

Intensity-based estimation of extreme loss event probability and value at risk

Kamal Hamidieh^{a*†}, Stilian Stoev^b and George Michailidis^b

We develop a methodology for the estimation of extreme loss event probability and the value at risk, which takes into account both the magnitudes and the intensity of the extreme losses. Specifically, the extreme loss magnitudes are modeled with a generalized Pareto distribution, whereas their intensity is captured by an autoregressive conditional duration model, a type of self-exciting point process. This allows for an explicit interaction between the magnitude of the past losses and the intensity of future extreme losses. The intensity is further used in the estimation of extreme loss event probability. The method is illustrated and backtested on 10 assets and compared with the established and baseline methods. The results show that our method outperforms the baseline methods, competes with an established method, and provides additional insight and interpretation into the prediction of extreme loss event probability. Copyright © 2012 John Wiley & Sons, Ltd.

Keywords: heavy tails; clustering of extremes; value at risk; self-exciting point processes; autoregressive conditional duration; generalized Pareto distribution

1. Introduction

The introduction and proliferation of sophisticated and complex financial products has made managing financial risk both an important and exceedingly challenging task for all market participants. For this issue to be addressed, a number of risk measures have been proposed in the literature over the years, including the standard deviation, value at risk (VaR), and expected shortfall [1]. From a statistical viewpoint, VaR at a level q , $q \in (0, 1)$, is the q th quantile of the loss distribution of the portfolio over a fixed time horizon, where typically q is 0.95 or 0.99 (see, e.g., [2]). VaR has been used in assessing financial risk, *credit risk* (see, e.g., [3]), *operational risk* (see, e.g., [4]), and insurance (see, e.g., [5]). For a comprehensive discussion on the strengths and weaknesses of VaR, see [6, 7]; for its drawbacks from a theoretical standpoint, see [5, 8]; and for quantitative aspects, see [9–11].

The standard definition of VaR incorporates only the marginal distributions of the data, which makes it impervious to the temporal dynamics of the markets. For temporal dependence to be accounted, GARCH filters are generally applied to historical data, and a type of conditional VaR is calculated in practice (see, e.g., [12]). Looking at conditional VaR is important because ignoring the dependence between past losses and the times of the arrival of future large losses can have significant consequences. Indeed, a cluster of extreme losses arriving in a short period can lead to the rapid decline of the value of a portfolio, potentially triggering margin calls and in severe cases leading to ruin and bankruptcy. This *clustering phenomenon* is not well accounted for by models that involve only the marginal distribution of the losses.

Another natural measure of risk is the *extreme loss event probability* (ELEP). That is, given the history of the asset and a threshold u , ELEP is the conditional probability that the asset will incur a loss higher than u in the next period. This indicator complements VaR by providing in some sense the inverse information, that is, the *probability* of encountering a loss larger than a pre-specified magnitude, rather than the *loss magnitude* for a pre-specified probability. In practice, it is important to be able to track both ELEP and VaR dynamically, as the market conditions change.

The main objective of this paper is to develop a model for the temporal dynamics of extreme losses in order to improve upon the estimation of financial risk. As particular applications, we propose a methodology for the estimation of ELEP and a novel methodology for the estimation of *conditional VaR*, given the past behavior of the extreme losses.

^aDepartment of Statistics, Rice University, Houston, TX, U.S.A.

^bDepartment of Statistics, The University of Michigan, Ann Arbor, MI, U.S.A.

*Correspondence to: Kamal Hamidieh, Department of Statistics, Rice University, Houston, TX, U.S.A.

†E-mail: gmichail@umich.edu

Fundamental contributions in extreme value theory (EVT) have established that the times of extreme events (exceeding a threshold u) converge (as u grows and time is rescaled) to a *cluster Poisson point process* [13]. In reality, it is difficult to assess whether a given threshold u is extreme enough for the asymptotic theory of *cluster Poisson* behavior to apply. Local non-stationarity, ubiquitous in applications, reinforces the need to explore alternative approaches to modeling the temporal dynamics of extremes directly. Thus, for a given threshold u , we propose to model the times of exceedance, along with the excesses by using a *self-exciting point process* (SEPP). The SEPP models were introduced in the seminal works of Hawkes [14–16]. They are in fact marked point process (MPP) where the intensity function is causally dependent on the SEPP itself (for more details, see [17–19]). As discussed in [19], SEPPs come ‘closest to fulfilling, for point processes, the kind of role that the autoregressive model plays for conventional time series.’

In our case, the times between consecutive losses (and hence the intensity of the point process) are described by using a type of log-autoregressive conditional duration (ACD) model. The ACD models, originally introduced by Engle and Russell [20], are in fact SEPP models. Here, we represent the magnitudes of the losses (i.e., the marks) by the generalized Pareto distribution (GPD). The resulting model allows *interactions* between the past loss magnitudes and the arrival intensity of future losses. This interaction mechanism offers an interpretable description of the observed clustering phenomenon, which can also be statistically tested in practice. For related works, see also [21–23].

The performance of the developed methodology is assessed via extensive backtesting. We examine 10 time series of asset returns and demonstrate that our estimator of conditional VaR clearly outperforms the baseline techniques, which ignore either the temporal dependence or the extreme behavior of the losses. Our methods compete with the widely used GARCH-based methods of McNeil and Frey [12]. At the same time, the underlying conditional intensity of extreme losses arrival process provides a new perspective to the temporal dynamics of extremes, not available from the knowledge of VaR alone.

The remainder of the paper is organized as follows. In Section 2, we introduce the mathematical framework based on point process modeling of the extreme losses and their intensity. In Section 3, we derive and interpret the intensity function. In Section 4, we obtain an expression for the ELEP. In Section 5, we apply our intensity approach to estimating 1-day VaR. Backtesting results are presented in Section 6. Practical aspects on the dynamics of extremes and the application of our methodology are discussed in Section 7. We conclude with some remarks in Section 8.

2. Formal definitions and setup

Consider the negative daily log returns $\{Z_t\}_{t \geq 0}$, $t \in \mathbb{Z}^+$, of an asset. We will assume that Z_t 's have an absolutely continuous distribution. We use the convention of working with the negative log returns so that the losses correspond to the right tail of the distribution rather than the left tail. From now on, we will refer to the negative log returns simply as the returns.

Fix a high *threshold* u and refer to all returns that exceed u as *extreme losses*. Denote by T_i , $i \in \mathbb{Z}^+$, the time of the *exceedance* corresponding to the occurrence of extreme loss event $\{Z_{T_i} > u\}$. Let $Y_i \equiv Z_{T_i} - u$ denote the *excess loss*, that is, the magnitude of the i th extreme loss over u . Thus, we obtain an MPP $\{(T_i, Y_i)\}_{i \geq 1}$ associated with the extreme losses of $\{Z_t\}_{t \geq 0}$ over u . The Y_i 's are also referred to as *marks*. Let $X_{i+1} = T_{i+1} - T_i$, $i \geq 1$ be the *interexceedance* times.

To define the history of the MPP denoted by \mathcal{H}_t , let $\mathcal{F}_n \equiv \sigma\{X_j, Y_j, j \leq n\}$ and let the counting process $N(t) \equiv \max\{n \geq 0 : T_n \leq t\}$ be such that for all $t > 0$, we have $T_{N(t)} \leq t < T_{N(t)+1}$. Therefore, $N(t)$ will be a stopping time. We formally define $\mathcal{H}_t = \mathcal{F}_{N(t)}$. Intuitively, \mathcal{H}_t can be interpreted as the extreme loss history of the returns.

As indicated in the introduction, we focus on a class of MPP called SEPP (see, e.g., [19]). An important function associated with the SEPP is the *conditional intensity function*. It characterizes the statistical structure of its arrival times and is defined as follows:

$$\lambda(t|\mathcal{H}_t) \equiv \lim_{h \rightarrow 0} \frac{\mathbb{P}(\{N(t, t+h) > 0\}|\mathcal{H}_t)}{h}, \tag{1}$$

with $N(t, t+1) \equiv N(t) - N(t+1)$ (see, e.g., [24]).

Observe that if $\lambda(t|\mathcal{H}_t) \equiv \text{constant}$, $t \in [0, \infty)$ instead of $t \in \mathbb{Z}^+$ and $T_0 \equiv 0$, then $\{T_i\}_{i \geq 1}$ becomes a simple Poisson point process on $[0, \infty)$ with a constant intensity. In general, depending on the specification of the stochastic model for the SEPP, the conditional intensity $\lambda(t|\mathcal{H}_t)$ may depend on multiple T_j 's and their marks Y_j 's, $j \leq t$.

In this setup, we define the *conditional VaR* of level q , $q \in (0, 1)$, at time t , as the conditional quantile $z_q(t)$ such that

$$\mathbb{P}(\{Z_{t+1} > z_q(t)\}|\mathcal{H}_t) = 1 - q. \tag{2}$$

Note that the information in \mathcal{H}_t is based on the MPP. Thus, the conditional VaR can be interpreted as the level of loss that will be exceeded only with probability $1 - q$, given the history of the extreme losses of the returns.

Next, we define the conditional ELEP, $p_u(t)$, at time t as follows:

$$p_u(t) = \mathbb{P}(\{Z_{t+1} > u\} | \mathcal{H}_t). \quad (3)$$

The conditional ELEP can be interpreted as the probability of an extreme loss in the next day, given the history of the extreme losses of the returns.

The key quantities defined in Equations (1)–(3) are all *conditional* on the information from the history of the MPP associated with the past losses. However, from now on, for the sake of brevity, we will drop the term conditional.

The main goals in this paper are twofold: (i) develop an interpretable and adequate SEPP model through specifying the intensity in (1) and (ii) the estimation of the resulting model from data and its application to the calculation and backtesting of VaR and ELEP in Equations (2) and (3), respectively. Our modeling framework allows us to statistically test and quantify the impact of the past extreme losses in increasing or decreasing the intensity of losses, VaR, and ELEP.

3. Intensity modeling and estimation

In this section, we develop a model for the SEPP by specifying a parametric form for the conditional intensity $\lambda(t | \mathcal{H}_t)$ in Equation (1). We adopt the so-called log-ACD model introduced in [25]. The ACD models, pioneered by Engle and Russell [20], are akin to the GARCH models; for more details, see, for example, [26, 27] and Bauwens and Hautsch in [28] and the references therein.

Recall that $X_{n+1} = T_{n+1} - T_n$ denotes the n th interexceedance time for the MPP $\{(T_i, Y_i)\}_{i \geq 1}$. We set

$$X_{n+1} = \exp(\Psi_{n+1})\epsilon_{n+1}, \quad (4)$$

and

$$\Psi_{n+1} = \omega + \alpha\epsilon_n + \beta\Psi_n + \eta Y_n, \quad (5)$$

where ϵ_i 's are independent and identically distributed (i.i.d.) exponential random variables with unit mean and independent from Y_i 's. The last excess Y_n is involved in the model, which in turn affects the future interexceedance times X_j , $j \geq n + 1$.

Remark 1

The aforementioned model does not imply that the interexceedance times are exponential; rather, they are *conditionally* exponential. Our extensive experiments with different error distributions such as the Weibull and the generalized Gamma and the inclusion of additional past excess values did not lead to better results.

The following result, the proof of which can be found in [29, p. 72] and it is based on the results in [30], gives the conditions under which the log-ACD model is strictly stationary.

Theorem 1

Consider the log-ACD model in (4) and (5), where $\{\epsilon_i\}_{i \geq 0}$ is an i.i.d. standard exponential sequence and $\{Y_i\}_{i \geq 0}$ is an i.i.d. almost surely positive sequence.

Equations (4) and (5) have a strictly stationary solution $\{X_i\}_{i \geq 1}$ provided

$$0 \leq \beta < 1, \quad \text{and} \quad \mathbb{E}[Y_i] < \infty.$$

Theoretical results on the existence of the moments of a log-ACD model can be found in [31]. Recently, pseudo-maximum likelihood estimation (MLE) properties of the log-ACD models have been studied in [32].

The intensity $\lambda(t | \mathcal{H}_t)$, in this setting, can be interpreted as the hazard function corresponding to $X_{N(t)+1}$ (see [19, p. 231] and [28, p. 960]). Using Equations (4) and (5), we get

$$\lambda(t | \mathcal{H}_t) = \exp \left\{ -(\omega + \alpha\epsilon_{N(t)} + \beta\Psi_{N(t)} + \eta Y_{N(t)}) \right\}. \quad (6)$$

Furthermore, the memoryless property of the exponential distribution implies

$$\mathbb{P}(\{N(t, t+h) = 0 | \mathcal{H}_t\}) = \exp \{-h\lambda(t | \mathcal{H}_t)\},$$

for all $h > 0$. Thus, for the day-to-day time differences ($h = 1$), we obtain the following:

$$\mathbb{P}(\{N(t, t+1) = 0 | \mathcal{H}_t\}) = \exp \{-\lambda(t | \mathcal{H}_t)\}. \quad (7)$$

The aforementioned equation will be used later in the estimation of ELEP.

An advantage of the log-ACD versus other ACD-type models is that the intensity is positive regardless of the signs of the coefficients ω , α , β , and η in Equation (5). These coefficients can be estimated via MLE, given the *initial intensity* $\lambda_0 := \exp\{-\psi_0\}$. The standard errors for the estimates can be computed numerically via the differentiated Hessian matrices. Let $f_{X_1}(x_1)$ and $f_{X_i | X_{[1:i-1]}, Y_{[1:i-1]}}(x_i | x_{[1:i-1]}, y_{[1:i-1]})$, where $x_{[1:i-1]} = (x_1, \dots, x_{i-1})$ and $y_{[1:i-1]} = (y_1, \dots, y_{i-1})$ represent the marginal and conditional densities of X_1 and X_i given past values of X_{i-1} and Y_{i-1} , $i = 2, \dots, n$, respectively. By using the conditional independence of the X_i 's given the past and Equations (4) and (5), we obtain the log-likelihood for the log-ACD model:

$$\log \left(f_{X_1}(x_1) \prod_{i=2}^n f_{X_i | X_{[1:i-1]}, Y_{[1:i-1]}}(x_i | x_{[1:i-1]}, y_{[1:i-1]}) \right) = - \sum_{i=1}^n \left(\frac{x_i}{e^{\psi_i}} + \psi_i \right), \quad (8)$$

where

$$\psi_i = \omega + \alpha \frac{x_{i-1}}{\exp(\psi_{i-1})} + \beta \psi_{i-1} + \eta y_{i-1}, \quad (i = 1, \dots, n).$$

The value $\Psi_0 = \psi_0$ must be specified in advance. Observe that the initial intensity $\lambda_0 = \exp\{-\psi_0\}$ cannot be estimated because of the lack of data prior to time $t = 1$. If no prior information about the intensity λ_0 is available, a reasonable initial value for $\psi_0 = -\log \lambda_0$ is $-\log \mathbb{P}(\{Z_1 > u\})$, which is the intensity of arrivals for independent Z_t 's.

Consider the MLE parameter estimates $\hat{\omega}$, $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\eta}$ obtained by maximizing Equation (8). These values can be used as plug-ins to derive the following *residuals* of the model:

$$\hat{\epsilon}_i := \frac{X_i}{\exp(\hat{\Psi}_i)}, \quad \text{where} \quad \hat{\Psi}_i = \hat{\omega} + \hat{\alpha} \frac{X_{i-1}}{\exp(\hat{\Psi}_{i-1})} + \hat{\beta} \hat{\Psi}_{i-1} + \hat{\eta} Y_{i-1}, \quad (9)$$

with $i = 1, \dots, n$. If the model is adequate and the parameters estimate the true values well, then the residuals $\hat{\epsilon}_i$'s should be approximately independent standard exponential random variables. This fact can be used for goodness-of-fit diagnostics of the model.

Figure 1 shows the autocorrelation function and quintile–quintile (Q–Q) plot of the residuals in Equation (9) for interexceedance times of the West Texas Intermediate (WTI) oil spot price returns from November 15, 2005 to November 4, 2008 when $u = 0.0265$ —the choice of u is discussed in Section 5—for a total of 1000 trading days. There were 100 excess values above u and 99 interexceedance times. The choice of WTI is arbitrary and for demonstration purpose. These diagnostic plots confirm that the model captures well the temporal dynamics of the extremes in the given data.

The MLE results of Equation (8), namely $\hat{\omega}$, $\hat{\alpha}$, $\hat{\beta}$, $\hat{\eta}$, and the residual from Equation (9) can be used to obtain a plug-in estimate of the intensity in Equation (6)

$$\hat{\lambda}(t | \mathcal{H}_t) = \exp \left\{ - \left(\hat{\omega} + \hat{\alpha} \hat{\epsilon}_{N(t)} + \hat{\beta} \hat{\Psi}_{N(t)} + \hat{\eta} Y_{N(t)} \right) \right\}. \quad (10)$$

This estimate of the intensity function will be involved in the estimation of ELEP and VaR.

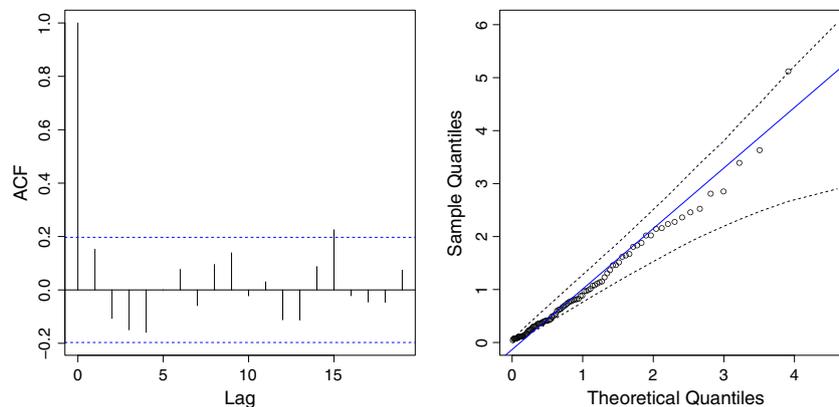


Figure 1. Autocorrelation function (ACF) plot of the log-autoregressive conditional duration model residuals of the oil data from November 15, 2005 to November 4, 2008 (left panel). Quantile–quantile plot of the residuals versus a unit rate exponential variable (right panel) along with pointwise 95% confidence intervals (dashed lines).

4. Estimation of extreme loss event probability

In this section, we focus on the estimation of ELEP defined in Equation (3). Additionally, we elaborate on the relationship between ELEP and the intensity function.

First, note that we have the following equality between events

$$\{N(t, t + 1) = 1\} | \mathcal{H}_t = \{Z_{t+1} > u\} | \mathcal{H}_t. \quad (11)$$

This, along with Equation (7), implies that

$$p_u(t) = 1 - \mathbb{P}(\{N(t, t + 1) = 0\} | \mathcal{H}_t) = 1 - \exp\{-\lambda(t | \mathcal{H}_t)\}. \quad (12)$$

An estimate of ELEP can be obtained by plugging in the estimated intensity from (10) into (12).

Note also that ELEP can be approximated via the intensity value because for large u , $p_u(t)$ and $\lambda(t | \mathcal{H}_t)$ are small and, by (12), we have $p_u(t) = \lambda(t | \mathcal{H}_t) + o(\lambda(t | \mathcal{H}_t))$. Thus, the intensity can be interpreted as an approximate ELEP for the next day, given the history of the extreme losses of the asset.

To be concrete, we give an example. We use the same oil data used in Section 3. The estimated parameters are $\hat{\omega} = 0.5355$, $\hat{\alpha} = 0.1663$, $\hat{\beta} = 0.7540$, and $\hat{\eta} = -11.4166$. The last excess loss value over the threshold of $u = 0.0265$ was $Y_{N(t)} = 0.0367$, which occurred on Monday, November 3, 2008. Using the residuals in Equation (9), we have $\hat{\epsilon}_{N(t)} = 1.8814$, and $\hat{\Psi}_{N(t)} = 1.1597$. Using these values, by plugging into equation (10), we obtain $\hat{\lambda}(t | \mathcal{H}_t) = 0.2714$. The data end on Tuesday, November 4, 2008. Thus, given the extreme loss history from the past 1000 (trading) days, the estimated ELEP for the oil returns—that is a loss over the threshold of 0.0265 or 2.65%—for Wednesday, November 5, 2008 is $1 - \exp(-0.2714) = 0.2376$. Note that this value is quite close to the estimated intensity.

We can quantify the relationship between a small change in ELEP $\Delta p_u(t)$ and a small change in intensity $\Delta \lambda(t | \mathcal{H}_t)$ from Equation (12) as follows:

$$\frac{\Delta p_u(t)}{p_u(t)} = \frac{1 - \exp(-\Delta \lambda(t | \mathcal{H}_t))}{\exp(\lambda(t | \mathcal{H}_t)) - 1} \approx \frac{\Delta \lambda(t | \mathcal{H}_t)}{\lambda(t | \mathcal{H}_t)}.$$

Thus, the relative change in the intensity is approximately equal to the relative change in ELEP.

A question of interest for a practitioner may be in what way and how much does an extreme loss affect the probability of sustaining another extreme loss in the next period, given the history. For a small change in the excesses losses $\Delta Y_{N(t)}$, we have

$$\frac{\Delta \lambda(t | \mathcal{H}_t)}{\lambda(t | \mathcal{H}_t)} = \exp(-\eta(\Delta Y_{N(t)})) - 1 \approx -\eta(\Delta Y_{N(t)}).$$

Applying the aforementioned approximation to our current values means, for example, that given a 0.01 unit increase in the excess losses $\Delta Y_{N(t)} = 0.01$, the relative ELEP increases by $-0.01 \times -11.4166 = 0.114166$ or 11.4166%. Direct computation of $\Delta p_u(t) / p_u(t)$ would have yielded 10.3553%.

Our presentation shows the dual purpose of the intensity function: it provides an exact expression for the ELEP and also approximates ELEP. In the next section, a third function of the intensity is demonstrated in the derivation of the VaR.

5. Estimation of value at risk

In this section, we obtain an estimate for the 1-day VaR, as defined by Equation (2). The distribution of the Y_i 's equals the conditional distribution of $Z_{T_i} - u$, given the event $\{Z_{T_i} > u\}$. We are concerned with extreme losses where the value of u is relatively large. A well-established result from EVT shows that this conditional distribution becomes asymptotically the GPD:

$$\mathbb{P}(\{Y \leq x\}) = H_{\kappa, \xi}(x) \equiv 1 - \left(1 + \xi \frac{x}{\kappa}\right)_+^{-1/\xi},$$

where $(x)_+ = \max\{0, x\}$ denotes the positive part of x . Here, $\kappa > 0$ is the scale, and $\xi \in \mathbb{R}$ is the shape parameter.

More precisely, we have the following:

Theorem 2 (Pickands–Balkema–De Haan)

Suppose that Z belongs to the domain of max attraction of an extreme value distribution. Let u^* denote the upper endpoint of the distribution of Z . Then, we have

$$\lim_{u \rightarrow u^*} \sup_{0 \leq x \leq u^* - u} |\mathbb{P}(\{Z - u \leq x\} | \{Z > u\}) - H_{\kappa(u), \xi}(x)| = 0, \quad (13)$$

for some function $\kappa(u) > 0$ and a fixed shape parameter ξ .

Note that a random variable Z is said to belong to the domain of max attraction of an extreme value distribution G if for i.i.d. Z, Z_1, \dots, Z_n , there exist constants $c_n > 0$ and $d_n \in \mathbb{R}$ such that

$$\mathbb{P}(\{c_n^{-1}(\max(Z_1, \dots, Z_n) - d_n) \leq z\}) \longrightarrow G(z) = \exp\{-H_{1,\xi}(z)\}, \quad \text{as } n \rightarrow \infty.$$

The distribution G is one of the three extreme value types: Gumbel, Fréchet, or Weibull.

Depending on the values of ξ in (13), one has three different classes of GPD: (i) *the classical Pareto* heavy-tailed distribution, when $\xi > 0$; (ii) *exponential* when $\xi = 0$; and (iii) *bounded above* when $\xi < 0$.

For a first proof of this result, see [33]; extensive theoretical and application-oriented discussions can be found in [34–36]. Risk management applications of the Pickands–Balkema–de Haan theorem are discussed in [9].

Practically, all well-known distributions used to model the returns, such as the Student t , the normal, and their skewed counterparts such as the normal inverse Gaussian distribution, satisfy the conditions of the aforementioned theorem. Therefore, the excess losses Y_i 's will follow approximately a GPD for sufficiently high threshold u . That is,

$$\mathbb{P}(\{Y_i > y\}) \approx 1 - H_{\kappa,\xi}(y) \equiv \left(1 + \xi \frac{y}{\kappa}\right)_+^{-1/\xi}. \quad (14)$$

We will model the Y_i 's as independent GPD random variables. The independence assumption may appear to be restrictive, but as argued in [23], explicit modeling of the dependence of excess losses can lead to irregular VaR estimates without much improvement in the accuracy of the estimation.

The log returns of the financial assets are not generally modeled with bounded distributions, although a theoretical loss of 100% of the value of an asset represents an upper bound. Nevertheless, the GPD gives us the flexibility to model a larger class of distributions.

Given a sample of excess values $Y_i, 1 \leq i \leq n$, one can estimate the parameters ξ and κ via MLE. The log-likelihood is given by

$$\log \left(\prod_{j=1}^n f_{Y_j}(y_j) \right) = -n \log \kappa - (1 + 1/\xi) \sum_{i=1}^n \log(1 + \xi y_i / \kappa)_+, \quad (15)$$

where $f_{Y_j}(y_j)$ is the density of a GPD and is non-regular because the support of GPD depends on the parameters. Nevertheless, Smith [37] has shown that the MLEs of ξ and κ maintain the asymptotic normality properties of the MLEs of regular models, as long as $\xi > -1/2$. Recently, Zhou [38] showed that the likelihood equations have a consistent solution for ξ , provided $\xi > -1$. Note that, for the values of $\xi \geq 1$, we have $\mathbb{E}[Y_i] = \infty$ [34]. See [39] for reducing the bias of the GPD estimates and [40] for a new MLE-based method that attempts to get around some of the issues encountered in MLE estimation.

In our work, we never encountered GPD with estimates of $\xi \leq -1/2$ or $\xi \geq 1$. Thus, the existing MLE theory ensures that we can obtain consistent estimates of the parameters ξ and κ of finite mean Y_i 's.

The choice of the appropriate threshold u above which the excess losses are well approximated by the GPD model is an open problem and remains an active research area. For recent developments, see [41, 42] and their corresponding references.

A number of practical approaches for the choice of u from data have been developed. One of them is to pick a relatively high initial threshold u_0 , then estimate $\xi \equiv \xi(u)$ as function of increasing thresholds $u > u_0$. One can then examine the plot of ξ versus u . Because by Equation (13), the value of ξ should be asymptotically constant, an adequate choice of u corresponds to the region in the plot where the ξ estimates start to stabilize as a function of the threshold. Similar analysis with κ cannot be carried out as κ is often a function of u as seen in Equation (13). Although reparametrization can be carried out to alleviate this issue, see, for instance, [35, p. 83], the examination of the plot of ξ versus u is most often performed.

Another very useful diagnostic tool for assessing the goodness of fit for the GPD model is to examine the following residuals:

$$\widehat{E}_i = \frac{\log \left(1 + \widehat{\xi} Y_i / \widehat{\kappa} \right)}{\widehat{\xi}}.$$

If the threshold u is chosen well, then Y_i 's are approximately GPD with parameters $\widehat{\xi}$ and $\widehat{\kappa}$. By a simple variable transformation, one can verify that \widehat{E}_i 's form an identically distributed sequence of exponential random variables with mean 1. In

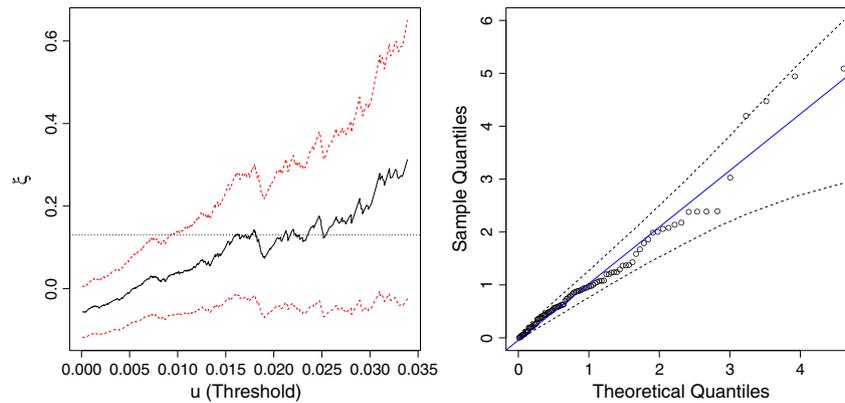


Figure 2. The data set is daily negative returns of West Texas Intermediate oil from November 15, 2005 to November 3, 2008, for a total of 1000 days. The left panel shows estimates of the shape parameter ξ as a function of the threshold u in the generalized Pareto distribution model. Observe that the ξ 's seem to stabilize for $u = 0.016$ to 0.0265 . The dotted horizontal line is equal to $\xi = 0.13$. The solid lines are the actual ξ estimates, and the dashed lines correspond to the 95% confidence intervals of ξ . The right panel shows the Q–Q plot of the generalized Pareto distribution residuals versus a unit rate exponential random variable for threshold $u = 0.0265$ along with pointwise 95% confidence intervals (dashed lines).

practice, a Q–Q plot of the residuals \widehat{E}_i versus standard exponential random variable should appear linear if the excesses are approximately GPD.

The left panel of Figure 2 shows that an adequate choice of u for the same oil returns discussed in Section (4) could be from 0.016 (1.6%) to 0.0265 (2.65%). The threshold of 0.0265 is approximately equal to the 0.90 quantile of the data. An estimate of the shape parameter ξ is 0.13. However, the Q–Q plot on the right panel of Figure 2 confirms that the choice of $u = 0.026$ illustrated on the left panel is adequate. For more examples, see [34, p. 360], [35, p. 85], and [43, p. 148].

In our analysis, we will always set u to the 0.90 quantile of the data as we will be performing thousands of estimations of the GPD parameters in Section 6. McNeil and Frey [12], via a simulation study, showed that this choice is reasonable in situations where checking of the residuals and the plots of the ξ estimates versus the thresholds is not practical.

To obtain an estimator for VaR, begin with Equation (16); subtracting u and conditioning on the event $\{Z_{t+1} > u\}$, we get

$$1 - q = \mathbb{P}(\{Z_{t+1} > z_q(t)\} | \mathcal{H}_t) = \mathbb{P}(\{Z_{t+1} - u > z_q(t) - u\} | Z_{t+1} > u, \mathcal{H}_t) \mathbb{P}(\{Z_{t+1} > u\} | \mathcal{H}_t),$$

where $z_q(t) > u$. Because $\{Y_{N(t)+1} > 0\} = \{Z_{t+1} > u\}$ and with the use of the event equality in Equation (11), we obtain

$$1 - q = \mathbb{P}(\{Y_{N(t)+1} > z_q(t) - u | Y_{N(t)+1} > 0, \mathcal{H}_t\}) \mathbb{P}(\{N(t, t + 1) = 1\} | \mathcal{H}_t).$$

The independence of $Y_{N(t)+1}$ from \mathcal{H}_t and the fact that $Y_{N(t)+1}$ follows a GPD with parameters ξ and κ , along with Equation (12), imply that

$$\begin{aligned} 1 - q &= \left(1 + \xi \frac{z_q(t) - u}{\kappa}\right)_+^{-1/\xi} (1 - \mathbb{P}(\{N(t, t + 1) = 0\} | \mathcal{H}_t)) \\ &= \left(1 + \xi \frac{z_q(t) - u}{\kappa}\right)_+^{-1/\xi} (1 - \exp(-\lambda(t | \mathcal{H}_t))). \end{aligned}$$

By solving the last relation for $z_q(t)$, we obtain

$$z_q(t) = u + \frac{\kappa}{\xi} \left[\left(\frac{1 - \exp(-\lambda(t | \mathcal{H}_t))}{1 - q} \right)^\xi - 1 \right]. \quad (16)$$

The aforementioned estimate is valid regardless of the sign of the shape parameter ξ , provided that $z_q(t) > u$ and $\lambda(t | \mathcal{H}_t) \geq -\log(q)$. The inequalities imply that Equation (16) is only valid for quantile estimates above the threshold u and that the intensity must satisfy a lower bound of $-\log(q)$. An estimate of $z_q(t)$ can now be obtained by plugging in the MLE results from Equation (15) and the intensity from Equation (6).

To put Equation (16) into perspective, we derive the unconditional VaR at level q . Assuming that the distributions of the excess losses $Z_t - u$ given $Z_t > u$ are approximately GPD, by conditioning on the event $\{Z_t > u\}$, we obtain the following:

$$z_q^{\text{un}} = u + \frac{\kappa}{\xi} \left[\left(\frac{\mathbb{P}(\{Z_1 > u\})}{1 - q} \right)^{\xi} - 1 \right]. \quad (17)$$

Comparing our expression for VaR in Equation (16) with Equation (17), we see that $1 - \exp\{-\lambda(t|\mathcal{H}_t)\}$ plays the role of $\mathbb{P}(\{Z_t > u\})$. This is in line with our intuition, because the rate at which one encounters losses Z_t exceeding level u is governed by the probability $\mathbb{P}(\{Z_t > u\})$. Our methodology allows one to keep track of this rate locally in time by using the intensity $\lambda(t|\mathcal{H}_t)$ of the associated MPP $\{(T_i, Y_i)\}_{i \geq 1}$. The unconditional expression for VaR in Equation (17) is equivalent to assuming a constant intensity of $\lambda = -\log(1 - \mathbb{P}(\{Z_1 > u\}))$. For further details on the theoretical properties of the aforementioned unconditional estimator, see [44].

6. Backtesting

To validate our methodology and assess its accuracy in estimating ELEP and VaR, we conduct backtesting. In Section 6.1, we describe the general backtesting procedure and present the results in Sections 6.2 and 6.3.

6.1. Description of the backtesting procedures

For each asset, we focus on the time series of its returns $\{Z_t\}_{1 \leq t \leq N}$, where N stands for the sample size rather than the point process $N(t)$ in Section 2. We consider a window of length $l = 1000$ days, corresponding to about 4 years of trading. The model under consideration is first fitted to a moving window of data $\{Z_{t-l+1}, Z_{t-l+2}, \dots, Z_t\}$, for each $t \in \{l, l+1, \dots, N-2, N-1\}$. The resulting model parameters yield an estimate of the intensity in Equation (6), which is then used to obtain ELEP from Equation (12) and the 1-day VaR from Equation (16).

Backtesting ELEP. We compare the estimated ELEP with the realized future value of Z_{t+1} . Note that we cannot observe the actual or realized probability of loss. However, if we are estimating ELEP accurately, then higher values of ELEP should be associated with events of extreme losses. We can assess this relationship via logistic regression, where the response is a sequence of 1's and 0's, with 1 corresponding to the occurrence of the Bernoulli event $\{Z_{t+1} > u\}$ with some probability θ_{t+1} and the predictor is the ELEP estimates. Our logistic regression model becomes the following:

$$\log \left(\frac{\theta_{t+1}}{1 - \theta_{t+1}} \right) = \phi_0 + \phi_1 \hat{p}_u(t), \quad (18)$$

which we estimate in the usual generalized linear model context (see, e.g., the function `glm` in the package R).

The coefficient ϕ_1 assesses the statistical relationship between the estimated ELEP and the event of extreme loss. We report the two-sided p -values for the parameter ϕ_1 , and following the convention in hypothesis testing, we say that ELEP successfully predicts extreme loss events when the p -values are below the critical level of 5%. Note that this is a version of the conditional autoregressive VaR test in [45].

Backtesting VaR. We first use the procedure described in [12, 46] (see also Appendix A). (We focused on this specific backtesting procedure because regulatory bodies, such as Basel Committee on Banking Supervision, impose a penalty on the bank (by requiring the bank to hold additional reserves) if the observed number of violations over a fixed period is too high. This is analogous to the backtesting procedure we have described.) We additionally use the well-known Ljung–Box (LB) of lag 1 as described in the following text.

We obtain VaR values $\hat{z}_q(t)$, for the quantiles $q \in \{0.95, 0.99, 0.995\}$, which are also compared with the realized future value Z_{t+1} . We say a *violation* has occurred whenever $\{\hat{z}_q(t) < Z_{t+1}\}$. To examine how well a certain methodology tracks VaR, we examine the *violation sequence*, that is, a sequence of 1's and 0's, with 1 representing a violation and 0 the occurrence of the event $\{\hat{z}_q(t) \geq Z_{t+1}\}$. Note that this sequence is different than the sequence produced by the extreme loss event occurrence. That sequence was based on the event of loss over the threshold u as opposed to loss greater than the estimated VaR.

If a model—for VaR, we will compare our results with that of other methods—accurately estimates the conditional quantiles, then the violations should be approximately Bernoulli random variables with probability of ‘success’ $1 - q$. Furthermore, a model that incorporates well the statistical dependence in the data would yield approximately independent violations. In the ideal case, the number of violations in the corresponding violation sequence should be approximately binomially distributed with expectation $(N - l)(1 - q)$. This suggests using a two-sided binomial test to assess whether a certain model leads to substantially overestimating or underestimating VaR. The normal approximation to the binomial

can also be used (see, e.g., [46]). Large p -values would support the model under study, whereas p -values close to zero indicate poor performance. We say that a method fails to track VaR if it yields p -values below the critical level of 5%.

Here, we use two-sided p -values to evaluate various methods. This is because both overestimation and underestimation of VaR can be problematic in practice. Indeed, overestimating VaR is detrimental to the risk taker as it may trigger overly conservative strategies and, thus, limit the profit potential. On the other hand, underestimating VaR clearly implies underestimating risk, which can lead to large losses.

To perform the LB test, we first computed the duration between violations of the VaR at each level $q \in \{0.95, 0.99, 0.995\}$. Then, the p -value for the autocorrelation of lag 1 are obtained. It is well known that the test statistics has a chi-squared distribution with the degrees of freedom equal to 1 under the null hypothesis that there is no autocorrelation. Large p -values again support the model under study.

There are other VaR backtesting procedures that can be used (see [11, 46, 47] for further details).

6.2. Results on backtesting ELEP

We apply the backtesting procedure to the data sets described in Table III. Table I summarizes our results by reporting the estimated parameters of the logistics regression in Equation (18) and the p -value associated with the parameter ϕ_1 . The intercept terms ϕ_0 's are all near $\log(0.10) = -2.30$. Our thresholds were set at the value corresponding to the 0.90 quantile of the data, which implies that, on average, an extreme loss event occurs with a probability of 0.10. The intercept term ϕ_0 captures this, whereas the ϕ_1 's provide additional adjustment depending on the values of ELEP. With the exception of the Standard & Poor (S&P) 500 data, the estimated ELEPs are highly statistically significant predictors of the extreme loss events. Note that all estimated coefficients ϕ_1 's are positive. This means that higher ELEP values are positively associated with an increase in the ELEP. This is consistent with intuition and confirms the success of the ELEP estimator in tracking the conditional probability of extremes.

6.3. Results on backtesting value at risk

We conduct two sets of backtesting for VaR. First, we apply our method to the same data sets used in [12] and are available from <http://www.ma.hw.ac.uk/~mcneil/>. This allows for a direct comparison between our approach and the established approach of [12]. Second, we perform extensive backtesting by using 10 data sets of various types of assets.

McNeil and Frey [12] obtained their estimates of VaR by first pre-whitening the volatility of the time series with a GARCH(1,1) model and the return series with an AR(1) process. They then applied GPD modeling to the residuals of the pre-whitened series to estimate the appropriate quantiles using Equation (17). The 'conditional EVT' is the flagship method proposed in [12]. For detailed descriptions, see [12]. A brief description of their study and their main models is given in Appendix A.

6.3.1. Results of comparisons with the competing methods. The data in [12] include the S&P 500 index from January 1960 to June 1993, the DAX index from January 1973 to July 1996, the BMW share price over the same period, the US dollar–British pound exchange rate from January 1980 to May 1996, and the price of gold from January 1980 to December 1997.

Table I. Results of the backtesting of extreme loss event probability.			
Asset	$\hat{\phi}_0$	$\hat{\phi}_1$	p -value of ϕ_1
Gold	-2.31	1.06	0.00
Silver	-2.36	1.72	0.00
Oil	-2.52	2.36	0.00
USD/GBP	-2.45	1.96	0.00
IBM	-2.16	0.24	0.01
SP500	-2.17	0.01	0.11
DAX	-2.26	0.73	0.00
JPM	-2.32	1.51	0.00
UKX	-2.68	4.25	0.00
Bonds	-2.14	1.24	0.02

The second column gives the estimated intercept values. The third column gives the estimated ϕ_1 parameter. The fourth column contains the p -values associated with the ϕ_1 estimates.

The expected number of violations, the observed number of violations, and the two-sided binomial test p -values obtained via the backtesting procedure described in Section 6.1 are reported in Table II, which is a reproduction of Table II in [12] with our intensity-based row added.

With the marginal exception of gold at the 0.95 level, the proposed ‘intensity-based’ method never ‘fails’, that is, it always yields binomial p -values greater than 5%. In 7 of 15 cases, our method outperforms or ties with the McNeil/Frey’s Conditional EVT in terms of how close the number of observed violations are to the number of expected ones. This first backtesting scenario shows that our method is competitive and performs as well as the conditional EVT method of [12].

The unconditional EVT and conditional normal fail to track conditional VaR well. This can be attributed to the fact that the former ignores the temporal dependence structure in the data, whereas the latter fails to capture the heavy tails in the returns. The conditional t method does better than the unconditional EVT and conditional normal because it models the dependence and also allows for heavy-tailed returns.

Ultimately, the McNeil/Frey’s conditional EVT and our intensity-based methods perform best, because they model well both the temporal dynamics and the extreme value behavior of the returns.

6.3.2. Application to a diverse set of asset classes. Here, we further test our methodology of estimating VaR. We have already demonstrated our method’s competitiveness with the widely referenced method of [12] in the previous section.

We consider time series of 10 diverse set of assets. The 20-year US Treasury yields return series was multiplied by -1 , so an increase in the yield of the treasury represents a gain to the holder of the short position. The stock and stock index data sets are adjusted for dividends and splits. For more details, see Table III.

Table IV shows the backtesting performance of our ‘intensity-based’ method. It performs uniformly well over essentially all data sets and for all quantile levels, failing (at level 5%) only once for the case of UKX. As far as the serial dependence of the VaR violations are concerned, with the exception of the index data (S&P500, DAX, and UKX), IBM, and silver, all at the 0.95 level, the LB results are good.

Table II. The expected number of violations, the observed number of violations, and the two-sided binomial test p -values obtained via the backtesting procedure.

Asset	S&P 500	DAX	BMW	USD–GBP	Gold
Length of test	7414	5146	5146	3274	3413
0.95 quantile					
Expected	371	257	257	164	171
Conditional EVT	366 (0.81)	258 (0.97)	261 (0.82)	151 (0.32)	155 (0.22)
Conditional normal	384 (0.49)	238 (0.22)	210 (0.00)	169 (0.69)	122 (0.00)
Conditional t	404 (0.08)	253 (0.80)	245 (0.44)	186 (0.08)	168 (0.84)
Unconditional EVT	402 (0.10)	266 (0.59)	251 (0.70)	156 (0.55)	131 (0.00)
Intensity based	392 (0.26)	259 (0.92)	247 (0.52)	161 (0.84)	146 (0.05)
0.99 quantile					
Expected	74	51	51	33	34
Conditional EVT	73 (0.91)	55 (0.62)	48 (0.67)	35 (0.72)	25 (0.12)
Conditional normal	104 (0.00)	74 (0.00)	86 (0.00)	56 (0.00)	43 (0.14)
Conditional t	78 (0.68)	61 (0.18)	52 (0.94)	40 (0.22)	29 (0.39)
Unconditional EVT	86 (0.18)	59 (0.29)	55 (0.62)	35 (0.72)	25 (0.12)
Intensity based	65 (0.29)	54 (0.73)	48 (0.67)	40 (0.22)	29 (0.39)
0.995 quantile					
Expected	37	26	26	16	17
Conditional EVT	43 (0.36)	24 (0.77)	29 (0.55)	21 (0.26)	18 (0.90)
Conditional normal	63 (0.00)	44 (0.00)	57 (0.00)	41 (0.00)	33 (0.00)
Conditional t	45 (0.22)	32 (0.23)	18 (0.14)	21 (0.26)	20 (0.54)
Unconditional EVT	50 (0.04)	36 (0.05)	31 (0.32)	21 (0.26)	11 (0.15)
Intensity based	45 (0.22)	31 (0.32)	25 (0.92)	21 (0.26)	18 (0.81)

The rows labeled ‘conditional EVT’ correspond to the flagship method of [12]. The results of our ‘intensity based’ method are indicated in boldface. The ‘conditional normal’ and ‘conditional t ’ correspond to pre-whitening the data series via AR(1)–GARCH(1,1) filters with standard normal and Student t innovations, respectively. The rows labeled ‘unconditional EVT’ is based on Equation (17).

Table III. Summary information of the data.

Item	Data	Start date	End date	Length	Source
1	S&P 500 index	03-23-70	11-03-08	9750	Bloomberg
2	DAX index	03-03-70	11-03-08	9718	Bloomberg
3	FTSE 100 index	01-04-84	11-03-08	6282	Bloomberg
4	Gold spot	01-02-80	11-03-08	7370	Bloomberg
5	Silver spot	01-03-84	11-03-08	6380	Bloomberg
6	WTI oil spot	01-02-86	11-04-08	5765	http://eia.doe.gov
7	20-Year US Treasury yields	10-01-93	11-03-08	3776	http://www.federalreserve.gov
8	USD/GBP	01-02-90	11-03-08	4742	http://www.federalreserve.gov
9	IBM	03-03-70	11-03-08	9763	Bloomberg
10	JPM	07-29-80	11-03-08	7134	Bloomberg

S&P 500, DAX, and FTSE are the US Standard & Poor 500, the German DAX, and the British 100 Large Cap indices, respectively. Gold and silver spot prices are expressed in US dollars per ounce. WTI oil is the benchmark West Texas Intermediate spot crude oil price in US dollars. The 20-year US Treasury yields are the US constant maturity yields on 20-year notes. USD/GBP is the US dollar currency rate per 1 unit of GBP. IBM and JPM are the US stocks of the IBM Corporation and J.P. Morgan.

Table IV. Backtesting performance of our intensity-based method.

Asset	Gold	Silver	Oil	USD/GBP	SP500	DAX	UKX	IBM	JPM	Bonds
Data length (N)	6369	5379	4764	3741	8749	8717	5281	8762	6133	2775
0.95 quantile										
Expected	318	269	238	187	437	436	264	438	307	139
Intensity based	286	246	223	193	417	409	248	468	307	154
Binom p -value	(0.06)	(0.15)	(0.32)	(0.65)	(0.33)	(0.19)	(0.31)	(0.15)	(0.98)	(0.19)
LB p -value	(0.27)	(0.00)	(0.73)	(0.40)	(0.00)	(0.00)	(0.00)	(0.01)	(0.24)	(1.00)
0.99 quantile										
Expected	64	54	48	37	87	87	53	88	61	28
Intensity based	64	55	47	40	97	89	69	96	68	38
Binom p -value	(1.00)	(0.89)	(1.00)	(0.68)	(0.31)	(0.87)	(0.03)	(0.39)	(0.40)	(0.07)
LB p -value	(1.00)	(0.51)	(0.42)	(0.19)	(0.31)	(0.36)	(0.55)	(0.27)	(0.96)	(0.17)
0.995 quantile										
Expected	32	27	24	19	44	44	26	44	31	14
Intensity based	32	31	24	19	62	52	30	49	37	17
Binom p -value	(0.93)	(0.44)	(1.00)	(1.00)	(0.01)	(0.22)	(0.49)	(0.45)	(0.28)	(0.42)
LB p -value	(0.86)	(0.10)	(0.42)	(0.37)	(0.18)	(0.96)	(0.66)	(0.70)	(0.43)	(0.72)

The format of this table is similar to that of Table II. The expected number of violations and the observed number of violations along with the two-sided binomial (Binom) p -values and the Ljung–Box (LB) p -values are reported for our intensity-based method.

7. On the temporal dynamics of extreme losses

An important advantage of the proposed method over existing techniques lies in its interpretability. Namely, it captures explicitly the interaction between the extreme losses and their intensities. To gain intuition about this self-exciting phenomenon, consider first the scatterplot of the excesses Y_i as a function of the next interexceedance times X_{i+1} . Observe that the points (X_{i+1}, Y_i) in Figure 3 (left panel) tend to ‘hug’ the two axes. This implies that an extreme loss is typically followed by a short interexceedance time and vice versa. Because, however, the distributions of the Y_i ’s and X_{i+1} ’s are highly skewed, this scatter plot alone does not clearly show the dependence between the variables. Moreover, the Y_i ’s are heavy tailed (Figure 3 (right panel)); and thus, the classical (Pearson) correlation analysis may not be adequate if $\mathbb{E}[Y_i^2] = \infty$. Instead, to quantify the dependence between the X_{i+1} ’s and Y_i ’s, we resort to the Spearman’s *rank correlation*. The *rank correlation* coefficients between the X_{i+1} ’s and the Y_i ’s for 10 assets for different thresholds and settings are shown in Table V. Observe that in only 3 of 60 cases, we have positive rank correlations, whereas the majority exhibiting large negative ones. This elementary analysis complements the empirical finding of clustered extreme losses. In fact, they also exhibit self-excitation; namely, the greater the loss magnitude, the higher the intensity of losses in the future. This is captured by our model through the presence of the coefficient η in Equation (5).

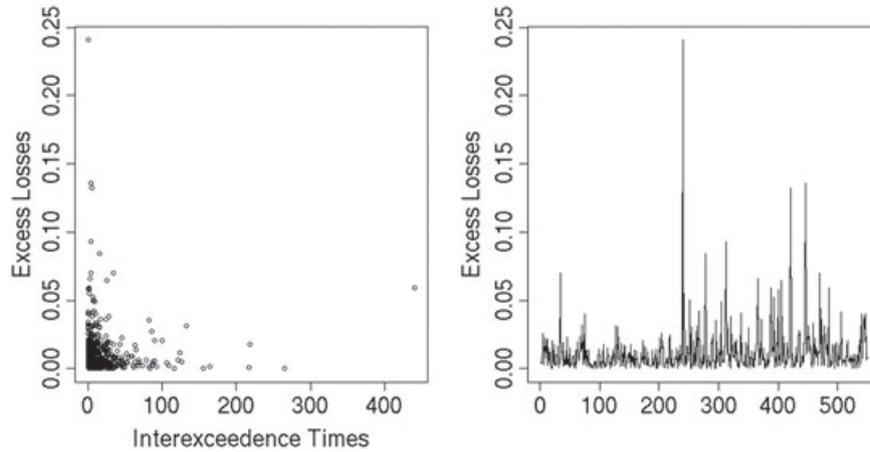


Figure 3. Left panel: excess losses $Y_i = Z_i - u$ versus interexceedence times X_i . Here, the Z_i 's are negative daily log returns for the IBM stock from March 1970 to November 2008. The threshold u equals the 0.95 quantile of the loss empirical distribution. Right panel: time series plot of the Y_i 's. Observe their highly skewed and heavy-tailed behavior.

Table V. Spearman's rank correlation coefficients between the excesses Y_i 's and the next interexceedence times X_{i+1} .

Asset	Level					
	90%	95%	99%	90%	95%	99%
Gold	-0.258	-0.225	-0.279	-0.095	-0.142	-0.079
Silver	-0.233	-0.189	0.006	-0.202	-0.092	-0.157
Oil	-0.161	-0.243	-0.224	-0.155	-0.197	0.013
USD/GBP	-0.100	-0.257	0.041	-0.085	-0.180	-0.277
IBM	-0.213	-0.189	-0.078	-0.132	-0.178	-0.187
SP500	-0.265	-0.220	-0.233	-0.238	-0.202	-0.183
DAX	-0.260	-0.321	-0.236	-0.209	-0.234	-0.161
JPM	-0.230	-0.202	-0.080	-0.188	-0.254	-0.195
UKX	-0.244	-0.272	-0.499	-0.193	-0.252	-0.326
Bonds	-0.065	-0.122	-0.120	-0.054	-0.139	0.009

The level u was chosen to be the 0.90th, 0.95th, and 0.99th quantiles. In the first three levels, the percentiles correspond to the entire data sets; in the last three levels, the percentiles u were computed from moving windows of the data of size 1000 data points (about 4 years of trading activity).

In the backtesting, our intensity model was fit to sliding windows of the data. The resulting approximations of the Hessian matrix at the MLE parameter estimates yield standard error estimates for the parameters in the model. It is of interest to test the significance of the *interaction* parameter η . For each of the 10 data sets and each of the data windows, we obtained two-sided p -values for the hypotheses

$$\begin{cases} H_0 & : \eta = 0, \\ H_a & : \eta \neq 0, \end{cases}$$

based on normal approximations. Table VI provides a summary with the values reported being the percentage of backtesting windows with significantly non-zero estimates $\hat{\eta}$ along with the fraction of negative $\hat{\eta}$'s, for each of the 10 data sets. Note that the majority of the $\hat{\eta}$'s are *negative*. This, in view of Equation (5), implies that an extreme Y_i increases the values of the intensity. That is, *self-excitation* occurs, and the subsequent interexceedence times are likely to be small. The results in Table VI support the elementary observations based on the rank correlations in Table V. This interaction phenomenon is statistically significant and should not be ignored when modeling and measuring risk.

Lastly, Figure 4 shows a plot of the estimated intensities for the oil data. We have labeled a few major events in the plot. It is evident from the plot that during crises, the intensity 'spikes', indicating higher ELEP. Similar plots for the rest of the data sets also exhibit spiking of the intensity during major events. This is further evidence that the estimated intensities are tracking the actual events corresponding to the extreme losses.

Table VI. The proportion of backtesting windows with statistically significant non-zero $\hat{\eta}$'s at level $\alpha = 5\%$ (second column) and the proportion of negative $\hat{\eta}$'s (third column).

Asset	% of significant η 's	% of negative η 's
Gold	65.89	75.51
Silver	43.09	84.09
Oil	51.78	82.77
USD/GBP	73.12	89.14
IBM	46.87	80.23
SP500	34.79	82.50
DAX	41.85	82.85
JPM	52.01	86.77
UKX	48.77	83.03
Bonds	54.96	62.58

Observe that for all 10 data sets, the η coefficient is statistically significant for substantial periods ranging from 34% to 73% of the backtesting windows. Also, the vast majority of the estimates of η are *negative*.

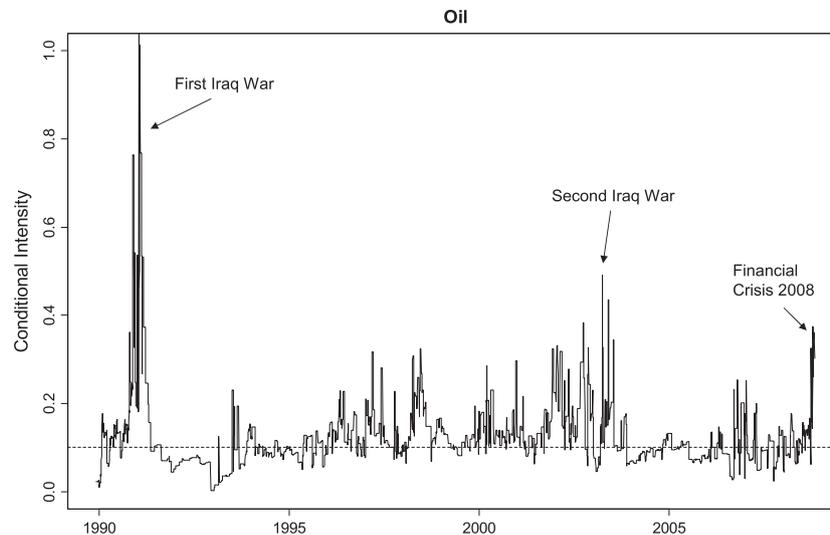


Figure 4. Conditional intensity of oil negative log returns. The horizontal line corresponds to the level of 0.10. A few spikes corresponding to major events have been labeled.

8. Concluding remarks and discussion

In this paper, we developed a methodology for the estimation of conditional ELEP and the VaR of an asset, given the history of past extreme losses. Motivated by the empirical fact that extreme losses cluster in time, we proposed to model directly the interplay between the extreme loss magnitudes and the times of their occurrence. Capturing the temporal dynamics allows us to improve upon the estimation of VaR, by taking into account the additional information available from the past behavior of the assets. Application to a large number of assets confirms that it performs well and competitive with existing methods. Further, it is easy to interpret and allows for direct testing of the interaction between extreme losses and their intensity of occurrence. Our findings strongly suggest the existence of a significant interaction, which should not be ignored when estimating risk measures, especially for the potential of extreme losses as recent events attest to.

An anonymous referee has raised the following valid point. Because the backtesting results of [12] are similar to ours, what is the case for using our methodology? Certainly from accurate VaR estimation point of view, our work competes but does not necessarily improve upon [12]. Our point process-based methodology, however, approaches the problem from a new angle and has the following new features: (i) we obtain an estimate of an intensity function, which can be interpreted as the probability of extreme loss in the next period; (ii) we can quantify the sensitivity of the probability of a future extreme loss to the changes in the current extreme losses; and (iii) our methodology could be extended to include many

other potentially useful covariates such as trading volume information, exchange rates, price of gold, and others. This would allow for modeling and estimation of conditional VaR and ELEP of large portfolios. In our framework, one can also test the significance of various covariates in a similar way as it was carried out in Section 7. We therefore hope for the proposed methodology to find practical applications to quantify as well as understand the structure of risk in practice.

APPENDIX A. Description of competing methods for value-at-risk estimation

The dynamic of the $\{Z_t\}_{t \geq 0}$ is assumed to follow

$$Z_t = \mu_t + \sigma_t W_t,$$

where the residuals W_t 's are an i.i.d. with mean zero, unit variance, and CDF of $F_W(w)$.

Furthermore, it is assumed that the conditional mean of the returns, μ_t , follows an AR(1) process:

$$\mu_t = \phi Z_{t-1}. \quad (19)$$

Letting $\varepsilon_t = Z_t - \mu_t$, the conditional variance is given by a GARCH(1,1) model:

$$\sigma_t^2 = b_0 + b_1 \varepsilon_{t-1}^2 + b_2 \sigma_{t-1}^2, \quad (20)$$

where b_0, b_1 , and b_2 are all positive.

The goal of McNeil and Frey [12] is also to estimate the same conditional quantile we defined in Equation (2). Letting $F_{Z_{t+1}|\sigma\{Z_s, s \leq t\}}(z)$ represent the conditional distribution of the Z_t , then

$$F_{Z_{t+1}|\sigma\{Z_s, s \leq t\}}(z) = \mathbb{P}(\{\sigma_{t+1} W_{t+1} + \mu_{t+1} \leq z | \sigma\{Z_s, s \leq t\}\}) = F_W((z - \mu_{t+1})/\sigma_{t+1}).$$

Therefore, a conditional quantile of the Z_t can be obtained as follows:

$$z_q^{\text{MF}}(t) = \mu_{t+1} + \sigma_{t+1} w_q, \quad (21)$$

where w_q is the upper q quantile of the marginal distribution of the i.i.d. sequence W_t and the superscript 'MF' stands for McNeil–Frey's conditional quantile estimate. The models in Equations (19) and (20) are fitted via pseudo-maximum likelihood method to obtain the parameter estimates $\hat{\phi}, \hat{b}_0, \hat{b}_1$, and \hat{b}_2 . The assumption is made that $\varepsilon_t \sim \mathcal{N}(0, 1)$ only to obtain the parameter estimates, but the normality assumption does not play a role in estimating the conditional quantile in Equation (21). For further details on pseudo-maximum likelihood method, see [48].

The pre-whitened residuals for a data of length N is obtained as follows:

$$(w_{t-N+1}, \dots, w_t) = \left(\frac{Z_{t-N+1} - \hat{\mu}_{t-N+1}}{\hat{\sigma}_{t-N+1}}, \dots, \frac{Z_t - \hat{\mu}_t}{\hat{\sigma}_t} \right).$$

The one-step forecasts are as follows:

$$\hat{\mu}_{t+1} = \hat{\phi} Z_t, \quad (22)$$

$$\hat{\sigma}_{t+1}^2 = \hat{b}_0 + \hat{b}_1 \varepsilon_t^2 + \hat{b}_2 \hat{\sigma}_t^2, \quad (23)$$

where $\hat{\varepsilon}_t = Z_t - \hat{\mu}_t$. Using the residuals, w_t 's, and fitting a GPD model to the excesses of these residuals over threshold u_w , we obtained a quantile estimate of w_q in the following form:

$$\hat{w}_q = u_w + \frac{\hat{\kappa}}{\hat{\xi}} \left[\left(\frac{\mathbb{P}(\{W > u_w\})}{1-q} \right)^{\hat{\xi}} - 1 \right]. \quad (24)$$

McNeil and Frey [12] always set u_w to the $q = 0.90$ quantile of the w_t 's, so $\mathbb{P}(\{W > u_w\}) = 0.1$. Finally, substituting the estimate forms in Equations (22)–(24) into Equation (21), one obtains the following estimate for the 1-day VaR

$$\hat{z}_q^{\text{MF}}(t) = \hat{\mu}_{t+1} + \hat{\sigma}_{t+1} \hat{w}_q.$$

The aforementioned equation is the 'conditional EVT' of [12].

Fitting a GARCH(1,1) model to Equation (20) with ε_t distributed as Student's t , with ν degrees of freedom, leads to the 'conditional t ' method. In this case, it can be shown that $\hat{w}_q = \sqrt{(\hat{\nu} - 2) / \hat{\nu}} F_T^{-1}(q)$, where F_T^{-1} is the inverse CDF of a Student's t -distribution with the estimated ν degrees of freedom. If one assumes that w_t 's are distributed as standard normal random variables, then $\hat{w}_q = \Phi^{-1}(q)$. This leads to the 'conditional normal' estimation. This model is similar to the standard RiskMetrics™ model widely used in industry (see, e.g., [26, p. 290]).

Acknowledgements

The authors thank the associate editor and three anonymous referees for their detailed comments, which helped improve the paper. Stilian Stoev and George Michailidis were partially supported by NSF grant DMS-0806094 and the NSF SCREMS grant DMS-0821196 at the University of Michigan.

References

1. Crouhy M, Galai D, Mark R. *The Essentials of Risk Management*. McGraw-Hill: New York, 2006.
2. Hull JC. *Risk Management and Financial Institutions*. Pearson Prentice Hall: Boston, 2007.
3. Bluhm C, Overbeck L, Wagner C. *An Introduction to Credit Risk Modeling*. Chapman and Hall: New York, 2002.
4. Panjer HH. *Operational Risk: Modeling Analytics*. Wiley-Interscience: Hoboken, 2006.
5. Dowd K, Blake D. After VaR: the theory, estimation, and insurance applications of quantile-based risk measures. *The Journal of Risk and Insurance* 2006; **73**(2):193–229.
6. Jorion P. Risk management for hedge funds with position information: both managers and investors could be satisfied. *Journal of Portfolio Management* 2007a; **34**(1):127–136.
7. Alexander C. *Market Risk Analysis, Value at Risk Models (Volume IV)*. John Wiley & Sons Inc.: West Sussex, England, 2008.
8. Artzner P, Delbaen F, Eber J-M, Heath D. Coherent measures of risk. *Mathematical Finance* 1999; **9**:203–228.
9. McNeil AJ, Frey R, Embrechts P. *Quantitative Risk Management: Concepts, Techniques, and Tools*. Pearson Addison Wesley: Princeton, 2005.
10. Jorion P. *Value at Risk: The New Benchmark for Managing Financial Risk*, (3rd edn). McGraw-Hill Companies: New York, 2007b.
11. Christoffersen PF. *Elements of Financial Risk Management*. Academic Press: Oxford, 2003.
12. McNeil A, Frey R. Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. *Journal of Empirical Finance* 2000; **7**:271–300.
13. Hsing T, Husler J, Leadbetter M. On the exceedance point process for a stationary sequence. *Probability Theory and Related Fields* 1988; **78**:97–112.
14. Hawkes AG. Point spectra of some mutually exciting point processes. *Journal of the Royal Statistical Society: Series B* 1971a; **33**:438–443.
15. Hawkes AG. Spectra of some self-exciting and mutually exciting point processes. *Biometrika* 1971b; **58**:83–90.
16. Hawkes AG. Spectra of some mutually exciting point processes with associated variables. In *Stochastic point processes: Statistical analysis, theory, and applications (Conference, IBM Research Center, Yorktown Heights, New York)*, Wiley-Interscience, New York, 1971; 261–271.
17. Snyder DL, Miller MI. *Random Point Processes in Time and Space*, (2nd edn). Springer: New York, 1991.
18. Karr AF. *Point Processes and their Statistical Inference*, (2nd edn). Marcel Dekker: New York, 1991.
19. Daley DJ, Vere-Jones D. *An Introduction to the Theory of Point Processes*, (2nd edn), Probability and its Applications (New York), Vol. I. Springer-Verlag: New York, 2003. Elementary theory and methods.
20. Engle RF, Russell JR. Autoregressive conditional duration: a new model for irregularly spaced transaction data. *Econometrica* 1998; **66**:1127–1162.
21. Christoffersen P, Pelletier D. Backtesting value-at-risk: a duration-based approach. *Journal of Financial Econometrics* 2004; **2**(1):84–108.
22. Focardi S, Fabozzi F. An autoregressive conditional duration model of credit-risk contagion. *Journal of Risk Finance* 2005; **6**:208–225.
23. Chaves-Demoulin V, Davison AC, McNeil A. Estimating value-at-risk: a point process approach. *Quantitative Finance* 2005; **5**(2):227–234.
24. Cox DR, Isham V. *Point Processes*. Chapman & Hall Ltd.: Boca Raton, 1988.
25. Bauwens L, Giot P. The logarithmic ACD model: an application to the bid-ask quote process of three NYSE stocks. *Annales d'Economie et de Statistique* 2000; **60**:117–149.
26. Tsay R. *Analysis of Financial Time Series*, (2nd edn). John Wiley & Sons: Hoboken, New Jersey, U.S.A., 2005.
27. Pacurar M. Autoregressive conditional duration (ACD) models in finance: a survey of the theoretical and empirical literature. *Journal of Economic Surveys* 2008; **22**(4):711–751.
28. Andersen T, Davis R, Kreiß J, Mikosch T (eds), *Handbook of Financial Time Series*. Springer: Berlin - Heidelberg, 2009.
29. Hamidieh K. Topics in statistical modeling and estimation of extremes and their dependence. *Ph.D. thesis*, University of Michigan, 2008.
30. Brandt A. The stochastic equation $Y_{n+1} = A_n Y_n + B_n$ with stationary coefficients. *Advances in Applied Probability* 1986; **18**(1):211–220.
31. Bauwens L, Galli F, Giot P. The moments of log-ACD models. *SSRN eLibrary*, 2003.
32. Allen D, Chan F, McAleer M, Peiris S. Finite sample properties of the QMLE for the log-ACD model: application to Australian stocks. *Journal of Econometrics* 2008; **147**(1):163–185. Econometric modelling in finance and risk management: An overview.
33. Pickands J. Statistical inference using the extreme order statistics. *The Annals of Statistics* 1975; **3**:119–131.
34. Embrechts P, Kluppelberg C, Mikosch T. *Modelling Extremal Events*. Springer-Verlag: Berlin - Heidelberg, 1997.
35. Coles SG. *An Introduction to Statistical Modeling of Extreme Values*. Springer-Verlag: Berlin - Heidelberg, 2001.
36. Beirlant J, Goegebeur Y, Segers J, Teugels J. *Statistics of Extremes*. Wiley: New York, 2004.
37. Smith RL. Maximum likelihood estimation in a class of nonregular cases. *Biometrika* 1985; **72**:67–90.
38. Zhou C. Existence and consistency of the maximum likelihood estimator for the extreme value index. *Journal of Multivariate Analysis* 2009; **100**(4):794–815.

39. Giles DE, Feng H. Bias-corrected maximum likelihood estimation of the parameters of the generalized Pareto distribution. *Econometrics Working Papers 0902*, Department of Economics, University of Victoria, 2009.
40. Zhang J, Stephens MA. A new and efficient estimation method for the generalized Pareto distribution. *Technometrics* 2009; **51**(3):316–325.
41. Tancredi A, Anderson C, O'Hagan A. Accounting for threshold uncertainty in extreme value estimation. *Extremes* 2006; **9**(2):87–106.
42. Gonzalo J, Olmo J. Which extreme values are really extreme? *Journal of Financial Econometrics* 2004; **2**(3):349–369.
43. Reiss R-D, Thomas M. *Statistical Analysis of Extreme Values: From Insurance, Finance, Hydrology, and Other Fields*, (3rd edn). Birkhäuser Verlag: Basel, Switzerland, 2007.
44. Smith RL. Estimating tails of probability distributions. *The Annals of Statistics* 1987; **15**:1174–1207.
45. Engle RF, Manganelli S. CAViaR: conditional autoregressive value at risk by regression quantiles. *Journal of Business and Economic Statistics* 2004; **22**(4):367–381.
46. Christoffersen P. *Encyclopedia of Quantitative Finance*. John Wiley and Sons.: New York, 2008. chapter Backtesting.
47. Berkowitz J, Christoffersen P, Pelletier D. Evaluating value-at-risk models with desk-level data. *Management Science* 2009; **57**(12):2213–2227.
48. Gouriéroux C, Monfort A, Trognon A. Pseudomaximum likelihood methods: applications to Poisson models. *Econometrica* 1984; **52**(3):701–720.