Multivariate Dynamic Intensity Peaks-Over-Threshold Models

Nikolaus Hautsch* and Rodrigo Herrera†

August 1, 2019

Abstract

We propose a multivariate dynamic intensity peaks-over-threshold model to capture extremes in multivariate return processes. The random occurrence of extremes is modeled by a multivariate dynamic intensity model, while temporal clustering of their size is captured by an autoregressive multiplicative error model. Applying the model to daily returns of three major stock indexes yields strong empirical support for a temporal clustering of both the occurrence and the size of extremes. Backtesting Value-at-Risk and Expected Shortfall forecasts shows that the consideration of clustering effects and of feedback between the magnitudes and the intensity of extremes results in better forecasts of risk.

Keywords: Extreme value theory, Value-at-Risk, Expected shortfall, Marked point process, Risk management.

JEL Codes: C58, C32, C53

*Corresponding author. Department of Statistics and Operations Research, University of Vienna, Research Platform "Data Science @ Uni Vienna", Vienna Graduate School of Finance, and Center for Financial Studies, Frankfurt. Hautsch acknowledged research support by the Wiener Wissenschafts-, Forschungs- und Technologiefonds (WWTF). Email: nikolaus.hautsch@univie.ac.at. Address: Oskar-Morgenstern-Platz 1, A-1090 Vienna. Tel: +43-1-4277-38680.

†Facultad de Economía y Negocios, Universidad de Talca. Herrera acknowledges the Chilean CONICYT funding agency for financial support (FONDECYT 1180672) for this project.

1 Introduction

Financial risk management has become an ubiquitous task for banks, companies, and financial institutions, especially during the last subprime mortgage crisis. The global crisis of 2008 has demonstrated the importance of modeling and forecasting of extreme events and their dynamic behavior during crisis periods. Classical extreme value theory (EVT) constitutes the mathematical and statistical foundation for the description of the distribution of extreme events. Traditional concepts to describe the tail of a loss distribution are Value-at-Risk (VaR) and Expected Shortfall (ES), see, e.g., McNeil and Frey (2000), Cotter and Dowd (2006) or Chavez-Demoulin et al. (2014). On the other hand, point process methods allow the dynamic behavior of (extreme) events to be captured and are typically applied in the context of portfolio credit risk, market microstructure analysis, contagion analysis, or jump-diffusion models, see, e.g. Engle and Russell (1998), Bauwens and Hautsch (2006), Errais et al. (2010), Bacry and Muzy (2014) or Aït-Sahalia et al. (2015). Moreover, point process theory provides an elegant formulation for the characterization of the limiting distribution of extreme value distributions (see Pickands, 1971 and Smith, 1989) and therefore, builds a natural complementary framework to extreme value analysis.

In this paper, we aim at bringing together both branches of the literature and propose a dynamic multivariate model capturing the occurrence and size of extremes in a multivariate time series. Important features of the proposed framework are to allow for (i) temporal clustering of both the occurrence of extremes and the size thereof, (ii) cross-sectional feedback between individual exceedance intensities, and (iii) feedback between the magnitude of exceedances and their intensity. On the one hand, we introduce an autoregressive conditional intensity peaks-over-threshold (ACI-POT) model, which, in its most basic form, corresponds to the combination of two known models: the ACI model introduced by Russell (1999) and the POT model by Davison and Smith (1990). Moreover, we propose a multivariate extension of a Hawkes-POT model, introduced in a univariate context by Chavez-Demoulin et al. (2005) and recently reviewed in different applications by Chavez-Demoulin and McGill (2012), Herrera and Schipp (2014), and Gresnigt et al. (2015).

Our approach complements the multivariate Hawkes-POT model proposed by Grothe et al. (2014). They treat the multivariate process of extreme events as a superposition of individual self-exciting Hawkes processes, whose exceedance times and exceedance magnitudes are coupled

together using an extreme value copula. The form of the latter is implied by specific parametric assumptions for the decay function and the impact function, capturing the effect of the marks of exceedances onto the conditional rate of future exceedances. Though the resulting model yields a unifying and parsimonious framework for the modeling of multivariate extremes, its drawback, however, is that the univariate processes do not allow for dynamic spillovers and that dynamic interdependence result from the copula only. Consequently, the multivariate point processes are exclusively coupled through *contemporaneous* effects but do not allow for spillovers across time and processes, making its use in the context of risk forecasting a difficult task.

The major difference in our setting is to explicitly allow for dynamic spillovers across the processes of extreme events. The key element of our framework is a feedback mechanism, accommodating temporal and cross-sectional interdependence between the arrival rates of extreme events and their magnitudes, thus allowing not only for self-excitation but also *cross-excitation*. The proposed class of processes generates a flexible and computationally tractable multivariate dependence structure, properties that in the last years have been empirically well-documented in other context by Bowsher (2007), Bacry et al. (2012), Bacry et al. (2013), Aït-Sahalia et al. (2014), and Aït-Sahalia et al. (2015), among others. Accordingly, our framework allows us to analyze how the effect of the occurrence of an extreme event is traced through the system and dynamically affects the other processes.

A further contribution, from an empirical perspective, is to identify and discuss typical features of cluster behavior of extreme events in financial markets. In our empirical application, we consider three major stock market indexes, DAX, S&P500, and FTSE100. By means of the multivariate ACI-POT approaches, we show that we can well capture these stylized facts and can produce reliable forecasts of VaR and ES. Corresponding results for Hawkes-POT approaches are presented in a Supplementary Appendix.

The remainder of the paper is organized as follows. In Section 2, we discuss some stylized facts associated with cluster behavior of extreme events in financial time series. Section 3 summarizes the concepts in EVT from the viewpoint of point process theory and introduces the proposed dynamic intensity POT models. In Section 4, we illustrate how to apply the proposed models to produce conditional risk measures such as Value-at-Risk and Expected Shortfall. Section 5 discusses estimation results and diagnostics which are based on applications of the proposed models to the daily returns of international stock indexes. Section 6 provides VaR and ES in-sample and

out-of-sample backtesting results. Finally, Section 7 concludes.

2 Clustering of Extreme Events

The clustering of extreme events is recognized as a prevalent feature in financial time series. To illustrate this stylized fact, we consider daily data of an equal-weighted portfolio based on the DAX, S&P500, and FTSE100 indices over the period January 3, 1994, to December 30, 2014. A flexible non-parametric tool for capturing different types of extremal dependence is the extremogram introduced by Davis et al. (2009), which can be considered as an analog of the autocorrelation function for extreme events. Let X_t be a strictly stationary $\mathbb{R}^d$-valued time series. Then, the extremogram for lag h is defined by

$$\rho_{AB}(h) = \lim_{x\to\infty} Pr\left(x^{-1}X_h \in A \mid x^{-1}X_0 \in B\right),$$

for $h = 0, 1, 2, \ldots$, provided that the limit exists for two sets A and B and are bounded away from 0.[1] Similarly, we can define the cross-extremogram as

$$\phi_{AB}(h) = \lim_{x\to\infty} Pr\left(x^{-1}Y_h \in A \mid x^{-1}X_0 \in B\right),$$

which can be straightforwardly extended to higher dimensions. In practice, the limits for x are typically replaced by corresponding quantiles of the processes.

Figure 1 displays extremograms and cross-extremograms with x chosen as the 91.5% empirical quantile of the portfolio returns.[2] We observe that the (cross-) extremograms decay hyperbolically as lags increase. The extremogram of losses and the cross-extremograms for losses conditional on gains reveal the most significant dependence on higher lags. Hence, losses are more strongly clustered than gains, confirming findings (e.g., by Hamidieh et al., 2009; Olmo, 2005; Jondeau and Rockinger, 2003.

Moreover, we also observe temporal clustering of extremes *across* markets. The two first

[1] Usually, in univariate time series, the choice of the sets is defined by $A = B = [1, \infty)$, and thus, the extremogram corresponds to the upper tail dependence coefficient between X_0 and X_h.

[2] For all (cross-) extremograms shown in this paper, we utilize a stationary bootstrap based on $10,000$ bootstrap replications to construct confidence intervals with block sizes given by an independent geometric distribution with mean 250 (which closely corresponds to the number of yearly trading days). For details on the estimation and construction of confidence intervals for extremograms, we refer to Davis et al. (2012).

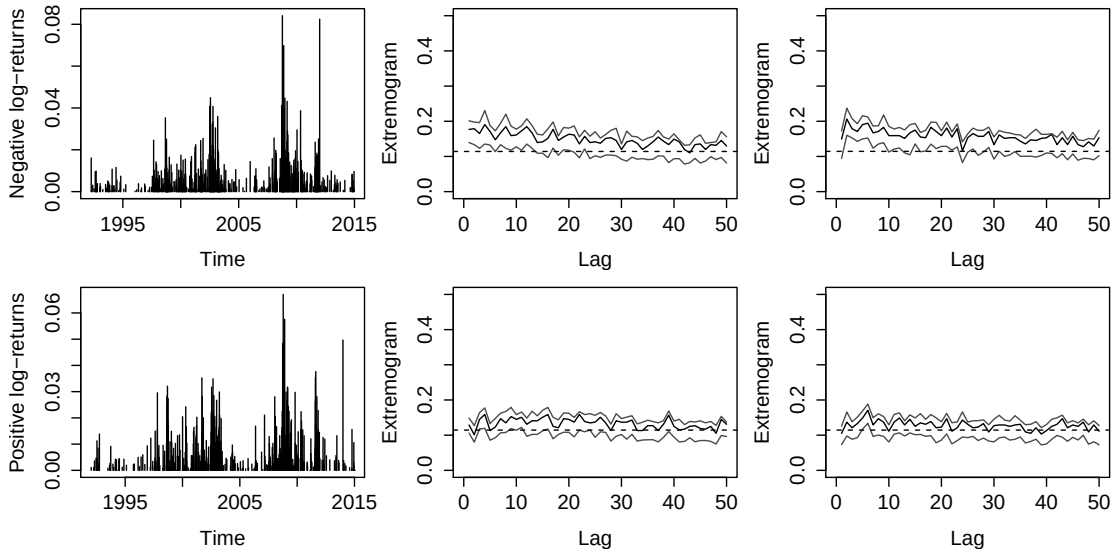

Figure 1: 8.5% of the most extreme losses (top left) and gains (bottom left) for an equal-weighted portfolio based on the DAX, S&P500, and FTSE100 indexes from January 3, 1994, to December 30, 2014. The sample extremograms are shown in the middle figures with losses in the top and gains in the bottom. The right figures show the cross-extremograms for losses conditional on gains (top right) and gains conditional on losses (bottom right) at different lags. The dashed line corresponds to the value of the extremogram under the null hypothesis of independence at a 95% confidence level obtained based on 100 permutations. The sampling distribution of the (cross-) extremogram and confidence intervals are obtained based on $10,000$ bootstrap replications.

panels of Figure 2 display the time series of 9% of the most negative log returns of the three indices as well as the associated intensities. The latter are estimated based on a ACI-POT model as proposed in the remainder of this paper. We observe a considerable amount of co-clustering of extremes. This is even more evident by analyzing trivariate cross-extremograms for negative log returns of the three indices. Denote the negative log returns of the DAX, S&P500, and FTSE100 index by X_t, Y_t, and Z_t, respectively. Then, the third panel displays the cross-extremogram

$$\hat{\phi}_A^1(h) = \lim_{x\to\infty} P\left(x^{-1}Y_h \in A \cup x^{-1}Z_h \in A \mid x^{-1}X_0 \in A\right),$$

with x being the 91% empirical quantiles of the negative log returns and $A = (1,\infty)$. Likewise, in the bottom panel of Figure 2, we depict the cross-extremogram

$$\hat{\phi}_A^2(h) = \lim_{x\to\infty} P\left(x^{-1}X_0 \in A \mid x^{-1}Y_h \in A \cup x^{-1}Z_h \in A\right).$$

We observe clear evidence for both types of cross-extremal dependence among extreme nega-

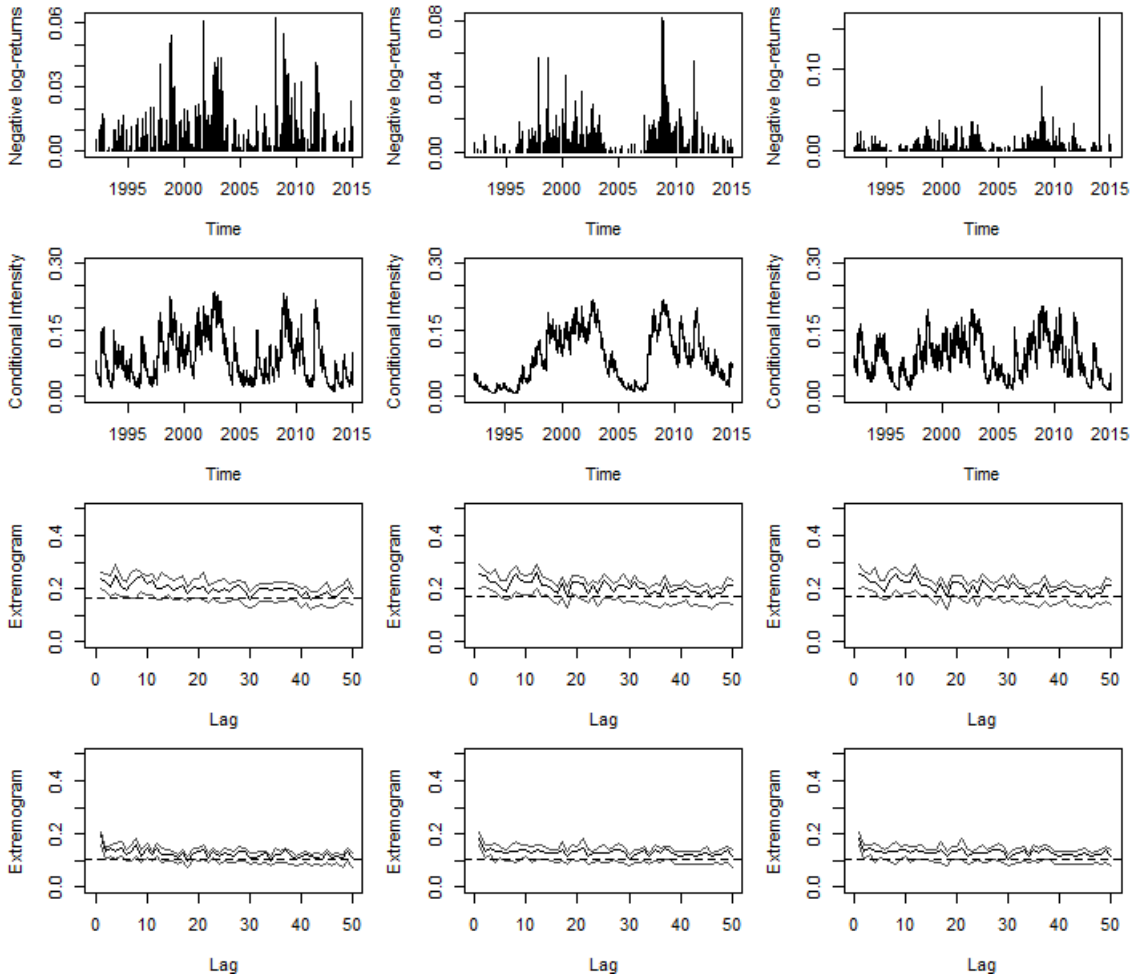

Figure 2: From top to bottom: Time series of 9% of the most extreme losses (the 91% empirical quantiles of the negative log returns), the conditional intensity for the occurrence of losses, the trivariate sample cross-extremograms corresponding to $\hat{\phi}_A^1(h)$ and $\hat{\phi}_A^2(h)$, respectively. From left to right for DAX, S&P 500, and FTSE 100 indexes.

tive events, confirming, e.g., Longin and Solnik (2001), Byström (2004), and Herrera and Eichler (2011). Clustering in extreme events can be moreover observed in the properties of the time between exceedances. Classical EVT assumes independent and identically distributed (i.i.d) observations implying that the timing of exceedances follows a Poisson point process, and thus that inter-exceedance times are exponentially distributed. The first panel of Figure 3 shows quantile-quantile plots, reflecting that the exponential distribution is clearly at odds with empirical observations. Moreover, autocorrelations in inter-exceedance times, as shown in the second panel, reveal a high degree of persistence. Finally, the bottom panel provides univariate empirical extremograms measuring the impact of a large loss on future extremes in the same market. All estimates are highly significant, and thus, confirm findings, e.g. from Davis et al. (2012) or Chang et al. (2013).

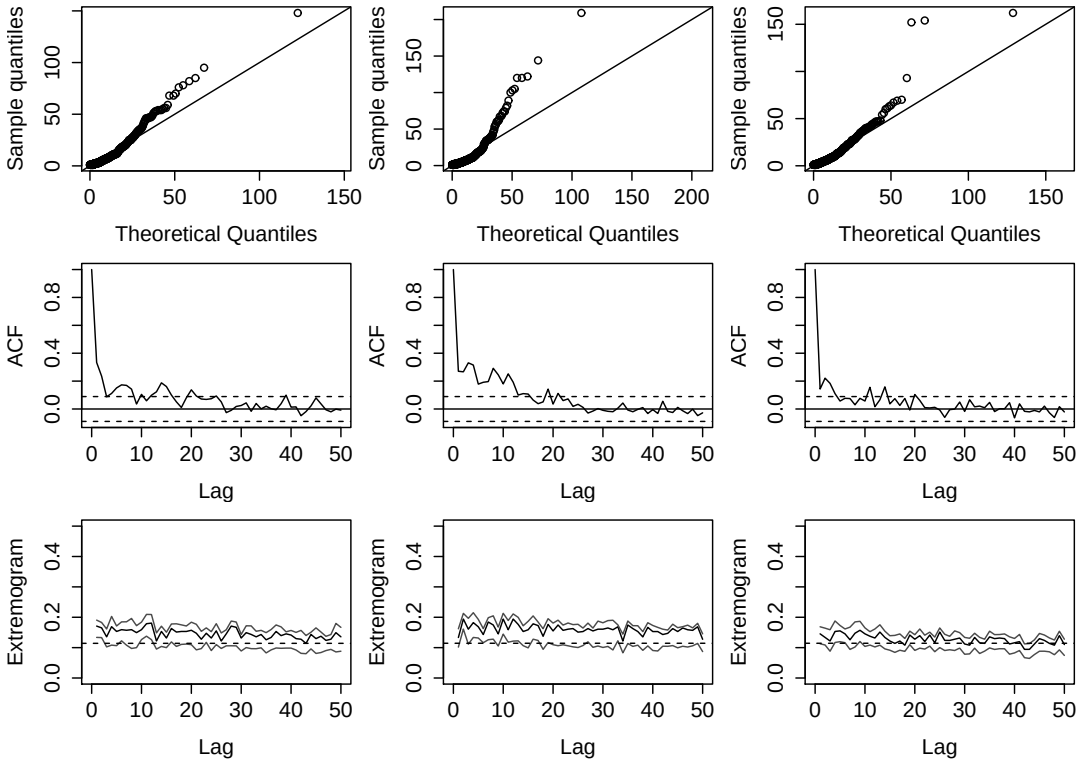

Figure 3: QQ-plots for inter-exceedance times, autocorrelation of inter-exceedance times, and sample extremograms for the 9% most extreme losses of the DAX, S&P 500, and FTSE 100 daily log returns.

We thus find strong empirical evidence for co-clustering, self-excitation and cross-excitation in extreme events. In the next section, we present an econometric model to capture this feature.

3 Dynamic Intensity Peaks-Over-Threshold Models

Denote by $\boldsymbol{z}_t := \left(z_t^1, \dots z_t^M\right)$ the $M \times 1$ vector of negative log returns at time t each one with an unknown cumulative distribution functions F^m with upper end point with distribution function $\epsilon_F^m := \sup\left\{Y_t^m \in \mathbb{R} : F^m < 1\right\} \leq \infty$ and sample maxima $\mathcal{M}_n^m = \max\left\{z_1^1, \dots, z_n^1\right\}$. Under some suitable normalization of the maxima , it can be shown that F^m is in the domain of attraction of the generalized extreme value (GEV) distribution

$$H\left(z^m\right) = \begin{cases} \exp\left\{-\left(1+\xi\frac{z^m-\mu}{\sigma}\right)_+^{-1/\xi}\right\}, & \xi \neq 0, \\ \exp\left\{-\exp\left(-\frac{z^m-\mu}{\sigma}\right)\right\}, & \xi = 0, \end{cases} \tag{1}$$

where $\mu, \xi \in \mathbb{R}$, and $\sigma > 0$, corresponding to location, shape, and scale parameters, respectively, and $a_+ = \max(a, 0)$.

Define $y_t^m := z_t^m - u^m$ as the so-called exceedance beyond a high threshold $0 < u^m < \epsilon_F^m$. By denoting the occurrence time of the j-th exceedance in process m by t_j^m, we define the corresponding magnitude $y_j^m := z_{t_j}^m - u^m$ as a mark. Furthermore, $x_j^m := t_j^m - t_{j-1}^m$ represents the inter-exceedance time between consecutive m-type extreme events, while $x^m(t) := t - t_j^m$ denotes the backward recurrence time to the most previous extreme event j. According to the Pickands-Balkema-de Haan theorem, when $u^m \to \epsilon_F^m$ the distribution of i.i.d. marks converges to a generalized Pareto distribution, whose density function is given by

$$g\left(y^m \mid \mathcal{H}_t\right) = \begin{cases} \frac{1}{\beta}\left(1 + \xi\frac{y^m}{\beta}\right)_+^{-1/\xi - 1}, & \xi \neq 0, \\ \frac{1}{\beta}\exp\left(-\frac{y^m}{\beta}\right), & \xi = 0, \end{cases} \tag{2}$$

where $\beta = \sigma + \xi\left(u - \mu\right)$ is a reparametrized scale parameter.

Define $N^m(t)$ as the right-continuous function, counting the number of events for which $y_j^m > 0$ until and including t. We therefore have $N^m(t) = N^m(t_j^m) = j$ if $t_j^m \leq t < t_{j+1}^m$. Thus, $j = 1, 2, \ldots$ is a counting sub-index for each dimension $m \in \{1, \ldots, M\}$. Accordingly, $\breve{N}^m(t)$ denotes the corresponding left-continuous function, counting the number of events for which $y_j^m > 0$ until but excluding t. Then, $\boldsymbol{N}(t) := \left(N^1(t), \ldots, N^M(t)\right)$ defines a multivariate marked point process (MPP) of exceedances. For each dimension $m \in \{1, \ldots, M\}$ the process is characterized by a double sequence $\left\{\left(t_j^m, y_j^m\right)\right\}_{j \geq 1} \in \Omega = (0, 1] \times (u^m, \infty)$ of unpooled arrival times and marks, where time is for convenience measured on a rescaled interval $(0, 1]$. Finally, let us define $\breve{\boldsymbol{N}}(t)$ as the left-continuous counting function of the pooled arrival times t_i of extreme events regardless of the event type, and therefore, $\mathrm{i} = 1, 2, \ldots$ corresponds to a counting sub-index for the pooled process.

The multivariate process of exceedance times and marks can be specified via an $M-$variate vector of conditional intensities

$$\begin{aligned} \lambda^m\left(t, y \mid \mathcal{H}_t\right) \mathtt{d}t\mathtt{d}y &= \mathbb{E}\left[N^m\left(\left[t, t + \mathtt{d}t\right] \times \left[y, y + \mathtt{d}y\right]\right) \mid \mathcal{H}_t\right] \\ &= \mathbb{E}\left[N^m\left(\mathtt{d}t \times \mathtt{d}y\right) \mid \mathcal{H}_t\right], \end{aligned} \tag{3}$$

where $\mathcal{H}_t = \left\{\left(t_j^m, y_j^m\right) \ \forall\ (m, j) : t_j^m < t\right\}$ denotes the internal history of $\boldsymbol{N}(t)$, while $\mathtt{d}t$ and

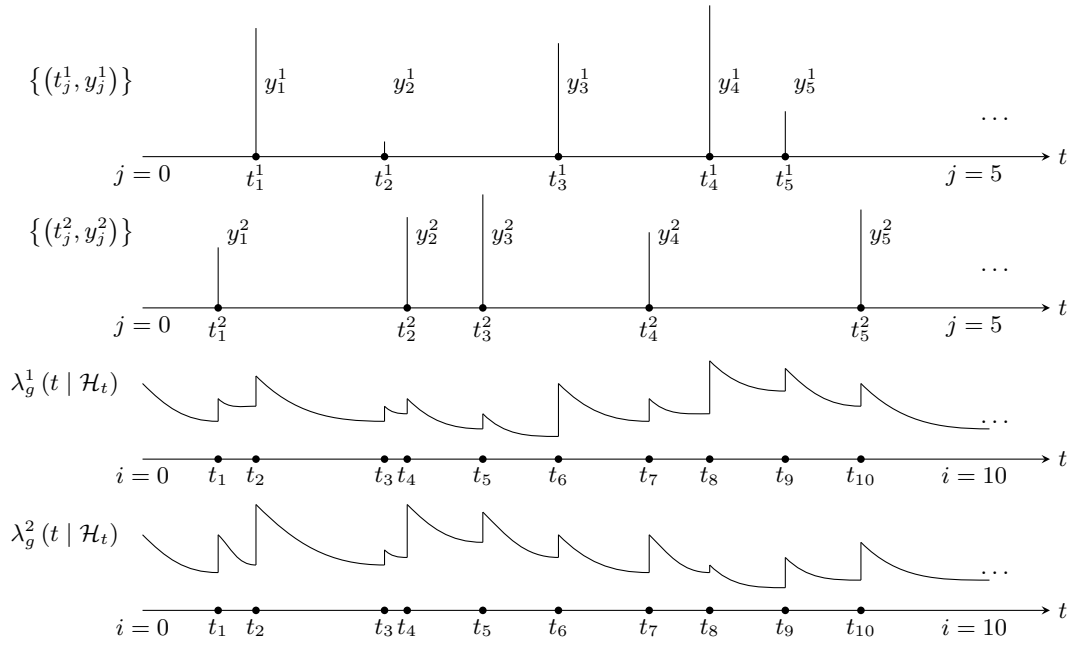

Figure 4: Graphical presentation of a bivariate MPP of exceedances. Observe that j is a counting sub-index for each dimension $m \in \{1, 2, \ldots\}$, while i is the sub-index for the pooled process .

$\mathtt{d}y$ reflect an infinitesimal amount around the time t and mark y, respectively. Following Daley and Vere-Jones (2003), the conditional intensity (3) can be decomposed into two parts: one component characterizing the so-called ground intensity $\lambda_g^m(t \mid \mathcal{H}_t)$, capturing the intensity of the occurrence of extreme events, and the conditional probability density function of the size of exceedances, $g^m(y \mid \mathcal{H}_t, t)$, i.e.,

$$\lambda^m(t, y \mid \mathcal{H}_t) = \lambda_g^m(t \mid \mathcal{H}_t)\, g^m(y \mid \mathcal{H}_t, t)\,. \tag{4}$$

Figure 4 gives a graphical representation of this framework for two dimensions. The first two top panels describe two stochastic processes of exceedances with their respective arrival times and exceedances, while the two bottom panels show their corresponding conditional ground intensities, which are updated at each event of the pooled process.

In order to capture the dynamics in the magnitude of exceedances, we propose a multiplicative error model in the spirit of Engle (2002). Let $\psi_j^m := \ln \mathbb{E}\left(y_j^m \mid \mathcal{H}_t\right)$ define the log of the conditional expectation of the exeeceedance y_j^m. Then, the size of exceedances in the m-th component

follows a logarithmic multiplicative error model (MEM) given by

$$\begin{aligned} y_j^m &= \exp\left(\varphi_j^m\right)\epsilon_j^m, \\ \varphi_j^m &= \psi_j^m - \ln(1-\xi_m), \\ \psi_j^m &= w_m + \rho_m \ln y_{j-1}^m + \beta_m \psi_{j-1}^m + \gamma_m x_j^m, \end{aligned}$$

where $x_j^m := t_j^m - t_{j-1}^m$ is the time elapsed since the last event, $\xi_m \in \mathbb{R}_+$ is the shape parameter of the GPD, and ρ_m, β_m, γ_m are parameters. This conditional autoregressive specification introduces a feedback between the inter-exceedance times and the conditional expectation of the size of exceedances, as governed by the parameter γ_m. The logarithmic specification ensures the non-negativity of the process without explicitly imposing corresponding parameter restrictions. The error terms ϵ_j^m are i.i.d generalized Pareto random variables with parameters ξ_m and scale parameter one, with p.d.f. given by

$$g_\epsilon(\epsilon_j^m) = \left(1 + \xi_m \epsilon_j^m\right)^{-1/\xi_m - 1}. \tag{5}$$

The conditional density of the exceedance y_j^m is therefore given by

$$g^m\left(y \mid \mathcal{H}_t, t\right) = \frac{1}{\exp\left(\varphi_j^m\right)} \left(1 + \xi_m \frac{y_j^m}{\exp\left(\varphi_j^m\right)}\right)^{-1/\xi_m - 1}, \tag{6}$$

corresponding to a generalized Pareto density with time-varying scale parameter $\exp\left(\varphi_j^m\right)$. Inserting (8) and (6) into (4) yields the intensity representation of the m-th component of a multivariate ACI-POT model

$$\lambda^m\left(t, y \mid \mathcal{H}_t\right) = \frac{\exp\left(\Phi^m_{\breve{N}(t)+1} + \tilde{y}^m_{\breve{N}^m(t)}\delta_m\right)\lambda_0^m\left(t\right)}{\exp\left(\varphi^m_{\breve{N}^m(t)}\right)} \left(1 + \xi_m \frac{y^m_{\breve{N}^m(t)}}{\exp\left(\varphi^m_{\breve{N}^m(t)}\right)}\right)^{-1/\xi_m - 1}. \tag{7}$$

Dynamic specifications of $\lambda_g^m\left(t \mid \mathcal{H}_t\right)$ will be provided in the next section.

3.1 The Multivariate ACI-POT Model

Capturing self- and cross-excitations in the ground intensity $\lambda_g^m\left(t \mid \mathcal{H}_t\right)$ in terms of an ACI process yields

$$\lambda_g^m\left(t \mid \mathcal{H}_t\right) = \exp\left(\Phi^m_{\breve{N}(t)+1} + \tilde{y}^m_{\breve{N}^m(t)}\delta_m\right)\lambda_0^m\left(t\right). \tag{8}$$

Here, $\Phi^m_{\breve{N}(t)+1}$ is a discrete-time dynamic process that is updated instantaneously *after* the occurrence of an extreme event of the pooled process and does not change until the next event. Note that $\Phi^m_{\breve{N}(t)+1} = \Phi^m_i$ if $t_{i-1} < t \leq t_i$ and $\Phi^m_{\breve{N}(t)+1} = \Phi^m_{i+1}$ if $t > t_i$. Furthermore, $\tilde{y}^m_{\breve{N}^m(t)} := y^m_{\breve{N}^m(t)}/u^m = (z^m_{\breve{N}^m(t)} - u^m)/u^m$ denotes the "standardized" exceedance of the most recent m-type extreme event.[3] Its effect on $\lambda_g^m\left(t \mid \mathcal{H}_t\right)$ is captured by the parameter δ_m. Finally, $\lambda_0^m(t) = \lambda_0^m\left(x^m(t)\right)$ corresponds to a baseline intensity that changes continuously in the time elapsed since the previous m-type extreme event.

By stacking the components Φ_i^m in a $M \times 1$ vector $\boldsymbol{\Phi}_i := \left(\Phi_i^1, \ldots, \Phi_i^M\right)'$, a VARMA($p$,$q$)–type specification of the form

$$\Phi_i^m = \sum_{k=1}^{M}\sum_{\ell=1}^{p} b_{mk}^{\ell}\Phi_{i-\ell}^k + \sum_{k=1}^{M}\sum_{r=1}^{q} a_{mk}^{r}(\varepsilon_{i-r}d_{i-r}^k) \tag{9}$$

is obtained. The model can also be written in a more compact matrix notation

$$\boldsymbol{\Phi}_i = \sum_{\ell=1}^{p} B^{\ell}\boldsymbol{\Phi}_{i-\ell} + \sum_{r=1}^{q} A^{r}(\varepsilon_{i-r}\boldsymbol{d}_{i-r}),$$

where $A^r := \{a_{mk}^r\}$ corresponds to an $M \times M$ coefficient matrix denoting the impact of $r-$lagged innovation term ε_{j-r} on the ground intensity, $B^{\ell} := \left\{b_{mk}^{\ell}\right\}$ is an $M \times M$ coefficient matrix of persistence parameters, and $\boldsymbol{d}_i := \left(d_i^1, \ldots, d_i^M\right)^{\top}$ is a vector of indicator variables taking on the value one if the i-th event of the pooled process is of type m, and zero otherwise.

One fundamental result of point process theory is the random time change theorem by Meyer (1971). Through this theorem we can transform a wide class of point process to a homogeneous Poisson process. Based on this idea we build the scalar innovation term ε_i. This is computed based

[3] Using a standardized exceedance instead of $y^m_{\breve{N}^m(t)}$ is advantageous to avoid numerical instabilities of estimates in the case of very high exceedances. Such standardizations are commonly used in EVT, see, e.g., Resnick (2006).

on the integrated intensity of the process for which we observed the most recent exceedance,

$$\varepsilon_{\breve{N}(t)} := \sum_{m=1}^{M} \left\{ 1 - \Lambda^m \left(t_{\breve{N}^m(t)-1}, t_{\breve{N}^m(t)} \right) \right\} d^m_{\breve{N}(t)} = \sum_{m=1}^{M} \left\{ 1 - \int_{t_{\breve{N}^m(t)-1}}^{t_{\breve{N}^m(t)}} \lambda^m_g \left(s \mid \mathcal{H}_s \right) ds \right\} d^m_{\breve{N}(t)}, \tag{10}$$

where $\Lambda^m \left(t^m_{j-1}, t^m_j \right)$ denotes the m-type integrated intensity. Accordingly, ε_i corresponds to a mixture of centered exponential random variables with $E\left[\varepsilon_i\right] = 0$.

The process $\mathbf{\Phi}_i$ is mean reverting with $E\left[\mathbf{\Phi}_i\right] = 0$ if all the eigenvalues of $\det|\mathbf{\Phi}_i|$ lie inside the unit circle (see Proposition 2 in Russell, 1999 and the discussion in Hautsch, 2012). In the VARMA(1,1)–type specification, $\mathbf{\Phi}_i$ follows a $MA\left(\infty\right)$ representation $\mathbf{\Phi}_i = \sum_{k=1}^{\infty} B^{k-1} A(\varepsilon_{i-k} \boldsymbol{d}_{i-k})$, allowing to construct impulse response functions to study how extreme events propagate through the multivariate process. Indeed, under the initial condition $\mathbf{\Phi}_0 = \mathbf{0}$, the vector of impulse responses h periods forward, induced by a marginal change in ε_i that is triggered by an event in process m, is given by

$$IRF^m(h) = B^{h-1} A e_m, \tag{11}$$

where e_m denotes the m-th column of an $M \times M$ identity matrix. Finally, the baseline intensity function $\lambda^m_0\left(t\right)$ is specified in form of a parametric function representing the shape of a hazard function. In this paper, we utilize the generalized gamma distribution with hazard function given by

$$\lambda^m_0\left(t\right) = \frac{|q_m|}{\sigma_m x^m(t) \Gamma\left(q_m^{-2}\right) S^m(t)} \varrho_m(t)^{q_m^{-2}} \exp\left(-\varrho_m(t)\right), \tag{12}$$

with location υ_m, scale $\sigma_m > 0$ and shape q_m, where $\varrho_m(t) := q_m^{-2} \left(e^{-\upsilon_m} x^m(t)\right)^{q_m^{-2}}$. Here $\Gamma\left(t;\gamma\right) := \int_0^t x^{\gamma-1} e^{-x} \mathrm{d}x / \Gamma\left(\gamma\right)$ and $S^m(t)$ denotes the survival function of the generalized gamma distribution given by

$$S^m(t) = \begin{cases} 1 - \Gamma\left(\varrho_m(t); q_m^{-2}\right) & \text{if } q_m < 0 \\ \Gamma\left(\varrho_m(t); q_m^{-2}\right) & \text{if } q_m > 0. \end{cases}$$

This type of hazard function is commonly used in the empirical literature since it exhibits both

monotonic and non-monotonic behavior. An explicit derivation of conditional moments of $\lambda_g^m\left(t \mid \mathcal{H}_t\right)$ is generally not feasible due to a nonlinear relationship between the conditional intensity at time t_j^m and the expected time until the next extreme event t_{j+1}^m.

3.2 The Multivariate Hawkes-POT Model

A Hawkes process is a self-exciting point process, which is primarily applied in seismology, e.g., Hawkes and Oakes (1974) or Ogata (1988), but more recently also in finance, e.g., Bowsher (2007), Aït-Sahalia et al. (2014) or Aït-Sahalia et al. (2015). In the context of EVT, a univariate Hawkes-POT process is introduced by Chavez-Demoulin et al. (2005), and more recently used in Chavez-Demoulin and McGill (2012) and Grothe et al. (2014). The latter authors show that a superposition of individual self-exciting Hawkes processes with specific decay and impact functions yield exceedance times and exceedance magnitudes which can be coupled together using an extreme value copula. This yields a parsimonious framework, which, however does not allow for spillovers across time and processes.

Using a Hawkes framework, the conditional intensity of arrival times in the m-th component is given by

$$\lambda_g^m(t \mid \mathcal{H}_t) = \mu_m + \sum_{k=1}^{M} b_{mk} \sum_{j=1}^{N^m(t)} h_k^m\left(t - t_j^k\right), \tag{13}$$

where $\mu_m > 0$ corresponds to the immigrant rate (or baseline intensity), $h : \mathbb{R} \to \mathbb{R}^+$ is a decay kernel describing the instantaneous influence of the k-th component, and how this deviates from the baseline μ_m through time. Finally, the parameters $b_{mk} > 0$ are coefficients defining the $M \times M$ branching matrix $\mathbf{B} = \{b_{mk}\}$. We assume that the decay kernel function corresponds to the product of two exponential functions: one puts exponential weights on the time elapsed since the last event. The other function scales the kernel by the size of the most recent standardized excess, $\tilde{y}_{j-1}^m = y_{j-1}^m / u^m$. Then, $h_k^m\left(t - t_j^k\right)$ is given by

$$h_k^m\left(t - t_j^k\right) = a_k \exp\left(\delta_k \tilde{y}_j^k - a_k\left(t - t_j^k\right)\right),$$

with $a_k > 0$ and $\delta_k \in \mathbb{R}$. The impact of spillovers between the individual processes is captured by the parameters b_{mk}. Thus, the intensities rise in response to the self- (b_{mm}) and cross-excitation (b_{mk} for $m \neq k$) mechanisms introduced by past extreme events in the own market and the other

markets, respectively. In addition, the feedback between intensities and the size of exceedances is captured by the parameter δ_k. Unlike in the copula approach by Grothe et al. (2014), spillover dynamics between markets are not necessarily symmetric (i.e., $b_{mk} \neq b_{km}$ in general).

Specifying the size of m-type exceedances based on the MEM specification according to Section 3 yields the multivariate Hawkes-POT model given by

$$\lambda^m\left(t, y \mid \mathcal{H}_t\right) = \frac{\mu_m + \sum_{k=1}^{M} b_{mk} \sum_{j=1}^{N^m(t)} a_k \exp\left(\delta_k \tilde{y}_j^k - a_k\left(t - t_j^k\right)\right)}{\exp\left(\varphi^m_{\breve{N}^m(t)}\right)} \left(1 + \xi_m \frac{y^m_{\breve{N}^m(t)}}{\exp\left(\varphi^m_{\breve{N}^m(t)}\right)}\right)^{-1/\xi_m - 1}. \tag{14}$$

As shown in the following proposition, the mean stationarity of the ground intensity of the non-marked Hawkes-POT model ($\delta_k = 0$) requires the eigenvalues of the branching matrix $\mathbf{B}$ to lie inside the unit circle.[4]

Proposition 1. *The ground intensity of the Hawkes-POT model is mean stationary, if and only if, the* $spr\left(\mathbf{B}\right) = \max\left\{|\varphi| : \det\left(\mathbf{B} - \varphi \boldsymbol{I_M}\right) = 0\right\} < 1$*, where* $\boldsymbol{I_M}$ *is the identity matrix,* $spr\left(\mathbf{B}\right)$ *and* φ *denote the spectral radius and the eigenvalues vector of* $\mathbf{B}$*, respectively.*

Proof. See Supplementary Appendix.

3.3 Specification Testing

Denoting the vector of MEM parameters by $\boldsymbol{\theta_\varrho}$ and the vector of ACI or Hawkes parameters, respectively, by $\boldsymbol{\theta_\lambda}$, and observing the process over the time interval $(0, T]$, the resulting log-likelihood function is given by

$$\begin{aligned} \ln L\left(t, y \mid \boldsymbol{\theta_\lambda}, \boldsymbol{\theta_\varrho}; \mathcal{H}_t\right) &= \sum_{m=1}^{M} \sum_{j=1}^{N^m(T)} \ln g(y_j^m | \mathcal{H}_t; \boldsymbol{\theta_\varrho}) \\ &\quad + \sum_{m=1}^{M} \sum_{j=1}^{N^m(T)} \left\{ d_j^m \ln \lambda_g^m\left(t_j \mid \mathcal{H}_t; \boldsymbol{\theta_\lambda}\right) - \int_{t_{j-1}^m}^{t_j^m} \lambda_g^m\left(s \mid \mathcal{H}_s; \boldsymbol{\theta_\lambda}\right) ds \right\}. \end{aligned} \tag{15}$$

Note that the two sets of parameters $\boldsymbol{\theta_\lambda}$ and $\boldsymbol{\theta_\varrho}$ are disjoint, which allows us to maximize the two log-likelihood components separately. Below we develop Lagrange Multiplier (LM) tests to test

[4] Proving the stability of the more general model for $\delta_k \neq 0$ is significantly more complicated and left for future research.

for self-excitations, cross-excitations and spillover effects. For sake of brevity, we illustrate the derivations for the ACI framework only. The results hold analogously in the Hawkes setting.

If the latest observed extreme event j is the type m, the contribution to the log-likelihood function of this event is given by

$$\ln L\left(t_j^m, y_j^m \mid \boldsymbol{\theta}_{\boldsymbol{\lambda}}, \boldsymbol{\theta}_{\boldsymbol{\varrho}}; \mathcal{H}_t\right) = \ln g(y_j^m|\mathcal{H}_t; \boldsymbol{\theta}_{\boldsymbol{\varrho}}) + \ln \lambda_g^m\left(t_j \mid \mathcal{H}_t; \boldsymbol{\theta}_{\boldsymbol{\lambda}}\right) - \int_{t_{j-1}^m}^{t_j^m} \lambda_g^m\left(s \mid \mathcal{H}_s; \boldsymbol{\theta}_{\boldsymbol{\lambda}}\right) ds.$$

By conditional independence between the rate and the magnitude of the exceedances, the score with respect to $\boldsymbol{\theta}_{\boldsymbol{\lambda}}^*$ and $\boldsymbol{\theta}_{\boldsymbol{\varrho}}^*$ is obtained as

$$\begin{aligned} s_j^m\left(\boldsymbol{\theta}_{\boldsymbol{\lambda}}^*\right) &= \frac{\partial \ln L\left(t_j^m, y_j^m \mid \boldsymbol{\theta}_{\boldsymbol{\lambda}}, \boldsymbol{\theta}_{\boldsymbol{\varrho}}; \mathcal{H}_t\right)}{\partial \boldsymbol{\theta}_{\boldsymbol{\lambda}}} \\ &= \left(\frac{\partial \lambda_g^m\left(t_j \mid \mathcal{H}_t; \boldsymbol{\theta}_{\boldsymbol{\lambda}}\right)}{\partial \boldsymbol{\theta}_{\boldsymbol{\lambda}}} \Big/ \lambda_g^m\left(t_j \mid \mathcal{H}_t; \boldsymbol{\theta}_{\boldsymbol{\lambda}}\right)\right) - \int_{t_{j-1}^m}^{t_j^m} \frac{\partial \lambda_g^m\left(s \mid \mathcal{H}_s; \boldsymbol{\theta}_{\boldsymbol{\lambda}}\right) ds}{\partial \boldsymbol{\theta}_{\boldsymbol{\lambda}}}\Bigg|_{\boldsymbol{\theta}_{\boldsymbol{\lambda}}=\boldsymbol{\theta}_{\boldsymbol{\lambda}}^*} \end{aligned} \tag{16}$$

and

$$\begin{aligned} s_j^m\left(\boldsymbol{\theta}_{\boldsymbol{\varrho}}^*\right) &= \frac{\partial \ln L\left(t_j^m, y_j^m \mid \boldsymbol{\theta}_{\boldsymbol{\lambda}}, \boldsymbol{\theta}_{\boldsymbol{\varrho}}; \mathcal{H}_t\right)}{\partial \boldsymbol{\theta}_{\boldsymbol{\varrho}}} \\ &= \left(\frac{\partial g(y_j^m|\mathcal{H}_t; \boldsymbol{\theta}_{\boldsymbol{\varrho}})}{\partial \boldsymbol{\theta}_{\boldsymbol{\varrho}}} \Big/ g(y_j^m|\mathcal{H}_t; \boldsymbol{\theta}_{\boldsymbol{\varrho}})\right)\Big|_{\boldsymbol{\theta}_{\boldsymbol{\varrho}}=\boldsymbol{\theta}_{\boldsymbol{\varrho}}^*}, \end{aligned} \tag{17}$$

evaluated under $\boldsymbol{\theta}_{\boldsymbol{\lambda}} = \boldsymbol{\theta}_{\boldsymbol{\lambda}}^*$ and $\boldsymbol{\theta}_{\boldsymbol{\varrho}} = \boldsymbol{\theta}_{\boldsymbol{\varrho}}^*$, respectively. The score vectors and information matrices for both sets of parameters are then given by

$$S\left(\boldsymbol{\theta}\right) = \sum_{m=1}^{M} \sum_{j=1}^{N(T)} s_j^m\left(\boldsymbol{\theta}\right) \quad \text{and} \quad I\left(\boldsymbol{\theta}\right) = \sum_{m=1}^{M} \sum_{j=1}^{N(T)} s_j^m\left(\boldsymbol{\theta}\right) s_j^m\left(\boldsymbol{\theta}\right)^\top,$$

where $\boldsymbol{\theta} \in \left\{\boldsymbol{\theta}_{\boldsymbol{\lambda}}^*, \boldsymbol{\theta}_{\boldsymbol{\varrho}}^*\right\}$. The LM test is then computed as $S\left(\boldsymbol{\theta}^*\right)^\top I\left(\boldsymbol{\theta}^*\right)^{-1} S\left(\boldsymbol{\theta}^*\right)$, which is asymptotically $\chi^2\left(q\right)$ distributed, with q denoting the number of restrictions.

To test for spill-over effects in the intensities across markets, we formulate the null hypothesis as $b_{mk}^\ell = 0$ and $a_{mk}^r = 0$ for all $m \neq k$. In this case, the corresponding LM test, denoted by LM_{cross}, requires only the estimates of the score (16). Hence, for $\left\{b_{mk}^\ell, a_{mk}^r\right\} = \{0, 0\}$ for all $m \neq k$, we obtain

$$\left.\frac{\partial \lambda_g^m\left(t_j \mid \mathcal{H}_t;\boldsymbol{\theta}_{\boldsymbol{\lambda}}\right)}{\partial b_{mk}^{\ell}}\right|_{\boldsymbol{\theta}_{\boldsymbol{\lambda}}=\boldsymbol{\theta}_{\boldsymbol{\lambda}}^*} = \exp\left(\Psi_j^m+\tilde{y}_{j-1}^m\delta_m\right)\lambda_0^m\left(t_j\right)\Phi_{j-\ell}^k$$

$$\left.\int_{t_{j-1}^m}^{t_j^m}\frac{\partial \lambda_g^m\left(s \mid \mathcal{H}_s;\boldsymbol{\theta}_{\boldsymbol{\lambda}}\right)}{\partial b_{mk}^{\ell}}ds\right|_{\boldsymbol{\theta}_{\boldsymbol{\lambda}}=\boldsymbol{\theta}_{\boldsymbol{\lambda}}^*} = \exp\left(\Psi_j^m+\tilde{y}_{j-1}^m\delta_m\right)\Lambda_0^m\left(t_j,t_{j-1}\right)\Phi_{j-\ell}^k,$$

and

$$\left.\frac{\partial \lambda_g^m\left(t_j \mid \mathcal{H}_t;\boldsymbol{\theta}_{\boldsymbol{\lambda}}\right)}{\partial a_{mk}^{r}}\right|_{\boldsymbol{\theta}_{\boldsymbol{\lambda}}=\boldsymbol{\theta}_{\boldsymbol{\lambda}}^*} = \exp\left(\Psi_j^m+\tilde{y}_{j-1}^m\delta_m\right)\lambda_0^m\left(t_j\right)\varepsilon_{j-r}d_{j-r}^k$$

$$\left.\int_{t_{j-1}^m}^{t_j^m}\frac{\partial \lambda_g^m\left(s \mid \mathcal{H}_s;\boldsymbol{\theta}_{\boldsymbol{\lambda}}\right)}{\partial a_{mk}^{r}}ds\right|_{\boldsymbol{\theta}_{\boldsymbol{\lambda}}=\boldsymbol{\theta}_{\boldsymbol{\lambda}}^*} = \exp\left(\Psi_j^m+\tilde{y}_{j-1}^m\delta_m\right)\Lambda_0^m\left(t_j,t_{j-1}\right)\varepsilon_{j-r}d_{j-r}^k,$$

where

$$\Psi_j^m=\sum_{w=1}^{p}b_{mm}^w\Phi_{j-w}^m+\sum_{r=1}^{q}a_{mm}^r(\varepsilon_{j-r}d_{j-r}^m).$$

Hence, the score contribution of the extreme event j in the m-th process is given by

$$s_j^m\left(\boldsymbol{\theta}_{\boldsymbol{\lambda}}^*\right) = \left[\begin{array}{c}\frac{\partial \ln L\left(t_j^m,y_j^m|\boldsymbol{\theta}_{\boldsymbol{\lambda}},\boldsymbol{\theta}_{\boldsymbol{\varrho}};\mathcal{H}_t\right)}{\partial b_{mk}^{\ell}}\\ \frac{\partial \ln L\left(t_j^m,y_j^m|\boldsymbol{\theta}_{\boldsymbol{\lambda}},\boldsymbol{\theta}_{\boldsymbol{\varrho}};\mathcal{H}_t\right)}{\partial a_{mk}^{r}}\end{array}\right]_{\boldsymbol{\theta}_{\boldsymbol{\lambda}}=\boldsymbol{\theta}_{\boldsymbol{\lambda}}^*} = \left[\begin{array}{c}\left\{1-\exp\left(\Psi_j^m+\tilde{y}_{j-1}^m\delta_m\right)\Lambda_0^m\left(t_j,t_{j-1}\right)\right\}\Phi_{j-\ell}^k\\ \left\{1-\exp\left(\Psi_j^m+\tilde{y}_{j-1}^m\delta_m\right)\Lambda_0^m\left(t_j,t_{j-1}\right)\right\}\varepsilon_{j-r}d_{j-r}^k\end{array}\right].$$

To test for a feedback from the magnitude of exceedances to the intensities, we impose the restriction $\delta_m = 0$, for $m = 1,\ldots,M$. To construct the corresponding LM test, denoted by LM_{int}, the corresponding score elements are obtained as

$$\left.\frac{\partial \lambda_g^m\left(t_j \mid \mathcal{H}_t;\boldsymbol{\theta}_{\boldsymbol{\lambda}}\right)}{\partial \delta_m}\right|_{\boldsymbol{\theta}_{\boldsymbol{\lambda}}=\boldsymbol{\theta}_{\boldsymbol{\lambda}}^*} = \exp\left(\Phi_j^m\right)\lambda_0^m\left(t_j\right)\tilde{y}_{j-1}^m$$

$$\left.\int_{t_{j-1}^m}^{t_j^m}\frac{\partial \lambda_g^m\left(s \mid \mathcal{H}_s;\boldsymbol{\theta}_{\boldsymbol{\lambda}}\right)}{\partial \delta_m}ds\right|_{\boldsymbol{\theta}_{\boldsymbol{\lambda}}=\boldsymbol{\theta}_{\boldsymbol{\lambda}}^*} = \exp\left(\Phi_j^m\right)\Lambda_0^m\left(t_j,t_{j-1}\right)\tilde{y}_{j-1}^m,$$

and therefore

$$s_j^m\left(\boldsymbol{\theta}_{\boldsymbol{\lambda}}^*\right) = \left\{1-\exp\left(\Phi_j^m\right)\Lambda_0^m\left(t_j,t_{j-1}\right)\right\}\tilde{y}_{j-1}^m.$$

Similarly, to test for a feedback from the time between extreme events and their magnitudes, we formulate the hypothesis that $\gamma_m = 0$, for $m = 1, \ldots, M$. The corresponding LM test, denoted by LM_{marks}, is then constructed based on

$$\begin{aligned}\left.\frac{\partial g(y_j^m|\mathcal{H}_t;\boldsymbol{\theta}_{\boldsymbol{\varrho}})}{\partial \gamma_m}\right|_{\boldsymbol{\theta}_{\boldsymbol{\varrho}}=\boldsymbol{\theta}^*_{\boldsymbol{\varrho}}} &= (1/\xi_m+1)\frac{x_j^m y_j^m \xi_m}{\exp\left(2\widetilde{\varphi}_j^m\right)}\left(1+\xi_m\frac{y_j^m}{\exp\left(\widetilde{\varphi}_j^m\right)}\right)^{-1/\xi_m-2} \\ &\quad -\frac{x_j^m}{\exp\left(\widetilde{\varphi}_j^m\right)}\left(1+\xi_m\frac{y_j^m}{\exp\left(\widetilde{\varphi}_j^m\right)}\right)^{-1/\xi_m-1},\end{aligned}$$

where

$$\widetilde{\varphi}_j^m = w_m + \rho_m \ln y_{j-1}^m + \beta_m \psi_{j-1}^m - \ln(1-\xi_m),$$

implying

$$s_j^m\left(\boldsymbol{\theta}^*_{\boldsymbol{\varrho}}\right) = x_j^m \frac{y_j^m - \exp\left(\widetilde{\varphi}_j^m\right)}{\xi_m y_j^m + \exp\left(\widetilde{\varphi}_j^m\right)}.$$

4 Improving Conditional Risk Measures

The Basel Committee on Banking Supervision has proposed using ES instead of the VaR as an internal model-based approach for regulatory market risk capital. Here, we illustrate how to derive both risk measures based on the proposed specifications. Consider all losses z_t defined as the negative log returns of a particular asset with underlying cumulative distribution function F. For ease of exposition, we omit the superscript m. ES is estimated by first obtaining the VaR at confidence level α, which is based on the predictive distribution $F_{z_t|\mathcal{H}_t}\left(VaR_\alpha^t\right)$. By inverting the uderlying distribution we can obtain the quantile function, which we call as a VaR at confidence level α

$$VaR_\alpha^t = q_\alpha^t\left(F_{z_t|\mathcal{H}_t}\right)$$

where $\mathcal{H}_t$ denotes the history of event times preceding time t. By computing the conditional survival function $\overline{F}_{z_t|\mathcal{H}_t}(z) = 1 - F_{z_t|\mathcal{H}_t}(z)$ as

$$\overline{F}_{z_t|\mathcal{H}_t}(z) = Pr\left(z_t > z \mid \mathcal{H}_t\right) = Pr\left(z_t > u \mid \mathcal{H}_t\right) Pr\left(z_t > z+u \mid z_t > u, \mathcal{H}_t\right), \tag{18}$$

where the probability $Pr\left(z_t > u \mid \mathcal{H}_t\right)$ can be derived as

$$Pr\left(N(t) - N(\bar{t}) > 0 \mid \mathcal{H}_t\right) = 1 - \exp\left(-\int_{\bar{t}}^{t} \lambda_g\left(s \mid \mathcal{H}_s\right) ds\right) \approx \lambda_g\left(t \mid \mathcal{H}_t\right), \tag{19}$$

where $\bar{t} < t$ is the occurrence time of the most recent observed extreme, and the last result is obtained by using the approximation $\ln(x) \approx x - 1$ as $x \to 1$. The conditional probability of exceedances is then computed as

$$\begin{aligned} Pr\left(z_t > z + u \mid z_t > u, \mathcal{H}_t\right) &= \frac{\int_{\bar{t}}^{t}\int_{z+u}^{\infty} \lambda\left(s, l \mid \mathcal{H}_s\right) dsdl}{\int_{\bar{t}}^{t}\int_{u}^{\infty} \lambda\left(s, l \mid \mathcal{H}_s\right) dsdl} = \left(1 + \xi\frac{z-u}{\exp\left(\varphi_{\breve{N}(t)}\right)}\right)^{-1/\xi} \\ &:= \overline{G}_{\xi,\exp\left(\varphi_{\breve{N}(t)}\right)}\left(z \mid \mathcal{H}_t, t\right), \end{aligned} \tag{20}$$

where $\overline{G}(\cdot)$ denotes the conditional generalized Pareto survival function. Finally, the VaR is defined by $Pr\left(z_t > VaR_\alpha^t \mid \mathcal{H}_t\right) = 1 - \alpha$, implying

$$VaR_\alpha^t = u + \frac{\exp\left(\varphi_{\breve{N}(t)}\right)}{\xi}\left\{\left(\frac{1-\alpha}{\lambda_g\left(t \mid \mathcal{H}_t\right)}\right)^{-\xi} - 1\right\}. \tag{21}$$

From this result, the conditional ES, corresponding to the conditional distribution of extreme events beyond the VaR, given $\mathcal{H}_t$, is computed as

$$ES_\alpha^t = \frac{1}{1-\alpha}\int_\alpha^1 q_s^t\left(F_{z_t|\mathcal{H}_t}\right) ds = \frac{VaR_\alpha^t}{1-\xi} + \frac{\exp\left(\varphi_{\breve{N}(t)}\right) - \xi u}{1-\xi}. \tag{22}$$

Note that

$$\lim_{\alpha\to 1}\frac{ES_\alpha^t}{VaR_\alpha^t} = \frac{1}{1-\xi}, \tag{23}$$

with the limit not depending on time. Recently, the Basel Committee proposed using the VaR at the 99% confidence level in internal model-based approaches with ES evaluated at the 97.5% confidence level (see BCBS, 2013). According to the Basel Committee, ES is less sensitive to extreme events than VaR, and therefore should account for the tail risk in a more comprehensive form. We analyze this proposition in the next section.

5 Applications

We employ the log returns of the DAX, S&P500, and FTSE100 indices through the sample period from January 2, 1994, to December 31, 2014, covering 5,128 trading days. Our first application is based on a bivariate model for the analysis of the clustering of extreme losses and gains of an equally-weighted portfolio based on the three indices. The second application considers a trivariate model to jointly model negative log returns of the three indexes. Here, we only illustrate applications of the ACI-POT model. The corresponding results for the Hawkes-POT model are comparable and are provided in the Supplementary Appendix.

In order to determine the tail threshold u, we follow the statistic proposed by Reiss and Thomas (2007) to determine the number of exceedances k by

$$\underset{k}{\operatorname{argmin}} f(k) = \frac{1}{k} \sum_{i=1}^{k} i^{\beta} \left| \hat{\xi}_i - \text{median} \left(\hat{\xi}_1, \ldots, \hat{\xi}_k \right) \right|, \tag{24}$$

where $\hat{\xi}_i$ is the estimate of the shape parameter for the sample fraction of extremes above the upper order statistic i, and $\beta \in [0, 0.5]$ is a tuning parameter. The idea is to find the sample proportion for which the distribution of the shape parameters is stable. Figure 5 displays the statistic in dependence of k and β for gains and losses. We find a proportion between 410 and 460 observations for gains and losses, respectively, to be a satisfactory size. We thus choose to proceed with 436 observations, corresponding to 8.5% of the most extreme events for losses and gains. For the trivariate application below, we proceed similarly and determine a threshold of 9% as a reasonable choice. In a robustness analysis in the Supplementary Appendix, we demonstrate that the results of our empirical analysis are pretty stable with respect to the choice of u.

5.1 Modeling Positive and Negative Extremes

Table 1 in the Appendix gives the estimation results for ACI-POT(1,1), (1,2) and (2,1) specifications. We find that the inclusion of a second lag improves the log likelihood only marginally, making the ACI-POT(1,1) the best fitting specification according to the BIC. Additional analysis (which is not reported here) shows that the inclusion of even higher lag orders considerably increases the complexity of the model without strongly improving its fit. This is in line with other studies in the literature employing dynamic intensity processes (see, e.g., Kehrle and Peter, 2013;

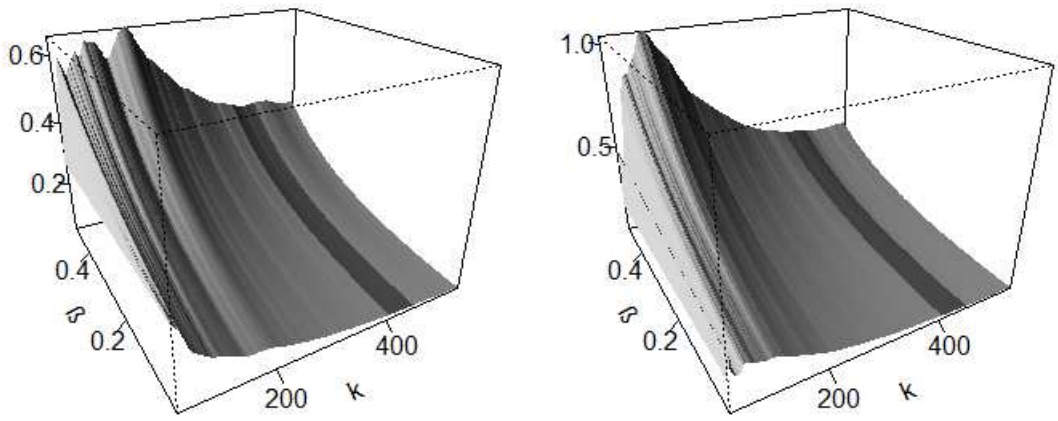

Figure 5: Threshold selection in the bivariate application. The left (right) panel shows the results of the statistic proposed in Eq. (24) for losses (gains) in dependence of k and β. The dark gray rectangle displays the subset of combinations of k and β yielding stable statistics.

Bacry and Muzy, 2014; Aït-Sahalia et al., 2015).

In Figure 6, we plot the estimated conditional intensity of the ground processes of positive and negative log returns, respectively, based on the different specifications. The bottom panel shows a barcode plot with black and gray colors depicting extreme events in gains and losses, respectively.

We can summarize the following findings: First, we find significant evidence for spillover effects between positive and negative extreme observations. This is particularly supported by the $LMcross$ test. Moreover, the estimates reflect some asymmetries in the spillover effects. Shocks in the intensities of negative extremes tend to *decrease* the probability of observing positive extremes ($a_{12} = -1.041$), while shocks in the intensities of positive extremes tend to increase the probability of observing negative extremes ($a_{21} = 0.644$). This kind of asymmetry can obviously not be captured by a Hawkes-type specification as the parameterization is more restrictive and does not allow for negative spillover effects. [5] This asymmetry is also reflected by the impulse-response functions shown in the bottom panel of Figure 7. The corresponding estimates indicate that shocks in positive extremes are less persistent than shocks in negative extremes.

Third, as depicted by Figure 6, the different ACI-POT specifications reflect a higher variability in negative extremes than in positive extremes (grey and black lines, respectively).

Fourth, the baseline hazard functions of the inter-exceedance times of both return processes reveal an inverted U-shape with the underlying densities being positively skewed as $q_m < 0$. As

[5] See the Supplemental Appendix for corresponding results for the Hawkes-POT approach.

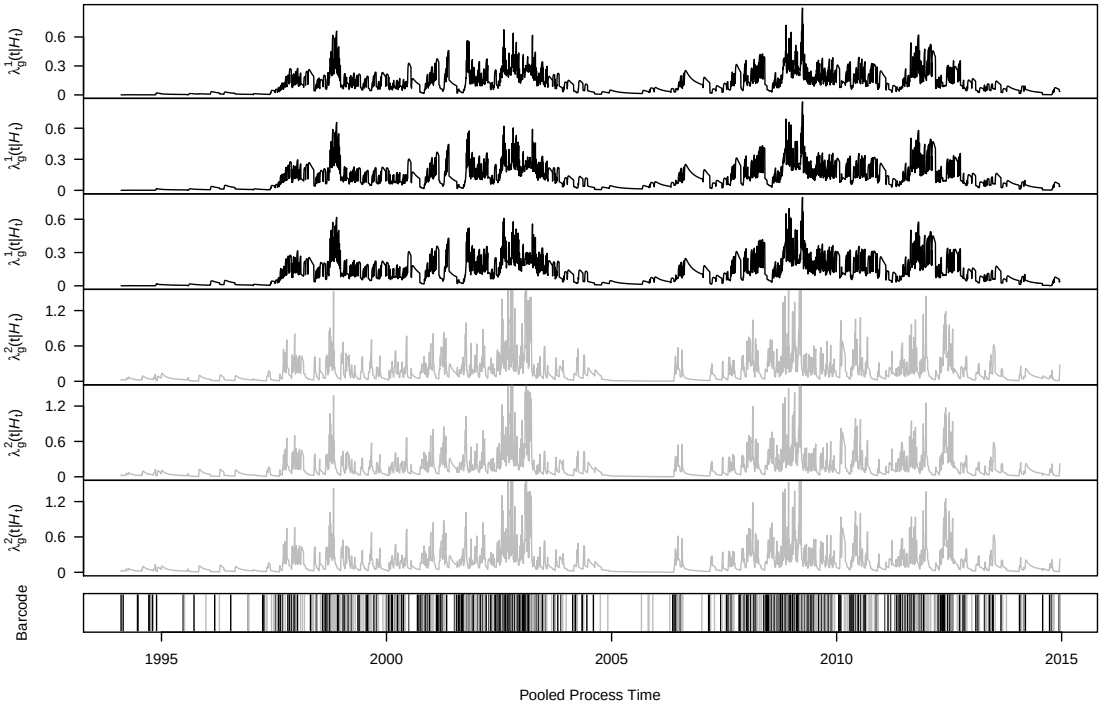

Figure 6: Bivariate conditional intensity of the ground process for the analyzed index portfolio. The three top panels show the estimated conditional intensities of the ground processes for positive log returns based on, from top to bottom, the ACI-POT (1,1), ACI-POT (2,1), and ACI-POT (1,2) specifications. Correspondingly, the next three panels exhibit the conditional intensity of the ground processes for negative log returns specifications. The bottom panel displays a barcode plot wherein the black or gray colors indicate the log returns causing the extreme observation.

shown by Figure 7, we observe that for positive extremes, the baseline function λ_0 decreases more slowly than for negative extreme returns. We moreover find that the baseline intensity function is generally higher for negative extremes than for positive extremes, indicating a higher temporal clustering.

Fifth, symmetries between positive and negative extremes are also reflected in the estimates of the coefficient δ_m, which captures the influence of the size of exceedances on the intensities. Though the effects are jointly significant according to the LM_{int} test, they are individually significant only for negative extremes, but not for positive extremes. Hence, higher (negative) exceedances particularly increase the intensity of negative extreme events. As a result of this self-enforcing behavior, negative returns are more clustered than positive ones. This is in line with, among others, Campbell and Hentschel (1992), suggesting that volatility is higher after stock markets exhibit losses, making stock market returns negatively correlated with future volatility.

Sixth, the estimates of the MEM specification provide clear evidence for a clustering of the

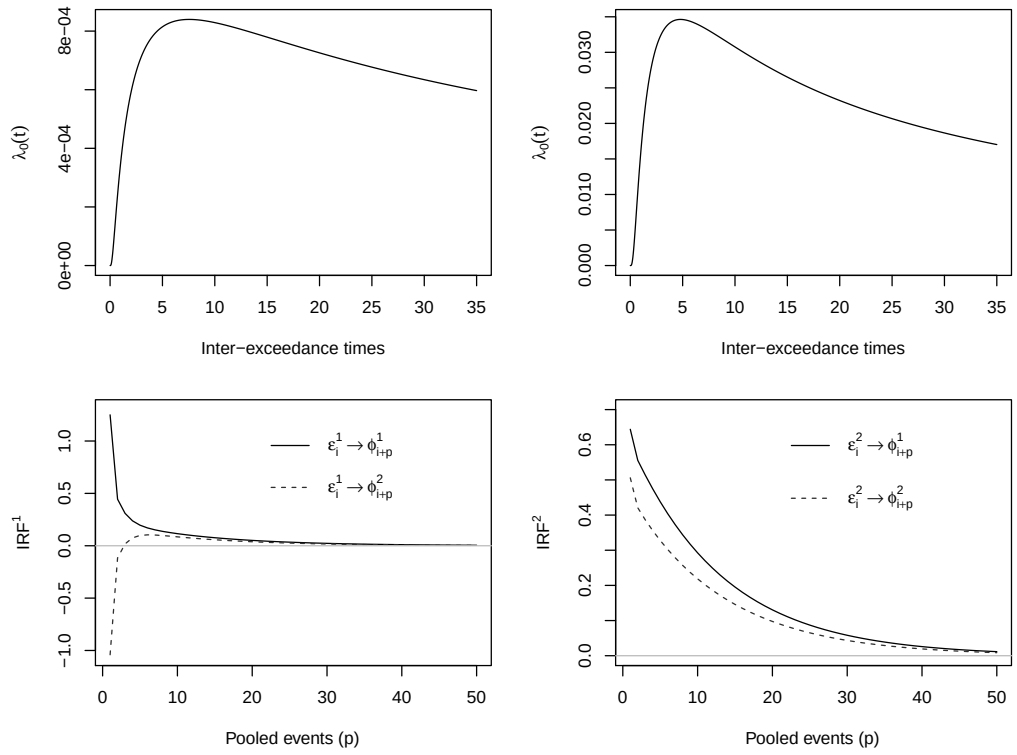

Figure 7: Top panel: Baseline hazard functions for the bivariate ACI-POT model. The left plot exhibits the baseline hazard function for positive returns, while the right plot shows the baseline hazard function for negative returns. Bottom panel: Impulse-response functions of innovations associated with a shock in the gains (IRF^1) and losses (IRF^2) in the pooled process. The left plot exhibits the impact of a standard deviation innovation on gains and its responses on the gains (solid line) and losses (dashed line) processes. The right plot shows the impact of a standard deviation innovation on losses and its responses on losses (solid line) and gains (dashed line) processes.

size of exceedances. Hence, small (large) exceedances are likely to be followed by small (large) exceedances. The coefficients γ_m are individually and jointly highly significant (according to the LM_{marks} test) and negative for gains. This indicates that high lagged inter-exceedance waiting times imply a reduction of the expected size of the marks. This is in line with Santos and Alves (2012), Hammoudeh et al. (2013) and Herrera and Schipp (2014).

Finally, diagnostics for the ACI-POT model can be straightforwardly performed by means of the de-meaned integrated intensities of the M individual ground processes $\varepsilon_j^m = 1 - \int_{t_{j-1}^m}^{t_j^m} \lambda_g^m \left(s \mid \mathcal{H}_s\right) ds$. Under correct specification and according to the random time change theorem Meyer (1971), each residual ε_j^m should be i.i.d. standard exponentially distributed with a mean of zero. Thus, by means of the distributional properties of this process, the model goodness-of-fit is evaluated. Engle and Russell (1998) proposes a test for excess dispersion which

builds on the statistic $\sqrt{n_\varepsilon}\left(\left(\tilde{\sigma}_\varepsilon^2 - 1\right)/\sqrt{8}\right)$, where n_ε corresponds to the number of residuals and $\tilde{\sigma}_\varepsilon$ is the empirical standard deviation of the residuals series ε_j^m, which should be 1 under the null hypothesis of correct model specification. Under this hypothesis, the test statistic is asymptotically standard normally distributed. In fact, in Table 1 we observe that the residuals are, on average, close to zero with standard deviations not far from unity. In addition, Ljung-Box statistics, testing against independence in the residuals, indicate that the model is able to capture the dynamics of the data fairly well.

5.2 Modeling Spillovers in Extremes

Table 2 gives the estimation results based on trivariate ACI-POT models for extremes in DAX, S&P500, and FTSE100 returns. The table reports an unrestricted and restricted ACI-POT(1,1) specification as well as an ACI-POT (1,2) specification.[6] The restricted ACI-POT(1,1) model is a specification without incorporating feedback between individual exceedance intensities and without feedback between the magnitude of exceedances and their conditional ground intensities.

Again, the best fit is provided by the *unrestricted* ACI-POT(1,1) model. The inclusion of higher lags improves the log likelihood only slightly but results into a worse BIC. Likewise, the restricted version of the ACI-POT model, which rules out feedback and cross-excitation effects, is not supported by the data. This is indicated by the information criteria as well as the LM tests. The latter indicate that spillover effects in intensities across markets as well as between intensities and magnitudes are statistically highly significant. The estimates therefore reveal clear evidence for self-excitation and cross-excitation effects as documented in Section 2, indicating the usefulness of the proposed model.

The ground processes reveal highly persistent though stationary dynamics in the intensities. According to the residual-based Ljung Box tests, the specification seems to capture the dynamics in the data pretty well. Moreover, the baseline functions reveal inverted U-shaped patterns. As illustrated in the top-panel of Figure 8, the baseline hazard functions increase until three to four days after the occurrence of the last extreme event, and decline thereafter. This non-monotonic pattern seems to be an important feature characterizing the time evolution of extreme events on financial markets and requires flexibility of the underlying parametrization. Observing peaks of

[6] An ACI-POT(2,1) models yields a very similar (but slightly worse) fit than the ACI-POT(1,2) specification and is omitted here. It is available upon request from the authors.

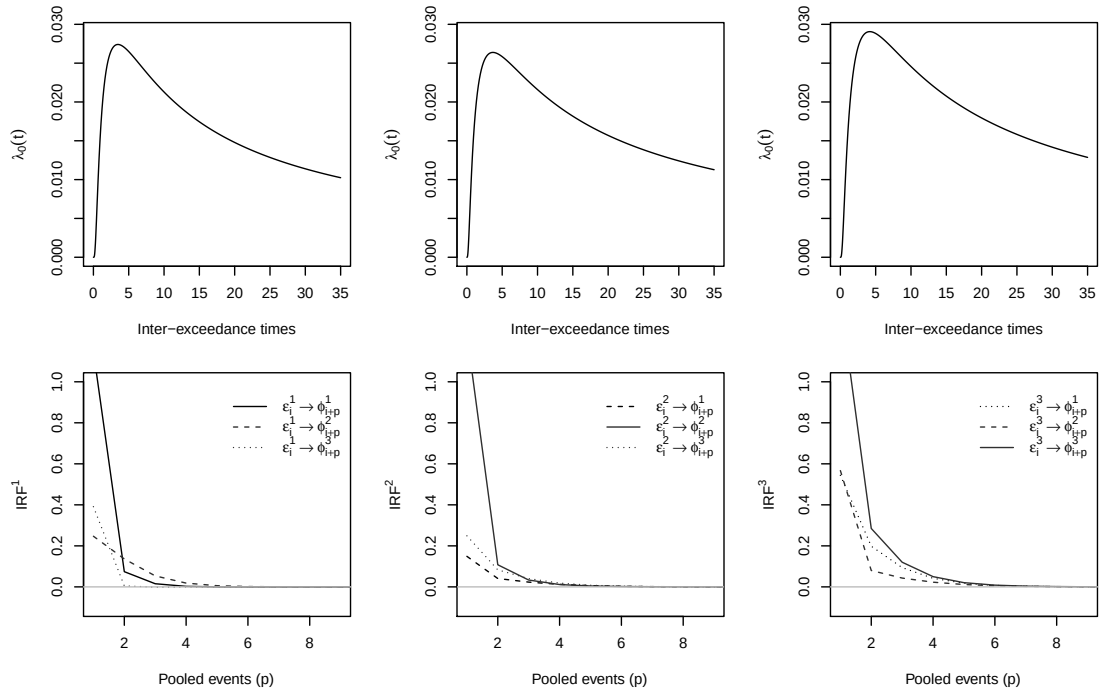

Figure 8: Top panel: Baseline hazard functions for the trivariate ACI-POT (1,1) model. From left to right the plots correspond to the baseline hazard functions of negative returns for the DAX, S&P 500, and FTSE 100 indices, respectively. Bottom panel: Impulse-response functions of innovations to the losses in the FTSE (right), DAX (middle), S&P 500 (left) markets for the ACI-POT model. Solid and dotted lines represent the response functions of losses to an innovation shock in the own and in the other markets, respectively.

the baseline functions around three to four exceedance periods, moreover, reflects underlying temporal clustering, making it more likely to observe a further extreme price movement just after a previous one than after a long period without exceedances. The tests on excess dispersion reveal no evidence of overdispersion and thus indicate that the flexibility provided by the baseline functions is sufficient to pick up the distributional properties of inter-exceedance times.

The estimated impulse responses shown in the bottom panel of Figure 8 indicate spillover effects in the sense that losses in one market increase the intensity of loss exceedances in other markets. Nevertheless, self-excitation effects in the own market tend to dominate cross-market effects. Overall, the impulse responses show similar patterns. This is likely due to periods of turmoil around the subprime mortage crisis, generating mechanisms of mutual excitation and, therefore, clustering of extreme events. This result is in line with studies of dependencies in extremes in international stock markets showing that extreme losses tend to affect several stock markets at the same time, creating co-movements and strong dependencies among their conditional

intensities (see, e.g., Poon et al., 2003 or Baltzer et al., 2008).

When interpreting cross-market effects, it should be taken into account that the U.S. market and the European markets do not open and close simultaneously. In particular, the U.S. market lags the European markets by 6 hours. Hence, any information occurring during the time when European markets are closed is incorporated in the daily return of the S&P500 index, while European indices can only react with a delay. Nevertheless, the current specification including all indices simultaneously provides a better fit to the data (and a better performance in terms of in-sample and out-of-sample VaR accuracy as discussed in the following section) than specifications including, for instance, the U.S. market in lagged form.[7] In accordance with the estimates for the bivariate specifications in Section 5.1, we find that the impact of the exceedance size on the conditional intensity of the ground process, as captured by the coefficient δ_m, is significant in all cases. Hence, extreme events in one series increase the conditional intensity for the next extreme event in the same series, but also in the other series. Finally, the estimates of the MEM process for the size of exceedances reveals a high persistence with coefficients $\beta_m < 1$. Hence, exceedance sizes are strongly autocorrelated and depend negatively on the length of past inter-exceedance waiting times, as reflected by the coefficient γ_m. The corresponding LM tests show that these effects are not only individually but also jointly significant.

Estimates of Hawkes-POT models reveal qualitatively similar results and are found in the Supplementary Appendix.

6 VaR and ES Forecasting using Dynamic Intensity models

An important advantage of a VaR-based risk assessment is the possibility of backtesting. Conversely, there is no consensus on how to backtest ES. Emmer et al. (2015) propose a framework to backtest ES based on a representation in terms of the integrated VaR,

$$\begin{aligned} ES_\alpha^t &= \frac{1}{1-\alpha}\int_\alpha^1 q_s^t\left(F_{z_t|\mathcal{H}_t}\right) ds \\ &\approx \frac{1}{4}\left[VaR_\alpha^t + VaR_{0.75\alpha+0.25}^t + VaR_{0.5\alpha+0.5}^t + VaR_{0.25\alpha+0.75}^t\right], \end{aligned} \tag{25}$$

[7] Corresponding results for such alternative specifications are available upon request from the authors.

where $q_s^t\left(F_{z_t|\mathcal{H}_t}\right)$ is approximated as in (21). This allows making use of backtesting techniques developed for VaR. In particular, if each of these confidence levels are successfully backtested, then, to a certain degree, the same is true for ES_α^t. In order to test the accuracy of VaR estimates, we utilize a battery of tests proposed in the literature, which are described in detail in Supplementary Appendix.

The first three tests are based on a binomial type test introduced by Christoffersen (1998): an unconditional coverage test (LR_{uc}), evaluating the expected fraction of exceptions (i.e., exceedances of the VaR), a test for the independence of exceptions (LR_{ind}), and a conditional coverage test (LR_{cc}), which is a combination of the latter two. Moreover, we implement the dynamic quantile tests proposed by Engle and Manganelli (2004) which rely on linear regressions. The first is the dynamic quantile hit test (DQ_{hit}), where de-meaned exceptions are regressed on their lags, while the second one, the dynamic quantile VaR (DQ_{VaR}) test, uses in addition the contemporaneous VaR estimates. Finally, we implement a loss measure V^{ES} which evaluates the potential loss between the forecasted ES ($\widehat{ES}_\alpha^t$) and the observed return Z_t, given that this return has exceeded the actual VaR

$$V^{ES} = \frac{\sum_{t=0}^{T}\left(Z_t - (-\widehat{ES}_\alpha^t)\right)\mathbf{1}_{\left\{Z_t<-\widehat{VaR}_\alpha^t\right\}}}{\sum_{t=0}^{T}\mathbf{1}_{\left\{Z_t<-\widehat{VaR}_\alpha^t\right\}}}.$$

An accurate estimate of ES should result in a low absolute value of this quantity. However, its weakness is that it depends on the accuracy of the preliminary VaR estimation, since only returns below the VaR are taken into account Embrechts et al. (2005). For instance, if the VaR estimates of a model do not generate any exceedances, this measure cannot be evaluated.

In order to assess the accuracy of the proposed approaches for the estimation and prediction of VaR and ES at different confidence levels, we estimate all models using the sample from January 3, 1994 to December 30, 2014. The estimated parameters are then used to compute one-day-ahead forecasts of the 99%-VaR and 97.5%- ES in the forecast period from January 2, 2015 to December 30, 2016.[8] The model parameters are not re-estimated each trading day since the additional information obtained from the additional day is negligible compared to the historical sample information and results would change only very mildly.

[8]The Basel Committee BCBS (2013) recommends changing the risk-based capital framework building on 99%-VaR to 97.5%-ES. These confidence levels are used in our empirical analysis.

Table 3 in the Appendix gives the test outcomes for the in-sample and out-of-sample VaR and ES estimates of the trivariate models jointly modeling extremes in all three index series. Recall that we need to estimate the VaR confidence levels $(0.975, 0.98125, 0.9875, 0.99375)$ in order to make use of the integral representation in (25) enabling us to backtest ES at the 97.5% level. For comparison purposes, we also report the VaR at the 99% confidence level.

In terms of in-sample predictive performance the ACI-POT(1,1) performs best. The explicit inclusion of mutual interactions between the point processes and the processes of exceedances results into a higher VaR accuracy. In fact, the *unrestricted* specification passes the tests 10% more often than the restricted specification. These findings are also confirmed by the out-of-sample analysis. The unrestricted specifications yield valid predictions in terms of VaR accuracy, with 90% of p-values exceeding 0.05. This proportion, however, is reduced to 81% when the restricted model is used.[9]

Figures 9 displays the estimated VaR and ES times series based on the ACI-POT(1,1) model. The figure also shows barcodes visualizing the extent to which extremes occur individually or jointly in the three series. We observe the highest VaR and ES estimates in all series during 2000–2002. During this period, the three stock market indices experienced large losses that were mainly due to the dot-com crash and the aftermath of the 9/11 terrorist attacks. After this period, the level of extreme risks declined until the subprime crisis in 2007, followed by the global crisis in 2008–2009.

As an additional evaluation metric, we analyze the difference between the predicted 99%-VaR and 97.5%-ES based on this approach. According to the Basel Committee BCBS (2013), 97.5%-ES is less sensitive to extreme events than the 99%-VaR and therefore should account for tail risk in a more comprehensive way. From a theoretical point of view, the ratio between both measures should be close to $\overline{ES}_{0.975}/\overline{VaR}_{0.99} \approx 0.4^{\xi}/(1-\xi)$, where ξ is the shape parameter of the GPD.[10] Since for all return series analyzed, we have $\xi > 0$, the ratio should be greater than one. Table 4 reports the time series average (through both estimation and backtesting periods) of this ratio for all models. The results indicate that the ratios are greater than one and nearly identical for all approaches, but are slightly higher in case of the ACI-POT(1,1) specification for the backtesting

[9] A VaR forecasting comparison with the Hawkes-POT model is provided in the Supplementary Appendix. We find that the ACI-POT model slightly outperforms the Hawkes-POT model not only in terms of in-sample VaR accuracy but also VaR accuracy.

[10] Note that $\frac{\overline{VaR}_{0.975}}{VaR_{0.99}} \approx 0.4^{\xi}$ and from (23) we know that $\frac{\overline{ES}_{0.975}}{\overline{VaR}_{0.975}} \approx \frac{1}{1-\xi}$.

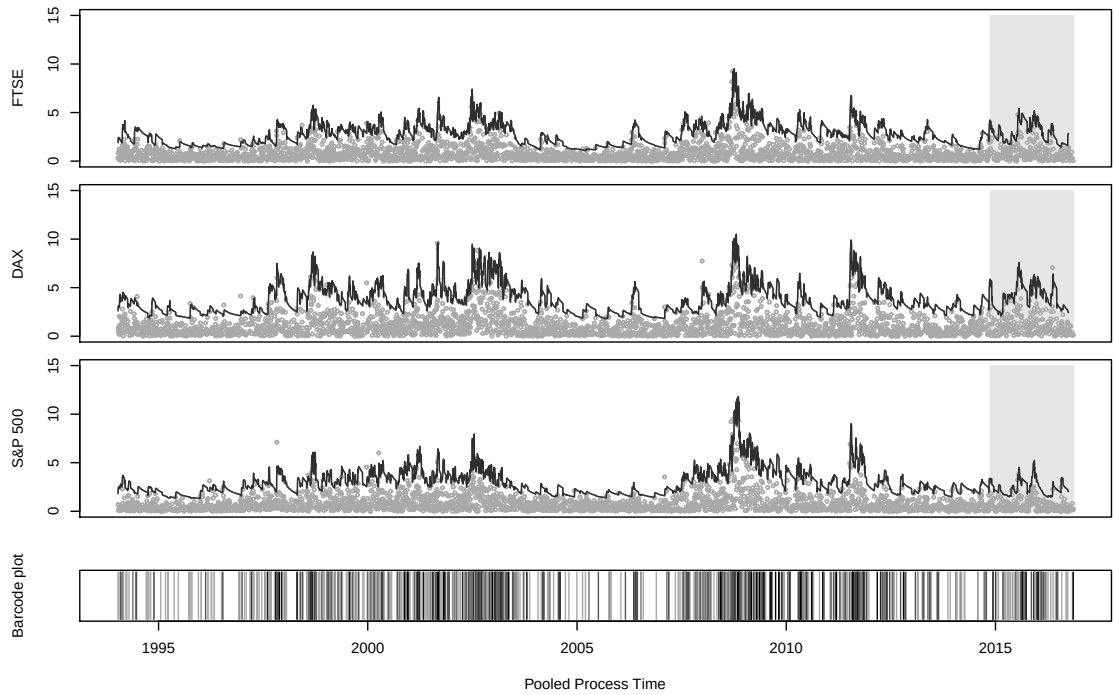

Figure 9: From top to bottom: estimated 99%-VaR (gray line) and 97.5%-ES (black line) for the trivariate ACI-POT model with generalized gamma hazard function applied to negative log returns of the FTSE 100, DAX, and S&P 500 indexes. In-sample period: January 3, 1994 to December 30, 2014. Out-of-sample period: January 2, 2015 to December 30, 2016 (marked by dark background). The bottom panel shows a barcode plot with light colors indicating extreme events in the FTSE 100, DAX, or S&P 500, and the mid-range dark colors indicating a simultaneous extreme event in any pair of negative log returns. The dark black color marks a simultaneous extreme event in all three negative log return series.

period.

Finally, to evaluate the sensitivity of our analysis with respect to the choice of the tail threshold u, we estimate the trivariate ACI-POT model for 100 different threshold levels u, ranging from the 90% to 94.999% quantile of returns. For each estimate, we then evaluate the in-sample and out-of-sample VaR accuracy. The results are provided in the Supplementary Appendix and document that our findings are widely stable with respect to the choice of u.

7 Conclusions

We propose a multivariate dynamic intensity framework to jointly model the occurrence of extreme observations (exceeding a certain threshold) in a multivariate time series of log returns. The event arrival is modeled as a MPP of exceedances where the marks are associated with the magnitude of (loss) exceedances. The major feature of these models is to allow for the clustering of the arrival of

extremes over both time and the cross-section and the clustering of the size of exceedances. This is achieved by combining a multivariate dynamic intensity process with a multiplicative error model based on a generalized Pareto distribution for the magnitude of exceedances. Both components are linked to allow for feedback effects between the arrival intensity of extremes and the size of exceedances above the threshold.

Empirical evidence based on the return series of the DAX, S&P500, and FTSE100 indices provides strong support for the models. We find significant evidence for (co-)cluster structures in extreme stock market losses, which are well-captured by the proposed approach. Furthermore, we demonstrate that the new models yield a good out-of-sample backtesting performance when they are applied to the prediction of VaR and ES.

We see it as a major advantage of the proposed framework that it can be easily extended in various directions and – depending on the chosen specification – is also tractable in higher dimensions. Consequently, it might be used as a valuable framework to analyze, for instance, systemic risk or tail dependencies.

References

Aït-Sahalia, Y., Cacho-Diaz, J., and Laeven, R. J. (2015). Modeling financial contagion using mutually exciting jump processes. *Journal of Financial Economics*, 177(3):585 – 606.

Aït-Sahalia, Y., Laeven, R. J., and Pelizzon, L. (2014). Mutual excitation in eurozone sovereign CDS. *Journal of Econometrics*, 183(2):151 – 167. Analysis of Financial Data.

Bacry, E., Dayri, K., and Muzy, J.-F. (2012). Non-parametric kernel estimation for symmetric hawkes processes. application to high frequency financial data. *The European Physical Journal B*, 85(5):1–12.

Bacry, E., Delattre, S., Hoffmann, M., and Muzy, J.-F. (2013). Modelling microstructure noise with mutually exciting point processes. *Quantitative Finance*, 13(1):65–77.

Bacry, E. and Muzy, J. (2014). Hawkes model for price and trades high-frequency dynamics. *Quantitative Finance*, 14(7):1147–1166.

Baltzer, M., Cappiello, L., Santis, R. A. D., and Manganelli, S. (2008). Measuring financial integration in new eu members states. *Occasional Papers Series, European Central Bank*, 81.

Bauwens, L. and Hautsch, N. (2006). Stochastic conditional intensity processes. *Journal of Financial Econometrics*, 4:450–493.

BCBS (2013). Fundamental review of the trading book: A revised market risk framework. *Basel Committee on Banking Supervision*.

Bowsher, C. G. (2007). Modelling security market events in continuous time: Intensity based, multivariate point process models. *Journal of Econometrics*, 141(2):876–912.

Byström, H. N. (2004). Managing extreme risks in tranquil and volatile markets using conditional extreme value theory. *International Review of Financial Analysis*, 13(2):133–152.

Campbell, J. Y. and Hentschel, L. (1992). No news is good news: An asymmetric model of changing volatility in stock returns. *Journal of Financial Economics*, 31(3):281–318.

Chang, L.-B., Geman, S., Hsieh, F., and Hwang, C.-R. (2013). Invariance in the recurrence of large returns and the validation of models of price dynamics. *Physical Review E*, 88(2):022116.

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

Chavez-Demoulin, V., Embrechts, P., and Sardy, S. (2014). Extreme-quantile tracking for financial time series. *Journal of Econometrics*, 181(1):44 –52.

Chavez-Demoulin, V. and McGill, J. (2012). High-frequency financial data modeling using hawkes processes. *Journal of Banking & Finance*, 36(12):3415–3426.

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

Cotter, J. and Dowd, K. (2006). Extreme spectral risk measures: An application to futures clearinghouse margin requirements. *Journal of Banking and Finance*, 30:3469–3485.

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

Davis, R. A., Mikosch, T., and Cribben, I. (2012). Towards estimating extremal serial dependence via the bootstrapped extremogram. *Journal of Econometrics*, 170(1):142–152.

Davis, R. A., Mikosch, T., et al. (2009). The extremogram: a correlogram for extreme events. *Bernoulli*, 15(4):977–1009.

Davison, A. and Smith, R. (1990). Models for exceedances over high thresholds. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 393–442.

Embrechts, P., Kaufmann, R., and Patie, P. (2005). Strategic long-term financial risks: Single risk factors. *Computational Optimization and Applications*, 32(1-2):61–90.

Emmer, S., Kratz, M., and Tasche, D. (2015). What is the best risk measure in practice? a comparison of standard measures. *Journal of Risk*, 18(2):31 –60.

Engle, R. (2002). New frontiers for arch models. *Journal of Applied Econometrics*, 17(5):425–446.

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.

Errais, E., Giesecke, K., and Goldberg, L. (2010). Affine point processes and portfolio credit risk. *SIAM Journal on Financial Mathematics*, 1(1):642–665.

Gresnigt, F., Kole, E., and Franses, P. H. (2015). Interpreting financial market crashes as earthquakes: A new early warning system for medium term crashes. *Journal of Banking & Finance*, 56:123 – 139.

Grothe, O., Korniichuk, V., and Manner, H. (2014). Modeling multivariate extreme events using self-exciting point processes. *Journal of Econometrics*, 182(2):269 – 289.

Hamidieh, K., Stoev, S., and Michailidis, G. (2009). On the estimation of the extremal index based on scaling and resampling. *Journal of Computational and Graphical Statistics*, 18(3):731–755.

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.

Hautsch, N. (2012). *Econometrics of financial high-frequency data*. Springer.

Hawkes, A. G. and Oakes, D. (1974). A cluster process representation of a self-exciting process. *Journal of Applied Probability*, pages 493–503.

Herrera, R. and Eichler, S. (2011). Extreme dependence with asymmetric thresholds: Evidence for the european monetary union. *Journal of Banking & Finance*, 35(11):2916 – 2930.

Herrera, R. and Schipp, B. (2014). Statistics of extreme events in risk management: The impact of the subprime and global financial crisis on the german stock market. *The North American Journal of Economics and Finance*, 29:218 – 238.

Jondeau, E. and Rockinger, M. (2003). Testing for differences in the tails of stock-market returns. *Journal of Empirical Finance*, 10(5):559 – 581.

Kehrle, K. and Peter, F. J. (2013). Who moves first? an intensity-based measure for information flows across stock exchanges. *Journal of Banking & Finance*, 37(5):1629 – 1642.

Longin, F. and Solnik, R. (2001). Extreme correlation of international equity markets. *Journal of Finance*, 56:649–676.

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.

Meyer, P.-A. (1971). Demonstration simplifiee d'un theoreme de knight. *Séminaire de probabilités de Strasbourg*, 5:191–195.

Ogata, Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical Association*, 83:9–27.

Olmo, J. (2005). Testing the existence of clustering in the extreme values. *Documentos de trabajo. Economic series (Universidad Carlos III. Departamento de Economía)*, 18(1).

Pickands, J. (1971). The two-dimensional poisson process and extremal processes. *Journal of applied Probability*, pages 745–756.

Poon, S.-H., Rockinger, M., and Tawn, J. (2003). Modelling extreme-value dependence in international stock markets. *Statistica Sinica*, 14:929 – 954.

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

Resnick, S. (2006). *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*. Springer, Berlin.

Russell, J. (1999). Econometric modeling of multivariate irregularly-spaced high-frequency data. *Manuscript, GSB, University of Chicago*.

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.

Smith, R. (1989). Extreme value analysis of environmetal time series: An application to trend detection in ground-level ozone. *Statistics Sience*, 4:367–393.

Model	ACI-POT (1,1)				ACI-POT (2,1)					ACI-POT (1,2)				
Log-return m	gains		losses			gains		losses			gains		losses	
	par	(p-value)	par	(p-value)		par	(p-value)	par	(p-value)		par	(p-value)	par	(p-value)
					Ground process									
a^1_{m1}	1.250	(0.000)	0.644	(0.000)	a^1_{m1}	1.320	(0.000)	0.556	(0.002)	a^1_{m1}	1.326	(0.000)	0.635	(0.002)
a^1_{m2}	-1.041	(0.000)	0.506	(0.000)	a^1_{m2}	-0.976	(0.000)	0.599	(0.000)	a^1_{m2}	-1.001	(0.000)	0.619	(0.000)
b^1_{m1}	0.755	(0.000)	0.224	(0.000)	b^1_{m1}	0.831	(0.000)	0.484	(0.002)	a^2_{m1}	-0.295	(0.200)	0.105	(0.620)
b^1_{m2}	0.208	(0.000)	0.644	(0.000)	b^1_{m2}	0.031	(0.867)	0.492	(0.000)	a^2_{m2}	0.440	(0.042)	-0.219	(0.274)
υ_m	9.121	(0.000)	3.081	(0.000)	b^2_{m1}	-0.073	(0.635)	-0.262	(0.082)	b^1_{m1}	0.821	(0.000)	0.123	(0.001)
σ_m	5.139	(0.000)	1.484	(0.000)	b^2_{m2}	0.146	(0.377)	0.192	(0.148)	b^1_{m2}	0.086	(0.053)	0.834	(0.000)
Q_m	-1.607	(0.023)	-0.367	(0.022)	υ_m	9.284	(0.000)	3.106	(0.000)	υ_m	9.951	(0.000)	3.086	(0.000)
δ_m	-0.009	(0.894)	0.121	(0.077)	σ_m	5.271	(0.001)	1.479	(0.000)	σ_m	5.875	(0.007)	1.495	(0.000)
					Q_m	-1.646	(0.038)	-0.346	(0.030)	Q_m	-1.831	(0.056)	-0.384	(0.018)
					δ_m	-0.005	(0.934)	-0.112	(0.144)	δ_m	-0.006	(0.930)	-0.089	(0.294)
LL_1	-2535.322					-2541.610					-2529.575			
					Ground process residuals									
Mean (ε_m)	0.003		-0.076			0.008		-0.032			0.000		-0.066	
$\tilde{\sigma}_\varepsilon$	1.212		1.250			1.134		1.075			1.215		1.167	
Excess.dis	1.588	(0.112)	1.180	(0.238)		2.110	(0.035)	1.152	(0.249)		3.521	(0.000)	2.667	(0.008)
LB_ε	5.652	(0.017)	0.566	(0.452)		10.189	(0.070)	20.235	(0.001)		5.489	(0.359)	11.581	(0.359)
					Mark process									
w^m	0.054	(0.003)	0.069	(0.000)										
ρ_m	0.084	(0.000)	0.115	(0.000)										
β_m	0.799	(0.000)	0.778	(0.000)										
γ_m	-0.003	(0.048)	-0.002	(0.058)										
ξ_m	0.088	(0.069)	-0.017	(0.748)										
LL_2	-335.863		-355.843											
	Diagnostics													
LM_{cross}	6362.754	(0.000)												
LM_{int}	22.232	(0.000)												
LM_{mark}	10.390	(0.001)	12.199	(0.000)										
AIC	6506.057					6526.633					6502.563			
BIC	6630.090					6669.748					6645.678			

Table 1: Estimates of the bivariate ACI-POT models for extreme events of losses and gains of a portfolio based on the log-returns of the FTSE 100, DAX and S&P 500 indexes from January 3, 1994 to December 30, 2014. P-values are in parentheses. LL_1corresponds to the log-likelihood of the ACI part, while LL_2 to the POT part. Mean (ε_m): mean of residuals, $\tilde{\sigma}_\varepsilon$ standard deviation of the residuals, LB_ε: Ljung-Box statistic, *Excess.dis:* excess dispersion test. p-values for the maximum likelihood estimates are obtained from a two-sided t-test.

Model	ACI-POT (1,1)						*Restricted* ACI-POT (1,1)							ACI-POT (1,2)					
Log-return m	FTSE		DAX		S&P500			FTSE		DAX		S&P500			FTSE		DAX		S&P500	
	par	(p-value)	par	(p-value)	par	(p-value)		par	(p-value)	par	(p-value)	par	(p-value)		par	(p-value)	par	(p-value)	par	(p-value)
Ground process																				
a^1_{m1}	1.148	(0.000)	0.149	(0.224)	0.544	(0.000)	a^1_{m1}	0.898	(0.000)	1.049	(0.000)	1.195	(0.000)	a^1_{m1}	1.233	(0.000)	0.160	(0.170)	0.574	(0.000)
a^1_{m2}	0.247	(0.033)	1.235	(0.000)	0.567	(0.000)	b^1_{m1}	0.677	(0.000)	0.596	(0.000)	0.633	(0.000)	a^1_{m2}	0.325	(0.003)	1.272	(0.000)	0.592	(0.000)
a^1_{m3}	0.391	(0.002)	0.249	(0.042)	1.402	(0.000)	υ_m	2.341	(0.000)	2.385	(0.000)	2.623	(0.000)	a^1_{m3}	0.396	(0.001)	0.230	(0.056)	1.384	(0.000)
b^1_{m1}	0.207	(0.052)	-0.007	(0.937)	0.188	(0.061)	σ_m	1.515	(0.000)	1.483	(0.000)	1.527	(0.000)	a^2_{m1}	6.525	(0.000)	-4.672	(0.000)	-2.208	(0.000)
b^1_{m2}	0.201	(0.145)	0.286	(0.004)	-0.069	(0.519)	Q_m	-0.716	(0.000)	-0.549	(0.004)	-0.609	(0.001)	a^2_{m2}	7.150	(0.000)	-5.078	(0.000)	-2.646	(0.000)
b^1_{m3}	-0.122	(0.301)	0.040	(0.645)	0.496	(0.000)								a^2_{m3}	7.194	(0.000)	-5.458	(0.000)	-2.091	(0.001)
υ_m	3.329	(0.000)	3.446	(0.000)	3.273	(0.000)								b^1_{m1}	-7.122	(0.000)	4.179	(0.000)	2.896	(0.000)
σ_m	1.897	(0.000)	1.810	(0.000)	1.688	(0.000)								b^1_{m2}	-7.723	(0.000)	4.576	(0.000)	3.144	(0.000)
Q_m	-0.759	(0.001)	-0.518	(0.021)	-0.532	(0.009)								b^1_{m3}	-8.088	(0.000)	4.786	(0.000)	3.228	(0.000)
δ_m	0.325	(0.000)	0.261	(0.000)	0.181	(0.001)								υ_m	3.641	(0.000)	3.759	(0.000)	3.422	(0.000)
														σ_m	2.029	(0.000)	1.943	(0.000)	1.765	(0.000)
														Q_m	-0.785	(0.001)	-0.577	(0.016)	-0.579	(0.007)
														δ_m	0.307	(0.000)	0.227	(0.002)	0.148	(0.010)
LL_1			-4114.845							-4200.991					-4099.062					
Spr			0.422							0.677					0.887					
Ground process residuals																				
Mean (ε_m)	0.003		-0.006		0.006			0.019		0.011		0.039			0.001		-0.010		-0.008	
$\tilde{\sigma}_\varepsilon$	1.076		1.045		1.047			1.019		1.037		1.010			1.059		1.026		1.037	
Excess.dis	1.205	(0.228)	0.696	(0.486)	0.726	(0.627)		0.295	(0.768)	0.571	(0.568)	0.147	(0.883)		0.925	(0.355)	0.401	(0.688)	0.568	(0.570)
LB_ε	0.900	(0.343)	0.063	(0.802)	3.381	(0.066)		0.642	(0.423)	0.720	(0.396)	4.318	(0.038)		0.979	(0.323)	0.515	(0.473)	3.478	(0.062)
Mark process																				
w^m	0.062	(0.001)	0.111	(0.000)	0.068	(0.001)														
ρ_m	0.100	(0.000)	0.077	(0.001)	0.115	(0.000)														
β_m	0.733	(0.000)	0.739	(0.000)	0.768	(0.000)														
γ_m	-0.003	(0.010)	-0.006	(0.006)	-0.003	(0.115)														
ξ_m	0.056	(0.221)	0.019	(0.684)	0.014	(0.753)														
LL_2	-370.704		-487.889		-385.169															
Diagnostics																				
LM_{cross}	514.946	(0.000)																		
LM_{int}	298.676	(0.000)																		
LM_{mark}	27.595	(0.000)	9.897	(0.000)	47.116	(0.000)														
AIC	10807.21							10949.51		.					10793.65					
BIC	11042.71							11106.50							11076.24					

Table 2: Estimates of trivariate ACI-POT models applied to the negative log-returns of the FTSE 100, DAX and S&P 500 indexes from January 3, 1994 to December 30, 2014. P-values are in parentheses. LL_1corresponds to the log-likelihood of the ACI part, while LL_2 to the POT part. Mean (ε_m): mean of residuals, $\tilde{\sigma}_\varepsilon$ standard deviation of the residuals, LB_ε: Ljung-Box statistic, *Excess.dis:* excess dispersion test. p-values for the maximum likelihood estimates are obtained from a two-sided t-test.

VaR in-sample

		ACI-POT (1,1)							*Restricted* ACI-POT (1,1)							ACI-POT (1,2)						
Stock Index	α	excep.	LRuc	LRind	LRcc	DQ_{hit}	DQ_{VaR}	V^{ES}	excep.	LRuc	LRind	LRcc	DQ_{hit}	DQ_{VaR}	V^{ES}	excep.	LRuc	LRind	LRcc	DQ_{hit}	DQ_{VaR}	V^{ES}
FTSE	0.975	121	0.52	0.10	0.20	0.10	0.24	0.18	116	0.27	0.18	0.22	0.19	0.42	0.14	124	0.71	0.28	0.52	0.28	0.36	0.18
	0.98125	82	0.14	0.05	0.05	0.06	0.02	0.12	99	0.76	0.18	0.39	0.18	0.39	0.24	80	0.09	0.05	0.03	0.05	0.02	0.11
	0.9875	49	0.05	0.50	0.11	0.50	0.10	0.05	60	0.61	0.20	0.38	0.20	0.14	0.15	52	0.12	0.56	0.25	0.56	0.17	0.08
	0.99	37	0.04	0.27	0.06	0.28	0.10	0.01	45	0.37	0.42	0.48	0.42	0.08	0.10	42	0.18	0.40	0.29	0.41	0.33	0.09
	0.99375	27	0.36	0.59	0.57	0.59	0.53	0.10	26	0.27	0.61	0.48	0.61	0.26	0.04	24	0.14	0.63	0.30	0.64	0.87	0.09
DAX	0.975	120	0.47	0.91	0.76	0.91	0.44	0.34	123	0.65	0.98	0.90	0.98	0.07	0.23	117	0.31	0.84	0.59	0.84	0.19	0.28
	0.98125	78	0.05	0.86	0.16	0.86	0.26	0.26	92	0.67	0.79	0.88	0.79	0.04	0.22	79	0.07	0.17	0.07	0.17	0.02	0.24
	0.9875	39	0.00	0.31	0.00	0.31	0.03	0.01	56	0.30	0.65	0.53	0.65	0.46	0.23	38	0.00	0.29	0.00	0.29	0.01	-0.01
	0.99	28	0.00	0.15	0.00	0.15	0.12	-0.07	38	0.05	0.29	0.09	0.29	0.27	0.07	28	0.00	0.15	0.00	0.15	0.03	-0.14
	0.99375	17	0.00	0.74	0.01	0.74	0.11	-0.22	23	0.09	0.09	0.06	0.09	0.02	-0.06	19	0.01	0.71	0.04	0.71	0.06	-0.14
S&P 500	0.975	114	0.20	0.26	0.23	0.27	0.21	0.12	111	0.12	0.03	0.03	0.03	0.04	0.07	113	0.17	0.27	0.21	0.28	0.24	0.09
	0.98125	82	0.14	0.10	0.09	0.11	0.13	0.11	89	0.46	0.08	0.16	0.08	0.02	0.07	87	0.34	0.08	0.14	0.09	0.06	0.12
	0.9875	53	0.15	0.29	0.21	0.30	0.00	0.00	63	0.90	0.21	0.45	0.21	0.01	0.05	54	0.19	0.28	0.24	0.29	0.01	0.03
	0.99	42	0.18	0.40	0.29	0.41	0.02	-0.01	49	0.75	0.33	0.59	0.33	0.01	-0.02	42	0.18	0.40	0.29	0.41	0.01	-0.02
	0.99375	26	0.27	0.61	0.48	0.61	0.50	0.02	30	0.72	0.55	0.78	0.55	0.36	-0.04	23	0.09	0.65	0.22	0.65	0.61	-0.10

VaR out-sample

		ACI-POT (1,1)							*Restricted* ACI-POT (1,1)							ACI-POT (1,2)						
Stock Index	α	excep.	LRuc	LRind	LRcc	DQ_{hit}	DQ_{VaR}	V^{ES}	excep.	LRuc	LRind	LRcc	DQ_{hit}	DQ_{VaR}	V^{ES}	excep.	LRuc	LRind	LRcc	DQ_{hit}	DQ_{VaR}	V^{ES}
FTSE	0.975	12	0.89	0.03	0.10	0.03	0.03	0.37	12	0.89	0.31	0.59	0.32	0.07	0.26	12	0.89	0.03	0.10	0.03	0.00	0.33
	0.98125	6	0.34	0.06	0.11	0.06	0.05	0.24	7	0.56	0.09	0.19	0.09	0.00	0.10	8	0.82	0.00	0.02	0.01	0.00	0.35
	0.9875	4	0.43	0.02	0.05	0.02	0.05	0.33	5	0.74	0.04	0.11	0.04	0.01	0.20	4	0.43	0.02	0.05	0.02	0.04	0.24
	0.99	3	0.42	0.84	0.71	0.84	0.94	0.43	3	0.42	0.84	0.71	0.84	0.17	0.14	4	0.77	0.02	0.07	0.02	0.04	0.39
	0.99375	1	0.20	0.95	0.44	0.95	0.41	0.48	1	0.20	0.95	0.44	0.95	1.00	-0.44	1	0.20	0.95	0.44	0.95	0.51	0.38
DAX	0.975	12	0.89	0.31	0.59	0.32	0.07	0.18	13	0.67	0.38	0.62	0.39	0.05	0.30	13	0.67	0.37	0.61	0.38	0.01	0.10
	0.98125	8	0.82	0.59	0.85	0.60	0.13	0.18	8	0.82	0.59	0.85	0.60	0.03	0.19	10	0.65	0.50	0.72	0.51	0.10	0.16
	0.9875	4	0.43	0.79	0.71	0.79	0.77	0.00	5	0.74	0.74	0.90	0.74	0.08	-0.07	5	0.74	0.74	0.90	0.74	0.63	0.07
	0.99	3	0.42	0.84	0.71	0.84	0.94	-0.09	3	0.42	0.84	0.71	0.84	0.42	-0.37	4	0.77	0.79	0.93	0.79	0.67	0.13
	0.99375	2	0.58	0.89	0.85	0.90	0.92	-0.03	3	0.94	0.84	0.98	0.84	0.43	0.14	2	0.58	0.89	0.85	0.90	0.91	-0.05
S&P 500	0.975	7	0.15	0.09	0.08	0.09	0.00	-0.14	7	0.15	0.09	0.08	0.09	0.06	-0.16	8	0.27	0.12	0.16	0.12	0.06	-0.09
	0.98125	6	0.34	0.06	0.11	0.06	0.01	-0.07	6	0.34	0.06	0.11	0.06	0.04	-0.22	6	0.34	0.06	0.11	0.06	0.01	-0.16
	0.9875	5	0.74	0.04	0.11	0.04	0.01	-0.03	5	0.74	0.04	0.11	0.04	0.00	-0.20	5	0.74	0.04	0.11	0.04	0.01	-0.11
	0.99	4	0.77	0.79	0.93	0.79	0.16	-0.05	5	0.86	0.04	0.11	0.04	0.00	-0.05	5	0.86	0.04	0.11	0.04	0.01	0.02
	0.99375	4	0.53	0.79	0.79	0.79	0.18	0.22	4	0.53	0.79	0.79	0.79	0.00	0.07	4	0.53	0.79	0.79	0.79	0.09	0.13

Table 3: VaR accuracy test for ACI-POT approaches, for the in-sample period (from January 3, 1994 to December 30, 2014) and the backtesting period (January 2, 2015 to December 30, 2016). Entries in the rows are the number of observations exceeding the VaR level ("exceptions") and p-values of the corresponding accuracy tests.

	ACI-POT (1,1)		*Restricted* ACI-POT (1,1)		ACI-POT (1,2)		Theoretical
	in-sample	out-sample	in-sample	out-sample	in-sample	out-sample	
FTSE	1.010	1.046	1.033	1.019	1.013	1.014	1.006
DAX	1.022	1.050	1.030	1.029	1.022	1.024	1.002
S&P500	1.021	1.052	1.036	1.028	1.021	1.023	1.001

Table 4: Ratio between both measures of risk for all specifications ($\overline{ES}_{0.975}/\overline{VaR}_{0.99}$) using the mean of these risk measures for the whole period (in-sample and backtesting periods).