# SELF-EXCITING POINT PROCESS MODELING OF FINANCIAL RISK

**Rodrigo Herrera**
Universidad de Talca
Departamento de Modelación y Gestión Industrial
rodriherrera@utalca.cl

**ABSTRACT**

The analysis of return series from financial markets is often based on the Peaks-over-threshold (POT) model. This model assumes independent and identically distributed observations and therefore a Poisson process is used to characterize the occurrence of extreme events. However, stylized facts such as clustered extremes and serial dependence typically violate the assumption of independence. In this paper we concentrate on an alternative approach to overcome these difficulties. We consider the stochastic intensity of the point process of exceedances over a threshold in the framework of irregularly spaced data. The main idea is to model the time between exceedances through an Autoregressive Conditional Duration (ACD) model, while the marks are still being modeled by generalized Pareto distributions. The main advantage of this approach is its capability to capture the short-term behavior of extremes without involving an arbitrary stochastic volatility model or a prefiltration of the data, which certainly impacts the estimation. We make use of the proposed model to obtain an improved estimate for the Value at Risk. The model is then applied and illustrated to transactions data from Bayer AG, a blue chip stock from the German stock market index DAX.

**KEYWORDS. Extreme value theory, Value at Risk, Selfexciting Point Process.**

**Main area: Financial Management, Probabilistic Models, Statistics**

## 1 Introduction

The current subprime crisis, together with its consequences for international markets, shows that a deeper understanding of extreme events in statistical data from economics, insurance and finance is of high priority. According to the Extreme Value Theory (EVT), extreme events refer, for example, to extraordinary claims to insurance companies, crashes of equity markets or extreme losses in credit portfolios due to borrower defaults. Hence, extreme events occur rarely, i.e., only few extreme observations are available; but probabilities and dependence structures have to be assigned to extreme events due to their economic impact.

The main contribution of this paper from the point of view of extreme value theory is that we can capture the short-term behaviour of extreme events in stock market returns without involving an arbitrary stochastic volatility model or the prefiltration of the data, which certainly impacts the measures of risk. We model the stochastic intensity of the point process of exceedances within the framework of irregularly spaced data. Contrary to the classical Peaks over Threshold (POT) methodology, where the time of occurrence of the extreme events is modeled, the methodology proposed models the inter-exceedance times between extreme events. To this end, we use a methodology similar to an Autoregressive Conditional Duration (ACD) model (see Engle and Russell, 1998 for more reference), while the marks still being modelled by generalized Pareto distributions. Like the GARCH models, the ACD models and their alternatives (see Hautsch, 2004) have proven to be very useful in capturing the clustering effects. Taking into account the dynamic aspects within the cluster of extremes is absolutely necessary if one seeks to exploit the full amount of information in financial data, such as the irregular spacing in time, the discreteness of price changes, the cluster of extremes as well as the presence of strong correlation among the duration time between extreme events and persistent dynamics.

The results of the empirical application to the Bayer stock index show that characteristics associated with previous extreme losses as well the time between these extreme events have a significant impact on the dynamic aspects and size of future extreme events. In a Value at Risk (VaR) context the results of our backtesting procedure, which dynamically adjusts quantiles incorporating the new information daily, allows us to statistically conclude that the models proposed are suitable for the estimation of different risk measures as the VaR, according to the restriction imposed by Basel Committee on Banking Supervision (1996, 2006).

This paper is organized as follows. In section 2 we outline relevant aspects of the classical POT model of EVT, and then we describe the ACD-POT model theory that is central to the paper and discusse a conditional generalized Pareto distribution based approach for the exceedances. In addition, we make use of the models proposed to obtain an expression and its estimate for the VaR one day ahead predictive distribution of the returns, conditionally on the past and current data. In section 3 the models are applied to transactions data from the Bayer index. Conclusions are resumed in section 4.

## 2 The Peaks-Over-Threshold method of extreme value theory

The normal distribution is the important limiting distribution for sample averages as summarized in a central limit theorem. However, we cannot hope that the magnitude of such an extreme event as the crash of 1987 could be modelled by such class of distributions. In fact, under the Gaussian hypothesis for any given stock, an observation more than five standard deviations from the mean should be observed about once every 7,000 years! Fortunately, the family of extreme value distributions is the one to study the limiting distributions of sample extrema. The Fisher and Tippet theorem suggests that the asymptotic distribution of the maxima of iid random variables under some norming constants, the resulting distribution, if it is non-degenerate distribution function, belong to the maximum domain of attraction of $H_{\xi,\mu,\sigma}$, where

$H_{\xi,\mu,\sigma}$ is the generalized extreme value distribution

$$H_{\xi,\mu,\sigma}(y) = \begin{cases} \exp\left\{-\left(1+\xi\frac{y-\mu}{\sigma}\right)^{-1/\xi}\right\} & \xi \neq 0, \\ \exp\left\{-\exp\left(-\frac{y-\mu}{\sigma}\right)\right\} & \xi = 0, \end{cases}$$

where $\xi, \mu \in \mathbb{R}$ and $\sigma > 0$ are the shape location and scale parameter respectively, and $1+\xi y > 0$. Essentially, all the common, continuous distributions of statistics and risk management are in MDA($H_{\xi,\mu,\sigma}$). In this paper the statistical approach is based on viewing the high level of exceedances (extreme events), $Y_i > u$ for a high threshold $u > 0$, as a marked point processes (MPP). In many stochastic process models, a point process arises as the component that carries the information about the events $t$ in time or space of objects that may themselves have a stochastic structure and stochastic dependency relations. We define a MPP $N$ as a set of observations, occurrence times and marks $\{(t_i, y_i)\}$ on the space $\mathcal{T} \times \mathcal{Y}$, whose history $\mathcal{H}_t = (\{t_1, y_1\}, \ldots, \{t_{t-1}, y_{t-1}\})$ consists only of the occurrence times and marks $\{t_1, y_1\}, \ldots, \{t_{t-1}, y_{t-1}\}$ up to time $t$ but not including $t$. Moreover, we define the point process $N_g$, *the ground process,* which refers to the stochastic process of the inter-exceedance times of the extreme events. This point process has a conditional density function $p(t \mid \mathcal{H}_t)$ and its corresponding survival distribution function $S(t \mid \mathcal{H}_t)$. The conditional (finite) intensity function (or hazard function) for the ground process $N_g$ is given by

$$\lambda_g(t \mid \mathcal{H}_t) = \frac{p(t \mid \mathcal{H}_t)}{S(t \mid \mathcal{H}_t)} \tag{1}$$

while the conditional intensity function for the MPP $N$ is given by

$$\lambda(t, y \mid \mathcal{H}_t) = \lambda_g(t \mid \mathcal{H}_t) f(y \mid \mathcal{H}_t, t) \tag{2}$$

where $f(y \mid \mathcal{H}_t, t)$ is the density function of the marks conditional on $t$ and $\mathcal{H}_t$. Thus, the conditional intensity function with respect to the internal history $\mathcal{H}_t$ determines the probability structure of the point process $N$ uniquely. The advantage is that we only have to concentrate on the distribution limit of the maxima more than on the entire sample of a distribution to estimate such type of probabilities.

## 3 The autoregressive conditional duration Peaks-Over-Threshold (ACD-POT) model

We propose a set of models, which show autocorrelation between inter-exceedance times, clustered extremes and non iid exceedances or marks size. Following Engle and Russell (1998) we define a model for the conditional intensity of the ground point process of exceedances depending only on a fixed number of the most recent inter-exceedance times $x_i = t_i - t_{i-1}$. Let $\psi_i$ be the expectation of the $i$-th inter-exceedance time given by

$$\mathbb{E}(x \mid x_{i-1}, \ldots, x_1) = \psi_i(x_{i-1}, \ldots, x_1; \theta) \equiv \psi_i, \tag{3}$$

where $\theta$ is a parameter vector. We assume that $\psi_i$ correspond to the ACD class of models. In general the assumption is based on that for a strictly positive function with positive support $\varphi(\cdot) : \mathbb{R}_+ \to \mathbb{R}_+$ the standardized durations

$$\varepsilon_i = \frac{x_i}{\varphi(\psi_i)} \tag{4}$$

are iid random variables. To derive a general expression for the conditional intensity let $p$ be the density function of (4)

$$p\left(\frac{x_i}{\varphi(\psi_i)} \mid \mathcal{H}_i; \theta\right) = p\left(\frac{x_i}{\varphi(\psi_i)} \mid \theta\right), \tag{5}$$

where $\theta$ is a parameter vector. This implies that the time dependence of the duration process is summarized by the conditional expected duration sequence. If we define again a MPP on $[t_0,T)\times\mathscr{Y}$ for some finite positive time $T$ and let $(t_1,y_1),\ldots,(t_{N(T)},y_{N(T)})$ be a realization of $N$ over the interval $[0,T)$, one can easily show that the conditional expected intensity of the interexceedances times between extreme events, the ground process, can be expressed as a multiplicative effect between the baseline hazard function and a shift given by the expected duration

$$\lambda_g(t\mid\mathscr{H}_t;\theta)=\lambda_0\left(\frac{t-t_{N(T)}}{\varphi\left(\psi_{N(T)}\right)}\right)\frac{1}{\varphi\left(\psi_{N(T)}\right)} \tag{6}$$

In addition, we also consider the case of predictable marks, i.e., the marks are conditionally generalized Pareto, given the history $\mathscr{H}_t$ up to the time of the mark. To this end, we parameterized a scaling parameter $\beta\left(t,y\mid\mathscr{H}_t\right)$ such that it depends on the history[1]. In this way, we assume that in a period of turmoil the temporal intensity of the inter-exceedance times and the magnitude of the marks increase. Replacing the form of the ground process by one of this type and plugging the a generalized pareto density for size of the exceedances or marks in (2) we obtain the ACD-POT model in its more general form[2].

$$\lambda\left(t,y\mid\mathscr{H}_t;\theta\right)=\frac{\lambda_0\left(\frac{t-t_{N(T)}}{\varphi\left(\psi_{N(T)}\right)}\right)}{\varphi\left(\psi_{N(T)}\right)\beta\left(t,y\mid\mathscr{H}_t\right)}\left(1+\xi\frac{y-u}{\beta\left(t,y\mid\mathscr{H}_t\right)}\right)_+^{-1/\xi-1}. \tag{7}$$

The conditional rate of crossing the threshold $y\geq u$ at time $t$, given the history $\mathscr{H}_t$ up to that time, is in this case

$$\tau\left(t,y\mid\mathscr{H}_t;\theta\right)=\int_y^\infty\lambda\left(t,s\mid\mathscr{H}_t;\theta\right)ds=\frac{\lambda_0\left(\frac{t-t_{N(T)}}{\varphi\left(\psi_{N(T)}\right)}\right)}{\varphi\left(\psi_{N(T)}\right)}\left(1+\xi\frac{y-u}{\beta\left(t,y\mid\mathscr{H}_t\right)}\right)_+^{-1/\xi},$$

while the implied distribution of the marks when an extreme observation occurs is given by

$$\frac{\tau(t,u+y\mid\mathscr{H}_t;\theta)}{\tau(t,u\mid\mathscr{H}_t;\theta)}=\left(1+\xi\frac{y-u}{\beta\left(t,y\mid\mathscr{H}_t\right)}\right)_+^{-1/\xi}=\overline{G}_{\xi,\beta(t,y\mid\mathscr{H}_t)}(y).$$

Note that the marginal distribution of the marks will now be a conditional generalized Pareto distribution.

One main purpose of this paper is to develop a methodology to obtain an expression and its estimate for the quantile of the one day ahead predictive distribution of the returns, conditionally on the past and current data. In particular, the Value-at-risk (VaR), which has become standard measure in financial risk management due to its conceptual simplicity, computational facility and ready applicability. For the ACD-POT models the VaR is defined for the $\alpha$-th quantile as follows

$$\mathrm{VaR}_\alpha^t=u+\frac{\beta\left(t,y\mid\mathscr{H}_t\right)}{\xi}\left(\left(\frac{1-\alpha}{1-\exp\left(-\lambda\left(t,s\mid\mathscr{H}_t;\theta\right)\right)}\right)^{-\xi}-1\right). \tag{8}$$

The last equation implies that the VaR is only defined for our models if $1-\exp\left(-\lambda\left(t,s\mid\mathscr{H}_t;\theta\right)\right)>1-\alpha$.

In the following sections we introduce the models that we want to utilize to parameterize; the expected conditional duration function $\psi_i$, the distribution of probability of the standardized durations $\varepsilon_i$ and the models for the scale parameter $\beta\left(t,y\mid\mathscr{H}_t\right)$.

[1] We can also parameterized the shape parameter $\xi$. However, the behaviour of the estimation is severely affected. For this reason it is reasonable to take the shape parameter to be constant.

[2] The features of this model immediately follow of the classical POT model and the parameteres are the same as for the generalized extreme value distribution $H_{\xi,\mu,\sigma}$ with scaling parameter constant and equal to $\beta=\sigma+\xi(u-\mu)$.

### 3.1 ACD models for the expected conditional duration

In this subsection, we consider models that allow for additive as well as multiplicative components in the conditional duration function $\psi$. In addition, we introduce parameterizations that allow not only for linear but also for more flexible innovation impact curves. For simplicity, we restrict our attention to ACD models of order (1,1). The most popular autoregressive conditional duration models are:

- (ACD) The first ACD model (Engle and Russell, 1998) : $\psi_i = w + ax_{i-1} + b\psi_{i-1}$.
- (Log-ACD) The logarithmic ACD model introduced by Bauwens and Giot (2000) : $\psi_i = \exp\{w + ax_{i-1} + b\psi_{i-1}\}$, where $w > 0$, $a, b \geq 0$.
- (BCACD) The Box-Cox-ACD model (Dufour and Engle (2000)): $\psi_i = w + \frac{a}{\delta}\left(\varepsilon_{i-1}^{\delta} - 1\right) + b\psi_{i-1}$.
- (EXACD) The EXponential ACD Model (Dufour and Engle, 2000): $\psi_i = w + \{a\varepsilon_{i-1} + \delta\,|\varepsilon_{i-1} - 1|\} + b\psi_{i-1}$.

In order to ensure stationarity and existence of the unconditional expected duration for the Log-ACD model we need $a + b < 1$. Strict stationarity of the conditional mean for the models Log-ACD, BCACD and EXACD is guaranteed when $|b| < 1$. This BCACD specification includes the Log-ACD model for the Box-Cox parameter $\delta \to 0$ and a linear specification for $\delta = 1$. For the EXCAD model, the news effects are modeled with a piece-wise linear specification. Thus, for durations shorter than the conditional mean $(\varepsilon_{i-1} < 1)$, the news impact curve has a slope $a - \delta$ and an intercept $w + \delta$. Durations longer than the conditional mean $(\varepsilon_{i-1} > 1)$, also have a linear effect, but with a slope $a + \delta$ and intercept $w - \delta$. For more references to ACD models we refer to Hautsch (2004); Bauwens and Hautsch (2009).

### 3.2 Distributional assumptions for the standardized durations

The second important ingredient in the parameterization of our ACD-POT model is the distributional assumption for the innovation process. In this paper we explore the generalized gamma distribution. The major advantage of this distribution is that this has non-monotonic hazard function taking bathtub shaped or inverted U-shaped forms. This feature is of particular importance if we are interested in modelling risk measures such as the VaR or the expected shortfall.

A three parameter generalized gamma density is given by

$$f(x \mid \gamma, k) = \frac{\gamma x^{k\gamma-1}}{\lambda^{k\gamma}\Gamma(k)} \exp\left\{-\left(\frac{x}{\lambda}\right)^{\gamma}\right\},\ x > 0.$$

It includes the exponential distribution ($\gamma = k = 1$), the Weibull distribution ($k = 1$), the half-normal ($\gamma = 1/2, k = 1$) and the ordinary gamma distribution ($k = 1$). Under the restriction that $\lambda = 1$ we chose $\varphi(\psi_i) = \phi_i = \psi_i \frac{\Gamma(k)}{\Gamma\left(k+\frac{1}{\gamma}\right)}$ which implies a conditional density of the standardized duration given by

$$p\left(\frac{x_i}{\phi_i} \mid \mathscr{H}_i; \theta\right) = \frac{\gamma\psi_i}{x_i\Gamma(k)}\left(\frac{x_i}{\phi_i}\right)^{k\gamma} \exp\left\{-\left(\frac{x_i}{\phi_i}\right)^{\gamma}\right\}$$

where $\theta$ is once more a parameter vector. Note that if $k = 1$, then we get the Weibull-ACD model, while for $k = \gamma = 1$ the model reduces to an Exponential-ACD model. The hazard function implied by the generalized gamma model may now be written as

$$\lambda_g(x_i \mid \mathscr{H}_i; \theta) = \frac{\frac{\gamma x_i^{k\gamma-1}}{\phi_i^{k\gamma}\Gamma(k)} \exp\left\{-\left(\frac{x_i}{\phi_i}\right)^{\gamma}\right\}}{I\left(k, \left(\frac{x_i}{\phi_i}\right)^{\gamma}\right)},$$

where is the upper incomplete gamma integral $I\left(k,\left(\frac{x_i}{\phi_i}\right)^{\gamma}\right)=\int_{\left(\frac{x_i}{\phi_i}\right)^{\gamma}}^{\infty}u^{k-1}\exp\left(-u\right)du$.

In addition, the shape properties of the conditional hazard function can be derived from its parameters values. If $k\gamma<1$, the hazard rate is decreasing for $\gamma\leq 1$ and U-shaped for $\gamma>1$. Conversely, if $k\gamma>1$, the hazard rate is increasing for $\gamma\geq 1$, and inverted U-shaped for $\gamma<1$. Finally, if $k\gamma=1$, the hazard is decreasing for $\gamma<1$, constant for $\gamma=1$, and increasing for $\gamma>1$.

The conditional intensity in this case takes the form

$$\lambda\left(t,y\mid\mathscr{H}_t;\theta\right)=\frac{\frac{\gamma x_i^{k\gamma-1}}{\phi_i^{k\gamma}\Gamma(k)}\exp\left\{-\left(\frac{x_i}{\phi_i}\right)^{\gamma}\right\}}{I\left(k,\left(\frac{x_i}{\phi_i}\right)^{\gamma}\right)}\frac{1}{\beta\left(t,y\mid\mathscr{H}_t\right)}\left(1+\xi\frac{y-u}{\beta\left(t,y\mid\mathscr{H}_t\right)}\right)_+^{-1/\xi-1}.$$

### 3.3 Models for the time varying scale parameter

In this section we consider different models to parameterize the scaling parameter $\beta\left(t,y\mid\mathscr{H}_t\right)$ such that it depends on the history. This feature implies that the marks are conditionally generalized Pareto, given the history $\mathscr{H}_t$ up to the time of the mark. Under these models we assume that in a period of turmoil the temporal intensity of the inter-exceedance times and the magnitude of the marks increase. We will specify and estimate two alternatives forms for the scaling parameter $\beta\left(t,y\mid\mathscr{H}_t\right)$, the lineal

$$\beta\left(t,y\mid\mathscr{H}_t\right)=\omega+\beta_1 y_{i-1}+\beta_2\varphi\left(\psi_i\right)$$

and the exponential

$$\beta\left(t,y\mid\mathscr{H}_t\right)=\omega+\beta_1 y_{i-1}+\beta_2^{\varphi\left(\psi_i\right)},$$

where $\omega,\beta_1,\beta_2\in\mathbb{R}_+$.

The two models have in common a natural interpretation; the scaling parameter $\beta\left(t,y\mid\mathscr{H}_t\right)$ depends on the last mark and the conditional mean function of the inter-exceedance times, given the information up to time $t$, but not including $t$[3].

### 3.4 Empirical application

For the empirical test we chose the transaction data from Bayer AG as announced already in the introduction. In this study we only concentrate on the left tail, so that the daily returns are obtained as $r_t=-100\ln(p_t/p_{t-1})$, where $p_t$ denotes the (closing) stock price at day $t$. The sample period spans from 2 January 1990 to January 18, 2008, two days before January 20, when the Global stock markets suffered their biggest falls since September 11, 2001. A second sample is used for backtesting the estimation of the VaR in the Bayer AG from 20 January 2008 to January 16, 2009. In the backtest we daily update the new information that becomes available for the parameter estimates previously obtained. Thus, we dynamically adjust quantiles, which allows us to improve as accurately as possible the estimation of the risk measures.

In order to summarize adequately the large quantity of empirical results obtained, we use a classification scheme for the ACD-POT models. The first letters denote the type of ACD model: ACD, Log-ACD, BCACD or EXACD[4]. The last small letter denotes the models for the time varying scale parameter: l (lineal), e (exponential) or u (for unpredictable marks with scale parameter $\beta$ constant). For example, a model

[3]In the first instance we take different models into account to find the best approach. Our analyses with financial time series have suggested that by model comparisons based on the likelihood ratio statistic, these models keep the formulation easy to understand.

[4]For a meaningful comparison of alternatives and for simplicity, we limit the dynamic structure of the ACD-POT models to the first lag order only.

Log-ACDu means that we are working with a Log-ACD model for the expected conditional duration with generalized gamma distribution and unpredictable marks.

Table 1: Results of the estimation of all ACD-POT models with distributional assumption generalize gamma for the standardized durations of the inter-exceedance times for the Bayer returns. Standard deviations are given in parentheses. Loglike are the results of the maximization of the log-likelihood estimation and AIC is the Akaike Information Criterion.

| Models | Parameters | | | | | | | | | | Loglik | AIC |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | ACD model | | | | | | POT Model | | | | | |
| | $w$ | $a_1$ | $b_1$ | $\delta$ | $\gamma$ | $k$ | $\xi$ | $\omega$ | $\beta_1$ | $\beta_2$ | | |
| ACDu | 0.833 | 0.135 | 0.775 | | 0.157 | 33.093 | 1.091 | 0.144 | | | -2042.72 | 4099.45 |
| | (0.326) | (0.035) | (0.057) | | (0.046) | (19.491) | (0.072) | (0.048) | | | | |
| ACDl | 0.754 | 0.157 | 0.767 | | 0.121 | 53.767 | 0.503 | 0.094 | 5.7e-07 | 4.492 | -2024.8 | 4067.59 |
| | (0.280) | (0.036) | (0.051) | | (0.041) | (36.539) | (0.100) | (0.044) | (0.010) | (0.826) | | |
| ACDe | 0.769 | 0.146 | 0.773 | | 0.097 | 86.674 | 0.657 | 0.123 | 0.055 | 0.513 | -2032.27 | 4082.55 |
| | (0.190) | (0.036) | (0.052) | | (0.021) | (41.234 | (0.166) | (0.046) | (0.045) | (0.212) | | |
| Log-ACDu | 0.175 | 0.122 | 0.831 | | 0.162 | 31.198 | 1.091 | 0.144 | | | -2042.61 | 4099.23 |
| | (0.078) | (0.031) | (0.053) | | (0.044) | (17.101) | (0.072) | (0.048) | | | | |
| Log-ACDl | 0.184 | 0.139 | 0.814 | | 0.179 | 24.668 | 0.474 | 0.097 | 3.2e-07 | 4.731 | -2025.03 | 4068.07 |
| | (0.072) | (0.030) | (0.049) | | (0.046) | (12.629) | (0.105) | (0.043) | (0.009) | (0.887) | | |
| Log-ACDe | 0.161 | 0.128 | 0.833 | | 0.108 | 69.870 | 0.653 | 0.124 | 0.055 | 0.510 | -2031.9 | 4081.8 |
| | (0.069) | (0.031) | (0.049) | | (0.029) | (37.581) | (0.165) | (0.046) | (0.045) | (0.208) | | |
| EXACDu | 0.643 | 0.343 | 0.599 | 0.095 | 0.974 | 0.949 | 0.952 | 0.346 | | | -2133.07 | 4282.15 |
| | (0.241) | (0.064) | (0.112) | (0.128) | (0.000) | (0.000) | (0.071) | (0.083) | | | | |
| EXACDl | 0.675 | 0.308 | 0.582 | 0.156 | 0.989 | 0.980 | 0.638 | 0.229 | 0.182 | 0.504 | -2138.13 | 4296.27 |
| | (0.102) | (0.075) | (0.040) | (0.083) | (0.000) | (0.000) | (0.063) | (0.053) | (0.055) | (0.027) | | |
| EXACDe | 0.053 | 0.129 | 0.916 | 0.001 | 0.132 | 47.045 | 0.653 | 0.120 | 0.055 | 0.504 | -2033.52 | 4087.05 |
| | (0.057) | (0.023) | (0.028) | (0.010) | (0.023) | (21.023) | (0.164) | (0.046) | (0.045) | (0.205) | | |
| BCACDu | 0.214 | 0.133 | 0.905 | 0.889 | 0.160 | 31.695 | 1.090 | 0.145 | | | -2043.57 | 4103.14 |
| | (0.081) | (0.044) | (0.033) | (0.394) | (0.046) | (18.085) | (0.072) | (0.048) | | | | |
| BCACDl | 0.205 | 0.157 | 0.916 | 0.776 | 0.139 | 41.114 | 0.502 | 0.093 | 2.6e-08 | 4.501 | -2025.24 | 4070.48 |
| | (0.074) | (0.039) | (0.029) | (0.314) | (0.043) | (25.380) | (0.100) | (0.043) | (0.008) | (0.830) | | |
| BCACDe | 0.197 | 0.145 | 0.915 | 0.838 | 0.120 | 56.519 | 0.664 | 0.123 | 0.053 | 0.520 | -2033.16 | 4086.31 |
| | (0.075) | (0.043) | (0.030) | (0.360) | (0.041) | (38.521) | (0.161) | (0.046) | (0.045) | (0.208) | | |

All parametric models are estimated using quasi maximum likelihood. An important point is the choice of the threshold, which implies a balance between bias and variance. The threshold must be set high enough so that the exceedances are distributed generalized Pareto. However, the choice of the optimal threshold is still considered an open problem and different approaches have been proposed to overcome this difficulty. In this paper we choose to work with the 10% of the maxima of the sample[5].

In relation to the measures of goodness of fit we utilize the W-statistics to assess our success in modelling the temporal behaviour of the exceedances of the threshold $u$. This statistic states that if the GPD parameter model is correct, then the residuals are approximately independent unit exponential variables.

[5]The choice of the threshold is done with help of the the mean excess (ME) function, which is a popular tool used to determine the adequacy of the GPD model in practice.

In addition, to check that there is no further time series structure the autocorrelation function (ACF) for the residuals is also included. Similarly, we provide empirical evidence on the accuracy of actual VaR measures derived from the models. The first of them is an unconditional coverage test proposed by McNeil and Frey (2000). The idea is to test if the fraction of violations obtained for a particular risk measure, is significantly different from the theoretical one. A violation of the VaR is defined as occurring when the ex-post return is lower than the VaR. The second approach proposed by Berkowitz et al. (2009) tests for uncorrelatedness of the violations. In particular, we suggest the well-known Ljung-Box test of the violation sequence's autocorrelation function.

### 3.5 Empirical results

The maximum log-likelihood estimates of the ACD-POT models proposed for the returns are displayed in Table 1. For the inter-exceedance times, the generalized gamma seems to be the best distribution between the two choices. The results on ACD models for the expected conditional duration lead to markedly fovour the Log-ACD specifications, followed by the ACD one. Finally, the models with time varying scale parameters lead to a better fit. Indeed, the results suggest that the models with predictable marks react more quickly to increasing and decreasing cluster of extremes, which means that the size of the exceedances has an effect on the probability of further exceedances in the near future.

According to the AIC of the models proposed, the best fitted model for the Bayer index is a ACD model with generalized gamma distribution and lineal form for the scale parameter (ACDl) with AIC of 4067.59. We observe further that $k = 53.767$ (36.539), $\gamma = 0.121$ (0.041), which implies that $k\gamma > 1$ and $\gamma < 1$ so that the hazard rate is inverted U-shaped. This should not come as a surprise if one is aware of the intimate relationship between durations and cluster of extremes. Furthermore, this is the sort of hazard function that earlier authors have found to be realistic in modeling the dynamics of "price durations" in stock markets (see for instance Zhang et al., 2001). In relation to the results of estimation of the conditional GPD model to the excedances we obtained $\xi$=0.503 (0.100), $\omega$= 0.094 (0.044), $\beta_1$=5.7e-07 (0.001) and $\beta_2$= 4.492 (0.826). This result indicates that the lineal form to parameterize the scaling parameter $\beta\,(t,y \mid \mathscr{H}_t)$, such that it depends on the history, was a good choice. Interestingly, the size of the last exceedance is not as important as the expectation of the i-th inter-exceedance time.

Although the model of choice identified by the AIC may be seen as the best among the existing models because it shows the best global fit, this does not mean that no better model is possible for backtesting. So, we usually check whether the major features of the given data can be reproduced by the estimated models, for instance, the cluster of extreme events. If this important feature is not reproduced, we must consider further models whose AIC values can be compared with those of the previous best model. To this end, we include two other models to have a comparison of different alternatives in the backtest. The second best alternative is a Log-ACD and the third is a BCACD, both with generalized gamma distribution and lineal form for the scale parameter. We concentrate on these three alternatives and test the reliability of these models by investigating the conditional GPD assumption of the marks in the models fitted, the quality of the times component of our model and the performance in-sample of the estimated VaRs.

The results on the goodness of fit in sample are displayed in Figure 1. Here, we first assess the conditional GPD assumption of the marks in the models fitted. To this end, we provide the W-statistic. This statistic forms an iid sequence of exponential random variables with mean one if the marks are GPD. According to the QQ-plots displayed in Figure 1, we do not observe a substantial deviation from an exponential distribution. In addition, to check that there is no further time series structure, the autocorrelation function (ACF) for the residuals (middle panel) is also included. The autocorrelations is negligible at nearly all lags. Finally, to appraise the quality of the times component of our model, we employ the residual analysis method for point processes. This is based on the change of time scale using the estimated conditional in-

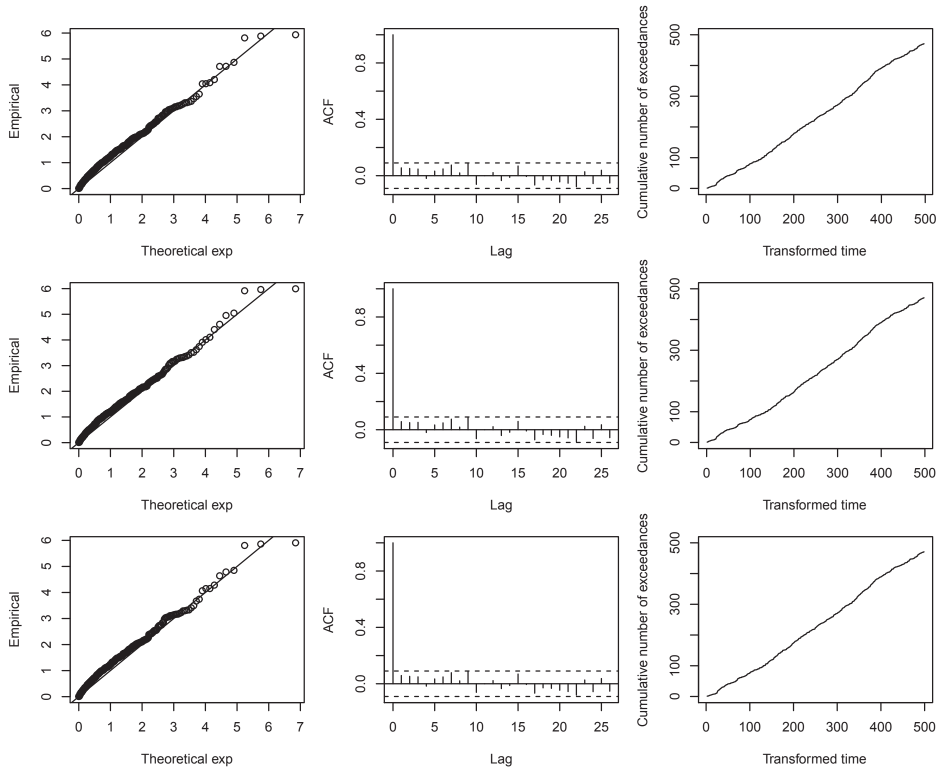


Figure 1: QQ-plots of the residuals (left), autocorrelation function of the residuals (middle) and cumulative numbers of the residual process versus the transformed time $\{\tau_i\}$ (right), for the returns of the Bayer index for the gACDl (upper panel), gLog-ACDl (middle panel) and gBCACDl (lower panel) models.

tensity. We investigated whether the transformed time-scale version of the data constitutes a homogeneous Poisson process. The residual analysis for the three models indicates that the ACD-POT models in their three alternatives in the changed time scale seems to be acceptable. Hence, for the returns the ACD-POT specification seems to be appropriate.

In relation to the performance in-sample of the estimates VaRs, Tables 2 displays the results for the unconditional coverage test and the Ljung-Box test, for all the models for three different VaR levels (0.05, 0.01, and 0.001). For each model proposed in the last section we give the number of violations or failures in the VaR, the unconditional coverage test and the Ljung-Box test (the last two in brackets). In the case of these three models fitted the conditional VaR is correctly estimated for all the confidence levels[6].

### 3.6 Backtesting the models

Backtesting provides invaluable feedback about the accuracy of the models proposed to risk managers. The performance of VaR with respect to backtesting has been carried out with the daily returns for one year, i.e., from January 20,2008 to January 16, 2009. The backtest method consists on comparing the estimated conditional VaR for one day time horizon $t$, given knowledge of returns up to and including $t$ for three different confidence levels (0.95, 0.99, and 0.999). For each day in the back test we reestimate the models, something that immediately reveals possible stability problems of a model. Then, we reestimated

[6]The null hypothesis is rejected whenever the p-value of the binomial test and the Ljung_Box test are less than 5 percent.

Table 2: Some VaR in-sample results all models. The number of violations for all VaR confidence levels is displayed for each model. Values in parentheses are p-values for the unconditional coverage test and the Ljung-Box statistic with 5 lags. The number of observations in the sample is 4607.

| Models | Number of violations | | | Models | Number of violations | | |
|---|---|---|---|---|---|---|---|
| | $VaR_{0.95}$ | $VaR_{0.99}$ | $VaR_{0.999}$ | | $VaR_{0.95}$ | $VaR_{0.99}$ | $VaR_{0.999}$ |
| EXACDu | 387 | 61 | 3 | ACDu | 386 | 95 (0.00) | 0 |
| | (0.00, 0.00) | (0.05, 0.09) | (0.64, 0.86) | | (0.00, 0.00) | (0.00, 0.00) | (0.02, - ) |
| EXACDl | 400 | 144 | 10 | ACDl | 245 | 54 | 6 |
| | (0.00, 0.00) | (0.00, 0.00) | (0.03, 0.00) | | (0.52, 0.22) | (0.30, 0.48) | (0.48, 0.51) |
| EXACDe | 288 | 43 | 0 | ACDe | 291 | 46 | 0 |
| | (0.00, 0.00) | (0.61, 0.92) | (0.02, - ) | | (0.00, 0.00) | (0.94, 0.88) | (0.02, - ) |
| BCACDu | 389 | 93 | 0 | Log-ACDu | 390 | 90 | 0 |
| | (0.00, 0.00) | (0.00, 0.00) | (0.02, - ) | | (0.00, 0.00) | (0.00, 0.00) | (0.02, - ) |
| BCACDl | 244 | 53 | 6 | Log-ACDl | 243 | 43 | 5 |
| | (0.57, 0.31) | (0.17, 0.29) | (0.48, 0.51) | | (0.60, 0.36) | (0.61, 0.92) | (0.82, 0.61) |
| BCACDe | 293 | 43 | 0 | Log-ACDe | 289 | 41 | 0 |
| | (0.00, 0.00) | (0.69, 0.92) | (0.02, - ) | | (0.00, 0.00) | (0.42, 0.93) | (0.02, - ) |

Table 3: Some VaR backtesting results for three models. Values in parentheses are p-values for the unconditional coverage test and the Ljung-Box test iwth 5 lags. Values smaller than a p-value 0.05 indicate failure. The number of observations in the sample is 288.

| Model | Number of violations | | |
|---|---|---|---|
| | $VaR_{0.95}$ | $VaR_{0.99}$ | $VaR_{0.999}$ |
| ACDl | 18 | 4 | 0 |
| | (0.34, 0.38) | (0.54, 0.81) | (1, -) |
| Log-ACDl | 20 | 3 | 0 |
| | (0.14, 0.72) | (0.77, 0.86) | (1, -) |
| BCACDl | 19 | 5 | 0 |
| | (0.22, 0.48) | (0.22, 0.76) | (1, -) |

the risk measures for each return series according to the formula (8).

Table 3 reports the results on the VaR backtesting exercise for all confidence levels, while the VaR violations for the 0.99 confidence level under the Log-ACDl, ACDl and BCACDl models are shown in Figure 2. The performances for the models are similar for the results on VaR forecasting, although we observe some differences. For instance, the gLog-ACDl model tends to lightly underestimate the $VaR_{0.95}$, while the gBCACDl do the same for the $VaR_{0.99}$. However, the unconditional coverage test and the Ljung-Box test for all the confidence levels indicate that no severe clustering of exceedances is present and that the VaR violations can be considered as independent, respectively. In addition, according to the "traffic light" approach the three models are all classified in the green zone (see for more reference Basel Committee on Banking Supervision (2006)). Finally, due to the shortness of the time horizon we do not find a VaR violation for the 0.999 quantile, and therefore the Box-Ljung test p-values are not reported.

To summarize, the results of our backtesting procedure with a dynamic adjustment of quantiles incorporating the new information daily allows us to statistically conclude that the models proposed are

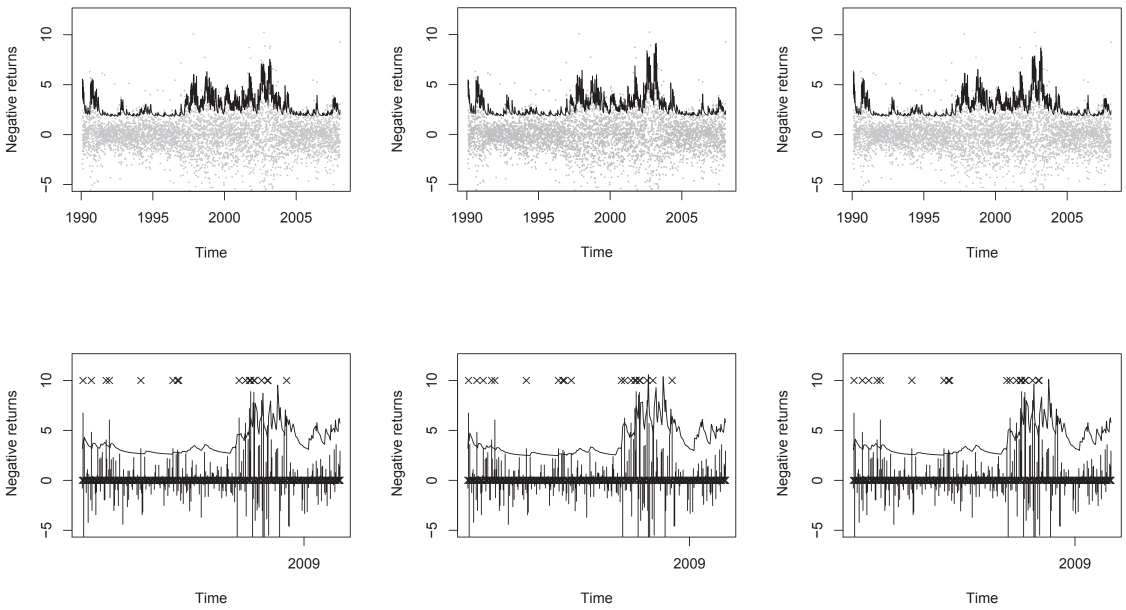


Figure 2: In-sample and Backtest estimation of the VaR for the 0.99 confidence level for the the models fitted to the Bayer returns. The black line is the VaR estimation. In the top panel from left to right we have the VaR estimates in-sample for the ACDl, Log-ACDl and BCACDl models. In the bottom panel from left to right we have the backtesting results for the VaR estimates for the ACDl, Log-ACDl and BCACDl models.The $\times$ marks at the top of the figures indicate the violations of the VaR at the 0.99 confidence level.

suitable for the estimation of different risk measures, as for example, the VaR according to the restriction imposed by Basel. Moreover, these models allow us to take the heavy-tailness or the stochastic nature of the cluster of extreme events into consideration.

## 4 Conclusions

This paper proposed a new technique for modelling extreme events of stationary sequences as is the case of the most financial returns. We make use of a new class of self-exciting point process models that seem particularly well suited. The idea was to create a model being able to incorporate stylized facts such as clustering of extreme events and autocorrelation of the inter-exceedance times of extreme events, i.e., properties that are observed in practice.

The model can be interpreted as a combination between the classical Peaks over Threshold (POT) model from Extreme Value Theory and the class of Autoregresive Conditional Duration (ACD) models that are popular in finance. For this reason we call it ACD-POT models.

We observe that under this methodology the estimation of such models can be straightforwardly derived through conditionals intensities. Different models were proposed having in mind the simplicity of the structure of the conditional intensities. However, other more complicated structures could also be adopted.

With regard to the empirical application the models and their estimations with returns from Bayer AG were more than satisfactory. Our empirical results show that characteristics associated with previous extreme losses as well the time between these extreme events have a significant impact on the dynamic aspects and size of future extreme events.

On average, the best three models fit well in-sample for the VaR for different levels of risk, i.e., in terms of capital requirement; the models keep necessary capital to guarantee the desired confidence level. For these models the VaR is backtested through a comparison with the actual losses over an out-of-the-sample period of one year. The backtesting results indicate that the proposed methodology performs well in forecasting the risk dynamically and provides therefore certainly more precise estimate as the information in the data sample is exploited more efficiently. This refers particularly to clustering of extreme events and the inter-exceedance times.

## References

Basel Committee on Banking Supervision (1996). Supervisory framework for the use of "backtesting" in conjunction with the internal models approach to market risk capital requirements. *Basel Committee on Banking Supervision*.

Basel Committee on Banking Supervision (2006). Basel II: International Convergence of Capital Measurement and Capital Standards: A Revised Framework - Comprehensive Version. *Basel Committee on Banking Supervision*.

Bauwens, L. and P. Giot (2000). The logarithmic ACD model: an application to the bid-ask quote process of three NYSE stocks. *Annales d'Economie et de Statistique 60*, 117–149.

Bauwens, L. and N. Hautsch (2009). Modelling financial high frequency data using point processes. In J.-P. K. T. G. Andersen, R. A. Davis and T. Mikosch (Eds.), *Handbook of Financial Time Series*, Volume 6, pp. 953–979. Springer.

Berkowitz, J., P. Christoffersen, and D. Pelletier (2009). Evaluating value-at-risk models with desk-level data. *Forthcoming in Management Science*.

Dufour, A. and R. Engle (2000). The ACD model: predictability of the time between consecutive trades. Technical report, University of Reading, ISMA Centre.

Engle, R. and J. Russell (1998). Autoregressive conditional duration: A new model for irregularly spaced transaction data. *Econometrica 66*, 1127–1162.

Hautsch, N. (2004). *Modelling irregularly spaced financial data: theory and practice of dynamic duration models*. Springer.

McNeil, A. and R. Frey (2000). Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. *Journal of Empirical Finance 7*, 271–300.

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