Statistics of extreme events in risk management: The impact of the subprime and global financial crisis on the German stock market

Rodrigo Herrera[1,], Bernhard Schipp[1]

[a]Facultad de Economía y Negocios, Universidad de Talca, Av. Lircay s/n- Talca, Chile
[b]Faculty of Business and Economics, Technische Universität Dresden, D-01062 Dresden, Germany.

Abstract

Given the growing need for managing financial risk and the recent global crisis, risk prediction is a crucial issue in banking and finance. In this paper, we show how recent advances in the statistical analysis of extreme events can provide solid methodological fundamentals for modeling extreme events. Our approach uses self-exciting marked point processes for estimating the tail of loss distributions. The main result is that the time between extreme events plays an important role in the statistical analysis of these events and could therefore be useful to forecast the size and intensity of future extreme events in financial markets. We illustrate this point by measuring the impact of the subprime and global financial crisis on the German stock market in extenso, and briefly as a benchmark in the US stock market. With the help of our fitted models, we backtest the Value at Risk at various quantiles to assess the likeliness of different extreme movements on the DAX, S&P 500 and Nasdaq stock market indices during the crisis. The results show that the proposed models provide accurate risk measures according to the Basel Committee and make better use of the available information.

JEL classification: C01; C58; C22; E44

Keywords: Extreme value theory, Value at Risk, Subprime crisis, German stock market.

1. Introduction

While the majority of European countries are experiencing a debt crisis, Germany has been enjoying the biggest economic boom with an extraordinary trade surplus sice the global crisis of 2008. Although Germany was initially hit hard by the global financial crisis, its exports helped the country's economy recover once the worst was over. The Deutscher Aktien Index, or DAX,

*Corresponding author

is Germany's primary blue chip stock index and is also on course to complete its best year since 2003, reaching a new peak of 8530.89 on May 20, 2013, surpassing its previous high mark seen on July 13, 2007.

Germany is the largest individual European economy; thus, the DAX is considered to be the most important index in Europe, which in turn makes it highly influential throughout the financial world. However, every financial crisis brings extreme losses to the world stock market, and attempts are made to minimize these as much as possible. For instance, one of these extreme losses occurred on June 4, 2012, when the DAX index dropped below 6000 points following the announcements of Spain and Italy's troubled banking sectors. This created extreme price movements not only in the stocks and indices of these countries but also to those across the European Union. A black year for the DAX stock market index was 2008, with five of the most extreme losses in its 25-year history. The DAX index lost more than 7% on each of those days, as for example, a drop of 523.98 points (7.16%) on January 21, 2008, which was attributable to steep losses in the financial industry of US$1.45 billion of investments by the German Bank WestLB.

Several authors (Chavez-Demoulin and McGill, 2012; Santos and Alves, 2012; Allen et al., 2013; McAleer et al., 2013; de Jesus et al., 2013; Hammoudeh et al., 2013; Santos et al., 2013) have argued that extreme value theory (EVT) allows us to explicitly take into account the extreme events contained in the tail distribution of losses. This provides three main advantages over classical methods, such as conditional models on volatility, historical simulations or the Gaussian distribution approach. First, as it is a parametric method, we can extrapolate the behavior of the tails to extreme levels, e.g., Value at Risk (VaR) estimations. Second, EVT focuses only on modeling the tail behavior of a loss distribution using only extreme values rather than the whole data set. Third, as we do not assume a particular model for returns, we propose a data-based approach to fit the distribution tails.

Unfortunately, EVT presumes independent identically distributed (i.i.d) observations, with stylized facts of stock market returns, such as clustered extremes and serial dependence, whereas the three worst daily losses in the DAX during October 2008 with losses greater than 7% typically violated this assumption. These problems are often addressed by the application of a declustering method, and then the standard model is fitted to the cluster maxima only. However, the drawback of these approaches is the information loss contained in these clusters (see Smith and Weissman, 1994; Laurini and Tawn, 2003; de Jesus et al., 2013). Another alternative is to apply a generalized autoregressive conditional heteroskedasticity (GARCH) or stochastic volatility model to the returns and then use EVT in the residuals (see, for example, McNeil and Frey, 2000; Allen et al., 2013; Santos et al., 2013).

In response to the stylized facts of extreme events in financial markets, a new line of research has emerged that incorporates the clustering behavior of these extreme events into the model. In particular, Chavez-Demoulin et al. (2005) propose a self-exciting version of the peaks over threshold (POT) model, the so-called Hawkes-POT model, while Herrera and Schipp (2013) introduce another alternative approach, the autoregressive conditional intensity POT (ACD-POT) model. Another similar idea is to include the inter-exceedance times (the time between extreme events) as covariates in the POT approach, the duration-POT (DPOT) approach, introduced by Santos and Alves (2012). The main advantage of these models is that the clustering behavior of extreme events is taken into account, making better use of the information given by the data to drive improvement in the estimation and forecast of different risk measures.

The contribution of this paper is twofold. First, we introduce new models whose main feature is to directly model extreme events above a high threshold previously defined, regardless of the cluster behavior demonstrated by extreme events in financial markets. Specifically, we raise the question whether the inter-exceedance times can contribute to the measurement accuracy of market risk in financial markets. In particular, whether the inter-exceedance times should be included as covariates to describe the size of the extreme events (DPOT, Hawkes-DPOT and ACD-DPOT model) or whether this should be included in terms of their intensities (Hawkes-POT and ACD-POT). To the best of our knowledge, no prior study has analyzed this issue to provide better measures of market risk. Second, Germany is one of the most important and the largest economies in the euro zone; therefore, we analyze the returns of the DAX index during the recent turmoil periods using the proposed models and compare these results with two leading US stock market indices, S&P 500 and Nasdaq, during the same time period.

The main conclusion is that while time observations play an important role in irregularly spaced data, the same occurs with inter-exceedance times in the statistical analysis of extreme events for financial markets. Roughly speaking, while Basel rules (Basel Committee on Banking Supervision, 2006) count the number of VaR exceptions, the proposed self-exciting marked point process (SEMPP) approach takes into account the time when the extreme events occur. In this way, the models capture the cluster behavior of these events correcting for possible inadequacy of risk models due to financial fluctuations, which by nature are extremely complicated during a crisis period. Furthermore, the SEMPP approach meets three desirable features that enhance the accuracy of the risk measures and raise the standard of risk management: the expected frequency of violations is in line with the selected confidence level, absence of dependence among VaR violations, and a quick reaction to changes in volatilities during the crisis period, avoiding violation clustering. Actually, these aspects are fully met by our methodological proposal and therefore provide strong support for the validity of the empirical analysis.

This paper is organized as follows. Section 2 introduces the framework for extreme events from the point of view of SEMPP. Section 3 presents the empirical analysis of the DAX index. This section includes a preliminary analysis of the stylized facts associated with extreme events in this market and the results of modeling and backtesting the returns of this stock market index during the subprime crisis. Section 4 includes a short analysis of US stock market indices during the same period with the aim of getting more research evidence and an in-depth understanding of the cluster behavior of extreme events during the crisis period in both countries. Conclusions and discussions are presented in Section 5.

2. Self-exciting marked point processes in EVT

In this section, we summarize the results for the POT approach from the point of view of the SEMPP theory, which underlies our modeling. General literature on the subject of EVT include Chavez-Demoulin et al. (2005); Herrera and Schipp (2009); Chavez-Demoulin and McGill (2012) and Herrera and Schipp (2013). From a practical point of view, suppose that we have observed the returns of a known stock market whose behavior is not necessarily i.i.d[1]. Now, imagine that we have only recorded the information of the most extreme events, that is, all the time events t whose magnitudes y have exceeded a high threshold u, previously defined. Thus, we obtain a set of events $\{(t_i, y_i)\}_{i\geq 1}$, which are defined in the space[2] $[0,1) \times [u, \infty)$. Under this point of view, the magnitudes y_i correspond to the marks, which carry information about the occurrence times t_i and whose internal history is given by $\mathscr{H}_t := \{(t_i, y_i) : i = 1, \ldots, N(t) - 1\}$ Thus, this set of events defines a marked point process $N(t) := \sum_{i\geq 1} \quad_{\{t_i \leq t, y_i\}}$ whose conditional intensity is defined by

$$\lambda(t, y \mid \mathscr{H}_t) = \lim_{\triangle t \to 0^+} \frac{1}{\triangle t} P\left(N\left([t, t + \triangle t) \times y\right) > 0 \mid \mathscr{H}_t\right)$$

or alternatively

$$\lambda\left(t, y \mid \mathscr{H}_t\right) = \lambda_g\left(t \mid \mathscr{H}_t\right) g\left(y \mid \mathscr{H}_t, t\right), \tag{2.1}$$

where $\lambda_g\left(t \mid \mathscr{H}_t\right)$ is a conditional intensity that describes only the behavior of the arrival times t of these extreme events, named the ground process, while $g\left(y \mid \mathscr{H}_t, t\right)$ corresponds to the probability density function of the exceedances, which in our case is a generalized Pareto density function according to the Pickands–Balkema–de Haan theorem (see Pickands III, 1975; Balkema and De Haan, 1974):

[1]As a convention in this paper, a negative value is treated as a positive number and extreme events take place when losses are part of the right tail of the distribution.

[2]Note that we defined the time interval between 0 and 1 for ease of exposition.

$$g\left(y \mid \mathscr{H}_t, t\right) = \begin{cases} \frac{1}{\beta(y|\mathscr{H}_t,t)} \left(1 + \xi \frac{y-u}{\beta(y|\mathscr{H}_t,t)}\right)^{-1/\xi - 1} &, \quad \xi \neq 0 \\ \frac{1}{\beta(y|\mathscr{H}_t,t)} \exp\left(-\frac{y-u}{\beta(y|\mathscr{H}_t,t)}\right)^{-1/\xi - 1} &, \quad \xi = 0, \end{cases} \qquad , \tag{2.2}$$

where $\beta\left(y \mid \mathscr{H}_t, t\right)$ is a scale parameter conditional on the history of the process and ξ is the shape parameter.[3] We consider only the scale parameter to be conditional because different empirical analyses have shown that it is reasonable to allow the shape parameter to remain constant (see Chavez-Demoulin et al., 2005; Chavez-Demoulin and McGill, 2012; Herrera, 2013; Santos and Alves, 2012; Santos et al., 2013). Finally, since the conditional intensity can also depend on the history of the process $\mathscr{H}_t$, these types of models are called self-excited (see Daley and Vere-Jones, 2003 for a more formal introduction to SEMPP).

Observe that for SEMPP models, the estimation of the VaR can be directly obtained. Indeed, the VaR at the α confidence level is the smallest value z_α^{t+1} for which the probability that the next observation period for Y_{t+1} will exceed z is not more than α using the information observed up to time t, which is the solution to $\mathbb{P}\left(Y_{t+1} > \mathrm{z}_\alpha^{t+1} \mid \mathscr{H}_t\right) = 1 - \alpha$. This probability can be estimated as follows:

$$\begin{aligned} \mathbb{P}\left(Y_{t+1} > \mathrm{z}_\alpha^{t+1} \mid \mathscr{H}_t\right) &= \mathbb{P}\left(Y_{t+1} > u \mid \mathscr{H}_t\right) \mathbb{P}\left(Y_{t+1} - u > \mathrm{z}_\alpha^{t+1} \mid Y_{t+1} > u, \mathscr{H}_t\right) \\ &= \left[1 - \mathbb{E}\left\{N\left([t, t+1) = 0 \mid \mathscr{H}_t\right)\right\}\right] G\left(y \mid \mathscr{H}_t, t\right) \\ &= \left\{1 - \exp\left(-\int_t^{t+1} \lambda_g\left(l \mid \mathscr{H}_l\right) dl\right)\right\} G\left(y \mid \mathscr{H}_t, t\right) \\ &\approx \lambda_g(t \mid \mathscr{H}_t) \times \overline{G}\left(y \mid \mathscr{H}_t, t\right), \end{aligned}$$

where $\overline{G}\left(y \mid \mathscr{H}_t, t\right)$ corresponds to the survival function of the cumulative distribution function of (2.2), the generalized Pareto distribution (GPD) function. Thus, to solve this equation for an event higher than u, the VaR is defined by

$$\mathrm{VaR}_\alpha^{t+1} = u + \frac{\beta\left(t, y \mid \mathscr{H}_t\right)}{\xi} \left\{ \left(\frac{\lambda_g(t \mid \mathscr{H}_t)}{1 - \alpha}\right)^{\xi} - 1 \right\}. \tag{2.3}$$

The last equation implies that the VaR is only defined for our models if $\lambda_g(t \mid \mathscr{H}_t) > 1 - \alpha$. Thus, the time when these extreme events occur as well as the dynamic behavior of these inter-exceedances can play a role in describing the future behavior of new extreme events. The log-likelihood L of the events $\{(t_i, y_i)\}_{i \geq 1}$ in a set $[0, T] \times [u, \infty)$ is given in terms of the conditional

[3]This density function is well-defined for $u \leq y < y_F$, where y_F is the right endpoint with values $y_F = \infty$ if $\xi > 0$ and $y_F = -\beta\left(y \mid \mathscr{H}_t, t\right)/\xi$ if $\xi < 0$.

intensity for the ground process and density for the marks as follows:

$$L = \sum_{i=1}^{N(T)} \log \lambda_g (t_i \mid \mathscr{H}_t) - \int_0^T \lambda_g (s \mid \mathscr{H}_s) \, ds + \sum_{i=1}^{N(T)} \log g (y_i \mid \mathscr{H}_t, t) \, . \tag{2.4}$$

Observe that in the case of independence between the ground process and the density of the marks, the log-likelihood could be split and estimated separately. Furthermore, in the case that our observations were i.i.d random variables, the number of exceedances over this threshold u should follow a Poisson process and, therefore, the intensity of the ground process $\lambda_g (t \mid \mathscr{H}_t)$ would be constant, while the marks are modeled by an unconditional GPD (for a detailed explanation see Herrera and Schipp, 2013).

The particular feature of the SEMPP approach is the representation of the conditional intensity for the ground process as a sum of contributions from all previous time events. In this paper, we consider two SEMPP models, the Hawkes-POT model introduced in Chavez-Demoulin et al. (2005)[4] and the ACD-POT model proposed by Herrera and Schipp (2013). Another interesting alternative, which we also wish to consider, is the so-called DPOT model introduced in Santos and Alves (2012). The DPOT is not directly a SEMPP but a one-dimensional point process with inter-exceedance time covariates, which introduce an autoregressive influence on the scale parameter of the marks.

2.1. *The Hawkes-POT model*

The Hawkes-POT model is obtained by parameterizing the intensity of the ground process $\lambda_g(t \mid \mathscr{H}_t)$ and the scale parameter $\beta\,(y \mid \mathscr{H}_t t)$ by means of a Hawkes process (Hawkes, 1971) as follows:

$$\lambda_g(t \mid \mathscr{H}_t) = k + \phi \sum_{i:t_i<t} \exp\{\delta y_i - \gamma(t - t_i)\} \tag{2.5}$$

and

$$\beta\,(y \mid \mathscr{H}_t, t) = \beta_0 + \eta \sum_{i:t_i<t} \exp\{\delta y_i - \gamma(t - t_i)\} \, ,$$

respectively. Under this parametrization all parameters are positive, k and β_0 represent the baseline rate of events, which in most applications is assumed to be constant in time, while ϕ and η are impact parameters related to the new extreme events. In other words, an extreme event increases

[4] See also Herrera and Schipp, 2009; Chavez-Demoulin and McGill, 2012 for new applications.

the chance of attaining other extreme events immediately afterward, then decreases exponentially to the baseline rate of events, displaying a monotone decreasing behavior.

2.2. The ACD-POT model

The second specification is the ACD-POT approach where the conditional intensity of the ground process is driven by a self-excited function that is updated at each extreme event of the process. These types of specifications were proposed by Herrera and Schipp (2013) in the context of EVT. This type of model is a mixture between the classical POT model and the autoregressive conditional duration (ACD) model (see Engle and Russell, 1998). The conditional intensity of the ground point process of exceedances for this approach depends only on a fixed number of the most recent inter-exceedance times $x_i = t_i - t_{i-1}$ and the expected conditional duration $\psi_i \equiv E\left[x_i \mid x_{i-1},\dots,x_1\right]$ of all information up to and including time t_{i-1}:

$$x_{N(t)} \quad = \quad \psi_{N(t)}\varepsilon_{N(t)},$$

where ε_i are i.i.d random variables. In particular, we consider a logarithmic ACD (Log-ACD) model, introduced by Bauwens and Giot (2000) in order to prevent ψ becoming negative, in which the autoregression bears on the logarithm of the conditional expected duration

$$\psi_{N(t)} = \exp\left\{w + a\log x_{N(t)-1} + b\log\psi_{N(t)-1}\right\}.$$

The ground process $\lambda_g(t \mid \mathscr{H}_t;\theta)$ for this type of model can be expressed as a multiplicative effect between the baseline hazard function $\lambda_0(\cdot)$ of the random variable ε and a shift given by the expected duration

$$\lambda_g(t \mid \mathscr{H}_t) = \lambda_0\left(\frac{t - t_{N(t)}}{\psi_{N(t)}}\right)\frac{1}{\psi_{N(t)}}. \tag{2.6}$$

In this paper, we propose the Burr distribution (see Grammig and Maurer, 2000) for the random variable ε, which displays a flexible non-monotonic hazard function taking a bathtub-shaped or inverted U-shaped form. The density function is defined by

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

Let $\lambda = 1$ and $\phi_i = \psi_i \frac{\gamma^{2\left(1+\frac{1}{k}\right)}\Gamma\left(\gamma^{-2}+1\right)}{\Gamma\left(1+\frac{1}{k}\right)\Gamma\left(\gamma^{-2}-\frac{1}{k}\right)}$, where $0 < \gamma^{-2} < k$. Then, the implied conditional intensity function for the ground process is given by

$$\lambda_g(t \mid \mathscr{H}_t) = \frac{k\phi_{N(t)}^{-k}\left(t-t_{N(t)}\right)^{k-1}}{1+\gamma^2\phi_{N(t)}^{-k}\left(t-t_{N(t)}\right)^{k}}. \tag{2.7}$$

In addition, for the scale parameter $\beta\,(y \mid \mathscr{H}_t, t)$, we consider a linear parametrization such that it depends on the history of the process, as in the Hawkes-POT model

$$\beta\,(y \mid \mathscr{H}_t, t) = \beta_0 + \beta_1\lambda_g(t \mid \mathscr{H}_t).$$

This type of parametrization has shown the best results in empirical studies (for example, see Herrera and Schipp, 2013; Herrera, 2013).

2.3. The Hawkes-DPOT and ACD-DPOT models

For the above models, we observe that one of the main ingredients is the scale modeling in the GPD. The economic interpretation for those specifications is that in a period of turmoil the scale parameter in the density of the marks is influenced by the temporal conditional intensity of the ground process; therefore, the estimated conditional GPD mean follows the path of the ground process.[5] Based on a similar idea, Santos and Alves (2012) propose characterizing the expected mean and variance of the marks by means of covariates of inter-exceedance times. In particular, they observed that short inter-exceedance times display a higher mean and variance than long inter-exceedance times, which in turn suggest an inverse relationship between the size of the marks and inter-exceedance times. For a general overview of EVT including the use of explanatory variables, see Coles (2001).

Once again, define $x_i = t_i - t_{i-1}$ as the most recent inter-exceedance time between two extreme events, with $x(t) = t - t_i$ being the backward recurrence time and $N(t)$ the counting process of exceedances at t. Notice that $x(t_i) = t_i - t_{i-1} = x_i$. Alternatively, the authors propose using the information of the last v inter-exceedance times,[6] that is, $x(t)_v = t - t_{N(t)-v}$, as covariates of the scale parameter as follows:

$$\beta\,(y \mid \mathscr{H}_t, t) = \frac{\beta_0}{\{x(t)_v\}^{\beta_1}},$$

where $\beta_0 > 0$ and $\beta_1 \geq 0$. Observe that they only model the marks while for the ground process they assume that this follows a Poisson process of exceedances with a constant rate given by the expected number of extreme events N_u divided by the size of the sample N, $\lambda_g = N_u/N$. We propose two other alternatives based on the DPOT model for the GPD of the marks but include the

[5]Indeed, the mean of the conditional GPD is given by $\left(\beta_0 + \eta\sum_{i:t_i<t}\exp\{\delta y_i - \gamma(t-t_i)\}\right)/(1-\xi)$ for the Hawkes-POT model and$\beta_0 + \beta_1\lambda_g(t \mid \mathscr{H}_t)/(1-\xi)$ for the ACD-POT model.

[6]Observe that $x(t)_v = x(t) + x_{N(t)-1} + \cdots + x_{N(t)-v+1} = t - t_{N(t)-v}$, and therefore, $x(t_i)_v = t_i - t_{i-v} = x_{i,v}$

dynamic behavior of the Hawkes and ACD approaches to model the inter-exceedance times in the ground process. The conditional intensity for these two alternatives is given by

$$\lambda\left(t,y \mid \mathscr{H}_t\right) = \lambda_g\left(t \mid \mathscr{H}_t\right) \frac{\{x(t)_v\}^{\beta_1}}{\beta_0} \left(1+\xi \frac{y-u}{\beta_0}\{x(t)_v\}^{\beta_1}\right)^{-1/\xi-1}, \tag{2.8}$$

where $\lambda_g\left(t \mid \mathscr{H}_t\right)$ is replaced here by (2.5) or (2.6) depending on whether we want to estimate the Hawkes-DPOT or the ACD-DPOT, respectively.[7]

3. Empirical analysis of the DAX index

3.1. Data description

Our data set consists of daily returns defined by $r_t = -100\log(p_t/p_{t-1})$, where p_t denotes the value of the DAX index at day t over a sample period from January 2, 1991 to January 18, 2008; on January 21, 2008, global stock markets suffered their biggest falls since September 11, 2001. Observe that we concentrate only on the losses. A second sample is used for backtesting the estimation of the different risk measures in the DAX index from January 21, 2008 to June 30, 2013. We update daily the new information that becomes available for the parameter estimates previously obtained. Thus, we dynamically adjust the models, which allows us to improve as accurately as possible the estimation of the risk measures.

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 exceedances have a GPD. In this paper we choose to work with 8% of the maxima of the sample. This threshold selection is based on the stability of the shape parameter, which influences directly the risk measures estimates (see Chavez-Demoulin and McGill, 2012; Herrera, 2013 for other alternatives). A detailed description of the methodology used can be found in Appendix A.

In order to better understand the empirical application, it is worth looking briefly at the basic characteristics of the extreme events that we want to analyze. In Table 1 in the Appendix, we find some descriptive statistics of the daily returns for the DAX index. The mean return is close to zero and it differs considerably in terms of standard deviation, skewness and kurtosis of a normally distributed random variable. In particular, the returns of the DAX index exhibit a high kurtosis. The assumption of normally distributed returns is strongly rejected through the Jarque-Bera test. Other assumptions, such as the null hypothesis that the returns series are i.i.d random variables as well as the returns have a unit root, are strongly rejected.

[7] Observe that for ease of exposition we assume that $\xi \neq 0$. The case where $\xi = 0$ can be obtained similarly.

<Insert Figure 1 about here>

Relating stylized facts of the most extreme events, Figure 1 displays some of the most important events for the DAX index during the in-sample period, for example, the Asian financial crisis in 1997, the collapse of the Long-Term Capital Management (LTCM) hedge fund in 1998, the dot.com Bubble in 2000 and the September 11, 2001 terror attacks, among others. On the right side of the top panel, we observe in detail the most extreme losses for this index (8% of the most important losses), whose clusters are evident around the year 2000. In particular, the left side of the bottom panel in Figure 1 shows strong evidence of an autocorrelation between the inter-exceedance times for the data analyzed. Another interesting stylized fact described by Santos and Alves (2012) is the apparent relationship between the size of the marks and the inverse of past inter-exceedance times. In our case, we found a strong positive correlation between them, in particular, this relationship (Pearson correlation 0.42) was stronger when we considered the whole inter-exceedance times preceding the last three events, $x_{i,3} = t_i - t_{i-3}$. Summarizing, stylized facts, such as clusters of extremes, dependence among inter-exceedance times and the size of marks, support the use of the proposed models.

3.2. Evaluation framework

We compare the models using the goodness of fit and a number of statistical accuracy tests for the VaR in-sample and in the backtest.

For the goodness of fit we employ

- W-statistic (Smith, 2003). This test assesses the success in modeling the temporal behavior of the exceedances. The W-statistic is defined by

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

  This statistic states that if the GPD parameter model is correct, the residuals are approximately independent unit exponential variables. For this reason, we test if the residuals are approximately independent by means of a Box-Ljung test (BL_W) and if they are unit exponential variables through a Kolmogorov-Smirnov test (KS_W).

For the statistical accuracy tests for the VaR we consider

- Unconditional coverage test (LR_{uc}). The idea is to test whether the fraction of violations obtained for a particular risk measure is significantly different from the theoretical one. A violation I_{t+1} of the VaR for the day $t+1$ is defined as occurring when the return r_{t+1} is higher than the VaR as follows:

$$I_{t+1} = \begin{cases} 1 & \text{if } r_{t+1} > VaR_{\alpha}^{t+1} \\ 0 & \text{if } r_{t+1} \leq VaR_{\alpha}^{t+1}. \end{cases}$$

- Test of independence between violations of the VaR (LR_{ind}). Under the null hypothesis, a violation today has no influence on the probability of a violation tomorrow. Christoffersen (1998) suggests this test of independence by modeling the number of violations as a binary first-order Markov chain.

- Conditional coverage test (LR_{cc}). This is a combination of the unconditional coverage test and the test of independence in order to test correct conditional coverage. For more details on the estimation of the LR_{uc}, LR_{ind} and LR_{cc} tests, we refer to Christoffersen (1998).

- Dynamic Quantile tests (DQ_{hit} and DQ_{VaR}). Engle and Manganelli (2004) propose examining whether the Hits on the VaR ($Hit_i = I_i - \alpha$) for the present period are uncorrelated with the above Hits and/or VaR estimates by means of a logit model.[8] In our approach for the first case, denoted by the DQ_{hit} the regressor vector contains only a lagged violation of the VaR

$$Hit_{t+1} = a + bHit_t + e_t,$$

while the second test, DQ_{VaR}, uses the contemporaneous VaR estimate

$$Hit_{t+1} = a + b_1 Hit_t + b_2 VaR_{\alpha}^{t} + e_t.$$

Under the null hypothesis, $H_0 = \mathbf{b} = 0$, the regressors should have no explanatory power.

3.3. *Empirical results in-sample for the DAX index*

Having investigated the characteristics of the financial series, we can now turn to a comparative analysis of the SEMPP proposed in the previous sections. We estimate all models proposed in Section 2, optimizing the log-likelihood function (2.4).[9]

We estimate eight models: one model each for Hawkes-POT and ACD-POT and two models for the DPOT, Hawkes-DPOT and ACD-DPOT approaches. According to Santos and Alves (2012)

[8]Observe that the sequence Hit_i is the de-meaned process on α associated with I_t.

[9]We use the optimx package in R (Nash and Varadhan, 2011), which allows for different strategies of optimization of the maximum likelihood.

and Santos et al. (2013), better results are obtained when we choose the parameter β_1 in the interval between 0.7 and 0.8 instead of estimating β_1 by means of the log-likelihood function. For this reason, we consider two models for DPOT, Hawkes-DPOT and ACD-DPOT, one with a fixed β_1 and the other estimated by maximizing the log-likelihood. In the empirical application, a $\beta_1 = 0.75$ exhibits better results according to the LR_{uc} test.

Results of the estimation are summarized in Table 2. Observe that the models are not directly comparable in terms of goodness of fit, especially the DPOT approaches. For this reason we compare them in terms of the performance in the estimation of the VaR and the goodness of fit for the GPD. However, we include the AIC and BIC statistics for completeness. Only the ACD-POT (DPOT) and Hawkes-POT(DPOT) are directly comparable. According to these results, the Hawkes-DPOT and the ACD-DPOT model exhibit the best fit.

<Insert Figure 2 about here>

Relating the statistical accuracy tests for the VaR, Figure 2 displays the results in-sample for the estimates of VaR at the 0.99 confidence level. At first glance, Figure 2 indicates that the VaR estimates are very similar while the violations are not found at the crisis period, which could indicate some kind of bias in the model. Table 3 reports the results for all tests at three confidence levels for the VaR: 0.95, 0.99 and 0.999. Entries in columns are the significance levels (p-values) of the respective tests, with the exception of level α and the number of violations at the VaR. We observe that the DPOT, Hawkes-DPOT and ACD-DPOT, estimated with parameter β_1 not constant, seem to be less variable through time in comparison with the rest of the models. However, the results of accuracy of the VaR estimates in Table 3 show that indeed two of these models, Hawkes-DPOT and ACD-DPOT, display the best performance with a total of 14 of 15 tests approved. The only model with similar performance is the standard ACD-POT. In relation to the goodness of fit of the GPD for the marks, the models where parameter β_1 was not estimated displayed a poor fit according to the W-statistic (BL_W and KS_W), in constrast to the results obtained by Santos and Alves (2012).

Summarizing, models with a ground process whose conditional intensity is characterized by a Hawkes or ACD and whose marks follow a DPOT approach display the best performance in-sample. Nevertheless, more important is the accuracy of the backtests produced by the proposed risk models. A systematic evaluation of the accuracy of the forecasts generated by these models is given in the next subsection.

3.4. *Backtesting the models*

To evaluate the performance of the VaR models, backtesting was carried out with the daily returns from January 21, 2008 to June 30, 2013. On each day during the backtesting, we fitted the eight models introduced above, then we reestimated the VaR daily for each return series according to (2.3) . Table 2 also reports the results on the VaR backtesting exercise.

An important year for the backtest period is 2008, which encompasses five of the worst trading days since the beginning of the DAX index in 1989. The first one is January 21 with -7.164%. The next three are October 6, 10 and 15, with a change in percentage of -7.073, -7.012 and -6.493, respectively. The last one was October 15 with a percent change of -6.838. Moreover, 36 extreme events of the whole backtesting sample are found in this year.

Figure 3 displays the results for the backtesting for a VaR with a confidence level 0.99. Overall, the performance of the models for at least the unconditional coverage seem to fit satisfactorily, especially for 2008, the year of the subprime crisis. Indeed, the number of violations for this year was never higher than three, with one, of course, in October 2008.

<Insert Figure 3 about here>

Deeper analysis of Table 2 shows that all models, even with a fixed parameter β_1, approve most tests for VaR accuracy. The most important difference is related to the results obtained for the dynamic quantile test DQ_{VaR} for the confidence level 0.95. Notice that for all models whose ground process were kept constant or had a Hawkes' type, we found some kind of autocorrelation between the violation and the most recent estimate for the VaR. However, according to the "traffic light" approach, the SEMPP models are all classified in the green zone (see Basel Committee on Banking Supervision, 2006).

4. Contrasting German and US stock markets

In view of the recent financial crisis, it seems to be clear that international linkages among financial institutions may explain contagion transmission cross country and the role of the US market as a leader among the world's stock markets (Mandilaras and Bird, 2007; Lee and Chang, 2013; Dimpfl and Peter, 2014; Yamamoto, 2014). Similarly, Germany is one of the most important and the largest economies in the euro zone.

Concerning linkages between the German and US stock markets, to the best of our knowledge there are few studies assessing the relative importance of contagion and interdependence between them. Baur and Jung (2006) investigate spillovers and correlations around the opening of both stock markets. Flad and Jung (2008) using high-frequency data identify a common trend shared by German and US stock markets and show that the US market is the driving force in the transatlantic system of stock indices. Bonfiglioli and Favero (2005) found evidence of short-term interdependence and contagion between both markets and show that the effect of fluctuations of the US stock market on the German stock market exhibits a non-linear behavior.

In this section we contrast our results obtained for the DAX index returns with two of the most important US stock market indices, the S&P 500 and Nasdaq, during the same period of study proposed for the German stock market, i.e., the in-sample estimation covers the period from January 2, 1991 to January 18, 2008, while a second sample is used for backtesting from January 21, 2008 to June 30, 2013.

The aim is to obtain more research evidence and an in-depth understanding of the cluster behavior of extreme events during the crisis period in both countries. Table 1 also includes summary statistics of the S&P 500 and Nasdaq returns. The data exhibit the usual stylized facts of stock market returns, in particular skewness and excess kurtosis. As expected, both US market indices reject the null hypothesis of normality.

As for DAX returns, we apply the eight models proposed, and results of the estimation are summarized in Table 2. In contrast to results obtained for the DAX returns, the S&P 500 and Nasdaq show a slight preference for the standard Hawkes-POT and ACD-POT models. As a result of the above, the estimations for both countries could not provide a marked preference for a model, indicating that the inclusion of inter-exceedance times as covariates is relevant for the models. However, the way in which this information is included or captured in the model (e.g., in terms of its intensity or duration) depends on the financial asset analyzed.

In order to shed light on the behavior of extreme events during the subprime crisis in the US market, we include the in-sample and backtesting estimations for the VaR. Tables 4 and 5 display the results. The in-sample results for the VaR estimates also confirm the Hawkes-POT and ACD-POT for S&P 500 returns but not for the Nasdaq returns, where according to the number of tests approved Hawkes-DPOT model seems to be preferred. Observe that the most crucial confidence level for the rest of the models is the 0.95 quantile, where the hypothesis tests of independence among the violations are mainly rejected. In line with these results, in the models where parameter β_1 was not estimated, the goodness of fit of the GPD for the marks showed the worst results. Finally, regarding the backtesting analysis performed in both US stock markets, the number of violations observed for all VaR confidence levels remained within the expected range complying, at the very least, with the Basel Committee recommendations.

5. Conclusions

We have illustrated how modern EVT can be used to model tail-related risk measures, such as VaR, applying it to negative daily log-returns on German and US stock markets during the recent crisis period. We propose an extension of the classic POT to model cluster behavior through the SEMPP for the inter-exceedance times and for the exceedance sizes. Maximum likelihood methods are used to calculate the parameters, where the self-exciting approach can follow eight different models.

In relation to the results for the DAX index, we observe that Hawkes-DPOT, ACD-DPOT and the simple ACD-POT display the best performance in-sample. This means that past inter-exceedance times can influence not only the frequency or intensity of how these extreme events occur but also the size of exceedances. The way the inter-exceedance times affect the size of exceedance was modeled through the scale parameter in the GPD by means of covariates. In the backtest, the results are impressive and almost superior to the in-sample, essentially due to the rapid adaption of models using the most recent information. A result that varies with the empirical application is the choice of working with the inter-exceedance times or with the conditional intensity of their ground processes.

In the case of the US stock market, the results confirm the Hawkes-POT and ACD-POT for S&P 500 returns but not for the Nasdaq returns, where the Hawkes-DPOT model seems to be preferred. Regarding the backtesting analysis performed in both US stock market indices during the crisis period, for all models the number of violations observed for all VaR confidence levels remained within the expected range, therefore complying with the Basel Committee recommendations.

Concerning the question whether the inter-exceedance times can contribute to the measurement accuracy of market risk in financial markets, our main conclusion is that inter-exceedance times play an important role in the statistical analysis of extreme events. Among the two possible strategies to incorporate these times, either as covariates to describe the size of the extreme events or in terms of their intensities, the estimations for both countries did not provide a clear preference for a specific model. Therefore, the way in which this information is included depends on the specific financial instrument investigated.

Three directions for future research emerge from the results. Being interested in long-term behavior rather than in short-term forecasting, the simulation of these models is possible for estimating risk measures with other time horizons. Alternatively, using a combination of these models to compare conservative and aggressive strategies for choosing between VaR models, as done by McAleer et al. (2013), may also be a useful risk management strategy. Finally, other flexible forms for the self-exciting function could be used incorporating other characteristics of the series, such as trends of increasing exceedances or different regimes, such as after shocks.

Acknowledgments

We thank the reviewers for their helpful comments. The usual disclaimer applies.

References

Allen, D. E., Singh, A. K., and Powell, R. J. (2013). EVT and tail-risk modelling: Evidence from market indices and volatility series . *The North American Journal of Economics and Finance*, 26:355 – 369.

Balkema, A. A. and De Haan, L. (1974). Residual life time at great age. *The Annals of Probability*, 5 (2):792 – 804.

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

Baur, D. and Jung, R. C. (2006). Return and volatility linkages between the us and the german stock market. *Journal of International Money and Finance*, 25(4):598 – 613.

Bauwens, L. and 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.

Bonfiglioli, A. and Favero, C. A. (2005). Explaining co-movements between stock markets: the case of US and Germany. *Journal of International Money and Finance*, 24(8):1299 – 1316.

Chavez-Demoulin, V., Davison, A., and McNeil, A. (2005). A point process approach to value-at-risk estimation. *Quantitative Finance*, 5 (2):227 – 234.

Chavez-Demoulin, V. and McGill, J. (2012). High-frequency financial data modeling using hawkes processes. *Journal of Banking & Finance*, 36:3415 – 3426.

Christoffersen, P. (1998). Evaluating interval forecasts. *International economic review*, 39 (4):841 – 862.

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

Daley, D. and Vere-Jones, D. (2003). *An Introduction to the Theory of Point Processes*. Springer Series in Statistics, New York.

de Jesus, R., Ortiz, E., and Cabello, A. (2013). Long run peso/dollar exchange rates and extreme value behavior: Value at risk modeling. *The North American Journal of Economics and Finance*, 24:139 – 152.

Dimpfl, T. and Peter, F. J. (2014). The impact of the financial crisis on transatlantic information flows: an intraday analysis. *Journal of International Financial Markets, Institutions and Money*, 31:1 – 13.

Engle, R. and Manganelli, S. (2004). Caviar. *Journal of Business and Economic Statistics*, 22(4):367 – 381.

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

Flad, M. and Jung, R. C. (2008). A common factor analysis for the us and the german stock markets during overlapping trading hours. *Journal of International Financial Markets, Institutions and Money*, 18(5):498 – 512.

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

Guillou, A. and Hall, P. (2001). A diagnostic for selecting the threshold in extreme value analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(2):293 – 305.

Hammoudeh, S., Santos, P. A., and Al-Hassan, A. (2013). Downside risk management and var-based optimal portfolios for precious metals, oil and stocks. *The North American Journal of Economics and Finance*, 25:318 – 334.

Hawkes, A. (1971). Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58:379 – 402.

Herrera, R. (2013). Energy risk management through self-exciting marked point process. *Energy Economics*, 38:64 – 76.

Herrera, R. and Schipp, B. (2009). Self-exciting extreme value models on stock market crashes. In Schipp, B. and Krämer, W., editors, *Statistical Inference, Econometric Analysis and Matrix Algebra*, pages 209 – 231. Physica.

Herrera, R. and Schipp, B. (2013). Value at risk forecasts by extreme value models in a conditional duration framework. *Journal of Empirical Finance*, 23:33 – 47.

Laurini, F. and Tawn, J. (2003). New estimators for the extremal index and other cluster characteristics. *Extremes*, 6:189 – 211.

Lee, H.-C. and Chang, S.-L. (2013). Spillovers of currency carry trade returns, market risk sentiment, and u.s. market returns. *The North American Journal of Economics and Finance*, 26:197 – 216.

Mandilaras, A. and Bird, G. (2007). Foreign exchange markets in south-east asia 1990- 2004: An empirical analysis of spillovers during crisis and non-crisis periods. *The North American Journal of Economics and Finance*, 18(1):41 – 57.

McAleer, M., Jimenez-Martin, J.-A., and Perez-Amaral, T. (2013). Has the basel accord improved risk management during the global financial crisis? *The North American Journal of Economics and Finance*, 26:250 – 265.

McNeil, A. and 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.

Nash, J. C. and Varadhan, R. (2011). Unifying optimization algorithms to aid software system users: optimx for r. *Journal of Statistical Software*, 43(9):1 – 14.

Pickands III, J. (1975). Statistical inference using extreme order statistics. *The Annals of Statistics*, pages 119 – 131.

Reiss, R.-D. and Thomas, M. (2007). *Statistical analysis of extreme values: with applications to insurance, finance, hydrology and other fields*. Birkhäuser Basel.

Santos, P. A., Alves, I. F., and Hammoudeh, S. (2013). High quantiles estimation with quasi-port and dpot: An application to value-at-risk for financial variables. *The North American Journal of Economics and Finance*, 26:487 – 496.

Santos, P. A. and Alves, M. F. (2012). Forecasting value-at-risk with a duration-based POT method. *Mathematics and Computers in Simulation*, 94:295 – 309.

Scarrott, C. and MacDonald, A. (2012). A review of extreme value threshold es-timation and uncertainty quantification. *REVSTAT–Statistical Journal*, 10(1):33 – 60.

Smith, R. L. (2003). *Statistics of extremes, with applications in environment, insurance and finance*.

Smith, R. L. and Weissman, I. (1994). Estimating the extremal index. *Journal of the Royal Statistical Society, Series B*, 56:515–528.

Tancredi, A., Anderson, C., and O'Hagan, A. (2006). Accounting for threshold uncertainty in extreme value estimation. *Extremes*, 9(2):87 – 106.

Yamamoto, S. (2014). Transmission of US financial and trade shocks to Asian economies: Implications for spillover of the 2007- 2009 US financial crisis. *The North American Journal of Economics and Finance*, 27:88 – 103.

Appendix A. Accounting for Threshold Uncertainty

Threshold uncertainty plays one of the most important roles in the utilization of EVT. The selection of the threshold level is not unique, and a number of approaches exist to this end (see Guillou and Hall, 2001; Tancredi et al., 2006; Scarrott and MacDonald, 2012 for a plethora of approaches). The basic idea of all these approaches is the same — optimize a trade-off between bias and variance. On one hand, a high threshold level will diminish the size of the sample and the variance parameter estimates will be high. On the other hand, a low threshold will enlarge the size of the sample, reducing the variance but inducing bias into parameter estimates, since we could be modeling the bulk of the sample instead of the tail of the distribution.

In this paper we follow the statistics proposed by Reiss and Thomas (2007) to determine the threshold level, or equivalently, the number of exceedances k

$$\arg\min_k f(k) = \frac{1}{k}\sum_{i=1}^{k} i^{\beta} \left| \hat{\xi}_i - \text{median}\left(\hat{\xi}_1, \ldots, \hat{\xi}_k\right) \right|,$$

where $\hat{\xi}_i$ is a shape parameter estimate for the sample fraction of the extremes above the upper order statistic i, and $\beta \in [0, 0.5]$ is a tuning parameter. This choice is motivated by the structural form of the SEMPP approach, where the only parameter that is not updated constantly is the shape parameter, and therefore, it should remain relatively invariant through different threshold levels and time.

In Figure 4 we display the results of this statistic for all stock market returns analyzed using the ACD-POT approach. The x-axis shows the number of exceedances from the 0.95 to the 0.90 quantile, while the y-axis exhibits the tuning parameter β. From top to bottom we observe the results for the DAX, S&P 500 and Nasdaq returns, respectively. The gray boxes show the interval where the shape parameter seems to be more stable for the entire spectrum of the tuning parameter. For the DAX returns this interval corresponds to the 0.925 - 0.915 quantile, while for the S&P 500 and Nasdaq returns these intervals correspond in both cases to 0.92 - 0.91. For ease of the exposition and to make the results comparable, we define the threshold levels for all the stock market returns analyzed at the 0.92 quantile.

Appendix B. Tables

	DAX	S&P 500	Nasdaq
N° Observations	5720	5688	5688
Std. dev	1.455	1.170	1.544
Minimum	- 9.871	-9.469	-10.168
Mean	0.032	0.028	0.039
Maximum	10.797	10.957	13.254
Kurtosis	4.671	8.691	5.856
Skewness	-0.099	-0.238	-0.077
Jarque-Bera test	5209.15 (0)	17976.35 (0)	8144.39 (0)
Phillips-Perron Unit Root Test	-17.723 (0.01)	-18.221 (0)	-17.199 (0)

Table 1: Descriptive statistics of daily log-returns for the DAX, S&P 500 and Nasdaq indices. p-values are in parentheses.

Models	μ	ψ	a	b	γ	δ	ξ	β_0	β_1	Log-like	AIC	BIC
					DAX Index							
Hawkes-POT	0.029	0.023			0.060	0.367	0.021	0.451	0.132	-1445.245	2904.490	2949.061
	(0.005)	(0.005)			(0.010)	(0.049)	(0.009)	(0.050)	(0.026)			
ACD-POT	0.129		0.034	0.733	1.175	1.592	0.048	0.426	4.171	-1456.153	2928.306	2979.244
	(0.082)		(0.027)	(0.059)	(0.361)	(0.202)	(0.043)	(0.075)	(0.670)			
DPOT							0.084	2.491	0.298	-264.554	535.108	554.210
							(0.056)	(0.478)	(0.058)			
Hawkes-DPOT	0.015	0.029			0.045	0.121	0.084	2.491	0.298	-1150.932	2315.864	2360.435
	(0.003)	(0.008)			(0.011)	(0.089)	(0.056)	(0.478)	(0.058)			
ACD-DPOT	0.257		0.049	0.626	1.034	1.453	0.084	2.491	0.298	-1151.509	2319.018	2369.956
	(0.207)		(0.049)	(0.105)	(0.398)	(0.206)	(0.056)	(0.477)	(0.058)			
DPOT ($\beta_1 = 0.75$)							0.238	8.593	0.75	-367.1289	740.258	759.356
							(0.057)	(0.668)				
Hawkes-DPOT ($\beta_1 = 0.75$)	0.015	0.029			0.045	0.121	0.238	8.593	0.75	-1582.745	3179.490	3224.061
	(0.005)	(0.006)			(0.009)	(0.086)	(0.057)	(0.669)				
ACD-DPOT ($\beta_1 = 0.75$)	0.068		0.010	0.769	1.620	1.850	0.238	8.593	0.75	-1480.023	2976.046	3026.984
	(0.103)		(0.017)	(0.111)	(0.470)	(0.267)	(0.057)	(0.668)				
					S&P 500 Index							
Hawkes-POT	0.021	0.018			0.036	0.459	0.001	0.342	0.054	-1304.995	2623.990	2668.551
	(0.005)	(0.004)			(0.008)	(0.056)	(0.001)	(0.041)	(0.014)			
ACD-POT	0.217		0.066	0.707	0.939	1.559	0.075	0.364	1.934	-1306.927	2629.854	2680.781
	(0.096)		(0.028)	(0.054)	(0.244)	(0.152)	(0.049)	(0.063)	(0.487)			
DPOT							0.067	1.413	0.263	-203.074	412.148	431.246
							(0.049)	(0.267)	(0.058)			
Hawkes-DPOT	0.014	0.014			0.020	0.240	0.067	1.413	0.263	-1315.699	2645.398	2689.959
	(0.004)	(0.004)			(0.005)	(0.136)	(0.049)	(0.266)	(0.058)			
ACD-DPOT	0.212		0.061	0.737	0.940	1.603	0.067	1.413	0.263	-1306.433	2628.866	2679.793
	(0.096)		(0.026)	(0.055)	(0.240)	(0.155)	(0.049)	(0.267)	(0.058)			
DPOT ($\beta_1 = 0.75$)							0.248	5.858	0.75	-232.817	471.634	490.73
							(0.0589)	(0.461)				
Hawkes-DPOT ($\beta_1 = 0.75$)	0.014	0.014			0.020	0.240	0.248	5.858	0.75	-1345.442	2704.884	2749.445
	(0.004)	(0.004)			(0.005)	(0.136)	(0.059)	(0.461)				
ACD-DPOT ($\beta_1 = 0.75$)	0.212		0.061	0.737	0.940	1.603	0.248	5.858	0.75	-1336.177	2688.354	2739.281
	(0.096)		(0.026)	(0.055)	(0.240)	(0.155)	(0.059)	(0.462)				
					Nasdaq Index							
Hawkes-POT	0.021	0.016			0.032	0.308	0.001	0.442	0.064	-1691.183	3396.366	3440.927
	(0.004)	(0.004)			(0.007)	(0.043)	(0.002)	(0.053)	(0.017)			
ACD-POT	0.065		0.021	0.767	1.689	2.102	0.006	0.522	2.910	-1693.214	3402.428	3453.355
	(0.046)		(0.020)	(0.041)	(0.544)	(0.334)	(0.042)	(0.080)	(0.448)			
DPOT							0.012	2.645	0.325	-461.037	928.074	947.172
							(0.041)	(0.360)	(0.048)			
Hawkes-DPOT	0.016	0.014			0.021	0.209	0.012	2.647	0.325	-1712.441	3438.882	3483.443
	(0.004)	(0.004)			(0.005)	(0.076)	(0.041)	(0.360)	(0.048)			
ACD-DPOT	0.034		0.008	0.798	2.161	2.413	0.012	2.647	0.325	-1696.316	3408.632	3459.559
	(0.045)		(0.012)	(0.052)	(0.588)	(0.349)	(0.041)	(0.360)	(0.048)			
DPOT ($\beta_1 = 0.75$)							0.185	7.868	0.75	-493.332	992.664	1011.762
							(0.046)	(0.521)				
Hawkes-DPOT ($\beta_1 = 0.75$)	0.016	0.014			0.021	0.209	0.185	7.868	0.75	-1744.736	3503.472	3548.033
	(0.004)	(0.004)			(0.005)	(0.076)	(0.046)	(0.521)				
ACD-DPOT ($\beta_1 = 0.75$)	0.034		0.008	0.798	2.161	2.413	0.185	7.868	0.75	-1728.611	3473.222	3524.149
	(0.045)		(0.012)	(0.052)	(0.588)	(0.349)	(0.046)	(0.521)				

Table 2: Results for the DAX, S&P500 and Nasdaq stock market returns. Standard deviations are shown in parentheses. Log-like are the results of the maximization of the log-likelihood estimation, while AIC and BIC are the Akaike information criterion and the Bayesian information criterion, respectively.

Models	GoF POT			VaR in-sample						VaR backtest					
	BL_W	KS_W	α	Viol.	LR_{uc}	LR_{ind}	LR_{cc}	DQ_{hit}	DQ_{VaR}	Viol.	LR_{uc}	LR_{ind}	LR_{cc}	DQ_{hit}	DQ_{VaR}
Hawkes-POT	0.43	0.06	0.95	204	0.43	0.65	0.66	0.66	0.78	77	0.39	0.95	0.69	0.93	0.99
			0.99	41	0.76	0.01	0.02	0.01	0.01	16	0.60	0.56	0.73	0.54	0.61
			0.999	3	0.51	0.95	0.80	0.95	0.56	0	0.09	1.00	0.25	-	-
ACD-POT	0.29	0.50	0.95	228	0.37	0.98	0.67	0.98	095	83	0.12	0.95	0.30	0.95	0.94
			0.99	43	1.00	0.08	0.21	0.08	0.00	17	0.43	0.52	0.60	0.52	0.80
			0.999	6	0.44	0.90	0.74	0.90	0.75	2	0.63	0.94	0.89	0.94	1.00
DPOT	0.06	0.10	0.95	233	0.22	0.00	0.00	0.00	0.00	68	0.81	0.34	0.62	0.36	0.00
			0.99	56	0.06	0.04	0.02	0.04	0.00	9	0.15	0.73	0.34	0.73	0.17
			0.999	3	0.51	0.95	0.80	0.95	0.74	0	0.09	1.00	0.25	-	-
Hawkes-DPOT	0.06	0.10	0.95	215	0.99	0.93	1.00	0.94	0.00	69	0.91	0.38	0.67	0.39	0.00
			0.99	38	0.43	0.05	0.10	0.05	0.14	9	0.15	0.73	0.34	0.73	0.17
			0.999	6	0.44	0.90	0.74	0.90	0.17	0	0.09	1.00	0.25	-	-
ACD-DPOT	0.06	0.10	0.95	223	0.59	0.31	0.51	0.32	0.00	77	0.39	0.92	0.69	0.93	0.94
			0.99	49	0.37	0.13	0.21	0.13	0.19	11	0.40	0.68	0.65	0.68	0.69
			0.999	6	0.44	0.90	0.74	0.90	0.09	0	0.09	1.00	0.25	-	-
DPOT ($\beta_1 = 0.75$)	0.00	0.00	0.95	246	0.03	0.00	0.00	0.00	0.00	74	0.62	0.28	0.50	0.30	0.00
			0.99	56	0.06	0.04	0.02	0.04	0.01	21	0.08	0.42	0.16	0.43	0.34
			0.999	8	0.11	0.86	0.28	0.86	0.06	1	0.72	0.97	0.94	0.97	0.43
Hawkes-DPOT ($\beta_1 = 0.75$)	0.00	0.00	0.95	246	0.03	0.00	0.00	0.00	0.00	74	0.62	0.28	0.50	0.30	0.00
			0.99	56	0.06	0.04	0.02	0.04	0.01	21	0.08	0.42	0.16	0.43	0.34
			0.999	8	0.11	0.86	0.28	0.86	0.06	1	0.72	0.97	0.94	0.97	0.43
ACD-DPOT ($\beta_1 = 0.75$)	0.00	0.00	0.95	187	0.04	0.51	0.11	0.52	0.04	80	0.23	0.81	0.47	0.82	0.61
			0.99	55	0.08	0.01	0.00	0.01	0.00	21	0.08	0.32	0.13	0.32	0.54
			0.999	14	0.00	0.04	0.00	0.04	0.00	2	0.63	0.94	0.89	0.94	0.24

Table 3: Summary of different statistics for comparing the models using the goodness of fit, and a number of statistical accuracy tests for the VaR in-sample and in the backtest for the DAX index. Entries in the columns are the significance levels (p-values) of the respective tests, with the exception of the α level and the number of violations at the VaR (Viol.). The cells with “-” values mean that the test cannot be estimated. Number of observations in the in-sample period is 4304. Number of observations in the backtest period is 1400.

Models	GoF POT			VaR in-sample						VaR backtest					
	BL_W	KS_W	α	Viol.	LR_{uc}	LR_{ind}	LR_{cc}	DQ_{hit}	DQ_{VaR}	Viol.	LR_{uc}	LR_{ind}	LR_{cc}	DQ_{hit}	DQ_{VaR}
Hawkes-POT	0.86	0.39	0.95	194	0.14	0.08	0.07	0.09	0.21	62	0.44	0.61	0.65	0.62	0.82
			0.99	37	0.35	0.42	0.47	0.42	0.34	13	0.86	0.63	0.88	0.63	0.07
			0.999	6	0.44	0.90	0.73	0.90	0.25	0	0.10	1	0.26	-	-
ACD-POT	0.73	0.51	0.95	213	0.90	0.10	0.25	0.11	0.25	67	0.76	0.44	0.71	0.45	0.72
			0.99	40	0.65	0.39	0.62	0.39	0.24	17	0.42	0.52	0.58	0.52	0.8
			0.999	8	0.11	0.87	0.28	0.86	0.64	2	0.63	0.94	0.89	0.94	1
DPOT	0.74	0.16	0.95	230	0.29	0	0	0	0	73	0.67	0.28	0.51	0.29	0
			0.99	45	0.76	0.09	0.23	0.09	0	5	0.01	0.85	0.02	0.85	0.01
			0.999	6	0.44	0.90	0.73	0.90	0.64	0	0.10	1	0.25	-	-
Hawkes-DPOT	0.74	0.16	0.95	193	0.12	0.02	0.02	0.02	0	71	0.85	0.23	0.47	0.24	0
			0.99	37	0.35	0.42	0.47	0.43	0.73	4	0	0.88	0.01	0.88	0.05
			0.999	8	0.11	0.86	0.28	0.86	0.99	0	0.10	1	0.25	-	-
ACD-DPOT	0.74	0.16	0.95	202	0.36	0.02	0.04	0.02	0.01	59	0.19	0.73	0.39	0.74	0.49
			0.99	42	0.88	0.36	0.65	0.36	0.35	14	0.98	0.59	0.87	0.6	0.74
			0.999	8	0.11	0.86	0.28	0.86	0.92	0	0.10	1	0.25	-	-
DPOT ($\beta_1 = 0.75$)	0.01	0	0.95	243	0.05	0	0	0	0	72	0.76	0.25	0.49	0.26	0
			0.99	45	0.76	0.50	0.76	0.50	0.77	9	0.16	0.73	0.35	0.73	0.65
			0.999	8	0.11	0.86	0.28	0.86	0.94	0	0.10	1	0.25	-	-
Hawkes-DPOT ($\beta_1 = 0.75$)	0.01	0	0.95	175	0	0.29	0.01	0.30	0.05	72	0.76	0.25	0.49	0.26	0
			0.99	60	0.01	0.06	0.01	0.06	0.02	9	0.16	0.73	0.35	0.73	0.65
			0.999	11	0.01	0.81	0.03	0.81	0.02	0	0.10	1	0.25	-	-
ACD-DPOT ($\beta_1 = 0.75$)	0.01	0	0.95	171	0	0.23	0	0.24	0.29	55	0.06	0.9	0.18	0.9	0.48
			0.99	64	0	0.34	0.01	0.34	0.01	15	0.77	0.57	0.81	0.57	0.75
			0.999	13	0	0.78	0	0.78	0.13	2	0.63	0.94	0.89	0.94	0.77

Table 4: Summary of different statistics for comparing the models using the goodness of fit, and a number of statistical accuracy tests for the VaR in-sample and in the backtest for the S&P 500 index. Entries in the columns are the significance levels (p-values) of the respective tests, with the exception of the α level and the number of violations at the VaR (Viol.). The cells with “-” values mean that the test cannot be estimated. Number of observations in the in-sample period is 4298. Number of observations in the backtest period is 1390.

Models	GoF POT			VaR in-sample						VaR backtest					
	BL_W	KS_W	α	Viol.	LR_{uc}	LR_{ind}	LR_{cc}	DQ_{hit}	DQ_{VaR}	Viol.	LR_{uc}	LR_{ind}	LR_{cc}	DQ_{hit}	DQ_{VaR}
Hawkes-POT	0.89	0.12	0.95	195	0.18	0.17	0.16	0.18	0.18	69	0.39	0.67	0.63	0.67	0.88
			0.99	40	0.67	0.01	0.02	0.01	0.02	10	0.47	0.07	0.14	0.07	0.17
			0.999	7	0.23	0.88	0.48	0.88	0.65	0	0.11	1	0.29	-	-
ACD-POT	0.03	0.60	0.95	186	0.04	0.02	0.01	0.02	0.03	83	0.11	0.64	0.24	0.65	0.09
			0.99	42	0.88	0.07	0.19	0.07	0.19	12	0.6	0.09	0.21	0.09	0.12
			0.999	4	0.88	0.93	0.99	0.93	0.30	4	0.07	0.88	0.20	0.88	0.97
DPOT	0.08	0.42	0.95	222	0.62	0	0	0	0	70	0.95	0.8	0.96	0.80	0
			0.99	55	0.08	0.01	0	0.01	0	16	0.58	0.18	0.34	0.18	0.01
			0.999	6	0.44	0.90	0.73	0.90	0.35	3	0.24	0.91	0.49	0.91	0.67
Hawkes-DPOT	0.08	0.42	0.95	190	0.08	0.12	0.06	0.13	0	63	0.42	0.93	0.72	0.93	0
			0.99	41	0.76	0.06	0.17	0.07	0.01	15	0.77	0.15	0.34	0.15	0.01
			0.999	5	0.74	0.91	0.94	0.91	0.78	3	0.24	0.91	0.49	0.91	0.67
ACD-DPOT	0.08	0.42	0.95	194	0.14	0	0	0	0	76	0.43	0.93	0.73	0.94	0.97
			0.99	52	0.18	0.03	0.04	0.03	0.05	15	0.77	0.15	0.34	0.15	0.36
			0.999	6	0.44	0.90	0.73	0.90	0.97	2	0.63	0.94	0.89	0.94	0.99
DPOT ($\beta_1 = 0.75$)	0.01	0.02	0.95	235	0.16	0	0	0	0	79	0.25	0.80	0.50	0.81	0
			0.99	64	0	0.02	0	0.02	0	17	0.42	0.20	0.32	0.20	0.08
			0.999	9	0.05	0.85	0.14	0.85	0.92	3	0.24	0.91	0.49	0.91	0.97
Hawkes-DPOT ($\beta_1 = 0.75$)	0.01	0.02	0.95	174	0	0.07	0	0.08	0.01	64	0.49	0.54	0.66	0.55	0.25
			0.99	53	0.14	0	0.01	0	0.02	16	0.58	0.18	0.34	0.18	0.40
			0.999	10	0.02	0.83	0.06	0.83	0.29	3	0.24	0.91	0.49	0.91	0.41
ACD-DPOT ($\beta_1 = 0.75$)	0.01	0.02	0.95	176	0.01	0.01	0	0.01	0	63	0.42	0.50	0.57	0.51	0.15
			0.99	56	0.06	0.04	0.02	0.04	0.13	16	0.58	0.18	0.34	0.18	0.38
			0.999	10	0.02	0.83	0.06	0.83	0.75	3	0.24	0.91	0.49	0.91	0.96

Table 5: Summary of different statistics for comparing the models using the goodness of fit, and a number of statistical accuracy tests for the VaR in-sample and in the backtest for the Nasdaq index. Entries in the columns are the significance levels (p-values) of the respective tests, with the exception of the α level and the number of violations at the VaR (Viol.). The cells with “-” values mean that the test cannot be estimated. Number of observations in the in-sample period is 4298. Number of observations in the backtest period is 1390.

Appendix C. Figures

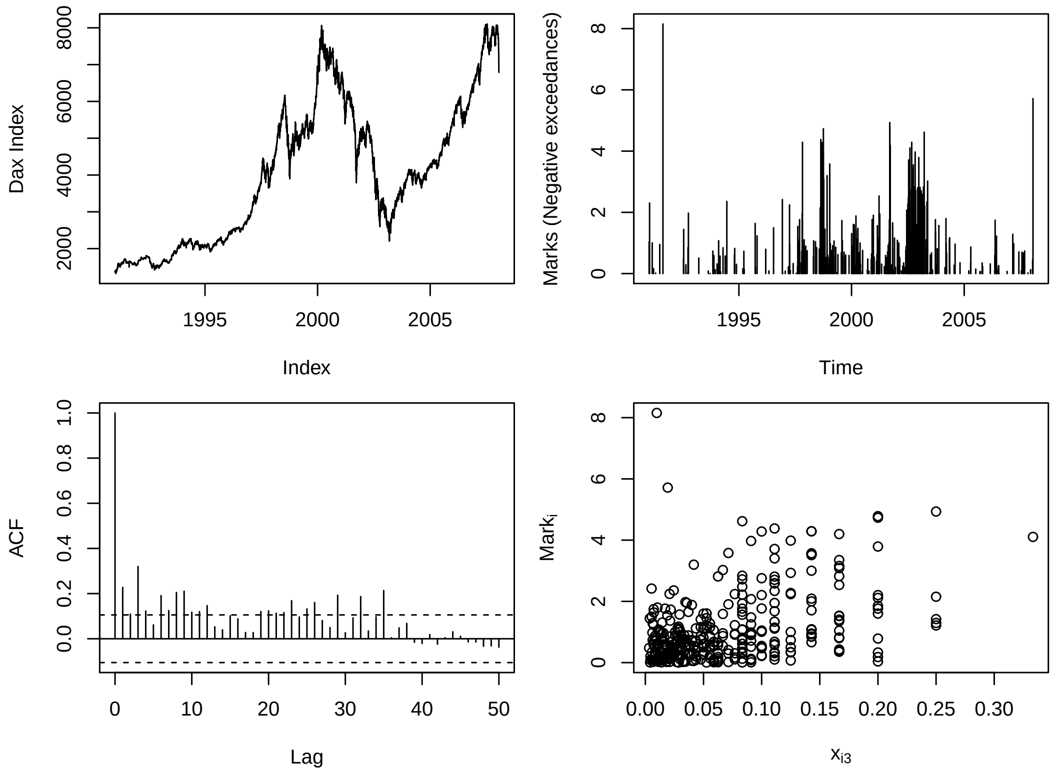

Figure 1: From top to bottom and from left to right, we observe the DAX index from January 2, 1991 to January 18, 2008, the most important losses for the returns of this index, the autocorrelation function for the inter-exceedance times of these losses, and a scatter plot of marks and inter-exceedance times preceding the last three events (i.e., $x_{i,3} = t_i - t_{i-3}$).

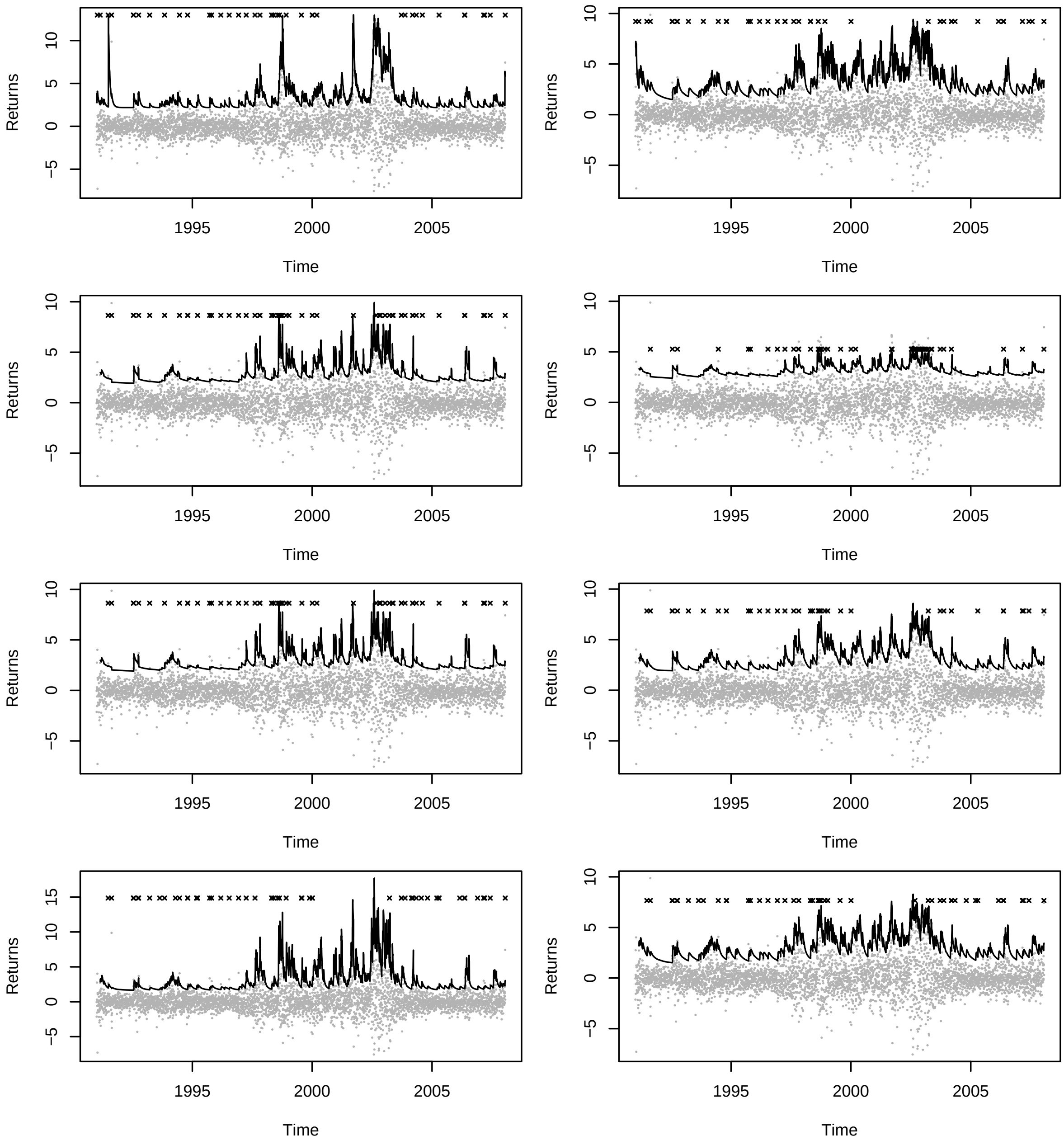

Figure 2: VaR estimation in-sample at the 0.99 confidence level for the eight models fitted to the DAX returns. The black lines are the VaR estimates, while $\times$ are the violations. From top to bottom and from left to right, the models are: Hawkes-POT, ACD-POT, DPOT ($\beta_1 = 0.75$), DPOT, Hawkes-DPOT ($\beta_1 = 0.75$), ACD-DPOT ($\beta_1 = 0.75$), Hawkes-DPOT and ACD-DPOT.

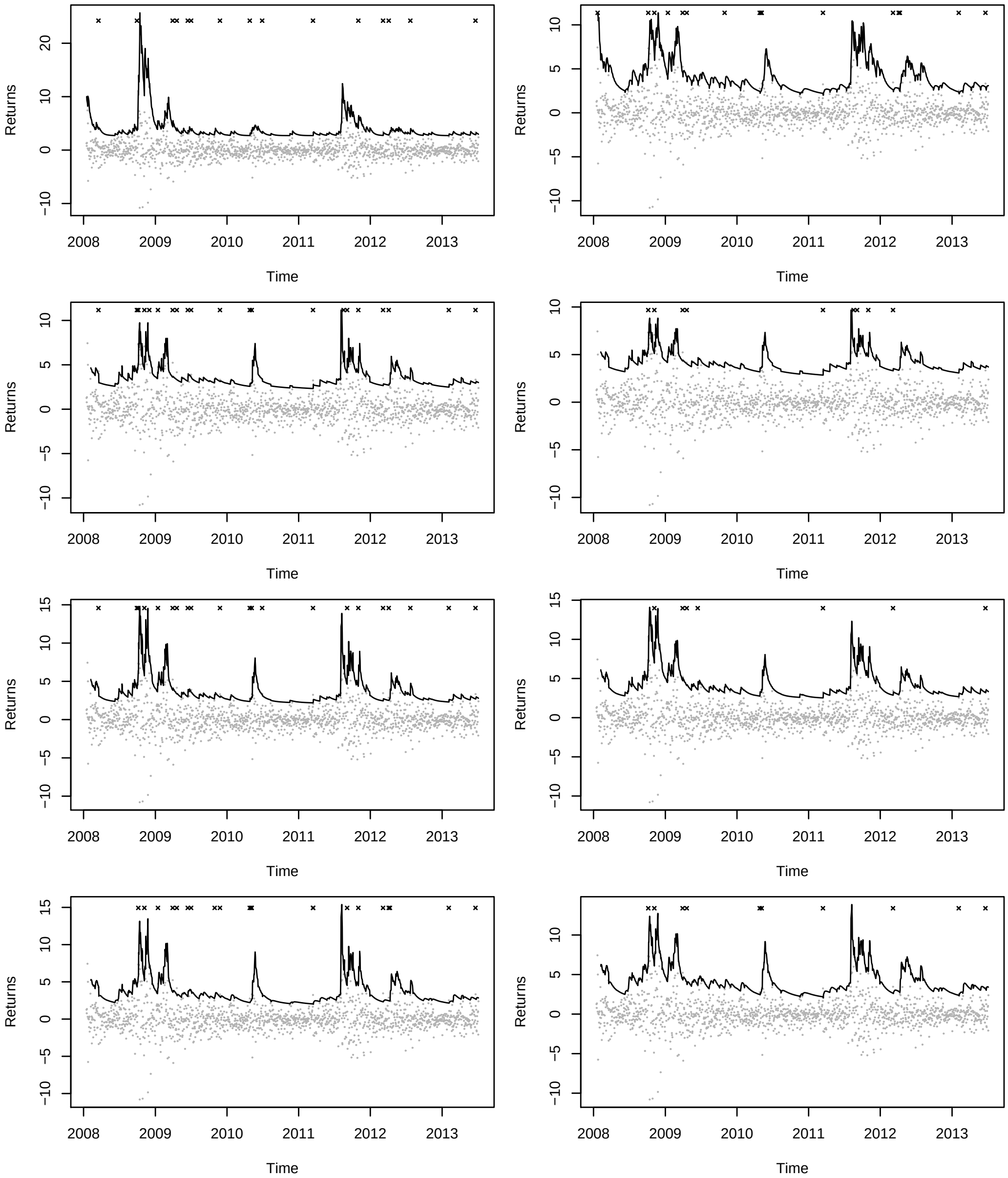

Figure 3: VaR estimation for the backtesting at the 0.99 confidence level for the eight models fitted to the DAX returns. The black lines are the VaR estimates, while $\times$ are the violations. From top to bottom and from left to right, the models are: Hawkes-POT, ACD-POT, DPOT ($\beta_1 = 0.75$), DPOT, Hawkes-DPOT ($\beta_1 = 0.75$), ACD-DPOT ($\beta_1 = 0.75$), Hawkes-DPOT and ACD-DPOT.

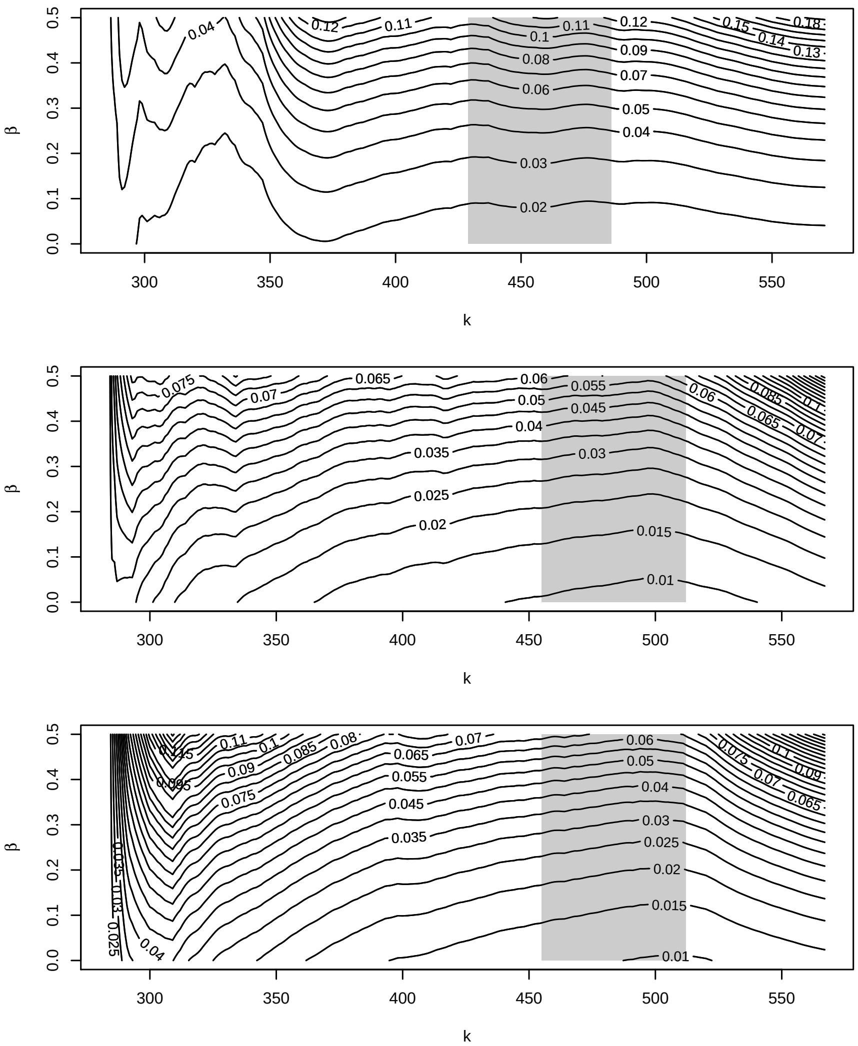

Figure 4: Results of the statistics proposed by Reiss and Thomas (2007) to determine the threshold level for the returns analyzed. From top to bottom the results for the DAX, S&P 500 and Nasdaq returns, respectively. The gray boxes show the interval where the shape parameter seems to be more stable for the entire spectrum of the tuning parameter. For the DAX returns this interval corresponds to the 0.925 - 0.915 quantile, while for the S&P 500 and Nasdaq returns these intervals correspond in both cases to the 0.92 - 0.91 quantile.