Statistics and Its Interface Volume 8 (2015) 109–122

# Marked point process adjusted tail dependence analysis for high-frequency financial data*

Alexander Malinowski†, Martin Schlather, and Zhengjun Zhang

Although the extremes of high-frequency financial transaction data have a huge economic impact, basic characteristics of the data have not been addressed up to now. To capture dependence between the tail behavior of inter-transaction returns and the pattern of transaction times, this paper combines marked point process (MPP) theory with extreme value analysis. Suitable measures of interaction are provided, based on second-order moments of MPPs. Applying these measures to financial transaction data, it is verified that the extreme value index of the return distribution is indeed locally increased, i.e., on the scale of minutes, by the existence of surrounding transactions. A simulation study underpins the observed effects and enables assessing the finite sample properties of the respective estimators. Further, asymptotic results on the estimators are given.

AMS 2000 subject classifications: Primary 60G55; secondary 60G70.
Keywords and phrases: Tail index, Mark-location dependence, Moment measure, Transaction data.

## 1. INTRODUCTION

The irregular spacing of financial data recorded at intraday frequency level has been inspiring an extensive usage of (marked) point process methods in econometric and financial applications. The seminal paper of Engle [15] and the contributions of Engle and Lunde [16], Bowsher [4] and Bauwens and Hautsch [3] are well-known examples. Yet, at the same time, movements of asset prices are commonly modeled via continuous-time stochastic processes—an approach that suggests to perceive transaction data as non-evenly spaced measurements of an underlying continuous-time process [e.g., 1, 22]. As long as the pattern of point locations in such a model is stochastically independent of the underlying process, global parameter estimation is well-established in literature [e.g., 20, 42] including various contributions on declustering and debiasing of non-evenly spaced measurements [e.g., 26]. Though, in the context of financial transaction data, the observed prices clearly depend on the frequency of trading events; executing a transaction impacts on both the instantaneous and on future prices. Ignoring these dependencies and applying standard tools can lead to severe biases. Here, marked point processes (MPPs) provide a commonly used framework that can capture arbitrary forms of dependency between point locations and observed values, called *marks* in this context [e.g., 3, 9, 11, 28, 35, 36, 40]. While the effect of single transactions is relevant on a rather small scale, on larger scales, continuous-time representations of asset prices might be advantageous; a general challenge is to model these processes across different scales [e.g., 12].

Another main issue of financial modeling and quantitative asset pricing is the assessment of financial risk. The rapid growth and globalization of financial markets together with the financial crises during the last decades have led to a strong demand for risk management systems. While some of the early models for asset returns, e.g., the capital asset pricing model (CAPM), are solely based on variances, risk measures that reflect the shareholders' preferences more adequately include the value at risk and the expected shortfall. Extreme value theory (EVT) goes one step further and considers the full tail of the distribution. It builds the theoretical framework for analyzing and modeling the univariate and the joint extremal behavior of multiple assets, which is of particular interest with regard to crashes and large portfolio losses [e.g., 13, 14].

This paper tries to bring together the two concepts, MPPs and EVT, in order to quantify interactions between the tail behavior and the pattern of transaction times. The core idea is to apply second-order moment measures of point processes, similar to those presented in Schlather, Ribeiro, Jr and Diggle [40]. Therein, first and second moments (i.e., mean and variance) of a point's mark distribution are conditioned on the existence of further points at a certain distance. In terms of financial high-frequency data, for instance, these quantities might be used to detect an increase of the log-returns' variance when the time distance to the previous transaction shrinks.

To enable statements of this kind for the thickness of tails of the return distribution, the second-order characteristics of [40] have to be generalized twofold:

*We are grateful to the two anonymous referees for providing valuable comments on the manuscript.
†Corresponding author.

In classic MPP theory, stationarity and ergodicity are two basic assumptions meaning that the stochastic behavior of the underlying process is constant in space or time and constant over different realizations. Many financial and economic processes, though, exhibit structural breaks due to abrupt changes in the underlying economic mechanisms and conditions, even after having corrected for seasonalities and trends [e.g., 2]. Structural breaks are commonly captured by means of regime-switching models [e.g., 7, 19], nowadays applied to diverse financial and economic processes [e.g., 21, 30, 37]. For intra-day data, also (hierarchical) non-linear models in the spirit of the autoregressive conditional duration model (ACD, [17]) are used, [e.g., 3, 5, 34, 43]. In an MPP framework, structural changes can easily be covered by dropping the ergodicity assumption. It seems both reasonable and natural to perceive financial transaction data as a concatenation of structurally different realizations of finite clock time length from a possibly non-ergodic MPP. This paper will revert to Malinowski, Schlather and Zhang [33], where (conditional) mark means and appropriate estimators for non-ergodic MPPs are provided.

The second direction of generalization of the characteristics in [40] is w.r.t. extreme values. Instead of mean and variance of the return distribution, here, the tail index will be of interest and a suitable representation in terms of second-order moment measures has to be found.

A further challenge is the assessment of the asymptotic distribution of MPP point estimates, which is in general not analytically tractable. Variance estimates and confidence intervals either have to be based on rather strong mixing or independence assumptions or non-parametric techniques such as subsampling and bootstrapping can be used. When estimating the tail dependence index for stationary time series, which is closely related to estimating the ordinary tail index, Laurini and Tawn [31] and Ledford and Tawn [32] state that confidence intervals based on iid assumptions will be too small when the extremes are dependent. They propose a block bootstrapping method to obtain proper variance estimates for their estimators. In our MPP set-up and in view of the massive amount of data when intra-daily financial data are used, the subsampling approach of Politis and Sherman [39] can be expected to yield reliable results. Subsampling-based variance estimates will be compared to those obtained from assuming independence between point locations and marks of the MPP.

The rest of the paper is organized as follows: In Section 2, some basic concepts and definitions from EVT and MPP theory are reviewed, including the generic form of the moment measure based summary statistics. Then the definition is tailored to the extreme value context, while the focus is on the tail index of the mark distribution and its interaction with point locations. An alternative to subsampling is proposed in order to assess the variability of the corresponding estimators. A central limit theorem (CLT) result for MPPs yields their asymptotic distribution. In Sections 3 and 4, the methods are applied to simulated data and to real high-frequency transaction data from the German stock index DAX, respectively. Section 5 closes with a summary and discussion of the results. Technical details and proofs are given in the appendix.

## 2. METHODS

### 2.1 Marked point processes

Throughout the paper, $\Phi = \{(t_i, y_i) : i \in \mathbb{N}\}$ is a stationary (not necessarily ergodic) and simple marked point process on $\mathbb{R}$ with real-valued marks $m(t_i) = y_i$, and $\Phi_{\mathrm{g}} = \{t : (t, y) \in \Phi\}$ denotes its ground process of point locations. Here, the $t_i$ can simply be regarded as time points of transactions. This section briefly reviews some of the definitions of second-order moment measures for MPPs and their estimators. For more details, the reader is referred to Malinowski, Schlather and Zhang [33] and, for the general theory of point processes, to Stoyan, Kendall and Mecke [41] and Daley and Vere-Jones [8, 9].

For $I \in \mathcal{B}(\mathbb{R})$ and $t \geq 0$, let

$$C(t, I) = \{(t_1, t_2) \in \mathbb{R}^2 : t_1 \in [0, t],\, t_2 \in I + t_1\},$$
$$C(I) = C(1, I).$$

For functions $f, h : \mathbb{R} \to \mathbb{R}$ and $C \in \mathcal{B}(\mathbb{R} \times \mathbb{R})$, let

$$\alpha^{(2)}_{f,h}(C) = \mathbb{E} \sum^{\neq}_{(t_1,y_1),(t_2,y_2)\in\Phi} f(y_1)h(y_1)\mathbf{1}_C((t_1, t_2)), \qquad (1)$$

where "$\neq$" indicates that the sum runs over all pairs of points with $(t_1, y_1) \neq (t_2, y_2)$. Throughout the paper it is assumed that $h$ and $f \cdot h$ are non-negative. If the derivative of $\alpha^{(2)}_{f,h}(C(\cdot))$ w.r.t. the Lebesgue measure exists, it is denoted by $\rho^{(2)}_{f,h}$. Extending the classical second-order factorial moment measure [41],

$$\mu^{(2)}_{f,h}(I) = \frac{\alpha^{(2)}_{f,h}(C(I))}{\alpha^{(2)}_{1,h}(C(I))}, \quad I \in \mathcal{B}(\mathbb{R}),$$

and

$$\mu^{(2)}_{f,h}(r) = \frac{\rho^{(2)}_{f,h}(r)}{\rho^{(2)}_{1,h}(r)}, \quad r \in \mathbb{R}, \qquad (2)$$

provide the notion of a conditional mark mean, subject to the conditioning that $\Phi$ has a further point at a distance contained in $I$ or at distance $r$, respectively. The choices $f(x) = x$ and $f(x) = x^2$, for instance, refer to the conditional mean and the non-centered second moment of the marks, respectively. The function $h$ allows to consider only a certain type of marks. If $h(\cdot) = \mathbf{1}_A(\cdot)$ is an indicator function with $A$ a Borel subset of $\mathbb{R}$, $\mu^{(2)}_{f,h}$ represents the conditional mean of $f(y)$, additionally conditioned on the event $\{y \in A\}$. If $h \equiv 1$, we simply write $\mu^{(2)}_f$, $\alpha^{(2)}_f$ and $\rho^{(2)}_f$.

If $\Phi$ is non-ergodic, its probability law $P$ can be decomposed according to a mixing measure $\lambda$ on the space $\mathcal{P}_{\text{erg}}$ of all ergodic MPP probability laws: $P(M) = \int_{\mathcal{P}_{\text{erg}}} q(M)\lambda(dq)$, $M \in \mathcal{M}_0$, where $\mathcal{M}_0$ is the canonical $\sigma$-algebra of counting measures. (For more details, the reader is referred to the appendix.) Let $Q \sim \lambda$ be the corresponding mixing random variable, i.e., $[\Phi \,|\, Q = q] \sim q$. Then

$$\mu^{(2)}_{f,h}(r) = \frac{\mathbb{E}\big[\mu^{(2)}_{f,h,\Phi|Q}(r) \cdot \rho^{(2)}_{1,h,\Phi|Q}(r)\big]}{\rho^{(2)}_{1,h}(r)},$$

i.e., $\mu^{(2)}_{f,h}(r)$ is a weighted average of its ergodic subclasses counterparts, with weights being proportional to the intensity of pairs of points with distance $r$. An alternative definition of conditional mark mean, which is proposed in Malinowski, Schlather and Zhang [33] and which avoids this implicit weighting, is given by

$$(3) \qquad \tilde{\mu}^{(2)}_{f,h} = \mathbb{E}\mu^{(2)}_{f,h,\Phi|Q} = \int_{\mathcal{P}_{\text{erg}}} \mu^{(2)}_{f,h,\Phi|Q=q}\, \lambda(dq).$$

While $\mu^{(2)}_{f,h}$ rather refers to a typical point out of the union of all ergodicity classes, $\tilde{\mu}^{(2)}_{f,h}$ reflects the marks' expectation within a typical ergodicity class, no matter how densely the points occur in each of the classes.

## 2.2 (Conditional) tail index for MPPs

The fundamental Fisher-Tippett-Gnedenko theorem states that there exist only three possible distributions for suitably standardized maxima of iid random variables. Out of these, the Fréchet distribution $\exp(-x^{-\alpha})$, $x > 0$, $\alpha > 0$, is the only heavy-tailed distribution and is therefore commonly applied in the context of financial data.

A standard problem is the estimation of the tail index $\xi = \alpha^{-1}$. For a random variable $Y$ (w.l.o.g., $Y \geq 0$) in the max-domain of attraction (MDA) of a Fréchet($\alpha$) distribution, it is well-known [e.g. 14, Sec. 6.4.2] that for the threshold $u$ tending to infinity,

$$(4) \qquad \xi(u) = \mathbb{E}(\log Y - u |\, \log Y > u) \to \xi = \alpha^{-1},$$

which is also the basis for the well-known Hill estimator [24] of the tail index.

An MPP analog of $\xi(u)$ as in (4), is given by

$$\frac{\mathbb{E}\sum_{(t,y)\in\Phi,\ t\in[0,1]} (\log y - u) \cdot \mathbf{1}_{\log y > u}}{\mathbb{E}\sum_{(t,y)\in\Phi,\ t\in[0,1]} \mathbf{1}_{\log y > u}}.$$

Taking limits for $u \to \infty$ gives a definition of the marks' tail index in an MPP setting, based on its mean excess representation.

The *conditional tail index* can now be defined by including an additional conditioning on the existence of a further point:

**Definition 2.1.** Let $\xi(I, u) = \mu^{(2)}_{f_u, h_u}(I)$ with $f_u(y) = \log y - u$ and $h_u(y) = \mathbf{1}_{\log y > u}$ for $y > 0$ and with $f_u(y) = h_u(y) = 0$ for $y \leq 0$. As in (2), we may define $\xi(r, u) = \mu^{(2)}_{f_u, h_u}(r)$ and consider $\xi$ as a function on $(\mathcal{B}(\mathbb{R}) \cup \mathbb{R}) \times \mathbb{R}^+$. Let $\tilde{\xi}(I, u)$ and $\tilde{\xi}(r, u)$ denote the analog of (3), i.e., the two-stage expectation, averaging within each ergodicity class first, and then pooling the different classes. Then $\xi(\cdot) = \lim_{u\to\infty} \xi(\cdot, u)$ and $\tilde{\xi}(\cdot) = \lim_{u\to\infty} \tilde{\xi}(\cdot, u)$ are called *conditional tail indices* of a mark of $\Phi$, conditional on the existence of a further point at a certain distance.

## 2.3 Estimation

Assuming the process is observed on the interval $[0, T]$, $T > 0$, the quantities $\mu^{(2)}_{f,h}(I)$ and $\tilde{\mu}^{(2)}_{f,h}(I)$ can naturally be estimated through

$$(5) \qquad \begin{aligned} \hat{\mu}^{n,\text{wght}}_{f,h}(I, \mathbf{w}) &= \hat{\mu}^{n,\text{wght}}_{f,h}(I, \mathbf{w}, (\Phi_1, \ldots, \Phi_n), T) \\ &= \Big(\sum w_i(\Phi_i, T)\Big)^{-1} \sum_{i=1}^{n} w_i(\Phi_i, T)\hat{\mu}_{f,h}(I, \Phi_i, T), \end{aligned}$$

with

$$\hat{\mu}_{f,h}(I, \Phi, T) = \frac{\hat{\alpha}_{f,h}(I, \Phi, T)}{\hat{\alpha}_{1,h}(I, \Phi, T)},$$

$$\hat{\alpha}_{f,h}(I, \Phi, T) = \sum\nolimits^{\neq}_{(t_1,y_1),(t_2,y_2)\in\Phi} f(y_1)h(y_1)\mathbf{1}_{(t_1,t_2)\in C(T,I)}.$$

Here, the weights $w_i(\Phi_i, T)$ are required to converge stochastically to some constant within each ergodicity class. For instance, with $w_i(\Phi_i, T) = T^{-1}\hat{\alpha}_{f,h}(I, \Phi_i, T)$ and $w_i(\Phi_i, T) = 1$, respectively, $\mu^{(2)}_{f,h}(I)$ and $\tilde{\mu}^{(2)}_{f,h}(I)$ can be estimated consistently.

Estimation of the tail behavior generally requires a trade-off between tail relevance and the amount of data. For estimation of $\tilde{\xi}$, a suitable threshold $u$ has to be chosen such that the estimator of $\tilde{\xi}(I, u)$ can be taken as an approximation of $\tilde{\xi}(I)$. Plugging in $f_u(y) = \log y - u$ and $h_u(y) = \mathbf{1}_{\log y > u}$ into (5), the canonical estimator of $\xi(I, u)$, based on a single realization of $\Phi$, is

$$\hat{\xi}(I, u, \Phi, T) = \frac{\sum^{\neq}_{(t_1,y_1),(t_2,y_2)\in\Phi} (\log y_1 - u)\mathbf{1}_{\log y_1 > u}\mathbf{1}_{(t_1,t_2)\in C(T,I)}}{\sum^{\neq}_{(t_1,y_1),(t_2,y_2)\in\Phi} \mathbf{1}_{\log y_1 > u}\mathbf{1}_{(t_1,t_2)\in C(T,I)}}.$$

For $n$ realizations of $\Phi$, the estimator

$$(6) \qquad \begin{aligned} \hat{\xi}^{n,\text{wght}}(I, u, \mathbf{w}) &= \hat{\xi}^{n,\text{wght}}(I, u, \mathbf{w}, (\Phi_1, \ldots, \Phi_n), T) \\ &= \hat{\mu}^{n,\text{wght}}_{f_u, h_u}(I, \mathbf{w}) \end{aligned}$$

will be considered, where the right-hand side (RHS) is given by (5). If all weights are chosen equal to 1, the estimator is already consistent for $\tilde{\xi}(I, u)$. Under some additional assumptions on the mark-location dependence, the estimator's variance can be improved by choosing different weights while retaining consistency. To this end, assume $\mathbb{E}[\hat{\xi}(I,u,\Phi_i,T)\,|\,\mathcal{A}_u^*]$ to be constant a.s., where $\mathcal{A}_u^*$ denotes the $\sigma$-algebra that contains all information about the point locations of $\Phi_1,\ldots,\Phi_n$ and about the locations of points whose log marks exceed the threshold $u$. A formal definition of $\mathcal{A}_u^*$ is given in the appendix. Then, the optimal weights in (6) are given by the inverse of the conditional variances, i.e., $w_i = \mathrm{Var}[\hat{\xi}(I,u,\Phi_i,T)\,|\,\mathcal{A}_u^*]^{-1}$ [33, Prop. 2], provided that they are stochastically independent of the mixing random variable $Q$. To exemplify these conditional variances, explicit expressions under some idealized independence and mixing assumptions are derived in the following. The worthiness of the following results for practical applications is discussed in the adjacent Section 2.4.

**Condition 2.2** (Independent-noise-marking)**.** Let $Y_i$, $i\in\mathbb{Z}$, be iid variables in the MDA of a standard Fréchet distribution. We say that an MPP $\Phi$ satisfies the condition (Independent-noise-marking), if $\Phi \stackrel{d}{=} \{(t_i,Y_i)\,|\,t_i\in\tilde{\Phi}\}$ for some stationary unmarked point process $\tilde{\Phi}$ on $\mathbb{R}$, for which neighboring points have some minimum distance $d_0>0$ and which is independent of the $Y_i$.

**Condition 2.3** (GRFM-trans)**.** Let $\tilde{\Phi}$ be as in condition (Independent-noise-marking), and let $\{Y(t):t\in\mathbb{R}\}$ be an independent random process which arises from a stationary Gaussian process $Z$ by a monotone transformation of the margins, i.e., $Y=g(Z)$, such that the marginals of $Y$ are in the Fréchet MDA. The covariance function $C$ of $Z$ is assumed to have a finite range, i.e., $C(h)=0$ for all $|h|>h_0$ for some $h_0>0$. Then, we say that an MPP $\Phi$ is a Gaussian random field model with transformed margins, for short: $\Phi$ satisfies the condition (GRFM-trans), if $\Phi \stackrel{d}{=} \{(t_i,Y(t_i))\,|\,t_i\in\tilde{\Phi}\}$.

**Theorem 2.4.** *For a stationary MPP as in (GRFM-trans),*

$$\mathrm{Var}[\hat{\xi}(I,u,T)\,|\,\mathcal{A}_u^*] = v_u\left[\frac{\sum_{t_1\in\Phi_{\mathrm{g}}\cap[0,T]} n(t_1,\Phi_{\mathrm{g}},I,u)^2}{\left[\sum_{t_1\in\Phi_{\mathrm{g}}\cap[0,T]} n(t_1,\Phi_{\mathrm{g}},I,u)\right]^2}+\varepsilon_u\right],$$

*where* $v_u = \mathrm{Var}[\log Y(0)\,|\,\log Y(0)>u]$, $n(t_1,\Phi_{\mathrm{g}},I,u) = \mathbf{1}_{\log Y(t_1)>u}\cdot\sum_{t_2\in\Phi_{\mathrm{g}}\setminus\{t_1\}}\mathbf{1}_{t_2-t_1\in I}$ *and* $\varepsilon_u$ *is an* $\mathcal{A}_u^*$-*measurable random variable with* $|\varepsilon_u|\to 0$ *a.s. and in* $L_1$*, as* $u\to\infty$.

The proof is given in the appendix. Since condition (Independent-noise-marking) is a special case of condition (GRFM-trans), the following corollary is obvious from the proof.

**Corollary 2.5.** *For an MPP $\Phi$ satisfying the condition (Independent-noise-marking), the assertion of Theorem 2.4 holds with $\varepsilon_u=0$.*

If $u$ is large enough, the term $\varepsilon_u$ in Theorem 2.4 can be neglected and the resulting optimal weights in $\hat{\xi}^{n,\mathrm{wght}}(I,u,\mathbf{w})$ are

$$(7)\qquad w_i(\Phi_i,I,u) = v_u^{-1}\frac{\left[\sum_{t_1\in\Phi_{i,\mathrm{g}}\cap[0,T]} n(t_1,\Phi_{i,\mathrm{g}},I,u)\right]^2}{\sum_{t_1\in\Phi_{i,\mathrm{g}}\cap[0,T]} n(t_1,\Phi_{i,\mathrm{g}},I,u)^2}.$$

For the continuous case $\xi(r,u)$, the same estimators as for $\xi(I,u)$ can be used, while the indicator function $\mathbf{1}_{(t_1,t_2)\in C(T,I)}$ might be replaced by a general kernel $K_h$. Then, the above formulae for the conditional variance and the weights are still valid if $n(t_1,\Phi_{\mathrm{g}},I,u)$ is replaced by

$$n(t_1,\Phi_{\mathrm{g}},r,u) = \mathbf{1}_{\log Y(t_1)>u}\sum_{t_2\in\Phi_{\mathrm{g}}\setminus\{t_1\}} K_h(r-(t_2-t_1)).$$

## 2.4 Confidence intervals

In the following, the asymptotic distribution of $\hat{\xi}^{n,\mathrm{wght}}(I,u,\mathbf{w})$ under the above assumptions (Independent-noise-marking) and (GRFM-trans) is derived. If these assumptions are violated, the reliability of the resulting confidence intervals (CIs) can be assessed, e.g., via the nonparametric subsampling approach according to Politis and Sherman [39].

### 2.4.1 Confidence intervals based on (Independent-noise-marking) and (GRFM-trans)

The estimator $\hat{\xi}^{n,\mathrm{wght}}(I,u,\mathbf{w})$ involves two levels of aggregation of independent or weakly dependent random terms: the outer summation over different realizations and the inner summation over all points of a particular realization.

First consider the inner level of aggregation and assume that $\Phi$ is ergodic throughout this paragraph. Then $\hat{\xi}(I,u) = \hat{\alpha}_{f_u,h_u}(I,\Phi,T)/\hat{\alpha}_{1,h_u}(I,\Phi,T)$ is an average of a random number of (dependent) summands and the following CLT-type result follows from [33, Thm. 1].

**Theorem 2.6.** *Let $\Phi$ be an MPP as in (Independent-noise-marking) or (GRFM-trans) and let $(u_T)_{T\geq 0}$ be a family of non-negative non-decreasing numbers such that* $u_\infty = \lim_{T\to\infty} u_T\in[0,\infty]$ *exists and*

$$\frac{T^{-1}\hat{\alpha}_{1,1}(I,\Phi,T)-\lambda}{\mathbb{E}\hat{\alpha}_{1,h_{u_T}}(I,\Phi,1)}\to 0 \text{ a.s.,}\quad \text{as } T\to\infty,$$

*where $\lambda$ is the intensity of point locations. Let*

$$\hat{\alpha}^*_{f_u,h_u}(I,\Phi,T) = \sum^{\neq}_{(t_1,y_1),(t_2,y_2)\in\Phi}\left(f_u(y_1)-\mu^{(2)}_{f_u,h_u}(I)\right)h_u(y_1)\mathbf{1}_{(t_1,t_2)\in C(T,I)}$$

*be a centered version of* $\hat{\alpha}_{f_u,h_u}(I,\Phi,T)$.

*Then, for $I \in \mathcal{B}(\mathbb{R})$ and $T \to \infty$,*

$$\frac{\hat{\alpha}^*_{f_{u_T},h_{u_T}}(I,\Phi,T)}{\sqrt{\hat{\alpha}_{1,h_{u_T}}(I,\Phi,T)}} \Rightarrow \mathcal{N}(0, s_{u_\infty}),$$

*where*

$$s_{u_\infty} = \lim_{T\to\infty} v_{u_T} \left[ \frac{\mathbb{E} \sum_{t_1 \in \Phi_g \cap [0,1]} n(t_1, \Phi_g, I, u_T)^2}{\mathbb{E}\hat{\alpha}_{1,h_{u_T}}(I,\Phi,1)} + \mathbb{E}\varepsilon_{u_T} \right],$$

$$v_u = \operatorname{Var}\big[\log Y(0) \mid \log Y(0) > u\big], \quad u \in [0,\infty),$$

*and $\varepsilon_u$ is given by Thm. 2.4 or Cor. 2.5. If the family $(u_T)_{T\ge 0}$ is eventually constant, then $u_T$ can be replaced by the limiting constant $u_\infty \in [0,\infty)$. Furthermore, for $u$ large (and $T>0$ arbitrary),*

$$\text{(8)} \qquad \operatorname{Var} \frac{\hat{\alpha}_{f_u,h_u}(I,\Phi,T)}{\hat{\alpha}_{1,h_u}(I,\Phi,T)} = \operatorname{Var} \frac{\hat{\alpha}^*_{f_u,h_u}(I,\Phi,T)}{\hat{\alpha}_{1,h_u}(I,\Phi,T)} \approx v_u \mathbb{E} \left\{ \frac{\sum_{t_1 \in \Phi_g \cap [0,T]} n(t_1,\Phi_g,I,u)^2}{\hat{\alpha}_{1,h_u}(I,\Phi,T)^2} \right\}.$$

For a proof, the reader is referred to the appendix.

Concerning the outer level of aggregation in $\hat{\xi}^{n,\mathrm{wght}}(I,u,\mathbf{w})$, again by a CLT argument, the finite sample distribution is approximately Gaussian. By assumption, $\mathbb{E}[\hat{\xi}(I,u,\Phi_i,T) \mid \mathcal{A}^*_u]$ is a.s. constant and the weights in (7) are $\mathcal{A}^*_u$-measurable. The corresponding variance is obtained by a straightforward calculation using Thm. 2.4:

$$\begin{aligned}
&\operatorname{Var}\left[\hat{\xi}^{n,\mathrm{wght}}(I,u,\mathbf{w})\right] \\
&= \mathbb{E}\operatorname{Var}\left[ \frac{\sum_{k=1}^n w(\Phi_k, \cup_j I_j, u) \frac{\hat{\alpha}_{f_u,h_u}(\Phi_k,I,T)}{\hat{\alpha}_{1,h_u}(\Phi_k,I,T)}}{\sum_{k=1}^n w(\Phi_k,\cup_j I_j,u)} \,\middle|\, \mathcal{A}^*_u \right] \\
&\approx v_u \cdot \mathbb{E}\left[ \left[\frac{1}{\sum_k w(\Phi_k,I,u)}\right]^2 \sum_{k=1}^n \frac{\left[\sum_{t_1\in\Phi_{k,g}\cap[0,T]} n(t_1,\Phi_{k,g},I,u)\right]^2}{\sum_{t_1\in\Phi_{k,g}\cap[0,T]} n(t_1,\Phi_{k,g},I,u)^2} \cdot \frac{\sum_{t_1\in\Phi_{k,g}\cap[0,T]} n(t_1,\Phi_{k,g},I,u)^2}{\left[\sum_{t_1\in\Phi_{k,g}\cap[0,T]} n(t_1,\Phi_{k,g},I,u)\right]^2} \right] \\
&= v_u \cdot n \cdot \mathbb{E}\left[ \sum_{k=1}^n \frac{\sum_{t_1\in\Phi_{k,g}\cap[0,T]} n(t_1,\Phi_{k,g},I,u)^2}{\left[\sum_{t_1\in\Phi_{k,g}\cap[0,T]} n(t_1,\Phi_{k,g},I,u)\right]^2} \right]^{-2}.
\end{aligned}$$

If equal weights are used,

$$\text{(9)} \qquad \operatorname{Var}\left[\hat{\xi}^n(I,u)\right] \approx v_u \cdot n^{-1} \cdot \mathbb{E}\left[ \frac{\sum_{t_1\in\Phi_{k,g}\cap[0,T]} n(t_1,\Phi_{k,g},I,u)^2}{\left[\sum_{t_1\in\Phi_{k,g}\cap[0,T]} n(t_1,\Phi_{k,g},I,u)\right]^2} \right].$$

In any of the two cases, the resulting CI is given by the Gaussian approximation. We will refer to the CIs based on this approach as *model-based confidence intervals*.

It should be mentioned here that biases known from classical theory on estimation of first and second order tail parameters might in principle also occur within the MPP framework of this paper. Under suitable second-order conditions, the Hill estimator, for instance, is known to be asymptotically normal, with a bias depending on the second order tail parameter and the threshold $u$ applied for the estimator [e.g. 10, Ch. 3]. To find appropriate conditions that would allow for similar results in the MPP framework is not straightforward due to the complex structure of dependencies. Hence, in Theorem 2.6, it is not the estimator's deviation from the conditional tail index $\xi(I)$ that is considered, but the deviation from its finite threshold counterpart $\mu^{(2)}_{f_u,h_u}(I)$.

A second type of bias typically arises when the tail index is considered as a function of a covariate and then estimated with a kernel-based approach, where the kernel averages over a certain range in the covariate space. Based on iid data, Goegebeur and De Wet [18] provide sound theoretical results for such a situation, including a bias correction for the corresponding tail index estimator. While the conditioning on the existence of points in an MPP, as considered in the paper at hand, might appear similar to the existence of covariate information, it can certainly not be treated within the same framework. This is again due to the fact that the distances within the point pattern of an MPP might be linked to the tail properties of the marks in a highly endogenous way; an analog approach for MPPs is not obvious.

### 2.4.2 Subsampling-based confidence intervals

While the assumptions (Independent-noise-marking) and (GRFM-trans) allow for a theoretical calculation of the tail index estimator's asymptotic variance, subsampling provides a fully non-parametric way of estimating the uncertainty of the estimator. A broad survey on bootstrapping and subsampling methods is given by Politis, Romano and Wolf [38]. For a general statistic $s(\Phi,T)$ for which $T\operatorname{Var}[s(\Phi,T)] \to V$ for some $V>0$ as $|T|\to\infty$, Politis and Sherman [39] showed that, under some mixing assumptions, $V$ is consistently estimated through

$$\hat{V} = [(1-c)T]^{-1} \int_{[0,(1-c)T]} cT \cdot \mathbb{E}\left[ s(\Phi_{-y}, cT) - \overline{s(\Phi,cT)} \right]^2 dy$$

if $c = c_T \to 0$ and $cT \to \infty$ as $T \to \infty$. Here, $\overline{s(\Phi, cT)} = [(1-c)T]^{-1} \int_{[0,(1-c)T]} s(\Phi_{-y}, cT)dy$ and $\Phi_x$ denotes the translation of the point pattern $\Phi$ by $x$.

The Riemann sum approximation of $\hat{V}$, is then given by

$$\hat{V}^* = n^{-1} \sum_{i=1}^{n} \frac{T}{n} \left[ s(\Phi, [\tfrac{i-1}{n}T, \tfrac{i}{n}T]) - \bar{s} \right]^2$$

with $\bar{s} = n^{-1} \sum_{i=1}^{n} s(\Phi, [\frac{i-1}{n}T, \frac{i}{n}T])$. Hence,

$$\mathrm{Var}[s(\Phi, T)] \approx T^{-1}\hat{V}^* = n^{-2} \sum_{i=1}^{n} \left[ s(\Phi, [\tfrac{i-1}{n}T, \tfrac{i}{n}T]) - \bar{s} \right]^2.$$

With regard to $\hat{\xi}^n(I, u)$, which already is an average of $n$ realizations, an additional splitting of the observation window is not needed if $n$ is sufficiently large. Then, $\mathrm{Var}\, \hat{\xi}^n(I, u)$ is naturally estimated through $n^{-2} \sum_{i=1}^{n} [\hat{\xi}(I, u, \Phi_i) - \bar{\xi}]^2$, where $\bar{\xi} = n^{-1} \sum_{i=1}^{n} \hat{\xi}(I, u, \Phi_i)$. Confidence intervals can again be based on the quantiles of the normal distribution since $\hat{\xi}^n(I, u)$ is asymptotically Gaussian (for $n \to \infty$) by the classical CLT. We will refer to these CIs as *subsampling-based confidence intervals*.

## 3. SIMULATION STUDY

### 3.1 The model

Doubly stochastic Poisson processes (DSPPs), also called Cox processes, are well-established in the modeling of high-frequency financial data [e.g., 6, 23, 29]. Here, a DSPP-based MPP model is considered, combined with an intensity-dependent marking [e.g., 25, 36]. Let $Z(t) = (Z_1(t), Z_2(t))$, $t \in \mathbb{R}$, be a bivariate stationary Gaussian field, where $Z_1$ generates the intensity and $Z_2$ drives the marks. This approach allows for a flexible management of dependencies between intensity and marks via the matrix-valued cross-covariance function $\mathrm{C}(r) = \begin{pmatrix} \mathrm{C}_{11}(r) & \mathrm{C}_{12}(r) \\ \mathrm{C}_{21}(r) & \mathrm{C}_{22}(r) \end{pmatrix}$, $r \in \mathbb{R}$, where $\mathrm{C}_{ij}(r) = \mathrm{Cov}(Z_i(0), Z_j(r))$, $i, j \in \{1, 2\}$. The mean of $Z$ is denoted by $(m_1, m_2)$. In particular, the random intensity of point locations is given by $\exp(Z_1(\cdot))$, i.e., the unmarked ground process $\Phi_{\mathrm{g}} = \{t_i : i \in \mathbb{N}\}$ is a log Gaussian Cox process (LGCP) with random intensity measure $\Lambda(B) = \int_B \exp(Z_1(t))dt$. In addition, let $(Y_i)_{i\in\mathbb{N}}$ be a sequence of independent random Fréchet variables with $Y_i \sim F_{\alpha(t_i)}$, where $F_\alpha(x) = \exp(-x^{-\alpha})$ denotes the Fréchet distribution function with parameter $\alpha$, and $\alpha(\cdot)$ is given by $\alpha(t) = \alpha_0 + \alpha_1 \exp(-Z_2(t))$, $\alpha_0, \alpha_1 \geq 0$, $t \in \mathbb{R}$. Let further $(S_i)_{i\in\mathbb{N}}$ be a sequence of iid random signs with $\mathbb{P}(S_1 = 1) = \mathbb{P}(S_1 = -1) = 0.5$. Then, conditionally on $Z$, let the marks be given by $y_i = m(t_i) = S_i Y_i$, i.e., their absolute values are Fréchet-distributed with an intensity-dependent tail parameter and their signs are random and independent of $Z$. To have finite first moments of the marks, $\alpha_0 \geq 1$ is assumed.

Since $F_\alpha^{-1}(V) \sim F_\alpha$ for $V \sim U[0,1]$, the marks $y_i = m(t_i)$ can be considered as a deterministic function of time provided that $Z$, the random field $V = \{V(t)\}_{t\in\mathbb{R}}$ of iid $U[0,1]$ variables and the random field $S = \{S(t)\}_{t\in\mathbb{R}}$ of iid signs are known:

$$\begin{aligned} &m_{v,s,\lambda}(t) \\ &= \left[ m(t) \,|\, V(\cdot) = v(\cdot),\ S(\cdot) = s(\cdot),\ \exp(Z_2)(\cdot) = \lambda_2(\cdot) \right] \\ &= s(t) \cdot F^{-1}_{\alpha_0 + \alpha_1/\lambda_2(t)}(v(t)). \end{aligned}$$

If the two components of $Z$ are positively correlated, the Fréchet parameter $\alpha$ tends to be small when the intensity of points is high. This will lead to increased conditional tail indices $\xi(r, u)$ for small temporal distances $r$.

### 3.2 Theoretical value of $\xi(r, u)$

Since $[\Phi|Z, V, S]$ is a Poisson point process with deterministic marks, the conditional tail index $\xi(r, u)$ can at least partially be treated analytically using an extended Campbell theorem and the fact that the reduced Palm measure of a Poisson process coincides with the probability measure $P^\Phi$ [e.g., 9, Prop. 13.1.IV and Prop. 13.1.VII, resp.]. For the second order moment measure $\alpha_{f,h}$ (cf. (1)), this yields

$$\begin{aligned} &\alpha^{(2)}_{f^\ell,h}(C(I)) \\ &= \iiint \int_{[0,1]} \mathbb{E}_{\Phi|Z,V,S}[(f^\ell h)(m_{v,s,\lambda}(t_1))\Phi_{\mathrm{g}}(I + t_1)] \\ &\qquad \cdot \lambda_1(t_1)dt_1\, \mathbb{P}^{(\exp(Z_1),\exp(Z_2))}(d\lambda_1, d\lambda_2)\mathbb{P}^V(dv)\mathbb{P}^S(ds) \\ &= \iiint \int_{[0,1]} (f^\ell h)(m_{v,s,\lambda}(t_1)) \int_{I+t_1} \lambda_1(r)dr \\ &\qquad \cdot \lambda_1(t_1)dt_1\, \mathbb{P}^{(\exp(Z_1),\exp(Z_2))}(d\lambda_1, d\lambda_2)\mathbb{P}^V(dv)\mathbb{P}^S(ds), \end{aligned}$$

for $\ell \in \{0, 1\}$. Due to the Cox-process-based construction of $\Phi$, the measures $\alpha^{(2)}_{f^\ell,h}(C(\cdot))$, $\ell \in \{0, 1\}$, are dominated by the Lebesgue measure $\nu$ on $\mathbb{R}$ and with Fubini's theorem,

$$\begin{aligned} &\frac{\partial \alpha^{(2)}_{f^\ell,h}(C(r))}{\partial \nu(r)} \\ &\quad = \iiint (f^\ell h)(m_{v,s,\lambda}(t_1))\lambda_1(r + t_1)\lambda_1(t_1) \\ &\qquad\qquad \mathbb{P}^{(\exp(Z_1),\exp(Z_2))}(d\lambda_1, d\lambda_2)\mathbb{P}^V(dv)\mathbb{P}^S(ds). \end{aligned}$$

Hence,

$$\begin{aligned} (10) \quad \xi(r, u) &= \frac{\partial \alpha^{(2)}_{f_u,h_u}(C(r))}{\partial \alpha^{(2)}_{1,h_u}(C(r))} \\ &= \frac{\partial \alpha^{(2)}_{f_u,h_u}(C(r))}{\partial \nu(r)} \cdot \left( \frac{\partial \alpha^{(2)}_{1,h_u}(C(r))}{\partial \nu(r)} \right)^{-1} \\ &= \frac{\int (f_u h_u)(m_{v,s,\lambda}(0))\lambda_1(r)\lambda_1(0)\, \mathbb{P}(d(\lambda_1, \lambda_2, v, s))}{\int h_u(m_{v,s,\lambda}(0))\lambda_1(r)\lambda_1(0)\, \mathbb{P}(d(\lambda_1, \lambda_2, v, s))}, \end{aligned}$$

where $\mathbb{P}(d(\lambda_1, \lambda_2, v, s))$ is short notation for $\mathbb{P}^{(\exp(Z_1),\exp(Z_2))}(d\lambda_1, d\lambda_2)\mathbb{P}^V(dv)\mathbb{P}^S(ds)$. As before, for $y > 0$, $f_u(y) = \log y - u$ and $h_u(y) = \mathbf{1}_{\log y > u}$, and $f_u(y) = h_u(y) = 0$ for $y \leq 0$. Note that the RHS of (10) is not an integral w.r.t. the law of the MPP anymore, but only w.r.t. on the law of the random fields that drive the intensity of points and the marking. Although it is analytically intractable, Monte-Carlo simulation of $Z$, $V$ and $S$ provides an approximation $\hat{\xi}^{\mathrm{MC}}(r, u)$ of $\xi(r, u)$ for the above model. This enables a direct comparison of the true conditional tail index with the estimated one based on realizations of the full point process.

## 3.3 Results

Since, by construction, locations and marks are dependent, the confidence intervals derived in Section 2.4, which are based on the assumption (GRFM-trans), are only approximate and possibly underestimate the true variance. Compared to this deviation through mark location interaction, the fact that the covariance function of the underlying random field does not meet the finite range condition in assumption (GRFM-trans) will be negligible. By this simulation study, the actual level of the confidence intervals is determined.

The particular set-up is the following: The random field $Z_1$ has a mean value of $m_1 = -\log(0.5)$ and the exponential covariance model $\mathrm{C}(h) = 0.1\exp(-|h|/4)$. Perceiving distances as being measured in minutes, this choice causes the average distance between consecutive observations to be approximately 0.5 minutes and interaction effects to range up to 10 minutes, which roughly corresponds to the respective numbers in real transaction data (cf. Section 4). The second component of $Z$ is a linear combination of shifts of $Z_1$: $Z_2(\cdot) = Z_1(\cdot) + \sum_{i=1}^n c_i[Z_1(\cdot) - Z_1(\cdot - s_i)]$ with $c_i, s_i \in \mathbb{R}$. This determines a particular form for the cross-covariance function of $(Z_1, Z_2)$. While $c_1 = \ldots = c_n = 0$ implies completely symmetric interaction effects between marks and locations, positive values of $c$ and $s$ introduce asymmetry: $Z_2(t)$ is positively correlated with $Z_1(t)$ but $Z_2(t)$ is particularly large if $Z_1$ is small at the locations $t - s_i$. Since the Fréchet parameters of the marks are given by $\alpha(t) = \alpha_0 + \alpha_1\exp(-Z_2(t))$, the larger the value of $Z_2$, the heavier the tail of the mark distribution. Hence, this specification of $Z_2$ with positive values of $s_i$ induces a heavy tail at time $t$ if the intensity of points at $t$ is large *or* if there is an increase in intensity immediately before $t$. We choose $n = 100$ for smoothness reasons and $(c_1, \ldots, c_{100}) = (0.100, 0.099, \ldots, 0.001)$ and $(s_1, \ldots, s_{100}) = (2, 4, \ldots, 200)/60$. Further, let $\alpha_0 = 3$ and $\alpha_1 = 0.1$.

The model is simulated on a 24,000 hour interval, which roughly corresponds to 3,000 days of trading, i.e., the point process contains approximately 3 million points. Figure 1 summarizes the behavior of the estimator $\hat{\xi}^n(r, u)$ based on such a realization, where $u$ is the 95% and the 99% sample quantile and $n$ is chosen to be 100, which means that the simulated dataset is split into 100 parts of a length

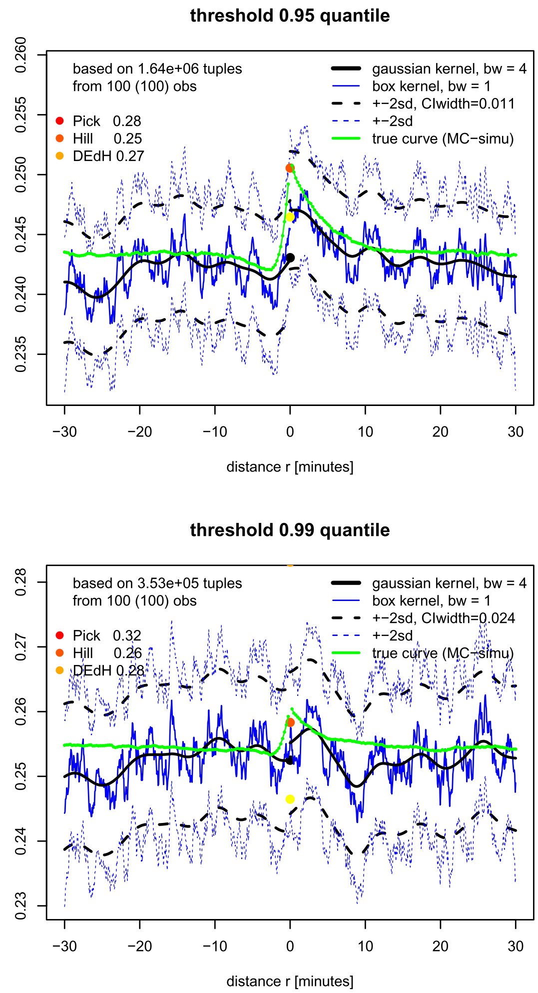


*Figure 1. Estimation of $\xi(r, u)$ for $u$ the 95% (top) and the 99% sample quantile (bottom) together with pointwise approximated 95%-CIs. The Gaussian kernel with a bandwidth of 4 and the rectangular kernel with a bandwidth of 1 are used. An approximation of the theoretical values $\xi(r, u)$ is based on Monte-Carlo simulation of the RHS of (10) with $10^6$ realizations of the random fields $Z$, $V$ and $S$. (Color figure online)*

roughly corresponding to one month. Note that in the simulation, there is no instationarity or regime-switching included; hence, the non-ergodic modeling does not play an important role, here. As kernels for the estimator $\hat{\xi}(r, u)$, the Gaussian and the rectangular kernel are used with a bandwidth of 4 and 1, respectively. The approximated pointwise 95%-confidence intervals according to (8) in Theorem 2.6 (model-based CIs) are included. For the approximation $\hat{\xi}^{\mathrm{MC}}(r, u)$ of $\xi(r, u)$, 1 million realizations of the random fields $Z$, $V$ and $S$ on $[-30, 30]$ are generated. Note that, once $Z$ is simulated, it is sufficient to simulate $V(\cdot) \sim U[u^*, 1]$ with $u^* = \inf_{t\in[-30,30]} F_{\alpha(t)}(\exp(u))$. For smaller values of $V$, $h_u(m(t))$ is zero and the corresponding points would not enter into the estimator.

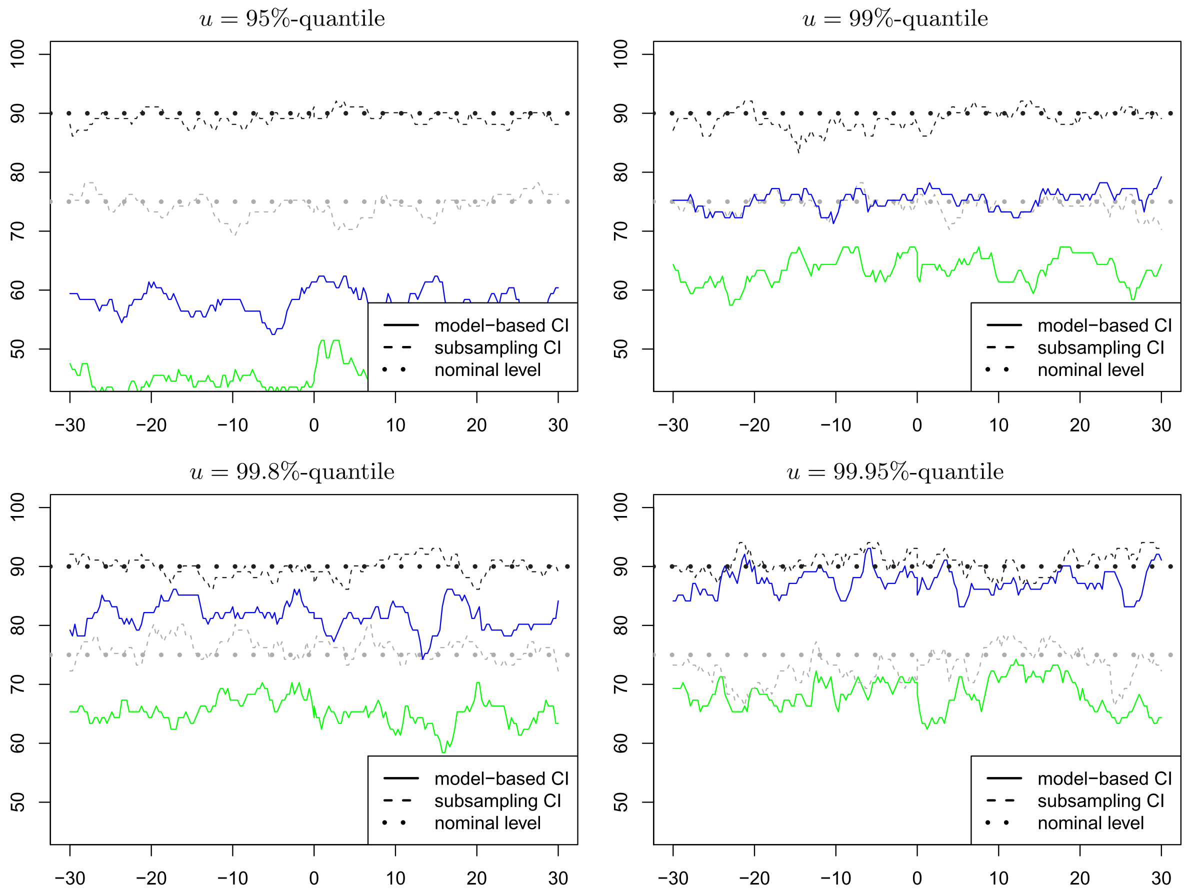


*Figure 2. Empirical level of the model-based CIs according to (8) in Theorem 2.6 (continuous lines) and of subsampling-based CIs from Section 2.4.2 (dashed lines) for different thresholds $u$ (from left to right, then top to bottom: 95%, 99%, 99.8% and 99.95%-quantile) and for the two nominal levels 75% and 90% (dotted lines).*

In order to validate the confidence intervals, a realization of the above model is simulated and the confidence intervals are calculated. Then, the model is simulated another 100 times and for each grid point $r \in [-30, 30]$, those realizations are counted whose respective values $\hat{\xi}^n(r, u)$ fall into the afore calculated CI. Then, the roles are interchanged 100 times such that each realization once becomes the center of the CI. As nominal levels, we choose 75% and 90%. Figure 2 shows the results for different values of the threshold $u$. It displays that for a relatively low threshold ($u$ = 95%-quantile), the variance of $\hat{\xi}^n(r, u)$ is considerably underestimated leading to an empirical level that is up to 25 percentage points below the nominal level. With increasing threshold, this error decreases. Already for the 99.8%-quantile, the confidence intervals hit the nominal level in average. The subsampling-based confidence intervals hit the nominal level for all thresholds, as expected.

Figure 3 shows the estimator $\hat{\xi}^n(r, u)$, applied to the union of all 101 simulated datasets, together with an approximation to the true function $\xi(r, u)$, obtained from Monte-Carlo simulation. The MC-estimate is fairly smooth since it is based on $10^8$ random field realizations on the interval $[-30, 30]$. The width of the confidence intervals is roughly one tenth of that in Figure 1 since there are 101 realizations instead of one.

## 4. APPLICATION TO TRANSACTION DATA FROM THE GERMAN STOCK EXCHANGE

The conditional tail index estimator $\hat{\xi}(r, u)$ is exemplarily applied to large transaction datasets from stock trading in Germany, processed via the Xetra trading system between 1997 and 2004. Blocks of size one year are considered separately in order to exclude possible long-term effects. The same data pre-processing as in Engle [15] is applied in order to account for diurnal patterns in the duration and return series. Further, the original returns are transformed to *returns per time unit* [15].

In correspondence with the various contributions on structural changes and nonlinear modeling of financial processes mentioned in the introduction, also in the transaction datasets, periods of trading can be observed that behave differently from the major part of the trading time. In the case

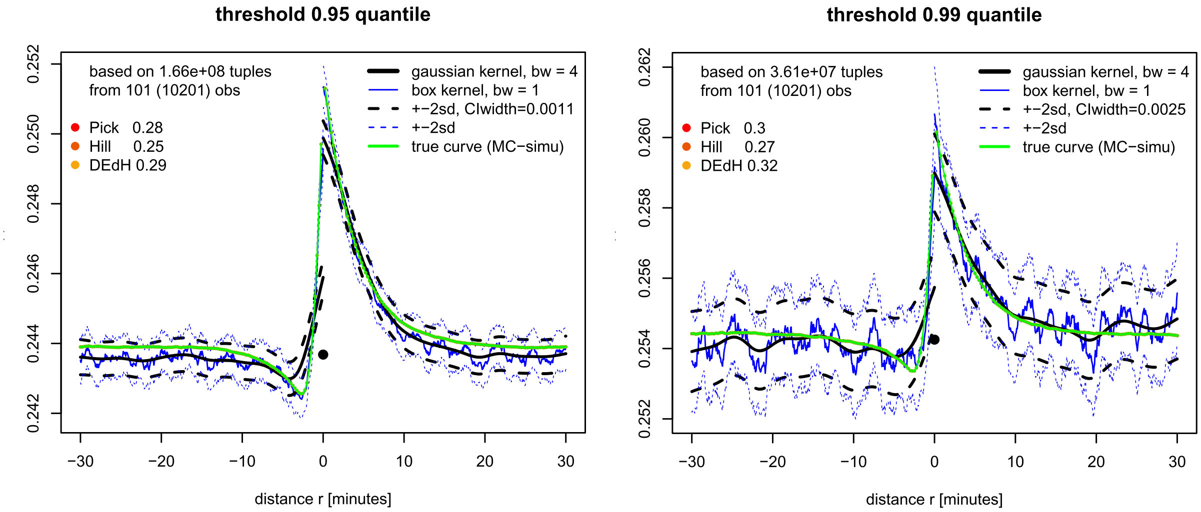


*Figure 3. Analogously to Figure 1, but averaged over all 101 realizations.*

that tail characteristics of the logreturns are of interest, only the set of extremal transactions is considered which may further strengthen the effects of structural breaks. Hence, we consider a realization $\varphi$ of the process $\Phi$, observed on a certain interval $B$, as a concatenation of multiple realizations $\varphi_1, \dots, \varphi_n$ of a possibly non-ergodic MPP, observed on smaller intervals $B_1, \dots, B_n$, respectively, where the intervals are pairwise disjoint and $B = \cup_i B_i$; the objective is then to estimate the conditional tail index $\tilde{\xi}(r, u)$. This means that each trading period is considered to belong to one randomly chosen regime (ergodicity class) out of a possibly infinite number of different regimes. Here, additionally independence between the concatenated parts is assumed, which is in general only an approximation to the truth. However, if the clock time length $D$ of each period is large compared to the average inter-event distance within each realization of the point process, dependence between events from different parts can be expected to be fairly small and the error of this independence assumption is negligible. This argument might be formalized via some weak mixing conditions guaranteeing that the estimators of $\mu_f^{(i)}$, applied to the small realizations, become asymptotically independent (as $D \to \infty$). By the same arguments, edge effects due to finite observation windows can be neglected [e.g., 41].

As regards the concrete choice of $D$, McCulloch and Tsay [34], for instance, assume that the length of each realization corresponds to one trading day, but also other choices of $D$ might be adequate, depending on the statistical questions at hand.

In order to test the results for being significant, the following null model is considered: Within each subsample of length $D$, the marks of the MPP are randomly permuted while the pattern of point locations is kept fixed. This procedure can be expected to destroy any mark-location dependence in the MPP. Hence, applying the above estimator to multiple realizations of this null model yields a set of reference curves that correspond to no interaction effects.

## 4.1 Results

Four different levels of disaggregation are applied, in particular, the data is split into blocks of length one year ($n = 1$), one month ($n = 12$), one week ($n = 52$) and one day ($n \approx 250$). It turns out that a choice $n > 1$ yields more stable results and smaller estimated variances, compared to $n = 1$. However, going below a length of one week (i.e., $n > 52$) does not seem to be sensible since, particularly through the selection of extreme transactions, data become sparse and many of the small blocks would not contain any observation exceeding the threshold. Moreover, the estimation results do not differ significantly between moderate choices of $n$, i.e., $n$ between 12 and 52. In the following, only the results for partitioning into blocks of length one week are shown, i.e. $n = 52$.

In Section 2.3, variance-minimizing weights were introduced that maintain the consistency property of $\hat{\xi}^{n,\mathrm{wght}}(r, u, \mathbf{w})$ for $\tilde{\xi}(r, u)$ under some suitable independence assumptions. Though, it turns out that non-equal weighting (i.e., the use of $\hat{\xi}^{n,\mathrm{wght}}(r, u, \mathbf{w})$ instead of $\hat{\xi}^{n}(r, u)$) only marginally improves the estimators variance since all weights turn out to be fairly similar in this particular dataset ($\exp(\mathrm{entropy}(\mathbf{w})) \approx n - 2$). Hence, in order to be able to compare the estimated variance to a subsampling-based variance estimate, we restrict to the unweighted estimator in the following.

Exemplarily, Figure 4 shows the conditional tail index estimator for a one-year period (2004) of transaction data of the Deutsche Telekom AG stock (ISIN DE0005557508) with a total of 898,000 transactions. Here, only the lower tail, i.e., negative log returns, are considered. While in the above simulation study, the tails were symmetric by con-

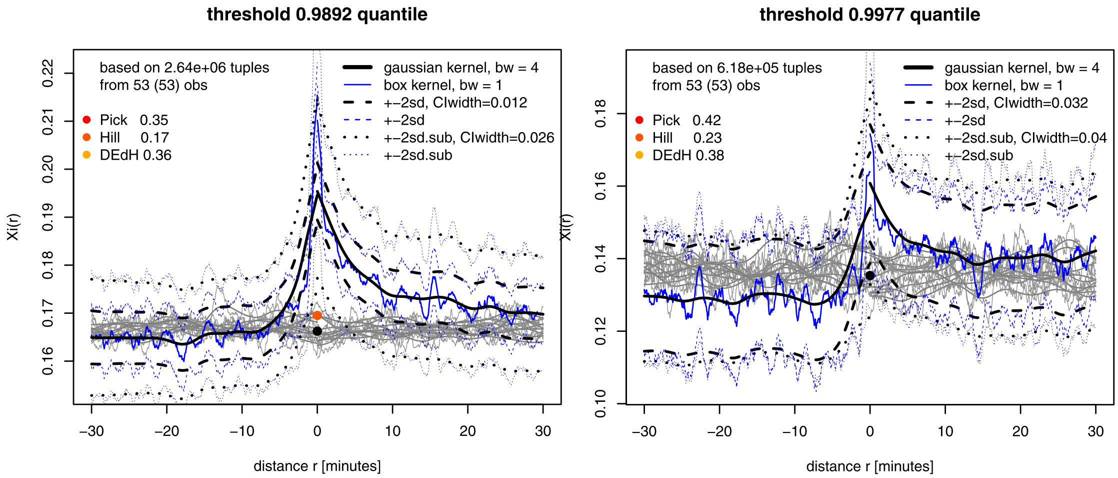


*Figure 4. Estimation of $\xi(r,u)$ for a one-year period of transaction data of the Deutsche Telekom AG stock (ISIN DE0005557508) with a total of $898{,}000$ transactions. $u$ being the $98.9\%$- (left) and the $99.77\%$-quantile (right). Pointwise $95\%$-CIs based on Theorem 2.6 and subsampling-based CIs in dashed and dotted lines, resp. (Compare also Figure 1.) The gray lines stem from realizations of a null model.*

struction, in real data the tails can be expected to behave differently. However, the basic characteristics of the conditional tail index $\xi(r)$ turn out to be the same for negative and positive log returns in our transaction datasets. Note that this is contrary to larger scale return data (e.g., daily data), for which the negative returns usually exhibit heavier tails than the positive returns.

Figure 4 exhibits that the tail index is significantly increased for small values of $r$ (from $-5$ to $+10$ minutes). The confidence intervals indicate the precision of the estimates. Qualitatively the same results can be shown for most of the other stocks of the German stock index that have sufficiently long records. The increase of the tail index $\xi(r,u)$ at the origin is not completely symmetric, the decay for $r > 0$ (conditioning on the future) is slower than for negative values of $r$. Assuming that causal influence can only be carried out by past events, this might sound counter-intuitive at first sight. Though, transactions are generally clustered, which causes a large overlap between the data that enter into $\hat{\xi}^n(r,u)$ and those entering $\hat{\xi}^n(-r,u)$. Furthermore, an extreme log return possibly induces further immediate transactions due to reactions of other market participants. Hence, for small positive values of $r$, $\hat{\xi}^n(r,u)$ might be even larger than $\hat{\xi}^n(-r,u)$, although there is no causal influence from future transactions to current log returns. The gray curves in Figure 4 stem from applying the estimator to multiple realizations of the null model. The fact that the estimated curve for the original data projects beyond the range of the null model curves confirms that mark-location interactions w.r.t. the tail index exist.

Another observation that can be made from Figure 4 and that also holds true for the other German stock index datasets, is that the model-based confidence intervals approach the subsampling-based intervals as the threshold

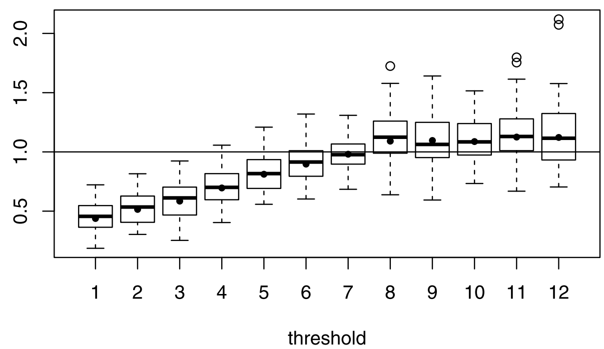


*Figure 5. Ratio of model-based standard deviation to subsampling standard deviation for different thresholds: The $i$-th boxplot corresponds to the $\left(1-0.05\cdot(\frac{1}{2})^{i-1}\right)$-quantile. Each boxplot contains the values for the 12 largest datasets ($\geq$500,000 transactions).*

increases to a sufficiently high level. Figure 5 shows the ratio of average model-based standard deviation to average subsampling standard deviation, averaged over all distances $r$ and upper and lower tails. Each boxplot represents a different threshold and contains the values for the 12 largest datasets ($\geq$500,000 transactions). For low thresholds, the true variance, represented by the subsampling variance, is substantially underestimated by the model-based variance. For large thresholds, the ratio is slightly larger than 1 in average, thus introducing a bit of conservatism in the resulting confidence intervals. In summary, the asymptotic confidence intervals derived in Section 2.4, based on the assumption (GRFM-trans), work reasonably well for our transaction datasets and yield reliable results for thresholds above the 99.5%-quantile.

## 5. DISCUSSION

Irregularly spaced financial data, particularly log returns between consecutive transactions of electronically traded assets, can naturally be perceived as MPPs, which are therefore well-established in financial and econometric literature. At the same time, modeling extreme financial events is of pivotal interest, for example, for insurance or risk management purposes. This paper tries to bring together these two concepts.

Based on existing second-order moment measures for marked point processes, an MPP analog of the extreme value index (tail index) is proposed as well as conditional versions thereof, to detect whether the tail behavior of a mark depends on the point pattern in its "neighborhood". MPP analogs for other summary statistics of (multivariate) extreme value distributions can be defined in a similar way. Conditional versions of extremal coefficients, for instance, allow for detection of interaction of multivariate mark distributions with the pattern of point locations and can also help to detect whether the sampling of a continuous-space process is independent of the process itself or not. However, multivariate extreme value parameters, and in particular their MPP analogs, are difficult to estimate consistently. Typically, estimators in that context are non-stable.

Since the above summary statistics are defined as mean values of certain mark functionals, the question arises, which mean is actually of interest in a practical situation. If there is an underlying continuous-time process from which the observed values are generated by a random sampling procedure, then the mean of interest will usually be reflected by the temporal average over the whole index space instead of the average over the sampling locations. Weighting procedures can then be used to compensate for the irregular distribution of point locations. Though, the assumption of a continuous-time background process seems to be problematic in the framework of financial transaction data since the observed values interact with each other and with the point pattern. Malinowski, Schlather and Zhang [33] suggest to proceed differently for different scales, i.e., to consider the data as a genuine MPP on the very small scale, but to assume an underlying random field on larger scales and to correct for the irregular distribution of locations by a weighting procedure based on the idea of variance minimization. This is closely related to including non-ergodicity into the model and to replace expectation functionals w.r.t. the point process by two-step expectations that average within each ergodicity class first and then aggregate the different classes.

When being faced with real data, assuming ergodicity or not is entirely discretionary since there will always be the restriction to finite observation windows. As regards financial transaction data, it might be sensible to perceive the data as a concatenation of multiple realizations of a possibly non-ergodic MPP. But also for realizations of an ergodic process, employing the estimators derived for the non-ergodic set-up can improve the statistical properties of the estimate.

By applying the conditional tail index estimator to real transaction data, it is shown that the tail index of inter-transaction log returns is significantly increased if there are other transactions close-by. Finite sample properties of the respective estimators, in particular the variability, have been assessed by exploiting that, due to the thresholding, consecutive events that exceed the threshold become stochastically independent under some weak assumptions. The variance estimates based on the assumption (GRFM-trans) and the subsampling-based variance estimates turn out to coincide for sufficiently high thresholds. Though, for general processes, variance estimates of $\hat{\xi}^{n,\mathrm{wght}}(I,u,\mathbf{w})$ based on the assumptions (Independent-noise-marking) or (GRFM-trans) can be highly biased.

The detection of an increase of risk caused by the existence of other transactions, might by itself be a valuable finding for risk management purposes or automated trading algorithms—referring to a very fine temporal scale. Yet, it also indicates that treating this type of data as measurements of a continuous-time process might be suboptimal because this does not capture physical interaction between the observed events.

## ACKNOWLEDGMENT

The data used within this study were provided by the 'Karlsruher Kapitalmarktdatenbank' (KKMDB). A. Malinowski has been financially supported the German Science Foundation (DFG), Research Training Group 1644 'Scaling problems in Statistics'.

## APPENDIX A. MPP THEORY AND PROOFS

Formally, an MPP $\Phi$ on $\mathbb{R}$ with real-valued marks is a measurable mapping

$$\Phi : (\Omega, \mathcal{A}, P) \to (\mathbb{M}_0, \mathcal{M}_0),$$

where $\mathbb{M}_0 = \mathbb{M}_0(\mathbb{R}\times\mathbb{R})$ is the set of all counting measures $\varphi$ on $\mathbb{R}\times\mathbb{R}$ with $\varphi(\,\cdot\,\times\mathbb{R})$ being locally finite, and $\mathcal{M}_0 = \mathcal{M}_0(\mathbb{R}\times\mathbb{R})$ is the smallest $\sigma$-algebra making all mappings $\Phi_{B\times L} : \mathbb{M}_0 \to \mathbb{N}_0$, $\varphi \mapsto \varphi(B\times L)$, $B, L \in \mathcal{B}(\mathbb{R})$, measurable, i.e.

$$\mathcal{M}_0 = \sigma\left(\{\Phi^{-1}_{B\times L}(k) : k\in\mathbb{N}_0, B, L \in \mathcal{B}(\mathbb{R})\}\right).$$

In order to condition on the location of all transactions and on exceeding the threshold $u$, we define the following sub-$\sigma$-algebra: Let $\mathcal{I}_u = \{\varnothing, [u,\infty), [u,\infty)^c, \mathbb{R}\} \subset \mathcal{B}(\mathbb{R})$ be the $\sigma$-algebra generated by the interval $[u,\infty)$ and let $\mathcal{M}^*$ be the smallest $\sigma$-algebra making all mappings $\Phi_{B\times L} : \mathbb{M}_0 \to \mathbb{N}_0$, $\varphi \mapsto \varphi(B\times L)$, $B \in \mathcal{B}(\mathbb{R}), L \in \mathcal{I}_u$, measurable, i.e.

$$\mathcal{M}^* = \sigma\left(\{\Phi^{-1}_{B\times L}(k) : k\in\mathbb{N}_0, B\in\mathcal{B}(\mathbb{R}), L \in \mathcal{I}_u\}\right).$$

Furthermore, let $\mathcal{A}_u^* = \Phi^{-1}(\mathcal{M}^*) \subset \mathcal{A}$.

Since $\sigma(\Phi^{-1}(\mathcal{E})) = \Phi^{-1}(\sigma(\mathcal{E}))$ holds true for any subset $\mathcal{E} \subset \mathcal{M}_0$, it is

$$\begin{aligned}
&\mathcal{A}_u^* = \Phi^{-1}(\mathcal{M}^*) \\
&= \sigma(\Phi^{-1}(\{\Phi_{B\times L}^{-1}(k) : k \in \mathbb{N}_0, B \in \mathcal{B}(\mathbb{R}), L \in \mathcal{I}_u\})) \\
&= \sigma\Big(\big\{\{\omega : \Phi(\omega)(B\times L) = k\} : k \in \mathbb{N}_0, B \in \mathcal{B}(\mathbb{R}), L \in \mathcal{I}_u\big\}\Big)
\end{aligned}$$

and

$$\begin{aligned}
&\Phi_{\mathrm{g}}^{-1}(\mathcal{M}_0(\mathbb{R})) = \sigma(\Phi_{\mathrm{g}}^{-1}(\{\Phi_B^{-1}(k) : k \in \mathbb{N}_0, B \in \mathcal{B}(\mathbb{R})\})) \\
&= \sigma\Big(\big\{\{\omega \in \Omega : \Phi(\omega)(B\times\mathbb{R}) = k\} : k \in \mathbb{N}_0, B \in \mathcal{B}(\mathbb{R})\big\}\Big).
\end{aligned}$$

Thus, $\Phi_{\mathrm{g}}^{-1}(\mathcal{M}_0(\mathbb{R})) \subset \mathcal{A}_u^*$ and the ground process $\Phi_{\mathrm{g}}$ is $(\mathcal{A}_u^*, \mathcal{M}_0(\mathbb{R}))$-measurable. By similar arguments, also the $\mathbb{N}_0$-valued random variable $\hat{\alpha}_{1,h_u}(I,\Phi,T)$ is $\mathcal{A}_u^*$-measurable.

The following lemma shows in which way the Gaussian dependence structure of the underlying random field enters into the proof of Theorem 2.4.

**Lemma A.1.** *A random field $Y$ on $\mathbb{R}^d$ as in assumption (GRFM-trans) has the following property: For all $t, s \in \mathbb{R}^d$, $t \neq s$, the conditional distribution*

$$\begin{aligned}
&F_u(x,y) \\
&= \mathbb{P}\big[\log Y(t) \le x, \log Y(s) \le y \,|\, \log Y(t) > u, \log Y(s) > u\big]
\end{aligned}$$

*becomes a product distribution in the limit $u \to \infty$.*

*Proof.* Follows directly from Juri and Wüthrich [27, Thm. 5.3] and the fact that a copula is invariant under monotone transformation of the margins. □

*Proof of Theorem 2.4 and extension.* With regard to the proof of Theorem 2.6, the more general case of $\frac{\hat{\alpha}_{f_u,h_u}(I,\Phi,T)}{\hat{\alpha}_{1,h_u}(I,\Phi,T)^\ell}$ with $\ell \ge 0$ is considered.

With $e(u) = \mathbb{E}\left[\log Y(0) - u | \log Y(0) > u\right]$ it is

(11)

$$\begin{aligned}
&\mathbb{E}\Big[\frac{\hat{\alpha}_{f_u,h_u}(I,\Phi,T)}{\hat{\alpha}_{1,h_u}(I,\Phi,T)^\ell}\,\Big|\,\mathcal{A}_u^*\Big] \\
&= \hat{\alpha}_{1,h_u}(I,\Phi,T)^{-\ell} \\
&\quad\cdot \mathbb{E}\left[\sum_{(t_1,y_1),(t_2,y_2)\in\Phi} (\log y_1 - u)\mathbf{1}_{\log y_1 > u}\mathbf{1}_{(t_1,t_2)\in C(T,I)}\,\middle|\,\mathcal{A}_u^*\right] \\
&= \hat{\alpha}_{1,h_u}(I,\Phi,T)^{-\ell} \\
&\qquad \cdot \textstyle\sum_{t_1\in\Phi_{\mathrm{g}}\cap[0,T]} \mathbf{1}_{\log Y(t_1)>u} \cdot \#\{t_2 \in \Phi_{\mathrm{g}} : t_2 - t_1 \in I\} \\
&\qquad\qquad \cdot \mathbb{E}\left[(\log Y(t_1) - u) | \mathcal{A}_u^*\right] \\
&= e(u) \cdot \hat{\alpha}_{1,h_u}(I,\Phi,T)^{1-\ell}.
\end{aligned}$$

Furthermore,

$$\begin{aligned}
(12)\quad &\mathrm{Var}\Big[\frac{\hat{\alpha}_{f_u,h_u}(I,\Phi,T)}{\hat{\alpha}_{1,h_u}(I,\Phi,T)^\ell}\,\Big|\,\mathcal{A}_u^*\Big] \\
&= \mathbb{E}\big[(\hat{\alpha}_{f_u,h_u}/\hat{\alpha}_{1,h_u}^\ell)^2 \,|\, \mathcal{A}_u^*\big] - \big(\mathbb{E}[\hat{\alpha}_{f_u,h_u}/\hat{\alpha}_{1,h_u}^\ell \,|\, \mathcal{A}_u^*]\big)^2 \\
&= \hat{\alpha}_{1,h_u}^{-2\ell} \cdot \mathbb{E}\big[\hat{\alpha}_{f_u,h_u}^2 \,|\, \mathcal{A}_u^*\big] - e(u)^2\hat{\alpha}_{1,h_u}^{2-2\ell}
\end{aligned}$$

with

(13)

$$\begin{aligned}
&\mathbb{E}\big[\hat{\alpha}_{f_u,h_u}(I,\Phi,T)^2 \,|\, \mathcal{A}_u^*\big] \\
&= \mathbb{E}\Big[\sum_{t_1\in\Phi_{\mathrm{g}}\cap[0,T]}\ \sum_{s_1\in\Phi_{\mathrm{g}}\cap[0,T]} (\log Y(t_1) - u)(\log Y(s_1) - u) \\
&\qquad \cdot \mathbf{1}_{\log Y(t_1)>u}\mathbf{1}_{\log Y(s_1)>u} \cdot \#\{t_2 \in \Phi_{\mathrm{g}} : t_2 - t_1 \in I\} \\
&\qquad\qquad \cdot \#\{s_2 \in \Phi_{\mathrm{g}} : s_2 - s_1 \in I\}\,\Big|\,\mathcal{A}_u^*\Big] \\
&= \sum_{t_1\in\Phi_{\mathrm{g}}\cap[0,T]}\ \sum_{s_1\in\Phi_{\mathrm{g}}\cap[0,T]} n(t_1,\Phi_{\mathrm{g}},I,u)n(s_1,\Phi_{\mathrm{g}},I,u) \\
&\qquad \cdot \mathbb{E}\big[(\log Y(t_1) - u)(\log Y(s_1) - u) \,|\, \mathcal{A}_u^*\big] \\
&= \sum_{t_1\in\Phi_{\mathrm{g}}\cap[0,T]}\ \sum_{s_1\in\Phi_{\mathrm{g}}\cap[0,T]} n(t_1,\Phi_{\mathrm{g}},I,u)n(s_1,\Phi_{\mathrm{g}},I,u) \\
&\qquad \cdot \Big[\mathbb{E}\left[\log Y(0) - u|\mathcal{A}_u^*\right]^2 + \mathrm{Cov}\big[\log Y(t_1), \log Y(s_1) \,|\, \mathcal{A}_u^*\big]\Big] \\
&= \sum_{t_1\in\Phi_{\mathrm{g}}\cap[0,T]}\ \sum_{s_1\in\Phi_{\mathrm{g}}\cap[0,T]} n(t_1,\Phi_{\mathrm{g}},I,u)n(s_1,\Phi_{\mathrm{g}},I,u) \\
&\qquad \cdot \mathrm{Cov}\big[\log Y(t_1), \log Y(s_1) \,|\, \log Y(t_1) > u, \log Y(s_1) > u\big] \\
&\qquad + e(u)^2 \cdot \hat{\alpha}_{1,h_u}(I,\Phi,T)^2.
\end{aligned}$$

Due to the finite range $h_0$ of the covariance function of $Y$ and the minimum distance $d_0$ between point locations

(14)

$$\begin{aligned}
&\mathbb{E}\big[\hat{\alpha}_{f_u,h_u}(I,\Phi,T)^2 \,|\, \mathcal{A}_u^*\big] \\
&\quad = v_u \sum_{t_1\in\Phi_{\mathrm{g}}\cap[0,T]} n(t_1,\Phi_{\mathrm{g}},I,u)^2 + e(u)^2\hat{\alpha}_{1,h_u}(I,\Phi,T)^2 \\
&\qquad + \varepsilon_u v_u \hat{\alpha}_{1,h_u}(I,\Phi,T)\frac{h_0}{d_0}
\end{aligned}$$

for some $\mathcal{A}_u^*$-measurable random variable $\varepsilon_u$ which is less than one in absolute value. It follows directly from Lemma A.1 that the conditional covariance terms for $t_1 \neq s_1$ vanish for $u \to \infty$. Hence, $\varepsilon_u \to 0$ a.s. and since $\varepsilon_u$ is dominated by 1, also $\mathbb{E}|\varepsilon_u| \to 0$, as $u \to \infty$. Note that, if additionally condition (Independent-noise-marking) is satisfied, all covariance terms in (13) vanish due to the iid assumption, up to those for which $t_1 = s_1$. Hence, $\varepsilon_u$ equals 0 in this case.

Plugging (14) into (12) yields

$$\begin{aligned}
(15)\qquad &\mathrm{Var}\big[\hat{\alpha}_{f_u,h_u}(I,\Phi,T)/\hat{\alpha}_{1,h_u}(I,\Phi,T)^\ell \,|\, \mathcal{A}_u^*\big] \\
&= v_u \cdot \frac{\sum_{t_1\in\Phi_{\mathrm{g}}\cap[0,T]} n(t_1,\Phi_{\mathrm{g}},I,u)^2}{\hat{\alpha}_{1,h_u}(I,\Phi,T)^{2\ell}} \\
&\qquad + \varepsilon_u v_u \hat{\alpha}_{1,h_u}(I,\Phi,T)^{1-2\ell}\frac{h_0}{d_0}
\end{aligned}$$

and the proof is complete. □

*Proof of Theorem 2.6.* Since the marginals of the underlying random field $Y$ are assumed to be in the Fréchet MDA, $\log Y(0)$ is in the Gumbel MDA. It is well known that being in the MDA of the generalized extreme value distribution with shape parameter $\xi$ is equivalent to the distribution of excesses over high thresholds converging to the generalized Pareto distribution (GPD) with the same shape parameter $\xi$ [e.g., 14, Thm. 3.4.13]. In the Gumbel case, where $\xi$ equals 0, the corresponding GPD reduces to the exponential distribution and therefore all moments of the excesses $Z_i = [\log Y(0) - u \mid \log Y(0) > u]$ exist and converge to some constant in $(0, \infty)$ as $u \to \infty$.

Then, application of [33, Thm. 1] yields the weak convergence of $\hat{\alpha}^*_{f_{u_T}, h_{u_T}}(I, \Phi, T)/\sqrt{\hat{\alpha}_{1,h_{u_T}}(I, \Phi, T)}$ to a centered Gaussian variable.

For the asymptotic variance of

$$\frac{\hat{\alpha}^*_{f_u,h_u}(I, \Phi, T)}{\hat{\alpha}_{1,h_u}(I, \Phi, T)^\ell}, \quad \ell \in \{0.5, 1\},$$

note that under the random field model assumption, $\xi(I, u) = e(u)$. Hence, applying the decomposition of variance w.r.t. $\mathcal{A}^*_u$ and replacing $f_u(y) = \log y - u$ by $\tilde{f}_u(y) = \log y - u - e(u)$ in (11)–(15), the terms $e(u)$ in these equations vanish and it follows directly that

$$\begin{aligned} (16)\qquad \operatorname{Var}\left[\frac{\hat{\alpha}^*_{f_u,h_u}(I, \Phi, T)}{\hat{\alpha}_{1,h_u}(I, \Phi, T)^\ell}\right] &= \operatorname{Var}\left[\frac{\hat{\alpha}_{f_u,h_u}(I, \Phi, T)}{\hat{\alpha}_{1,h_u}(I, \Phi, T)^\ell}\right] \\ &\quad - \operatorname{Var}\left[e(u) \cdot \hat{\alpha}_{1,h_u}(I, \Phi, T)^{1-\ell}\right], \end{aligned}$$

for arbitrary $u$ and $T$.

With (11),

$$\begin{aligned} \operatorname{Var}\mathbb{E}\left[\frac{\hat{\alpha}_{f_u,h_u}(I, \Phi, T)}{\hat{\alpha}_{1,h_u}(I, \Phi, T)^\ell} \,\Big|\, \mathcal{A}^*_u\right] &= \operatorname{Var}\left[e(u)\hat{\alpha}_{1,h_u}(I, \Phi, T)^{1-\ell}\right], \end{aligned}$$

which equals 0 if and only if $\ell = 1$. Together with (15), one obtains

$$\begin{aligned} (17)\qquad \operatorname{Var}\left[\frac{\hat{\alpha}_{f_u,h_u}(I, \Phi, T)}{\hat{\alpha}_{1,h_u}(I, \Phi, T)^\ell}\right] &= \mathbb{E}\operatorname{Var}\left[\frac{\hat{\alpha}_{f_u,h_u}(I, \Phi, T)}{\hat{\alpha}_{1,h_u}(I, \Phi, T)^\ell} \,\Big|\, \mathcal{A}^*_u\right] \\ &\quad + \operatorname{Var}\mathbb{E}\left[\frac{\hat{\alpha}_{f_u,h_u}(I, \Phi, T)}{\hat{\alpha}_{1,h_u}(I, \Phi, T)^\ell} \,\Big|\, \mathcal{A}^*_u\right] \\ &= v_u \mathbb{E}\left[\frac{\sum_{t_1 \in \Phi_g \cap [0,\, T]} n(t_1, \Phi_g, I, u)^2}{\hat{\alpha}_{1,h_u}(I, \Phi, T)^{2\ell}}\right] \\ &\quad + \mathbb{E}\left[\varepsilon_u v_u \hat{\alpha}_{1,h_u}(I, \Phi, T)^{1-2\ell}\frac{h_0}{d_0}\right] \\ &\quad + e(u)^2 \operatorname{Var}\left[\hat{\alpha}_{1,h_u}(I, \Phi, T)^{1-\ell}\right] \end{aligned}$$

with some random function $\varepsilon_u$ satisfying $|\varepsilon_u| \le 1$ and $\varepsilon_u \to 0$ a.s. and in $L_1$ for $u \to \infty$ (see proof of Theorem 2.4).

Then, with (16), (17) and $\ell = 0.5$, the asymptotic variance of $\frac{\hat{\alpha}^*_{f_{u_T}, h_{u_T}}(I,\Phi,T)}{\sqrt{\hat{\alpha}_{1,h_{u_T}}(I,\Phi,T)}}$ (for $T \to \infty$) is

$$\begin{aligned} s_{u_\infty} = \lim_{T\to\infty} v_{u_T} &\\ \cdot\, \mathbb{E}&\left[\frac{\sum_{t_1 \in \Phi_g \cap [0,\, T]} n(t_1, \Phi_g, I, u_T)^2}{\hat{\alpha}_{1,h_{u_T}}(I, \Phi, T)} + \varepsilon_{u_T}\frac{h_0}{d_0}\right], \end{aligned}$$

where the expectation can also be applied to numerator and denominator separately due to the a.s. convergence of $\hat{\alpha}_{1,h_{u_T}}(I, \Phi, T)/\mathbb{E}\hat{\alpha}_{1,h_{u_T}}(I, \Phi, T)$ [cf. 33]. For fixed $u$, the asymptotic variance of $\frac{\hat{\alpha}^*_{f_u, h_u}(I,\Phi,T)}{\sqrt{\hat{\alpha}_{1,h_u}(I,\Phi,T)}}$ is

$$\begin{aligned} (18)\qquad s_u &= \lim_{T\to\infty} v_u \left[\mathbb{E}\left[\frac{\sum_{t_1 \in \Phi_g \cap [0,\, T]} n(t_1, \Phi_g, I, u)^2}{\hat{\alpha}_{1,h_u}(I, \Phi, T)} + \varepsilon_u \frac{h_0}{d_0}\right]\right] \\ &= v_u\left[\frac{\mathbb{E}\sum_{t_1 \in \Phi_g \cap [0,\, 1]} n(t_1, \Phi_g, I, u)^2}{\mathbb{E}\hat{\alpha}_{1,h_u}(I, \Phi, 1)} + \frac{h_0}{d_0}\mathbb{E}\varepsilon_u\right], \end{aligned}$$

where the second equation follows by again applying the pointwise ergodic theorem to both numerator and denominator, and by noting that the ratio on the RHS of (18) is bounded by $(|I|/d_0)^2$.

The last assertion concerning the variance of $\frac{\hat{\alpha}_{f_u, h_u}(I,\Phi,T)}{\hat{\alpha}_{1,h_u}(I,\Phi,T)}$ and $\frac{\hat{\alpha}^*_{f_u, h_u}(I,\Phi,T)}{\hat{\alpha}_{1,h_u}(I,\Phi,T)}$ also follows from (16) and (17) with $\ell = 1$. ☐

*Received 29 September 2014*

## REFERENCES


[1] Aït-Sahalia, Y., Mykland, P. A. and Zhang, L. (2005). How often to sample a continuous-time process in the presence of market microstructure noise. *Review of Financial Studies* **18** 351–416.

[2] Andreou, E. and Ghysels, E. (2009). Structural breaks in financial time series. In *Handbook of Financial Time Series* (T. Mikosch, J. P. Kreiss, R. Davis and T. Andersen, eds.) 37, 839–870. Springer, Berlin, Heidelberg.

[3] Bauwens, L. and Hautsch, N. (2009). Modelling financial high frequency data using point processes. In *Handbook of Financial Time Series* (T. G. Andersen, R. A. Davis, J. P. Kreiss and T. Mikosch, eds.) Springer, Berlin.

[4] Bowsher, C. G. (2007). Modelling security market events in continuous time: intensity based, multivariate point process models. *Journal of Econometrics* **141** 876–912. MR2413490

[5] Cartea, A. and Jaimungal, S. (2013). Modeling asset prices for algorithmic and high frequency trading. *Applied Mathematical Finance* **20**. MR3169865

[6] Centanni, S. and Minozzo, M. (2006). A Monte Carlo approach to filtering for a class of marked doubly stochastic poisson processes. *Journal of the American Statistical Association* **101** 1582–1597. MR2279481

[7] Cosslett, S. R. and Lee, L. F. (1985). Serial correlation in latent discrete variable models. *Journal of Econometrics* **27** 79–97.

[8] Daley, D. J. and Vere-Jones, D. (2003). *An Introduction to the Theory of Point Processes. Vol. I: Elementary Theory and Methods*, 2nd ed. *Probability and its Applications*. Springer, New York. MR1950431
[9] Daley, D. J. and Vere-Jones, D. (2008). *An Introduction to the Theory of Point Processes. Vol. II: General Theory and Structure*, 2nd ed. *Probability and its Applications*. Springer, New York. MR2371524
[10] de Haan, L. and Ferreira, A. (2006). *Extreme Value Theory. Springer Series in Operations Research and Financial Engineering*. Springer, New York. MR2234156
[11] Diggle, P. J., Menezes, R. and Su, T. (2010). Geostatistical inference under preferential sampling. *Journal of the Royal Statistical Society: Series C* **59** 191–232. MR2744471
[12] Duval, C. and Hoffmann, M. (2011). Statistical inference across time scales. *Electronic Journal of Statistics* **5** 2004–2030. MR2870155
[13] Embrechts, P. (2000). *Extremes and Integrated Risk Management*. Risk Books.
[14] Embrechts, P., Mikosch, T. and Klüppelberg, C. (1997). *Modelling Extremal Events: for Insurance and Finance (Stochastic Modelling and Applied Probability)*. Springer, London. MR1458613
[15] Engle, R. F. (2000). The econometrics of ultra-high-frequency data. *Econometrica* **68** 1–22.
[16] Engle, R. F. and Lunde, A. (2003). Trades and quotes: a bivariate point process. *Journal of Financial Econometrics* **1** 159–188.
[17] Engle, R. F. and Russell, J. R. (1998). Autoregressive conditional duration: a new model for irregularly spaced transaction data. *Econometrica* **66** 1127–1162. MR1639411
[18] Goegebeur, Y. and De Wet, T. (2012). Local estimation of the second-order parameter in extreme value statistics and local unbiased estimation of the tail index. *Communications in Statistics – Theory and Methods* **41** 3575–3607. MR3003842
[19] Goldfeld, S. M. and Quandt, R. E. (1973). A Markov model for switching regressions. *Journal of Econometrics* **1** 3–15.
[20] Hamilton, J. D. (1993). *Time-Series Analysis*, 1 ed. Princeton Univerity Press. MR1278033
[21] Hamilton, J. D. (2008). Regime-switching models. In *The New Palgrave Dictionary of Economics* (S. N. Durlauf and L. E. Blume, eds.) Palgrave Macmillan, Basingstoke.
[22] Hansen, P. R. and Lunde, A. (2006). Realized variance and market microstructure noise. *Journal of Business & Economic Statistics* **24** 127–161. MR2234447
[23] Hautsch, N. (2012). *Econometrics of Financial High-Frequency Data*. Springer, Berlin. MR2848661
[24] Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution. *The Annals of Statistics* **3** 1163–1174. MR0378204
[25] Ho, L. P. and Stoyan, D. (2008). Modelling marked point patterns by intensity-marked Cox processes. *Statistics & Probability Letters* **78** 1194–1199. MR2441462
[26] Isaaks, E. H. and Srivastava, R. M. (1989). *Applied Geostatistics*. Oxford University Press, New York.
[27] Juri, A. and Wüthrich, M. V. (2003). Tail dependence from a distributional point of view. *Extremes* **6** 213–246. MR2081852
[28] Karr, A. F. (1991). *Point Processes and their Statistical Inference*. Marcel Dekker, Inc, New York. MR1113698
[29] Lando, D. (1998). On Cox processes and credit risky securities. *Review of Derivatives Research* **2** 99–120.
[30] Lange, T. and Rahbek, A. (2009). An introduction to regime switching time series models. In *Handbook of Financial Time Series* (T. Mikosch, J. P. Kreiss, R. Davis and T. Andersen, eds.) 38, 871–887. Springer, Berlin, Heidelberg.
[31] Laurini, F. and Tawn, J. A. (2008). Regular variation and extremal dependence of GARCH residuals with application to market risk measures. *Econometric Reviews* **28** 146–169. MR2655622
[32] Ledford, A. W. and Tawn, J. A. (2003). Diagnostics for dependence within time series extremes. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* **65** 521–543. MR1983762
[33] Malinowski, A., Schlather, M. and Zhang, Z. (2014). Intrinsically weighted means and non-ergodic marked point processes. *Annals of the Institute of Statistical Mathematics*. Accepted. DOI:10.1007/s10463-014-0485-6.
[34] McCulloch, R. E. and Tsay, R. S. (2001). Nonlinearity in high-frequency financial data and hierarchical models. *Studies in Nonlinear Dynamics & Econometrics* **5**.
[35] Møller, J. and Waagepetersen, R. P. (2003). *Statistical Inference and Simulation for Spatial Point Processes*. Chapman & Hall/CRC, Boca Raton, FL.
[36] Myllymäki, M. and Penttinen, A. (2009). Conditionally heteroscedastic intensity-dependent marking of log Gaussian Cox processes. *Statistica Neerlandica* **63** 450–473. MR2598981
[37] Piger, J. (2009). Econometrics: models of regime changes. In *Encyclopedia of Complexity and Systems Science* (R. A. Meyers, ed.) 2744–2757. Springer.
[38] Politis, D. N., Romano, J. P. and Wolf, M. (1999). *Subsampling*. Springer, New York. MR1707286
[39] Politis, D. N. and Sherman, M. (2001). Moment estimation for statistics from marked point processes. *Journal of the Royal Statistical Society, Series B* **63** 261–275. MR1841414
[40] Schlather, M., Ribeiro, Jr, P. J. and Diggle, P. J. (2004). Detecting dependence between marks and locations of marked point processes. *Journal of the Royal Statistical Society, Series B* **66** 79–93. MR2035760
[41] Stoyan, D., Kendall, W. S. and Mecke, J. (1995). *Stochastic Geometry and its Applications*, 2nd ed. John Wiley & Sons Ltd, Chichester. MR0895588
[42] Wackernagel, H. (2003). *Multivariate Geostatistics: An Introduction with Applications*. Springer.
[43] Zhang, M. Y., Russell, J. R. and Tsay, R. S. (2001). A nonlinear autoregressive conditional duration model with applications to financial transaction data. *Journal of Econometrics* **104** 179–207. MR1862032

Alexander Malinowski
University Mannheim
Institute for Mathematics
A5 6, 68131 Mannheim
Germany
E-mail address: malinows@math.uni-goettingen.de

Martin Schlather
University Mannheim
Institute for Mathematics
A5 6, 68131 Mannheim
Germany
E-mail address: schlather@math.uni-mannheim.de

Zhengjun Zhang
University of Wisconsin at Madison
Department of Statistics
1300, University Avenue
Madison, WI 53706
USA
E-mail address: zjz@stat.wisc.edu