	0.0091	0.0112	0.1654	0.0093	0.0030	0.0013	0.0017	0.0322	0.0004	-40.4175	-4215.7251	8437.4501	4175.3076	-8346.6151
GE	0.0248	0.0064	0.0096	0.0115	0.2072	0.0099	0.0014	0.0014	0.0018	0.0488	0.0006	-407.3793	-2436.8765	4879.7530	2029.4972	-4054.9944
GE	0.0418															
IBM	0.0180	0.0142	0.0089	0.0111	0.1607	0.0098	0.0030	0.0012	0.0016	0.0330	0.0004	-103.9476	-4225.9307	8457.8615	4121.9831	-8239.9662
IBM	0.0253	0.0071	0.0095	0.0118	0.1884	0.0105	0.0015	0.0015	0.0019	0.0491	0.0007	-451.1066	-2454.7326	4915.4652	2003.6260	-4003.2521
IBM	0.0451															
CO	0.0173	0.0160	0.0065	0.0084	0.0867	0.0100	0.0038	0.0010	0.0014	0.0284	0.0004	-91.0121	-4274.3804	8554.7608	4183.3682	-8362.7365
CO	0.0245	0.0085	0.0082	0.0107	0.1190	0.0102	0.0018	0.0013	0.0018	0.0421	0.0006	-429.0501	-2495.4619	4996.9237	2066.4118	-4128.8236
CO	0.0424	0.0022	0.0094	0.0136	0.2118	0.0111	0.0005	0.0021	0.0031	0.1061	0.0015	-261.4463	-652.5435	1311.0869	391.0972	-778.1943
BO	0.0243	0.0173	0.0065	0.0086	0.0910	0.0133	0.0039	0.0010	0.0014	0.0313	0.0006	-447.8303	-4287.3364	8580.6727	3839.5061	-7675.0121
BO	0.0336	0.0084	0.0082	0.0106	0.0567	0.0149	0.0018	0.0015	0.0021	0.0384	0.0008	-631.6557	-2506.4161	5018.8323	1874.7604	-3745.5209
BO	0.0584	0.0025	0.0050	0.0075	0.1727	0.0144	0.0007	0.0016	0.0027	0.1069	0.0020	-318.7601	-683.4006	1372.8013	364.6406	-725.2811
GM	0.0205	0.0090	0.0088	0.0099	0.2989	0.0107	0.0024	0.0013	0.0015	0.0366	0.0005	-309.9954	-4164.7919	8335.5838	3854.7965	-7705.5930
GM	0.0288	0.0050	0.0093	0.0105	0.3300	0.0125	0.0014	0.0015	0.0018	0.0523	0.0008	-616.0400	-2433.6878	4873.3756	1817.6478	-3631.2956
GM	0.0534	0.0012	0.0106	0.0111	0.4807	0.0193	0.0003	0.0023	0.0024	0.1621	0.0035	-284.9197	-578.5871	1163.1742	293.6674	-583.3348

Table B.19.: Parameter estimates, standard errors, and maximum log–likelihoods for the self–exciting POT model. The ground process exhibits no mark influence, the marks are assumed to be unpredictable (iid GPD).

dataset	threshold	$\hat{\tau}$	$\hat{\phi}$	$\hat{\gamma}$	$\hat{\delta}$	$\hat{\xi}$	$\hat{\beta}$	$se_{\hat{\tau}}$	$se_{\hat{\phi}}$	$se_{\hat{\gamma}}$	$se_{\hat{\delta}}$	$se_{\hat{\xi}}$	$se_{\hat{\beta}}$	ll_{max}	$ll_{max,times}$	AIC_{times}	$ll_{max,marks}$	AIC_{marks}
DJIA	0.0108	0.0146	0.0064	0.0116	64.2254	0.1809	0.0057	0.0032	0.0013	0.0017	25.1580	0.0340	0.0003	546.5357	-4187.3266	8382.6531	4733.8622	-9463.7244
DJIA	0.0152	0.0083	0.0064	0.0147	101.9488	0.1903	0.0063	0.0015	0.0016	0.0024	43.1643	0.0478	0.0004	-111.0294	-2414.2312	4836.4623	2303.2017	-4602.4035
DJIA	0.0266	0.0023	0.0095	0.0321	119.9545	0.1455	0.0096	0.0005	0.0045	0.0082	84.1105	0.0988	0.0013	-186.5788	-602.8682	1213.7364	416.2894	-828.5788
GE	0.0179	0.0205	0.0047	0.0118	74.9136	0.1654	0.0093	0.0034	0.0014	0.0019	33.6885	0.0322	0.0004	-28.0510	-4203.3586	8414.7172	4175.3076	-8346.6151
GE	0.0248	0.0090	0.0049	0.0126	80.8951	0.2072	0.0099	0.0017	0.0015	0.0020	40.0630	0.0488	0.0006	-396.9630	-2426.4602	4860.9205	2029.4972	-4054.9944
GE	0.0418																	
IBM	0.0180	0.0249	0.0023	0.0127	227.9853	0.1607	0.0098	0.0036	0.0015	0.0019	178.1554	0.0330	0.0004	-85.5092	-4207.4923	8422.9845	4121.9831	-8239.9662
IBM	0.0253	0.0096	0.0040	0.0135	112.6478	0.1884	0.0105	0.0016	0.0016	0.0021	63.7021	0.0491	0.0007	-439.7661	-2443.3922	4894.7843	2003.6260	-4003.2521
IBM	0.0451																	
CO	0.0173	0.0284	0.0023	0.0109	165.1237	0.0867	0.0100	0.0048	0.0013	0.0020	126.9078	0.0284	0.0004	-79.2009	-4262.5691	8533.1382	4183.3682	-8362.7365
CO	0.0245	0.0133	0.0024	0.0131	209.0250	0.1190	0.0102	0.0022	0.0016	0.0024	177.9379	0.0421	0.0006	-416.3030	-2482.7148	4973.4296	2066.4118	-4128.8236
CO	0.0424	0.0024	0.0057	0.0139	44.9049	0.2118	0.0111	0.0005	0.0022	0.0031	34.3229	0.1061	0.0015	-258.6192	-649.7163	1307.4327	391.0972	-778.1943
BO	0.0243	0.0226	0.0030	0.0086	63.4317	0.0910	0.0133	0.0043	0.0012	0.0015	38.8063	0.0313	0.0006	-440.8567	-4280.3628	8568.7256	3839.5061	-7675.0121
BO	0.0336	0.0090	0.0038	0.0096	57.5278	0.0567	0.0149	0.0019	0.0015	0.0019	34.5397	0.0384	0.0008	-625.1668	-2499.9273	5007.8546	1874.7604	-3745.5209
BO	0.0584	0.0024	0.0028	0.0061	28.1862	0.1727	0.0144	0.0007	0.0018	0.0025	32.6921	0.1069	0.0020	-317.6871	-682.3277	1372.6554	364.6406	-725.2811
GM	0.0205	0.0114	0.0069	0.0100	17.1901	0.2989	0.0107	0.0026	0.0013	0.0016	7.1816	0.0366	0.0005	-303.9898	-4158.7863	8325.5726	3854.7965	-7705.5930
GM	0.0288	0.0063	0.0068	0.0105	19.4391	0.3300	0.0125	0.0016	0.0014	0.0019	8.1361	0.0523	0.0008	-609.1274	-2426.7752	4861.5503	1817.6478	-3631.2956
GM	0.0534	0.0013	0.0052	0.0100	30.9313	0.4807	0.0193	0.0004	0.0022	0.0027	19.5366	0.1621	0.0035	-279.7211	-573.3885	1154.7770	293.6674	-583.3348

Table B.20.: Parameter estimates, standard errors, and maximum log–likelihoods for the self–exciting POT model. The ground process accounts for possible mark influence, the marks are assumed to be unpredictable (iid GPD).

dataset	threshold	$\hat{\tau}$	$\hat{\phi}$	$\hat{\gamma}$	$\hat{\xi}$	$\hat{\beta}$	$\hat{\alpha}$	$se_{\hat{\tau}}$	$se_{\hat{\phi}}$	$se_{\hat{\gamma}}$	$se_{\hat{\xi}}$	$se_{\hat{\beta}}$	$se_{\hat{\alpha}}$	ll_{max}	$ll_{max,times}$	AIC_{times}	$ll_{max,marks}$	AIC_{marks}
DJIA	0.0108	0.0101	0.0098	0.0113	0.1269	0.0025	0.0004	0.0026	0.0012	0.0014	0.0306	0.0003	0.0001	583.7804	-4196.1612	8398.3224	4779.9416	-9553.8833
DJIA	0.0152	0.0066	0.0121	0.0146	0.1446	0.0035	0.0008	0.0013	0.0017	0.0021	0.0445	0.0004	0.0001	-97.1711	-2425.5854	4857.1708	2328.4143	-4650.8287
DJIA	0.0266	0.0021	0.0214	0.0298	0.1040	0.0086	0.0010	0.0004	0.0052	0.0076	0.1000	0.0014	0.0009	-191.5713	-608.6323	1223.2646	417.0610	-828.1220
GE	0.0179	0.0143	0.0099	0.0123	0.0997	0.0045	0.0007	0.0029	0.0013	0.0017	0.0309	0.0005	0.0001	0.8670	-4215.7251	8437.4501	4216.5920	-8427.1841
GE	0.0248	0.0067	0.0103	0.0124	0.1095	0.0056	0.0011	0.0014	0.0014	0.0018	0.0465	0.0007	0.0002	-381.0557	-2436.8765	4879.7530	2055.8208	-4105.6415
GE	0.0418																	
IBM	0.0180	0.0141	0.0089	0.0111	0.1280	0.0051	0.0006	0.0030	0.0012	0.0015	0.0315	0.0006	0.0001	-73.7103	-4225.9307	8457.8615	4152.2205	-8298.4409
IBM	0.0253	0.0070	0.0094	0.0116	0.1641	0.0067	0.0008	0.0015	0.0014	0.0018	0.0479	0.0008	0.0002	-438.7600	-2454.7326	4915.4652	2015.9726	-4025.9452
IBM	0.0451																	
CO	0.0173	0.0173	0.0074	0.0099	0.0677	0.0048	0.0006	0.0038	0.0011	0.0015	0.0273	0.0006	0.0001	-60.4267	-4274.3804	8554.7608	4213.9537	-8421.9073
CO	0.0245	0.0088	0.0087	0.0115	0.1013	0.0063	0.0009	0.0018	0.0014	0.0020	0.0400	0.0008	0.0002	-415.2155	-2495.4619	4996.9237	2080.2463	-4154.4926
CO	0.0424	0.0022	0.0095	0.0138	0.2163	0.0101	0.0006	0.0005	0.0022	0.0032	0.1057	0.0020	0.0009	-261.1722	-652.5435	1311.0869	391.3713	-776.7425
BO	0.0243	0.0184	0.0073	0.0098	0.0910	0.0084	0.0006	0.0039	0.0011	0.0016	0.0301	0.0010	0.0001	-434.5510	-4287.3364	8580.6727	3852.7854	-7699.5708
BO	0.0336	0.0085	0.0084	0.0110	0.0769	0.0127	0.0004	0.0019	0.0015	0.0021	0.0405	0.0014	0.0003	-630.2109	-2506.4161	5018.8323	1876.2052	-3746.4104
BO	0.0584	0.0025	0.0052	0.0079	0.2035	0.0102	0.0025	0.0007	0.0015	0.0025	0.1103	0.0026	0.0015	-317.1069	-683.4006	1372.8013	366.2938	-726.5876
GM	0.0205	0.0090	0.0088	0.0099	0.1769	0.0045	0.0008	0.0024	0.0011	0.0013	0.0329	0.0006	0.0001	-254.1315	-4164.7919	8335.5838	3910.6603	-7815.3207
GM	0.0288	0.0045	0.0081	0.0089	0.1605	0.0051	0.0015	0.0013	0.0012	0.0014	0.0420	0.0008	0.0002	-562.0219	-2433.6878	4873.3756	1871.6659	-3737.3317
GM	0.0534	0.0012	0.0102	0.0105	0.1983	0.0115	0.0035	0.0003	0.0021	0.0022	0.1203	0.0026	0.0010	-270.6021	-578.5871	1163.1742	307.9850	-609.9700

Table B.21.: Parameter estimates, standard errors, and maximum log–likelihoods for the self–exciting POT model. The ground process does not account for possible mark influence, the marks are assumed to be predictable (GPD with conditional scale parameter).

dataset	threshold	$\hat{\tau}$	$\hat{\phi}$	$\hat{\gamma}$	$\hat{\delta}$	$\hat{\xi}$	$\hat{\beta}$	$\hat{\alpha}$	$se_{\hat{\tau}}$	$se_{\hat{\phi}}$	$se_{\hat{\gamma}}$	$se_{\hat{\delta}}$	$se_{\hat{\xi}}$	$se_{\hat{\beta}}$	$se_{\hat{\alpha}}$	ll_{max}	$ll_{max,times}$	AIC_{times}	$ll_{max,marks}$	AIC_{marks}
DJIA	0.01	0.02	0.01	0.01	140.87	0.10	0.00	0.00	0.00	0.00	0.00	46.86	0.03	0.00	0.00	612.58	-4187.33	8382.65	4799.90	-9593.81
DJIA	0.02	0.01	0.01	0.02	133.32	0.13	0.00	0.00	0.00	0.00	0.00	52.34	0.04	0.00	0.00	-78.37	-2414.23	4836.46	2335.86	-4665.72
DJIA	0.03	0.00	0.01	0.03	126.75	0.09	0.01	0.00	0.00	0.00	0.01	88.28	0.10	0.00	0.00	-185.33	-602.87	1213.74	417.53	-829.07
GE	0.02	0.02	0.00	0.01	166.62	0.06	0.01	0.00	0.00	0.00	0.00	79.00	0.03	0.00	0.00	35.29	-4203.36	8414.72	4238.65	-8471.30
GE	0.02	0.01	0.00	0.01	98.49	0.08	0.01	0.00	0.00	0.00	0.00	46.04	0.05	0.00	0.00	-365.45	-2426.46	4860.92	2061.01	-4116.02
GE	0.04																			
IBM	0.02	0.03	0.00	0.01	244.56	0.12	0.01	0.00	0.00	0.00	0.00	183.16	0.03	0.00	0.00	-50.48	-4207.49	8422.98	4157.02	-8308.03
IBM	0.03	0.01	0.00	0.01	99.51	0.16	0.01	0.00	0.00	0.00	0.00	53.32	0.05	0.00	0.00	-427.77	-2443.39	4894.78	2015.62	-4025.24
IBM	0.05																			
CO	0.02	0.03	0.00	0.01	244.05	0.06	0.01	0.00	0.00	0.00	0.00	197.04	0.03	0.00	0.00	-38.87	-4262.57	8533.14	4223.70	-8441.40
CO	0.02	0.01	0.00	0.01	180.29	0.10	0.01	0.00	0.00	0.00	0.00	135.41	0.04	0.00	0.00	-401.20	-2482.71	4973.43	2081.52	-4157.03
CO	0.04	0.00	0.01	0.01	55.45	0.21	0.01	0.00	0.00	0.00	0.00	40.21	0.10	0.00	0.00	-257.53	-649.72	1307.43	392.19	-778.38
BO	0.02	0.03	0.00	0.01	140.77	0.08	0.01	0.00	0.00	0.00	0.00	102.18	0.03	0.00	0.00	-419.70	-4280.36	8568.73	3860.66	-7715.32
BO	0.03	0.01	0.00	0.01	84.04	0.07	0.01	0.00	0.00	0.00	0.00	52.47	0.04	0.00	0.00	-621.10	-2499.93	5007.85	1878.83	-3751.66
BO	0.06	0.00	0.00	0.01	26.64	0.20	0.01	0.00	0.00	0.00	0.00	28.04	0.11	0.00	0.00	-315.88	-682.33	1372.66	366.45	-726.89
GM	0.02	0.02	0.00	0.01	53.44	0.11	0.01	0.00	0.00	0.00	0.00	15.31	0.03	0.00	0.00	-217.59	-4158.79	8325.57	3941.19	-7876.38
GM	0.03	0.01	0.00	0.01	38.73	0.11	0.01	0.00	0.00	0.00	0.00	12.66	0.04	0.00	0.00	-539.91	-2426.78	4861.55	1886.87	-3767.74
GM	0.05	0.00	0.00	0.01	33.60	0.18	0.01	0.00	0.00	0.00	0.00	20.29	0.12	0.00	0.00	-264.24	-573.39	1154.78	309.15	-612.30

Table B.22.: Parameter estimates, standard errors, and maximum log–likelihoods for the self–exciting POT model. The ground process accounts for possible mark influence, the marks are assumed to be predictable (GPD with conditional scale parameter).

C Likelihood ratio tests

C.1. Losses, thresholds 0.01, 0.02, 0.03

Question 1: Is the intensity time–dependent?

dataset	threshold	pvalue
DJIA	0.01	0.00
DJIA	0.02	0.00
DJIA	0.03	0.00
GE	0.01	0.00
GE	0.02	0.00
GE	0.03	0.00
IBM	0.01	0.00
IBM	0.02	0.00
IBM	0.03	0.00
KO	0.01	0.00
KO	0.02	0.00
KO	0.03	0.00
BA	0.01	0.00
BA	0.02	0.00
BA	0.03	0.00
GM	0.01	0.00
GM	0.02	0.00
GM	0.03	0.00

Table C.1.: P–values for the likelihood ratio tests for time–varying intensities. The model under the null hypothesis is the homogeneous Poisson model, the alternative model is the simplest self–exciting ground model (simple Hawkes model without mark influence). In this case, the LR distribution exhibits two degrees of freedom.

Question 2: Do exceedance sizes affect the intensity?

dataset	threshold	pvalue
dji	0.01	0.03
dji	0.02	0.07
dji	0.03	0.86
ge	0.01	0.00
ge	0.02	0.03
ge	0.03	1.00
ibm	0.01	0.00
ibm	0.02	0.22
ibm	0.03	0.74
coco	0.01	0.00
coco	0.02	0.15
coco	0.03	0.69
bo	0.01	0.00
bo	0.02	0.02
bo	0.03	0.16
gm	0.01	0.00
gm	0.02	0.01
gm	0.03	0.31

Table C.2.: P–values for the likelihood ratio tests for the presence of an influence of mark sizes on the intensity. Therefore, the simple Hawkes model without mark influence is tested against the simple Hawkes model with mark influence (represented by the parameter delta). For the lowest of the investigated thresholds such an influence clearly exists, contrary to the the highest threshold. Evidence for $u = 0.02$ is not definite.

Question 3: Are exceedances time–dependent?

dataset	threshold	pvalue
dji	0.01	0.00
dji	0.02	0.00
dji	0.03	0.00
ge	0.01	0.00
ge	0.02	0.00
ge	0.03	0.00
ibm	0.01	0.00
ibm	0.02	0.00
ibm	0.03	0.00
coco	0.01	0.00
coco	0.02	0.00
coco	0.03	0.03
bo	0.01	0.00
bo	0.02	0.00
bo	0.03	0.00
gm	0.01	0.00
gm	0.02	0.00
gm	0.03	0.00

Table C.3.: P–values for the likelihood ratio tests for the predictability property using the marks part of the likelihood which is conditionally independent of the times part (ground model without mark influence). For all datasets and thresholds, there is clear evidence in favour of the predictability property, i.e. the excess distribution appears to be time–dependent.

C.2. Losses, thresholds equalling 0.90 (0.95, 0.99) quantile

Question 1: Is the intensity time–dependent?

dataset	threshold	pvalue
DJIA	0.01	0.00
DJIA	0.01	0.00
DJIA	0.03	0.00
GE	0.02	0.00
GE	0.02	0.00
GE	0.04	0.00
IBM	0.02	0.00
CO	0.02	0.00
CO	0.02	0.00
CO	0.04	0.00
BO	0.02	0.00
BO	0.03	0.00
GM	0.02	0.00
GM	0.03	0.00
GM	0.05	0.00

Table C.4.: P–values for the likelihood ratio tests for time–varying intensities. The model under the null hypothesis is the homogeneous Poisson model, the alternative model is the simplest self–exciting ground model (simple Hawkes model without mark influence). In this case, the LR distribution exhibits two degrees of freedom.

Question 2: Do exceedance sizes affect the intensity?

dataset	threshold	pvalue
DJIA	0.01	0.04
DJIA	0.01	0.03
DJIA	0.03	0.10
GE	0.02	0.04
GE	0.02	0.10
GE	0.04	1.00
IBM	0.02	0.07
CO	0.02	0.01
CO	0.02	0.20
CO	0.04	1.00
BO	0.02	0.02
BO	0.03	0.57
GM	0.02	0.01
GM	0.03	0.21
GM	0.05	1.00

Table C.5.: P–values for the likelihood ratio tests for the presence of an influence of mark sizes on the intensity. Therefore, the simple Hawkes model without mark influence is tested against the simple Hawkes model with mark influence (represented by the parameter delta). For the lowest of the investigated thresholds such an influence clearly exists, contrary to the the highest threshold. Evidence for $u = 0.02$ is not definite.

Question 3: Are exceedances time–dependent?

dataset	threshold	pvalue
DJIA	0.01	0.00
DJIA	0.01	0.00
DJIA	0.03	0.00
GE	0.02	0.00
GE	0.02	0.00
GE	0.04	0.02
IBM	0.02	0.00
CO	0.02	0.00
CO	0.02	0.00
CO	0.04	0.16
BO	0.02	0.00
BO	0.03	0.01
GM	0.02	0.00
GM	0.03	0.00
GM	0.05	0.00

Table C.6.: P–values for the likelihood ratio tests for the predictability property using the marks part of the likelihood which is conditionally independent of the times part (ground model without mark influence). For all datasets and thresholds, there is clear evidence in favour of the predictability property, i.e. the excess distribution appears to be time–dependent.

C.3. Gains, thresholds 0.01, 0.02, 0.03

Question 1: Is the intensity time–dependent?

dataset	threshold	pvalue
DJIA	0.01	0.00
DJIA	0.02	0.00
DJIA	0.03	0.00
GE	0.02	0.00
GE	0.03	0.00
CO	0.02	0.00
CO	0.03	0.00
GM	0.01	0.00
GM	0.02	0.00
GM	0.03	0.00

Table C.7.: P–values for the likelihood ratio tests for time–varying intensities. The model under the null hypothesis is the homogeneous Poisson model, the alternative model is the simplest self–exciting ground model (simple Hawkes model without mark influence). In this case, the LR distribution exhibits two degrees of freedom.

Question 2: Do exceedance sizes affect the intensity?

dataset	threshold	pvalue
DJIA	0.01	0.00
DJIA	0.02	0.00
DJIA	0.03	0.01
GE	0.02	0.00
GE	0.03	0.00
CO	0.02	0.00
CO	0.03	0.00
GM	0.01	0.00
GM	0.02	0.00
GM	0.03	0.00

Table C.8.: P–values for the likelihood ratio tests for the presence of an influence of mark sizes on the intensity. Therefore, the simple Hawkes model without mark influence is tested against the simple Hawkes model with mark influence (represented by the parameter delta). For the lowest of the investigated thresholds such an influence clearly exists, contrary to the the highest threshold. Evidence for $u = 0.02$ is not definite.

Question 3: Are exceedances time–dependent?

dataset	threshold	pvalue
DJIA	0.01	0.00
DJIA	0.02	0.00
DJIA	0.03	0.04
GE	0.02	0.00
GE	0.03	0.00
CO	0.02	0.00
CO	0.03	0.00
GM	0.01	0.00
GM	0.02	0.00
GM	0.03	0.00

Table C.9.: P–values for the likelihood ratio tests for the predictability property using the marks part of the likelihood which is conditionally independent of the times part (ground model without mark influence). For all datasets and thresholds, there is clear evidence in favour of the predictability property, i.e. the excess distribution appears to be time–dependent.

C.4. Gains, thresholds equalling 0.90 (0.95, 0.99) quantile

Question 1: Is the intensity time–dependent?

dataset	threshold	pvalue
DJIA	0.01	0.00
DJIA	0.02	0.00
DJIA	0.03	0.00
GE	0.02	0.00
GE	0.02	0.00
IBM	0.02	0.00
IBM	0.03	0.00
CO	0.02	0.00
CO	0.02	0.00
CO	0.04	0.00
BO	0.02	0.00
BO	0.03	0.00
BO	0.06	0.00
GM	0.02	0.00
GM	0.03	0.00
GM	0.05	0.00

Table C.10.: P–values for the likelihood ratio tests for time–varying intensities. The model under the null hypothesis is the homogeneous Poisson model, the alternative model is the simplest self–exciting ground model (simple Hawkes model without mark influence). In this case, the LR distribution exhibits two degrees of freedom.

Question 2: Do exceedance sizes affect the intensity?

dataset	threshold	pvalue
DJIA	0.01	0.00
DJIA	0.02	0.00
DJIA	0.03	0.00
GE	0.02	0.00
GE	0.02	0.00
IBM	0.02	0.00
IBM	0.03	0.00
CO	0.02	0.00
CO	0.02	0.00
CO	0.04	0.02
BO	0.02	0.00
BO	0.03	0.00
BO	0.06	0.14
GM	0.02	0.00
GM	0.03	0.00
GM	0.05	0.00

Table C.11.: P–values for the likelihood ratio tests for the presence of an influence of mark sizes on the intensity. Therefore, the simple Hawkes model without mark influence is tested against the simple Hawkes model with mark influence (represented by the parameter delta). For the lowest of the investigated thresholds such an influence clearly exists, contrary to the the highest threshold. Evidence for $u = 0.02$ is not definite.

Question 3: Are exceedances time–dependent?

dataset	threshold	pvalue
DJIA	0.01	0.00
DJIA	0.02	0.00
DJIA	0.03	0.21
GE	0.02	0.00
GE	0.02	0.00
IBM	0.02	0.00
IBM	0.03	0.00
CO	0.02	0.00
CO	0.02	0.00
CO	0.04	0.46
BO	0.02	0.00
BO	0.03	0.09
BO	0.06	0.07
GM	0.02	0.00
GM	0.03	0.00
GM	0.05	0.00

Table C.12.: P–values for the likelihood ratio tests for the predictability property using the marks part of the likelihood which is conditionally independent of the times part (ground model without mark influence). For all datasets and thresholds, there is clear evidence in favour of the predictability property, i.e. the excess distribution appears to be time–dependent.

D

Properties of the homogeneous Poisson process

To illustrate the connection of exponentially distributed durations and Poisson distributed event numbers within some period A, we start with the gamma distribution. The density of the gamma (Pearson–type II) distribution with shape parameter $\alpha > 0$ and (inverse) scale parameter β can be written

$$g_{\alpha,\beta}(x) = x^{\alpha-1}\frac{\beta^{\alpha}e^{-\beta x}}{\Gamma(\alpha)}, \qquad (x > 0). \tag{D.1}$$

The exponential density is a special case for $\alpha = 1$.

Gamma distribution functions $G_{\alpha,\beta}$ are *sum–reproductive*, i.e. the sum of m independent gamma rvs with parameters α_i, β_i is a gamma rv with parameters $\alpha_1 + \ldots + \alpha_m$ and $\beta_1 + \ldots + \beta_m$. Moreover, gamma dfs approach the normal df as $\alpha \to \infty$.

For positive integers $\alpha = n+1$, the sum–reproductivity property leads to the distribution of the sum of $n+1$ iid exponential rvs X_i. When assuming $\beta_i = \beta$, the gamma df can be written

$$G_{n+1,\beta}(x) = 1 - e^{-\beta x}\sum_{i=0}^{n}\frac{(\beta x)^i}{i!}, \qquad (x > 0), \tag{D.2}$$

and is also called the Erlang distribution in this special case.

We now consider the homogeneous Poisson process $N(t) = \{\text{number of events in } (0,t]\}$ with parameter λ, whereas t_i denotes the time of the ith event. From (D.2) we can deduce

$$\begin{aligned} P(N(t) = k) &= P(t_k \le t, t_{k+1} > t) \\ &= P(t_k \le t) - P(t_{k+1} \le t) \\ &= \frac{(\lambda t)^k}{k!}e^{-\lambda t}. \end{aligned}$$

Hence, the number of events $N(t)$ is a Poisson rv with parameter λt. The expected number of events $E[N(t)] = \lambda t$ grows linearly with the width of the observed interval.

Bibliography

Artzner, D. J., F. Delbain, J. M. Eber, and D. Heath (1999): "Coherent Measures of Risk," *Mathematical Finance*, 9, 203–228.

Barberis, N., and R. Thaler (2002): "A Survey of Behavioral Finance," *NBER Working Paper no. 9222*.

Bauwens, L., and P. Giot (2001): *Econometric Modelling of Stock Market Intraday Activity*. Kluwer, Dordrecht, The Netherlands.

Bauwens, L., and N. Hautsch (2009): *Modelling Financial High Frequency Data Using Point Processes*pp. 953–981, Handbook of Financial Time Series. Springer.

Beirlant, J., Y. Goegebeur, J. Segers, and J. Teugels (2004): *Statistics of Extremes*. Wiley, Chichester.

Benartzi, S., and R. Thaler (1995): "Myopic Loss Aversion and the Equity Premium Puzzle," *The Quarterly Journal of Economics*, 110(1), 73–92.

Campbell, J. Y., A. W. Lo, and A. C. MacKinlay (1997): *The Econometrics of Financial Markets*. Princeton University Press, Princeton, NJ.

Chavez-Demoulin, V., A. C. Davison, and A. J. McNeil (2005): "Estimating Value-at-Risk: A Point Process Approach," *Quantitative Finance*, 5(2), 227–234.

Christoffersen, P. F. (1998): "Evaluating Interval Forecasts," *International Economic Review*, 39(4), 841–862.

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

Cox, D. R., and V. Isham (1980): *Point Processes*. Chapman and Hall, London.

Daley, D. J., and D. Vere-Jones (2003): *An Introduction to the Theory of Point Processes*, vol. I. Springer, New York, 2nd edn.

——— (2008): *An Introduction to the Theory of Point Processes*, vol. II. Springer, New York, 2nd edn.

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

Engle, R. F. (1982): "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of UK Inflation," *Econometrica*, 50, 987–1008.

Engle, R. F., and J. R. Russell (1994): "Forecasting Transaction Rates: The Autoregressive Conditional Duration Model," *NBER Working Paper no. 4966.*

——— (1997): "Forecasting the Frequency of Changes in Quoted Foreign Exchange Prices with Autoregressive Conditional Duration Model," *Journal of Empirical Finance*, 4, 187–212.

——— (1998): "A New Model for Irregularly Spaced Transaction Data," *Econometrica*, 66(5), 1127–1162.

Fama, E. F. (1970): "Efficient Capital Markets: A Review of Theory and Empirical Work," *Journal of Finance*, 25(2), 383–417.

Ferro, C. A. T., and J. Segers (2003): "Inference for Clusters of Extreme Values," *Journal of the Royal Statistical Society B*, 65, 545–556.

Grammig, J., and K.-O. Maurer (2000): "Non-monotonic hazard functions and the autoregressive conditional duration model," *Econometrics Journal*, 3(1), 16–38.

Gutenberg, B., and C. F. Richter (1954): *Seismicity of the Earth and Associated Phenomena*. Hafner Publishing, New York.

Hautsch, N. (2004): *Modelling Irregularly Spaced Financial Data*, Lecture Notes in Economics and Mathematical Systems. Springer, Berlin.

Hawkes, A. G. (1971): "Spectra of Some Self-Exciting and Mutually Exciting Point Processes," *Biometrika*, 58(1), 83–90.

Hawkes, A. G., and D. Oakes (1974): "A Cluster Process Representation of a Self-Exciting Process," *Journal of Applied Probability*, 11(3), 493–503.

Helmstetter, A., and D. Sornette (2002): "Sub-critical and Super-critical Regimes in Epidemic Models of Earthquake Aftershocks," *Journal of Geophysical Research*, 107(10), 1–21.

——— (2003a): "Importance of direct and indirect triggered seismicity in the ETAS model of seismicity," *Geophysical Research Letters*, 30(11), 1–4.

——— (2003b): "Predictability in the ETAS Model of Interacting Triggered Seismicity," *Journal of Geophysical Research*, 108(10), 1–10.

Karr, A. F. (1991): *Point Processes and Their Statistical Inference*. Marcel Dekker Inc, New York.

Kuester, K., S. Mittnik, and M. S. Paolella (2006): "Value-at-Risk Prediction: A Comparison of Alternative Strategies," *Journal of Financial Econometrics*, 4, 53–89.

McNeil, A. J., R. Frey, and P. Embrechts (2005): *Quantitative Risk Management.* Princeton University Press, Princeton, New Jersey.

Møller, J., and J. G. Rasmussen (2005): "Perfect Simulation of Hawkes Processes," *Adv. Appl. Prob.*, 37, 629–646.

Nocedal, J., and S. J. Wright (1999): *Numerical Optimization.* Springer, New York, USA.

Ogata, Y. (1988): "Statistical Models for Earthquake Occurrences and Residual Analysis for Point Processes," *Journal of the American Statistical Association*, 83, 9–27.

Omori, F. (1894): "On the aftershocks of earthquakes," *Journal of the College of Science, Imperial University of Tokyo*, 7, 111–200.

Papangelou, F. (1972): "Integrability of Expected Increments of Point Processes and a Related Random Change of Scale," *Transactions of the American Mathematical Society*, 165, 483–506.

Peng, R. (2003): "Multi-dimensional Point Process Models in R," *Journal of Statistical Software*, 8(16), 1–27.

Tsay, R. S. (2002): *Analysis of Financial Time Series*, Wiley Series in Probability and Statistics. Wiley, New York.

Venables, W., and B. Ripley (2003): *Modern Applied Statistics with S.* Springer, New York, 4th edn.

Vere-Jones, D. (1970): "Stochastic Models for Earthquake Occurrence," *Journal of the Royal Statistical Society. Series B (Methodological)*, 32(1), 1–62.

Vere-Jones, D., and T. Ozaki (1982): "Some Examples of Statistical Estimation Applied to Earthquake Data," *Annals of the Institute of Statistical Mathematics*, 34, 189–207.

Zeileis, A., and K. Hornik (2007): "Generalized M-Fluctuation Tests for Parameter Instability," *Statistica Neerlandica*, 61, 488–508.

Zhang, M. Y., J. R. Russell, and R. S. Tsay (2001): "A nonlinear autoregressive conditional duration model with applications to financial transaction data," *Journal of Econometrics*, 104, 179–207.

Zivot, E., and J. Wang (2006): *Modeling Financial Time Series with S-Plus.* Springer, New York, 2nd edn.

Declaration of Authorship

I hereby confirm that I have authored this master thesis independently and without use of others than the indicated resources.

Andreas Fuest

München, June 25, 2009