Modeling Multivariate Extreme Events Using Self-Exciting Point Processes

Oliver Grothe[a], Volodymyr Korniichuk[b,*], Hans Manner[a]

[a] *University of Cologne, Department of Economic and Social Statistics*
[b] *University of Cologne, Cologne Graduate School*

Abstract

We propose a model that can capture the typical features of multivariate extreme events observed in financial time series, namely clustering behavior in magnitudes and arrival times of multivariate extreme events, and time-varying dependence. The model is developed in the framework of the peaks-over-threshold approach in extreme value theory and relies on a Poisson process with self-exciting intensity. We discuss the properties of the model, treat its estimation, deal with testing goodness-of-fit, and develop a simulation algorithm. The model is applied to return data of two stock markets and four major European banks.

Keywords: Time Series, Peaks Over Threshold, Hawkes Processes, Extreme Value Theory.
JEL classification: C32, C51, C58, G15.

1. Introduction

A characteristic feature of financial time series is their disposition towards sudden extreme jumps. As an empirical illustration consider Figure 1, which shows occurrence times and magnitudes of exceedances of MSCI-USA and MSCI-EU indices' negated returns over a high quantile of their distributions. It is apparent from the figure that both occurrence times and magnitudes of the exceedances resemble a certain clustering behavior, namely, large negative returns tend to be followed by large ones and vice versa. Additionally, this clustering behavior is observed not only in time but also across the markets, which is manifested, among others, in the occurrence of joint exceedances. This synchronization of large returns' occurrences may be attributed to the information transmission across financial markets, see, for example, Wongswan (2006), where, based on high-frequency data, international transmission of economic fundamental announcements is studied on the example of the US, Japan, Korean and Thai equity markets. Other channels of the informational transmission are described in Bekaert, Ehrmann, Fratzscher, and Mehl (2012), where, in particular, the authors provide a strong support for the validity of the "wake-up call" hypothesis, which states that a local crisis in one market may prompt investors to reexamine their views on the vulnerability of other market segments, which in turn may cause spreading of the local shock to other markets. Clustering of extreme events may also be caused by intra-day volatility spillovers both within one market and across different markets, see Golosnoy, Gribisch, and Liesenfeld (2012) for a recent study of this topic. In general, it is not clear whether the joint exceedances are triggered by a jump in one component or just caused by a common factor – both scenarios occur in financial markets and are interesting to analyze. The behavior of extreme asset-returns presents an important topic for the risk management research and is of great relevance especially in view of the latest financial crisis.

The problem of modelling jumps or exceedances above high thresholds in asset-returns was considered in many papers. For example, Bollerslev, Todorov, and Li (2013) approach this problem by partitioning jumps into idiosyncratic and systemic components, and by further direct modeling of the jumps' distributional properties based on the results of extreme value theory. Aït-Sahalia, Cacho-Diaz, and Laeven (2011) propose a Hawkes jump-diffusion model in which self-exciting processes (with mutual excitement) are used for modelling clustering of extreme events both in time and

*Corresponding author
Email addresses: grothe@statistik.uni-koeln.de (Oliver Grothe), korniichuk@wiso.uni-koeln.de (Volodymyr Korniichuk), manner@statistik.uni-koeln.de (Hans Manner)

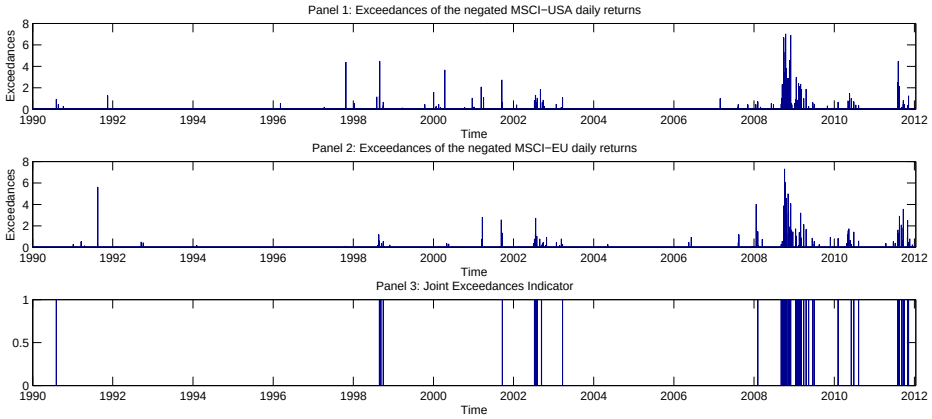

Figure 1: Exceedances of negated MSCI-USA (Panel 1) and MSCI-EU (Panel 2) daily log-returns over the respective 0.977th quantiles. Bar plot indicating times of the joint exceedances (Panel 3).

across assets. That paper develops a feasible estimation approach based on the generalized method of moments and provides a strong evidence of self-excitation and asymmetric cross-excitation in the financial markets. Modeling multivariate exceedances above high thresholds is also a topic of intensive research in extreme value theory. For example, it was shown in the literature that the multivariate generalized Pareto distribution is the natural distribution for multivariate extreme exceedances, Smith, Tawn, and Coles (1997) and Rootzen and Tajvidi (2006). Recent studies considering estimation of the probability that a random vector falls in some remote region are Einmahl, de Haan, and Krajina (2013), Drees and de Haan (2012). Note, however, that those methods are not directly applicable for the case when the extremes are clustering in time. Extensive references for EVT methods can be found in de Haan and Ferreira (2006) or Resnick (2007). Studies that are related to modeling clusters in financial data are Bowsher (2007) who introduces a new class of generalized Hawkes process (including non-linear models) and studies by its bivariate version the transaction times and mid-quote changes at high-frequency data for a NYSE stock, as well as Errais, Giesecke, and Goldberg (2010) who employ self-exciting processes for modeling portfolio credit risk, in particular, for valuation of credit derivatives.

Considering the recent developments in modelling extreme asset-returns, there is still a demand for a model that can provide insights into the temporal- and cross-dependence structure of multivariate extreme events in view of their clustering and specific dependence structure in the tails of (multivariate) distributions. In this paper we develop a model that can fill this gap. Working in the framework of marked self-exciting point processes and extreme value theory (EVT), we model multivariate extreme events as a univariate point process being constructed as a superposition of marginal extreme events. For modeling the marginal processes of exceedances we revise the existing specification of the univariate self-exciting peaks-over-threshold model of Chavez-Demoulin, Embrechts, and Nešlehová (2006) and McNeil, Frey, and Embrechts (2005), which is able to cope with the clustering of extremes (in both times and magnitudes) in the univariate case. After this revision, we are able to formulate stationarity conditions, not discussed in the literature before, and to analyze the distributional properties of the model. This constitutes a separate contribution of this paper.

We show that the only way how the marginal rates can be coupled into the multivariate rate of the superposed process is through the exponent measure of an extreme value copula. The copula used for the construction of the multivariate rate follows naturally from the EVT arguments, and is the same extreme value copula that governs the (conditional) multivariate distribution of the marginal exceedances at the same point of time. This result provides an integrated approach to modelling occurrence times and sizes of multivariate extreme events, because those two characteristics are inherently connected. Furthermore, the results provides insight into the dependence between point processes that jointly subject to EVT. This is in contrast to alternative approaches in the literature, where the dependence between marginal point processes is incorporated through an affine mutual excitement, see, for example, Aït-Sahalia, Cacho-Diaz, and Laeven (2011), Em-

brechts, Liniger, and Lin (2011), and magnitudes of the jumps (if considered) are modelled in a separate way.

Concerning the advantages of our method, it is worth noting that we use the data explicitly only above a high threshold. This allows us to leave the time series model for the non-extreme parts of the data unspecified. We consider the dependence structure of multivariate exceedances only in regions where the results from multivariate extreme value theory (MEVT) are valid. Furthermore, the MEVT enables us to extrapolate exceedance probabilities far into remote regions of the tail where hardly any data is available. With such a model we are able to extract the probabilities of arbitrary combinations of the dimensions in any sufficiently remote region. Since the model captures clustering behavior in (multivariate) exceedances, and accounts for the fact that not only times but also sizes of exceedances may trigger subsequent extreme events, the model provides asymmetric influences of marginal exceedances so that spill-over and contagion effects in financial market may be analyzed. This model may be of great interest for risk management purposes. For example, we can estimate the probabilities that from a portfolio of, say, d assets, a certain subset fall in a remote (extreme) set conditioned on the event that some other assets (or at least one them) from that portfolio take extreme values at the same point of time. We shortly discuss other possible risk management applications of the model and provide real data examples.

To estimate our proposed model, we derive the closed form likelihood function and describe the goodness-of-fit and simulation procedures. As noted earlier, our model treats a multivariate extreme exceedance as a realization of a univariate point process. This property is advantageous for the estimation, because, as it is mentioned in Bowsher (2007), there are currently no results concerning the properties of the maximum likelihood estimation (MLE) for multivariate point processes. For the univariate case, on the other hand, it is shown in Ogata (1978), that under some regularity conditions, the MLE for a stationary, simple point process is consistent and asymptotically normal. Inspired by Aït-Sahalia, Cacho-Diaz, and Laeven (2011), we consider also the model estimation based on method of moments, which, however, seems to underperform the MLE in the case of our model. The reason for this may lie in both the choice of moment conditions and in the fact that all moment conditions are based on the goodness-of-fit statistics, which cannot be directly calculated from the sample independently from the unknown parameters of the models.

In the empirical part of the paper, we apply our model to study extreme negative returns on the financial markets (USA, Europe) and in the European banking sector (Deutsche Bank, RBS, HSBC, and UBS). The results of goodness-of-fit demonstrate a reasonable fit of the model and suggest an empirical importance of the self-exciting feature for modeling both occurrence times, magnitudes, and interdependencies of the extreme returns. We find that conditional multivariate distributions of the returns are close to symmetric with the strength of dependence strongly responding to individual jumps. Despite the symmetrical structure of the distribution, there are still asymmetric effects coming from the self-exciting structure of the conditional marginal distributions of the exceedances' magnitudes. This self-exciting structure provides also a natural way how to model time-varying volatility of the magnitudes and, hence, their heavy tails.

The rest of the paper is structured as follows. The model and its properties are derived in Section 2. In Section 3 we describe estimation of the model, along with the goodness-of-fit and simulation procedures. Section 4 presents applications of the model to financial data and Section 5 concludes. Finally, some of the proof are relegated to the Appendix.

2. Model

The major challenges in constructing the model presented in this section are twofold. First, the model should capture the distinctive features of multivariate extreme events typically observed in financial markets, namely, clustering and spillover effects. Second, the model should be able to account for the specific distributional properties of magnitudes of extreme observations (i.e., for the distributions over the threshold). For both reasons, our model is developed in the framework of extreme value theory and marked point processes.

Throughout the text we use the following notation. Consider a random vector $\mathbf{X}_t = (X_{1,t}, \ldots, X_{d,t})$ which may, e.g., represent daily (negated) log-returns of d equities at time t. By $\mathbf{u} = (u_1, \ldots, u_d)$, the *initial threshold*, we denote a vector with components relating to sufficiently high quantiles of the marginal distributions of $\mathbf{X}_t$. We focus on the occurrence times as well as the magnitudes of multivariate extreme observations, which we define as situations when $\mathbf{X}_t$ exceeds $\mathbf{u}$ in at least one

component. Under a i-th marginal extreme event we understand the situation when $X_{i,t} > u_i$. We refer to such extreme events as marginal *exceedances* and characterise them by occurrence times $T_{i,1}, T_{i,2}, \ldots$ and magnitudes (the *marks*) of realizations $\tilde{X}_{i,1}, \tilde{X}_{i,2}, \ldots$, i.e., $\tilde{X}_{i,k} = X_{i,T_{i,k}}$. The history that includes both the times and magnitudes of exceedances of $(X_{i,s})_{s<t}$ above u_i will be denoted as $\mathbb{H}_{i,t}$ and the combined history over all marginal exceedances is denoted as $\mathbb{H}_t = \bigcup_{i=1}^{d} \mathbb{H}_{i,t}$.

This section is structured as follows. Section 2.1 deals with the univariate self-exciting peaks-over-threshold model, which is the basis for our multivariate model developed in Section 2.2. Section 2.3 provides some properties of the multivariate model.

2.1. Univariate model

This section deals with the univariate self-exciting peaks-over-threshold model. After a short review of this model, we reconsider some parts of its construction to enrich it with some new useful properties. In particular, we suggest a new specification for the impact function which, contrary to its existing specification, provides an intuitively reasonable mechanism how past exceedances trigger the future ones (Section 2.1.2), allows us to set a stationarity condition and to develop some distributional properties of the univariate model (Section 2.1.3). Finally, in Section 2.1.4 we consider the relationship of the univariate self-exciting peaks-over-threshold model to the general framework of the extreme value theory.

2.1.1. The self-exciting POT model

The basic setup to model univariate exceedances is to assume independent and identically distributed (iid) data and to use a peaks-over-threshold (POT) model developed in Davison and Smith (1990) and Leadbetter (1991). In the framework of EVT, the POT model is based on the asymptotic behavior of the threshold exceedances for iid or stationary data if these are in the maximum domain of attraction of some extreme value distribution. If the threshold is high enough, then the exceedances occur in time according to a homogeneous Poisson process and the mark sizes are independently and identically distributed according to the generalized Pareto distribution (GPD).

The *self-exciting POT model* presented in Chavez-Demoulin, Davison, and McNeil (2005) extends the standard set-up of the POT model by allowing for temporal dependence between extreme events. This temporal dependence is introduced into the model by modelling the rate of occurrences in the standard POT method with self-exciting processes, see Hawkes (1971).

Definition (Self-exciting point process) A point process $N(t)$, representing the cumulative number of events up to time t, is called a (linear) self-exciting process with the conditional rate $\tau(t)$, if

$$\mathrm{P}\left(N(t+\Delta) - N(t) = 1 \mid \mathbb{H}_t\right) = \tau(t)\Delta + o(\Delta), \qquad \mathrm{P}\left(N(t+\Delta) - N(t) > 1 \mid \mathbb{H}_t\right) = o(\Delta)$$

with

$$\tau(t) = \tau + \psi \int_{-\infty}^{t} c\left(\tilde{X}_s\right) g\left(t-s\right) dN(s), \qquad \tau > 0, \psi \geq 0,$$

where $\tilde{X}_s$ indicates the event's mark at time s. The impact function $c(\cdot)$ determines the contribution of events to the conditional rate and the decay function $g(\cdot)$ determines the rate how an influence of events decays in time. When no mark is associated with the event $c\left(\tilde{X}_s\right) \equiv 1$.

Choices of impact and decay functions are discussed in Section 2.1.2. The self-exciting POT is further extended in McNeil, Frey, and Embrechts 2005, where the temporal dependence is incorporated also into the conditional distribution of the marks, i.e., also the distribution of the marks depends on past information. We refer to this model as the *self-exciting POT model with predictable marks* (SE-POT). For convenience and consistency of notation we present the model using subindizes $i = 1 \ldots d$ which will later refer to the dimensions of our multivariate model. In the SE-POT model, the rate of crossing the initial threshold u_i is modelled by a self exciting point process where the rate is parametrized as

$$\tau_i(t, u_i) = \tau_i + \psi_i v_i^*(t), \qquad \tau_i > 0, \psi_i \geq 0, \tag{1}$$

with

$$v_i^*(t) = \int_{-\infty}^{t} c_i\left(X_{i,s}\right) g_i\left(t-s\right) dN_i(s), \tag{2}$$

where again $c_i(\cdot)$ and $g_i(\cdot)$ denote, respectively, the impact and decay functions, and $N_i(s)$ is a counting measure of i-th margin exceedances.

Additionally, the excesses over the threshold u_i are now assumed to follow the GPD with shape parameter ξ_i and time varying scale parameter $\beta_i + \alpha_i v^*(t)$. In particular, for $x_i > u_i$,

$$\mathrm{P}\left(X_{i,t} \leq x_i \mid X_{i,t} > u_i, \mathbb{H}_{i,t}\right) = 1 - \left(1 + \xi_i \frac{x_i - u_i}{\beta_i + \alpha_i v_i^*(t)}\right)^{-1/\xi_i} =: F_{i,t}\left(x_i\right), \qquad \beta_i > 0, \alpha_i \geq 0. \tag{3}$$

This distribution covers the cases of Weibull ($\xi_i < 0$), Gumbel ($\xi_i = 0$) and Fréchet ($\xi_i > 0$) tails, corresponding to distributions with finite endpoints, light tails, and heavy tails, respectively. For $\xi_i = 0$, the distribution function in (3) should be interpreted as $F_{i,t}\left(x_i\right) = 1 - e^{-x_i}$. Finally, due to the GPD as the conditional distribution of the marks, the conditional rate of exceeding a higher threshold $x_i \geq u_i$ scales in the following way

$$\tau_i(t, x_i) = \tau_i(t, u_i)\left(1 + \xi_i \frac{x_i - u_i}{\beta_i + \alpha_i v_i^*(t)}\right)^{-1/\xi_i}, \qquad x_i \geq u_i, \tag{4}$$

where $\tau_i(t, u_i)$ is the rate of crossing the initial threshold u_i given by Equation (1). The conditional rate $\tau_i(t, x_i)$ explicitly describes the conditional distribution of times of exceedances above any threshold $x_i \geq u_i$ in the following way.

$$\mathrm{P}\left(T_{i,k+1}\left(x_i\right) \leq t \mid \mathbb{H}_{i,T_{i,k}(x_i)}\right) = 1 - \exp\left(-\int_{T_{i,k}(x_i)}^{t} \tau_i(s, x_i) ds\right), \qquad t \geq T_{i,k}\left(x_i\right), \tag{5}$$

where $T_{i,k}\left(x_i\right)$ denotes (random) time of the k-th exceedance of $\left(X_{i,s}\right)_{s \in \mathbb{R}}$ above x_i. The above relationship is a direct consequence of the definition of the conditional intensity as the combination of hazard rates of the time intervals between exceedances, see Daley and Vere-Jones (2005), p. 231. There is a small abuse of notation in the equation above, as, to make the notation easy, we interchange the use of a hazard rate, a deterministic function, with the conditional intensity, a piecewise determined amalgam of hazard rates.

Note that the self-exciting component $v_i^*(t)$ enters both $\tau_i(t, u_i)$ in (1) and $F_{i,t}$ in (3) and thus provides a specific "clustering mechanism" into the conditional distribution of both times and marks of exceedances. After an exceedance occurs at time t_0 with mark x_0, the function $v_i^*\left(\cdot\right)$ jumps by $c_i\left(x_0\right)$ and increases the instantaneous probability of the exceedance's occurrence and the marks' volatility (through time-varying scale parameter $\beta_i(t)$). In absence of exceedances $v_i^*\left(\cdot\right)$ tends towards zero through function $g_i\left(\cdot\right)$. Being a transmitter of information of past exceedances to the future ones, the function $v_i^*\left(\cdot\right)$ may be interpreted as a kind of volatility measure of extreme exceedances. This interpretation may be found also in Bowsher (2007), where the estimated mid-quote intensity is used as approximation to the stock price's instantaneous volatility.

The clustering mechanism of the SE-POT model how past exceedances may trigger the occurrence of future exceedances can quite accurately describe the cluster behaviour of extreme exceedances observed on financial markets, see Chavez-Demoulin and McGill (2012). That is why the SE-POT model is chosen as a stepping stone for our multivariate model developed in Section 2.2.

Because of the overall importance of the SE-POT model for our multivariate model, in the next sections we develop some of its distributional properties, including a stationarity condition, and reconsider the existing specifications for the decay and impact functions.

2.1.2. Decay and impact functions

Considering functional specification of the decay and impact functions in (2) there are advantages in some specific forms. The decay function chosen in this paper is $g(s) = e^{-\gamma s}$, $\gamma > 0$ (the subindex "i" is dropped), which is a popular specification suggested in Hawkes (1971). This specification makes the self-exciting process a Markov process (Oakes, 1975) and leads to a simple formula for the covariance density (derived in Appendix B). This choice is also motivated in view of Boltzman's theory of elastic aftereffects, see Ogata (1988), p.11. An alternative is the function $g(s) = (s + \gamma)^{-(1+\rho)}$, with $\gamma, \rho > 0$. This specification originally comes from seismology, where is known as Omori law, see Helmstetter and Sornette (2002). Due to the substantial advantages in deriving the analytical formulas, we will stick to $g(s) = e^{-\gamma s}$ throughout the paper.

The aim of the impact function $c(\cdot)$ is to capture the effect of the marks of exceedances onto the conditional rate of future exceedances. A popular choice is $c(x) = e^{\delta x}$, see for example Chavez-Demoulin and McGill (2012) or McNeil, Frey, and Embrechts (2005) (Section 7.4.3). However, an important point to consider when specifying that function is to ensure its ability to accurately extract information from the marks. Provided the conditional distribution of the marks is time-varying (as it is indeed the case with the SE-POT model, see (3)), one expects $c(\cdot)$ to account not only for the magnitudes of the marks but also for the conditional distribution from which they were drawn. To put it differently, not the size of the mark but its quantile in the corresponding conditional distribution is decisive in determining the effect of the mark onto the conditional rate. Thus, instead of specifying $c(\cdot)$ as a fixed function, we suggest the following specification

$$c(x_t) = c^* \left(F_t(x_t) \right),$$

where F_t is the marks' conditional distribution (3) and $c^*(\cdot)$ is an increasing function $[0, 1] \to [1, \infty]$. This specification can properly capture the time-varying impact of an exceedance on the conditional rate. An easy way to construct $c^*(\cdot)$ is as $c^*(\cdot) = 1 + G^{\leftarrow}(\cdot)$, where $G^{\leftarrow}(\cdot)$ is the inverse of a distribution function G of some continuous positive random variable with finite mean δ. With such $c^*(\cdot)$ the impact function takes the form

$$c(x_t) = 1 + G^{\leftarrow} \left(F_t(x_t) \right). \tag{6}$$

We will use the above specification for the impact function throughout the text. In the empirical part of the paper, we will use $G^{\leftarrow}$ of an exponential distribution, which yields $c^*(u) = 1 - \delta \log(1-u)$.

Besides the appropriate extraction of information from the marks, the choice (6) for the impact function is advantageous over $c(x) = e^{\delta x}$, because (6) allows us to set the stationarity condition for the SE-POT model and to develop its distributional properties. In the next section we discuss those properties.

2.1.3. Stationarity condition and properties of the SE-POT model

As it was noted in Chavez-Demoulin, Davison, and McNeil (2005), the SE-POT model relates to the class of general self-exciting Hawkes processes and constitutes by its construction a branching process. A comprehensible explanation of the Hawkes process' representation as a branching process can be found in Møller and Rasmussen (2005) or Hawkes and Oakes (1974).

According to the branching process representation, there are two types of exceedances above the initial threshold in the SE-POT model: immigrants, that arrive as a homogeneous Poisson process with a constant rate τ, and descendants (triggered events), that follow a finite Poisson process with decaying rate determined by function $v^*(\cdot)$, see Daley and Vere-Jones (2005) (see Example 6.3(c)). Since both immigrants and descendants can trigger further descendants, for setting stationarity conditions it is necessary to consider the average number of the first-generation descendants trigged by one exceedance (whether by an immigrant or descendant).

That average number of triggered descendants is known as *a branching coefficient* and we denote it as ν. It is usual to consider $\nu = 1$ as a certain level of stability of the exceedance process: if $\nu \geq 1$ the development of the process could explode, i.e., the number of events in finite time interval tends to infinity. Clearly, in that case the process is non-stationary. In seismological literature the situation of $\nu > 1$ is called super-critical regime, see Helmstetter and Sornette (2002).

For practical application the case $\nu < 1$ is the most important because then the process of exceedances becomes *stationary*, provided the process of immigrants is stationary as well (which is the case in the SE-POT model). In the SE-POT model with $\nu < 1$, exceedances occur in finite clusters of length $(1 - \nu)^{-1}$, where exceedances within the cluster are temporally dependent but the clusters themselves are independent. In Proposition 2.1 we provide a formula for the branching coefficient and the stationarity condition of the SE-POT model.

Proposition 2.1. *The process of exceedances with the conditional intensity $\tau(t, u)$ of the SE-POT model, where $\tau(t, u)$ is as in (1)-(2) (dropping the subindex i), with decay function $g(s) = e^{-\gamma s}$, and the impact function as in (6), has the branching coefficient $\nu = \frac{\psi(1+\delta)}{\gamma}$ and is stationary if $\nu < 1$ with an average rate $\bar{\tau} := E[\tau(t, u)] = \frac{\tau}{1-\nu}$.*

A proof is given in Appendix A.

Under the stationarity condition of Proposition 2.1, the moments of the counting measure $N(t, t+s)$ of marginal exceedances above the initial threshold in time interval $(t, t+s)$ can be expressed as follows

$$\begin{aligned}
\mathrm{E}\left[N(t,t+s)\right] &= s\bar{\tau}, \qquad s>0,\\
\mathrm{Var}\left[N(t,t+s)\right] &= s\bar{\tau} + 2\int_0^s (s-z)\,\mu(z)dz, \qquad s>0,\\
\mathrm{Cov}\left[N(t_1,t_2),N(t_3,t_4)\right] &= \int_{t_1}^{t_2}\int_{t_3}^{t_4} \mu\left(z_1-z_2\right)dz_1dz_2, \qquad t_1<t_2<t_3<t_4,
\end{aligned}$$

where $\mu(u)$ is the process' covariance density defined as

$$\mu(z) = \frac{\mathrm{E}\left[dN(t+z)dN(t)\right]}{(dt)^2} - \bar{\tau}^2, \qquad z>0.$$

A reference for the above formulas can be found in, e.g, Vere-Jones and Davies (1966), p.253.

Proposition 2.2. *Setting the decay function as $g(s) = e^{-\gamma s}$ and the impact function as in (6) the covariance density of the SE-POT model takes the form*

$$\mu(u) = Ae^{-bz}, \qquad z>0, \tag{7}$$

where

$$b = \gamma - \psi(1+\delta), \qquad\qquad A = \frac{\bar{\tau}\psi(1+\delta)\left(2\gamma-\psi(1+\delta)\right)}{2\left(\gamma-\psi(1+\delta)\right)}.$$

A proof of can be found in Appendix B.

With (7) the above formulas for second moments of the counting process N take the forms

$$\mathrm{Var}\left[N(t,t+s)\right] = s\bar{\tau} + \frac{2A}{b^2}\left(bs + e^{-bs} - 1\right)$$

and

$$\mathrm{Cov}\left[N(t_1,t_2),N(t_3,t_4)\right] = \frac{A}{b^2}\left(e^{-b(t_3-t_2)} - e^{-b(t_3-t_1)} - e^{-b(t_4-t_2)} + e^{-b(t_4-t_1)}\right), \tag{8}$$

for $t_1<t_2<t_3<t_4$.

From this one can conclude, first, that the variance of $N(t,t+s)$ grows for large s linearly with s – a feature similar to Brownian motion. Second, the covariance between $N(t_1,t_2)$ and $N(t_3,t_4)$ reduces exponentially to zero as $t_3-t_2\to\infty$. This property corresponds with the earlier statement that exceedances occurring within one cluster are serially dependent but those lying in different clusters are uncorrelated.

In this section we analysed the SE-POT model from the perspective of self-exciting point-processes. In the next section we discuss the relationship of the SE-POT model to EVT models.

2.1.4. Relationship of SE-POT and EVT

As it was noted in Section 2.1.3, the SE-POT model relates to the class of general self-exciting Hawkes processes. On the other hand, by setting $\alpha=\psi=0$ the SE-POT transforms to the standard POT model for iid exceedances. Hence, one can expect that the SE-POT can be regarded as a special representation of point process of non-independent extremes.

According to the extremal index' theory (see Section 4 in Leadbetter, 1983), the extremal clusters of exceedances of a stochastic processes with an extremal index $\theta<1$ (e.g., GARCH) have an average cluster size θ^{-1} and occur in time according to a homogeneous Poisson process, i.e., individual exceedances follow a *Poisson cluster process*. The parallels to the SE-POT model are that there individual exceedances occur also according to a Poisson cluster process with an average cluster size $(1-\nu)^{-1}$ and cluster arrival rate τ, see Hawkes and Oakes (1974).

To set the relationship between the properties of SE-POT model and the extremal index, note that the latter is an asymptotic concept and the former model is rather a finite sample empirical representation of the possible asymptotic dependence. This difference precludes the formalization

of that relationship, but still note that the SE-POT model meets all the assumptions required for the extremal index. Those assumptions include stationarity of the process, heavy-tailedness of the marks, and a mixing condition $D(w_n)$ that restricts the "long range" dependence in the process, for details consult Leadbetter (1988), Section 2. While the first two conditions are discussed earlier in the text, we refer to Daley and Vere-Jones (1988), Proposition 10.3.IX, for the general proof of the last condition. That proposition states that a cluster process is mixing if the process of cluster-centers is mixing itself. Indeed, this is the case for the SE-POT model where the the process of cluster-centers is a homogeneous Poisson process with rate τ.

2.2. Multivariate Model

In this section we introduce our approach to model the process of multivariate exceedances of $\mathbf{X}_t = (X_{1,t}, \ldots, X_{d,t})$, $t \in \mathbb{R}$ above any $\mathbf{x} \geq \mathbf{u}$, where $\mathbf{u} = (u_1, \ldots, u_d)$ is the initial threshold, conditioned on the history of the past realizations. The model is subject to multivariate extreme value theory (MEVT), accounting for the specific multivariate dependence structure between exceedances, and to the SE-POT model, accounting for the clustering in times of occurrence and marks of marginal extreme events.

Under a multivariate exceedance at time t we understand a situation when $\mathbf{X}_t$ exceeds $\mathbf{x}$ in at least one component. Our model provides an instantaneous conditional rate that $\mathbf{X}_t$ exceeds any $\mathbf{x} \geq \mathbf{u}$ in at least one component. Each of the univariate components is modeled as a SE-POT process as described in Section 2.1. This conditional multivariate rate and the rates of the univariate processes can then be used to extract the probabilies of all combinations of exceedances, e.g., that $\mathbf{X}_t$ exceeds $\mathbf{x} \geq \mathbf{u}$ in all components simultaneously.

2.2.1. Model Construction

Assuming that the point process of marginal exceedances of $\mathbf{X}_t = (X_{1,t}, \ldots, X_{d,t})$ above $\mathbf{x} \geq \mathbf{u}$ is governed by the conditional rates $\tau_i(t, x_i)$, $i = 1, \ldots, d$ of the SE-POT model, the ultimate aim of the multivariate model is to provide an instantaneous conditional rate of at least one exceedance of $\mathbf{X}_t$ above $\mathbf{x}$. We denote this rate as $\tau(t, \mathbf{x})$. From the interpretation of the conditional rate as a combination of hazard rates, see Daley and Vere-Jones (2005), Section 7.2, it should hold for $\tau(t, \mathbf{x})$ that

$$\mathrm{P}\left(T_k(\mathbf{x}) > t \mid \mathbb{H}_{t_0}\right) = \exp\left(-\int_{t_0}^{t} \tau(s, \mathbf{x}) ds\right), \qquad t \geq t_0 \geq T_{k-1}(\mathbf{x}), \tag{9}$$

where $T_{k-1}(\mathbf{x})$ is the time of the last event, when $(\mathbf{X}_s)_{s \leq t_0}$ exceeds $\mathbf{x}$ in at least one component. The following proposition shows the relationship of the joint rate $\tau(t, \mathbf{x})$ and the marginal rates $\tau_i(t, x_i)$, $i = 1, \ldots, d$.

Proposition 2.3. *If the marginal point processes of exceedances of $\mathbf{X}_t = (X_{1,t}, \ldots, X_{d,t})$, $t \in \mathbb{R}$ above threshold $\mathbf{x} = (x_1, \ldots, x_d)$, that lies above the initial threshold $\mathbf{u} = (u_1, \ldots, u_d)$, are governed by conditional rates $\tau_i(t, x_i)$, $i = 1, \ldots, d$ of the SE-POT model, then the conditional rate of the event that $\mathbf{X}_t$ exceeds $\mathbf{x}$ in at least one component is of the form*

$$\tau(t, \mathbf{x}) = V_t\left(\frac{1}{\tau_1(t, x_1)}, \ldots, \frac{1}{\tau_d(t, x_d)}\right), \tag{10}$$

where V_t is the exponent measure of extreme value copula C_t^*, i.e.,

$$V_t\left(y_1, \ldots, y_d\right) = -\log C_t^*\left(e^{-1/y_1}, \ldots, e^{-1/y_d}\right). \tag{11}$$

Proof For the ease of presentation, we employ the following notation. $N(t, \mathbf{x})$ denotes a counting measure of events when $(\mathbf{X}_s)_{s \leq t}$ exceeds $\mathbf{x}$ in at least one component. $T_1(\mathbf{x}), T_2(\mathbf{x}), \ldots$ stand for the consecutive times of those multivariate events. The corresponding notation for the marginal processes of exceedances are denoted as, respectively, $N_i(t, x_i)$ and $T_{i,1}(x_i), T_{i,2}(x_i), \ldots$

Provided there is no exceedance at time t, the conditional distribution of $T_{N(t_0, \mathbf{x})+1}(\mathbf{x})$ given $\mathbb{H}_{t_0}$ can be expressed in terms of marginal times of exceedances in the following way

$$\mathrm{P}\left(T_{N(t_0, \mathbf{x})+1}\left(\mathbf{x}\right) > t \mid \mathbb{H}_{t_0}\right) = \mathrm{P}\left(T_{1, N_1(t_0, x_1)+1}\left(x_1\right) > t, \ldots, T_{d, N_d(t_0, x_d)+1}\left(x_d\right) > t \mid \mathbb{H}_{t_0}\right). \tag{12}$$

Recall that $\mathbb{H}_{t_0} = \bigcup_{i=1}^{d} \mathbb{H}_{i,t_0}$, where $\mathbb{H}_{i,t_0}$ is a history of point process of exceedances of $(X_{i,s})_{s \le t_0}$ above the *initial* threshold u_i.

The conditional set $\mathbb{H}_{t_0}$ contains information both on the conditional distribution of the margins (i.e., distribution of $T_{i,N_i(t_0,x_i)+1}(x_i) > t_0$ given $\mathbb{H}_{i,t_0}$) and on their joint distribution (i.e., distribution of $T_{i,N_i(t_0,x_i)+1}(x_i) > t_0$ given $\mathbb{H}_{t_0}$). Since we know from (5) that

$$\mathrm{P}\left(T_{i,N_i(t_0,x_i)+1}(x_i) \ge t \mid \mathbb{H}_{i,t_0}\right) = \exp\left(-\int_{t_0}^{t} \tau_i(s,x_i)ds\right), \qquad t \ge t_0,$$

we decompose the right-hand side of (12) into the conditional distribution of the margins given their own histories and joint distribution of those conditional distributions (with help of a copula[1]):

$$\begin{aligned}\mathrm{P}\left(T_{1,N_1(t_0,x_1)+1}(x_1) > t, \ldots, T_{d,N_d(t_0,x_d)+1}(x_d) > t \mid \mathbb{H}_{t_0}\right) = \\ \mathrm{P}\left(\bar{U}_{1,N_1(t_0,x_1)+1} \le e^{-\int_{t_0}^{t} \tau_1(s,x_1)ds}, \ldots, \bar{U}_{d,N_d(t_0,x_d)+1} \le e^{-\int_{t_0}^{t} \tau_d(s,x_d)ds}\right) \\ =: C_{t_0}\left(e^{-\int_{t_0}^{t} \tau_1(s,x_1)ds}, \ldots, e^{-\int_{t_0}^{t} \tau_d(s,x_d)ds}\right), \quad (13)\end{aligned}$$

where $\bar{U}_{i,N_i(t_0,x_i)+1} := \exp\left(-\int_{t_0}^{T_{i,N_i(t_0,x_i)+1}(x_i)|\mathbb{H}_{i,t_0}} \tau_i(s,x_i)ds\right)$, with $T_{i,N_i(t_0,x_i)+1}(x_i) \mid \mathbb{H}_{i,t_0}$ denoting a random variable $T_{i,N_i(t_0,x_i)+1}(x_i)$ conditioned on $\mathbb{H}_{i,t_0}$; C_{t_0} is a copula function conditioned on the information set $\mathbb{H}_{t_0}$. The use of copula function is justified because $\bar{U}_{i,N_i(t_0,x_i)+1} \sim \mathrm{Unif}[0,1]$.

Comparing (13) and (9), the condition on the intensity $\tau(t, x_1, \ldots, x_d)$ reads

$$\exp\left(-\int_{t_0}^{t} \tau(s,x_1,\ldots,x_d)ds\right) = C_{t_0}\left(e^{-\int_{t_0}^{t} \tau_1(s,x_1)ds}, \ldots, e^{-\int_{t_0}^{t} \tau_d(s,x_d)ds}\right). \quad (14)$$

To solve the above equation, note that for $t_0 \uparrow t$ the following holds

$$\int_{t_0}^{t} \tau(s,x_1,\ldots,x_d)ds \approx (t-t_0)\tau(t_0,x_1,\ldots,x_d) \text{ and } \int_{t_0}^{t} \tau_i(s,u_i)ds \approx (t-t_0)\tau_i(t_0,u_i),\ i = 1,\ldots d,$$

which, in turn, allows us to write

$$\tau(t,x_1,\ldots,x_d) = -\lim_{t_0 \uparrow t} \frac{\log C_{t_0}\left(e^{-(t-t_0)\tau_1(t_0,x_1)}, \ldots, e^{-(t-t_0)\tau_d(t_0,x_d)}\right)}{t-t_0}.$$

Denoting $s = 1/(t-t_0)$ and setting $t_0 = t$ in the subscript of copula and in $\tau_i(t_0,u_i)$, we obtain

$$\tau(t,x_1,\ldots,x_d) = -\log \lim_{s\to\infty} C_t\left(e^{-\frac{\tau_1(t,x_1)}{s}}, \ldots, e^{-\frac{\tau_d(t,x_d)}{s}}\right)^s. \quad (15)$$

With the following intuitive boundaries on $\tau(t,x_1,\ldots,x_d)$

$$0 < \max\{\tau_1(t,x_1),\ldots,\tau_d(t,x_d)\} \le \tau(t,x_1,\ldots,x_d) \le \sum_{i=1}^{d} \tau_i(t,u_i) < \infty,$$

we conclude that a non-trivial limit in (15) exists for all $0 < \tau_i(t,u_i) < \infty$, $i = 1,\ldots,d$, namely, there exists a function $C_t^*(w_1,\ldots,w_d) \in [0,1]$ defined on $(w_1,\ldots,w_d) \in [0,1]^d$, such that

$$\lim_{s\to\infty} C_t\left(e^{-\frac{\tau_1(t,x_1)}{s}}, \ldots, e^{-\frac{\tau_d(t,x_d)}{s}}\right)^s = C_t^*\left(e^{-\tau_1(t,x_1)}, \ldots, e^{-\tau_d(t,x_d)}\right) \in (0,1). \quad (16)$$

From the results of multivariate extreme value theory, see for example Gudendorf and Segers (2010),

[1] In continuous case, a copula is a multivariate distribution function with uniformly on $[0,1]$ distributed marginal distributions. A detailed introduction to copulas can be found in Nelsen (2006). For an excellent review of copula based models for econometric time series see Patton (2012).

it is well known that if there exists a copula $C(w_1, \ldots, w_d)$ such that

$$\lim_{s\to\infty} C\left(w_1^{1/s}, \ldots, w_d^{1/s}\right)^s = C^*\left(w_1, \ldots, w_d\right), \qquad \forall\left(w_1, \ldots, w_d\right) \in [0,1]^d, \tag{17}$$

then C^* is an extreme value copula and C is said to be in the domain of attraction of C^*. Surely, this definition applies to (16) with $w_i = e^{-\tau_i(t,u_i)}$. Hence, we conclude that C_t^* is an extreme value copula.

Note that for any extreme value copula C^* there exists a measure V, the **exponent measure**, such that

$$C^*\left(w_1, \ldots, w_d\right) = \exp\left(-V\left(\frac{1}{-\log w_1}, \ldots, \frac{1}{-\log w_d}\right)\right), \tag{18}$$

where for the exponent measure, we have the homogeneity property

$$V\left(ay_1, \ldots, ay_d\right) = \frac{1}{a} V\left(y_1, \ldots, y_d\right) \qquad \forall a > 0. \tag{19}$$

Denoting the exponent measure of C_t^* as V_t, the final conditional rate of at least one exceedance above $(x_1, \ldots, x_d)$ takes the form

$$\tau(t, x_1, \ldots, x_d) = V_t\left(\frac{1}{\tau_1(t, x_1)}, \ldots, \frac{1}{\tau_d(t, x_d)}\right).$$

Note that since C_t in (16) is conditioned on the information set $\mathbb{H}_t$ we allow the copula C_t^*, and hence V_t, to evolve with time. ■

Remark The presence of the extreme-value copula in rate (10) is quite surprising, as the derivation considers only the conditional time intervals between exceedances, not their magnitudes. Furthermore, if $\tau_i(t, x_i)$ and the dependence parameters of V_t do not evolve with time, we obtain a standard EVT model. Note, that extreme value copulas include a broad spectrum of dependence structures, including the independence case. Popular extreme value copulas are Gumbel, Galambos, Hüsler-Reiss, and t-EV copulas, see Gudendorf and Segers (2010).

The extreme value copula in Proposition 2.3 has a direct relationship to the copula that governs the sizes of exceedances of $\mathbf{X}_t$ above $\mathbf{u}$. The following proposition formulates that relationship.

Proposition 2.4. *The extreme value copula that governs the multivariate conditional rate (10) from Proposition 2.3 is the same copula that governs the limiting distribution of the normalized magnitudes of $\mathbf{X}_t$, i.e., for every fixed t there exist sequences $\mathbf{a}_t(n) = (a_{1,t}(n), \ldots, a_{d,t}(n)) \in \mathbb{R}_+^d$ and $\mathbf{b}_t(n) = (b_{1,t}(n), \ldots, b_{d,t}(n)) \in \mathbb{R}^d$, such that*

$$\lim_{n\to\infty} nP\left(\frac{X_{1,t} - b_{1,t}(n)}{a_{1,t}(n)} > y_1 \cup \ldots \cup \frac{X_{d,t} - b_{d,t}(n)}{a_{d,t}(n)} > y_d\right) = W_t\left(\frac{1}{y_1}, \ldots, \frac{1}{y_d}\right), \tag{20}$$

with $W_t\left(\frac{1}{y_1}, \ldots, \frac{1}{y_d}\right)$ being equal to $V_t\left(\frac{1}{y_1}, \ldots, \frac{1}{y_d}\right)$ from Proposition 2.3.

Proof Since the point process of exceedances of $(\mathbf{X}_s)_{s\in\mathbb{R}}$ above a sufficiently high threshold is assumed to follow the SE-POT model, the magnitudes of exceedances of $\mathbf{X}_t$ have the generalized Pareto distribution (3). Clearly, the GPD belongs to the maximum domain of attraction of the extreme value distribution, which reassures the existence of the normalized constants $\mathbf{a}_t(n)$ and $\mathbf{b}_t(n)$ such that (20) holds.

It remains to be shown that $W_t = V_t$, or, in terms of copulas, $C_t^\circ = C_t^*$, where C_t° is an extreme value copula associated with exponent measure W_t, see (11). It suffices to prove that relationship for the case when $X_{i,s}$, $i = 1, \ldots, d$, does not depend on its own history, because C_t^* models the (limiting) **conditional** survival distributions of time intervals between marginal exceedances, see (13) and (16). Although the marginal exceedances of $(X_{i,s})_{s\in\mathbb{R}}$, $i = 1, \ldots, d$ are serially dependent, due to conditioning of those survival distributions with the SE-POT model, which captures that serial dependence, the functional form of C_t^* is the same as if the marginal exceedances were serially independent.

Consider iid random vectors $\mathbf{Y}_j = (Y_{1,j}, \ldots, Y_{d,j})$, $j = 1, 2, \ldots$ that have the same dependence structure as $\mathbf{X}_t$. We assume that the margins of $\mathbf{Y}_j$ are unit Frechet distributed, $\mathrm{P}(Y_{i,j} \leq y) = e^{-1/y}$, $i = 1, \ldots, d$. This assumption does not deprive of the proof's generality, because the marginal tail distribution of $\mathbf{X}_t$ is known and the relationship $W_t = V_t$ defines the equivalence only in the corresponding dependence structures of X_t and of the one suggested by exponent measure V_t. Relationship (20) for $\mathbf{Y}_j$ reads

$$\lim_{n\to\infty} n\mathrm{P}\left(\frac{Y_{1,j}}{n} > y_1 \cup \ldots \cup \frac{Y_{1,j}}{n} > y_d\right) = W_t\left(\frac{1}{y_1}, \ldots, \frac{1}{y_d}\right).$$

It is a well known result from the EVT that for $A \in [0,1]$ and $B = \mathbb{R}_+^d \setminus [(0, y_1) \times \ldots \times (0, y_d)]$ the point process

$$N_n(S = A \times B) = \sum_{j=1}^{n} I_{\{(j/n, \mathbf{Y}_j/n) \in S\}}, \text{ as } n \to \infty, \tag{21}$$

of events when $(\mathbf{Y}_j/n)$ exceeds threshold $\mathbf{y} = (y_1, \ldots, y_d)$ in at least one component converges in distribution to a Poisson point process with rate $\lambda(A) \times W_t\left(\frac{1}{y_1}, \ldots, \frac{1}{y_d}\right)$, where $\lambda(A)$ is the Lebesgue measure of $A \in [0,1]$ and W_t is the *exponent measure* of the set B. For details consult Coles and Tawn (1991) and Theorem 6.1.11 in de Haan and Ferreira (2006).

Using the same notation for counting measures and times of exceedances of $(\mathbf{Y}_j/n)$ above threshold $(y_1, \ldots, y_d)$ as in the proof of Proposition 2.3 and recalling that intervals between homogeneous Poisson events are exponentially distributed (with the mean equal to the inverse rate), the conditional distribution of $T_{N(t_0,\mathbf{y})+1}(\mathbf{y})$ given $\mathbb{H}_{t_0}$ takes the form

$$\mathrm{P}\left(T_{N(t_0,\mathbf{y})+1}(\mathbf{y}) > t \mid \mathbb{H}_{t_0}\right) = \exp\left[-(t - t_0)W_t\left(\frac{1}{y_1}, \ldots, \frac{1}{y_d}\right)\right]. \tag{22}$$

On the other hand, from straightforward arguments one can conclude that marginal processes of exceedances of $(Y_{i,j}/n)$ above y_i also converges to the homogeneous Poisson process on $[0,1]$ with rate $\lim_{n\to\infty} n\mathrm{P}(Y_{i,j}/n > y_i) = \lim_{n\to\infty} n\left(1 - e^{-1/(y_i n)}\right) = 1/y_i$, i.e.,

$$\mathrm{P}\left(T_{i,N_i(t_0,y_i)+1}(y_i) \geq t \mid \mathbb{H}_{i,t_0}\right) = \exp\left(-\frac{t - t_0}{y_i}\right), \qquad t \geq t_0.$$

The above distribution allows us to express the right-hand side of marginal decomposition (12) as

$$\begin{aligned}&\mathrm{P}\left(T_{1,N_1(t_0,y_1)+1}(y_1) > t, \ldots, T_{d,N_d(t_0,y_d)+1}(y_d) > t \mid \mathbb{H}_{t_0}\right) = \\ &\quad \mathrm{P}\left(\bar{U}_{1,N_1(t_0,y_1)+1} \leq e^{-\frac{t-t_0}{y_1}}, \ldots, \bar{U}_{d,N_d(t_0,y_d)+1} \leq e^{-\frac{t-t_0}{y_d}}\right) =: C_{t_0}^{\circ}\left(e^{-\frac{t-t_0}{y_1}}, \ldots, e^{-\frac{t-t_0}{y_d}}\right),\end{aligned} \tag{23}$$

where $\bar{U}_{i,N_i(t_0)+1} = \exp\left(-\frac{T_{i,N_i(t_0,y_i)+1}(y_i) \mid \mathbb{H}_{i,t_0} - t_0}{y_i}\right)$ with $T_{i,N_i(t_0,y_i)+1}(y_i) \mid \mathbb{H}_{i,t_0}$ denoting a random variable $T_{i,N_i(t_0,y_i)+1}(y_i)$ conditioned on $\mathbb{H}_{i,t_0}$, and $C_{t_0}^{\circ}$ is a copula function.

Equating (23) to (22) we obtain the following condition on the copula $C_{t_0}^{\circ}$:

$$C_{t_0}^{\circ}\left(e^{-z_1}, \ldots, e^{-z_d}\right) = \exp\left[-aW_t\left(\frac{a}{z_1}, \ldots, \frac{a}{z_d}\right)\right], \qquad a > 0, \forall t_0 < t \tag{24}$$

where we denote $z_i = (t - t_0)/y_i$ and $a = t - t_0$. Existence of the unique copula $C_{t_0}^{\circ}$ is guaranteed only in the case when the right-hand side of (24) does not change with a. Since this condition is nothing else than the homogeneity property of the exponent measure: $aW_t\left(\frac{a}{z_1}, \ldots, \frac{a}{z_d}\right) = W_t\left(\frac{1}{z_1}, \ldots, \frac{1}{z_d}\right)$, $\forall a > 0$, – we conclude that $C_{t_0}^{\circ}$ exists and it can be expressed as

$$C_{t_0}^{\circ}(w_1, \ldots, w_d) = \exp\left[-W_t\left(\frac{1}{-\log w_1}, \ldots, \frac{1}{-\log w_d}\right)\right], \tag{25}$$

where w_i denotes e^{-z_i} from (24). From decomposition (18) it follows that $C_{t_0}^{\circ}$ is an **extreme value copula**. Note that $C_{t_0}^{\circ}$ is independent of the conditional set $\mathbb{H}_{t_0}$.

Due to the characteristic property (see Theorem 7.44 in McNeil, Frey, and Embrechts (2005)) of extreme value copulas:

$$C^{\circ}_{t_0}\left(u_1^s,\ldots,u_d^s\right) = C^{\circ}_{t_0}\left(u_1,\ldots,u_d\right)^s, \qquad s>0. \tag{26}$$

the limit copula in (16) with $C_t := C_t^{\circ}$ yields the same limiting copula $C_t^* = C_t^{\circ}$. Hence, we conclude that exponent measure V_t of C_t^* equals exponent measure W_t of C_t°. ■

The above proposition presents an interesting result, because it relates the multivariate rate (10) to the extreme value copula of the properly normalized magnitudes of $\mathbf{X}_{t^*}$. This relationship may be regarded as an extension of the EVT result (21) for multivariate exceedances in the independence case to the special case of dependence provided by the SE-POT model and the (possible) time-variation of the exponent measure V_t.

2.2.2. A closer look at the model implied dependence

Proposition 2.3 develops a natural way how dynamics of marginal exceedances can be incorporated through extreme value copula C_t^* into the construction of the multivariate rate (10). Combining the point processes of marginal exceedances, the copula C_t^* provides a specific type of dependence between them. In this section we attempt to clarify the concept of dependence between the point processes provided by our multivariate model and we suggest a way of modelling that dependence.

Consider a specific form of rate (10) with exponent measure $V_t(y_1,y_2) = \left(y_1^{-\theta}+y_2^{-\theta}\right)^{1/\theta}$ of the two-dimensional symmetric Gumbel copula. Rate (10) takes the form

$$\tau(t,x_1,x_2) = \left[\tau_1(t,x_1)^{\theta}+\tau_2(t,x_2)^{\theta}\right]^{1/\theta},$$

where θ is the dependence parameter of the Gumbel copula. In the limit this copula approaches the two-dimensional comonotonicity copula as $\theta\to\infty$ and the independence copula as $\theta=1$. It is easy to verify that for $\tau(t,x_1,x_2)$ it holds

$$\tau(t,x_1,x_2)\big|_{\theta>1} = \left[\tau_1(t,x_1)^{\theta}+\tau_2(t,x_2)^{\theta}\right]^{1/\theta} < \tau_1(t,x_1)+\tau_2(t,x_2) = \tau(t,x_1,x_2)\big|_{\theta=1}. \tag{27}$$

The above inequality illustrates the concept of the point processes' dependence suggested by our model: if the marginal processes of exceedances are "dependent" ($\theta>1$), then the (instantaneous) expected number of events when at least one of the margin exceeds the threshold is always smaller than the expected number of the events in the "independence" case ($\theta=1$). Thus, if the two marginal processes of exceedances are dependent, then the probability that the margins exceed any threshold (x_1,x_2) *simultaneously* is strictly larger than in the independent case, i.e., strictly larger than zero. Note that by holding for any threshold (x_1,x_2), this describes dependence not only of the time of the exceedances, but also of their marks.

Note further that the feature of simultaneous marginal exceedances is a direct consequence of the point process interpretation of the extreme value theory as it described in Propositions 2.3 and 2.4 and that it presents a characteristic property of our model. It provides a specific definition of the dependence between point processes which are subject to EVT.

From a practical point of view, the concept of dependence between point processes of exceedances may be particularly useful in modelling moderately aggregated data, where simultaneous extreme events are observable, such as daily data. That concept may also be applicable for high-frequency data. This, however, may require a specific definition of simultaneous exceedances, e.g, those exceedances that occur in time interval of, say, 10 seconds can be regarded as simultaneous. Nevertheless, in recent studies, see Bollerslev, Todorov, and Li (2013) for an overview, it was argued that in high-frequency data the occurrence of common jumps across different assets may be induced by strong dependencies in the "extreme".

Returning to the modelling aspects, note that due to the conditioning on $\mathbb{H}_t$, the exponent measure V_t, and hence copula C_t^*, in the specification of conditional rate (10) may evolve through time. There are three possible ways for this time-evolution: first, C_t^* remains the same through time; second, the functional form of C_t^* remains the same but its parameters vary through time; finally, both the functional form and the parameters vary through time. In this paper we focus on the second option and fix the functional form of C_t^* (this choice is supported by Proposition

2.4) leaving its dependence parameter to change over time. The time-variation of the dependence parameter is a useful property, which can be justified from the financial point of view by contagion, frailty, and clustering of marginal exceedances which cause the strength of dependence described by C_t^* to change.

From the interpretation of the dependence between the point process, it is plausible to parameterize the time-varying dependence parameter $\theta(t)$ as a finite function of the number of simultaneous exceedances of $(\mathbf{X}_s)_{s<t}$ above the initial threshold. The relationship between the dependence and the number of simultaneous exceedances is also encountered in the literature, see, e.g., Bae, Karolyi, and Stulz (2003). On the other hand, it is reasonable to allow for changes in the dependence parameter when at least one of the margins jumps. This accounts for the fact that contagion between markets may occur in response to losses from only one of the markets. Finally, the dependence parameter may include (exogenous) information of some risk factors $\mathbf{z}_t$ such as the CBOE Volatility Index (VIX), which are believed to influence or forecast the strength of dependence between the margins. One way to construct the time varying dependence parameter $\theta(t) \geq 0$ (assuming that its size is proportional to the strength of dependence) is given by the multivariate Hawkes process

$$\theta(t) = \theta_m + \sum_{i=1}^{d} \psi_{m,i} \int_{-\infty}^{t} g_{m,i}(t-s) dN_i(s) + \Upsilon^{'} \mathbf{z}_t, \qquad \theta_m \geq 0, \Upsilon \geq 0, \psi_{m,i} \geq 0, \forall i, \tag{28}$$

where $N_i(s)$ is a counting measure of ith marginal exceedances, $g_{m,i}(\cdot)$ is the decay function, e.g., $g_{m,i}(t) = e^{-\gamma_{m,i} t}$, and Υ is a vector of coefficients. This parametrization is not derived from any theoretical arguments, but is simply one possible parametrization of many others that could be thought of. However, we believe (28) is one of the easiest and most practical ways to account for time-varying changes in the degree of the extreme dependence described by C^*. Estimating the model with dependence parameter (28), one can carry out statistical tests to identify the most influential factors for dependence modelling and place restrictions on the parameters. Note that the multivariate model discussed in this paper captures the dependence in two aspects: marginally, through the self-exciting structure of the conditional intensities, and cross-sectionally, through $\theta(t)$. This allows for a certain "decoupling" of the dependence and hence contributes to its more effective modelling. Note that the precise functional forms of $\theta(t)$ chosen for our applications are discussed in later sections.

2.3. Properties of the multivariate model

Using the multivariate model of exceedances, in this section we infer the joint conditional distribution of the marks (Section 2.3.1) and develop an approach to estimate probabilities to fall into any sufficiently remote region (Section 2.3.2). In Section 2.3.3 we investigate the mechanism how single marginal exceedances may trigger the other margins to exceed the initial threshold. Finally, Section 2.3.4 outlines useful implications of the model for risk management.

2.3.1. Joint conditional distribution of the marks

Proposition 2.4 intuitively suggests that from the conditional multivariate rate $\tau(t, x_1, \ldots, x_d)$ one should be able to infer the conditional joint distribution of the sizes of exceedances. Indeed, from the interpretation of $\tau(t, x_1, \ldots, x_d)$ as an instantaneous (i.e., as the time interval tends to zero) conditionally expected number of events when $(\mathbf{X}_s)_{s\in(t,t+1)}$ exceeds $(x_1, \ldots, x_d)$ in at least one marginal component per unit time, it follows

$$\mathrm{P}\left(\bigcup_{i=1}^{d} X_{i,t} > x_i \middle| \bigcup_{i=1}^{d} X_{i,t} > u_i, \right) = \frac{\tau(t, x_1, \ldots, x_d)}{\tau(t, u_1, \ldots, u_d)}, \qquad x_i \geq u_i, i = 1, \ldots, d$$

and hence

$$\mathrm{P}\left(X_{1,t} \leq x_1, \ldots X_{d,t} \leq x_d \middle| \bigcup_{i=1}^{d} X_{i,t} > u_i, \right) = 1 - \frac{\tau(t, x_1, \ldots, x_d)}{\tau(t, u_1, \ldots, u_d)}. \tag{29}$$

This conditional distribution of the marks incorporates the dynamics of marginal exceedances and of the dependence shifts. (29) may be useful for estimation of the conditional marginal expected shortfall, see Section 2.3.4.

A characteristic property of distribution (29) is that it provides a positive coefficient χ_U of upper tail dependence between the all marginal pairs of the distribution. Considering the two-dimensional version of (29), the coefficient of upper tail dependence χ_U may be expressed as

$$\chi_U = \lim_{q \to 1^-} \frac{2 - 2q - \tau\left(t, (1-q)\tau\left(t, u_1, u_2\right), (1-q)\tau\left(t, u_1, u_2\right)\right) / \tau\left(t, u_1, u_2\right)}{1-q} = 2 - \tau(t, 1, 1). \quad (30)$$

For example, with the Gumbel symmetric dependence structure, $\chi_U = 2 - 2^{1/\theta(t)}$. Note that χ_U is time-varying. This feature enables to approach the modelling of extreme dependencies with more flexibility. Furthermore, the upper tail dependence allows us to model clustering of the exceedances' sizes, which we observe in real data.

2.3.2. Probabilities of exceedances in a remote region

Proposition 2.3 provides a conditional multivariate rate of an event that $\mathbf{X}_t$ exceeds $\mathbf{x}$ in at least one component. Equation (10) can also be used to construct the conditional rate that $\mathbf{X}_t$ falls in any "remote" set $\mathbf{A} \in \mathbb{R}_+^d \setminus [(0, u_1) \times \ldots \times (0, u_d)]$. We denote this rate as $\tau\left(t, \mathbf{A}\right)$. This can be achieved by calculating the model's conditional intensity $\lambda\left(t, x_1, \ldots, x_d\right)$ which is defined as follows

$$\int_{x_1}^{\infty} \ldots \int_{x_d}^{\infty} \lambda\left(t, s_1, \ldots, s_d\right) ds_d \ldots ds_1 = \tau\left(t, x_1, \ldots, x_d\right),$$

where $\tau\left(t, x_1, \ldots, x_d\right)$ is the multivariate rate (10). With this intensity, the conditional rate $\tau\left(t, \mathbf{A}\right)$ takes the form

$$\tau\left(t, \mathbf{A}\right) = \int_{\mathbf{x} \in \mathbf{A}} \lambda\left(t, x_1, \ldots, x_d\right) dx_d \ldots dx_1. \quad (31)$$

Formulation (31) is very useful for practical applications, because with only one rate (10) we can infer the extremal behaviour of $(\mathbf{X}_t)_{t \in \mathbb{R}}$ in any remote region without a need of re-estimating the model.

2.3.3. Contagion mechanism

Clustering of (multivariate) extreme events in both times and magnitudes is a ubiquitous feature of financial time-series. The sources for that clustering may be attributed to financial contagion and/or exposure to (unobservable) common risk factors that determine the extreme behaviour of the returns. Without distinguishing the reason for the clustering, our model, being decisively a reduced-form one, provides a specific clustering mechanism in occurrence of multivariate extreme events. This mechanism is propagated in three channels. First, making the dependence parameter depend on the history of past exceedances allows an adjustment of the strength of that dependence, which in turn may accelerate occurrence of joint extreme events. Second, our model provides upper tail dependence in distribution of the marks of exceedances, see Section 2.3.1. Due to this feature our model reproduces clustering in the magnitudes of exceedances. Finally, the sheer possibility of joint extreme events, implied by our model, induces the univariate extremes to occur jointly, which consequently triggers further joint exceedances.

To highlight the last point on a two-dimensional version of our model, consider an event that at least one univariate exceedance happens at time t. Conditioned on this event, there is a non-negligible probability that both margins of the model jumps jointly. Figure 2 illustrates the level of this conditional probability for different values of marginal rates $\tau_1\left(t, u_1\right)$ and $\tau_2\left(t, u_2\right)$. For the calculations, the two-dimensional model with the Gumbel symmetric exponent measure and $\theta(t) = 2$ is used. The conditional probability is calculated as $\left(\tau_1\left(t, u_1\right) + \tau_2\left(t, u_2\right) - \tau\left(t, u_1, u_2\right)\right) / \tau\left(t, u_1, u_2\right)$, where $\tau\left(t, u_1, u_2\right) = \left(\tau_1\left(t, u_1\right)^{\theta(t)} + \tau_2\left(t, u_2\right)^{\theta(t)}\right)^{1/\theta(t)}$. The occurrence of a joint exceedance increases the conditional intensities of both margins, which, in turn, increases the conditional probability of a next co-exceedance. This interplay leads to the clustering in occurrence of simultaneous extreme events.

Figure 2 may also serve to describe the contagion mechanism implied by our model. For example, consider a single exceedance by the first margin at time t. This event instantaneously increases the conditional rate of the first margin, i.e., $\tau_1(t^+, u_1) > \tau_1(t, u_1)$, and hence the multivariate rate, i.e., $\tau(t^+, u_1, u_2) > \tau(t, u_1, u_2)$, where we denote $t^+ = t + \Delta$ for very small $\Delta > 0$. Investigating the contagion, it is of interest to note that the first margin's exceedance does not affect the conditional

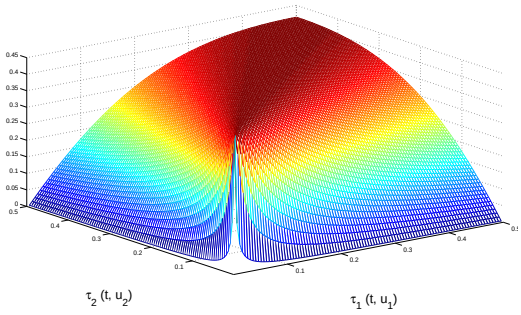

Figure 2: Probability of a joint extreme event at time point t conditioned on the event that at least one of the margins jumps at t.

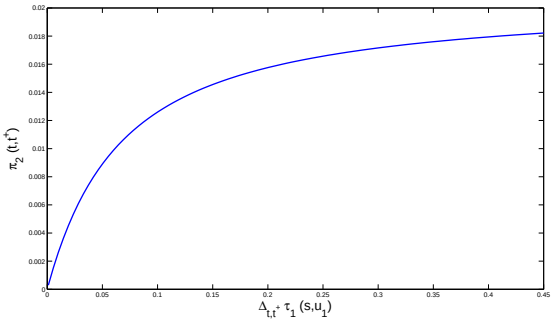

Figure 3: $\pi_2\left(t,t^+\right)$: instantaneous average number of second margin exceedances in the unit interval triggered by the increase of $\Delta_{t,t^+}\tau_1(s,u_1)$ (x-axis) in the first margin's conditional rate.

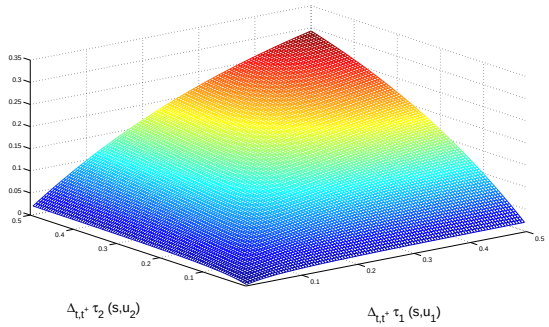

Figure 4: $\pi\left(t,t^+\right)$: increase in the rate of the joint exceedances triggered by a joint exceedance at time t.

rate of the second margin exceedances but affects indirectly the time of their occurrences. This indirect contagion mechanism can be described as follows: an increase $\tau(t,u_1,u_2)\to\tau(t^+,u_1,u_2)$ triggers the occurrence of multivariate (when at least one of the margins jumps) exceedances, which, in turn, may trigger, with the conditional probabilities in Figure 2, the occurrence of simultaneous exceedances, i.e., the exceedance when both the margins jump, which through the self-exciting structure of the processes trigger future jumps. To quantify this contagion impact, we suggest to consider the difference, $\pi_2\left(t,t^+\right)$, between the increments of the conditional rates. $\pi_2\left(t,t^+\right)$ is defined as follows

$$\pi_2\left(t,t^+\right):=\Delta_{t,t^+}\tau_1(s,u_1)-\Delta_{t,t^+}\tau(s,u_1,u_2) \tag{32}$$

with $\Delta_{t_1,t_2}\tau(s):=\tau(t_2)-\tau(t_1)$, t is the time when the first margin jumps, and $t^+=t+\Delta$ for very small $\Delta>0$. From the interpretation of a conditional rate as an instantaneous (i.e., as the time interval tends to zero) conditionally expected number of events in a unit time interval, it follows that $\Delta_{t,t^+}\tau_1(s,u_1)$ and $\Delta_{t,t^+}\tau(s,u_1,u_2)$ express the instantaneous expected number of, respectively, the first margin's or multivariate (when at least one of the margins jumps) exceedances triggered by the first margin's jump at time t. Hence, $\pi_2\left(t,t^+\right)$ describes the instantaneous expected (incremental) number of exceedances by the second margin, triggered by the first margin's event at time t. It is straightforward to show that $\pi_2\left(t,t^+\right)\geq 0$, with $\pi_2\left(t,t^+\right)=0$ in case when the margins are independent. Note that $\pi_2\left(t,t^+\right)$ is defined only for those t's when the first margin's exceendances occur. Figure 3 illustrates $\pi_2\left(t,t^+\right)$ for different values of $\Delta_{t,t^+}\tau_1(s,u_1)$. For the calculations we used the symmetric Gumbel exponent measure with $\theta(t)=2$, $\tau_1(t,u_1)=\tau_2(t,u_1)=0.05$. Similarly to $\pi_2\left(t,t^+\right)$ one can straightforwardly construct $\pi_1\left(t,t^+\right)$. To quantify the effect of the joint exceedances, we suggest to consider $\pi\left(t,t^+\right):=\Delta_{t,t^+}\tau_1(s,u_1)+\Delta_{t,t^+}\tau_2(s,u_2)-\Delta_{t,t^+}\tau(s,u_1,u_2)$, which determines the increase in the rate of the joint exceedances triggered by a joint exceedance at time t. With the settings used for construction of the previous figure, Figure 4 illustrates $\pi\left(t,t^+\right)$ for different values of $\Delta_{t,t^+}\tau_1(s,u_1)$ and $\Delta_{t,t^+}\tau_2(s,u_2)$.

The contagion mechanism provided by our model differs from the one discussed in the literature,

e.g., Aït-Sahalia, Cacho-Diaz, and Laeven (2011), where a jump in one margin directly increases the conditional intensity of the other margins. In our model, this relationship is not automatic but stochastically determined based on the initial level of the intensities, see Figure 2.

2.3.4. Risk Management implications

Conditional marginal expected shortfall. The ability of our model to capture tail co-movements and effects of spreading distress caused by a single marginal exceedance can well be exploited in measuring systemic risk of a financial system. Acharya, Pedersen, Philippon, and Richardson (2010) introduce *marginal expected shortfall* (MES) as an important factor in constructing those systemic risk measures. The MES of a firm is defined as the expected shortfall of the firm X in a crisis. Defining a crisis as the occurrence of an extreme loss Y of the aggregate return on the market, the MES of the firms can be expressed as

$$MES = \mathrm{E}\left(X \mid Y > y^*\right)$$

where X and Y are the loss of the firm and the entire market, respectively, and y^* denotes the loss threshold above that the market is assumed to be in a critical state. By considering the conditional distributions in the above formula, one can obtain the *conditional MES* (notation MES_t) of the firm

$$MES_t = \mathrm{E}\left(X_t \mid Y_t > y^*\right),$$

with X_t and Y_t denoting the conditional up to time t loss distribution of the returns. MES_t can be estimated in the framework of the multivariate model by using conditional distribution (29) of the marks. Indeed, express MES_t as

$$MES_t = \mathrm{E}\left(X_t \mid X_t > u, Y_t > y^*\right) + \mathrm{E}\left(X_t \mid X_t < -u, Y_t > y^*\right) + \mathrm{E}\left(X_t \mid X_t \in [-u, u], Y_t > y^*\right),$$

where u $(-u)$ corresponds to the high (low) quantile of the right (left) tail of X_t. Estimating the bivariate model for point processes of exceedances of $X_t > u$ $(X_t < -u)$ and $Y_t > y^*$, and exploiting the conditional distribution (3) of the marks, the first two expectations in the above equation can easily be found. For estimation of $\mathrm{E}\left(X_t \mid X_t \in [-u, u], Y_t > y^*\right)$, i.e., the non-extreme body of the distribution, one can employ conventional methods.

Portfolio risk management. Consider d financial assets with the corresponding, say, daily negated returns $(R_{1,j}, \ldots, R_{d,j})_{j=1,2,\ldots}$. The daily negated return R_j^P of the portfolio constructed from those assets takes the form $R_j^P = w_1 R_{1,j} + \ldots + w_d R_{d,j}$, where $(w_1, \ldots, w_d)$ are weights of the assets in the portfolio. A typical problem that risk managers face is to estimate the probability

$$\mathrm{P}\left(R_j^P > b \mid \mathbb{H}_{j-1}\right) \tag{33}$$

for some high level of the portfolio loss b, where $\mathbb{H}_{j-1}$ denotes a history including information about past returns $(R_{1,k}, \ldots, R_{d,k})_{k=1,\ldots,j-1}$. Using (31) one can estimate a lower bound for $\mathrm{P}\left(R_j > b \mid \mathbb{H}_{j-1}\right)$. Indeed, assume that the multivariate model with rate (10) is applicable for the point process of exceedances of $(R_{1,j}, \ldots, R_{d,j})_{j=1,2,\ldots}$ above some initial threshold $(u_1, \ldots, u_d)$. Setting b such that $b > w_1 u_1 + \ldots + w_d u_d$, there exists the conditional rate

$$\tau_P(t,b) = \int_{\mathbf{x} \in \{s_1,\ldots,s_d \mid s_1 w_1 + \ldots s_d w_d > b\}} \lambda\left(t, x_1, \ldots, x_d\right) dx_1 \ldots dx_d =$$
$$\int_{u_1}^{\frac{b - w_2 u_2 - \ldots w_d u_d}{w_1}} \ldots \int_{u_k}^{\frac{b - \sum_{l=1}^{k-1} w_l x_l - \sum_{l=k+1}^{d} w_l u_l}{w_k}} \ldots \int_{\frac{b - \sum_{l=1}^{d-1} w_l x_l}{w_d}}^{\infty} \lambda\left(t, x_1, \ldots, x_d\right) dx_d \ldots dx_1,$$

which is derived from (31). Here, $\tau_P(t,b)$ is the conditional rate of the point process $N_P(t,b)$ that counts the number of exceedances of level b by $R_j^P = w_1 R_{1,j} + \ldots + w_d R_{d,j}$, $j = 1, \ldots, [t]$, for $R_{i,j} \geq 0$, $i = 1, \ldots, d$. Note that $N_P(t,b)$ is a continuous process, but we observe R_j only at discrete times $j = 1, 2, \ldots$. Taking this discreteness into account, we can approximate the lower bound for the conditional probability that R_j^P crosses the level b as the probability of at least one

exceedance of $N_P(t,b)$ in period $(j, j+1]$. Formally,

$$\mathrm{P}\left(R_j^P > b \mid \mathbb{H}_{j-1}\right) \geq 1 - \mathrm{P}\left(N_P(j,b) - N_P(j-1,b) = 0 \mid \mathbb{H}_{j-1}\right) = 1 - \exp\left(-\int_{j-1}^{j} \tau_P(t,b)dt\right).$$

3. Estimation, Goodness-of-Fit and Simulation

Maximum likelihood (ML) method is the most obvious and convenient approach for the estimation of intensity-based models, see Daley and Vere-Jones (2005), Section 7. Statistical fitting with ML of the multivariate model of exceedances presented in Section 2.2 can proceed through multi-stage ML or construction of the full likelihood function. In the multi-stage ML the parameters of the marginal point processes of exceedances $(\tau_i(t,x_i)$, for $i = 1,\ldots,d)$ are estimated via univariate ML, and then the remaining parameters of $\tau(t,x_1,\ldots,x_d)$ are estimated conditional on the estimated parameters for $\tau_i(t,x_i)$, for $i = 1,\ldots,d$. Similarly as for copula-based models, see Patton (2012), the multi-stage ML is computationally tractable, but it suffers from the loss of efficiency compared to the estimation via the full likelihood function. Besides MLE, one may consider estimation with the generalized method of moments (GMM), see Aït-Sahalia, Cacho-Diaz, and Laeven (2011). We relegate the details of the GMM estimation procedure to Appendix C and discuss it in more details in the empirical part of the paper.

We consider the univariate ML estimation in Section 3.1. Construction of the full likelihood function is described in Section 3.2. The goodness-of-fit and the simulation algorithm are presented, respectively, in Section 3.3 and Section 3.4.

Recall the notation: $(X_{1,j},\ldots,X_{d,j})$, $j = 1,2,\ldots,T^*$ are the observations, $\mathbf{u} = (u_1,\ldots,u_d)$ is the initial threshold, i.e., u_i is a sufficiently high quantile of $(X_{i,j})_{j=1,2,\ldots,T^*}$; $T_{i,k}$ and $\tilde{X}_{i,k}$, with $k = 1,\ldots,N_{u_i}$, stand for the times and marks of the marginal exceedances over u_i. By T_k, $k = 1,\ldots,N_u$, we denote the times when $\mathbf{X}_t$ exceeds $\mathbf{u}$ in at least one component.

3.1. Univariate model estimation

The likelihood function, denote it as L_i, of the SE-POT model for the marginal rate of exceedances $\tau_i(t,x_i)$ is of the form, see McNeil, Frey, and Embrechts (2005),

$$L_i = \exp\left(-T^*\tau_i - \psi_i \int_0^{T^*} v_i^*(s)ds\right) \prod_{j=1}^{N_{u_i}} \lambda_i\left(T_{i,j}, \tilde{X}_{i,j}\right), \tag{34}$$

where N_{u_i} is the number of the marginal exceedances above u_i, and

$$\lambda_i(t,x_i) = \frac{\tau_i + \psi_i v_i^*(t)}{\beta_i + \alpha_i v_i^*(t)} \left(1 + \xi_i \frac{x_i}{\beta_i + \alpha_i v_i^*(t)}\right)^{-1/\xi_i - 1}$$

is the conditional intensity of the self-exciting POT model with predictable marks. The intensity $\lambda_i(t,x_i)$ is derived from the equation

$$\int_{x_i}^{\infty} \lambda_i(t,s)ds = \tau_i(t,x_i).$$

With function v_i^* as in (2), $g_i(s) = e^{-\gamma_i s}$, and $c(x)$ as in (6), the integral in (34) takes the form

$$\int_0^{T^*} v_i^*(s)ds = \int_0^n \sum_{j:0<T_{i,j}<s} e^{-\gamma_i(s-T_{i,j})}\left(1 + G^{\leftarrow}\left(F_{i,j}\left(\tilde{X}_{i,j}\right)\right)\right)ds =$$
$$\frac{1}{\gamma_i}\sum_{k=1}^{N_{u_i}}\left[\left(e^{-\gamma_i T_{i,k}} - e^{-\gamma_i T_{i,k+1}}\right)\sum_{j=1}^{k} e^{\gamma_i T_{i,j}}\left(1 + G^{\leftarrow}\left(F_{i,j}\left(\tilde{X}_{i,j}\right)\right)\right)\right],$$

where $F_{i,j}(\cdot)$ is the conditional distribution function of the marks defined in (3) and $G^{\leftarrow}(\cdot)$ is an inverse of the distribution function G of some continuous positive random variable (compare to Section 2.1.2). In the above equation $T_{i,N_{u_i}+1}$ should be interpreted as T^*. For the theoretical

background behind the likelihood function (34) consult Daley and Vere-Jones (2005), Proposition 7.2.III. Consistency and asymptotic normality of the MLE estimator are shortly discussed at the end of the next section.

3.2. Multivariate model estimation

As the first step before expressing the complete likelihood function for the multivariate model of exceedances, we consider the procedure of constructing the densities of the observed events. Consider an event that the k-th multivariate exceedance occurs at time $T_k = t$ with, for example, only the first and second (out of $d > 2$) margins exceeding the initial threshold with the corresponding marks $\tilde{X}_{1,1} > u_1$ and $\tilde{X}_{2,1} > u_2$. Conditioning on the fact that the previous event occurred at time T_{k-1}, the density of the event at time T_k reads

$$\mathrm{p}\left(T_k = t \mid \mathbb{H}_{T_{k-1}}\right) \mathrm{p}\left(X_{1,t} = \tilde{X}_{1,1}, X_{2,t} = \tilde{X}_{2,1}, X_{3,t} \leq u_3, \ldots, X_{d,t} \leq u_d \mid T_k = t\right), \tag{35}$$

where $\mathrm{p}(A)$ denotes the density of A.

The first part of the above density can easily be obtained from (9) in the following form

$$\mathrm{p}\left(T_k = t \mid \mathbb{H}_{T_{k-1}}\right) = \tau(t, u_1, \ldots, u_d) \exp\left(-\int_{T_{k-1}}^{t} \tau(s, u_1, \ldots, u_d) ds\right)$$

and, due to (29), the second term of (35) reads

$$\mathrm{p}\left(X_{1,t} = \tilde{X}_{1,1}, X_{2,t} = \tilde{X}_{2,1}, X_{3,t} \leq u_3, \ldots, X_{d,t} \leq u_d \mid T_k = t\right) = \\ - \left.\frac{\partial^2 \tau(t, x_1, x_2, u_3 \ldots, u_d)}{\partial x_1 \partial x_2}\right|_{x_1 = \tilde{X}_{1,1}, x_2 = \tilde{X}_{2,1}} \frac{1}{\tau(t, u_1, \ldots, u_d)}.$$

Finally, (35) takes the form

$$- \left.\frac{\partial^2 \tau(t, x_1, x_2, u_3 \ldots, u_d)}{\partial x_1 \partial x_2}\right|_{x_1 = \tilde{X}_{1,1}, x_2 = \tilde{X}_{2,1}} \exp\left(-\int_{T_{k-1}}^{t} \tau(s, u_1, \ldots, u_d) ds\right).$$

The form of the above density is typical for all multivariate exceedances. To write down the complete likelihood function of the whole sample observed multivariate exceedances occurred in time interval $[0, T^*]$ at times $T_1, T_2, \ldots, T_{N_u}$, we adopt the following notation. Let

$$\mathbf{ind}(t) = \left\{i_1, \ldots, i_p : X_{i_1,t} > u_{i_1}, \ldots, X_{i_p,t} > u_{i_p}\right\} \in (1, 2, \ldots, d)$$

be the set of indices of the margins which exceed the initial margin at time t. We use $\mathbf{ind}(t)$ to define the following sets: $\mathbf{x}_{\mathbf{ind}(t)} = \left(x_{i_1}, \ldots, x_{i_p}\right)$, $\mathbf{x}_{-\mathbf{ind}(t)} = \mathbf{x} \setminus \mathbf{x}_{\mathbf{ind}(t)}$, where $\mathbf{x} = (x_1, \ldots, x_d)$. With those sets, the final likelihood function takes the form

$$L = \exp\left(-\int_0^{T^*} \tau\left(s, u_1, \ldots, u_d\right) ds\right) \prod_{j=1}^{N_u} \left[- \left.\frac{\partial \tau(t, x_1, x_2, \ldots, x_d)}{\partial \mathbf{x}_{\mathbf{ind}(T_j)}}\right|_{\mathbf{x}_{\mathbf{ind}(T_j)} = \mathbf{X}_{\mathbf{ind}(T_j)}, \mathbf{x}_{-\mathbf{ind}(T_j)} = \mathbf{u}_{-\mathbf{ind}(T_j)}}\right], \tag{36}$$

where $\frac{\partial \tau(t, x_1, x_2, \ldots, x_d)}{\partial \mathbf{x}_{\mathbf{ind}(T_j)}}$ is a multiple partial derivative of $\tau(t, x_1, x_2, \ldots, x_d)$ by all x_i with $i \in \mathbf{ind}\left(T_j\right)$. The integral in (36) cannot be solved explicitly. Hence, in practice, this integral is approximated by a sum over all observations.

Considering the properties of the MLE estimators, note that the multivariate model of exceedances treats the data as a realization of a univariate point process on $(0, T^*]$. This property is advantageous for estimation, because, as it is mentioned in Bowsher (2007), there are currently no results concerning the properties of the MLE for multivariate point processes. For the univariate case, it is shown in Ogata (1978), that under some regularity conditions, the MLE for a stationary,

simple[2] point process is consistent and asymptotically normal as $T^* \to \infty$. Concerning these conditions, note that our multivariate model is stationary if both the marginal processes of exceedances (Section 2.1.3) and the dependence parameter (namely risk factors $\mathbf{z}_t$) in (28) are stationary. Employing the results from Ogata (1978), there is a subtle point considering the conditional set used for the intensity evaluation in the likelihood, namely, whether the complete information set from $(-\infty, T^*)$ or rather incomplete (practically available) information set from $(0, T^*)$ is used. Nevertheless, it was noted in the paper that the likelihood evaluated on $(0, T^*)$ for the Hawkes self-exciting process with the *exponential* decay function satisfies the required regularity conditions. It should also hold for our model, because exactly the exponential decay function is suggested for both the SE-POT model and the dependence parameter. Another result from Ogata (1978), which we will intensively use in the empirical part of the paper, is that under the null hypothesis the likelihood ratio test statistics satisfies asymptotically the standard χ^2 distribution.

3.3. Goodness-of-fit

Applying the models of marginal and joint exceedances in practice, where true probabilities are unknown, it is vital to perform a goodness-of-fit procedure to check the performance of the model. Our approach to the goodness-of-fit test is based on the probability integral transformation (Diebold, Gunther, and Tay, 1998) of the sample of times of marginal exceedances $T_{i,1}, \ldots, T_{i,N_{u_i}}$ with the estimated conditional intensity $\tau_i(t, u_i)$. Recalling that for a continuous random variable X with distribution function F, $F(X)$ is uniformly distributed on the unit interval, we obtain from (5) that residuals $\chi_{i,j}$, $j = 1, \ldots, N_{u_i} - 1$ defined as

$$\chi_{i,j} = \int_{T_{i,j}}^{T_{i,j+1}} \tau_i(s, u_i) ds, \qquad j = 1, 2, \ldots, N_{u_i} - 1, \qquad i = 1, \ldots, d,$$

are independent realizations from the standard exponential distribution. Since the goodness-of-fit looks for evidence that the model is misspecified, the test of the estimated model can be limited to checking a hypothesis that the residuals are independent realizations from the standard exponential distribution.

With the analogous consideration as for the times of marginal exceedances, due to (9) we obtain a similar result for the times of mulativariate exceedances. For $T_1, \ldots, T_{N_u}$ the residuals defined as

$$\chi_j = \int_{T_j}^{T_{j+1}} \tau(s, u_1, \ldots, u_d) ds, \qquad j = 1, 2, \ldots, N_u - 1, \tag{37}$$

are also independent realizations from the standard exponential distribution. We will refer to the sample of $\chi_{i,j}$ or χ_j as *residual (marginal) inter-exceedance intervals*. Note that the goodness-of-fit test based the standardized (marginal) inter-exceedance times directly corresponds to the random time transformation of point processes, for details, see Section 7.4 in Daley and Vere-Jones (2005).

To test the fit of the model in describing the marks of exceedances, we employ the concept of the probability integral transformation as well. With the conditional distribution (3), the *residual marks* defined as

$$m_{i,j} = -\frac{1}{\xi_i} \log\left(1 + \frac{X_{i,T_{i,j}} - u_i}{\beta_i + \alpha_i v_i^*(T_{i,j})}\right), \qquad j = 1, 2, \ldots, N_{u_i}, \qquad i = 1, \ldots, d,$$

should be independent realizations from the standard exponential distribution, if the estimated model is suitable. The goodness-of-fit can be checked either graphically using QQplots, or using a formal goodness-of-fit test such as the Kolmogorov-Smirnov or Anderson-Darling tests to test whether the estimates of χ_j (or $\chi_{i,j}$ or $m_{i,j}$) follow the standard exponential distribution. As part of a goodness-of-fit procedure one can also analyze the ability of random data simulated from the model to reproduce certain characteristics of the data. The simulation procedure is described in the next section.

Note that the residual inter-exceedance intervals and marks are not only useful for measuring the model's fit, but they also form the basis for the moment estimator discussed in Appendix C.

[2] A point process is simple, if no two events occur at the same time.

3.4. Simulation

To simulate from the multivariate model of exceedances, we need to obtain both times and marks of those exceedances. Equation (37) offers an easy simulation procedure for the times of multivariate exceedances. It follows that conditioned on T_j, a realization of T_{j+1} can be found solving

$$\int_{T_j}^{T_{j+1}} \tau(s, u_1, \ldots, u_d)ds = E, \qquad j = 1, 2, \ldots,$$

where E is the standard exponential random variable. To find T_1, we set $T_0 = 0$ in the above equation. The above simulation method is known in the literature as the inverse method, see Daley and Vere-Jones (2005) (Algorithm 7.4.III). Alternatively, in order to omit calculation of the integral in the equation above, one can simulate the times of exceedances by the thinning algorithm, see Ogata (1981). It is a simple and efficient method that require the specification of the conditional intensity only (without a need of solving any integrals).

To simulate the marks at times of multivariate exceedance, first, it should be identified which margin exceeds the initial threshold, and then the marks should be simulated from the appropriate conditional distribution. In the two-dimensional model, for example, the marks should be simulated as follows:

i) with probability $\mathrm{P}\left(X_{1,t} > u_1, X_{2,t} \leq u_2 \mid T_{j+1} = t\right) = 1 - \frac{\tau_2(t,u_2)}{\tau(t,u_1,u_2)}$, only the first margin exceeds the initial threshold. The mark $X_{1,t}$ should be simulated from the following distribution

$$\mathrm{P}\left(X_{1,t} \leq x_1 \mid T_{j+1} = t, X_{2,t} \leq u_2\right) = \frac{\tau\left(t, u_1, u_2\right) - \tau\left(t, x_1, u_2\right)}{\tau\left(t, u_1, u_2\right) - \tau_2\left(t, u_2\right)}, \qquad x_1 > u_1.$$

ii) with probability $\mathrm{P}\left(X_{1,t} \leq u_1, X_{2,t} > u_2 \mid T_{j+1} = t\right) = 1 - \frac{\tau_1(t,u_1)}{\tau(t,u_1,u_2)}$, only the second margin exceeds the initial threshold. The mark $X_{2,t}$ should be simulated from the following distribution

$$\mathrm{P}\left(X_{2,t} \leq x_2 \mid T_{j+1} = t, X_{1,t} \leq u_1\right) = \frac{\tau\left(t, u_1, u_2\right) - \tau\left(t, u_1, x_2\right)}{\tau\left(t, u_1, u_2\right) - \tau_1\left(t, u_1\right)}, \qquad x_2 > u_2.$$

iii) with probability $\mathrm{P}\left(X_{1,t} > u_1, X_{2,t} > u_2 \mid T_{j+1} = t\right) = \frac{\tau_1(t,u_1)+\tau_2(t,u_2)-\tau(t,u_1,u_2)}{\tau(t,u_1,u_2)}$, both the first and the second margin exceed the initial threshold. The mark $(X_{1,t}, X_{2,t})$ should be simulated from the following distribution

$$\mathrm{P}\left(X_{1,t} \leq x_1, X_{2,t} \leq x_2 \mid T_{j+1} = t, X_{1,t} > u_1, X_{2,t} > u_2\right) = \frac{\tau\left(t, x_1, u_2\right) + \tau\left(t, u_1, x_2\right) - \tau\left(t, x_1, x_2\right) - \tau\left(t, u_1, u_2\right)}{\tau_1\left(t, u_1\right) + \tau_2\left(t, u_2\right) - \tau\left(t, u_1, u_2\right)},$$

for $x_1 > u_1$ and $x_2 > u_2$.

The above cases for marginal exceedances are a direct consequence of the conditional distributions of marks in (29) and hence can be extended for multivariate cases with any $d > 2$.

4. Application to Financial Data

In this section we illustrate an application of the model of multivariate exceedances to financial data. For illustration we consider a two- and four-dimensional application of the model, and focus on describing the behavior of extreme negative returns in financial markets worldwide and in the European banking sector. The choice of the data for and dimension of the models is motivated by illustrative reasons only.

4.1. Data and Preliminary Analysis

To illustrate the application of the two dimensional model we consider extreme negative returns in European and the US financial markets, which we approximate by daily log-returns of the Morgan Stanley Capital International index for the US (MSCI-USA) and Europe (MSCI-EU) covering the

Table 1: Summary statistics

	MSCI-USA	MSCI-EU	DB	HSBC	RBS	UBS
Mean(%)	0.0232	0.0155	-0.0025	0.0192	-0.0246	-0.0100
St.Deviation	0.0116	0.0126	0.0234	0.0190	0.0330	0.0231
Skewness	-0.2468	-0.1844	0.1968	-0.2152	-8.2565	0.1289
Excess Kurtosis	8.7890	8.0630	8.9337	8.6815	272.8115	12.5427

period January 1, 1990 to January 13, 2012. The data consists of 5749 observations. The MSCI-USA index is designed to measure large and mid cap equity performance of the US equity market, whereas the MSCI Europe Index measures the equity performance of the developed markets in Europe and consists of the following country indices: Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, the Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, and the United Kingdom[3].

The four-dimensional model is applied to negative equity returns of four major European banks: Deutsche Bank (DB), HSBC Holdings (HSBC), Royal Bank of Scotland (RBS), and United Bank of Switzerland (UBS), – embracing the period October 20, 1993 to January 13, 2012. The sample consists of 4768 observations. The summary statistics for all of the time series can be found in Table 1. Note that the extremely high skewness and kurtosis for RBS is the effect of several extremely large negative returns. We decided to keep these observations since the methods we apply have a certain robustness to outliers and we are in fact interested in very extreme events. Furthermore, the estimate of the tail index of left tail of RBS return time series (see Table 4) is larger than 0.25 suggesting that kurtosis does not exist for this time series. Note that for estimation of our model we use negated daily log-returns on the equity, allowing us to look at the upper rather than the lower tail.

For estimation of the multivariate model the initial threshold was set on the 97.7% quantile of the empirical distributions of MSCI-USA and MSCI-EU series, which corresponds, respectively, to 2.4922% and 2.8601% and results in 132 marginal exceedances for the two indexes and in 53 joint exceedances. For the bank data the initial thresholds are the following: 5.3201% for DB, 4.2021% for HSBC, 5.8534% for RBS, and 5.1799% for UBS. Those threshold correspond to the 97.9% quantile of the empirical distributions and result in 100 marginal (for all four indexes) and 16 joint exceedances. A preliminary analysis motivating this choice of the thresholds and verifying the extreme value condition can be found in Appendix D.

4.2. Copula Choice

The parametric specification for exponent measure V_t in (10) is still open and there are many parametric families of dependence structure in multivariate EVT. With respect to applications, the dependence structure should both be as flexible as possible and be able to capture an asymmetric dependence structure, in the sense that $V_t(y_1, y_2) \neq V_t(y_2, y_1)$. This allows for asymmetric responses of the probability of joint exceedances to exceedances of the individual variables implying interesting economic interpretations. For example, the stock market of a small country may react strongly to shocks to the US stock market, but not vice versa. We suggest to use the exponent measure of the Gumbel copula[4]. It has a simple structure with only one parameter $\theta \geq 1$, which makes it easy to add the time-dependent part and to extend it to an asymmetric form. It also can be extended to dimensions beyond two, which is advantageous for the multivariate application in Section 4.3.2. Furthermore, its dependence function in the tail is almost identical to the one of the t-copula for any choice of the parameters of the t-copula and is thus very flexible, see Demarta and McNeil (2005) for details.

The non-exchangeable Gumbel copula, see Tawn (1990), has the following exponent measure

$$V(y_1, \ldots, y_d) = \sum_{s \in S} \left\{ \left(\sum_{i \in s} w_{i,s}/y_i \right)^{\theta_s} \right\}^{1/\theta_s}, \tag{38}$$

[3]See www.msci.com for details.

[4]We initially also considered the Galambos copula, but its fit was inferior for all applications we considered.

where S is the set of all non-empty subsets of $\{1,\dots,d\}$ and the parameters are constrained by $\theta_s \geq 1$ for all $s \in S$, $w_{i,s} = 0$ if $i \notin s$, $w_{i,s} \geq 0$ (asymmetry parameters), $i = 1,\dots,d$ and $\sum_{s\in S} w_{i,s} = 1$ see also Coles and Tawn (1991). V in (38) is overparameterized for most applications, as it contains $2^{d-1}(d+2)-(2d+1)$ parameters. The task of estimating such a model is very similar to the estimation of high-dimensional copula-based models, for which pair copula construction is an effective solution to overcome a proliferation of parameters, while maintaining the flexible dependence structure of the model, see, for example, Aas, Czado, Frigessi, and Bakken (2009), Okhrin, Okhrin, and Schmid (2013). The idea of pair copula construction may also be transferred to construction of multivariate point processes with intensity (10). The only condition to preserve is that the exponent measure associated with the final rate must be one of an extreme value copula. We leave this estimation topic for future research as it is beyond the scope of this paper.

4.3. Applying the Model

4.3.1. Two-dimensional Model

In this section, we focus on extreme negative log-returns of MSCI-USA and MSCI-EU indexes and estimate the multivariate model of exceedances with the two-dimensional version of (38):

$$V_t(y_1,y_2) = \frac{(1-w_1)}{y_1} + \frac{(1-w_2)}{y_2} + \left(\left(\frac{w_1}{y_1}\right)^{\theta(t)} + \left(\frac{w_2}{y_2}\right)^{\theta(t)}\right)^{1/\theta(t)}, \tag{39}$$

where w_1 and w_2 denote the asymmetry parameters in the dependence structure. Setting $w_1 = w_2 = 1$, the symmetric version of (39) is obtained. Based on the discussion in Section 2.2.2, we parametrize dependence parameter $\theta(t)$ in the equation above as the Hawkes process of different exceedances and, in order to keep the model closed, we keep the specification free from exogenous risk factors. We set

$$\theta(t) = \theta_{2,0} + \int_0^t e^{-\gamma_{2,0}(t-s)}\left[\psi_{2,1}dN_{2,1}(s) + \psi_{2,2}dN_{2,2}(s) + \psi_{2,3}dN_{2,3}(s)\right], \tag{40}$$

with $\theta_{2,0} \geq 1$, $\gamma_{2,0} > 0$, $\psi_{2,i} \geq 0$, $i = 1,2,3$, where $N_{2,1}(s)$ and $N_{2,2}(s)$ are counting measures of exceedances of negated log-returns of, respectively, MSCI-US and MSCI-EU above the corresponding initial thresholds. $N_{2,3}(s)$ is a counting measure of the simultaneous exceedances. Such a construction of the dependence parameter, along with the asymmetry of the dependence structure, provides a certain level of flexibility in the dependence modelling. We conduct six likelihood tests in order to identify an appropriate model. In particular, we test the hypothesis that $w_1 = w_2 = 1$ (p-value 0.6574), $w_1 = 1$ (p-value 0.9999), $w_2 = 1$ (p-value 0.3597), $\psi_{2,1} = 0$ (p-value 0.9999), $\psi_{2,2} = 0$ (p-value 0.0243), and, finally, $\psi_{2,3} = 0$ (p-value 0.5401). This hypothesis testing allows for a certain simplification in the dependence structure, namely, symmetry, and provides a surprising insight that large negative exceedances of MSCI-EU have a decisive influence on the strength of tail dependence between negative log-returns of MSCI-EU and MSCI-US indices. Note that estimation of the model was conducted according to the one-step MLE procedure discussed in Section 3.

Having estimated the bivariate model in one step, we report the parameter estimates of the SE-POT model in Table 2 and estimates of the dependence parameter in Table 3. To compare, we report also in the tables the estimates obtained by the one-step method of moments (MM), see Appendix C. Note that the influence and decay functions of the SE-POT models were set in the way as discussed in Section 2.1.2, namely, $g(s) = e^{-\gamma s}$ and $c^*(u) = 1 - \delta\log(1-u)$.

The estimated branching coefficients of the SE-POT model are all smaller than one, which suggests that the processes are indeed stationary, although the large estimates of tail index (ξ_i) by both the MLE and the MM suggests that exceedances' heavy-tails can only poorly be explained by the time-varying volatility component $v_i(t)$ indicating a substantial downward potential caused by *jumps* on the US and European financial markets. Without distinguishing the reason for exceedances, our model, being by definition a jump process, can well account for time-varying volatility by incorporating this feature into the conditional distribution of the marks. Consider Appendix E for the goodness-of-fit statistics.

Furthermore, comparing the MLE estimates of δ_i and γ_i, note that the marks of MSCI-EU exceedances trigger the occurrence of future exceedances much stronger than MSCI-USA exceedances do, which corresponds to our previous conclusion that the MSCI-EU exceedances are decisive in

Table 2: Parameter estimates of the SE-POT model by the MLE and the MM. An inverse Hessian of the likelihood function is used to obtain the standard errors reported in parentheses right to the MLE estimates.

	MSCI-US		MSCI-EU	
Parameter	MLE	MM	MLE	MM
τ_i	0.0068 (0.0016)	0.0068 (0.0019)	0.0055 (0.0014)	0.0066 (0.0027)
ψ_i	0.0173 (0.0052)	0.0208 (0.0458)	0.0149 (0.0040)	0.0184 (0.0219)
γ_i	0.0404 (0.0103)	0.1428 (0.3476)	0.0463 (0.0132)	0.0845 (0.0904)
δ_i	0.6387 (0.1480)	3.8415 (3.5759)	1.1710 (0.2000)	2.2767 (1.9937)
ξ_i	0.2169 (0.1158)	0.2376 (0.0296)	0.2311 (0.1238)	0.2637 (0.0693)
β_i	0.4623 (0.0903)	0.3353 (0.0916)	0.4145 (0.0811)	0.3327 (0.1229)
α_i	0.1236 (0.0382)	0.1626 (0.4356)	0.1042 (0.0337)	0.1149 (0.1345)
branch. coeff.	0.7024	0.7053	0.6996	0.7141

modelling the dependence parameter. Note that compared to the MLE estimates, the MM ones suggest a completely different mechanism of the marks' impact. Although their impact is very large, its "trigger" power diminishes every day (in absence of further exceedances) with a much faster rate than one provided by the MLE estimates. For MSCI-US estimates, for example, those rates are, respectively, $1-\exp(-0.1428) \approx 13.31\%$ and $1-\exp(-0.0404) \approx 3.96\%$. This interplay between the impact and decay functions determines one mechanism of asymmetric responses of marginal events on the rate of multivariate exceedances. The marginal conditional rates of exceeding the initial threshold (calculated with the MLE estimates) are illustrated in Figure 5. For the corresponding figure for the MM estimates see Appendix F.

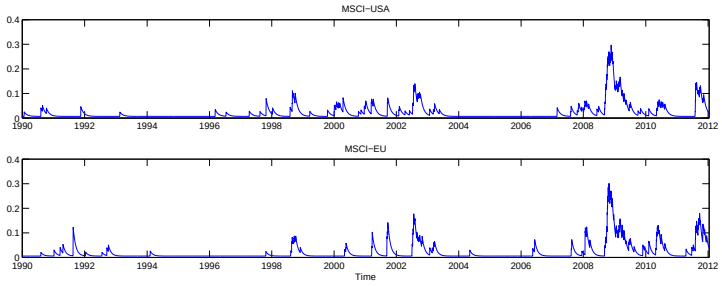

Figure 5: Estimated conditional rate of the marginal exceedances over the initial threshold for MSCI-USA and MSCI-EU. MLE estimates from Table 2.

Table 3: Parameter estimates of the dependence parameter. An inverse Hessian of the likelihood function is used to obtain the standard errors reported in parentheses right to the MLE estimates.

Parameter	MLE	MM
$\theta_{2,0}$	1.0294 (0.0285)	1.3534 (1.1089)
$\gamma_{2,0}$	0.0147 (0.0051)	0.3345 (9.0492)
$\psi_{2,2}$	0.0946 (0.0269)	1.4941 (47.6108)

The MLE estimates of the dependence parameter indicate a strong influence of the MSCI-EU exceedances on the strength of dependence: $\theta(\cdot)$ jumps by about 10% of its base level of 1.0294 at every time MSCI-EU exceeds its initial threshold. Based on the MLE estimates, Figure 6 plots the time varying dependence parameter $\theta(\cdot)$ (left panel) and the estimated conditional probabilities of multivariate events when at least one margins exceed the initial threshold compared with the (constant) empirical probability of those events (right panel). The MM estimates suggest even a higher influence of the MSCI-EU exceedances on $\theta(\cdot)$, but, similarly to the case of the SE-POT model, that influence diminishes very fast due to a large estimate of $\gamma_{2,0}$ causing an erratic behavior in the dependence parameter, see Figure F.19 in Appendix F. Note that extremely high standard errors for the MM estimates correspond to our general finding based on the simulation studies that MM is inappropriate for our multivariate model. The reason may lie both in the choice of moment conditions and in the fact that all moment conditions are based on the goodness-of-fit statistics,

which cannot be calculated from the sample independently from the unknown parameters of the models.

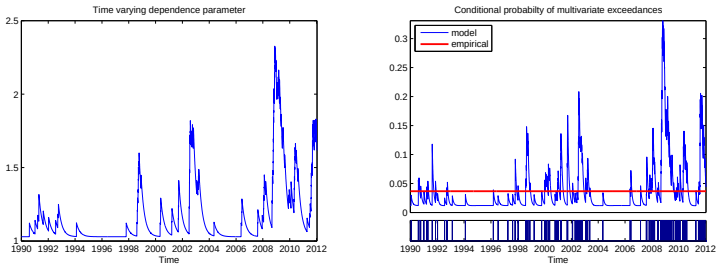

Figure 6: The estimated time-varying dependence parameter (left-hand panel) and the conditional probability of multivariate events when at least one margins exceed the initial threshold (right-hand panel) in the two dimensional model. The tick marks at the bottom of the right panel denote times of multivariate events.

It is important to be aware of the fact that the symmetrical dependence structure, with which the bivariate model was estimated, does not mean that the marks of exceedances have a fixed effect on the conditional multivariate rate. Along with the estimates of the impact function, this effect strongly depends on the exceedance history of the marginal processes, which is propagated by self-exciting component $v_i(t)$. That history affects the marks' conditional distribution through (3), which, in turn, may provide asymmetry in the marks' influence. For illustration, consider Figure 7 which displays how the conditional rate of joint exceedances might be influenced by different values of MSCI-EU and MSCI-US negated returns that could have happened on, say, 01.03.2009 (left panel) and 15.02.2010 (right panel). The calculation are based on the MLE estimates. One can clearly observe the change of the marks' influence depending on the time of the analysis.

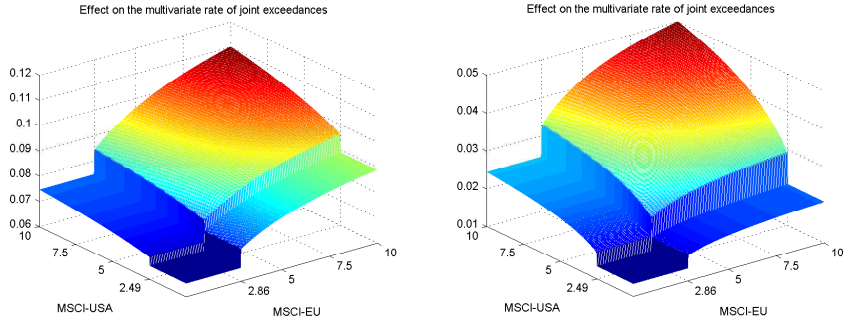

Figure 7: Effects of different values of MSCI-EU and MSCI-US negated returns, that could have happened on 01.03.2009 (left panel) and 15.02.2010 (right panel), on the next day's conditional rate of joint exceedances.

Based on the MLE estimates, Figure 8 shows the exponential QQ-plot of the residual inter-exceedances intervals of the bivariate model and their empirical autocorrelation functions. The figures clearly illustrate the residual inter-exceedances intervals do not deviate much from the standard exponential distribution suggesting the theoretical consistency of the model. This suggestion is further supported by analysing the residual intervals with the Kolmogorov-Smirnov and Ljung-Box (15 lags) tests, which failed to reject the null of, respectively, the standard exponential distribution and no autocorrelation with p-values 0.5673 and 0.3817. The goodness-of-fit tests for the marginal exceedance processes are reported in Appendix E.

4.3.2. Four-dimensional Model

In this Section, we focus on extreme negative equity returns of four major European banks: Deutsche Bank (DB), HSBC Holdings (HSBC), Royal Bank of Scotland (RBS), and UBS. To save space, we restrict ourselves to maximum likelihood estimation. Furthermore, fitting the model we prefer the sequential estimation procedure over the estimation in one step, see Section 3.2. The reason for this choice is that a proliferation of the model parameters hinders handling of the likelihood function. To cope with this problem, we estimate first the SE-POT model for marginal

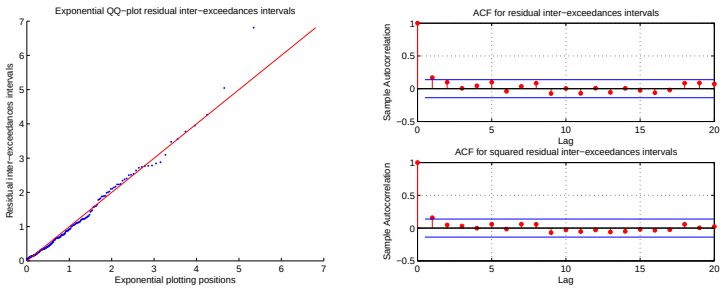

Figure 8: Exponential QQ-plot of the residual inter-exceedance intervals (left-hand panel) in the bivariate model. The sample autocorrelation function of those (squared) intervals (right-hand panel).

exceedances, and then, conditioned on those estimates, the parameters relating to the dependence parameter. Parameter estimates of the SE-POT model are reported in Table 4. Note that the influence and decay functions of the SE-POT models were set in the way as discussed in Section 2.1.2, namely, $g(s) = e^{-\gamma s}$ and $c^*(u) = 1 - \delta \log(1-u)$.

Table 4: MLE parameter estimates of the SE-POT model. An inverse Hessian of the likelihood function is used to obtain the standard errors reported in parentheses right to the estimates.

Parameter	DB	HSBC	RBS	UBS
τ_i	0.0050 (0.0035)	0.0046 (0.0020)	0.0028 (0.0062)	0.0031 (0.0011)
ψ_i	0.0289 (0.0219)	0.0179 (0.0087)	0.0175 (0.0210)	0.0332 (0.0101)
γ_i	0.0584 (0.0829)	0.0318 (0.0116)	0.0281 (0.0557)	0.0482 (0.0395)
δ_i	0.5444 (0.2235)	0.3963 (0.1131)	0.4100 (0.1356)	0.2360 (0.1438)
ξ_i	0.0744 (0.7601)	0.1053 (0.0267)	0.3150 (0.2530)	-0.0045 (0.4936)
β_i	1.1457 (0.2414)	1.1037 (0.1925)	1.0846 (0.1740)	1.3582 (1.6565)
α_i	0.2721 (0.1342)	0.1750 (0.0840)	0.3271 (0.3973)	0.3686 (0.1677)
branch. coeff.	0.7548	0.7836	0.8773	0.8508

Considering the efficiency of the estimates of dependence parameter, which are of concern here, note that the majority of the loss of efficiency associated with multi-stage estimation is attributed to estimation of shape parameter ξ_i, see Tawn (1990).

The estimate of tail parameter ξ_i for the UBS's returns is effectively zero, which corresponds to the tail index of a normal (light-tailed) distribution. This observations does not mean, however, that UBS's negative returns are unconditionally light-tailed: the phenomenon of fat tails and serial dependence often go together, see Andriani and McKelvey (2007). A serially dependent process with light-tailed marks and time-varying volatility may still be estimated to have heavy tails, if treated as an i.i.d sample. For example, an unconditional estimate of the UBS's tail parameter is $\hat{\xi} = 0.1145(0.1249)$. Along with the time-varying volatility, heavy tails may result from jumps in the underlying process (Bollerslev, Todorov, and Li, 2013), which, in our model, are accounted for by the tail index estimate. For the UBS's returns, it seems like the marks' heavy tails are well captured through incorporating the self-exciting component $v_i(t)$ into the marks' conditional distribution. The large tail parameter estimate of the RBS's returns suggests however that $v_i(t)$ cannot explain all variation in the tails, hence (assuming the model is correct) there is a significant (unexpected) jump component with tail parameter $\hat{\xi} = 0.3150(0.2530)$ determining the distribution of the marks. Note that an unconditional estimate of the RBS's tail parameter is $\hat{\xi} = 0.5452(0.1513)$. The conditional rates of exceeding the initial threshold are reported in Figure 9.

To estimate dependence structure of our model, we employ the Gumbel exponent measure (38) in the four-dimensional case, which provides 39 parameters responsible for the dependence structure. Clearly, this model is overparameterized for most applications and therefore some preliminary analysis must be conducted to simplify that dependence structure. To do that, we first analyse the bivariate dependencies for asymmetry. The procedure is the following: we fit the bivariate model of excedances with the dependence parameter as in (40) for all pairs of the four-dimensional data and then conduct a likelihood test on the hypothesis that the dependence structure is symmetric. Table 5 reports p-values of those likelihood tests.

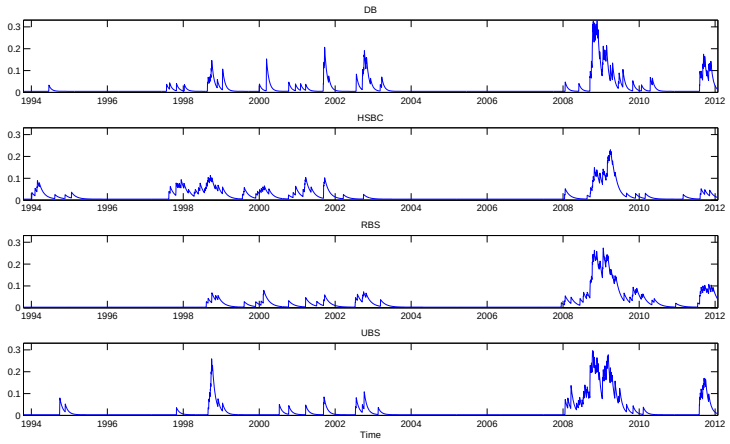

Figure 9: The estimated conditional rates of the marginal exceedances over the initial threshold in the SE-POT model for negated log-returns of DB, HSBC, RBS, and UBS stocks.

Table 5: p-values of the likelihood tests testing hypothesis that the bivariate dependence structure in the four-dimensional model is symmetric.

Pair	pValue	Pair	pValue
DB and HSBC	0.9219	HSBC and RBS	0.9343
DB and RBS	0.9999	HSBC and UBS	0.8160
DB and UBS	0.9974	RBS and HSBC	0.9999

It is clear from the table, the bivariate models display no evidence against symmetric dependencies. Extrapolating this observation also to the trivariate models and assuming the dependence parameter between different combinations of the banks' returns stays the same, we employ the following rate of exceedances for the four-dimensional model:

$$\tau(t,x_1,x_2,x_3,x_4) = \left(\tau_1(t,x_1)^{\theta(t)} + \tau_2(t,x_2)^{\theta(t)} + \tau_3(t,x_3)^{\theta(t)} + \tau_4(t,x_4)^{\theta(t)}\right)^{1/\theta(t)}. \qquad (41)$$

Similarly to the bivariate application, we parametrize dependence parameter $\theta(t)$ as the Hawkes process. In this case, however, there are much more different combination of marginal events on which $\theta(t)$ may depend. To be in line with Section 2.2.2, we suggest the following parametrization:

$$\theta(t) = \theta_{4,0} + \int_0^t e^{-\gamma_{4,0}(t-s)} \left[\psi_{4,1} dN_{4,1}(s) + \psi_{4,2} dN_{4,2}(s) + \psi_{4,3} dN_{4,3}(s) + \psi_{4,4} dN_{4,4}(s)\right], \qquad (42)$$

for $\theta_{4,0} \geq 1$, $\psi_{4,1}, \ldots, \psi_{4,4} \geq 0$, $\gamma_{4,0} > 0$, where $N_{4,i}(s)$ is a counting measure of events when at least two banks' negated returns exceed the initial threshold simultaneously and one of those banks is DB for $i = 1$, HSBC for $i = 2$, RBS for $i = 3$, or UBS for $i = 4$. As it is discussed in Section 2.2.2, one can design a number of alternatives to this specification, but still, in order to keep the model closed, we stick with (42).

After estimating the four-dimensional model, we conduct a likelihood test on the null hypothesis that the dependence parameter is not affected by the events described by counting measure $N_{4,i}(\cdot)$, i.e., $\psi_{4,i} = 0$. Hypotheses that $\psi_{4,1} = 0$ and $\psi_{4,4} = 0$ are not rejected with p-values close to 1. For parameters $\psi_{4,2}$ and $\psi_{4,3}$, p-values of the tests are, respectively, 0.0002 and 0.3269. Hence, we decided to keep only parameter $\psi_{4,2}$ in the model. This results of the likelihood tests effectively mean that multivariate events only with HSBC's exceedances contribute to the dependence structure (it should not be interpreted that only extreme events of that bank affect the dependence structure). Parameter estimates can be found in Tables 6.

The time varying dependence parameter $\theta(t)$ and the estimated conditional probabilities of multivariate exceedances are depicted in Figure 10. The results for the goodness-of-fit can be found in Figure 11. Again the hypothesis that residual inter-arrival intervals are independent standard

Table 6: Parameter estimates of the four-dimensional model of exceedances. An inverse Hessian of the likelihood function is used to obtain the standard errors reported in parentheses right to the estimates.

Parameter	Estimate
$\theta_{4,0}$	1.2304 (0.0298)
$\gamma_{4,0}$	0.0427 (0.0132)
$\psi_{4,2}$	0.1249 (0.0156)

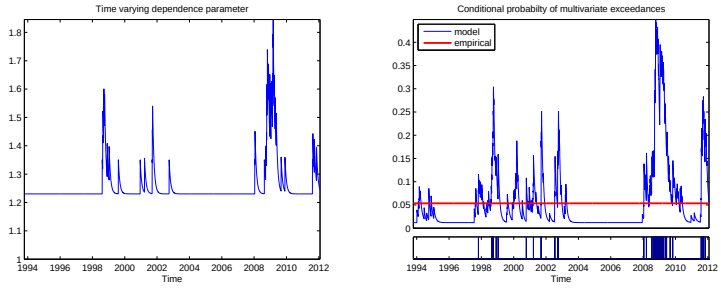

Figure 10: The estimated time-varying dependence parameter (left-hand panel) and the conditional probability of multivariate events when at least one margins exceed the initial threshold (right-hand panel) in the four-dimensional model. The tick marks at the bottom of the right panel denote times of multivariate events.

exponentially distributed is supported by Kolmogorov-Smirnov and Ljung-Box (15 lags) tests, which failed to reject the corresponding null hypothesis with p-values, respectively, 0.4798 and 0.1855.

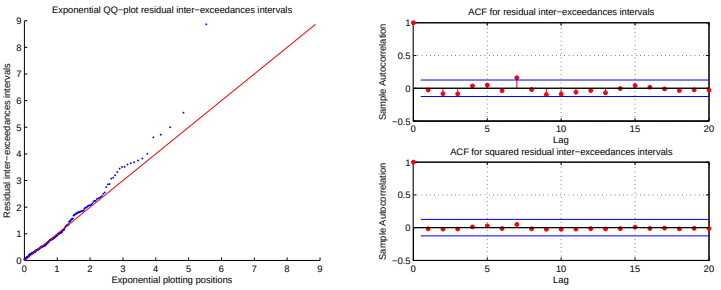

Figure 11: Exponential QQ-plot of the residual inter-exceedances intervals in the four-dimensional model (left-hand panel). The sample autocorrelation function of those (squared) intervals (right-hand panel).

To test accurateness of the assumption that the dependence parameter between different combinations of the banks' returns can be modelled as in (42), we consider goodness-of-fit for the two- and three-dimensional sub-models of rate (41). For example, one of its three-dimensional rates is of the form

$$\tau(t, x_1, x_2, x_3) = \left(\tau_1(t, x_1)^{\theta(t)} + \tau_2(t, x_2)^{\theta(t)} + \tau_3(t, x_3)^{\theta(t)}\right)^{1/\theta(t)}. \tag{43}$$

Treating those lower dimensional versions of (41) as the rates of the corresponding sum of marginal point process, we employ the same goodness-of-fit procedure as for the main model. Table 7 reports the p-values of the Kolmogorov-Smirnov and Ljung-Box (15 lags) tests for residual inter-exceedances intervals for different combinations of the considered banks. The corresponding exponential QQ-plots can be found in Appendix G. It is evident from the table that all sub-models describe the data well supporting by this our suggestion that the extreme dependence between the banks can be described by relationship (42). In case the number of considered banks (firms) is large, one can consider the extreme dynamics of the market index to construct the dependence parameter between those banks.

The fit of the model of joint exceedances in the two and four-dimensional cases suggest that the model provides an efficient way to quantify the effects that cause the clustering of extreme financial returns. Among others, these effects are the reaction of markets to common economic factors and interplay between markets through time-varying linkages. Note that while being able to quantify the effects, our model cannot explain the source of clustering and contagion, because our model

Table 7: p-values the Kolmogorov-Smirnov (KS) and Ljung-Box (LB) with 15 lags tests for residual inter-exceedances intervals for the two- and three-dimensional sub-models of the four-dimensional model.

Combination	KS	LB	Combination	KS	LB
DB and HSBC	0.3311	0.7799	RBS and HSBC	0.5676	0.9322
DB and RBS	0.1835	0.8049	DB and HSBC and RBS	0.3231	0.4384
DB and UBS	0.5641	0.8803	DB and HSBC and UBS	0.6107	0.5896
HSBC and RBS	0.3108	0.7943	DB and RBS and UBS	0.4333	0.5212
HSBC and UBS	0.2807	0.6523	HSBC and RBS and UBS	0.4024	0.7976

is decidedly reduced-form. To show when and where exactly the shock occurs, a pure qualitative analysis is required.

5. Conclusion

In the present paper, we develop a multivariate approach to model extreme asset returns considering the conditional distributional properties of both their magnitudes and occurrence times. The approach is developed in the framework of self-exciting point processes and multivariate extreme value theory, and follows naturally from treating the multivariate process of extreme events as a univariate process constructed as a superposition of individual extreme events. This is an intensity-based model which incorporates a feasible possibility to get updated in continuous time and lends itself to likelihood inference.

The paper provides its major statistical contribution in that it shows that, provided the marginal processes of extreme events follow the self-exciting peaks-over-threshold model, the functional form of the multivariate rate of extreme exceedances should follow the exponent measure of an extreme value copula. This exponent measure combines the marginal rates into the multivariate one. Due to its construction, the model can capture typical features of the financial time series, namely, heavy tails, extreme dependences, and clustering of extreme events in both time and across the assets (markets). A separate contribution of this paper is a derivation of the stationarity conditions for the self-exciting peaks-over-threshold model with predictable marks (the marginal process of exceedances).

To estimate the proposed model, we derive the closed form likelihood function and describe the goodness-of-fit and simulation procedures. Additionally, we consider the model estimation based on method of moments, which, however, turned out to be inferior to the MLE. The reason lies in the fact that the moment conditions cannot be directly calculated from the sample independently from the unknown parameters of the models.

We implemented the estimation procedure on extreme negative returns, studying MSCI-EU and MSCI-USA indexes, and equity of Deutsche Bank, RBS, HSBC, and UBS. The goodness-of-fit procedure demonstrates a reasonable fit of the model and suggests an empirical importance of the self-exciting feature for modeling both occurrence times, magnitudes, and interdependencies of the extreme returns. While the dependence structure of the model can account for asymmetry relationships, we find that conditional multivariate distributions of the returns are close to symmetric. Nevertheless, there are still asymmetric effects coming from the self-exciting structure of the conditional marginal distributions of the magnitudes of the exceedances. In the bivariate application with MSCI-EU and MSCI-USA indexes, we find that the extreme return exceedances of MSCI-EU are decisive (statistically) for modeling strength of the dependence between the two indexes. For the banking data, the multivariate exceedances with HSBC equity are the most important for modeling the dependence.

In the future research, it would be interesting to consider high-dimensional applications of the model for, e.g., risk management purposes, and to compare its performance to existing alternatives. This comparison would be particularly interesting if the application to high frequency data is considered. In order to overcome the problem of parameters' proliferation and to provide a flexible estimation approach, we plan also to apply the principle of pair copula construction to our multivariate point-process model.

Appendix A. Stationarity condition

Proof of Proposition 2.1

Proof Due the branching process's representation of the SE=POT model, the sufficient condition for stationarity of the SE-POT with conditional intensity $\tau(t)$ requires $\mathrm{E}\tau(t) = \bar{\tau} \in (0, \infty)$, see Daley and Vere-Jones (2005), Ex.6.3(c). From (1) $\bar{\tau}$ can be expressed as

$$\bar{\tau} = \tau + \psi \mathrm{E} \int_{-\infty}^{t} c\left(\tilde{X}_s\right) g\left(t-s\right) dN(s). \tag{A.1}$$

Note, that from the interpretation of the branching coefficient in Hawkes and Oakes (1974) and Daley and Vere-Jones (2005) (Example 6.3(c)) it follows that $\nu = \psi \mathrm{E} \int_{-\infty}^{t} c\left(\tilde{X}_s\right) g\left(t-s\right) dN(s)$.

Since the integral on the right-hand side of the above equation is just a sum of random variables, we can write

$$\mathrm{E} \int_{-\infty}^{t} c\left(\tilde{X}_s\right) g\left(t-s\right) dN(s) = \int_{-\infty}^{t} g\left(t-s\right) \mathrm{E}\left[c\left(\tilde{X}_s\right) dN(s)\right]. \tag{A.2}$$

From construction of the SE-POT model, see (1) and (3), it immediately follows that random variables $\tilde{X}_s$ and $dN(s)$ are dependent in general but conditional $v^*(s)$ (or even $\mathbb{H}_s$) they are independent. Hence it follows,

$$\mathrm{E}\left\{c\left(\tilde{X}_s\right) dN(s)\right\} = \mathrm{E}\left\{\mathrm{E}\left[c\left(\tilde{X}_s\right) dN(s) \middle| \mathbb{H}_s\right]\right\} = \mathrm{E}\left\{\mathrm{E}\left[c\left(\tilde{X}_s\right) \middle| \mathbb{H}_s\right] \mathrm{E}\left[dN(s) | \mathbb{H}_s\right]\right\}, \tag{A.3}$$

where $\mathrm{E}\left[dN(s) \mid \mathbb{H}_s\right] = \tau(s)ds$ and, considering the conditional distribution of $\tilde{X}_s$ in (3),

$$\mathrm{E}\left[c\left(\tilde{X}_s\right) \middle| \mathbb{H}_s\right] = \int_0^{\infty} c(x) f_s(x) dx,$$

where $f_s(x) = \frac{dF_s(x)}{dx} = \frac{1}{\beta + \alpha v^*(s)} \left(1 + \xi \frac{x}{\beta + \alpha v^*(s)}\right)^{-1/\xi - 1}$ is the conditional distribution density function of $\tilde{X}_s$.

Note that the integral in the above equation tends to infinity in all cases when the order of $c(x)$ exceeds $1/\xi$. In particular, the integral tends to infinity with $c(x) = e^{\delta x}$, which is a commonly used specification for $c(x)$ in the literature Chavez-Demoulin, Davison, and McNeil (2005) and McNeil, Frey, and Embrechts (2005). With the specification (6), however, we get

$$\mathrm{E}\left[c\left(\tilde{X}_s\right) \middle| \mathbb{H}_s\right] = \int_0^{\infty} c^*\left(F_s(x)\right) f_s(x) dx = \int_0^1 c^*\left(u\right) du.$$

In Section 2.1.2 it was suggested to construct $c^*(\cdot)$ as $c^*(\cdot) = 1 + G^{\leftarrow}(\cdot)$, where $G^{\leftarrow}(\cdot)$ is an inverse of the distribution function G of some continuous positive random variable with mean δ. Using this construction to calculate integral in the above equation we get

$$\mathrm{E}\left[c\left(\tilde{X}_s\right) \middle| \mathbb{H}_s\right] = 1 + \delta. \tag{A.4}$$

Substituting this result and $\mathrm{E}\left[dN(s) \mid \mathbb{H}_s\right] = \tau(s)ds$ into (A.3) we get

$$\mathrm{E}\left[c\left(\tilde{X}_s\right) dN(s)\right] = \bar{\tau}(1+\delta)ds,$$

which with (A.2) provides a formula for the expected value

$$\mathrm{E} \int_{-\infty}^{t} c\left(\tilde{X}_s\right) g\left(t-s\right) dN(s) = \bar{\tau}(1+\delta) \int_0^{\infty} g\left(s\right) ds.$$

Substituting the above equation into (A.1), finally yields

$$\bar{\tau} = \frac{\tau}{1 - \psi(1+\delta)\int_0^\infty g(s)\,ds}. \tag{A.5}$$

and

$$\nu = \psi(1+\delta)\int_0^\infty g(s)\,ds.$$

Thus, under the assumption of stationarity, we must have

$$\nu = \psi(1+\delta)\int_0^\infty g(s)\,ds < 1.$$

With, $g(s) = e^{-\gamma s}$, the above condition takes the form

$$\frac{\psi(1+\delta)}{\gamma} < 1. \qquad \blacksquare$$

Appendix B. Derivation of the covariance density

Proof of Proposition 2.2.

Proof The covariance density $\mu(z)$ of the SE-POT process of exceedances above the initial threshold is defined for $z > 0$ as

$$\mu(z) = \frac{\mathrm{E}\left[dN(t+z)dN(t)\right]}{(dt)^2} - \bar{\tau}^2, \qquad z > 0,$$

and for $z < 0$ the covariance density reads $\mu(z) = \mu(-z)$.

Note that for the case $z = 0$, the situation is slightly different, because $\mathrm{E}\left[(dN(t))^2\right] = \mathrm{E}\left[dN(t)\right] = \bar{\tau}ds$, i.e., the covariance density for $z = 0$ equals $\bar{\tau}$. The complete covariance density $\mu^{(c)}(z)$ (we use the same notation as in Hawkes (1971)) takes the form

$$\mu^{(c)}(z) = \bar{\tau} I_{z=0} + \mu(z), \tag{B.1}$$

where I_A denotes an indicator of event A.

To obtain an explicit formula for the covariance density $\mu(z)$, we follow the procedure described in Hawkes (1971). For $z > 0$ we write

$$\begin{aligned}\mu(z) = \mathrm{E}\left\{\mathrm{E}\left[\frac{dN(t)}{dt}\frac{dN(t+z)}{dt}\middle|\mathbb{H}_{t+z}\right]\right\} - \bar{\tau}^2 = \mathrm{E}\left\{\frac{dN(t)}{dt}\mathrm{E}\left[\frac{dN(t+z)}{dt}\middle|\mathbb{H}_{t+z}\right]\right\} - \bar{\tau}^2 = \\ \mathrm{E}\left\{\frac{dN(t)}{dt}\left[\tau + \psi\int_{-\infty}^{t+z} c\left(\tilde{X}_s\right) g\left(t+z-s\right) dN(s)\right]\right\} - \bar{\tau}^2 = \\ \bar{\tau}\tau - \bar{\tau}^2 + \psi\int_{-\infty}^{t+z} g\left(t+z-s\right)\mathrm{E}\left[c\left(\tilde{X}_s\right)\frac{dN(t)}{dt}\frac{dN(s)}{ds}\right] ds. \end{aligned} \tag{B.2}$$

Recalling (A.3) and (A.4) we can write

$$\begin{aligned}\mathrm{E}\left[c\left(\tilde{X}_s\right)\frac{dN(t)}{dt}\frac{dN(s)}{ds}\right] = \mathrm{E}\left\{\mathrm{E}\left[c\left(\tilde{X}_s\right)\middle|\mathbb{H}_s\right]\mathrm{E}\left[\frac{dN(t)}{dt}\frac{dN(s)}{ds}\middle|\mathbb{H}_s\right]\right\} = \\ (1+\delta)\mathrm{E}\left[\frac{dN(t)}{dt}\frac{dN(s)}{ds}\right] = (1+\delta)\left(\mu^{(c)}(s-t) + \bar{\tau}^2\right),\end{aligned}$$

which substituted in (B.2) yields

$$\mu(z) = \bar{\tau}\tau - \bar{\tau}^2 + \psi(1+\delta)\int_{-\infty}^{t+z} g\left(t+z-s\right)\left(\mu^{(c)}(s-t) + \bar{\tau}^2\right)ds =$$
$$\bar{\tau}\tau - \bar{\tau}^2 + \psi(1+\delta)\int_{-\infty}^{z} g\left(z-v\right)\left(\mu^{(c)}(v) + \bar{\tau}^2\right)dv =$$
$$\bar{\tau}\tau - \bar{\tau}^2\left(1 - \psi(1+\delta)\int_{0}^{\infty} g\left(z-v\right)dv\right) + \psi(1+\delta)\int_{-\infty}^{z} g\left(z-v\right)\mu^{(c)}(v)dv.$$

Together with (A.5) and (B.1), the above equation transforms

$$\mu(z) = \psi(1+\delta)\left(g(z)\bar{\tau} + \int_{-\infty}^{z} g\left(z-v\right)\mu(v)dv\right), \tag{B.3}$$

or, exploiting the symmetry of $\mu(z)$,

$$\mu(z) = \psi(1+\delta)\left(g(z)\bar{\tau} + \int_{0}^{\infty} g\left(z+v\right)\mu(v)dv + \int_{0}^{z} g\left(z-v\right)\mu(v)dv\right). \tag{B.4}$$

As it was noted in Hawkes (1971), the above equation is difficult to solve analytically in general. But for the case when $g(\cdot)$ decays exponentially the analytical solution may be obtained. Setting $g(v) = e^{-\gamma v}$ and taking the Laplace transform (denote it as μ^*) of (B.4), we get

$$\mu^*(y) = \frac{\bar{\tau}\psi(1+\delta)\left(2\gamma - \psi(1+\delta)\right)}{2(\gamma - \psi(1+\delta))}\frac{1}{y+\gamma-\psi(1+\delta)}.$$

Recalling that the Laplace transform $f^*(y)$ of $f(z) = e^{az}$ equals $f^*(y) = \frac{1}{y-a}$, it is easy to see from the above equation that

$$\mu(z) = \frac{\bar{\tau}\psi(1+\delta)\left(2\gamma - \psi(1+\delta)\right)}{2(\gamma - \psi(1+\delta))}e^{-(\gamma-\psi(1+\delta))z}. \qquad \blacksquare$$

Appendix C. Method of Moments estimation

The goodness-of-fit characteristics χ_j and $m_{i,j}$, which are standard exponentially distributed under a correctly specified model, can be used for construction of moment conditions for the Method of Moments (MM) estimation of both the SE-POT and multivariate models. To estimate seven parameters of the SE-POT model, one can set the following seven moment conditions

$$\begin{cases} \frac{1}{N_{u_i}}\sum_{j=1}^{N_{u_i}} \chi_{j,i} = 1 \\ \frac{1}{N_{u_i}}\sum_{j=1}^{N_{u_i}} \chi_{j,i}^2 = 2\left(\frac{1}{N_{u_i}}\sum_{j=1}^{N_{u_i}} \chi_{j,i}\right)^2 \\ \frac{1}{N_{u_i}}\sum_{j=1}^{N_{u_i}} \chi_{j,i}^3 = 6\left(\frac{1}{N_{u_i}}\sum_{j=1}^{N_{u_i}} \chi_{j,i}\right)^3 \\ \frac{1}{N_{u_i}}\sum_{j=1}^{N_{u_i}} Z_{j,i} = \frac{1}{1-\xi_i} \\ \frac{1}{N_{u_i}}\sum_{j=1}^{N_{u_i}} Z_{j,i}^2 = \frac{2}{(1-\xi_i)(1-2\xi_i)} \\ \frac{1}{N_{u_i}}\sum_{j=1}^{N_{u_i}} m_{j,i} = 1 \\ \frac{1}{N_{u_i}}\sum_{j=1}^{N_{u_i}} \left(T_{j,i} - T_{j-1,i}\right) = \frac{1}{\bar{\tau}_i} =: \frac{\gamma_i - \psi_i(1+\delta_i)}{\tau_i\gamma_i} \end{cases} \tag{C.1}$$

where $Z_{j,i} := \frac{X_{i,T_{i,j}} - u_i}{\beta_i + \alpha_i v_i^*(T_{i,j})}$ follows a GPD distribution with shape parameter ξ_i and scale parameter 1. Solving (numerically) the above system one can easily obtain estimates for the seven parameters of the SE-POT model. Extending this approach one can add some extra moment conditions and employ Generalized Method of Moments (GMM) with the common procedure of estimating the weighting matrix as the inverse of the covariance matrix, see, e.g., Greene (2003). However, Monte Carlo simulations (not reported in this paper) suggest that the GMM provides inferior estimates to

the MM ones. The reason for the poor performance is attributed to numerical instability and high sensitivity to the starting values of the GMM estimators of the SE-POT model.

One can proceed similarly for construction of the moments condition for the MM estimation the multivariate model. In particular, those may include moment conditions on (standard exponential distributed) χ_j and its variants for lower dimensional models. For the MM estimation of the bivariate model with, say, four parameters driving $\theta(t)$, the following conditions were used

$$\begin{cases} \frac{1}{N}\sum_{j=1}^{N}\chi_j = 1 \\ \frac{1}{N}\sum_{j=1}^{N}\chi_j^2 = 2\left(\frac{1}{N}\sum_{j=1}^{N}\chi_j\right)^2 \\ \frac{1}{N}\sum_{j=1}^{N}\chi_j^3 = 6\left(\frac{1}{N}\sum_{j=1}^{N}\chi_j\right)^3 \\ \frac{1}{N}\sum_{j=1}^{N}\exp\left(-\chi_j\right) = \left(1+\left(\frac{1}{N}\sum_{j=1}^{N}\chi_j\right)\right)^{-1} \end{cases} \tag{C.2}$$

where χ_j is from (37) and N is the number of events in the sample when at least one margin jumps. Although the method of moments is intuitive and easily programmed, in most cases, however, the method of moments estimators are not efficient, Greene (2003). Furthermore, unreported simulation results suggest that the maximum likelihood is superior in finite samples for our model.

Appendix D. Extreme value condition and the initial threshold

The rate of multivariate extreme exceedances (10) is governed by the extreme value copula, which corresponds, as it is suggested by Proposition 2.4, to the dependence structure of extreme magnitudes of observations. Considering the extreme dependence one distinguishes two possible cases: asymptotic dependence and asymptotic independence, – which require two different estimation procedure of the dependence structure. Hence, it is important for a correct parametrization of the multivariate model to find the appropriate type of the asymptotic dependence. The condition of heavy-tailedness of the observations must also be tested, because it is a prerequisite for the use of the SE-POT model.

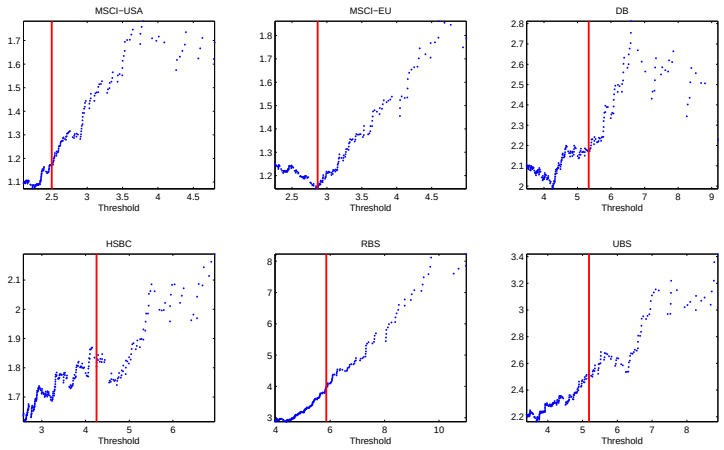

Figure D.12: Sample mean excess plots of negated daily log-returns of the MSCI-USA, MSCI-EU, DB, HSBC, RBS, and UBS. Solid red vertical lines indicate the initial threshold chosen for the model estimation.

Addressing the later condition, we rely on the mean-excess function to verify if the data is heavy-tailed and if GPD is an appropriate distribution. Details on this and other methods may be found, e.g., in McNeil, Frey, and Embrechts (2005), Embrechts, Klüppelberg, and Mikosch (1997), Resnick and Starica (1995), Chavez-Demoulin and Embrechts (2011). For positive-valued data $X_1, X_2, \ldots, X_n$ and its high threshold v, the mean-excess function is defined as

$$e_n(v) = \frac{\sum_{i=1}^{n}(X_i - v)I_{\{X_i>v\}}}{\sum_{i=1}^{n} I_{\{X_i>v\}}}. \tag{D.1}$$

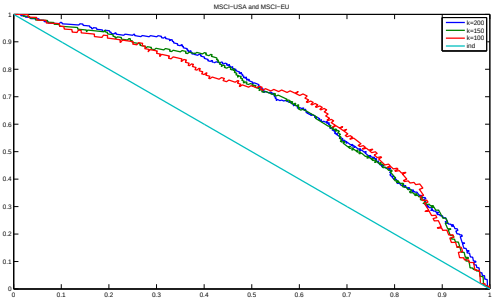

Figure D.13: Estimated Q-curves on negated returns of MSCI-USA and MSCI-EU: k denotes the number of upper order statistics used for estimation.

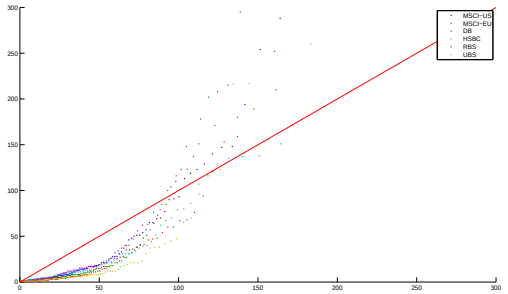

Figure D.14: Exponential QQ-plots of time intervals, measured in days, between consecutive marginal exceeances above the initial threshold.

Plotting $\{X_{i,n}, e_n(X_{i,n})\}$, where $X_{i,n}$ denotes the ith order statistic, we consider a shape of the mean-excess function for i close to n. If the shape looks approximately linear then this suggests that GPD is an appropriate distribution for the excesses over that threshold. The point where the mean-excess function visually becomes close to linear can be set as a threshold for GPD estimation.

Figures D.12 plots the estimates of mean-excess function for the last 6% of the sample upper order statistics. Solid vertical lines on the figure denote the marginal initial thresholds chosen for estimation of the SE-POT. For MSCI-USA and MSCI-EU series the initial threshold was set on the 97.7% quantile of the empirical distributions, which corresponds, respectively, to 2.4922% and 2.8601% and results in 132 marginal exceedances for the two indexes and in 53 joint exceedances. For the bank data the initial thresholds are the following: 5.3201% for DB, 4.2021% for HSBC, 5.8534% for RBS, and 5.1799% for UBS. Those threshold correspond to the 97.9% quantile of the empirical distributions, with result in 100 marginal (for all four indexes) and 16 joint exceedances. Figure D.14 illustrates the exponential QQ-plots for the time intervals between consecutive marginal exceedances above the initial thresholds. If the exceedances occurred independently then one would observe the exponential distribution of the intervals. Clearly, the exceedances do not occur independently: there are higher than expected frequencies of the small time intervals, i.e., there is a clustering of high losses. This observation, along with the heavy-tails of the returns, justifies the use of the SE-POT model.

Considering the asymptotic dependence, we employ a graphical illustration called a Q-curve, see de Haan and Ferreira (2006) for details. Figure D.13 illustrates the Q-curve of the negative log returns of MSCI-USA and MSCI-EU, estimated on different number (k) of upper order statistics of the return series. The line labeled "*ind*" indicates the Q-curve in the independence case. Note that flat Q-curves indicate asymptotic independence. The curves on Figure D.13 differ significantly from a straight line indicating that there is no asymptotic independence between negative log returns of MSCI-USA and MSCI-EU indexes.

To visualize the Q-curve in the four-dimensional case, we report its three-dimensional projections. Analogously to the bivariate case, a flat convex shape of the Q-curve in the three dimensional case indicates the presence of asymptotic independence. If the shape is concave, one expects no asymptotic independence. Figure D.15 illustrates the three dimensional Q-curves of DB, HSBC, RBS, UBS return series, estimated on 200 upper order statistics of the return series. The curves on Figure D.15 differ significantly from a flat curve indicating that there is no asymptotic independence between negative log returns of DB, HSBC, RBS, UBS equity prices.

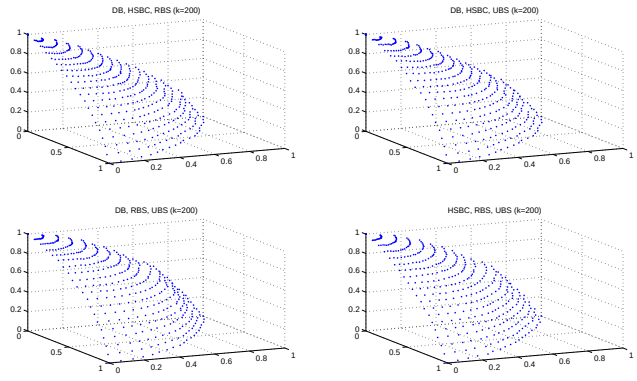

Figure D.15: Estimated Q-curves on negated log-returns of DB, HSBC, RBS, and UBS.

Appendix E. Marginal goodness-of-fit tests

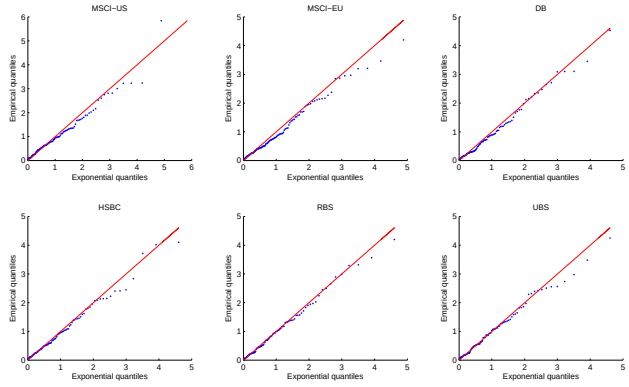

Figure E.16: Exponential QQ-plot of the residual marginal inter-exceedances intervals.

Table E.8: p-values of Kolmogorov-Smirnov (KS) and Ljung-Box (LB) tests checking the hypothesis of exponentially distributed and uncorrelated residual inter-exceedance intervals and marks of the marginal processes of exceedances.

	residual intervals		residual marks	
Data	KS	LB	KS	LB
MSCI-US	0.5498	0.9185	0.9558	0.9869
MSCI-EU	0.2566	0.4211	0.2358	0.9862
DB	0.3788	0.5971	0.4439	0.4809
HSBC	0.8852	0.2812	0.6144	0.6251
RBS	0.9349	0.7653	0.7534	0.3143
UBS	0.9237	0.6603	0.9563	0.0314

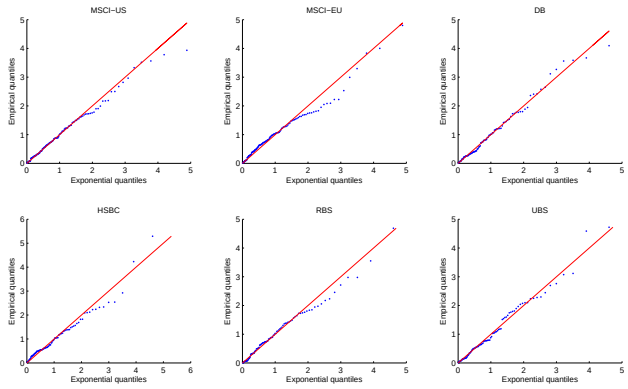

Figure E.17: Exponential QQ-plot of the residual marks.

Appendix F. Illustrative graphs and goodness-of-fit for the bivariate model with the MM estimates.

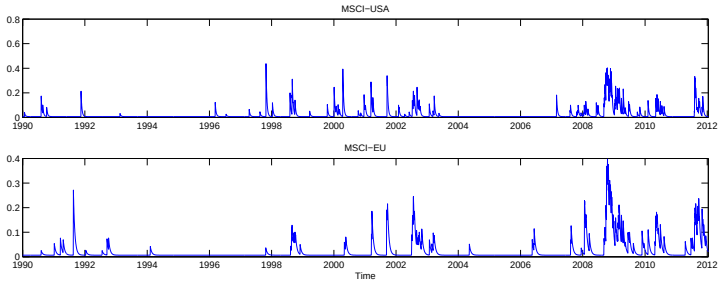

Figure F.18: Estimated conditional rate of the marginal exceedances over the initial threshold for MSCI-USA and MSCI-EU. MM estimates from Table 2.

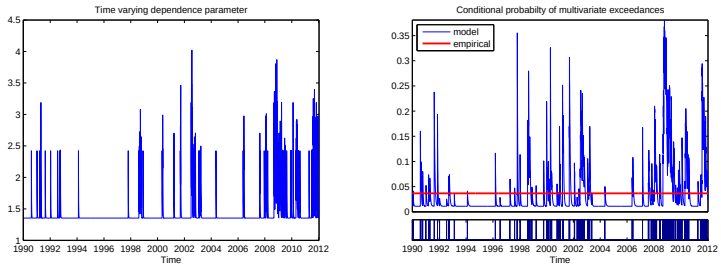

Figure F.19: The estimated time-varying dependence parameter (left-hand panel) and the conditional probability of multivariate events when at least one margins exceed the initial threshold (right-hand panel) in the two dimensional model. The tick marks at the bottom of the right panel denote times of multivariate events. MM estimates.

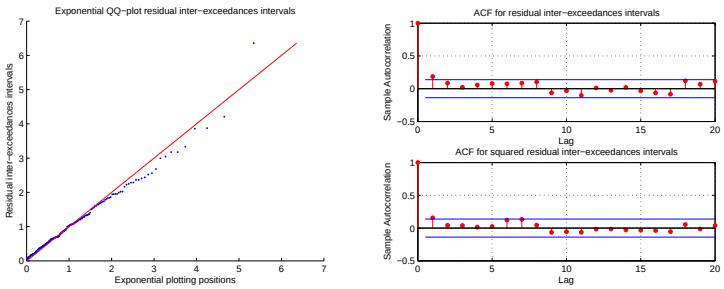

Figure F.20: Exponential QQ-plot of the residual inter-exceedance intervals (left-hand panel) in the bivariate model. The sample autocorrelation function of those (squared) intervals (right-hand panel). MM estimates.

Appendix G. Goodness-of-fit test for the bi- and trivariate sub-models of the four-dimensional model

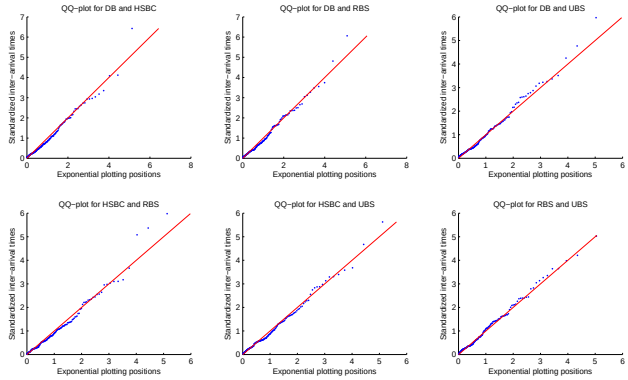

Figure G.21: Exponential QQ-plot for the residual inter-exceedance intervals of the bivariate sub-models of the four-dimensional model.

References

Aas, K., C. Czado, A. Frigessi, and H. Bakken (2009): "Pair-copula constructions of multiple dependence," *Insurance: Mathematics and Economics*, 44(2), 182–198.

Acharya, V., L. Pedersen, T. Philippon, and M. Richardson (2010): "Measuring Systematic Risk," *Working paper*.

Aït-Sahalia, Y., J. Cacho-Diaz, and R. J. Laeven (2011): "Modeling Financial Contagion Using Mutually Exciting Jump Processes," *NBER Working Paper No. w15850*.

Andriani, P., and B. McKelvey (2007): "A negative binomial model for time series of counts," *Journal of International business Studies*, 38, 1212–1230.

Bae, K.-H., G. A. Karolyi, and R. M. Stulz (2003): "A New Approach to Measuring Financial Contagion," *The Review of Financial Studies*, 16(3), 717–763.

Bekaert, G., M. Ehrmann, M. Fratzscher, and A. Mehl (2012): "Global crisis and equity market contagion," *Working Paper*.

Bollerslev, T., V. Todorov, and S. Z. Li (2013): "Jump tails, extreme dependencies, and the distribution of stock returns," *Journal of Econometrics*, 172, 307–324.

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

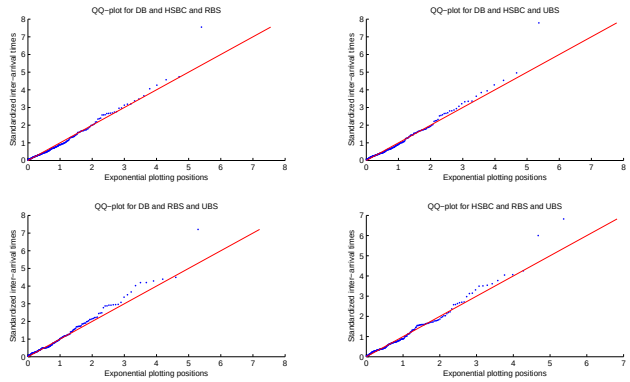

Figure G.22: Exponential QQ-plot for the residual inter-exceedance intervals of the trivariate sub-models of the four-dimensional model.

Chavez-Demoulin, V., A. C. Davison, and A. J. McNeil (2005): "Estimating Value-at-Risk: A point process approach," *Quantitative Finance*, 5(2), 227–234.

Chavez-Demoulin, V., and P. Embrechts (2011): "An EVT primer for credit risk," *The Oxford Handbook of Credit Derivatives*, 73(1), 500–532.

Chavez-Demoulin, V., P. Embrechts, and J. G. Nešlehová (2006): "Quantitative models for operational risk: Extremes, dependence and aggregation," *Journal of Banking & Finance*, 30, 2635–2658.

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

Coles, S. G., and J. A. Tawn (1991): "Modelling extreme multivariate events," *Journal of Royal Statistical Society B*, 53(2), 377–392.

Daley, D. J., and D. Vere-Jones (1988): *An Introduction to the Theory of Point Processes*. New York: Springer.

——— (2005): *An Introduction to the Theory of Point Processes. Volume 1. Elementary Theory and Methods*. Springer.

Davison, A. C., and R. L. Smith (1990): "Models for Exceedances over High Thresholds," *Journal of the Royal Statistical Society B*, 52, 393–442.

de Haan, L., and A. Ferreira (2006): *Extreme Value Theory: An Introduction*. New York: Springer.

Demarta, S., and A. J. McNeil (2005): "The t copula and related copulas," *International Statistical Review*, 73(1), 111–129.

Diebold, F. X., T. A. Gunther, and A. S. Tay (1998): "Evaluating Density Forecasts with Applications to Financial Risk Management," *International Economic Review*, 39(4), 863–883.

Drees, H., and L. de Haan (2012): "Estimating Failure Probabilities," *Working Paper*.

Einmahl, J. H. J., L. de Haan, and A. Krajina (2013): "Estimating extreme bivariate quantile regions," *Extremes*, 16, 121–145.

Embrechts, P., C. Klüppelberg, and T. Mikosch (1997): *Modelling Extremal Events for Insurance and Finance*. Springer.

Embrechts, P., T. Liniger, and L. Lin (2011): "Multivariate Hawkes Processes: an Application to Financial Data," *Journal of Applied Probability*, 48(A), 367–378.

Errais, E., K. Giesecke, and L. R. Goldberg (2010): "Affine Point Processes and Portfolio Credit Risk," *SIAM Journal of Financial Mathematics*, 1, 642–665.

Golosnoy, V., B. Gribisch, and R. Liesenfeld (2012): "Intra-Daily Volatility Spillovers between the US and German Stock Markets," *Working Paper*.

Greene, W. H. (2003): *Econometric Analysis*. Prentice Hall.

Gudendorf, G., and J. Segers (2010): "Extreme-Value Copulas," in *Copula Theory and Its Applications*, ed. by P. Jaworski, F. Durante, W. K. Härdle, and T. Rychlik, pp. 127–145.

Hawkes, A. G. (1971): "Point Spectra of Some Mutually Exciting Point Processes," *Journal of the Royal Statistical Society B*, 33(3), 438–443.

Hawkes, A. G., and D. Oakes (1974): "A cluster process representation of a self-exciting process," *Journal of Applied Probability*, 11, 493–503.

Helmstetter, A., and D. Sornette (2002): "Sub-critical and Super-critical Regimes in Epidemic Models of Earthquake Aftershocks," *Journal of Geophysical Research*, 107(10), 1–21.

Leadbetter, M. R. (1983): "Extremes and local dependence in stationary sequences," *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 65, 291–306.

——— (1988): "Extremal theory for stochastic processes," *The Annals of Probability*, 16(2), 431–478.

——— (1991): "On a basis for "Peaks over Threshold" modeling," *Statistics and Probability Letters*, 12, 357–362.

McNeil, A. J., R. Frey, and P. Embrechts (2005): *Quantitative Risk Management: Concepts, Techniques, Tools*. Princeton University Press.

Møller, J., and J. G. Rasmussen (2005): "Perfect simulation of Hawkes processes," *Advances in Applied Probability*, 37, 629–646.

Nelsen, R. B. (2006): *An Introduction to Copulas*. Springer.

Oakes, D. (1975): "A Markovian self-exciting process," *Journal of Applied Probability*, 12, 69–77.

Ogata, Y. (1978): "The asymptotic behaviour of maximum likelihood estimators for stationary point processes," *Annals of the Institute of Statistical Mathematics*, 30(A), 223–261.

——— (1981): "On Lewis' Simulation Method for Point Processes," *IEEE Transactions on Information Theory*, 27(1), 23–31.

——— (1988): "Statistical Models for Earthquake Occurrences and Residual Analysis for Point Processes," *Journal of the American Statistical Association*, 83(401), 9–27.

Okhrin, O., Y. Okhrin, and W. Schmid (2013): "On the structure and estimation of hierarchical Archimedean copulas," *Journal of Econometrics*, 173, 189–204.

Patton, A. J. (2012): "A review of copula models for econometric time series," *Journal of Multivariate Analysis*, 110, 4–18.

Resnick, S. I. (2007): *Extreme Values, Regular Variation, and Point Processes*. Springer.

Resnick, S. I., and C. Starica (1995): "Consistency of Hill's estimator for dependent data," *Journal of Applied Probability*, 32.

Rootzen, H., and N. Tajvidi (2006): "Multivariate generalized Pareto distributions," *Bernoulli*, 12, 917–930.

Smith, R. L., J. A. Tawn, and S. G. Coles (1997): "Markov chain models for threshold exceedances," *Biometrika*, 84(2), 249–268.

Tawn, J. A. (1990): "Modelling multivariate extreme value distributions," *Biometrika*, 77,2, 245–253.

Vere-Jones, D., and R. B. Davies (1966): "A statistical survey of earthquakes in the main seismic region of New Zealand," *New Zealand Journal of Geology and Geophysics*, 9(3), 251–284.

Wongswan, J. (2006): "Transmission of information across international equity markets," *Review of Financial Studies*, 19, 1157–1189.