TOWARDS ESTIMATING EXTREMAL SERIAL DEPENDENCE VIA THE BOOTSTRAPPED EXTREMOGRAM

Richard A. Davis[1], Columbia University
Thomas Mikosch, University of Copenhagen
Ivor Cribben, Columbia University

Abstract. Davis and Mikosch [3] introduced the extremogram as a flexible quantitative tool for measuring various types of extremal dependence in a stationary time series. There we showed some standard statistical properties of the sample extremogram. A major difficulty was the construction of credible confidence bands for the extremogram. In this paper, we employ the stationary bootstrap to overcome this problem. The use of the stationary bootstrap for the extremogram and the resulting interpretations are illustrated with several financial time series.

[1]coordinating author: Department of Statistics, 1255 Amsterdam Avenue, Columbia University, New York, NY 10027, USA; email: rdavis@stat.columbia.edu
Keywords: Extremogram, extremal dependence, stationary bootstrap, financial time series.
Primary JEL Classification Code: C50.

1. Introduction

With the wild swings recently seen in the financial markets and climatic conditions, there has been renewed interest in understanding and modeling extreme events. The extremogram, developed in Davis and Mikosch [3], is a flexible tool that provides a quantitative measure of dependence of extreme events in a stationary time series. In many respects, one can view the extremogram as the extreme-value analog of the autocorrelation function (ACF) of a stationary process. In classical time series modeling the ACF, and its sample counterpart, are the workhorses for measuring and estimating linear dependence in the family of linear time series processes. While the ACF has some use in measuring dependence in non-linear time series models, especially when applied to non-linear functions of the data such as absolute values and squares, it has limited value in assessing dependence between extreme events. On the other hand, the extremogram only considers observations, or groups of observations, which are large.

For a strictly stationary $\mathbb{R}^d$-valued time series (X_t), the *extremogram* is defined for two sets A and B bounded away from zero[2] by

$$\rho_{A,B}(h) = \lim_{x\to\infty} P(x^{-1}X_h \in B \mid x^{-1}X_0 \in A), \quad h = 0, 1, 2, \ldots, \tag{1.1}$$

provided the limit exists. Since A and B are bounded away from zero, the events $\{x^{-1}X_0 \in A\}$ and $\{x^{-1}X_h \in B\}$ are becoming extreme in the sense the probabilities of these events are converging to zero with $x \to \infty$. For the special choice in the $d = 1$ case of $A = B = (1, \infty)$, the extremogram reduces to the (upper) tail dependence coefficient between X_0 and X_h that is often used in extreme value theory and quantitative risk management; see e.g. McNeil et al. [11]. In this case, one is interested in computing the impact of a large value of the time series on a future value h time-lags ahead. With creative choices of A and B, one can investigate interesting sources of extremal dependence that may arise not only in the upper and lower tails, but also in other extreme regions of the sample space and for estimating cross-extremal dependence in a multivariate time series setting. For the latter, see [3] and [7] where the extremogram is extended to the multivariate case and illustrated in some multivariate financial applications.

We would like to emphasize that the extremogram is a *conditional* measure of extremal serial dependence. Therefore it is particularly suited for financial applications where one is often interested in the persistence of a shock (an extremal event on the stock market say) at future instants of time. Another good reason for using the extremogram for financial time series is a statistical one: for large x, the quantities $P(x^{-1}X_h \in B \mid x^{-1}X_0 \in A)$ are rare event probabilities; their non-parametric estimation cannot be based on standard empirical process techniques and requires large sample sizes. Fortunately, long financial time series are available and therefore the study of their extremal serial behavior is not only desirable but also possible. Financial time series often have the (from a statistical point of view) desirable property that they are heavy-tailed, i.e., extreme large and small values are rather pronounced and occur in clusters. The extremogram and its modifications discussed in this paper allow one to give clear quantititave descriptions of the size and persistence of such clusters.

In estimating the extremogram, the limit on x in (1.1) is replaced by a high quantile a_m of the process. Defining a_m as the $(1 - 1/m)$-quantile of the stationary distribution of $|X_t|$ or related

[2] A set C is bounded away from zero if $C \subset \{y : |y| > r\}$ for some $r > 0$.

quantity, the sample extremogram based on the observations $X_1, \ldots, X_n$ is given by

$$\widehat{\rho}_{A,B}(h) = \frac{\sum_{t=1}^{n-h} I_{\{a_m^{-1} X_{t+h} \in B, a_m^{-1} X_t \in A\}}}{\sum_{t=1}^{n} I_{\{a_m^{-1} X_t \in A\}}} . \tag{1.2}$$

In order to have a consistent result, we require $m = m_n \to \infty$ with $m/n \to 0$ as $n \to \infty$. There is some flexibility in the choice of a_m; the main condition is that (4.1) must be satisfied. In the univariate case ($d = 1$), the $(1 - 1/m)$-quantiles of $|X_t|$ or X_t, or the minus $1/m$-quantile of X_t are all possible candidates for a_m and the respective limit theory for $\hat{\rho}_{A,B}$ remains unchanged. Under suitable mixing conditions and other distributional assumptions that ensure the limit in (1.1) exists, it was shown in Davis and Mikosch [3] that $\widehat{\rho}_{A,B}(h)$ is asymptotically normal; i.e.,

$$\sqrt{n/m}\,(\widehat{\rho}_{A,B}(h) - \rho_{A,B:m}(h)) \xrightarrow{d} N(0, \sigma^2_{A,B}(h)), \tag{1.3}$$

where

$$\rho_{A,B:m}(h) = P(a_m^{-1} X_h \in B \mid a_m^{-1} X_0 \in A)\,. \tag{1.4}$$

We refer to (1.4) as the pre-asymptotic extremogram (PA-extremogram).

There are several obstacles in directly applying (1.3) for constructing confidence bands for the extremogram:

(i) The asymptotic variance $\sigma^2_{A,B}(h)$ is based on an infinite sum of unknown quantities and typically does not have a closed-form expression.
(ii) Estimating $\sigma^2_{A,B}(h)$ is similar to estimating the asymptotic variance of a sample mean from a time series and is often difficult in practice.
(iii) The PA-extremogram $\rho_{A,B:m}(h)$ cannot always be replaced by its limit $\rho_{A,B}(h)$ in (1.3).

For (i) and (ii), we turn to bootstrap procedures to approximate the distribution of $(\hat{\rho}_{A,B}(h) - \rho_{A,B:m}(h))$. This will allow us to construct *credible* (asymptotically correct) confidence bands for the PA-extremogram. As for (iii), a non-parametric bootstrap does not allow us to overcome the bias concern. We note, however, that the PA-extremogram is a conditional probability that is often the quantity of primary interest in applications. That is, one is typically interested in estimating conditional probabilities of *extreme* events as a measure of extremal dependence so that it is not necessary, and perhaps not even desirable to replace the PA-extremogram with the extremogram in (1.3).

The objective of this paper is to apply the bootstrap to the sample extremogram in order to overcome these limitations. By now, there are many non-parametric bootstrap procedures in the literature that are designed for use with stationary time series. Many of these involve some form of resampling from blocks of observations. That is, in constructing a bootstrap replicate of the time series, long stretches of the time series are stitched together in order to replicate the joint distributions of the process. While for a finite sample size n, it is impossible to replicate all the joint distributions, we can only sample from at most the m-variate distributions (for $m < n$) by sampling blocks of m consecutive observations. In order to obtain consistency of the procedure, m is allowed to grow with n at a suitable rate. In this paper, we adopt the stationary bootstrap approach as described in Politis and Romano [12] in which the block sizes are given by independent geometric random variables. Since the blocks are of random length, the stationary bootstrap is useful as an exploratory device in which dependence beyond a fixed block length can be discovered.

In our case, there are significant differences in the extremogram setting of our bootstrap application from the traditional one. First, the summands in the numerator and denominator of (1.2) form a triangular array of random variables and cannot be cast as a single stationary sequence. Second, most bootstrapping applications in extreme value theory, even in the iid case, require the replicate time series to be of smaller order than the original sample size n; see e.g. Section 6.4 of Resnick [13]. On the other hand, we are able to overcome these drawbacks and show that the bootstrapped sample extremogram, based on the replicates of size n provides an asymptotically correct approximation to the left-hand side of (1.3) provided the blocks grow at a proper rate.

In addition to providing a non-parametric estimate of the nature of extremal dependence as a function of time-lag, the extremogram can also provide valuable guidance in various phases of the typical time series modeling paradigm. For example, the sample extremogram might provide insight into the choice of models for the data with the goal of delivering models that are compatible with the extremal dependence. In the standard approach, models are often selected to fit the center of the distribution and can be inadequate for describing the extremes in the data. On the other hand, if the primary interest is on modeling extremes, then the modeling exercise should focus on this aspect. The quality of fit could be judged by assessing compatibility of the sample extremogram with the fitted model extremogram. Moreover, the sample extremogram from the residuals of the model fit can be used to check compatibility with the lack of extremal dependence.

The remainder of the paper is organized as follows. The theory of the bootstrapped extremogram is given in Section 2 and its use is demonstrated with several financial time series in Section 3. In conjunction with the bootstrapped extremogram, we present a quick and clean method for testing *significant* serial extremal dependence using a random permutation procedure. This procedure is similar in spirit to using the block bootstrap procedure, but with block size equal to 1. In Section 3.4, we connect the empirical extremogram with the distribution of return times between rare events studied in [9]. Consistent with the findings of Geman and Chang the presence of extremal clustering can be detected easily using the bootstrap applied to the extremogram of return times. Finally, background on regular variation and the proof of the main theorems in Section 2 are provided in the Appendix.

2. The Bootstrapped Sample Extremogram

In this section we will construct confidence bands for the sample extremogram based on a bootstrap procedure which takes into account the serial dependence structure of the data. The resulting confidence bands closely follow the sample extremogram. Moreover, assuming regular variation of the underlying time series, we will be able to show that the bootstrap confidence bands are asymptotically correct.

2.1. **Stationary bootstrap.** This resampling scheme was introduced by Politis and Romano [12]. It is an adaptation of the block bootstrap which allows for randomly varying block sizes.

For any strictly stationary sequence (Y_t) the *stationary bootstrap procedure* consists of generating pseudo-samples $Y_1^*,\dots,Y_n^*$ from the sample $Y_1,\dots,Y_n$ by taking the first n elements from

$$Y_{K_1},\dots,Y_{K_1+L_1-1},\dots,Y_{K_N},\dots,Y_{K_N+L_N-1}\,, \tag{2.1}$$

where (K_i) is an iid sequence of random variables uniformly distributed on $\{1,\dots,n\}$, (L_i) is an iid sequence of geometrically distributed random variables with distribution $P(L_1=k)=p\,(1-p)^{k-1}$,

$k = 1, 2, \ldots$, for some $p = p_n \in (0,1)$ such that $p_n \to 0$ as $n \to \infty$, and

$$N = N_n = \inf\{i \geq 1 : L_1 + \cdots + L_i \geq n\}\,.$$

The upper limits of the random blocks $\{K_i, \ldots, K_i + L_i - 1\}$ may exceed the sample size n. Therefore, in (2.1) we replace the (unobserved) Y_t's with $t > n$ by the observations $Y_{t \bmod n}$. Finally, the three sequences (Y_t), (K_i) and (L_i) are also supposed independent. The dependence of the sequences (K_i) and (L_i) on n is suppressed in the notation. The generated pseudo-sample can be extended to an infinite sequence (Y_t^*) by extending (2.1) to an infinite sequence. For every fixed $n \geq 1$, (Y_t^*) constitutes a strictly stationary sequence.

2.2. **Main results.** We want to apply the stationary bootstrap procedure to an array of strictly stationary sequences of indicators functions $I_t = I_{\{a_m^{-1} X_t \in C\}}$, $t \in \mathbb{Z}$, $n \geq 1$, where the underlying sequence (X_t) is strictly stationary $\mathbb{R}^d$-valued and regularly varying with index α (see Appendix for definition), a_m has the interpretation as a high quantile of the distribution of $|X_0|$ or related quantity (see (4.1) for definition) with $m = m_n \to \infty$ as $n \to \infty$, and C is a set bounded away from zero. To simplify notation, we have suppressed the dependence of I_t on n. We will work with a multivariate time series in this section for two reasons. First, the technique we employ is essentially dimensionless. Second, even in the univariate case, it will be necessary to establish bootstrap consistency based on a multivariate time series consisting of the series and its selected lagged components. The application of the stationary bootstrap in this context is rather unconventional since the sequences (I_t) constitute triangular arrays of strictly stationary sequences since they depend on a_m and hence n.

We write (I_t^*) for a bootstrap sequence generated from the sample $I_1, \ldots, I_n$ by the stationary bootstrap procedure described above. In what follows, P^*, E^* and var^* denote the probability measure generated by the bootstrap procedure, the corresponding expected value and variance. This means that $P^*(\cdot) = P(\cdot \mid (X_t))$ is the infinite product measure generated by the distributions of (K_i) and (L_i).

The bootstrap sample mean $\overline{I}_n^* = n^{-1} \sum_{i=1}^n I_i^*$ satisfies the following elementary properties:

$$\begin{aligned} E^*(\overline{I}_n^*) &= E^*(I_1^*) = \overline{I}_n = n^{-1} \sum_{i=1}^n I_i\,, \\ s_n^2 = \operatorname{var}^*(n^{1/2} \overline{I}_n^*) &= C_n(0) + 2 \sum_{h=1}^{n-1} (1 - h/n)\,(1-p)^h\, C_n(h)\,, \end{aligned}$$

where

$$C_n(h) = n^{-1} \sum_{i=1}^n (I_i - \overline{I}_n)(I_{i+h} - \overline{I}_n)\,, \quad h = 0, \ldots, n\,,$$

are the *circular sample autocovariances.* Here we again made use of the circular construction $I_j = I_{j \bmod n}$. Writing

$$\gamma_n(h) = n^{-1} \sum_{i=1}^{n-h} (I_i - \overline{I}_n)(I_{i+h} - \overline{I}_n)\,, \quad h \geq 0\,,$$

for the ordinary *sample autocovariances,* we have

$$C_n(h) = \gamma_n(h) + \gamma_n(n-h)\,, \quad h = 0, \ldots, n\,. \tag{2.2}$$

Recall the notion of a strictly stationary regularly varying sequence from Section 4.1, in particular the sequence of limiting measures (μ_h); see (4.1). For convenience, we write $\mu=\mu_1$. For any subset $C\subset \overline{\mathbb{R}}_0^d$, define the quantities

$$\sigma^2(C)=\mu(C)+2\sum_{h=1}^\infty \tau_h(C)\,, \tag{2.3}$$

with $\tau_h(C)=\mu_{h+1}(C\times \overline{\mathbb{R}}_0^{d(h-1)}\times C)$ and, suppressing the dependence on C in the notation,

$$\begin{aligned}\widehat P_m &= m\,\overline I_n\,,\quad p_0=EI_1=P(a_m^{-1}X_1\in C)\,,\\ p_{0h} &= E(I_0I_h)=P(a_m^{-1}X_0\in C\,,a_m^{-1}X_h\in C)\,,\quad h\ge 1\,.\end{aligned}$$

We also need the following mixing condition:

(M) The sequence (X_n) is strongly mixing with rate function (α_t). Moreover, there exist $m=m_n\to\infty$ and $r_n\to\infty$ such that $m_n/n\to 0$ and $r_n/m_n\to 0$ and

$$\lim_{n\to\infty} m_n\sum_{h=r_n}^\infty \alpha_h=0\,, \tag{2.4}$$

and for all $\epsilon>0$,

$$\lim_{k\to\infty}\limsup_{n\to\infty} m_n\sum_{h=k}^\infty P(|X_h|>\epsilon\,a_m\,,|X_0|>\epsilon\,a_m)=0\,. \tag{2.5}$$

Our next goal is to show that the stationary bootstrap is asymptotically correct for the bootstrapped estimator of $\widehat P_m$ given by

$$\widehat P_m^*=m\,\overline I_n^*=\frac{m}{n}\sum_{t=1}^n I_t^*\,.$$

The following result is the stationary bootstrap analog of Theorem 3.1 in [3]. It shows that the bootstrap estimator $\widehat P_m^*$ of $\widehat P_m$ is asymptotically correct.

Theorem 2.1. *Assume that the following conditions hold for the strictly stationary regularly varying sequence (X_t) of $\mathbb{R}^d$-valued random vectors:*

(1) *The mixing condition* (M) *and in addition*

$$\sum_{h=1}^\infty k\,\alpha_k<\infty\,. \tag{2.6}$$

(2) *The growth conditions*

$$p=p_n\to 0\,,\quad and\quad n\,p^2/m\to\infty. \tag{2.7}$$

(3) *The sets C and $C\times\overline{\mathbb{R}}_0^{d(h-1)}\times C\subset \overline{\mathbb{R}}_0^{d(h+1)}$ are continuity sets with respect to μ and μ_{h+1} for $h\ge 1$, C is bounded away from zero and $\sigma^2(C)>0$.*

(4) *The central limit theorem, $(n/m)^{1/2}\,(\hat P_m-mp_0)\stackrel{d}{\to} N(0,\sigma^2(C))$ holds.*

Then the following bootstrap consistency results hold:

$$E^*(\widehat P_m^*)\ \stackrel{P}{\to}\ \mu(C)\,, \tag{2.8}$$

$$m\,s_n^2=\mathrm{var}^*((n/m)^{1/2}\widehat P_m^*)\ \stackrel{P}{\to}\ \sigma^2(C)\,, \tag{2.9}$$

with $\sigma^2(C)$ *given in* (2.3). *In particular,*

$$P^*(|\widehat{P}_m^* - \mu(C)| > \delta) \stackrel{P}{\to} 0\,, \quad \delta > 0\,, \tag{2.10}$$

and the central limit theorem holds

$$\sup_x \Big| P^*\big((n/m)^{1/2}(m s_n^2)^{-1/2}(\widehat{P}_m^* - \widehat{P}_m) \le x\big) - \Phi(x)\Big| \stackrel{P}{\to} 0\,, \tag{2.11}$$

where Φ *denotes the standard normal distribution function.*

Politis and Romano [12] proved a corresponding result for the sample mean of the stationary bootstrap sequence (X_t^*) for a finite variance strictly stationary sequence (X_t). They also needed the growth conditions $p_n \to 0$ and $n\,p_n \to \infty$. Our additional condition $(n\,p_n)(p_n/m) \to \infty$, which implies $n\,p_n \to \infty$ is needed since $\widehat{P}_m^*$ is an average in the triangular scheme $I_t = I_{\{a_m^{-1} X_t \in C\}}$, $t = 1, \ldots, n$. Although various steps in the proof are similar to those in Politis and Romano [12], the triangular nature of the bootstrapped sequence (I_t) requires some new ideas. We found it surprising that the full stationary bootstrap works in this context. In the context of extreme value statistics the bootstrap often needs to be modified even when the data are iid.

We now turn our attention to the sample extremograms for which both the numerator and denominator are estimators of the type $\widehat{P}_m$. Therefore our next objective are the asymptotic properties of these ratio estimators which we study in a general context. We consider general sets $D_1, \ldots, D_h \subset \overline{\mathbb{R}}_0^d$ and $C = D_{h+1} \subset \overline{\mathbb{R}}_0^d$, $h \ge 1$. Since we deal with several sets D_i we need to indicate that the indicator functions I_t depend on these sets:

$$I_t(D_i) = I_{\{a_m^{-1} X_t \in D_i\}}\,, \quad t \in \mathbb{Z}\,,$$

and we proceed similarly for the estimators $\widehat{P}_m(D_i)$ of $\mu(D_i)$. Now we define the corresponding *ratio estimators*

$$\widehat{\rho}_{C,D_i} = \frac{\widehat{P}_m(D_i)}{\widehat{P}_m(C)} = \frac{\sum_{t=1}^n I_{\{a_m^{-1} X_t \in D_i\}}}{\sum_{t=1}^n I_{\{a_m^{-1} X_t \in C\}}}\,, \quad i = 1, \ldots, h\,.$$

Davis and Mikosch [3], Corollary 3.3, proved the joint asymptotic normality of these estimators: Note that there is a misprint for the expression for r_{D_i,D_j} in [3], which we now correct here as

$$r_{D_i,D_j} = \mu(D_i \cap D_j) + \sum_{h=1}^{\infty} [\mu_{h+1}(D_j \times \overline{\mathbb{R}}_0^{d(h-2)} \times D_i) + \mu_{h+1}(D_i \times \overline{\mathbb{R}}_0^{d(h-2)} \times D_j]\,.$$

The centering in the central limit theorem of Corollary 3.3 of [3] uses the PA-extremogram as opposed to the extremogram. In general the PA-extremograms cannot be replaced by their limits

$$\rho_{C,D_i} = \frac{\mu(D_i)}{\mu(C)}\,, \quad i = 1, \ldots, h\,,$$

unless the following additional condition holds

$$\lim_{n\to\infty} \sqrt{n m_n}\, [\mu(D_i) P(a_m^{-1} X_0 \in C) - \mu(C)\, P(a_m^{-1} X_0 \in D_i)] = 0\,, \quad i = 1, \ldots, h\,. \tag{2.12}$$

In addition to the complex form of the asymptotic variance which can hardly be evaluated, condition (2.12) points at another practical problem when applying the central limit theorem to the ratio estimators. The next result will show that these problems will be overcome by an application of the stationary bootstrap.

We construct bootstrap samples $I_1^*(D_i), \ldots, I_n^*(D_i)$, $i = 1, \ldots, h+1$, from the samples $I_1(D_i), \ldots, I_n(D_i)$, $i = 1, \ldots, h+1$, by a simultaneous application of the stationary bootstrap procedure, i.e., we use the same sequences (K_i) and (L_i) for the construction of the $h+1$ bootstrap samples; see Section 2.1. From the bootstrap samples the bootstrap versions $\widehat{P}_m^*(D_i)$ of $\widehat{P}_m(D_i)$, $i=1,\ldots,h+1$, and $\widehat{\rho}^*_{C,D_i}$ of $\widehat{\rho}_{C,D_i}$, $i=1,\ldots,h$, are constructed.

The following result shows that the bootstrapped ratio estimators $\widehat{\rho}^*_{C,D_i}$ are asymptotically correct estimators of their sample counterparts $\widehat{\rho}_{C,D_i}$, $i=1,\ldots,h$.

Theorem 2.2. *Assume that the following conditions hold for the strictly stationary regularly varying sequence (X_t) of $\mathbb{R}^d$-valued random vectors:*

(1) *The mixing conditions* (M) *and* (2.6) *hold.*
(2) *The growth conditions* (2.7) *on $p=p_n$ hold.*
(3) *The sets $D_1,\ldots,D_{h+1}$, $D_i\times\overline{\mathbb{R}}_0^{d(i-1)}\times D_i\subset\overline{\mathbb{R}}_0^{(i+1)d}$ are continuous with respect to μ and μ_{i+1}, $i=1,\ldots,h+1$, $\mu(C)>0$ and $\sigma^2(D_i)>0$, $i=1,\ldots,h$.*
(4) *The central limit theorem $(n/m)^{1/2}\left(\widehat{\rho}_{C,D_i}-\rho_{C,D_i:m}\right)_{i=1,\ldots,h}\stackrel{d}{\to}N(\mathbf{0},\Sigma)$ holds, where the asymptotic variance is defined in Corollary 3.3 of* [3].

Then the bootstrapped ratio estimators satisfy the following bootstrap consistency result

$$P^*\left(|\widehat{\rho}^*_{C,D_i}-\rho_{C,D_i}|>\delta\right)\stackrel{P}{\to}0\,,\quad \delta>0\,,\tag{2.13}$$

and the central limit theorem holds

$$P^*((n/m)^{1/2}\left(\widehat{\rho}^*_{C,D_i}-\widehat{\rho}_{C,D_i}\right)_{i=1,\ldots,h}\in A)\stackrel{P}{\to}\Phi_{\mathbf{0},\Sigma}(A)\,,\tag{2.14}$$

where A is any continuity set of the normal distribution $\Phi_{\mathbf{0},\Sigma}$ with mean zero and covariance matrix Σ.

2.3. **Consistency for the bootstrapped sample extremogram.** Recall that for observations $X_1,\ldots,X_n$ from a stationary regularly varying time series (X_t), the definition of the sample extremogram $(\widehat{\rho}_{A,B}(i))$ is given in (1.2). This estimate can be recast as a ratio estimator by introducing the $\mathbb{R}^{h+1}$-valued vector process

$$Y_t=(X_t,\ldots,X_{t+h})',\quad t\in\mathbb{Z}\,,\tag{2.15}$$

consisting of stacking $h+1$ consecutive values of the time series (X_t). Now the sets C and $D_0,\ldots,D_h$ specified in Theorem 2.2 are defined through the relations $C=A\times\overline{\mathbb{R}}_0^h$, $D_0=A\cap B\times\overline{\mathbb{R}}_0^h$ and for $D_i=A\times\overline{\mathbb{R}}_0^{i-1}\times B\times\overline{\mathbb{R}}_0^{h-i}$ for $i\ge1$. With this convention, Theorem 2.2 can be applied to the (Y_t) and (Y_t^*) sequences directly. We formulate here the consistency result for the bootstrapped sample extremogram

$$\widehat{\rho}^*_{AB}(i)=\frac{\sum_{t=1}^{n-i}I_{\{a_m^{-1}X_t^*\in A,a_m^{-1}X^*_{t+i}\in B\}}}{\sum_{t=1}^{n}I_{\{a_m^{-1}X_t^*\in A\}}}\,,\quad i\ge0\,.$$

We use the same notation as in Theorem 2.2.

Corollary 2.3. *Assume that the conditions of Theorem 2.2 are satisfied for the sequence (Y_t) and the sets $C,D_0,\ldots,D_h$ defined above. Then, conditionally on (X_t),*

$$(n/m)^{1/2}\left(\widehat{\rho}^*_{AB}(i)-\widehat{\rho}_{AB}(i)\right)_{i=0,1,\ldots,h}\stackrel{d}{\to}N(\mathbf{0},\Sigma)\,.$$

3. Examples of the Bootstrapped Sample Extremogram

3.1. **FTSE.** The first application of the bootstrapped extremogram is to the 6,443 daily log-returns of the FTSE 100 Index from April 4, 1984 to October 2, 2009. The sample extremogram of the FTSE for lags 1 to 40 corresponding to the left tail ($A = B = (-\infty, -1)$) with a_m equal to the negative of the .04 empirical quantile (considered to be fixed) is displayed as the bold line in the left panel of Figure 3.1. Approximate 95% confidence intervals for the PA-extremogram, produced using 10,000 stationary bootstrapped replicates with $p_n = 1/200$, are displayed in this graph as dashed lines. These bounds were found using the .025 and .975 quantiles from the empirical distribution of the bootstrapped replicates of $\hat{\rho}^*_{A,B}(h) - \hat{\rho}_{A,B}(h)$ and the sample extremogram. Notice that due to the bias in the bootstrapped distribution, the sample extremogram does not fall in the center of the intervals. Using a small p_n in the bootstrapped replicates helps reduce this bias. Observe that the solid horizontal line at height .04, corresponding to a PA-extremogram under an independence assumption, is well outside these confidence bands confirming the serial extremal dependence in the lower tail.

Typically, financial returns are modeled via a multiplicative model given by $X_t = \sigma_t Z_t$, where (Z_t) is an iid sequence of mean 0 variance 1 random variables and for each t, the volatility σ_t is independent of Z_t. Most financial time series models such as GARCH and stochastic volatility have this form. With such a model, an extreme value of the process occurs at time t if σ_t is large or if there is a large *shock* in the noise (i.e., Z_t) at time t. After estimating the volatility process (σ_t), the estimated devolatilized process is defined by $\hat{Z}_t = X_t/\hat{\sigma}_t$. If the multiplicative model is a reasonable approximation to the return series, then the devolatilized series should be free of extremal dependence. The right graph in Figure 3.1 shows the sample extremogram (solid line) for the devolatilized process obtained as the residuals in fitting a GARCH(1,1) model and 95% confidence bounds produced using the stationary bootstrap with $p_n = 1/200$. This time the solid horizonal line at height .04, which represents the value of the PA-extremogram under the null of no extremal dependence, falls within these bounds for most lags h. Hence the extremal dependence as measured by the extremogram has been effectively removed in this devolatilization process.

3.2. **A random permutation procedure.** Before considering a second example, we note that permutation procedures can offer a clean and effective alternative to the bootstrap for testing significant extremal dependence. Specifically, a permutation test procedure can be used to produce *confidence bands* for the sample extremogram under the assumption that the underlying data are in fact independent. These bounds can be viewed as the analogue of the standard $\pm 1.96/\sqrt{n}$ bounds used for the sample autocorrelation function (ACF) of a time series. Values of the sample ACF that extend beyond these bounds lend support that the ACF at the corresponding lags are non-zero. The bounds for the sample ACF are based on well-known asymptotic theory. Unfortunately, such bounds are not easily computable for the sample extremogram $\hat{\rho}_{A,B}(h)$. For a fixed lag h, if the value of the sample extremogram for the original data is not extreme relative to the values of the sample extremogram based on random permutations of the data, then the sample extremogram is impervious to the time order of the data. On the other hand, if the $\hat{\rho}_{A,B}(h)$ is more extreme (either larger or smaller than all the extremograms computed for 99 random permutations of the data), then we conclude the presence of extremal dependence at lag h with probability .98=98/100. Aside from boundary effects, (i.e., the numerator of the sample extremogram is a sum over $n - h$

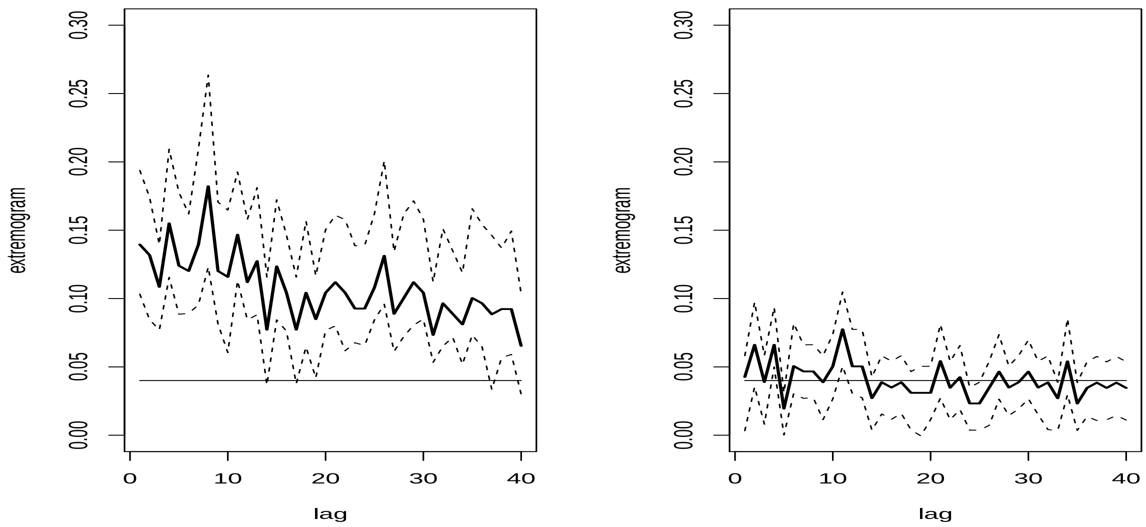

Figure 3.1. Left: *95% confidence bands (dashed lines) for the PA-extremogram of the FTSE based on the stationary bootstrap; sample extremogram (solid line); PA-extremogram (horizontal line at .04) based on the data being independent.* Right: *95% confidence bands (dashed lines) for the PA-extremogram of the residuals from a* GARCH(1, 1) *fit to FTSE based on the stationary bootstrap; sample extremogrom of residuals (solid line); PA-extremogram (horizontal line at .04) based on the data being independent.*

terms and hence depends mildly on h), the permutation distribution of the sample extremogram is virtually the same for all lags h. The bold lines in the graphs of Figure 3.2 correspond to the maximum and minimum of the sample extremogram at lag 1 based on 99 random permutations of the FTSE (left panel) and the residuals from the GARCH fit (right panel). If the data are in fact independent, then the value of the PA-extremogram would be .04 for all lags $h > 0$, which as expected, sits nearly in the middle of the confidence bounds in both graphs. Based on the permutation procedure, we would conclude the existence of extremal dependence of the FTSE data for the first 40 lags and the absence of extremal dependence in the residuals over the same set of lags.

3.3. **General Electric.** The next example illustrates how the stationary bootstrap for the sample extremogram performs as a function on the choice of the mean block size. Recall from Section 2 that the condition $p_n \to 0$ is needed in order to achieve consistency of the bootstrap estimators of the extremogram. Figure 3.3 (top left) shows the sample extremogram of the left tail for the 5-minute log-returns of General Electric (GE) from December 1, 2004 to July 26, 2006. We chose the sets $A = B = (-\infty, -1)$ corresponding to the lower tail and a_m to be the negative of the .01 empirical quantile of the log-returns. Since the interesting lags in this data set occur at 1, 79, 158, we focused our estimates at just these lags in the remaining panels of Figure 3.3. The sample extremogram has a large spike at lags 79 and 158. The New York Stock Exchange (NYSE) is open daily from 9:30am to 4pm. Hence, there are 78 5-minute spells each day and so we can conclude that there is evidence of strong extremal dependence between returns a day apart. In the top right

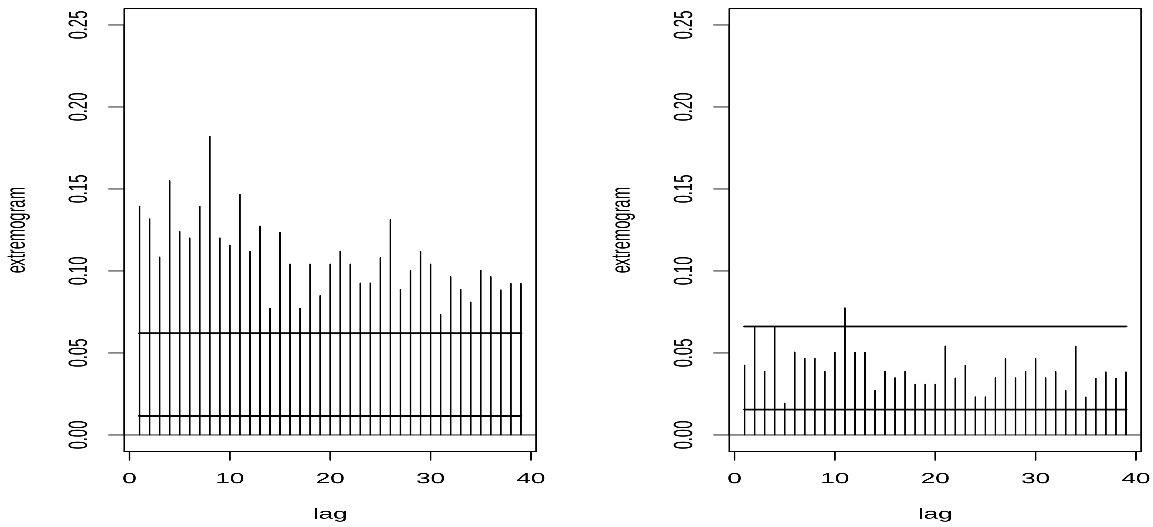

Figure 3.2. Left: *The sample extremogram for the lower tails of the FTSE. Bold horizontal lines represent confidence bounds for the PA-extremogram based on 99 random permutations of the data.* Right: *The sample extremogram for the lower tails of the residuals from a* GARCH(1, 1) *fit to FTSE. Bold horizontal lines represent confidence bounds for the PA-extremogram based on 99 random permutations of the data.*

graph, we have displayed the box-plots of the sample extremogram at lags 1, 79, and 158 based on 10,000 random permutations of the data. In this case, the serial dependence has been destroyed in each replication. The triangles on the plot correspond to the values of the sample extremogram (at lags 1, 79, 158) based on the original data. Note that the triangles are well outside these boxplots signifying extremal dependence in the data. The bottom two graphs in Figure 3.3 contain boxplots of the bootstrapped extremogram corresponding to $p_n = 50$ and $p_n = 200$ respectively. By resampling blocks with mean block size 50, the assumption is made that the dependence in observations X_t and X_{t+k} for $k > 50$ has little, if any, impact on the distribution of the sample extremogram. In particular, the dependence structure is broken for lags greater than 50 and so the bootstrapped sample extremogram cannot capture the extent of the extremal dependence at lag 79 and beyond, and certainly not at lag 158.

The fixed block bootstrap is a method that resamples blocks of data of fixed length in order to keep the dependence structure intact. If such a bootstrap with fixed block size 50 was used, the opportunity to detect the extremal dependence at lags 79 and 158 would be impossible. Thus, the randomly chosen block size is a potential advantage of the stationary bootstrap: it is not unlikely to get blocks of size 90, say, if the mean block size is 50. As the mean block sizes increase from 50 to 200, the stationary bootstrap captures more and more of the dependence at lags 79 and 158.

3.4. **Bootstrapping the extremogram of return times between rare events.** In their presentation [9], Geman and Chang consider the waiting or return times between rare (or extreme) events for financial time series. They define a rare event by a large excursion relative to observed returns: a return X_t is *rare* (or *extreme*) if $X_t \le \xi_p$ or $X_t \ge \xi_{1-p}$, where ξ_q is the q-quantile of the

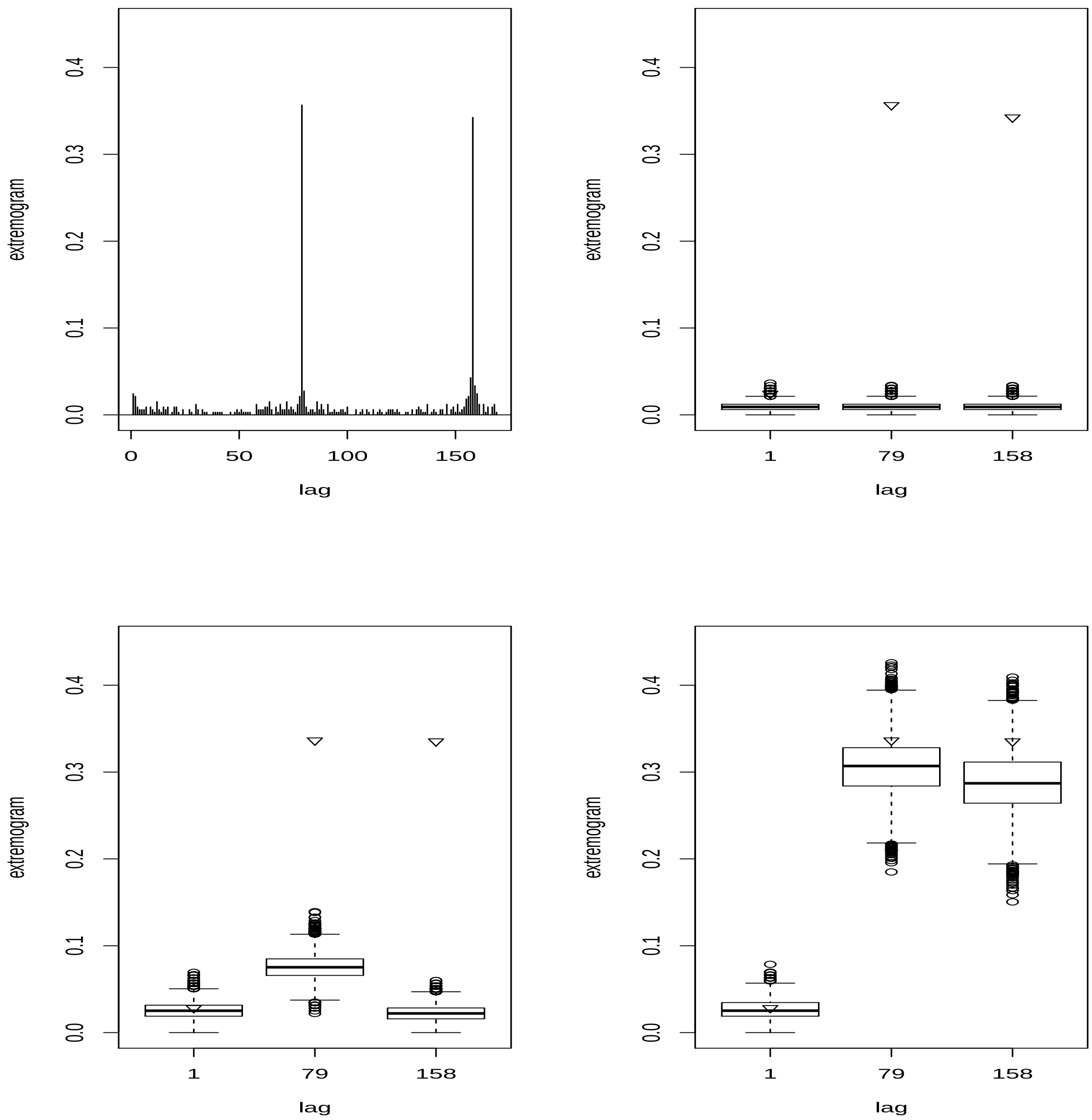

Figure 3.3. *The sample extremogram for the* 5*-minute log-returns of GE* (top left)*. Boxplots of the corresponding replicates of the extremograms at lags 1, 79 and 158, using permutations* (top right)*; bootstrapping with mean block size of* 50 (bottom left) *and bootstrapping with a mean block size of* 200 (bottom right)*.*

distribution of returns. Typical choices for p are 0.1 and 0.05. Denoting occurrences of rare events by the binary sequence

$$W_j = \begin{cases} 1, & \text{if } X_t \text{ is extreme,} \\ 0, & \text{otherwise,} \end{cases} \tag{3.1}$$

Geman and Chang study return times T_j, $j = 1, 2, \ldots$, between successive 1's of the W_j sequence. If the return times were truly iid, the successive waiting times between 1's should be iid geometric.

Using the histogram of waiting times, the geometric assumption can be examined. In order to perform inference and in particular hypothesis testing on the extremal clustering of returns, Geman and Chang [9] calculate the observed entropy of the excursion waiting times and compare it to the entropies of random permutations of the excursion waiting times. If the observed entropy behaves similarly to the random permutations, then one may conclude that the returns do not exhibit extremal clustering.

We now introduce an analog of the extremogram for the return times between rare events in a strictly stationary regularly varying sequence (X_t). Denoting the rare event by $A \subset \overline{\mathbb{R}}_0$, the corresponding return times extremogram is given by

$$(3.2)\quad \rho_A(h) = \lim_{x\to\infty} P(x^{-1}X_1 \notin A, \ldots, x^{-1}X_{h-1} \notin A, x^{-1}X_h \in A \mid x^{-1}X_0 \in A)\,, \quad h \ge 1\,.$$

Using the regular variation of the sequence (X_t) (see Section 4.1) and assuming that A and $A \times \overline{\mathbb{R}}_0^{h-1} \times A$ are continuity sets with respect to μ and μ_{h+1}, $h \ge 1$, and A is bounded away from zero, we can calculate the return times extremogram

$$\rho_A(h) = \frac{\mu_{h+1}(A \times \mathbb{R}^{h-1} \times A)}{\mu(A)}\,, \quad h \ge 0\,.$$

The return times sample extremogram is then defined as

$$(3.3)\quad \widehat{\rho}_A(h) = \frac{\sum_{t=1}^{n-h} I_{\{a_m^{-1}X_{t+h}\in A, a_m^{-1}X_{t+h-1}\notin A, \ldots, a_m^{-1}X_{t+1}\notin A, a_m^{-1}X_t\in A\}}}{\sum_{t=1}^{n} I_{\{a_m^{-1}X_t\in A\}}}\,, \quad h = 1, 2, \ldots, n-1.$$

An asymptotic theory for the return times sample extremogram and its bootstrap version is given at the end of Section 2.3. This theory shows that the stationary bootstrap is asymptotically correct for this sample extremogram.

We will now illustrate an example of the return times of extreme events. The graphs in Figure 3.4 show the histograms for the return times of extreme events for the daily log-returns (May 29, 1986 to October 2, 2009) of Bank of America (BAC) for the rare event corresponding to $A = \mathbb{R}\backslash[\xi_{0.05}, \xi_{0.95}]$. The left panel contains the histogram (solid vertical lines) of the log-returns, whereas the right panel shows the corresponding histograms for the filtered time series after fitting a GARCH(1,1) model to the data; cf. Section 3.1. Since the heights of the histogram correspond exactly to the return times sample extremogram, we can apply the bootstrap procedures of Section 3. The dashed lines that overlay the graphs in Figure 3.4 represent the .975 (upper) and .025 (lower) confidence bands computed from the bootstrap approximation to the sampling distribution of the sample extremogram. The solid curve is the geometric probability mass function with success probability $p = 0.1$. As seen in the left panel, the geometric probability mass function falls outside the confidence bands at nearly every value ranging from 1 to 20. This complements earlier findings of the presence of extremal clustering in the original daily returns. On the other hand, the confidence bands in the right panel contains the geometric probability mass function for all but a few values, which suggests that extremal clustering has been effectively removed in the devolatilized series.

4. Appendix: Regular Variation and Proofs

4.1. **Brief Interlude into Regular Variation.** The extremogram (1.1) is a limit of conditional probabilities and therefore not always defined. In this section we show that regular variation of the time series is sufficient for the existence of the extremogram for any choice of sets A and B bounded away from 0. The condition of regular variation is rather technical; the interested reader is

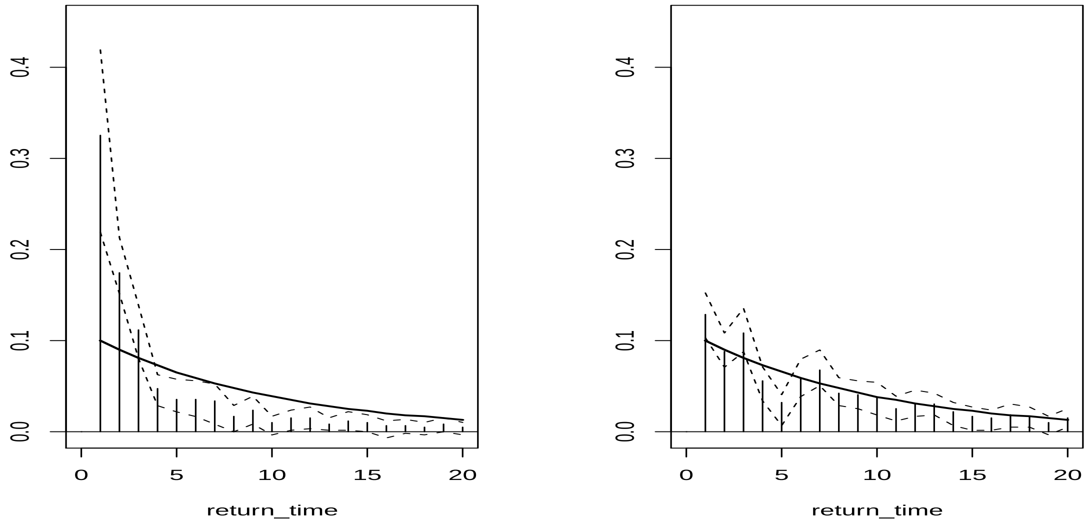

Figure 3.4. *Histograms (solid vertical lines) for the return times of extreme events using bootstrapped confidence intervals (dashed lines), geometric probability mass function (light solid) for* $A = \mathbb{R}\backslash[\xi_{0.05}, \xi_{0.95}]$ *for the daily log-returns of BAC* (left) *and for the residuals from a* GARCH(1, 1) *fit to the log-returns.*

referred to Davis and Mikosch [3] for more details. We also mention that this condition is satisfied for some of the standard financial time series models such as GARCH and stochastic volatility; see [2, 3, 4, 5, 6, 10].

In this paper we focus on strictly stationary sequences whose finite-dimensional distributions have power law tails in some generalized sense. In particular, we will assume that the finite-dimensional distributions of the d-dimensional process (X_t) have *regularly varying distributions with index* $\alpha > 0$. This means that for any $h \geq 1$, the radial part $|Y_h|$ of the lagged vector $Y_h = \text{vec}(X_1, \ldots, X_h)$ is *regularly varying with tail index* $-\alpha$:

$$\frac{P(|Y_h| > t\,x)}{P(|Y_h| > x)} \to t^{-\alpha} \quad \text{as } x \to \infty, \quad t > 0\,,$$

and the angular part $Y_h/|Y_h|$ is asymptotically independent of the radial part $|Y_h|$ for large values of $|Y_h|$: for every $h \geq 1$, there exists a random vector $\Theta_h \in \mathbb{S}^{hd-1}$ such that

$$P(Y_h/|Y_h| \in \cdot \mid |Y_h| > x) \stackrel{w}{\to} P(\Theta_h \in \cdot) \quad \text{as } x \to \infty.$$

Here $\stackrel{w}{\to}$ denotes weak convergence on the Borel σ-field of $\mathbb{S}^{hd-1}$, the unit sphere in $\mathbb{R}^{hd}$ with respect to a given norm $|\cdot|$. The distribution $P(\Theta_h \in \cdot)$ is called the *spectral measure* and α the *index of the regularly varying vector* Y_h. We also refer to (X_t) as a *regularly varying sequence with index* α.

For our purposes it will be convenient to use a sequential definition of a regularly varying sequence (X_t) which is equivalent to the definition above: there exists a sequence $a_n \to \infty$, an $\alpha > 0$ and a sequence of non-null Radon measures (μ_h) on the Borel σ-field of $\overline{\mathbb{R}}_0^{hd} = \overline{\mathbb{R}}^{hd}\backslash\{\mathbf{0}\}$ such that for $h \geq 1$,

$$n\,P(a_n^{-1} Y_h \in \cdot) \stackrel{v}{\to} \mu_h(\cdot)\,, \tag{4.1}$$

where $\stackrel{v}{\to}$ denotes vague convergence on the same σ-field; see Resnick [13], Section 6.1. The limiting measures have the property $\mu_h(t\cdot) = t^{-\alpha}\mu_h(\cdot)$, $t > 0$. We refer to Basrak and Segers [1] who give an enlightening interpretation of the structure of a regularly varying sequence.

Now to connect the measure in (4.1) with the extremogram, for suitably chosen sets A and B in $\mathbb{R}^d$ bounded away from the origin, set for $h \geq 2$, $\tilde{A} = A \times \mathbb{R}^{(h-1)d}$ and $\tilde{B} = A \times \mathbb{R}^{(h-2)d} \times B$. Provided $\tilde{A}$ and $\tilde{B}$ are μ_h continuity sets, then

$$\rho_{A,B}(h-1) = \lim_{n\to\infty} P(a_n^{-1}X_h \in B \mid a_n^{-1}X_1 \in A) = \lim_{n\to\infty} \frac{nP(a_n^{-1}Y_h \in \tilde{B})}{nP(a_n^{-1}Y_h \in \tilde{A})} = \frac{\mu_h(\tilde{B})}{\mu_h(\tilde{A})}\,. \tag{4.2}$$

It is worth noting that standard arguments in regular variation allow one to replace a_n in the first limit appearing in (4.2) with any sequence of numbers x_n tending to ∞.

4.2. **Proof of Theorem 2.1.** Notice that $E^*(\widehat{P}_m^*) = \widehat{P}_m$. Now it follows from Theorem 3.1 in Davis and Mikosch [3] that

$$\widehat{P}_m \stackrel{P}{\to} \mu(C)\,. \tag{4.3}$$

This proves (2.8).

Next we prove (2.9), i.e., we study the asymptotic behavior of $m\,s_n^2$. Write $\widetilde{I}_t = I_t - p_0$ and

$$\begin{aligned} \widetilde{C}_n(h) &= n^{-1}\sum_{i=1}^{n} \widetilde{I}_i\widetilde{I}_{i+h}\,, \quad h = 0,\ldots,n\,, \\ \widetilde{\gamma}_n(h) &= n^{-1}\sum_{i=1}^{n-h} \widetilde{I}_i\widetilde{I}_{i+h}\,, \quad h = 0,\ldots,n-1\,, \end{aligned}$$

where the I_j's in the summands are again defined circularly, i.e., $I_j = I_{j \bmod n}$. Since

$$C_n(h) = n^{-1}\sum_{i=1}^{n} I_i\,I_{i+h} - (\overline{I}_n)^2 = \widetilde{C}_n(h) + p_0^2 - (\overline{I}_n)^2$$

and from the central limit theorem in (4) of Theorem 2.1 and (4.3),

$$(n/m)^{1/2}\,m((\overline{I}_n)^2 - p_0^2) = \left[(n/m)^{1/2}\,(\widehat{P}_m - m\,p_0)\right]\left[m^{-1}(\widehat{P}_m + m\,p_0)\right] = O_P(m^{-1})\,,$$

we have

$$\begin{aligned} m\,s_n^2 - m&\left(\widetilde{C}_n(0) + 2\sum_{h=1}^{n-1}(1-h/n)\,(1-p)^h\,\widetilde{C}_n(h)\right) \\ &= m((\overline{I}_n)^2 - p_0^2)\left(1 + 2\sum_{h=1}^{n-1}(1-h/n)\,(1-p)^h\right) \\ &= O_P((mn)^{-1/2}p^{-1}) = o_P(1)\,. \end{aligned} \tag{4.4}$$

In the last step we used assumption (2.7). It follows from (2.2) and Lemma 5.2 in Davis and Mikosch [3] that

$$m\,\widetilde{C}_n(0) \stackrel{P}{\to} \mu(C) \quad \text{and} \quad m\,\widetilde{C}_n(h) \stackrel{P}{\to} \tau_h(C)\,, \quad h \geq 1\,. \tag{4.5}$$

We also have by assumption (2.5),

$$\lim_{k\to\infty}\limsup_{n\to\infty} E\left|m\sum_{h=k}^{r_n}(1-h/n)\,(1-p)^h\widetilde{C}_n(h)\right| \;\le\; \lim_{k\to\infty}\limsup_{n\to\infty} m\sum_{h=k}^{r_n} p_{0h}=0\,. \tag{4.6}$$

Combining (4.4)–(4.6) and recalling the definition of $\sigma^2(C)$ from (2.3), it suffices for (2.9) to show that for every $\delta>0$,

$$\lim_{k\to\infty}\limsup_{n\to\infty} P\left(m\left|\sum_{h=r_n}^{n-1}(1-h/n)\,(1-p)^h\,\widetilde{C}_n(h)\right|>\delta\right)=0\,.$$

We have

$$m\sum_{h=r_n}^{n-1}(1-h/n)\,(1-p)^h\,\widetilde{C}_n(h)=\sum_{h=r_n}^{n-r_n} b_n(h)\,\widetilde{\gamma}_n(h)+o_P(1)=T_n+o_p(1)\,,$$

where

$$b_n(h)=m\,\left[(1-h/n)\,(1-p)^h+(h/n)\,(1-p)^{n-h}\right]\,.$$

Then

$$\begin{aligned}
\operatorname{var}(T_n) &= \sum_{h_1=r_n}^{n-r_n}\sum_{h_2=r_n}^{n-r_n} b_n(h_1)b_n(h_2)\operatorname{cov}(\widetilde{\gamma}_n(h_1),\widetilde{\gamma}_n(h_2))\\
&\le \sum_{h_1=r_n}^{n-1}\sum_{h_2=r_n}^{n-1} b_n(h_1)b_n(h_2)\max_{r_n\le h_1,h_2<n-1}|\operatorname{cov}(\widetilde{\gamma}_n(h_1),\widetilde{\gamma}_n(h_2))|\\
&\le c\,p^{-2}m^2\max_{r_n\le h_1,h_2<n-1}|\operatorname{cov}(\widetilde{\gamma}_n(h_1),\widetilde{\gamma}_n(h_2))|\,. \qquad (4.7)
\end{aligned}$$

Here and in what follows, c denotes any positive constants whose value is not of interest. We have for fixed k,

$$\begin{aligned}
m^2|\operatorname{cov}(\widetilde{\gamma}_n(h_1),\widetilde{\gamma}_n(h_2))| &= (m/n)^2\left|\sum_{t=1}^{n-h_1}\sum_{s=1}^{n-h_2}\operatorname{cov}(I_tI_{t+h_1},I_sI_{s+h_2})\right|\\
&\le c\,m^2/n\sum_{r=1}^{n}|\operatorname{cov}(I_0I_{h_1},I_rI_{r+h_2})|\\
&= c\,(m/n)\,m\left(\sum_{|r-h_1|\le k}+\sum_{k<|r-h_1|\le r_n}+\sum_{|r-h_1|>r_n}\right)|\operatorname{cov}(I_0I_{h_1},I_rI_{r+h_2})|\\
&= c\,(m/n)[J_1+J_2+J_3]\,.
\end{aligned}$$

Using the sequential definition of the regular variation of (X_t) (see (4.1)), we have for fixed k

$$\limsup_{n\to\infty} J_1\le\limsup_{n\to\infty}\sum_{|r-h_1|\le k} m\,[p_{0,|r-h_1|}+p_0^2]\le\sum_{h\le k}\tau_h(C)\le\sigma^2(C)<\infty\,.$$

Next we use condition (2.5):

$$\lim_{k\to\infty}\limsup_{n\to\infty} J_2\le\lim_{k\to\infty}\limsup_{n\to\infty} m\sum_{k<|r-h_1|\le r_n} p_{0,|r-h_1|}+c\lim_{n\to\infty} m\,r_n\,p_0^2=0\,.$$

Finally, the mixing condition (2.4) yields

$$\limsup_{n\to\infty} J_3 \le c\,\limsup_{n\to\infty} m \sum_{|r-h_1|>r_n} \alpha_{|r-h_1|} = 0\,.$$

Thus we proved that

$$m^2|\mathrm{cov}(\widetilde{\gamma}_n(h_1),\widetilde{\gamma}_n(h_2))| \le c\,(m/n)$$

uniformly for $h_1,h_2\ge r_n$ and large n. We conclude from (4.7) and assumption (2.7) that

$$\mathrm{var}(T_n)\le c\,m/(np^2)\to 0\,.$$

Using (2.4), it also follows that $ET_n\to 0$. Thus we proved that $E(T_n^2)\to 0$. Combining the bounds above, we conclude that (2.9) is satisfied.

Next we prove the central limit theorem (2.11). Since both sums $\widehat{P}_m^*$ and $\widehat{P}_m$ contain the same number of summands and we consider the difference $(n/m)^{1/2}(\widehat{P}_n^*-\widehat{P}_m)$ we will assume in what follows that all summands $(m/n)I_t$ in $\widehat{P}_m^*$ and $\widehat{P}_m$ are replaced by their centered versions $(m/n)\widetilde{I}_t=(m/n)(I_t-p_0)$. We write $\widetilde{P}_m^*$ and $\widetilde{P}_m$ for the corresponding centered versions.

Write

$$\begin{aligned} S_{K_i,L_i} &= \widetilde{I}_{K_i}+\cdots+\widetilde{I}_{K_i+L_i-1}\,,\quad i=1,2,\ldots,\\ S_{nN} &= S_{K_1,L_1}+\cdots+S_{K_N,L_N}\,. \end{aligned} \tag{4.8}$$

Lemma 4.1. *Under the conditions of Theorem 2.1,*

$$P^*\big((n/m)^{1/2}|\widetilde{P}_m^*-(m/n)S_{nN}|>\delta\big)\to 0\,,\quad \delta>0\,. \tag{4.9}$$

Proof. By the argument in Politis and Romano [12] on p. 1312, using the memoryless property of the geometric distribution, $(m/n)S_{nN}-\widetilde{P}_m^*$ has the same distribution as $(m/n)S_{K_1,L_1}$ with respect to P^*. Hence it suffices for (4.9) to show that $(m/n)E^*(|S_{K_1,L_1}|^2)\stackrel{P}{\to}0$. An application of Markov's inequality shows that the latter condition is satisfied if

$$(m/n)E(|S_{K_1,L_1}|^2)\to 0\,. \tag{4.10}$$

We have by stationarity that

$$\begin{aligned} (m/n)E(|S_{K_1,L_1}|^2) &= (m/n)\,E(|\widetilde{I}_1+\cdots+\widetilde{I}_L|^2)\\ &= (m/n)\sum_{l=1}^{\infty}\mathrm{var}(I_1+\cdots+I_l)\,(1-p)^{l-1}\,p\,. \end{aligned}$$

We have (sums over empty index sets being zero) for every fixed $k,l\ge 1$,

$$\begin{aligned} m\,\mathrm{var}(I_1+\cdots+I_l) &= l\,\Big[m\,\mathrm{var}(I_1)+2m\sum_{h=1}^{l-1}(1-h/l)\,\mathrm{cov}(I_0,I_h)\Big]\\ &= l\,\Big[m\,\mathrm{var}(I_1)+2m\Big(\sum_{1\le h\le k}+\sum_{k<h\le r_n}+\sum_{r_n<h\le l-1}\Big)(1-h/l)\,\mathrm{cov}(I_0,I_h)\Big]\,. \end{aligned}$$

The right-hand side is $O(l)$ uniformly for l by virtue of regular variation of (X_t) and in view of the mixing condition (M). Since $np\to\infty$ we have

$$(m/n)E(|S_{K_1,L_1}|^2)\le c\,n^{-1}\sum_{l=1}^{\infty} l\,(1-p)^{l-1}\,p\le c\,(np)^{-1}\to 0\,.$$

This proves (4.10) and finishes the proof of the lemma. □

In view of (4.9) it suffices for (2.11) to show that $(n/m)^{1/2}[(m/n)S_{nN} - \widetilde{P}_m] \stackrel{d}{\to} N(0,\sigma^2(C))$ conditional on (X_i). Recall that $E^*(S_{K_1,L_1}) = p^{-1}(\overline{I}_n - p_0)$. Then

$$(4.11)\qquad (n/m)^{1/2}\Big[(m/n)S_{nN} - \widetilde{P}_m\Big] - (m/n)^{1/2}\sum_{i=1}^{N}(S_{K_i,L_i} - p^{-1}(\overline{I}_n - p_0))$$

$$= (n\,p)^{-1/2}\frac{N - n\,p}{\sqrt{n\,p}}\,\big[(n/m)^{1/2}\widetilde{P}_m\big]$$

The quantities $((N-np)/\sqrt{np})$ are asymptotically normal as $np\to\infty$ and have variances which are bounded for all n. Therefore these quantities are stochastically bounded in P^*-probability. Moreover, the quantities $(n/m)^{1/2}\widetilde{P}_m$ have bounded variances in P-probability, and therefore

$$(np)^{-1/2}[(n/m)^{1/2}\widetilde{P}_m] \stackrel{P}{\to} 0\,.$$

These arguments applied to (4.11) yield for $\delta>0$,

$$P^*\Big(\Big|(n/m)^{1/2}\big[(m/n)S_{nN} - \widetilde{P}_m\big] - (m/n)^{1/2}\sum_{i=1}^{N}(S_{K_i,L_i} - p^{-1}(\overline{I}_n - p_0))\Big| > \delta\Big) \stackrel{P}{\to} 0\,.$$

Therefore it suffices to prove that

$$(m/n)^{1/2}\sum_{i=1}^{N}(S_{K_i,L_i} - p^{-1}(\overline{I}_n - p_0)) \stackrel{d}{\to} N(0,\sigma^2(C))$$

conditional on (X_i). Now, an Anscombe type argument (e.g. Embrechts et al. [8], Lemma 2.5.8) combined with the asymptotic normality of $((N-np)/\sqrt{np})$ as $np\to\infty$ show that the random index N in the sum above can be replaced by any integer sequence $\ell=\ell_n\to\infty$ satisfying the relation $np/\ell\to 1$. Given (X_i), the triangular array

$$(m/n)^{1/2}(S_{K_i,L_i} - p^{-1}(\overline{I}_n - p_0))\,,\quad i=1,\ldots,\ell\,,\quad n=1,2,\ldots\,,$$

consists of row-wise iid mean zero random variables, hence it satisfies the assumption of infinite smallness conditional on (X_i). Therefore it suffices to apply a classical central limit theorem for triangular arrays of independent random variables conditional on (X_i). In view of (2.9), $\sigma^2(C)$ is the asymptotic variance of the converging partial sum sequence. Thus it suffices to prove the following Lyapunov condition conditional on (X_i):

$$(m/n)^{3/2}\ell E^*\big(|S_{K_1,L_1} - p^{-1}(\overline{I}_n - p_0)|^3\big)\quad\sim\quad m^{3/2}n^{-1/2}\,p\,E^*\big(|S_{K_1,L_1} - p^{-1}(\overline{I}_n - p_0)|^3\big) \stackrel{P}{\to} 0\,.$$

An application of the C_r-inequality yields

$$\begin{aligned}
& m^{3/2}n^{-1/2}\,p\,E^*\big(|S_{K_1,L_1} - p^{-1}(\overline{I}_n - p_0)|^3\big)\\
\le\;& 4\,m^{3/2}\,n^{-1/2}\,p\,\Big[E^*\big(|S_{K_1,L_1}|^3\big) + p^{-3}|\overline{I}_n - p_0|^3\Big]\\
=\;& 4m^{3/2}n^{-1/2}\,p\,E^*\big(|S_{K_1,L_1}|^3\big) + 4(n\,p)^{-2}|(n/m)^{1/2}\widetilde{P}_m|^3\,.
\end{aligned}$$

Since the last expression is $o_P(1)$ it suffices to show that

$$(4.12)\qquad m^{3/2}n^{-1/2}\,p\,E^*\big(|S_{K_1,L_1}|^3\big) \stackrel{P}{\to} 0\,.$$

An application of Markov's inequality shows that it suffices to switch to unconditional moments in the last expression. Writing $S_l = \widetilde{I}_1 + \cdots + \widetilde{I}_l$, we have by stationarity of (X_i)

$$E\big(|S_{K_1,L_1}|^3\big) = E\big(|S_{L_1}|^3\big) = \sum_{l=1}^{\infty} E(|S_l|^3)\,(1-p)^{l-1}\,p\,.$$

Next we employ a moment bound due to Rio [14], p. 54,

$$E(|S_l|^3) \le 3\widetilde{s}_l^3 + 144\,l \int_0^1 [\alpha^{-1}(x/2)\wedge l]^2\, Q^3(x)\,dx\,,$$

where α^{-1} denotes the generalized inverse of the rate function $\alpha(t)=\alpha_{[t]}$, Q is the quantile function of the distribution of $|\widetilde{I}_1| = |I_1 - p_0|$ and

$$\widetilde{s}_l^2 = \sum_{i=1}^{l}\sum_{j=1}^{l} |\mathrm{cov}(I_i,I_j)| = l\,p_0\,(1-p_0) + 2\sum_{h=1}^{l-1}(l-h)\,|p_{0h}-p_0^2|\,.$$

For every fixed $k\ge 1$,

$$\widetilde{s}_l^2 \;\le\; (l/m)\left[m\,p_0 + 2\left(\sum_{h=1}^{k} + \sum_{h=k+1}^{r_n} + \sum_{h=r_n+1}^{l-1}\right)(1-h/l)\,m\,|p_{0h}-p_0^2|\right]\,,$$

where sums over empty index sets are zero. Using regular variation of (X_i) and the mixing condition (M), we conclude that the right-hand side is of the order $O(l/m)$ uniformly for l. Hence

$$\begin{aligned} m^{3/2}\,n^{-1/2}\,p\sum_{l=1}^{\infty}\widetilde{s}_l^3\,(1-p)^{l-1}\,p \;&\le\; c\,n^{-1/2}\,p\sum_{l=1}^{\infty} l^{3/2}\,(1-p)^{l-1}\,p \\ &\le\; c\,(n\,p)^{-1/2}\to 0\,. \end{aligned}$$

Direct calculation with the quantile function of $|\widetilde{I}_t|$ shows that

$$\begin{aligned} \int_0^1 (\alpha^{-1}(x/2)\wedge l)^2 Q^3(x)dx \;&=\; p_0^3\int_0^{1-p_0}[\alpha^{-1}(x/2)\wedge l]^2 dx + (1-p_0)^3\int_{1-p_0}^{1}[\alpha^{-1}(x/2)\wedge l]^2\,dx \\ &\le\; c\left[m^{-3}\sum_{k=1}^{\infty} k\alpha_k + m^{-1}\right] = O(m^{-1})\,. \end{aligned}$$

In the last step we used condition (2.6). Combining the estimates above, we obtain

$$m^{3/2}n^{-1/2}\,p\sum_{l=1}^{\infty}E(|S_l|^3)(1-p)^{l-1}\,p \;\le\; c\,[(n\,p)^{-1/2} + (m/n)^{-1/2}]\to 0\,.$$

This proves relation (4.12) and concludes the proof of the theorem.

4.3. **Proof of Theorem 2.2.** From (2.10) we know that

$$P^*(|\widehat{P}_m^*(D_i)-\mu(D_i)|>\delta)\stackrel{P}{\to}0\,,\quad \delta>0\,,\quad i=1,\ldots,h+1, \tag{4.13}$$

therefore (2.13) follows.

Relation (4.13) implies that for each $i = 1, \ldots, h$, in P^*-probability,

$$\begin{aligned}\widehat{\rho}^*_{C,D_i} - \widehat{\rho}_{C,D_i} &= \frac{\widehat{P}^*_m(D_i)\widehat{P}_m(C) - \widehat{P}^*_m(C)\widehat{P}_m(D_i)}{\widehat{P}^*_m(C)\widehat{P}^*_m(C)} \\ &= \frac{1+o_P(1)}{\mu^2(C)}\Big[\mu(C)(\widehat{P}^*_m(D_i) - \widehat{P}_m(D_i)) - \mu(D_i)(\widehat{P}^*_m(C) - \widehat{P}_m(C))\Big]\,.\end{aligned}$$

Therefore it suffices for the central limit theorem (2.14) to prove a multivariate central limit theorem for the quantities $\widehat{P}^*_m(D_i) - \widehat{P}_m(D_i)$, $i = 1, \ldots, h+1$. We will show the result for $h = 1$; the general case is analogous. It will be convenient to write $D = D_1$ and $C = D_2$.

Lemma 4.2. *The following central limit theorem holds in P^*-probability*

$$\mathbf{S}_n = (n/m)^{1/2}\begin{pmatrix} \widehat{P}^*_m(D) - \widehat{P}_m(D) \\ \widehat{P}^*_m(C) - \widehat{P}_m(C)\end{pmatrix} \stackrel{d}{\to} N(\mathbf{0}, \mathbf{\Sigma})\,,$$

where the asymptotic covariance matrix is given by

$$\begin{aligned}\Sigma &= \begin{pmatrix} \sigma^2(D) & r_{DC} \\ r_{DC} & \sigma^2(C)\end{pmatrix}\,, \\ r_{DC} &= \mu(C\cap D) + \sum_{i=1}^\infty [\mu_{i+1}(D\times \overline{\mathbb{R}}_0^{d(i-2)}\times C) + \mu_{i+1}(C\times \overline{\mathbb{R}}_0^{d(i-2)}\times D)]\,.\end{aligned}$$

Proof. We show the result by using the Cramér-Wold device, i.e.,

$$\mathbf{z}'\mathbf{S}_n \stackrel{d}{\to} N(0, \mathbf{z}'\Sigma\mathbf{z})\,, \quad \mathbf{z}\in\mathbb{R}^2\,.$$

We indicate the main steps in the proof in which we follow the lines of the proof of Theorem 2.1. We observe that $E^*(\mathbf{z}'\mathbf{S}_n) = 0$. Next we show that, conditional on (X_t)

$$\begin{aligned}\mathrm{var}^*(\mathbf{z}'\mathbf{S}_n) &= (n/m)\Big[z_1^2\mathrm{var}^*(\widehat{P}^*_m(D)) + z_2^2\mathrm{var}^*(\widehat{P}^*_m(C)) + 2z_1z_2\mathrm{cov}^*(\widehat{P}^*_m(C), \widehat{P}^*_m(D))\Big] \\ &\stackrel{P}{\to} \mathbf{z}'\Sigma\mathbf{z}\,. \qquad (4.14)\end{aligned}$$

By (2.9), $(n/m)\mathrm{var}^*(\widehat{P}^*_m(D)) \stackrel{P}{\to} \sigma^2(D)$ and $(n/m)\mathrm{var}^*(\widehat{P}^*_m(C)) \stackrel{P}{\to} \sigma^2(C)$. Hence it suffices to show that

$$(n/m)\mathrm{cov}^*(\widehat{P}^*_m(C), \widehat{P}^*_m(D)) \stackrel{P}{\to} r_{DC}\,. \qquad (4.15)$$

We observe that

$$\mathrm{cov}^*(\widehat{P}^*_m(C), \widehat{P}^*_m(D)) = \frac{1}{4}\Big[\mathrm{var}^*(P^*_m(C) + \widehat{P}^*_m(D)) - \mathrm{var}^*(P^*_m(C) - \widehat{P}^*_m(D))\Big]\,.$$

Observe that $P^*_m(C) \pm \widehat{P}^*_m(D)$ contain the bootstrap sequences $I^*_t(C) \pm I^*_t(D) = (I_t(C) \pm I_t(D))^*$, $t = 1, \ldots, n$. Therefore the same ideas as for Lemma 5.2 in Davis and Mikosch [3] and in the proof of Theorem 2.1 above apply to show (4.15). We omit the details.

It immediately follows from Lemma 4.1 and the argument following it that the multivariate central limit theorem can be reduced to the central limit theorem for the triangular array

$$\begin{aligned}&(m/n)^{1/2}\Big[z_1\big(S_{K_i,L_i}(D) - p^{-1}(\overline{I}_n(D) - p_0(D))\big) + z_2\big(S_{K_i,L_i}(C) - p^{-1}(\overline{I}_n(C) - p_0(C))\big)\Big]\,, \\ &i = 1, \ldots, \ell\,, \quad n = 1, 2, \ldots\,, \qquad (4.16)\end{aligned}$$

where $\ell = \ell_n$ satisfies the relation $np/\ell \to 1$. This array consists of row-wise iid mean zero random variables, conditional on (X_t). Relation (4.14) yields the correct asymptotic variance for the central limit theorem of the quantities (4.16). Therefore it again suffices to apply a Lyapunov condition of order 3 to the summands (4.16) conditional on (X_t). However, an application of the C_r-inequality yields that, up to a constant multiple, this Lyapunov ratio is bounded by the sum of the Lyapunov ratios of $S_{K_1,L_1}(C)$ and $S_{K_1,L_1}(D)$ which, conditional on (X_t), were shown to converge to zero in the proof of Theorem 2.1. This finishes the sketch of the proof of the theorem. □

Acknowledgment. We wish to thank the AE and referee for their helpful comments through the course of the review which led to an improved paper. This research was supported in part by NSF DMS-1107031 (Davis) and by the Danish Natural Science Research Council (FNU) Grants 09-072331 "Point process modelling and statistical inference" and 10-084172 "Heavy tail phenomena: Modeling and estimation" (Mikosch). The first author also gratefully acknowledges the kind support of the University of Copenhagen, where he holds the position of Villum Kan Rasmussen Visiting Professorship (2011-2013).

References

[1] Basrak, B. and Segers, J. (2009) Regularly varying multivariate time series. *Stoch. Proc. Appl.* **119**, 1055–1080.

[2] Davis, R.A. and Mikosch, T. (2001) Point process convergence of stochastic volatility processes with application to sample autocorrelations. *J. Appl. Probab.* **38A**, 93–104.

[3] Davis, R.A. and Mikosch, T. (2009) The extremogram: a correlogram for extreme events. *Bernoulli* **15**, 977–1009.

[4] Davis, R.A. and Mikosch, T. (2009) Probabilistic properties of stochastic volatility models. In: Andersen, T.G., Davis, R.A., Kreiss, J.-P. and Mikosch, T. (Eds.) *The Handbook of Financial Time Series.* Springer, Heidelberg, pp. 255–268.

[5] Davis, R.A. and Mikosch, T. (2009) Extreme value theory for GARCH processes. In: Andersen, T.G., Davis, R.A., Kreiss, J.-P. and Mikosch, T. (Eds.) *The Handbook of Financial Time Series.* Springer, Heidelberg, pp. 187–200.

[6] Davis, R.A. and Mikosch, T. (2009) Extremes of stochastic volatility models. In: Andersen, T.G., Davis, R.A., Kreiss, J.-P. and Mikosch, T. (Eds.) *The Handbook of Financial Time Series.* Springer, Heidelberg, pp. 355–364.

[7] Davis, R.A., Mikosch, T., and Cribben, I. (2011) Estimating extremal dependence in univariate and multivariate time series via the extremogram. arXiv:1107.5592v1 [stat.ME].

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

[9] Geman, S. and Chang, L.-B. (2009) Rare events in the financial markets. *http://www.dam.brown.edu/people/geman/Homepage/Some recent stuff/Rare Events.pdf.*

[10] Lindner, A. (2009) Stationarity, mixing, distributional properties and moments of GARCH(p,q)-processes. In: Andersen, T.G., Davis, R.A., Kreiss, J.-P. and Mikosch, T. (Eds.) (2009), *The Handbook of Financial Time Series.* Springer, Heidelberg, pp. 43–70.

[11] McNeil, A., Frey, R. and Embrechts, P. (2005) *Quantitative Risk Management: Concepts, Techniques, and Tools.* Princeton Series in Finance. Princeton University Press, Princeton NJ.

[12] Politis, D.N. and Romano, J.P. (1994) The stationary bootstrap. *J. Amer. Statist. Assoc.* **89**, 1303–1313.

[13] Resnick, S.I. (2007) *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling.* Springer, New York.

[14] Rio, E. (1994) About the Lindeberg method for strongly mixing sequences. *ESAIM: Probability and Statistics* **1**, 35–61.

Department of Statistics, Columbia University, 1255 Amsterdam Ave. New York, NY 10027, U.S.A.
E-mail address: rdavis@stat.columbia.edu , www.stat.columbia.edu/∼rdavis

Department of Mathematics, University of Copenhagen, Universitetsparken 5, DK-2100 Copenhagen, Denmark
E-mail address: mikosch@math.ku.dk , www.math.ku.dk/∼mikosch

Department of Statistics, Columbia University, 1255 Amsterdam Ave. New York, NY 10027, U.S.A.
E-mail address: ivor@stat.columbia.edu