

VaR forecasts by extreme value models in a conditional duration intensity framework

Abstract

The analysis of extremes in financial return series is often based on the Peaks-Over-Threshold (POT) model. This model assumes independent and identically distributed observations and a Poisson process is accordingly 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 propose an alternative approach to overcome these difficulties by considering 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 would certainly affect the estimation. We make use of the proposed model to obtain an improved estimate for the Value at Risk. The model is then benchmarked to various competing approaches like Engle and Marianelli's CAViaR or the GARCH-EVT model. Finally we present a comparative empirical illustration to transaction data from Bayer AG, a blue chip stock from the German stock market index DAX, the DAX index itself and a hypothetical portfolio of international equity indexes .

JEL classification: C22, C58, F30.

Keywords: Extreme value theory, autoregressive conditional duration, value at risk, self-exciting point process, conditional intensity.

1. Introduction

In recent years there has been a noticeable increase in the frequency and impact of extreme events and financial crises. These events range from currency crashes (East Asia in 1997, Russia in 1998, Argentina in 2001), to liquidity crises (LTCM in 1998), to stock market crashes (Black Monday in 1987, Dot.com in 2000), and to the US subprime market spillovers from 2007 through

to 2009. An important aspect of these extreme events is that their impact is exacerbated by simultaneous occurrence in a multiple class of assets.

Critical questions are being asked concerning some of the quantitative methods used in risk management under the Basel II proposals. Why do extreme events occur? What measures are being taken to deal with extreme crises? Can researchers study the mechanics of extreme events in history and learn how to avoid them in the future? Both theoretical and more practically oriented questions are on the actual agenda of academics, practitioners and regulators when the task is to understand the dynamics of asset markets under stress (see Embrechts, 2009 and references therein).

The estimation of the Value at Risk (VaR) and related risk measures is a current topic of interest in finance, for which many approaches of varying sophistication have been derived. According to Chavez-Demoulin et al. (2005) two main approaches can be distinguished: the time series and the extreme value approach. The first emphasizes modeling the temporal features (e.g., volatility clustering and fat tails) with ARCH-type and stochastic volatility models. However, the study of extreme dependence may reveal contrasts which are obscured when concentrating on examining only the conditional second moment of a time series. Interestingly, and unlike the situation for GARCH processes (see Davis and Mikosch, 2009), there is no extremal clustering for stochastic volatility processes in either the light- or heavy-tailed cases. That is, large values of the processes do not come in clusters, which means that the large sample behavior of maxima is the same as that of the maxima of the associated iid sequence. On the other hand, Mikosch (2003) showed in a simulation study that for the GARCH case the expected cluster size in a set of various log-return series is smaller than for the fitted GARCH model, i.e., there is less dependence in the tails for the returns and volatilities than for the prescribed GARCH model. While these models imply some information about extreme events, still little is known about the extremes per se.

The extreme value approach makes inferences on the VaR using results from Extreme Value Theory (EVT), which only focuses on the tail of the distribution (see Embrechts et al. 1997 for an introduction). The majority of the approaches on EVT for VaR estimation concern the estimation of unconditional quantiles (see for example Danielsson and De Vries, 2000, Coles, 2001 and Cotter and Dowd, 2006). An exception is the work of McNeil and Frey (2000), which addresses the conditional quantile problem and proposes a method for applying EVT to the conditional return distribution by using a two-stage method, combining GARCH models for forecasting volatility and EVT techniques applied to the residuals from the GARCH analysis. Although this methodology works quite well in practice it has major drawbacks, as addressed by Mikosch (2003). Thus, one should be cautious with the interpretation of the results of this method, since there is no theory in

the extremal clustering behavior based on the residuals of a GARCH model.

A novel form to deal with the cluster on extremes is to use a cluster point process version of Peaks Over Threshold (POT) model introduced preliminarily in McNeil et al. (2005) and Chavez-Demoulin et al. (2005), where the clusters of extreme data are modeled as self-exciting point processes without involving a prefiltration of data. The main characteristic of these models is that the intensity of occurrence of extreme events can depend on past extreme events and the size of the exceedances, thus allowing more realistic models.

In this paper we concentrate on a different alternative. We model the stochastic intensity of the point process of exceedances within the framework of irregularly spaced data. Contrary to the classical POT methodology, where the time of occurrence of the extreme events is modeled, the proposed methodology is able to model the inter-exceedance times between extreme events. To this end, we use a technique similar to an Autoregressive Conditional Duration (ACD) model (see Engle and Russell, 1998 for more reference), while the marks still being modeled by generalized Pareto distributions. Like the GARCH models, the ACD models and their alternatives (see Engle and Russell, 1998; Ghysels and Jasiak, 1998; Engle, 2000; Zhang et al., 2001; Russell and Engle, 2005; Bauwens and Hautsch, 2006) have proven to be very useful in capturing the clustering effects. For this reason, it seems natural to model the cluster behavior of extreme observations by means of this class of processes.

The main contribution of this paper from the point of view of extreme value theory is that we are able to capture the short-term behavior of extremes without involving an arbitrary stochastic volatility model or the prefiltration of the data, which certainly impacts the measures of risk. Furthermore and contrary to the models proposed in McNeil et al. (2005) and Chavez-Demoulin et al. (2005), whose self-exciting functions are restricted to monotone decreasing functions, the models proposed in this paper allow hazard functions that are both monotonically decreasing and increasing. This has a logical interpretation in periods of financial turmoil, where the VaR typically increases in an initial period, then becomes close to constant before finally decreasing. To the best of our knowledge, this is the first research which takes the incidence of the inter-exceedance times and models for irregularly spaced data in extreme value models into account.

The results of the application to the stock market data from Bayer AG, the DAX index and a hypothetical portfolio indicate that the estimation of such models can be straightforward, derived through conditionals intensities. Different models were proposed, having in mind the simplicity of the structure of the conditional intensities. The empirical results show that characteristics associated with previous extreme losses as well as the time between these extreme events have a significant impact on the dynamic aspects and the size of future extreme events. In a VaR context

the results of our backtesting procedure, which dynamically adjusts quantiles to incorporate 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). Further, in comparison with others competitive models, in most of the cases the ACD-POT models outperform the basic specifications of CAViar models introduced by Engle and Manganelli (2004). Finally, the ACD-POT and the two-stage GARCH-EVT methodology (McNeil and Frey, 2000) were the only methods that eradicate the threat of VaR violation clustering in a great many situations.

This paper is organized as follows. In the next section we give a brief motivation for modeling the inter-exceedances times between extreme events. In section 3 we outline relevant aspects of the classical POT model and describe the ACD-POT model theory that is central to the paper and discuss 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 4 the models are applied to transactions data from Bayer AG, the DAX index and a hypothetical portfolio. Conclusions and proposals for future work are resumed in section 5.

2. Motivation

In the following we will explain our motivation in investigating extreme events in a stock market as a marked point process of exceedances. The classical POT model for iid data assumes that if a threshold u has been chosen highly enough then the exceedances over this threshold, the extreme events, occur in time according to a homogeneous Poisson process. In addition, the size of the excess returns over the threshold, the mark sizes, are independently and identically distributed according to the generalized Pareto distribution (GPD). Nevertheless, the result of more than half a century of empirical studies on financial time series indicates that this is not the case.

Figure 1 shows in the top panel the negative daily percentage log returns of Bayer shares between 2 January 1990 and 18 January 2008, and the times and sizes of the negative daily percentage log returns exceeding a threshold $u = 1,5$. Observe that this contradicts the classical model assumption of no cluster at the extremes. Indeed, under a homogeneous Poisson process the inter-exceedance times should be independent exponential random variables. The lower left picture shows an exponential probability plot for the inter-event times, these are clearly far from exponential, giving evidence against a Poisson process of exceedances. Furthermore, the auto-correlogram plot suggests clustering of the inter-exceedance times. This hypothesis is moreover reaffirmed by the Ljung-Box statistic using 10 lags. The null hypothesis of white noise is easily

rejected with the Ljung-Box statistic of 217.63 well above the critical value of 18.307 at the 5% level, rejecting the Null hypothesis.

Since extreme events are inherently irregularly spaced in time and their inter-exceedance times provide strong evidence of correlation, it seems natural to study the timing of transactions as an ACD model. Since the introduction of this model by Engle and Russell (1998), a plethora of modifications and alternatives have been proposed. Zhang et al. (2001) introduce a threshold ACD (TACD). Drost and Werker (2004) provide a method for obtaining efficient estimators of the ACD model with no need to specify the distribution. Fernandes and Grammig (2006) introduce the augmented ACD model, a very general model that covers almost all the existing ones. Meitz and Teräsvirta (2006) introduce smooth transition ACD models and testing evaluation procedures. Alternative models based on latent variables are the Stochastic Conditional Duration (SCD) model of Bauwens et al. (2004) and the Stochastic Volatility Duration (SVD) model of Ghysels et al. (2004).

Often, there will be additional information associated with the arrival times, as for example the stock market prices, which depending on the economic question at hand may be of interest. In the literature of point process, this extra information is called mark, as they identify or further describe the event which occurred. In extreme value theory, the point of time is the time at which an extreme event occurs, and the marks are the size of exceedances given a high threshold. For instance, Engle (2000) introduce a framework to estimate the dynamics of events (that are inherently irregularly spaced) and the associated prices conditional on the times. Ghysels and Jasiak (1998) propose to model the volatility of irregularly spaced data. Russell and Engle (2005) propose to use the ACD model for durations and the Autoregressive Conditional Multinomial (ACM) model for the conditional distribution of the discrete price changes.

Following the same direction as the last three models, in the next section we shall propose an extension of the classical ACD model to the inter-exceedance times between extreme events that is more flexible and provides a better adjustment than the existent classical extreme value models for VaR prediction, considering that all the temporal dependence in the inter-exceedance times is captured by a conditional mean function. Although, the immediate application is to financial transactions data we believe the model could prove useful in a variety of other settings, e.g. prices from energy markets.

3. Methodology

The use of EVT in risk management is a fairly recent innovation, but there is a much longer history of its use in the natural science and the insurance industry. Embrechts et al. (1997) survey

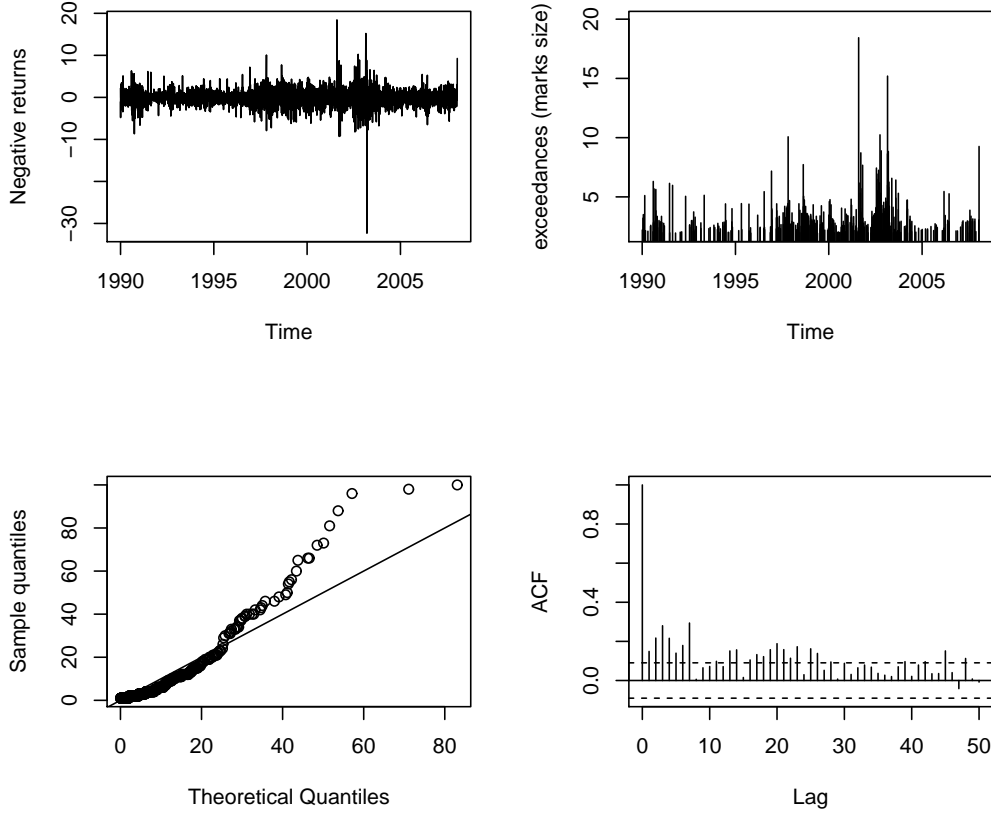


Figure 1: Upper left panel shows daily percentage loss data of Bayer between 02.01.1990 and 18.01.2008. Upper right panel displays the 732 largest losses (i.e. exceedances). Lower left panel shows a QQ-plot of inter-exceedance times against an exponential reference distribution. Finally, the lower right panel displays the autocorrelogram for the inter-exceedance times for the exceedances.

the mathematical theory of EVT in an excellent way and discuss its applications to both financial and insurance risk management. In this section we briefly outline the theoretical aspects of EVT and highlight some aspects that are specific to financial data. As a consequence, modifications and extensions of a direct EVT approach turn out to be beneficial.

3.1. Classical extreme value theory

Suppose $\{Y_t\}_{t=1}^n$ are random variables with distribution function F which belongs to the maximum domain of attraction of $H_{\xi,\mu,\sigma}$. Then $H_{\xi,\mu,\sigma}$ is a generalized extreme value distribution

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

where $1 + \xi y > 0$, $\xi, \mu \in \mathbb{R}$ and $\sigma > 0$ are the shape, location and scale parameter respectively. Classical EVT is sometimes applied directly, for example by fitting this distribution to the annual or monthly maxima of a financial series and much historical work was devoted to this approach (e.g. Smith, 2003). From a modern viewpoint, however, the classical approach is too narrow to be applied to a wide range of problems.

An alternative approach is to pursue the idea to interpret exceedances over thresholds as a point process. (Smith, 1989) view a point process N as a random distribution of indistinguishable points in a defined state space. We display this idea in Figure 2. For instance, the basic model for threshold exceedances in extreme value theory is based on constructing a two-dimensional point process $\{(t_i, y_i)\}_{i=1}^T$ with state space $\mathcal{T} \times \mathcal{Y} = [0, 1) \times (y, \infty)$. The time events t_i are the time t of the i -th peak exceedance and we shall refer to this process as the ground process, while $y_i - u$ is the value of the exceedances for a sufficiently high threshold u and we will call this process the process of marks (in Figure 2 these are observations $t = 2, 6, 8, \dots$ or equivalently $i = 1, 2, 3, \dots$). For iid observations¹, each data point has the same chance to exceed the threshold, and therefore, the two dimensional point process will look like as a non-homogeneous Poisson process with intensity defined for all subsets of the form $A = [t_1, t_2) \times (y, \infty)$, where t_1 and t_2 are times of occurrence of extreme events. This leads to the following representation

$$\lambda(t, y) = \lambda(y) = \frac{1}{\sigma} \left(1 + \xi \frac{y - \mu}{\sigma} \right)_+^{-1/\xi - 1}, \quad (2)$$

where $y_+ = \max(y, 0)$ and μ, σ, ξ are precisely the parameters of the generalized extreme value distribution.

The intensity measure of the subset A for any $y \geq u$ may be expressed in the form of an one-dimensional Poisson process with intensity

$$\tau(y) = \int_y^\infty \lambda(s) ds = -\ln H_{\xi, \mu, \sigma}(y).$$

If we accept that the point process of exceedances is one-dimensional Poisson with intensity $\tau > 0$, then the process has independent increments, i.e., the number of events t_i that occur in disjoint time intervals are mutually independent, which implies lack of memory in the evolution

¹A sufficient condition for the process is already its (weak) stationarity together with the condition that there are asymptotically no clusters among the high-level exceedances.

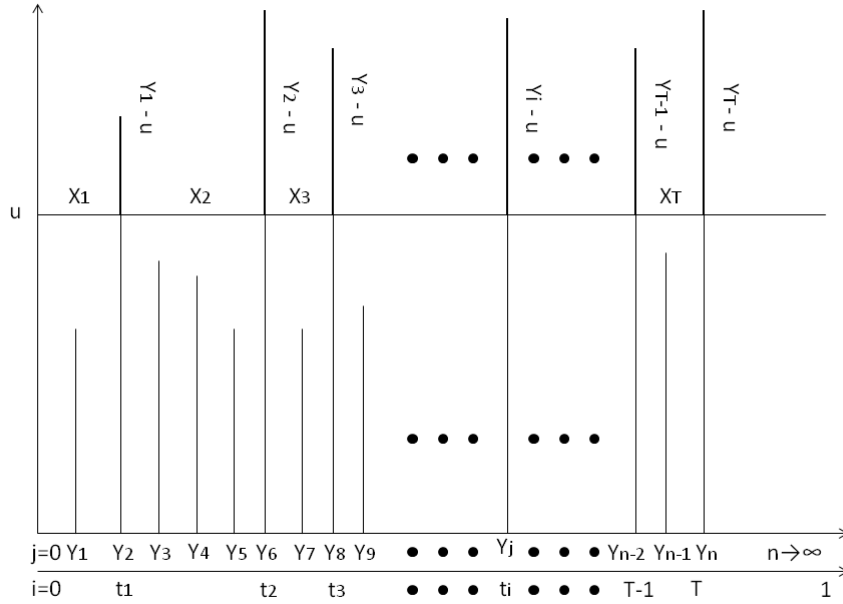


Figure 2: Illustration of high-level exceedances represented as a marked point process.

of the process. In addition, the number of extreme events t_i in any interval of length $(t_2 - t_1) \geq 0$ is distributed as Poisson with mean

$$\int_{t_1}^{t_2} \int_y^{\infty} \lambda(s) ds dt = \tau(y)(t_2 - t_1).$$

Indeed, if the time scale in (1) is measured, for example in months, then the corresponding version of (2) is precisely the probability that a set $A = [t_1, t_2) \times (y, \infty)$ is empty, or in other words, that the maximum of this month is smaller than u . Another important result of extreme value theory is the limiting conditional probability that $Y > u + y$ given $Y > u$

$$\frac{\tau(u+y)}{\tau(u)} = \left(1 + \frac{\xi y}{\sigma + \xi(u-\mu)}\right)^{-1/\xi} = \bar{G}_{\xi, \beta}(x),$$

which is just the survival function of the GPD, i.e., $\bar{G} = 1 - G$, with scaling parameter $\beta = \sigma + \xi(u - \mu)$ for $0 \leq y < y_F$. Here y_F is the right endpoint with values $y_F = \infty$ if $\xi > 0$ and $y_F = -\beta/\xi$ if $\xi < 0$.

3.2. The marked point process viewpoint of extreme value theory

A more general framework for EVT, which allows for time-dependent behavior, is based on viewing the high level of exceedances as a *marked point processes* (MPP). In many stochastic process models, a marked 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. In this way, dependence on covariates or other time dependent variables may be incorporated into the model.

In this paper we define a MPP N as a set of observations, occurrence times and marks $\{(t_i, y_i)\}_{i=1}^T$ on the space $\mathcal{T} \times \mathcal{Y}$, whose history $\mathcal{H}_t = (\{t_1, y_1\}, \dots, \{t_{t-1}, y_{t-1}\})$ consists only of the occurrence times and marks $\{t_1, y_1\}, \dots, \{t_{t-1}, y_{t-1}\}$ up to time t but not including t . Moreover, we define a point process N_g “the ground process” which refers to the stochastic process of the inter-exceedance times. This point process has a conditional density function $p(t | \mathcal{H}_t)$ and its corresponding survival distribution function $S(t | \mathcal{H}_t)$. The conditional (finite) intensity function (or hazard function) for the ground process N_g is given by

$$\lambda_g(t | \mathcal{H}_t) = \frac{p(t | \mathcal{H}_t)}{S(t | \mathcal{H}_t)}, \quad (3)$$

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

$$\lambda(t, y | \mathcal{H}_t) = \lambda_g(t | \mathcal{H}_t) f(y | \mathcal{H}_t, t), \quad (4)$$

where $f(y | \mathcal{H}_t, t)$ 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 N uniquely. Furthermore, we say that a MPP N on $\mathcal{T} \times \mathcal{Y}$ has independent marks, if given the ground process N_g the marks y_i are mutually independent random variables such that its distribution depends only on the corresponding location t_i . In addition, we define a MPP as having unpredictable marks for \mathcal{T} , if the distribution of the marks at t_i is independent of the locations and marks $\{(t_j, y_j)\}$ for which $t_j < t_i$. For a more formal introduction to marked point processes we refer to Daley and Vere-Jones (2003, p. 246).

According to our definition of MPP, the marks are conditionally independent of the associated ground process. Therefore, the product of mark densities has to be simply multiplied with the likelihood of the ground process. Letting N be a MPP on $[t_0, T) \times \mathcal{Y}$ for some finite positive T and let $(t_1, y_1), \dots, (t_{N(T)}, y_{N(T)})$ be a realization of N , we can obtain the log-likelihood L of such

a realization in terms of the conditional densities or intensities as

$$\begin{aligned}
L &= \sum_{i=1}^{N(T)} \log p_i(t_i | \mathcal{H}_t) + \sum_{i=1}^{N(T)} \log f_i(y_i | \mathcal{H}_t, t) \\
&= \sum_{i=1}^{N(T)} \log \lambda_g(t_i | \mathcal{H}_t) - \int_{t_0}^T \lambda_g(s | \mathcal{H}_t) ds + \sum_{i=1}^{N(T)} \log f_i(y_i | \mathcal{H}_t, t).
\end{aligned} \tag{5}$$

Observe that an alternative description of the non-homogeneous Poisson process (2) is by rewriting this as a special case of a MPP in terms of a ground process N_g with rate of the one-dimensional Poisson process of exceedances of the level u , i.e., $\tau = \lambda_g(t | \mathcal{H}_t) = -\ln H_{\xi, \mu, \sigma}(u)$, and a GPD function for the marks $f(y | \mathcal{H}_t, t) = \frac{1}{\beta} \left(1 + \xi \frac{y-u}{\beta}\right)^{-1/\xi-1}$

$$\lambda(t, y) = \tau \frac{1}{\beta} \left(1 + \xi \frac{y-u}{\beta}\right)^{-1/\xi-1}. \tag{6}$$

This is exactly the idea that we want to explore in the next section. We will concentrate on models where the conditional intensity for the ground process will be parametrized in terms of an interval between two consecutive extreme events $x_i = t_i - t_{i-1}$ such that the impact of a duration between successive events depends upon the number of intervening extreme events. The main area of application of these models has traditionally been in modeling of high frequency financial data. Their structure, however, would also seem appropriate for modeling extreme events and the tremors that follow these.

3.3. The autoregressive conditional duration peaks over threshold model (ACD-POT)

As mentioned in the introduction, exceedances of a high threshold for daily financial returns do (in contrast to iid data) not necessarily occur according to a homogeneous Poisson process. Thus, the classical POT model is not directly applicable to financial return data.

Therefore, we consider autoregressive conditional duration models for the conditional intensity of the ground process $\lambda_g(t | \mathcal{H}_t)$. In particular, we propose a set of models, which allow for 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 ψ_i be the expectation of the i -th inter-exceedance time given by

$$\mathbb{E}(x | x_{i-1}, \dots, x_1) = \psi_i(x_{i-1}, \dots, x_1; \theta) \equiv \psi_i, \tag{7}$$

where θ is a parameter vector. We assume that ψ_i corresponds to the ACD class of models. In general, the assumption is based on the fact that the standardized durations

$$\varepsilon_i = \frac{x_i}{\psi_i} \quad (8)$$

are iid random variables. Thus, the key idea is that the time dependence between the inter-exceedance times can be subsumed in their conditional expectations ψ_i , in such way that $\frac{x_i}{\psi_i}$ is independent and identically distributed. To derive a general expression for the conditional intensity let p be the density function of (8)

$$p\left(\frac{x_i}{\psi_i} \mid \mathcal{H}_i; \theta\right) = p\left(\frac{x_i}{\psi_i} \mid \theta\right), \quad (9)$$

where θ 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 $[t_0, T) \times \mathcal{Y}$ for some finite positive time T and let $(t_1, y_1), \dots, (t_{N(T)}, y_{N(T)})$ be a realization of N over the interval $[0, T)$, one can easily show that the conditional expected intensity of the inter-exceedances 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

$$\lambda_g(t \mid \mathcal{H}_t; \theta) = \lambda_0 \left(\frac{t - t_{N(T)}}{\psi_{N(T)}} \right) \frac{1}{\psi_{N(T)}}. \quad (10)$$

Furthermore, we also consider the case where the marks are conditionally generalized Pareto, given the history \mathcal{H}_t . To this end, we parameterize $\beta(t, y \mid \mathcal{H}_t)$ such that it depends on the history². 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. The ACD-POT model is also defined as follows

$$\lambda(t, y \mid \mathcal{H}_t; \theta) = \frac{\lambda_0 \left(\frac{t - t_{N(T)}}{\psi_{N(T)}} \right)}{\psi_{N(T)} \beta(t, y \mid \mathcal{H}_t)} \left(1 + \xi \frac{y - u}{\beta(t, y \mid \mathcal{H}_t)} \right)_+^{-1/\xi - 1}. \quad (11)$$

Effectively we have combined the one-dimensional intensity in (10) with a generalized Pareto density. Under this model the conditional rate of crossing the threshold $x \geq u$ at time t , given the

²We can also parameterize the shape coefficient ξ . However, the behavior of the estimation is severely affected. For this reason it is reasonable to assume the shape parameter to be constant.

history \mathcal{H}_t up to that time, is

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

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

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

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

In the following subsections we introduce specific models that enable to parametrize the expected conditional duration function ψ_i , the distribution of probability of the standardized durations ε_i and the models for the scale parameter $\beta(t, y | \mathcal{H}_t)$.

3.3.1. 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 ψ . In addition, we introduce parametrizations that allow not only for linear but also for more flexible innovations 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) : $\psi_i = w + a x_{i-1} + b \psi_{i-1}$.
- (Log-ACD) The logarithmic ACD model introduced by Bauwens and Giot (2000) : $\psi_i = \exp \{ w + a x_{i-1} + b \psi_{i-1} \}$, where $w > 0, a, b \geq 0$.
- (BCACD) The Box-Cox-ACD model (Dufour and Engle (2000)): $\psi_i = w + \frac{a}{\delta} (\varepsilon_{i-1}^\delta - 1) + b \psi_{i-1}$.
- (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 $\delta \rightarrow 0$ 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).

3.3.2. Distributional assumptions for the standardized durations

Besides the specification of the conditional mean function, another important issue in the parametrization of our ACD-POT model is the choice of the innovation process. There are many options for choosing the distribution for ε_i , as long as it is a probability distribution on the real positive line with zero. Engle and Russell (1998) used the exponential distribution and considered using the Weibull distribution.

In this subsection we explore two alternatives given by the Burr (Grammig and Maurer, 2000.) and the generalized gamma distribution (Lunde, 1999). The major advantage of these distributions over the exponential and Weibull distribution, is that they have non-monotonic hazard functions taking bathtub shaped or inverted U-shaped forms. This feature is of particular importance if we are interested in modeling risk measures such as the VaR or the expected shortfall.

The first alternative is the generalized gamma distribution introduced by Lunde (1999) in the context of ACD models to characterize the standardized durations. A three parameter generalized gamma density is given by

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

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 $\varphi(\psi_i) = \phi_i = \psi_i \frac{\Gamma(k)}{\Gamma(k + \frac{1}{\gamma})}$ which implies a conditional density of the standardized duration given by

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

where θ 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

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

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

In addition, the shape properties of the conditional hazard function can be derived from its parameter 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 of an ACD-POT model under this distributional assumption takes the form

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

The conditional log-likelihood function of this model on a set of observed inter-exceedance times and of marks or sizes of the exceedances can be derived easily from (5)

$$\begin{aligned} L = & \sum_{i=1}^{N(T)} \left\{ \log \gamma + (k\gamma - 1) \log\left(\frac{x_i}{\phi_i}\right) - \log(\Gamma(k) \phi_i) - \left(\frac{x_i}{\phi_i}\right)^\gamma \right\} \\ & - (1 + 1/\xi) \sum_{i=1}^{N(T)} \log\left(1 + \xi \frac{y_i - u}{\beta(t, y | \mathcal{H}_t)}\right)_+. \end{aligned}$$

The second alternative considered in this paper is the Burr distribution introduced in the context of ACD models by Grammig and Maurer (2000). The density function is defined by

$$f(x | \lambda, k, \gamma) = \frac{\lambda k t^{k-1}}{(1 + \gamma^2 \lambda t^k)^{\gamma^{-2}+1}}.$$

In this case we define $\varphi(\psi_i) = \phi_i = \psi_i \frac{\gamma^{2(1+\frac{1}{k})} \Gamma(\gamma^{-2}+1)}{\Gamma(1+\frac{1}{k}) \Gamma(\gamma^{-2}-\frac{1}{k})}$, where $0 < \gamma^{-2} < k$. We choose the density (9) to be Burr under the restriction that $\hat{\lambda} = 1$,

$$p\left(\frac{x_i}{\phi_i} | \mathcal{H}_t; \theta\right) = \frac{k \phi_i^{1-k} x_i^{k-1}}{(1 + \gamma^2 \phi_i^{-k} x_i^k)^{\gamma^{-2}+1}}.$$

The implied conditional hazard function in this case is

$$\lambda_g(x_i | \mathcal{H}_t; \theta) = \frac{k \phi_i^{-k} x_i^{k-1}}{1 + \gamma^2 \phi_i^{-k} x_i^k}, \quad (12)$$

which is non-monotonic function with respect to duration. From (12) can be obtained: the Weibull-ACD model if $\gamma^2 \rightarrow 0$, Exponential-ACD model if $\gamma^2 \rightarrow 0$ and $k = 1$ and Log-Logistic ACD for

$\gamma^2 = 1$. The conditional intensity in this case takes the form

$$\lambda(t, y | \mathcal{H}_t; \theta) = \frac{k\phi_i^{-k}x_i^{k-1}}{1 + \gamma^2\phi_i^{-k}x_i^k} \frac{1}{\beta(t, y | \mathcal{H}_t)} \left(1 + \xi \frac{y - u}{\beta(t, y | \mathcal{H}_t)}\right)_+^{-1/\xi - 1}.$$

The conditional log-likelihood function of an ACD-POT model under this distributional assumption on a set of observed inter-exceedance times and of marks can also be obtained from (5), i.e.,

$$\begin{aligned} L = & \sum_{i=1}^{N(T)} \left\{ \log k - k \log \phi_i + (k-1) \log x_i - (1 + \gamma^{-2}) \log \left(1 + \gamma^2 \phi_i^{-k} x_i^k\right) \right\} \\ & - (1 + 1/\xi) \sum_{i=1}^{N(T)} \log \left(1 + \xi \frac{y_i - u}{\beta(t, y | \mathcal{H}_t)}\right)_+. \end{aligned}$$

3.3.3. Models for the time varying scale parameter

In this section we consider different models to parametrize the scaling parameter $\beta(t, y | \mathcal{H}_t)$ such that it depends on the history. This feature implies that the marks are conditionally generalized Pareto, given the history \mathcal{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. Let t_* y_* correspond to the time and mark of the last extreme events that occur before the time t . We will specify and estimate five alternatives forms for the scaling parameter $\beta(t, y | \mathcal{H}_t)$.

1. The constant scale: $\beta(t, y | \mathcal{H}_t) = \beta_1$.
2. The lineal scale: $\beta(t, y | \mathcal{H}_t) = \omega + \beta_1 y_* + \beta_2 \psi_*$.
3. The polynomial scale: $\beta(t, y | \mathcal{H}_t) = \omega + \beta_1 y_* + \beta_2 \psi_*^{\beta_3}$.
4. The Hawkes³ scale: $\beta(t, y | \mathcal{H}_t) = \omega + \beta_1 \sum_{i:t_i < t} (1 + \beta_2 y_i) \exp(\beta_3 (t - t_i))$.
5. The autoregressive realized duration (ARD) scale: $\beta(t, y | \mathcal{H}_t) = \omega + \beta_1 \beta(t_*, y_* | \mathcal{H}_{t_*}) + \frac{\beta_2}{(t - t_*)^{\beta_3}}$.

For all models ω and $\forall_i \beta_i \in \mathbb{R}_+$. The first specification considers the constant scale case. The second and third of these correspond to a scaling parameter, which depends on the last mark and (linearly or polynomially) on the conditional mean function of the inter-exceedance times. The ARD scale specification directly includes the realized durations between excesses as covariates in contrast to the first two specifications. Finally, the Hawkes scale specification increases with

³Hawkes models type are frequently used in seismological modeling. See Ogata (1988) and Daley and Vere-Jones (2003) for additional references.

mark occurrences and their magnitudes, and decreases in time away from each event. All of these models have the feature that whenever an extreme event is observed the scale parameter should dynamically adapt. This strategy obviously should result in an improved estimation on measures of risk. In case of VaR, these specifications will typically lead to an increase depending on the size of the marks, the expected durations and/or the realized duration between the last extreme events.

3.4. Risk measures

One main purpose of this paper is to develop a suitable methodology to obtain an expression and its estimate for the quantile of the one day ahead predictive distribution of the returns, conditional on past and current data. In particular, we focus on Value-at-risk (VaR) and Expected shortfall (ES). These measures have become standard measures in financial risk management due to their conceptual simplicity, computational facility and ready applicability. In what follows we derive these measures for the ACD-POT models.

The VaR is defined as the q -th quantile of a distribution F given by

$$\text{VaR}_\alpha^t = y_\alpha^t = \inf \{y \in \mathbb{R} : F_{y_{t+1}|\mathcal{H}_t}(y) = \alpha\},$$

which is solution to $\mathbb{P}(y_{t+1} > y_\alpha^t | \mathcal{H}_t) = 1 - \alpha$. Observe that

$$\begin{aligned} \mathbb{P}(y_{t+1} > y_\alpha^t | \mathcal{H}_t) &= \mathbb{P}(y_{t+1} - u > y_\alpha^t - u | \mathcal{H}_t) \\ &= \mathbb{P}(y_{t+1} - u > y_\alpha^t - u | y_{t+1} > u, \mathcal{H}_t) \mathbb{P}(y_{t+1} > u | \mathcal{H}_t). \end{aligned} \quad (13)$$

The first term in the right hand side of equation (13) can be approximated via

$$\mathbb{P}(y_{t+1} - u > y_\alpha^t - u | y_{t+1} > u, \mathcal{H}_t) = \left(1 + \xi \frac{y_\alpha^t - u}{\beta(t, y | \mathcal{H}_t)}\right)_+^{-1/\xi},$$

while

$$\begin{aligned} \mathbb{P}(y_{t+1} > u | \mathcal{H}_t) &= \mathbb{P}(N(t, t+1) = 1 | \mathcal{H}_t) \\ &= 1 - \exp(-\lambda(t, s | \mathcal{H}_t; \theta)). \end{aligned}$$

Thus the VaR is defined by

$$\text{VaR}_\alpha^t = u + \frac{\beta(t, y | \mathcal{H}_t)}{\xi} \left(\left(\frac{1 - \alpha}{1 - \exp(-\lambda(t, s | \mathcal{H}_t; \theta))} \right)^{-\xi} - 1 \right). \quad (14)$$

The last equation implies that the VaR is only defined for our models if $1 - \exp(-\lambda(t, s | \mathcal{H}_t; \theta)) > 1 - \alpha$. In the case of expected shortfall (ES), it is defined as the average of all losses which are

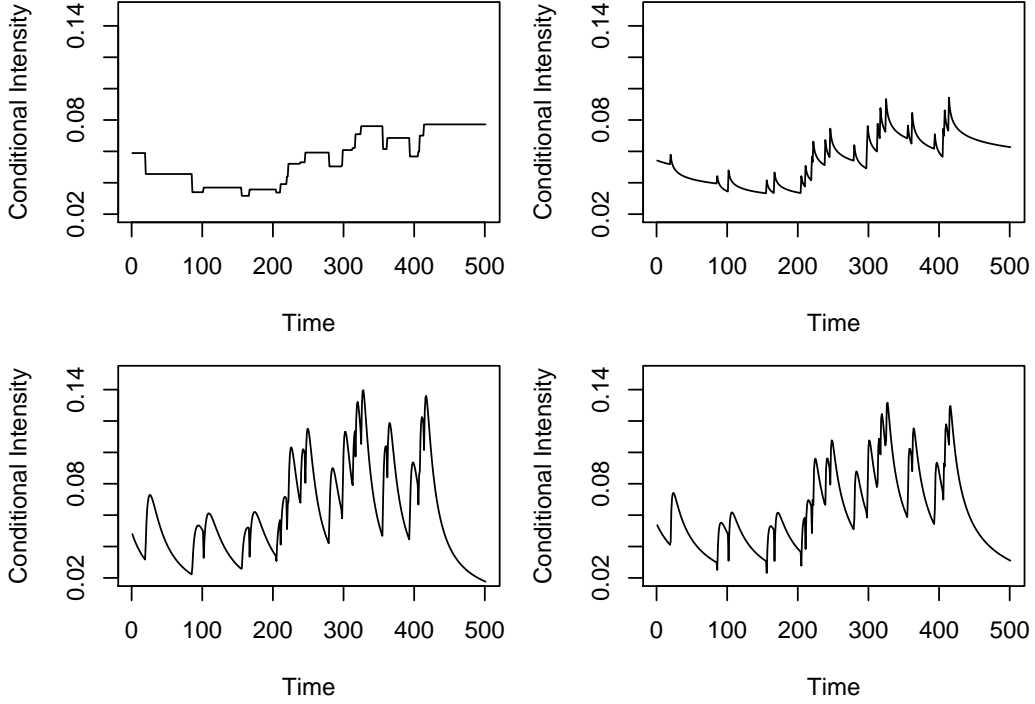


Figure 3: The path of the conditional intensities of an ACD-POT model under four types of distributional assumptions. Top panel: exponential (left) and Weibull; Bottom panel: Burr (left) and generalized gamma

greater or equal to VaR, i.e. the average loss in the worst $(1 - \alpha)\%$ cases $ES_{\alpha}^t = E[Y | Y > \text{VaR}_{\alpha}^t]$. In the models proposed the ES is given by

$$ES_{\alpha}^t = \frac{\text{VaR}_{\alpha}^t}{1 - \xi} + \frac{\beta(t, y | \mathcal{H}_t) - \xi u}{1 - \xi}. \quad (15)$$

3.5. A toy example

It is quite evident that the performance of the models depends on the distributional assumptions and the estimated time varying scale parameter. To gain understanding about their influence we visually analyze the behavior of the estimated conditional intensity of a small sample based on Bayer returns (the entire sample will be analysed later in detail).

At first we only concentrate on the kind of distributional assumption. Figure 3 displays the path of a ACD-POT model with four types of distributions: in the top panel the exponential (left) and the Weibull (right), as subcases of the Burr or generalized gamma distribution, and in the bottom panel the Burr (left) and the generalized gamma (right) distribution. Observe that in case of the exponential and Weibull distributions we have a flat or monotone conditional intensity, respectively. On the other hand, both the Burr and the generalized gamma distribution show a non-

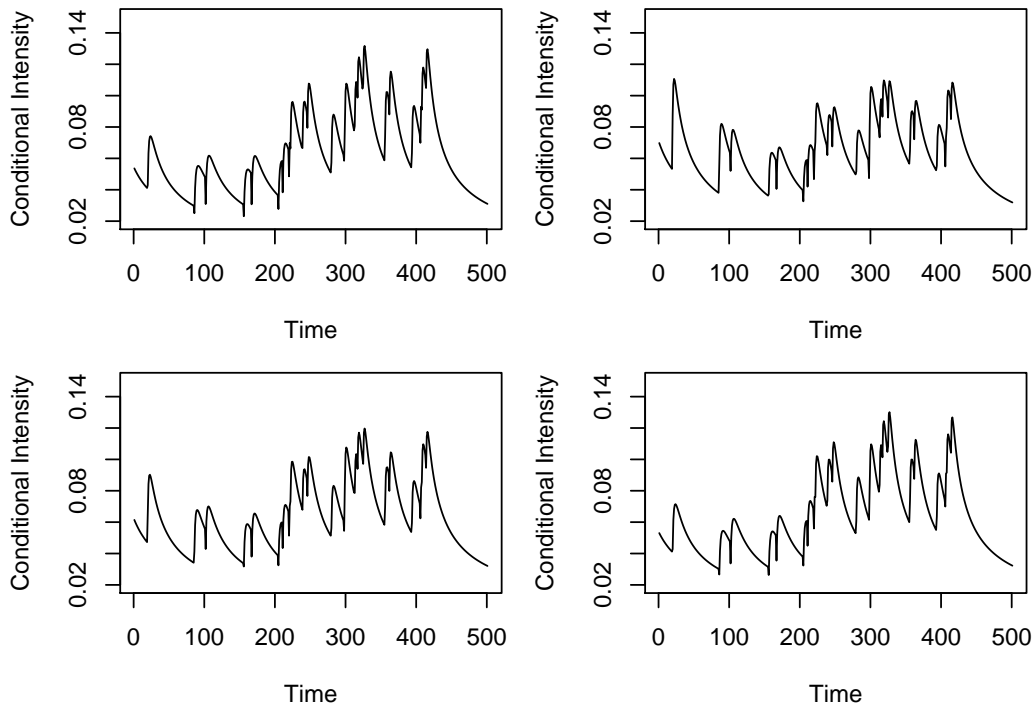


Figure 4: The path of the conditional intensities for the four models proposed under the assumption of a generalized gamma distribution for the innovations: the gACD1 (top left), the gLog-ACD1 (top right), the gEXACD1 (bottom left) and the gBCACD1model (bottom right).

monotone conditional intensity. This important feature allows the last two distributions rapidly adapting the conditional intensity to reach periods of high volatility which are associated with clustering of short inter-exceedance times.

In relation to the type of ACD model for conditional mean duration, Figure 4 shows conditional intensities for the four models, proposed under the assumption of a generalized gamma distribution for the innovations: the ACD (top left), the Log-ACD (top right), the EXACD (bottom left) and the BCACD model (bottom right). At this stage, it is not yet possible to reach a conclusion on the appropriateness of each model. However, before we make a choice, there are some important features of the models to keep in mind. On the one hand, the Log-ACD allows for nonlinear effects of short and long durations in the conditional mean, without requiring the estimation of additional parameters in comparison to the standard ACD model. While on the other hand, the BCACD and the EXACD models offer a captivating compromise between the need of greater flexibility and the burden of higher complexity.

4. Empirical application

In this section we present the main empirical results obtained by testing the models introduced in the previous sections.

4.1. Data Description

For the empirical test we chose the transaction data from three sources:

- Bayer AG, blue chip stock from Germany's DAX.
- DAX index, a market value-weighted portfolio of 30 major German companies traded at the Frankfurt Stock Exchange.
- A portfolio of international equity indices previously analysed in McNeil et al. (2005) and Chavez-Demoulin et al. (2011). This portfolio value is standardized to have weights 30% FTSE100, 40% S&P 500 and 30% SMI. The portfolio is assumed to have domestic currency sterling (GBP) and consequently has currency exposure to US dollar (USD) and Swiss franc (CHF). The value of the portfolio is therefore influenced by five risky factors: three log-index values and two log-exchange rates.

The sample period for Bayer and DAX 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 from 20 January 2008 to January 16, 2009. These data were obtained from Datastream. In case of the portfolio, the data comprises daily closing prices, from January 3, 1991 to December 29, 2000. For backtesting we consider a sample from January 3, 2001 to December 28, 2001. This data was obtained from the R library QRMLib, see McNeil et al. (2005).

In this study we concentrate only 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 . 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 the estimation of the risk measures as accurately as possible.

Table B.1 presents some relevant summary statistics about the unconditional distribution of the returns. The statistics show that all returns exhibit skewness to the losses as well as excess of kurtosis. Serial correlation was not rejected for DAX and Bayer returns, but for the portfolio returns with p-value of basically zero. For the squared returns the null hypothesis of no heteroscedasticity was clearly rejected for the three series. This probably anticipates the existence of volatility clusters.

4.2. Model fit and model selection

In order to summarize adequately the large quantity of empirical results, we use a classification scheme for the ACD-POT models. The first lower-case letter describes the type of distributional assumption with respect to the ACD model (generalized gamma or Burr). The following capital letters denote the type of ACD model: ACD, Log-ACD, BCACD or EXACD⁴. The number after the ACD model denotes the models for the time varying scale parameter: 1 (constant), 2 (linear), 3 (polynomial), 4 (Hawkes) and 5 (ARD). For example, a model gLog-ACD1 means that we are working with a Log-ACD model for the expected conditional duration with generalized gamma distribution and constant scale parameter. In total we have 40 models.

All parametric models are estimated using quasi maximum likelihood. We adopt different models according to our scheme classification in order to test the different ACD models with different distributional assumptions regardless of the possibility of a model with constant scale parameter⁵ (marks with iid GPD) or varying scale parameter (marks with GPD whose scale parameter is time-dependent).

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. There is no unique choice of the threshold level. A number of diagnostic techniques exist for this purpose, including graphical, bootstrap methods (see Embrechts et al. 1997; Falk et al. 2004). However, the choice of the optimal threshold is still considered an open problem and different approaches have been proposed to overcome this difficulty. The optimal choice is inspected via a sensitivity analysis which allow us to verify whether the tail index of the ACD-POT models remains stable among different thresholds. The procedure is detailed in Appendix A. In this paper we choose to work with the 10% of the maxima of the sample.

Since we report the empirical results for a large number of models⁶, we decided to reduce the number of models, so that we concentrate on the best results and the models of interest. However, the complete set of results is available from the authors upon request. Notwithstanding the above, some comments about the results can be made. For the inter-exceedance times, the generalized gamma distribution seems to give better estimations than that of the Burr distribution. The results on ACD models for the expected conditional duration seem to markedly favor the Log-ACD specifications, followed by the ACD. Finally, the models with time varying scale parameters lead to a

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

⁵Observe that this is equivalent to fitting separately a ACD model to the inter-exceedance times and a GPD to the excess losses over the threshold.

⁶For each return series we obtain 40 different models together with a large number of measures of GoF for different levels of VaR.

better fit. The Hawkes specification provides an uniformly better fit among the returns, followed by the linear and polynomial approach. In short, the results suggest that the models with time varying scale parameter have a better fit and 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.

We follow a pragmatic approach in choosing the model that is best suited to return modeling and the VaR requirements. The key idea is to select the best models fitted according to AIC and that, in relation to the VaR in-sample and the different Goodness of fit (GoF) measures, display the most accuracy. To this end, we propose a modified AIC (mAIC), which considers this idea:

$$AIC_* = AIC - 2f,$$

where AIC is the classical estimation and f is the number of tests that correctly accept the Null hypothesis for the GoF of inter-exceedance times, of the marks or exceedances, and of the accuracy measures for the VaR at levels ($\alpha = 0.05, 0.01, 0.001$), with a statistical significance at 5 %.

For each return series we select the best three models for each distributional assumption for the standardized durations. The maximum log-likelihood estimates of these ACD-POT models and their parameters are displayed in Table B.2. Regarding the dynamic of the VaR, we display the results on the estimation of the conditional 99% VaR in-sample in Figure B.7. Observe how this estimates provides a time-dependent VaR that is sensitive to short and large time scale volatility changes.

For the Bayer return series, under the assumption of that the standardized durations are distributed Burr, the best models fitted according to mAIC are bEXACD4, bLog-ACD3 and bACD3, while that under the assumption of that the standardized durations are distributed generalized gamma are gLog-ACD2, gLog-ACD3 and gACD3. Regarding the AIC of the models proposed, the best fitted model for the Bayer returns is a ACD model with generalized gamma distribution and Hawkes specification for the scale parameter (gACD4) with AIC of 4056.19. However, the model that shows the best accuracy in relation to the measures of GoF is an gLog-ACD model with polynomial specification for the scale parameter. Observe that in this model the expected duration has a direct (no-lineal) relation with the accuracy of the VaR (see Table B.3). Indeed the parameter $\beta_3 = 2.353$ (0.867) is the most significant coefficient in the specification of the time varying scaling parameter $\beta(t, y | \mathcal{H}_t)$.

In addition, for the gACD4 model we observe that $k = 60.007$ (38.884), $\gamma = 0.117$ (0.038), which implies that $k\gamma > 1$ and $\gamma < 1$ so that the hazard rate is inverted U-shaped⁷. In relation to

⁷This remains true for all the models with generalized gamma distribution.

the results of estimation of the conditional GPD model to the exceedances we obtained $\xi=0.091$ (0.088), $\omega= 0.663$ (0.041), $\beta_1=170$ (0.098), $\beta_2= 0.075$ (0.027) and $\beta_3= 0.279$ (0.366). This result indicates that the Hawkes form to parametrize the scaling parameter $\beta(t,y | \mathcal{H}_t)$, such that it depends on the history, was a good choice. Interestingly, the size of the exceedances are as important as the expectation of the old inter-exceedance times.

In relation to the DAX returns, the best models fitted are bACD4, bLog-ACD and bEXACD for the case where the standardized durations are distributed Burr. Under the generalized gamma assumption the best models are gACD2, gLog-ACD2 and gLog-ACD3. The overall best fitted model is a Log-ACD model with generalized gamma distribution and polynomial specification for the scale parameter (gLog-ACD3) with AIC of 3775.03. Also for this observations the best models fitted have a generalized gamma distribution and show a inverted U-shaped hazard rate.

Finally we consider the results of the ACD-POT models fit to the hypothetical portfolio of international equity indexes. In this case the best models are the ACD2, Log-ACD2 and Log-ACD4 either the standardized durations are distributed Burr or generalized gamma. Like the results for the other returns, the best specification for the standardized durations is the generalized gamma distribution. However, it is not easy to select the best choice among the models for the time varying scale parameter. Slightly better performance is given by the gACD4 model according to the mAIC.

For the three returns analysed, we show how the intimate relationship between durations and cluster of extremes can be used to obtain a dynamical better fit of extreme events; by means of a time varying scale parameter and a flexible distributional assumption, which allows the sort of hazard functions 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 and Grammig and Maurer, 2000).

4.3. Comparison with alternative strategies in-sample

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 it is the best model for backtesting. Therefore, we generally 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 could consider further models that can be compared with the previous best model. To this end, we include other models to have a comparison of different alternatives later in the backtest.

GARCH-EVT

The first alternative is the two step method (GARCH-EVT) introduced by McNeil and Frey (2000). The first step consists in fitting a time-varying volatility model, as a GARCH model, to

the data and estimating the tail of the filtered or standardized residuals, by an EVT model, yielding an estimate for the standardized quantile. This procedure is a conditional extreme value approach, so that a correct model specification of the volatility and mean dynamics is necessary. In this paper we will adopt three conditional models:

- ARMA-GARCH-EVT with Gaussian errors (CondN) as the original model introduced by McNeil and Frey (2000). The best specifications according to the AIC for the Bayer, DAX and the portfolio returns are ARMA(2,0)-GARCH(1,1)-EVT, ARMA(2,2)-GARCH(2,1)-EVT, and ARMA(3,3)-GARCH(1,1)-EVT, respectively.
- ARMA-GARCH-EVT with t-Students errors (CondS). In this case the specifications for the Bayer, DAX and the portfolio returns are ARMA(2,0)-GARCH(1,1)-EVT, AR(2,2)-GARCH(2,1)-EVT, and ARMA(2,2)-GARCH(1,1)-EVT, respectively.
- ARMA-APARCH-EVT with Skew t-Students errors (CondST) in order to better account for conditional asymmetry and heavy-tailedness⁸. In this case the specifications for the Bayer, DAX and the portfolio returns are ARMA(0,0)-APARCH(1,1)-EVT, AR(2,2)-APARCH(2,1)-EVT, and ARMA(1,1)-GARCH(1,1)-EVT, respectively.

In comparison with unconditional EVT, different authors (McNeil and Frey 2000; Gencay et al. 2003) demonstrate that this conditional methodology produces the most accurate forecasts of extreme losses both for standard and more extreme VaR quantiles and not only in normal market conditions but also in extreme market conditions. This is due to the fact that conditional VaR estimates embrace different volatility regimes, varying a lot more than unconditional ones. Thus, this model captures the benefits of both EVT and conditional volatility methodology.

CAViaR

The second alternative approach is to use quantile regression based methods as in Engle and Manganelli (2004) who consider an autoregression of the estimated VaRs. Thus, while statistical volatility models rely on the assumption that the shape of the conditional distribution is fixed over time and that it is only the volatility that varies. The recently proposed Conditional Autoregressive Value at Risk (CAViaR) model requires no such assumption, and allows quantiles to be modeled directly in an autoregressive framework. The key assumption is that the linear regression model is $r_t = \mathbf{x}_t' \beta_\alpha + \varepsilon_{t,\alpha}$, where r_t are the returns, and the conditional quantile function is given by

⁸The APARCH, or APGARCH, model of Ding et al. (1993) nests several of the most popular univariate parameterizations.

$Q_\alpha(r_t | x_t) = \mathbf{x}_t' \beta_\alpha$. Note that the distribution of the error term is left unspecified. Engle and Manganelli (2004) show that α -th regression quantile is defined as any $\hat{\beta}_\alpha$ that solves the following generalized objective function

$$\min_{\beta} \frac{1}{T} \left\{ \sum_{r_t \geq VaR_t} \alpha |r_t + VaR_t| + \sum_{r_t < -VaR_t} (1 - \alpha) |r_t + VaR_t| \right\}$$

with $VaR_t = Q_\alpha(r_t | x_t)$. The main advantage of this methodology is that no explicit distributional assumptions need to be made, guarding against this source of model misspecification. In this paper we consider four alternatives initially proposed by Engle and Manganelli (2004):

- The adaptive ($CAViaR_{ad}$): $VaR_t = VaR_{t-1} + \beta_2 \left\{ \left[1 - \exp(\beta_2 [r_{t-1} - VaR_{t-1}])^{-1} - \alpha \right] \right\}$.
- The symmetric absolute value ($CAViaR_{sa}$): $VaR_t = \beta_1 + \beta_2 VaR_{t-1} + \beta_3 |r_{t-1}|$.
- The asymmetric slope ($CAViaR_{as}$): $VaR_t = \beta_1 + \beta_2 VaR_{t-1} + \beta_3 \max(r_{t-1}, 0) + \beta_4 \max(-r_{t-1}, 0)$
- The indirect GARCH(1, 1) approach ($CAViaR_{GARCH}$): $VaR_t = \sqrt{\beta_1 + \beta_2 VaR_{t-1}^2 + \beta_3 r_{t-1}^2}$

The adaptive model, as the name suggests changes itself depending on whether or not VaR is exceeded. It takes a higher value when VaR is exceeded but decreases slightly otherwise. These last three models are similar to GARCH models in structure, the second and the fourth model are symmetrical, and hence respond symmetrically to past returns, while the third model responds asymmetrically to returns and captures the asymmetric leverage effect.

4.4. Measures of goodness of fit

To assess the predictive performance in-sample and out-sample (backtest) of the models under consideration, we divide these into: measures of Goodness of fit (GoF) for inter-exceedance times (GoF ACD), GoF for the marks or exceedances (GoF POT) and measures of accuracy for the VaR.

For the Gof for inter-exceedance times we employ density forecasting techniques for ACD models introduced by Bauwens et al. (2004) by means of a Pearson statistic (χ^2) together with the Kolmogorov-Smirnov test (KS_{ACD}) and the Anderson-Darling test (AD) for the standardized durations. In addition, in order to check that there is no further time series structure the Ljung-Box test (LB_{ACD}) is also included.

In the case of the GoF for the marks we utilize the W-statistics (Smith, 2003) to assess our success in modeling the temporal behavior 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. Further, we utilize the Kolmogorov-Smirnov test (KS_{POT}) statistic to test that the residuals are approximately unit exponential variables.

Similarly, we provide empirical evidence on the accuracy of actual VaR measures derived from the models. The first of them is an unconditional coverage (LR_{uc}) test (Christoffersen, 1998). 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 or Hit is defined as occurring when the ex-post return is lower than the VaR. A second test proposed by Christoffersen (1998) is a test of independence (LR_{ind}) between violations of the VaR, where under the null hypothesis a violation today has no influence on the probability of a violation tomorrow. The third test is a combination of the last two test which is known as the conditional coverage (LR_{cc}) test. The fourth approach proposed by Berkowitz et al. (2009) tests for uncorrelatedness of the violations. In particular, we suggest the well-known Ljung-Box (BT) test of the violation sequence's autocorrelation function. The last two tests, named the Dynamic Quantile (DQ) tests, were introduced by Engle and Manganelli (2004). The idea is to regress the violations on the VaR for the present period on a judicious choice of explanatory variables. In the first case, denoted by the DQ_{hit} , the regressor vector contains a constant and lagged violations of the VaR, while the second, DQ_{VaR} , additionally uses, the contemporaneous VaR estimate. All of these methods of GoF are reviewed briefly in the Appendix Appendix B.

4.5. Relative performance of the ACD-POT models in-sample

The results on the goodness of fit in-sample under the alternative modeling assumptions are reported in Tables B.3, B.4, and B.5, for Bayer-, DAX- and portfolio returns, respectively.

In relation to the ACD-POT models, the performance varies substantially across the modeling approaches as well as the distributional assumptions, however, some clear patterns emerge. In particular, the superior performance of the generalized gamma distribution relative to the Burr is clearly present in all the ACD-POT models. Observe that the main benefit of modeling extreme events by means of the ACD-POT methodology over other EVT specifications or the competitive models used here, lies in the explicit modeling of the inter-exceedance times, which is particularly reflected in the GoF ACD. Regarding the GoF POT related to the distribution of the marks or exceedances, all ACD-POT models show an acceptable level in any of the testing categories. Finally, the measures of accuracy for the VaR for the ACD-Models display a top performance at all the VaR levels with respect to correct unconditional and conditional coverage, as well as for any dynamic quantile test.

In what follows, we briefly discuss the performance of the CAViar models. Based on the results reported, we can observe that for all the returns analysed this methodology is a simple and accurate

approach to forecast VaR. However, a pitfall of the CAViar models is (with one exception) that their functional forms do not account for the dynamics in the clustering of extreme events. Therefore, CAViar is not able to produce iid VaR violations, which leads to strong rejection of independence of the violation sequences for all CAViar models used by mean of the DQ_{VaR} , under which the contemporaneous VaR estimate is also included in the regression.

Next, we turn to the results of the GARCH-EVT models, which deliver mixed results. In particular, the proper specification of the volatility dynamics is clearly the key point for the accuracy of the VaR estimates. Of course, it is far from obvious as to which specification will be optimal, and the decision should be based on out-of-sample VaR forecasting performance rather than in-sample. Regardless, the empirical results for all returns clearly demonstrate that the most flexible approach, in this case the CondST model, performs best. Specially, using a fat-tailed and asymmetric distribution we would expected to improve VaR forecast. Nevertheless, according to the GoF POT measures, we obtain an adequate W-statistic only for the portfolio returns, which state that the residuals are approximately independent unit exponential variables in the case that the GPD specification for the exceedances is correct. This is also confirmed by means of the Ljung-Box test (LB_{POT}) statistic which examines the null hypothesis of independence of the exceedances. Regarding the accuracy of the VaR estimate, our findings indicate that as far as the unconditional and conditional coverage is concerned, the normal assumption for the residuals do not seem to be the most adequate, while the skew Student's-t and Student's-t assumptions yield quite accurate results for most of the returns analysed for VaR at 0.01 and 0.001 levels. Although, for a 0.05 VaR level, all of these models tend to perform worse. Regarding the independence of the VaR violations, the results are mixed for returns analysed. According to the LR_{ind} and DQ tests applied, the explanation is that variation in volatility results in substantial variation in the mean inter-exceedance time between the marks or exceedances, so that no homogeneous result for independence at each VaR level can be obtained.

Summarizing the results for the ACD-POT models: Major improvements in VaR predictions are achieved in all aspects when the clustering dynamic of extreme events are taken into consideration. Indeed, by means of the GoF ACD we can test misspecification of the ACD models for the expected conditional duration and the distributional assumption standardized durations. Further, through the GoF POT we can control the correct conditional distribution of the marks or exceedances. Finally, regarding the returns analysed, as a result of these two improvements, the VaR violations are reasonably independent when using an ACD-POT model either the standardized durations are assumed Burr or generalized gamma - the latter being preferred overall. The above implicates that the ACD-POT based approaches outperform the basic specifications of

CAViar models. According to the VaR estimations, ACD-POT and GARCH-EVT methodology were the only methods that, more often than not, eradicates the threat of violation clustering.

4.6. Backtesting the models

Backtesting provides invaluable feedback about the accuracy of the models proposed to risk managers. The performance of VaR w.r.t. backtesting has been carried out with the daily returns series, which are scaled by 100, for one year of the three return series. The DAX and Bayer data are backtested from 20 January 2008 to January 16, 2009, while in the case of the portfolio we consider a sample from January 3, 2001 to December 28, 2001.

The backtest method consists of 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 re-estimate the models, something that immediately reveals a models' possible stability problems. Then, we re-estimated the risk measures for each return series.

For the backtest we only consider the best performing models in the in-sample estimation. To this end, we choose the best two models in each methodology in relation to each return series. In total we will consider six models per return series.

The results for 0.1% to 5% VaRs of the Bayer returns, which also covers the financial crisis period, are presented in Table B.6 and includes the VaR violations and the p-values for the different tests of unconditional and conditional coverage, and dynamic and no-dynamic independence. In relation to the ACD-POT models, we consider the gLog-ACD3 and the gLog-ACD4 specification. The performances for the models are similar to the results on VaR forecasting, although we observe some differences. For instance, the models tend to overestimate the 0.1% VaR. This was most likely due to the impact of the subprime crisis in the German financial markets. Indeed, two violations correspond to the beginning of the market turmoil. The results for the GARCH-EVT methodology are not so different. In this case again the model specifications have difficulties to make a correct estimation of the VaR at all confidence level. However, the 0.1% VaR a correct estimate, while that for 1% and 5% VaRs is not efficient enough. Finally, the results of the CAViar specification suggest that these models display the best performance for the Bayer returns.

In relation to the DAX index, the main result in Table B.6 is that, considering the models proposed, the ACD-POT models and the GARCH-EVT specifications are always more precise than the CAViar methodology. This result holds true across different tests, as does, for example, unconditional and conditional VaR coverage, independence tests and VaR confidence levels chosen. The differences in some test are sometimes very large. For instance, in Table B.6, using a 5% significance level for any VaR the correct confidence level the VaR coverage is rejected. Further,

the DQ_{VaR} test clearly appears to pick up dependence in the VaR violations, which is ignored by the Markov test. In relation to the ACD-POT approach, the models with the best performance are the gACD2 and the gLog-ACD3. Observe that the two models seem to be correctly specified and highly precise, as measured by the accuracy of the VaR.

We finally consider the hypothetical portfolio of international equity, which are influenced by five risk factors. In this case both ACD-POT models show a more accurate fit and a superior performance in relation to the VaR prediction than the other two alternatives. This result holds true across VaR coverage rates and independence tests for the violations at all significance levels chosen. In Table B.6, the gACD4 and gLog-ACD4 models were the only methods that did not underestimate the 5% VaR while the other two methods clearly underestimated the probability of violation. Finally, due to the shortness of the time horizon we do not find a VaR violation for the 0.001 quantile, and therefore the p-values for the tests of independence 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 ACD-POT models proposed are suitable for the estimation of different risk measures, as for example, the VaR according to the restriction imposed by Basel Committee on Banking Supervision (1996, 2006). Moreover, these models allow us to take the heavy-tailedness or the stochastic nature of the cluster of extreme events into consideration. Regarding the returns analysed at a 5% level of significance, the ACD-POT approaches pass 86 out of the 102 tests, the GARCH-EVT specifications pass 26 out of the 102, whereas the CAViar approaches only pass 33 out of the 102 tests.

In relation to the CAViar methodology, our results are quite striking in relation to the Engle and Manganelli (2004) results. Contrary to the results reported by them, a weaker CAViar performance is estimated for index data than for individual stock returns. The fact that the Bayer stock market and the DAX index comprise a highly volatile backtesting year may presumably deteriorate the performance of the established CAViar models with the more positive findings of Engle and Manganelli (2004). The key point is that most of the CAViar specifications used in this investigation do not account for volatility clustering, and therefore, they are not able to produce iid VaR violations, causing us to strongly reject independence of the violations and the contemporaneous VaR estimates for most of the models.

Regarding the GARCH-EVT approach, the performance varies substantially across the modeling approaches as well as the distributional assumptions. The main findings indicate that, at least for the returns analysed, the 5% quantile is still not large enough to be analysed by means of this approach. However, major improvements in VaR predictions are achieved in all aspects when we account for the volatility dynamics and the extreme events.

5. Conclusions and proposals for future work

This paper proposed a new technique for modeling extreme events of stationary sequences as is the case for 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 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 Autoregressive Conditional Duration (ACD) models which are popular in finance. For this reason we describe it as the ACD-POT model.

We observe that under this methodology the estimation of such models can be straightforwardly derived through conditional intensities. With the simplicity of the conditional intensities' structure in mind, different models were proposed. However, other more complicated structures could also be adopted, for instance, the stochastic volatility duration model of Ghysels et al. (2004) could be an interesting alternative.

With regard to the empirical application, the models and their estimations with returns from Bayer AG, DAX index and the hypothetical portfolio were more than satisfactory. The 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 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 some selected ACD-POT 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 therefore certainly provides a more precise estimate as the information in the data sample is exploited more efficiently. This particularly refers to clustering of extreme events and the inter-exceedance times. Furthermore, in comparison with others competitive models, in most of the cases the ACD-POT models outperform the basic specifications of CAViar models introduced by Engle and Manganelli (2004). Finally, the ACD-POT and the two stage GARCH-EVT methodology (McNeil and Frey, 2000) most often eradicate the threat of VaR violation clustering.

In summary the ACD-POT models do a very good job of modeling the inter-exceedance times associated with waiting times between extreme events and the size of exceedances. These models may therefore serve as a useful starting point for further extensions. Other possible directions for

future research emerge from the results. For instance, being interested in long term behavior rather than in short term forecasting, the simulation of ACD-POT models enables to calculate measures of risk over other time horizons. Other research options would be different distributional assumptions for the standardized durations or other flexible forms of the self-exciting models, which could be used by incorporating other characteristics of the series, such as trend of increasing exceedances or different regimes as aftershocks. Another idea is to combine ACD with another class of self-exciting models, such as Hawkes- or ETAS- (Epidemic Type Aftershock Sequence) models. This could help to characterize other important features such as slow decay of autocorrelation or a power-law decay between jumps. Finally, the application of these models is not only limited to daily returns. A natural extension is to use this methodology to attain high frequency data in order to estimate intraday measures of risk.

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.
- Bauwens, L., Giot, P., 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., Giot, P., Grammig, J., Veredas, D., 2004. A comparison of financial duration models via density forecasts. *International Journal of Forecasting* 20, 589–609.
- Bauwens, L., Hautsch, N., 2006. Stochastic conditional intensity processes. *Journal of Financial Econometrics* 4, 450–493.
- Bauwens, L., Hautsch, N., 2009. Modelling financial high frequency data using point processes. In: T. G. Andersen, R. A. Davis, J.-P. K., Mikosch, T. (Eds.), *Handbook of Financial Time Series*. Vol. 6. Springer, pp. 953–979.
- Berkowitz, J., Christoffersen, P., Pelletier, D., 2009. Evaluating value-at-risk models with desk-level data. *Forthcoming in Management Science*.
- Chavez-Demoulin, V., Davison, A., McNeil, A., 2005. A point process approach to value-at-risk estimation. *Quantitative Finance* 5, 227–234.
- Chavez-Demoulin, V., Embrechts, P., Sardy, S., 2011. Extreme-quantile tracking for financial time series. *Journal of Econometrics (to appear)*.
- Christoffersen, P., 1998. Evaluating interval forecasts. *International economic review*, 841–862.
- Coles, S., 2001. *An introduction to Statistical Modelling of Extreme Values*. Springer.
- Cotter, J., Dowd, K., 2006. Extreme spectral risk measures: An application to futures clearinghouse margin requirements. *Journal of Banking and Finance* 30, 3469–3485.
- Daley, D., Vere-Jones, D., 2003. *An Introduction to the Theory of Point Processes*. Springer.
- Danielsson, J., De Vries, C., 2000. Value-at-risk and extreme returns. *Annales d'Economie et de Statistique* 60, 239–270.
- Davis, R., Mikosch, T., 2009. Extreme value theory for GARCH processes. In: T. G. Andersen, R. A. Davis, J.-P. K., Mikosch, T. (Eds.), *Handbook of Financial Time Series*. Vol. 6. Springer, pp. 187–200.
- Ding, Z., Granger, C., Engle, R., 1993. A long memory property of stock market returns and a new model. *Journal of empirical finance* 1 (1), 83–106.
- Drost, F., Werker, B., 2004. Semiparametric duration models. *Journal of Business and Economic Statistics* 22 (1), 40–50.
- Dufour, A., Engle, R., 2000. The ACD model: predictability of the time between consecutive trades. Tech. rep., University of Reading, ISMA Centre.

- Embrechts, P., 2009. Linear Correlation and EVT: Properties and Caveats. *Journal of Financial Econometrics* 7, 30–39.
- Embrechts, P., Klüppelberg, C., Mikosch, T., 1997. *Modelling Extremal Events*. Springer.
- Engle, R., 2000. The econometrics of ultra-high-frequency data. *Econometrica* 68 (1), 1–22.
- Engle, R., Manganelli, S., 2004. Caviar. *Journal of Business and Economic Statistics* 22 (4), 367–381.
- Engle, R., Russell, J., 1998. Autoregressive conditional duration: A new model for irregularly spaced transaction data. *Econometrica* 66, 1127–1162.
- Falk, M., Hüsler, J., R., R., 2004. *Laws of Small Numbers: Extremes and Rare Events*. Birkhäuser, Basel.
- Fernandes, M., Grammig, J., 2006. A family of autoregressive conditional duration models. *Journal of Econometrics* 130 (1), 1–23.
- Gencay, R., Selcuk, F., Ulugulyagci, A., 2003. High volatility, thick tails and extreme value theory in value-at-risk estimation. *Insurance: Mathematics and Economics* 33 (2), 337–356.
- Ghysels, E., Gouriéroux, C., Jasiak, J., 2004. Stochastic volatility duration models. *Journal of Econometrics* 119 (2), 413–433.
- Ghysels, E., Jasiak, J., 1998. Garch for irregularly spaced financial data: the acd-garch model. *Studies in Nonlinear Dynamics & Econometrics* 2 (4), 133–149.
- Grammig, J., Maurer, K., 2000. Non-monotonic hazard functions and the autoregressive conditional duration model. *Econometrics Journal* 3, 16–38.
- Hautsch, N., 2004. *Modelling irregularly spaced financial data: theory and practice of dynamic duration models*. Springer.
- Lunde, A., 1999. A generalized gamma autoregressive conditional duration model. Tech. rep., Aalborg University.
- McNeil, A., Frey, R., 2000. Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. *Journal of Empirical Finance* 7, 271–300.
- McNeil, A. J., Frey, R., Embrechts, P., 2005. *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton University Press.
- Meitz, M., Teräsvirta, T., 2006. Evaluating models of autoregressive conditional duration. *Journal of Business & Economic Statistics* 24, 104 – 124.
- Mikosch, T., 2003. Modeling dependence and tails of financial time series. In: Finkenstaedt, B., Rootzen, H. (Eds.), *Extreme values in finance, telecommunications and the environment*. Chapman & Hall, pp. 185 – 286.
- Ogata, Y., 1988. Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical Association* 83, 9–27.
- Russell, J., Engle, R., 2005. A discrete-state continuous-time model of financial transactions prices and times. *Journal of Business and Economic Statistics* 23 (2), 166–180.
- Smith, R., 1989. Extreme value analysis of environmental time series: An application to trend detection in ground-level ozone. *Statistics Science* 4, 367–393.
- Smith, R. L., 2003. Statistics of extremes, with applications in environment, insurance and finance. In : Finkenstädt and Rootzén, 1–78.
- Zhang, M., Russell, J., Tsay, R., 2001. A nonlinear conditional autoregressive duration model with applications to financial transactions data. *Journal of Econometrics* 104, 179–207.

Appendix A. Threshold choice for EVT

The problem of finding an optimal threshold is very subjective due to fact that we need to find a sufficiently high threshold u above which the distribution of the excesses may be approximated by a GPD. The parameters of the GPD may be estimated by using, for example, maximum likelihood once the threshold u has been chosen. However, this choice is subject to a trade-off between variance and bias.

Observe that under the ACD-POT methodology only the tail index ξ remains constant, while the scale parameter varies throughout the time. From the point of view of the risk measures, a robust fit to a sample of extreme events and a good estimate of risk measures, as for example VaR, would be relatively insensitive to departures from the model. This is valuable in actual financial problems where one of the most important objectives is to obtain a robust measure of risk. However, EVT implementation faces many challenges, one of the most important being the fact that EVT is designed for independent data and financial data. Stock market returns in our case tend to be dependent, and therefore, a standard methodology for threshold selection does not exist.

The critical point in threshold selection is that by increasing the number of observations for the series of maxima, some observations from the centre of the distribution are introduced in the series, and that the tail index as well as the VaR estimate are more precise but biased (i.e., there is less variance). On the other hand, choosing a high threshold reduces the bias but makes the estimates more unstable (i.e., there are fewer observations).

Thus, the main objective in this section is to determine how sensitive the ACD-POT framework is to the choice of the threshold u , and in particular the VaR estimates obtained through these models. To this end, we choose the optimal threshold indirectly, by choosing an interval where the threshold quantile seems to be more stable in relation to the VaR estimate. To compare the different intervals, we computed the mean squared error (MSE) pointwise of the estimators as follows:

- We fix in advance a grid of size 100 of possible threshold values $\mathbf{q} \in [0.85, \dots, 0.94]$, i.e., $k = 1, \dots, 100$ with $q_1 = 0.85$ to $q_{100} = 0.94$, $q_k < q_{k+1}$ for all k . Further, we select different levels for the VaR that will be estimate through an ACD-Model. In this paper we choose $\alpha_j = (0.95, 0.96, 0.97, 0.98, 0.99, 0.999)$ for $j = 1, \dots, 7$.
- We choose a quantile threshold q_k and estimate a suitable ACD-POT model. Since the estimate of the VaR are time varying we compute a mean value $\overline{VaR}(q_k, \alpha_j)$ for each VaR level α_j .

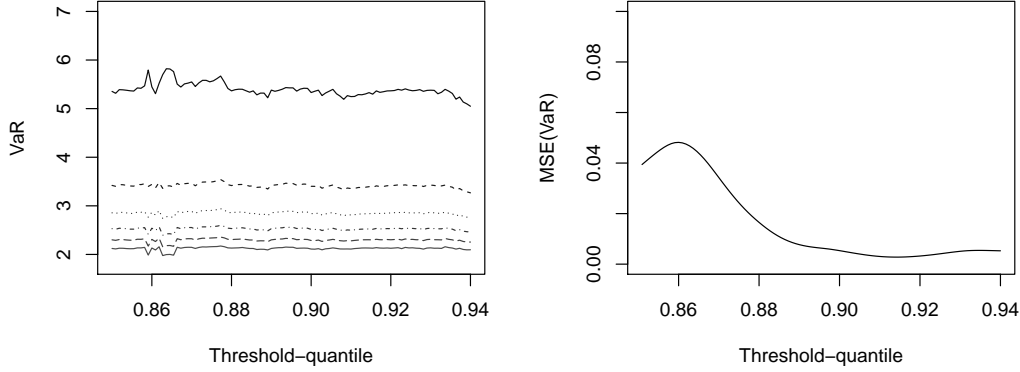


Figure A.5: On the left the mean value of the VaR, $\overline{VaR}(q_k, \alpha_j)$, for each VaR level α_j from bottom ($\alpha_1 = 0.95$) to top ($\alpha_7 = 0.999$) for the Bayer stock market returns. On the right side, estimated mean squared error $MSE(q_k - q_{k+1})$ for different thresholds.

- To compare the different estimates, we calculate the following MSE

$$MSE(q_k - q_{k+1}) = \frac{1}{m} \sum_{j=1}^m (\overline{VaR}(q_k, \alpha_j) - \overline{VaR}(q_{k+1}, \alpha_j))^2$$

- Finally, we choose the threshold by choosing an interval where the threshold quantile seems to be more stable.

For example, in Figure A.5 we display the analysis of threshold selection for the Bayer stock market returns in relation to gACD2 model. On the left, to compare the VaR sensitivity of the estimates with the threshold sensitivity, we plot the mean value of the VaR, $\overline{VaR}(q_k, \alpha_j)$, for each VaR level α_j from bottom ($\alpha_1 = 0.95$) to top ($\alpha_7 = 0.999$). For this example, and in general for all the returns that we have considered, the threshold selection seems to have limited influence on the VaR estimates. On the right side, we display the values of the $MSE(q_k - q_{k+1})$ respect to the thresholds intervals $q_k - q_{k+1}$. Graphical inspection of this figure shows that a threshold between the quantiles 0.90 and 0.92 may be the most justified. In general for most of the models estimated, a threshold selection between these two quantiles showed a great stability.

Appendix B. Goodness of fit

Appendix B.1. Goodness of fit ACD

Kolmogorov-Smirnov test (KS_{ACD}): This test quantifies a distance between the empirical distribution function of the standardized residuals of the sample and the true cumulative distribu-

tion function assumed.

Anderson-Darling test (AD): The Anderson-Darling test is the other alternative used to test if a semprirical distribution function of the standardized residuals came from true cumulative distribution function assumed.

Density forecasting (χ^2): we employ density forecasting techniques introduced by Bauwens et al. (2004). They assumed parametric densities for corresponding specification of the duration models and compared financial duration models based on the results with parametric densities. They incorporated the following procedure for comparison. Under the null of iid $U(0;1)$ behavior of z_i sequence, the joint distribution of the heights of the z_i histogram is multinomial, i.e.,

$$p(n_i) = \binom{n}{p} p^{n_i} (1-p)^{n-n_i},$$

where n is the sample size, n_i is the number of observations in the i -th bin, and $p = 1/m$ with m equal to the number of histogram bins. By means of this idea one can compute the Pearson's goodness of fit statistic (χ^2)

$$\chi^2 = \sum_{i=1}^m \frac{(n_i - np)^2}{np}$$

which is under the null hypothesis asymptotically distributed Chi-squared with $m - 1$ degrees of freedom.

Ljung-Box test (LB_{ACD}): In addition, to check that there is no further time series structure the Ljung-Box test is also included at lag 5.

Appendix B.2. Goodness of fit POT

W-statistics (W): we provide the W-statistics to assess our success in modelling the temporal behaviour of the exceedances of the threshold u . The W-statistic is defined by

$$W = \xi^{-1} \ln \left(1 + \xi \frac{y-u}{\beta(t,y | \mathcal{H}_t)} \right).$$

This statistic states that if the GPD parameter model is correct, then the residuals are approximately independent unit exponential variables. In practice, the independence assumption is checked via an Ljung-Box test at lag 5.

Kolmogorov-Smirnov test (KS_{ACD}): This test quantifies a distance between the empirical distribution function of the residuals of the POT model (the exceedances) and the exponential

distribution function. The residuals should be approximately independent unit exponential variables.

Appendix B.3. Accuracy of VaR

Test of Unconditional Coverage (LR_{uc}): Christoffersen (1998) terms the sequence of VaR forecasts efficient with respect to the history \mathcal{H}_{t-1} if $E[I_t | \mathcal{H}_{t-1}] = \alpha$, where $I_t = \mathbb{I}(r_t < -VaR_t)$ with \mathbb{I} being the indicator function. Due to the fact that $I_t | \mathcal{H}_{t-1} \sim Ber(\alpha)$, $t = 1, 2, \dots, T$. Applying iterated expectations, implies that I_t is uncorrelated (unconditional coverage) with any function of a variable in the information set available. This can be tested by means of a likelihood-ratio test

$$LR_{uc} = 2[\mathcal{L}(\hat{\alpha}; I_1, \dots, I_T) - \mathcal{L}(\alpha; I_1, \dots, I_T)] \sim \chi_1^2,$$

where \mathcal{L} is the log binomial likelihood. The maximum likelihood estimation $\hat{\alpha}$ is the ratio of number of violations, n_1 , to the total number of observations, $T = n_0 + n_1$.

Test of Independence (LR_{ind}): Christoffersen (1998) suggests a test of independence by modeling the number of violations I_t as a binary first order Markov chain with transition probability matrix

$$\Pi = \begin{bmatrix} 1 - \pi_{01} & \pi_{01} \\ 1 - \pi_{11} & \pi_{11} \end{bmatrix}, \quad \pi_{ij} = P(I_t = j | I_{t-1} = i),$$

as the alternative hypothesis of dependence. The joint likelihood, conditional on the first observation is given by

$$L(\pi^*; I_2, \dots, I_T | I_1) = (1 - \pi_{01})^{n_{00} + n_{10}} \pi_{01}^{n_{01} + n_{11}},$$

where n_{ij} represents the number of transitions from state i to state j . The maximum-likelihood estimators under the alternative hypothesis are

$$\hat{\pi}_{01} = \frac{n_{01}}{n_{00} + n_{01}} \quad \text{and} \quad \hat{\pi}_{11} = \frac{n_{11}}{n_{10} + n_{11}}.$$

Under the null hypothesis of independence, we have $\pi = \pi_{01} = \pi_{11}$, from which the conditional binomial joint likelihood is defined as

$$L(\pi; I_2, \dots, I_T | I_1) = (1 - \pi_{01})^{n_{00}} \pi_{01}^{n_{01}} (1 - \pi_{11})^{n_{10}} \pi_{11}^{n_{11}}.$$

Similar to the unconditional coverage test, the likelihood ratio test is given by

$$LR_{ind} = 2[\mathcal{L}(\hat{\pi}^*; I_2, \dots, I_T | I_1) - \mathcal{L}(\hat{\pi}; I_2, \dots, I_T | I_1)] \sim \chi_1^2.$$

Conditional Coverage (LR_{cc}): Christoffersen (1998) suggests combining the unconditional coverage test and the test of independence in order to test correct conditional coverage, because π^* is unconstrained. Then, we have

$$LR_{cc} = LR_{uc} + LR_{ind} \sim \chi_2^2.$$

We can jointly test for independence and correct coverage using the conditional coverage test.

Ljung-Box test (LB_{VaR}): we implement a test statistics proposed by Berkowitz et al. (2009) for the autocorrelations of de-meanded violations $Hit_t(\alpha) = I_t - \alpha$, which form a martingale difference sequence. This is a Ljung-Box statistic, which is a joint test of whether the first m autocorrelations of $Hit_t(\alpha)$ are zero by calculating

$$LB_{VaR}(m) = T(T+2) \sum_{k=1}^m \frac{\gamma_k^2}{T-k}$$

where T is the sample size, γ_k is the sample autocorrelation at lag k and $LB_{VaR}(m)$ is asymptotically chi-square with m degrees of freedom.

Dynamic quantile test (DQ_{hit} and DQ_{VaR}): A relevant VaR model should also feature a sequence of VaR violations which are not serially correlated. Engle and Manganelli (2004) suggest the Dynamic Quantile (DQ), which can jointly test the hypothesis that $E[Hit_t(\alpha)] = 0$ and that $Hit_t(\alpha)$ is uncorrelated with the variables included in the information set, where $Hit_t(\alpha) = I_t - \alpha$. Both tests can be done using the following artificial regression

$$Hit_t = \mathbf{X}\beta + \mathbf{u}, \quad \begin{cases} -\alpha, & \text{with probability } 1 - \alpha \\ 1 - \alpha, & \text{with probability } \alpha \end{cases},$$

where, under the null hypothesis, $H_0 = \beta = \mathbf{0}$, i.e, the regressors should have no explanatory power. Considering that the regressors are not correlated with the dependent variables under the null hypothesis, invoking a suitable central limit theorem Engle and Manganelli (2004) deduce the test statistic

$$DQ = \frac{\hat{\beta}\mathbf{X}'\mathbf{X}\hat{\beta}'}{\alpha(1-\alpha)} \sim \chi_{p+2}^2,$$

where p is the number of explanatory variables \mathbf{X} . In the empirical application, we use two specifications: the dynamic quantile hit (DQ_{hit}) test and the dynamic quantile VaR (DQ_{VaR}) test: In the first test the regressor matrix \mathbf{X} contains a constant and one lagged hits, while

the second test, DQ_{VaR} , uses, in addition, the contemporaneous VaR estimate.

Tables and Figures

Indexes	mean	sd	min	max	skewness
Bayer	0.027	1.826	-18.432	32.310	0.761
DAX	0.030	1.375	-9.871	7.553	-0.265
Portfolio	0.072	0.966	-6.558	6.110	-0.165
Indexes	kurtosis	Q(x)	Q(x ²)	Jarque-Bera	ADW
Bayer	26.12	5.41	76.78*	134489.50*	-17.59*
DAX	4.32	4.86	1237.71*	3724.86*	-16.08*
Portfolio	4.03	42.90*	222.48*	1629.01*	-12.91*

Table B.1: Summary statistics for the returns. Asymptotic p-value are shown in the brackets. *, **, *** denote statistical significance at the 1, 5 and 10 % level respectively. The Ljung-Box test statistic (Q) for serial correlation is calculated up to the 5-th order.

Models	Parameters										Loglike	AIC (mAIC)	
	ACD Model				POT model								
	w	a_1	b_1	δ	γ	k	ξ	ω	β_1	β_2	β_3		
<i>Bayer returns</i>													
bACD3	1.050 (0.4957)	0.220 (0.059)	0.733 (0.058)		0.951 (0.122)	1.443 (0.135)	0.085 (0.104)	0.761 (0.042)	3.8e-07 (0.01)	19.45 (18.72)	2.15 (0.68)	-2035.40	4090.79 (4056.79)
bLog-ACD3	0.276 (0.109)	0.157 (0.032)	0.777 (0.054)		0.951 (0.115)	1.455 (0.133)	0.093 (0.114)	0.762 (0.042)	2.7e-07 (0.01)	16.755 (16.38)	2.072 (0.712)	-2034.74	4089.49 (4055.49)
bEXACD4	0.139 (0.101)	0.162 (0.038)	0.887 (0.041)	1e-04 (0.009)	0.992 (0.124)	1.533 (0.148)	0.088 (0.089)	0.665 (0.112)	0.174 (0.040)	0.077 (0.028)	0.291 (0.416)	-2030.90	4083.80 (4047.8)
gACD3	0.738 (0.274)	0.165 (0.036)	0.760 (0.050)		0.192 (0.046)	21.219 (10.243)	0.086 (0.104)	0.728 (0.042)	5.5e-06 (0.008)	19.477 (15.189)	2.078 (0.547)	-2024.44	4068.88 (4032.88)
gLog-ACD3	0.208 (0.077)	0.138 (0.029)	0.804 (0.050)		0.107 (0.038)	68.860 (49.283)	0.089 (0.123)	0.765 (0.042)	3.8e-06 (0.008)	28.094 (33.925)	2.353 (0.867)	-2021.46	4062.93 (4026.93)
gLog-ACD4	0.187 (0.083)	0.122 (0.031)	0.826 (0.056)		0.117 (0.038)	60.007 (38.884)	0.091 (0.088)	0.663 (0.041)	0.170 (0.098)	0.075 (0.027)	0.279 (0.366)	-2018.10	4056.19 (4032.19)
<i>DAX returns</i>													
bACD4	1.273 (0.633)	0.189 (0.071)	0.789 (0.054)		1.237 (0.141)	1.938 (0.225)	0.025 (0.066)	0.583 (0.034)	0.104 (0.086)	0.122 (0.036)	1.572 (1.779)	-1880.416	3780.831 (3770.83)
bLog-ACD4	0.234 (0.093)	0.117 (0.029)	0.829 (0.046)		1.156 (0.127)	1.831 (0.200)	0.025 (0.066)	0.583 (0.034)	0.103 (0.086)	0.122 (0.036)	1.578 (1.790)	-1881.022	3782.044 (3770.04)
bEXACD4	0.147 (0.088)	0.173 (0.063)	0.900 (0.034)	4.2e-09 (0.005)	1.217 (0.143)	1.898 (0.222)	0.025 (0.066)	0.583 (0.086)	0.103 (0.034)	0.122 (0.036)	1.578 (1.783)	-1883.375	3788.749 (3752.75)
gACD3	0.793 (0.253)	0.195 (0.043)	0.716 (0.056)		0.132 (0.049)	47.998 (35.808)	0.021 (0.070)	0.580 (0.037)	0.071 (0.054)	32.091 (29.011)	2.706 (0.626)	-1880.545	3781.09 (3749.09)
gLog-ACD2	0.156 (0.052)	0.152 (0.030)	0.815 (0.042)		0.100 (0.019)	83.770 (32.351)	0.033 (0.039)	0.373 (0.076)	0.094 (0.056)	3.298 (0.574)		-1881.169	3780.337 (3752.34)
gLog-ACD3	0.200 (0.061)	0.167 (0.029)	0.783 (0.043)		0.121 (0.038)	57.682 (36.433)	0.021 (0.076)	0.569 (0.037)	0.077 (0.055)	22.461 (19.376)	2.433 (0.602)	-1877.519	3775.037 (3747.04)
<i>Portfolio returns</i>													
bACD2	0.709 (0.397)	0.213 (0.057)	0.741 (0.062)		0.682 (0.158)	1.223 (0.133)	0.093 (0.071)	0.176 (0.089)	0.053 (0.069)	3.167 (0.977)		-894.858	1807.717 (1779.72)
bACD4	0.859 (0.504)	0.183 (0.054)	0.747 (0.072)		0.651 (0.176)	1.224 (0.140)	0.087 (0.078)	0.248 (0.071)	0.075 (0.038)	0.026 (0.017)	7.3e-06 (0.027)	-893.091	1806.182 (1778.18)
bLog-ACD4	0.224 (0.106)	0.140 (0.035)	0.799 (0.058)		0.703 (0.154)	1.268 (0.138)	0.087 (0.077)	0.249 (0.071)	0.075 (0.038)	0.026 (0.017)	8.4e-05 (0.028)	-893.339	1806.679 (1770.68)
gACD2	0.799 (0.403)	0.197 (0.051)	0.739 (0.063)		0.146 (0.062)	37.661 (31.544)	0.093 (0.071)	0.184 (0.081)	0.058 (0.067)	2.964 (0.844)		-887.878	1793.756 (1759.76)
gACD4	0.962 (0.523)	0.170 (0.051)	0.744 (0.074)		0.168 (0.073)	29.118 (25.022)	0.088 (0.124)	0.251 (0.081)	0.073 (0.093)	0.026 (0.017)	0.025 (1.085)	-886.505	1793.01 (1757.01)
gLog-ACD4	0.249 (0.116)	0.132 (0.034)	0.791 (0.063)		0.174 (0.076)	27.432 (23.756)	0.087 (0.075)	0.249 (0.072)	0.075 (0.038)	0.026 (0.016)	3.7e-05 (0.024)	-886.966	1793.931 (1757.93)

Table B.2: Results of the estimation of all ACD-POT models with distributional assumption Burr for the standardized durations of the inter-exceedance times for 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	GoF ACD				Gof POT		Accuracy VaR							
	KS_{ACD}	AD	Chisq	LB_{POT}	W	KS_{POT}	α	failures	LRuc	LRind	LRcc	LB_{VaR}	DQ_{hit}	DQ_{VaR}
bACD3	0.07	0.06	0	0.68	0.43	0.11	0.05	238	0.84	0.15	0.35	0.13	0.15	0.3
							0.01	43	0.55	0.4	0.59	0.32	0.4	0.49
							0.001	9	0.08	0.86	0.21	0.91	0.86	0.01
bLog-ACD3	0.19	0.07	0.01	0.64	0.41	0.12	0.05	238	0.84	0.25	0.50	0.23	0.25	0.51
							0.01	39	0.23	0.33	0.30	0.22	0.33	0.56
							0.001	8	0.17	0.88	0.38	0.92	0.88	0.02
bEXACD4	0.05	0.06	0	0.50	0.15	0.11	0.05	264	0.06	0.06	0.03	0.05	0.06	0.04
							0.01	49	0.77	0.01	0.05	0	0.01	0.02
							0.001	10	0.03	0.84	0.10	0.90	0.84	0.26
gACD3	0.23	0.20	0	0.70	0.47	0.16	0.05	223	0.42	0.1	0.18	0.08	0.10	0.25
							0.01	38	0.17	0.31	0.23	0.20	0.31	0.5
							0.001	7	0.32	0.89	0.61	0.93	0.89	0.07
gLog-ACD3	0.54	0.24	0.09	0.73	0.42	0.12	0.05	230	0.74	0.09	0.22	0.07	0.09	0.15
							0.01	37	0.13	0.29	0.18	0.18	0.29	0.31
							0.001	6	0.56	0.91	0.84	0.94	0.91	0.23
gLog-ACD4	0.25	0.15	0.01	0.72	0.16	0.08	0.05	249	0.35	0.03	0.07	0.02	0.03	0
							0.01	42	0.46	0.06	0.12	0.01	0.06	0.01
							0.001	8	0.17	0.88	0.38	0.92	0.88	0.79
$CAViaR_{ad}$	-	-	-	-	-	-	0.05	236	0.97	0.51	0.8	0.5	0.52	0
							0.01	47	0.99	0.48	0.78	0.42	0.49	0
							0.001	4	0.74	0.94	0.94	0.97	0.94	0
$CAViaR_{sa}$	-	-	-	-	-	-	0.05	235	0.98	0.83	0.98	0.83	0.84	0
							0.01	48	0.89	0.51	0.79	0.45	0.51	0
							0.001	5	0.89	0	0.01	0	0	0
$CAViaR_{as}$	-	-	-	-	-	-	0.05	238	0.86	0.76	0.94	0.76	0.77	0
							0.01	47	0.99	NA	NA	0.5	0.34	0
							0.001	6	0.57	0.91	0.84	0.94	0.91	0
$CAViaR_{GARCH}$	-	-	-	-	-	-	0.05	248	0.41	0.57	0.6	0.56	0.58	0
							0.01	50	0.67	0.55	0.77	0.51	0.55	1
							0.001	9	0.08	0.86	0.21	0.91	0.86	0
CondN	-	-	-	0	0	-	0.05	224	0.44	0	0.01	0	0	0.01
							0.01	35	0.06	0.03	0.01	0	0.03	0.07
							0.001	8	0.17	0.88	0.38	0.92	0.88	0.73
CondT	-	-	-	0	0	-	0.05	231	0.77	0.01	0.02	0	0.01	0.03
							0.01	35	0.06	0.03	0.01	0	0.03	0.07
							0.001	9	0.08	0.86	0.21	0.91	0.86	0.46
CondST	-	-	-	0	0	-	0.05	238	0.86	0.01	0.04	0.01	0.01	0.05
							0.01	34	0.04	0.24	0.07	0.12	0.24	0.22
							0.001	8	0.17	0.88	0.38	0.92	0.88	0.01

Table B.3: Goodness of fit (GoF), to assess the predictive performance in-sample of the models under consideration for the Bayer returns, we divide these into: GoF for inter-exceedante times (GoF ACD), GoF for the marks or exceedances (GoF POT) and measures of accuracy for the VaR. Entries in the columns are the significance levels (p-values) of the respective tests, with exception of the level α and the number of violations at the VaR. The cells with values NA means that the test can not be estimated.

Models	GoF ACD				GoF POT		Accuracy VaR							
	KS_{ACD}	AD	Chisq	LB_{ACD}	W	KS_{POT}	α	failures	LRuc	LRind	LRcc	LB_{VaR}	DQ_{hit}	DQ_{VaR}
bACD4	0	0	0	0.72	0.91	0.77	0.05	335	0	0.01	0	0	0.01	0.02
							0.01	75	0	0	0	0	0	
							0.001	10	0.03	0.84	0.10	0.88	0.84	0.72
bLog-ACD4	0	0	0	0.71	0.92	0.76	0.05	327	0	0.01	0	0.01	0.01	0.03
							0.01	71	0	0	0	0	0	
							0.001	9	0.08	0.85	0.21	0.90	0.85	0.77
bEXACD4	0	0	0	0.78	0.92	0.76	0.76	334	0	0	0	0	0	0.01
							0.01	73	0	0	0	0	0	
							0.001	10	0.03	0.84	0.1	0.88	0.84	0.72
gACD3	0.32	0.15	0.00	0.44	0.81	0.77	0.05	244	0.54	0.22	0.38	0.20	0.22	0.44
							0.01	45	0.77	0.08	0.20	0.02	0.08	0.04
							0.001	7	0.32	0.89	0.61	0.92	0.89	0.34
gLog-ACD2	0.24	0.16	0.00	0.61	0.86	0.84	0.05	239	0.77	0.26	0.51	0.24	0.26	0.01
							0.01	41	0.37	0.05	0.10	0.01	0.05	0.13
							0.001	7	0.32	0.89	0.61	0.92	0.89	0.08
gLog-ACD3	0.22	0.12	0.00	0.49	0.71	0.58	0.05	246	0.45	0.24	0.38	0.22	0.24	0.09
							0.01	37	0.13	0.03	0.03	0.00	0.03	0.08
							0.001	5	0.89	0.92	0.99	0.94	0.92	0.34
$CAViaR_{ad}$	-	-	-	-	-	-	0.05	236	0.97	0.35	0.65	0.33	0.36	0
							0.01	47	0.99	NA	NA	0.49	0.33	0
							0.001	6	0.57	0.90	0.84	0.93	0.90	0
$CAViaR_{sa}$	-	-	-	-	-	-	0.05	235	0.98	0.59	0.86	0.59	0.60	0
							0.01	48	0.89	0.52	0.80	0.46	0.52	0
							0.001	6	0.57	0.90	0.84	0.93	0.90	0
$CAViaR_{as}$	-	-	-	-	-	-	0.05	237	0.91	0.75	0.94	0.74	0.75	0
							0.01	48	0.89	NA	NA	0.48	0.32	0
							0.001	4	0.74	0.93	0.94	0.95	0.93	0
$CAViaR_{GARCH}$	-	-	-	-	-	-	0.05	238	0.86	0.09	0.23	0.07	0.10	0
							0.01	48	0.89	0.52	0.80	0.46	0.52	0
							0.001	5	0.89	0.92	0.99	0.94	0.92	0
CondN	-	-	-	0	0	-	0.05	219	0.14	0.18	0.11	0.15	0.05	0.01
							0.01	40	0	0.01	0	0	0.01	0.07
							0.001	3	0.95	0.70	0.97	0.95	0.26	0.73
CondT	-	-	-	0	0	-	0.05	220	0.30	0.04	0.07	0.03	0.05	0.14
							0.01	38	0.17	0.04	0.05	0	0.04	0.01
							0.001	7	0.32	0.89	0.61	0.92	0.89	0
CondST	-	-	-	0	0	-	0.05	225	0.48	0.11	0.23	0.09	0.12	0.31
							0.01	42	0.45	0.01	0.02	0	0.01	0
							0.001	7	0.32	0.89	0.61	0.92	0.89	0.57

Table B.4: Goodness of fit (GoF), to assess the predictive performance in-sample of the models under consideration for the DAX returns, we divide these into: GoF for inter-exceedante times (GoF ACD), GoF for the marks or exceedances (GoF POT) and measures of accuracy for the VaR. Entries in the columns are the significance levels (p-values) of the respective tests, with exception of the level α and the number of violations at the VaR. The cells with values NA means that the test can not be estimated.

Models	GoF ACD				GoF POT		Accuracy VaR							
	KS_{ACD}	AD	Chisq	LB_{POT}	W	KS_{POT}	α	failures	LRuc	LRind	LRcc	LB_{VaR}	DQ_{hit}	DQ_{VaR}
bACD2	0	0	0	0.32	0.82	0.61	0.05	119	0.95	0.02	0.07	0.01	0.02	0.02
							0.01	26	0.63	0.45	0.67	0.59	0.45	0.50
							0.001	5	0.14	0.88	0.33	0.92	0.88	0.98
bACD4	0	0	0	0.38	0.67	0.97	0.05	114	0.68	0.03	0.08	0.01	0.03	0.03
							0.01	23	0.89	0.50	0.79	0.63	0.50	0.49
							0.001	4	0.33	0.91	0.62	0.93	0.91	0.75
bLog-ACD4	0	0	0	0.27	0.67	0.96	0.76	113	0.61	0.13	0.28	0.10	0.13	0.19
							0.01	23	0.89	0.50	0.79	0.63	0.50	0.49
							0.001	4	0.33	0.91	0.62	0.93	0.91	0.77
gACD2	0.35	0.37	0.06	0.20	0.74	0.89	0.05	113	0.61	0.13	0.28	0.10	0.13	0.05
							0.01	30	0.21	0.38	0.31	0.53	0.38	0.62
							0.001	3	0.69	0.93	0.92	0.95	0.93	0.94
gACD4	0.40	0.29	0.05	0.21	0.68	0.97	0.05	114	0.68	0.14	0.31	0.11	0.14	0.11
							0.01	22	0.73	0.52	0.77	0.65	0.52	0.55
							0.001	4	0.33	0.91	0.62	0.93	0.91	0.72
gLog-ACD4	0.50	0.38	0.17	0.17	0.68	0.95	0.05	110	0.43	0.10	0.18	0.07	0.10	0.09
							0.01	22	0.73	0.52	0.77	0.65	0.52	0.60
							0.001	4	0.33	0.91	0.62	0.93	0.91	0.75
$CAViaR_{ad}$	-	-	-	-	-	-	0.05	120	0.92	0.24	0.49	0.21	0.25	0
							0.01	24	0.96	0.48	0.78	0.62	0.49	0
							0.001	3	0.70	0.93	0.92	0.95	0.93	0
$CAViaR_{sa}$	-	-	-	-	-	-	0.05	119	0.99	0.99	1.00	0.99	0.98	0
							0.01	23	0.87	0.50	0.79	0.63	0.50	0
							0.001	2	0.80	0.95	0.97	0.97	0.95	0
$CAViaR_{as}$	-	-	-	-	-	-	0.05	119	0.99	0.40	0.70	0.38	0.41	0
							0.01	24	0.96	0.48	0.78	0.62	0.49	0
							0.001	2	0.80	0.95	0.97	0.97	0.95	0
$CAViaR_{GARCH}$	-	-	-	-	-	-	0.05	120	0.92	0.43	0.72	0.41	0.44	0
							0.01	24	0.96	0.24	0.50	0.12	0.24	0
							0.001	3	0.70	0.93	0.92	0.95	0.93	0
CondN	-	-	-	0.31	0.30	-	0.05	106	0.22	0	0.01	0	0	0.01
							0.01	20	0.42	0.56	0.61	0.68	0.56	0.58
							0.001	3	0.70	0.93	0.92	0.95	0.93	0.84
CondT	-	-	-	0.31	0.42	-	0.05	107	0.26	0	0	0	0	0
							0.01	18	0.21	0.60	0.40	0.71	0.60	0.65
							0.001	3	0.70	0.93	0.92	0.95	0.93	0.74
CondST	-	-	-	0.31	0.28	-	0.05	110	0.4	0.01	0.01	0	0.01	0.01
							0.01	19	0.31	0.58	0.51	0.69	0.58	0.78
							0.001	3	0.70	0.93	0.92	0.95	0.93	0.76

Table B.5: Goodness of fit (GoF), to assess the predictive performance in-sample of the models under consideration for the Portfolio returns, we divide these into: GoF for inter-exceedante times (GoF ACD), GoF for the marks or exceedances (GoF POT) and measures of accuracy for the VaR. Entries in the columns are the significance levels (p-values) of the respective tests, with exception of the level α and the number of violations at the VaR. The cells with values NA means that the test can not be estimated.

α	Accuracy VaR																										
	Boyer returns									DAX returns									Portfolio returns								
	Models	failures	LRuc	LRind	LRcc	LB _{VaR}	DQ _{int}	DQ _{VaR}	DQ _{VaR}	Models	failures	LRuc	LRind	LRcc	LB _{VaR}	DQ _{int}	DQ _{VaR}	DQ _{VaR}	Models	failures	LRuc	LRind	LRcc	LB _{VaR}	DQ _{int}	DQ _{VaR}	
0.05	gLog-ACD3	15	0.87	0.75	0.94	0.75	0.81	0.52	gACD2	22	0.06	0.8	0.16	0.79	0.8	0.31	gACD4	14	0.47	0.22	0.36	0.16	0.26	0.26	0.3		
0.01		5	0.26	0.05	0.07	0	0.07	0.03		6	0.11	0.61	0.24	0.72	0.62	0		1	0.33	0.93	0.62	0.95	0.93	0.99			
0.001		3	0	0.01	0	0	0.02	0		1	0.3	0.93	0.58	0.95	0.93	0.49		0	0.5	1	0.79	-	-	-			
0.05	gLog-ACD4	16	0.67	0.24	0.46	0.19	0.29	0.25	gLog-ACD3	21	0.09	0.7	0.23	0.68	0.7	0.39	gLog-ACD4	13	0.67	0.16	0.34	0.1	0.2	0.22			
0.01		5	0.26	0.05	0.07	0	0.07	0.03		6	0.11	0.61	0.24	0.72	0.62	0.37		2	0.83	0.85	0.96	0.89	0.85	0.6			
0.001		3	0	0.01	0	0	0.02	0		1	0.3	0.93	0.58	0.95	0.93	0.98		0	0.5	1	0.79	-	-	-			
0.76	CondT	33	0	0	0	0.03	0	0	CondT	23	0.03	0.06	0.02	0.18	0.06	0.14	CondT	22	0.01	0.01	0	0	0.01	0.02			
0.01		16	0	0.18	0	0.35	0.17	0		2	0.58	0.91	0.85	0.95	0.87	0.81		4	0.31	0.71	0.56	0.79	0.71	0.91			
0.001		2	0.04	0.91	0.11	0.95	0.87	0.04		1	0.3	1	0.59	1	0.93	0.59	CondT	4	0.5	1	0.79	-	-	-			
0.05	CondST	30	0	0.14	0	0.2	0.14	0	CondT	22	0.06	0.06	0.03	0.18	0.06	0.17	CondT	18	0.07	0.04	0.03	0.02	0.05	0.13			
0.01		13	0	0.29	0	0.46	0.27	0		2	0.58	0.91	0.85	0.95	0.87	0.71	CondT	1	0.33	0.93	0.61	0.95	0.93	0.88			
0.001		2	0.04	0.87	0.11	0.9	0.87	0		1	0.3	1	0.59	1	0.93	0.7		0	0.5	1	0.79	0.79	0.79	0.88			
0.05	CAV _{VaR} _{ind}	23	0.03	0.51	0.08	0.54	0.48	0.49	CAV _{VaR} _{ind}	36	0	0.04	0	0.08	0.04	0	CAV _{VaR} _{ind}	28	0	0.35	0	0.32	0.35	0			
0.01		5	0.26	0.71	0.49	0.81	0.68	0.89		10	0	0.42	0	0.58	0.4	0		7	0.01	0.51	0.04	0.63	0.51	0			
0.001		1	0.3	1	0.59	1	0.93	0.51	CAV _{VaR} _{ind}	2	0.04	0.91	0.11	0.95	0.87	0	CAV _{VaR} _{ind}	0	0.5	1	0.79	0.79	0.79	0.88			
0.05	CAV _{VaR} _{CAKCH}	25	0.01	0.37	0.02	0.42	0.35	0.58	CAV _{VaR} _{CAKCH}	33	0	0.69	0	0.7	0.65	0	CAV _{VaR} _{CAKCH}	29	0	0.06	0	0.04	0.07	0			
0.01		4	0.54	0.77	0.79	0.86	0.74	0.61		13	0	0.56	0	0.53	0.6	0		4	0.31	0.71	0.56	0.79	0.71	0			
0.001		1	0.3	1	0.59	1	0.93	0.66		2	0.04	0.91	0.11	0.95	0.87	0		0	0.5	1	0.79	0.79	0.71	0			

Figure B.6: VaR prediction performance in the backtest for the three returns analysed. We consider two best models in each category for each return series.

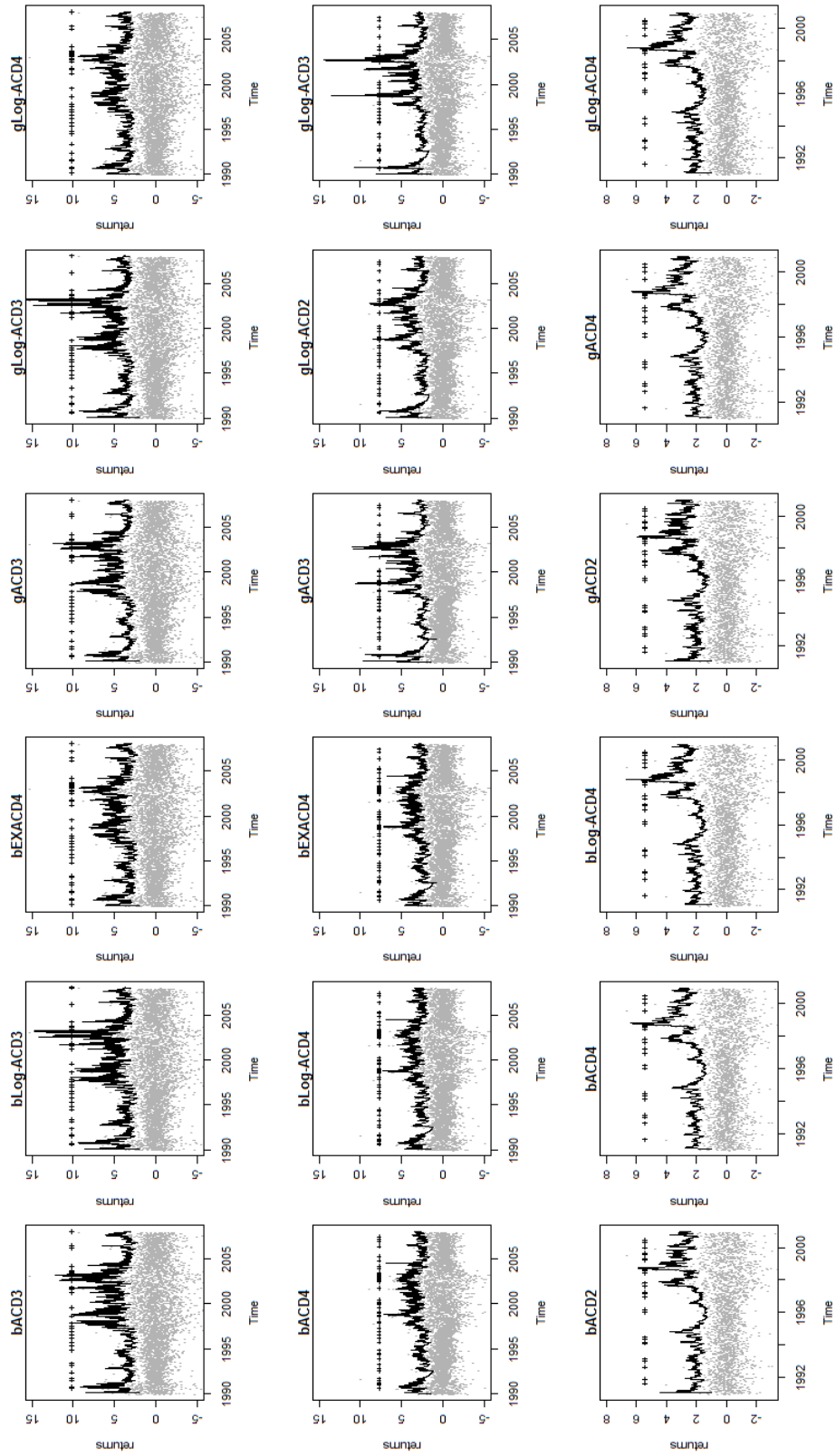


Figure B.7: In-sample estimation of the 99% conditional VaR for the Bayer returns (top panel), the DAX index returns (middle panel) and a hypothetical portfolio (bottom panel). The black line is the VaR estimation. The x marks at the top of the figures indicate violations of the VaR at the 0.01 confidence level.