*The Annals of Applied Probability*
2001, Vol. 11, No. 3, 878–921

# LONG STRANGE SEGMENTS OF A STOCHASTIC PROCESS[1]

BY PETER MANSFIELD, SVETLOZAR T. RACHEV AND GENNADY SAMORODNITSKY

*University of Tasmania, University of Karlsruhe and Cornell University*

We study long strange intervals in a linear stationary stochastic process with regularly varying tails. It turns out that the length of the longest strange interval grows, as a function of the sample size, at different rates in different parts of the parameter space. We argue that this phenomenon may be viewed in a fruitful way as a phase transition between short- and long-range dependence. We prove a limit theorem that may form a basis for statistical detection of long-range dependence.

**1. Introduction.** Imagine that a stationary time series $X_1, X_2, \dots$ represents input in a stochastic system, for example, the amount of work arriving to a service station in a unit of time, or the total of claims arriving to an insurance company in a unit of time. Theoretically speaking, a stochastic system is designed well if its "capacity" exceeds the average "load" generated by the input to the system. In a queuing system context, this usually means that the amount of work the servers are capable of processing in a unit of time exceeds the average amount of work arriving in a unit of time. In an insurance context, this means that the premium income the company receives per unit of time exceeds the average total amount of claims arriving per unit of time. Of course, it is well understood that even well-designed stochastic systems are affected by the intrinsic randomness in their input. Random fluctuations can cause a queue buildup in a service station and could imply disaster for an insurance company.

It is somewhat paradoxical that what really affects a stochastic system is not so much the "pure chaos" often associated with randomness, but rather certain kinds of "order" found in it. Even if the input $X_1, X_2, \dots$ is a nondegenerate iid sequence with a finite mean, chance will create periods of time when the observed (sample) mean is significantly different from the theoretical mean. Such periods will become longer when the time series $X_1, X_2, \dots$ has (positive) dependence, or memory; the longer the memory, the longer the length of such intervals. In the insurance company context, long periods of time when the claims arrive at a rate significantly higher than expected may well lead to ruin. Similarly, long periods of time when observed traffic intensity

Received December 1999; revised October 2000.

[1]This research was done with support of SEW-EURODRIVE-Foundation (SEW-Group) that funded a visit of G. Samorodnitsky to University of Karlsruhe in summer of 1999. Samorodnitsky's research was also partially supported by NSF Grants DMS-97-04982 and DMI-97-13549 as well as by NSA Grant MDA904-98-1-0041 at Cornell University.

*AMS* 2000 *subject classifications*. Primary 60G10, 60F15; secondary 60G70.

*Key words and phrases*. Long-range dependence, stationary process, large deviations, heavy tails, infinite moving average, maxima, regular variation, extreme value distribution, applications in finance, insurance, telecommunications.

is significantly higher than expected may cause extreme delays at a service station. If a queuing system is highly loaded to begin with, it will be hit even harder. This is the case, for example, with manufacturing systems in certain capital-intensive industries, such as the semiconductor industry, where the system has to run very close to capacity to produce the needed profit margin.

Today, it is not easy to find an area of applications of stochastic models where one does not believe in presence of dependence in the data relevant to that area. Indeed, many data sets are believed to exhibit presence of a special kind of dependence, the so-called *long-range dependence,* or *long memory*. Long-range dependence has been found in financial data; see, for example, Ding, Granger and Engle (1993), Bollerslev and Mikkelsen (1996) or Breidt, Crato and de Lima (1996), in communication networks; see, for example, Willinger et al. (1995), Beran et al. (1995) or Crovella and Bestavros (1996), and many other areas. Additional references can be found in Beran (1994). Long-range dependence is supposed to be the kind of dependence that not only dissipates slowly with time, but is qualitatively different from ordinary, short-range dependence.

Making the above statement precise has proved to be difficult. Early on, long-range dependence has become associated with particular kinds of scaling of a stochastic process, the simplest among which is self-similarity. This originated with the pioneering work of Mandelbrot and his co-workers [Mandelbrot and Van Ness (1968), Mandelbrot and Wallis (1968, 1969a, b, c)], which explained the *Hurst phenomenon*, for example, the empirical findings of Hurst (1951), who studied the water level of the Nile River. For this purpose, Mandelbrot and his co-workers used fractional Gaussian noise, which is increment process of the *fractional Brownian motion*, that is, the self-similar Gaussian process with stationary increments. Fractional Brownian motion is characterized by a single parameter (apart from its scale) $H \in (0, 1)$, sometimes called the Hurst parameter; the covariance function of fractional Gaussian noise is not summable when the Hurst parameter $H$ is greater than 0.5, which is precisely the range required to explain the Hurst phenomenon. Looking at the rate of decay of correlations is an attractive way of thinking of length of dependence in a stationary Gaussian sequence. Since then, nonsummability of correlations has become a common way of defining long-range dependence, even when the stochastic process is not believed to be Gaussian [see, e.g., Taqqu and Teverovsky (1998)], with possible adjustments consisting of requiring actual regular variation of correlations as in Beran (1994), or allowing any hyperbolic-type of decay of correlations as in Taqqu (1987).

It is difficult, however, to justify such concentration on the rate of decay of correlations. First, this does not allow one to talk about long-range dependence in a stochastic process with infinite variance; infinite-variance models have acquired prominence in the last 10 years. Furthermore, even if the variance is finite, the information carried by correlations is fairly limited if the actual process is far from being a Gaussian one. In fact, some researchers argue that there are simpler ways to explain the observed slowly decaying empirical correlations than by introducing sophisticated models viewed as having long memory. Certain kinds of nonstationary models will have a similar property. See, for example, Mikosch and Starica (1999) [but this line of thought can be traced back to Bhattacharya, Gupta and Waymire (1983)].

A possible alternative approach to the phenomenon of long-range dependence is to look at implications of the latter. That is, one looks at a particular important functional of a stochastic process. One looks then for a kind of a "phase transition" in the behavior of this functional. This approach has both advantages and drawbacks. Its main drawback is in not offering a unique definition of long-range dependence, for doing so ties the analysis to a given functional, or a family of functionals. Its advantage is in concentrating on an object of a priori importance. This makes the discussion of whether or not nonstationary models can have a similar property somewhat redundant, for it is the property itself that is of a greater interest than the model *per se*.

In the present paper we concentrate on one such functional of interest. Let $X_1, X_2, \dots$ be a stochastic process. For a Borel set $A \subset \mathbb{R}$, we define, for every $n = 1, 2, \dots,$

$$R_n(A) = \sup\left\{ j - i\colon 0 \le i < j \le n,\ \frac{X_{i+1} + \cdots + X_j}{j - i} \in A \right\} \tag{1.1}$$

(defined to be equal to zero if the supremum is taken over the empty set). If $X_1, X_2, \dots$ is a stationary ergodic process with a finite mean $\mu = EX_1$, then of particular interest are sets of the type

$$A = (\theta, \infty) \qquad \text{with a } \theta > \mu$$

and

$$A = (-\infty, \theta) \qquad \text{with a } \theta < \mu.$$

Indeed, $R_n((\theta, \infty))$ and $R_n((-\infty, \theta))$ are the greatest lengths of time intervals when *the system runs under effective load that is different from the nominal load*. We have already mentioned that such time intervals can be of a crucial importance in manufacturing and insurance applications, but these functionals are also important in finance, comparative analysis of DNA sequences and analysis of computer search algorithms. In this paper we will concentrate on the sets of the type $(\theta, \infty)$, and we use the notation

$$R_n(\theta) := R_n((\theta, \infty)).$$

It is clear that one can analyze the sets of the type $(-\infty, \theta)$ by changing the sign of the whole stochastic process.

A word on the terminology. The intervals whose length the functional $R_n(A)$ measures are sometimes called *long rare* intervals; see, for example, Dembo and Zeitouni (1993). We prefer to call them *long strange* intervals, reflecting the fact that, even though from a certain point of view, we are talking about a typical length, in such time intervals the system seems to overcome the law of large numbers when the mean $\mu$ lies outside of the closure of the set $A$.

Here is the specific model we will consider. Let

$$X_n = \mu + \sum_{j=-\infty}^{\infty} \varphi_{n-j} Z_j, \qquad n = 1, 2, \dots, \tag{1.2}$$

where $\dots, Z_{-1}, Z_0, Z_1, \dots$ is a sequence of zero mean iid random variables and $\mu$ is a constant. That is, $X_1, X_2, \dots$ is a two-sided linear process (or a two-sided infinite moving average).

In this paper we assume that $Z = Z_0$ satisfies the following regular variation and tail balance conditions:

$$\begin{gathered} P(|Z| > \lambda) = L(\lambda)\lambda^{-\alpha}, \\ \lim_{\lambda\to\infty} \frac{P(Z > \lambda)}{P(|Z| > \lambda)} = p, \qquad \lim_{\lambda\to\infty} \frac{P(Z < -\lambda)}{P(|Z| > \lambda)} = q, \end{gathered} \tag{1.3}$$

as $\lambda \to \infty$, for some $\alpha > 1$ and $0 < p = 1 - q \le 1$. Here $L$ is a slowly varying function at infinity. The coefficients $(\varphi_j)$, not all of which are equal to zero, have to satisfy certain assumptions to make sure that the infinite sum in (1.2) is well defined. Sufficient conditions for convergence are

$$\begin{cases} \displaystyle\sum_{j=-\infty}^{\infty} \varphi_j^2 < \infty, & \text{for } \alpha > 2, \\ \displaystyle\sum_{j=-\infty}^{\infty} |\varphi_j|^{\alpha-\varepsilon} < \infty \quad \text{for some } \varepsilon > 0 & \text{for } \alpha \le 2. \end{cases} \tag{1.4}$$

Moreover, under the conditions (1.4), $X_1, X_2, \dots$ is an ergodic stationary process whose marginal distribution satisfies

$$\begin{aligned} P(X > \lambda) &\sim \sum_{j=-\infty}^{\infty} |\varphi_j|^\alpha [p\, I_{\{\varphi_j>0\}} + q\, I_{\{\varphi_j<0\}}] P(|Z| > \lambda) \\ &\sim \sum_{j=-\infty}^{\infty} |\varphi_j|^\alpha [p\, I_{\{\varphi_j>0\}} + q\, I_{\{\varphi_j<0\}}] L(\lambda)\lambda^{-\alpha}, \end{aligned} \tag{1.5}$$

as $\lambda \to \infty$. Here and in what follows, $X$ stands for a generic random variable with the same distribution as $X_1$. See Mikosch and Samorodnitsky (2000), Lemma A3.7. See also Brockwell and Davis (1991) for an extensive treatment of linear processes and Resnick (1987) for a discussion of regular variation in the context of linear processes.

A strictly stronger assumption than (1.4) is that of absolute summability of the coefficients in (1.2):

$$\sum_{j=-\infty}^{\infty} |\varphi_j| < \infty. \tag{1.6}$$

We will see that there are important differences in the ways the functional $R_n(\theta)$ with $\theta > \mu$ behaves when the assumption (1.6) holds, and when this assumption does not hold but the weaker assumption (1.4) holds. In this sense we have a phase transition.

Several remarks are in order.

REMARK 1.1. (i) Linear processes are, arguably, the single most common class of stochastic models used in science. Traditionally, they have been used because of their flexibility to account for almost any possible behavior of covariance functions. On uses moments to fit a linear model to data. See Brockwell and Davis (1991). This argument does not explain why one should use linear models when the variance is infinite [this is the case, for example, when $\alpha < 2$ in (1.3)], or if one is not interested in covariances. See also Resnick (1997) for the case against blind usage of linear processes. We use linear processes because we are interested in the tails, not covariances, and the joint tails of a linear process provide a very good approximation to any multivariate regular varying distribution. See the discussion in Mikosch and Samorodnitsky (2000).

(ii) It is interesting that, in the case when $EZ^2 < \infty$, the condition (1.6) is also the one that assures absolute summability of the covariances of the process $X_1, X_2, \dots$.

Clearly, for any stochastic process and any fixed number $\theta$, the sequence of random variables $R_1(\theta), R_2(\theta), \dots$ is nondecreasing. How fast does it increase? This question, apart from its obvious importance in applications, is of theoretical interest in its own right. If $X_1, X_2, \dots$ is an iid sequence of random variables with finite exponential moments, then $R_n(\theta)$ grows logarithmically fast; see Dembo and Zeitouni (1993), Section 3.2. It turns out that with regularly varying tails, as in our setup, $R_n(\theta)$ itself grows as a regularly varying function. This is not, by itself, surprising. It is, perhaps, more surprising to see the effect of the dependence in the process on the rate of growth of $R_n(\theta)$.

We begin the next section by providing some intuition that may help the reader see why our results are stated the way they are, and where they come from. We also state our main result and give some possible applications of it. Section 3 contains the proof of the main result; it uses additional technical results that are provided in Section 4. Finally, in Section 5, we try to give the reader some idea of what may happen when one uses the lenses provided by our results to look at real data; our data comes from financial applications.

**2. Intuition and the main results.** The idea underlying our results is really very simple. It follows directly from the definition of $R_n(\theta)$ that, for any $1 \le m \le n$,

$$R_n(\theta) \ge m \qquad \text{if and only if } X_{i+1} + \cdots + X_{i+k} > k\theta \text{ for some } k = m, \; m+1, \dots, n \quad \text{and some } i = 0, \dots, n-k. \tag{2.1}$$

Now, by definition (1.2) of the linear process,

$$X_{i+1} + \cdots + X_{i+k} = k\mu + \sum_{j=-\infty}^{\infty} \left( \sum_{d=i+1-j}^{i+k-j} \varphi_d \right) Z_j. \tag{2.2}$$

To develop some intuition, we will use the logic of large deviations (even though this is not really a large-deviation situation, as we will discuss in a moment). Let $\theta > \mu$, so that, for a fixed $i$, the event

$$\{X_{i+1} + \cdots + X_{i+k} > k\theta\}$$

is an unlikely event if $k$ is large. The logic of large deviations is that unlikely events happen in the most likely way, and in the case of power-like tails, the most likely way is often that of the least number of causes. See, for example, Resnick and Samorodnitsky (1999) or Mikosch and Samorodnitsky (2000) for a discussion. This logic tells us that, for large sample sizes $n$ and large $m$ in (2.1), the event $\{R_n(\theta) \geq m\}$ is basically a consequence of a single large positive or negative value of a noise variable. How large this value has to be is determined by the coefficients (2.2). Then, intuitively,

$$P(R_n(\theta) \geq m) \sim P\Bigg(\Bigg(\sum_{d=i+1-j}^{i+k-j} \varphi_d\Bigg) Z_j > k(\theta - \mu), \quad \text{for some } j = \ldots, -1, 0, 1, \ldots \quad \text{some } k = m, m+1, \ldots, n \text{ and some } i = 0, \ldots, n-k\Bigg). \tag{2.3}$$

Now, what does (2.3) mean exactly? We are, obviously, thinking of the sample size $n$ going to infinity, and the number $m$ should increase to infinity as well (as a function of $n$). If the number $m = m(n)$ increases too fast, and the probability in the left-hand side of (2.3) goes to zero, then we are talking about an overall rare event, and then the equivalence of the left- and right-hand sides of (2.3) is, perhaps, not surprising. We are, however, interested in $m = m(n)$ increasing at "just the right rate," so that the probability in the left-hand side of (2.3) should not go to zero, and the event it describes is not a rare event. Why should one expect that the equivalence in (2.3) is still valid? To answer that question, let us look at the simplest possible case, that of an iid sequence $X_1, X_2, \ldots$.

The iid case corresponds to the choice of $\varphi_0 = 1$ and $\varphi_j = 0$ for $j \neq 0$ in (1.2). Assume temporarily for notational simplicity that $L(\lambda) \to L \in (0, \infty)$ as $\lambda \to \infty$ in (1.3), Clearly, we may (and will) assume that $\mu = 0$ (and $\theta > 0$). We will see later that the appropriate choice of $m$ in that case is $m = m(n) = n^{1/\alpha}$ in the sense that $n^{-1/\alpha} R_n(\theta)$ converges weakly to a nonzero limit. Now, it is *typical* for the maximum of $n$ iid random variables with a Pareto-like tail to be of the order of $n^{1/\alpha}$ and, in fact, there is likely to be more than one observation that large. However, for exactly the same reason, observations that large are likely to be separated in time by more than $n^{1/\alpha}$ observations and, hence, one does not expect that more than one of such large observations will contribute to the largest strange interval. This is the intuitive reason for the equivalence in (2.3).

Let us start by noting that the probability in the right-hand side of (2.3) can be understood through fairly straightforward computations. Still assuming that $\mu = 0$ and $\theta > 0$, we can rewrite it in the form

$$
\begin{aligned}
P\Bigg(&\text{for some } j = \dots, -1, 0, 1, \dots, \quad \text{either} \\
&Z_j \sup_{m \le k \le n} \frac{1}{k} \sup_{i=0,\dots,n-k} \left( \sum_{d=i+1-j}^{i+k-j} \varphi_d \right)_+ > \theta \\
&\text{or } Z_j \sup_{m \le k \le n} \frac{1}{k} \sup_{i=0,\dots,n-k} \left( \sum_{d=i+1-j}^{i+k-j} \varphi_d \right)_- < -\theta \Bigg).
\end{aligned}
\tag{2.4}
$$

Here, $a_+ = \max(a, 0)$ and $a_- = (-a)_+$ for any real number $a$.

If one believes that (2.4) gives the main term in the probability $P(R_n(\theta) \ge m)$ (at least, for appropriate choice of $m$), then it is a reasonably straightforward computation (which we will actually perform later) to figure out from (2.4) the behavior of the probability $P(R_n(\theta) \ge m)$ as $n \to \infty$ for a given sequence $m = m(n)$ and, alternatively, to figure out the right choice of $m = m(n)$ so that $P(R_n(\theta) \ge m)$ converges to a limit in $(0, 1)$. At this point it is useful to start introducing the appropriate notation.

Let $F$ denote the distribution function of the random variable $|Z|$ and, for $n \ge 1$, define

$$a_n = \left( \frac{1}{1 - F} \right)^{\leftarrow} (n). \tag{2.5}$$

Here, for a nondecreasing function $U$, we use the notation $U^{\leftarrow}$ to denote the left continuous inverse of $U$:

$$U^{\leftarrow}(y) = \inf\{s\colon U(s) \ge y\}.$$

It follows immediately from (1.3) that the sequence $(a_n)$ is regularly varying at infinity with exponent $1/\alpha$. Here we are following the notation in Resnick (1987), which should also be consulted for facts about regular varying tails and their quantile functions.

It turns out that under the condition (1.6) (the short-memory case in our approach), the appropriate choice for the sequence $m = m(n)$ is $m(n) = a_n$, $n \ge 1$, in the sense that $a_n^{-1} R_n(\theta)$ converges weakly to a nondegenerate limit. In this sense, in the short-memory case, the length $R_n(\theta)$ of the longest strange interval grows as $a_n$ (i.e., as a regularly varying function of the sample size $n$ with exponent $1/\alpha$). This is the same rate as that achieved in the iid case (but the actual weak limit of $a_n^{-1} R_n(\theta)$ does, in general, depend on the coefficients in (1.2) even in the short-memory case).

In contrast to that, the rate of growth of $R_n(\theta)$ is, in general, higher than that of $a_n$ in (2.5) if the condition (1.6) fails. In fact, we will show in a future paper that if the coefficients $(\varphi_n)$ in (1.2) have themselves a certain regular variation property, then the right choice for the sequence $m = m(n)$ is, actually, that of a regularly varying function with exponent *strictly greater* than $1/\alpha$, This suggests that one way to see if the data should be viewed as coming from a long-range dependent model is to plot $\log R_n(\theta)$ versus $\log n$. Long-range dependence corresponds to the case when the plot is close to being linear with a slope $H$ greater than $1/\alpha$, while a slope of $1/\alpha$ signifies short memory (or, rather, absence of indication for long-range dependence). This approach is, hence, akin to the $R/S$ statistic of Hurst (1951); note, however, that the approach using the rate of growth of the longest strange interval has the advantage of relating to an a priori important quantity. It is also interesting that, unlike the Hurst exponent in the $R/S$ statistic, our slope $H$ is not bounded by 1/2 from below.

Even though the above way of trying to detect long-range dependence looks attractive, it is difficult to convert it into a statistical test; the same problem also plagues the $R/S$ statistic. An additional difficulty involved is that the tail index $\alpha$ in (1.5) is not usually known and, hence, has itself, to be estimated from the sample. This should not be called an easy task; see, for example, Embrechts, Klüppelberg and Mikosch (1997) for some of the pitfalls. Therefore, it is desirable to have a statistic that does not rely on the tail index.

Let $M_n = \max(X_1, \dots, X_n)$ be the largest of the first $n$ observations, $n \geq 1$. It is well known that under fairly general conditions, $M_n$ grows as $a_n$ in (2.5); see, for example, Resnick (1987) [under conditions more stringent than (1.6)]; in the case of $\alpha$-stable noise $\dots, Z_{-1}, Z_0, Z_1, \dots$ in (1.2), Leadbetter, Lindgren and Rootzen (1983) show the same thing without the assumption (1.6). One of the by-products of our results is that $M_n$ grows as $a_n$ under the assumption (1.6). One of the grows as $a_n$ even under long-range dependence. We would like to exploit this fact.

Specifically, the statistic we would like to concentrate on is

$$W_n(\theta) = \frac{R_n(\theta)}{M_n}. \tag{2.6}$$

It is the self-normalized nature of the statistic $W_n(\theta)$ that makes it attractive. It turns out that under the assumption (1.6), that is, in the short-range dependence case, the ratio $W_n(\theta)$ has a weak limit; we have, basically already alluded to it. This will not be the case when the assumption (1.6) fails (the long-range dependence case); in this case, $R_n(\theta)$ grows at a faster rate than $M_n$, which is the subject of a subsequent paper.

The natural approach to proving weak convergence of $W_n(\theta)$ is via weak convergence of the sequence of random vectors in $\mathbb{R}^2$:

$$V_n(\theta) = a_n^{-1}\bigl(R_n(\theta), M_n\bigr), \qquad n \geq 1, \tag{2.7}$$

and the continuous mapping theorem. Hence the formulation of our main result that we present now.

THEOREM 2.1. *Let* $\mu = 0$ *and* $\theta > 0$. *Assume* (1.6). *For any* $x > 0$ *and* $y > 0$,

$$\begin{aligned} &P(a_n^{-1}R_n(\theta) \leq x,\ a_n^{-1}M_n \leq y) \\ &\quad \to \exp\{-p\max(M_+(\varphi)^\alpha \theta^{-\alpha} x^{-\alpha},\ m_+(\varphi)^\alpha y^{-\alpha}) \\ &\qquad - q\max(M_-(\varphi)^\alpha \theta^{-\alpha} x^{-\alpha},\ m_-(\varphi)^\alpha y^{-\alpha})\}, \end{aligned} \tag{2.8}$$

*where $p$ and $q$ are the tail weights in* (1.3),

$$M_+(\varphi) = \max\left\{ \sup_{-\infty<k<\infty} \left( \sum_{j=-\infty}^{k} \varphi_j \right)_+, \sup_{-\infty<k<\infty} \left( \sum_{j=k}^{\infty} \varphi_j \right)_+ \right\},$$

$$M_-(\varphi) = \max\left\{ \sup_{-\infty<k<\infty} \left( \sum_{j=-\infty}^{k} \varphi_j \right)_-, \sup_{-\infty<k<\infty} \left( \sum_{j=k}^{\infty} \varphi_j \right)_- \right\},$$

$$m_+(\varphi) = \sup_{-\infty<k<\infty} (\varphi_k)_+$$

*and*

$$m_-(\varphi) = \sup_{-\infty<k<\infty} (\varphi_k)_-.$$

REMARK 2.2. It is clear that for a general mean $\mu$ and any $\theta > \mu$, the statement of Theorem 2.1 remains valid if one replaces $\theta$ by $\theta - \mu$ in the right-hand side of (2.8).

REMARK 2.3. Observe that if the coefficients $\varphi_j$ are all nonnegative, then

$$M_+(\varphi) = \sum_{j=-\infty}^{\infty} \varphi_j, \qquad M_-(\varphi) = 0,$$

and similarly with nonpositive coefficients.

REMARK 2.4. A well-known form of representing the distribution function of a random vector, say, $(X, Y)$, with a general bivariate extreme value distribution (with the so-called $\Phi_\alpha$ marginals) is

$$P(X \le x, Y \le y) = \exp\left\{ -\int_{B_2^+} \max\left( \frac{s_1^\alpha}{x^\alpha}, \frac{s_2^\alpha}{y^\alpha} \right) m(ds_1, ds_2) \right\}, \tag{2.9}$$

where $m$ is a finite measure on $B_2^+ = \{(s_1, s_2)\colon s_1 \ge 0, s_2 \ge 0, s_1^2 + s_2^2 = 1\}$ (the so-called *spectral measure*). See, for example, Resnick (1987). In other words, Theorem 2.1 says that the vector $V_n(\theta)$ in (2.7) converges weakly, as $n \to \infty$, to a bivariate extreme value distribution (2.9) with a two-point spectral measure $m$ that puts the weight

$$p\left( \left( \frac{M_+(\varphi)}{\theta} \right)^2 + (m_+(\varphi))^2 \right)^{\alpha/2} \tag{2.10}$$

at the point

$$\left( \frac{M_+(\varphi)/\theta}{((M_+(\varphi)/\theta)^2 + (m_+(\varphi))^2)^{1/2}}, \frac{m_+(\varphi)}{((M_+(\varphi)/\theta)^2 + (m_+(\varphi))^2)^{1/2}} \right) \tag{2.11}$$

and the weight

$$q\left( \left( \frac{M_-(\varphi)}{\theta} \right)^2 + (m_-(\varphi))^2 \right)^{\alpha/2} \tag{2.12}$$

at the point

$$\left(\frac{M_-(\varphi)/\theta}{((M_-(\varphi)/\theta)^2+(m_-(\varphi))^2)^{1/2}}, \frac{m_-(\varphi)}{((M_-(\varphi)/\theta)^2+(m_-(\varphi))^2)^{1/2}}\right). \tag{2.13}$$

For example, if $q = 0$ or if the coefficients $\varphi_j$ are all nonnegative, then the spectral measure $m$ of the limiting distribution of $V_n(\theta)$ is a point mass given by (2.10) and (2.11). As a matter of fact, an inspection of the proof of Theorem 2.1 shows that, in the case of nonnegative coefficients, one does not need the full strength of the part of the assumption (1.3) that applies to the left tail of the distribution of the noise, and a much weaker assumption will suffice; for example, the left tail being bounded from above by a regularly varying function with any exponent greater than 1.

REMARK 2.5. Clearly, the story described in Theorem 2.1 breaks down at the boundary $\alpha = 1$. Indeed, once the boundary is crossed, that is, if $\alpha < 1$, then the mean $\mu$ is not defined. More importantly, the asymptotic behavior of the statistic $W_n(\theta)$ will be very different in this case. Specifically, for any real $\theta$, $R_n(\theta)$ will grow like $n$ (it is very likely that the sample mean of all $n$ observations exceeds $\theta$), while the largest observation $M_n$ will still grow as a regular varying function with exponent $1/\alpha > 1$. Hence, $W_n(\theta)$ will go to zero in this case. In fact, once the tail exponent $\alpha$ gets close to 1, one is likely to see in practice unusually low values of the statistic $W_n(\theta)$, that are caused not by short memory but, rather, by the proximity of the boundary $\alpha = 1$.

The form of the limiting distribution described in Theorem 2.1 allows one to compute easily the limiting distribution of the statistic $W_n(\theta)$ in (2.6). Let us introduce first some notation. Denote

$$a_+(\theta) = \frac{M_+(\varphi)/\theta}{m_+(\varphi)}, \qquad a_-(\theta) = \frac{M_-(\varphi)/\theta}{m_-(\varphi)}, \tag{2.14}$$

provided the denominators are not zero.

COROLLARY 2.6. *Assume* (1.6) *and*

$$pm_+(\varphi) + qm_-(\varphi) > 0. \tag{2.15}$$

*Then, as* $n \to \infty$,

$$W_n(\theta) \Rightarrow W_\infty(\theta). \tag{2.16}$$

*The limiting law of* $W_\infty(\theta)$ *is described as follows.*

(i) *Let* $0 < p < 1$, $m_+(\varphi) > 0$ *and* $m_-(\varphi) > 0$. *If*

$$a_+(\theta) > a_-(\theta) \tag{2.17}$$

*in* (2.14), *then the limiting distribution in* (2.16) *is concentrated on the interval* $[a_-(\theta), a_+(\theta)]$ *and*

$$P(W_\infty(\theta) = a_-(\theta)) = \frac{1}{1 + (p/q)(M_+(\varphi)/M_-(\varphi))^\alpha},$$

$$P(W_\infty(\theta) = a_+(\theta)) = \frac{1}{1 + (q/p)(m_-(\varphi)/m_+(\varphi))^\alpha},$$

*and on the interval* $(a_-(\theta), a_+(\theta))$ *the law of* $W_\infty(\theta)$ *is absolutely continuous with the density*

$$f_{W_\infty(\theta)}(w) = \frac{(q/p)(m_-(\varphi)/(M_+(\varphi)\theta^{-1}))^\alpha \alpha w^{\alpha-1}}{(1 + (q/p)(m_-(\varphi)/(M_+(\varphi)\theta^{-1}))^\alpha w^\alpha)^2}.$$

*If*, *on the other hand*,

$$a_+(\theta) < a_-(\theta), \tag{2.18}$$

*then the limiting distribution in* (2.16) *is concentrated on the interval* $[a_+(\theta), a_-(\theta)]$ *and*

$$P(W_\infty(\theta) = a_+(\theta)) = \frac{1}{1 + (q/p)(M_-(\varphi)/M_+(\varphi))^\alpha},$$

$$P(W_\infty(\theta) = a_-(\theta)) = \frac{1}{1 + (p/q)(m_+(\varphi)/m_-(\varphi))^\alpha},$$

*and on the interval* $(a_+(\theta), a_-(\theta))$ *the law of* $W_\infty(\theta)$ *is absolutely continuous with the density*

$$f_{W_\infty(\theta)}(w) = \frac{(p/q)(M_-(\varphi)\theta^{-1}/m_+(\varphi))^\alpha \alpha w^{\alpha-1}}{\left(1 + (p/q)(M_-(\varphi)\theta^{-1}/m_+(\varphi))^\alpha w^\alpha\right)^2}.$$

*Finally*, *if*

$$a_+(\theta) = a_-(\theta), \tag{2.19}$$

*then the limiting distribution in* (2.16) *is concentrated at the point* $a_+(\theta)$.

(ii) *Let* $p = 1$ *or* $m_-(\varphi) = 0$. *Then the limiting distribution in* (2.16) *is concentrated at the point* $a_+(\theta)$.

(iii) *Let* $m_+(\varphi) = 0$, $0 < p < 1$, *and* $m_-(\varphi) > 0$. *Then the limiting distribution in* (2.16) *is concentrated at the point* $a_-(\theta)$.

It is interesting to note that, while the distribution of $W_\infty(\theta)$ depends, in general, on the tail index $\alpha$, the support of that distribution *does not* depend on $\alpha$. Given, once again, the difficulty of measuring reliably the tail index from the sample, this is a welcome fact that may prove handy in constructing a statistical test for detecting long-range dependence.

EXAMPLE 2.7. Suppose that the underlying model has, actually, geometrically decaying coefficients:

$$\varphi_j = \rho^{|j|}, \qquad j = \dots, -1, 0, 1, \dots, \quad -1 < \rho < 1, \quad \rho \neq 0.$$

It is elementary to check that, in this case,

$$M_+(\varphi) = \frac{1}{1-\rho}, \qquad m_+(\varphi) = 1,$$

$$M_-(\varphi) = \frac{-\rho}{1-\rho}, \qquad m_-(\varphi) = -\rho,$$

if $-1 < \rho < 0$ and

$$M_+(\varphi) = \frac{1+\rho}{1-\rho}, \qquad m_+(\varphi) = 1,$$

$$M_-(\varphi) = m_-(\varphi) = 0,$$

if $0 < \rho < 1$. An immediate application of Corollary 2.6 tells us that

$$W_\infty(1) = \left\{ \begin{array}{ll} 1/(1-\rho), & \text{if } -1 < \rho < 0, \\ (1+\rho)/(1-\rho), & \text{if } 0 < \rho < 1, \end{array} \right\} := h(\rho) \tag{2.20}$$

almost surely.

Here is a possible, though provocative, way of looking at the significance of an observed value of the statistic $W_n(\theta)$. It is often suggested that long-range dependence–like phenomena observed in data can be explained by using a short-range dependent, for example, autoregressive, linear model, but with the autoregressive polynomial having a root very close to the unit circle. In our example here, this idea amounts to taking $\rho$ close to $\pm 1$.

Imagine that "nature" selects, unknowingly to us, $\rho$ uniformly between $-1$ and 1, and we get to observe a set of observations drawn from the resulting model. If, for large $n$, $W_n(\theta)$ should be about equal to its limit, $W_\infty(\theta) = h(\rho)/\theta$ [with $h(\rho)$ given in (2.20)], then an unusually large value of $W_n(\theta)$ has to result from an extreme value of $\rho$. Recalling that $\rho$ is chosen uniformly in $(-1, 1)$, we obtain the significance of an observed value $w > 1/(2\theta)$:

$$\phi(w, \theta) = \begin{cases} 1/(2w\theta), & \text{if } 1/(2\theta) < w \le 1/\theta, \\ 1/(w\theta + 1), & \text{if } w > 1/\theta. \end{cases} \tag{2.21}$$

Of course, this procedure is not a substitute for a standard statistical test, and we will present such a test in a future publication. However, the above procedure does provide an indication of how difficult it is to describe the observed long strange intervals by short-range dependent models.

Both Theorem 2.6 and Corollary 2.6 are proved in the next section. We take here the opportunity to point out that the idea underlying Theorem 2.1 is the same one as that described in the discussion around (2.3) in a simpler case of a one-dimensional convergence. Indeed, if one believes that the event whose probability we compute in the right-hand side of (2.3) describes how one can expect the event $\{R_n(\theta) \geq m\}$ to occur, then it is also believable that

$$
\begin{aligned}
&P(a_n^{-1} R_n(\theta) \leq x,\ a_n^{-1} M_n \leq y) \\
&\quad \sim P\Bigg(\text{for all } j = \dots, -1, 0, 1, \dots, \quad \Bigg(\sum_{d=i+1-j}^{i+k-j} \varphi_d\Bigg) Z_j \leq k\theta \\
&\qquad\qquad \text{for all } x a_n \leq k \leq n \text{ and all } i = 0, \dots, n-k, \\
&\qquad\qquad \text{and for all } j = \dots, -1, 0, 1, \dots, Z_j \varphi_{i-j} \leq y a_n \\
&\qquad\qquad \text{for all } i = 1, \dots, n\Bigg) \\
&\quad := P_M(\theta, x, y, n).
\end{aligned}
\tag{2.22}
$$

However, a fairly straightforward computation presented in Lemma 4.2 shows that the latter probability converges, as $n \to \infty$, to the right-hand side of (2.8). Therefore, what remains is to argue that $P_M(\theta, x, y, n)$ represents, indeed, the most likely way the event whose probability we are computing in the left-hand side of (2.8) happens. All that the proof in the next section does is to make that intuitive argument rigorous.

## 3. Proofs of the main results.

We start with the proof of Theorem 2.1, and it takes most of this section. The section concludes with the proof of Corollary 2.6.

PROOF. We begin by introducing a notation that will simplify some of the expressions below. For a $j = \dots, -1, 0, 1, \dots$, let

$$
\begin{aligned}
k_+(j) &= \operatorname*{argsup}_{x a_n \leq k \leq n} \frac{1}{k} \sup_{i=0,\dots,n-k} \Bigg(\sum_{d=i+1-j}^{i+k-j} \varphi_d\Bigg)_+, \\
k_-(j) &= \operatorname*{argsup}_{x a_n \leq k \leq n} \frac{1}{k} \sup_{i=0,\dots,n-k} \Bigg(\sum_{d=i+1-j}^{i+k-j} \varphi_d\Bigg)_-, \\
i_+(j) &= \operatorname*{argsup}_{i=0,\dots,n-k_+(j)} \Bigg(\sum_{d=i+1-j}^{i+k_+(j)-j} \varphi_d\Bigg)_+, \\
i_-(j) &= \operatorname*{argsup}_{i=0,\dots,n-k_-(j)} \Bigg(\sum_{d=i+1-j}^{i+k_-(j)-j} \varphi_d\Bigg)_-,
\end{aligned}
\tag{3.1}
$$

with the ties broken in, say, lexicographical order.

We fix for a moment $\varepsilon > 0$, and bound the probability in the left-hand side of (2.8) from above as follows (note that the first term in the right-hand side represents the main term):

$$
\begin{aligned}
&P(a_n^{-1} R_n(\theta) \le x, a_n^{-1} M_n \le y)\\
&\quad \le P\Bigg(\text{for all } j = \dots, -1, 0, 1, \dots, \quad \Bigg(\sum_{d=i+1-j}^{i+k-j} \varphi_d\Bigg) Z_j \le k\theta(1+\varepsilon),\\
&\qquad\qquad \text{for all } x a_n \le k \le n \text{ and all } \quad i = 0, \dots, n-k,\\
&\qquad\qquad \text{and for all } j = \dots, -1, 0, 1, \dots, \quad Z_j \varphi_{i-j} \le y a_n (1+\varepsilon)\\
&\qquad\qquad \text{for all } \quad i = 1, \dots, n\Bigg)\\
&\qquad + P\Bigg(a_n^{-1} R_n(\theta) \le x, \text{ and for some } j = \dots, -1, 0, 1, \dots, \Bigg(\sum_{d=i+1-j}^{i+k-j} \varphi_d\Bigg) Z_j\\
&\qquad\qquad > k\theta(1+\varepsilon) \text{ for some } x a_n \le k \le n \text{ and some } i = 0, \dots, n-k\Bigg)\\
&\qquad + P(a_n^{-1} M_n \le y, \text{ and for some } j = \dots, -1, 0, 1, \dots,\\
&\qquad\qquad Z_j \varphi_{i-j} > y a_n (1+\varepsilon) \text{ for some } i = 1, \dots, n)\\
&\quad := P(A_1) + P(A_2) + P(A_3).
\end{aligned}
\tag{3.2}
$$

We immediately identify $P(A_1)$ as $P_M(\theta(1+\varepsilon), x, y(1+\varepsilon), n)$ in (2.22) and, hence, by Lemma 4.2,

$$
\begin{aligned}
\lim_{n\to\infty} P(A_1) = \exp\Big\{-(1+\varepsilon)^{-\alpha}\Big(&p \max(M_+(\varphi)^\alpha \theta^{-\alpha} x^{-\alpha}, m_+(\varphi)^\alpha y^{-\alpha})\\
&- q \max(M_-(\varphi)^\alpha \theta^{-\alpha} x^{-\alpha}, m_-(\varphi)^\alpha y^{-\alpha})\Big)\Big\}.
\end{aligned}
\tag{3.3}
$$

To estimate $P(A_2)$ in (3.2), let

$$
\begin{aligned}
T := \inf\Bigg\{&j = \dots, -1, 0, 1, \dots, \quad \Bigg(\sum_{d=i+1-j}^{i+k-j} \varphi_d\Bigg) Z_j > k\theta(1+\varepsilon)\\
&\text{for some } x a_n \le k \le n \text{ and some } i = 0, \dots, n-k\Bigg\}\\
= \inf\Bigg\{&j = \dots, -1, 0, 1, \dots, \quad \Bigg(\sum_{d=i_+(j)+1-j}^{i_+(j)+k_+(j)-j} \varphi_d\Bigg)_+ Z_j > k_+(j)\theta(1+\varepsilon)\\
&\text{or} \quad \Bigg(\sum_{d=i_-(j)+1-j}^{i_-(j)+k_-(j)-j} \varphi_d\Bigg)_- Z_j < -k_-(j)\theta(1+\varepsilon)\Bigg\},
\end{aligned}
$$

and note that, by Corollary 4.3, $T$ can take the value $\infty$ but not $-\infty$. We claim that

$$\lim_{n\to\infty} \sup_{j_0} P(A_2 | T = j_0) = 0, \tag{3.4}$$

which will obviously imply that

$$\lim_{n\to\infty} P(A_2) = 0. \tag{3.5}$$

Observe that, for any $j_0 = \dots, -1, 0, 1, \dots,$

$$\begin{aligned} P(A_2 | T = j_0) = {}& P\Biggl(a_n^{-1} R_n(\theta) \le x, \Biggl(\sum_{d=i_+(j_0)+1-j_0}^{i_+(j_0)+k_+(j_0)-j_0} \varphi_d\Biggr)_+ Z_{j_0} \\ &\qquad > k_+(j_0)\theta(1+\varepsilon) \Bigg| T = j_0\Biggr) \\ &+ P\Biggl(a_n^{-1} R_n(\theta) \le x, \Biggl(\sum_{d=i_-(j_0)+1-j_0}^{i_-(j_0)+k_-(j_0)-j_0} \varphi_d\Biggr)_- Z_{j_0} \\ &\qquad < -k_-(j_0)\theta(1+\varepsilon) \Bigg| T = j_0\Biggr) \\ := {}& P_{1,\,j_0}(A_2) + P_{2,\,j_0}(A_2). \end{aligned}$$

We will prove that

$$\lim_{n\to\infty} \sup_{j_0} P_{1,\,j_0}(A_2) = 0. \tag{3.6}$$

Since the corresponding statement for $P_{2,\,j_0}(A_2)$ can be proved similarly, this will be enough to establish (3.4).

Let $i_0 = i_+(j_0)$ and $k_0 = k_+(j_0)$. Then

$$\begin{aligned} P_{1,\,j_0}(A_2) \le {}& P\Biggl(\sum_{j=-\infty}^{\infty} Z_j \sum_{d=i_0+1-j}^{i_0+k_0-j} \varphi_d \le k_0\theta, \\ &\qquad \Biggl(\sum_{d=i_0+1-j_0}^{i_0+k_0-j_0} \varphi_d\Biggr)_+ Z_{j_0} > k_0\theta(1+\varepsilon) \Bigg| T = j_0\Biggr) \\ \le {}& P\Biggl(\sum_{j=j_0+1}^{\infty} Z_j \sum_{d=i_0+1-j}^{i_0+k_0-j} \varphi_d \le -\varepsilon k_0\theta/2\Biggr) \qquad (3.7) \\ &+ P\Biggl(\sum_{j=-\infty}^{j_0-1} Z_j \sum_{d=i_0+1-j}^{i_0+k_0-j} \varphi_d \le -\varepsilon k_0\theta/2 \Bigg| T = j_0\Biggr) \\ := {}& P_{11,\,j_0}(A_2) + P_{12,\,j_0}(A_2). \end{aligned}$$

For a $K \ge 1$, we denote by

$$\tilde{\varphi}_j = \varphi_j \mathbf{1}(|j| \le K), \qquad j = \dots, -1, 0, 1, \dots, \tag{3.8}$$

the coefficients truncated at $\pm K$, and write

$$\begin{aligned} P_{11,\, j_0}(A_2) \;\le\; & P\Biggl(\sum_{j=j_0+1}^{\infty} |Z_j| \sum_{d=i_0+1-j}^{i_0+k_0-j} |\varphi_d - \tilde{\varphi}_d| > \varepsilon k_0 \theta/4\Biggr) \\ & + P\Biggl(\sum_{j=j_0+1}^{\infty} Z_j \sum_{d=i_0+1-j}^{i_0+k_0-j} \tilde{\varphi}_d \le -\varepsilon k_0 \theta/4\Biggr) \\ := {} & P_{111,\, j_0}(A_2) + P_{112,\, j_0}(A_2). \end{aligned} \tag{3.9}$$

Clearly,

$$\begin{aligned} P_{111,\, j_0}(A_2) &\le \bigl(\varepsilon k_0 \theta/4\bigr)^{-1} E\Biggl(\sum_{j=j_0+1}^{\infty} |Z_j| \Biggl(\sum_{d=i_0+1-j}^{i_0+k_0-j} |\varphi_d - \tilde{\varphi}_d|\Biggr)\Biggr) \\ &\le k_0 E|Z_1| \Biggl(\sum_{j=K+1}^{\infty} |\varphi_j| + \sum_{j=-\infty}^{-K} |\varphi_j|\Biggr) \bigl(\varepsilon k_0 \theta/4\bigr)^{-1}, \end{aligned}$$

and so

$$\lim_{K\to\infty} \limsup_{n\to\infty} \sup_{j_0} P_{111,\, j_0}(A_2) = 0. \tag{3.10}$$

Furthermore,

$$\begin{aligned} P_{112,\, j_0}(A_2) &= P\Biggl(\sum_{d=-\infty}^{i_0-j_0+k_0-1} \tilde{\varphi}_d \Biggl(\sum_{j=d-k_0+1}^{d} Z_j\Biggr) \le -\varepsilon k_0 \theta/4\Biggr) \\ &= P\Biggl(\sum_{d=-K}^{\min(K,\, i_0-j_0+k_0-1)} \varphi_d \Biggl(\sum_{j=d-k_0+1}^{d} Z_j\Biggr) \le -\varepsilon k_0 \theta/4\Biggr) \\ &\le 2KP\Biggl(\frac{1}{k_0} \sum_{j=1}^{k_0} Z_j \le -\frac{\varepsilon\theta}{8K \max(|\varphi_j|)}\Biggr), \end{aligned}$$

and by the law of large numbers, we immediately have that

$$\lim_{n\to\infty} \sup_{j_0} P_{112,\, j_0}(A_2) = 0 \tag{3.11}$$

for every $K \ge 1$. It follows from (3.9)–(3.11) that

$$\lim_{n\to\infty} \sup_{j_0} P_{11,\, j_0}(A_2) = 0. \tag{3.12}$$

We treat the second term in the right-hand side of (3.7) similarly. Write

$$P_{12,\, j_0}(A_2) = P\Biggl(\sum_{j=-\infty}^{j_0-1} \widetilde{Z}_j \sum_{d=i_0+1-j}^{i_0+k_0-j} \varphi_d \le -\varepsilon k_0 \theta/2\Biggr), \tag{3.13}$$

where $\widetilde{Z}_j, \ldots, -1, 0, 1, \ldots$ are independent, and each $\widetilde{Z}_j$ has the law of $Z_j$ conditioned on belonging to the interval

$$S_j = \left[ -\frac{k_-(j)\theta(1+\varepsilon)}{\left(\sum_{d=i_-(j)+1-j}^{i_-(j)+k_-(j)-j} \varphi_d\right)_-},\ \frac{k_+(j)\theta(1+\varepsilon)}{\left(\sum_{d=i_+(j)+1-j}^{i_+(j)+k_+(j)-j} \varphi_d\right)_+} \right],$$

and note that

$$S_j \supset \left[ -\frac{a_n x \theta(1+\varepsilon)}{\sum_{d=-\infty}^{\infty} |\varphi_d|},\ \frac{a_n x \theta(1+\varepsilon)}{\sum_{d=-\infty}^{\infty} |\varphi_d|} \right] := S,$$

which is a set that does not depend on $j$ and increases, as $n \to \infty$, to the whole real line. With the truncated coefficients $(\tilde{\varphi}_j)$ as in (3.8), we have

$$\begin{aligned} P_{12,\, j_0}(A_2) \le\ & P\left( \sum_{j=-\infty}^{j_0-1} |\widetilde{Z}_j| \sum_{d=i_0+1-j}^{i_0+k_0-j} |\varphi_d - \tilde{\varphi}_d| > \varepsilon k_0 \theta/4 \right) \\ & + P\left( \sum_{j=-\infty}^{j_0-1} \widetilde{Z}_j \sum_{d=i_0+1-j}^{i_0+k_0-j} \tilde{\varphi}_d \le -\varepsilon k_0 \theta/4 \right) \\ := {} & P_{121,\, j_0}(A_2) + P_{122,\, j_0}(A_2). \end{aligned} \tag{3.14}$$

Since $E|\widetilde{Z}_j| \le E|Z_1|$ for all $j$, we conclude as above that

$$P_{121,\, j_0}(A_2) \le k_0 E|Z_1| \left( \sum_{j=K+1}^{\infty} |\varphi_j| + \sum_{j=-\infty}^{-K} |\varphi_j| \right) (\varepsilon k_0 \theta/4)^{-1},$$

implying that

$$\lim_{K\to\infty} \limsup_{n\to\infty} \sup_{j_0} P_{121,\, j_0}(A_2) = 0. \tag{3.15}$$

Furthermore,

$$\begin{aligned} P_{122,\, j_0}(A_2) &= P\left( \sum_{d=\max(-K,\, i_0-j_0+2)}^{K} \varphi_d \left( \sum_{j=d-k_0+1}^{d} \widetilde{Z}_{-j+i_0+1} \right) \le -\varepsilon k_0 \theta/4 \right) \\ &\le \sum_{d=-K}^{K} P\left( \frac{1}{k_0} \sum_{j=d-k_0+1}^{d} \widetilde{Z}_{-j+i_0+1} \le -\frac{\varepsilon\theta}{8K \max(|\varphi_j|)} \right). \end{aligned}$$

Note that (enlarging, if necessary, our probability space) we can construct a copy of the original noise sequence $(Z_j^*)$ such that, for every $j$,

$$\begin{aligned} E|Z_j^* - \widetilde{Z}_j| \le E(|Z_1|) \max\Big( & |P(Z_j > 0) - P(\widetilde{Z}_j > 0)|, \\ & |P(Z_j < 0) - P(\widetilde{Z}_j < 0)| \Big) \\ & + \big| E(|Z_1|) - E(|Z_1| | Z_1 \in S) \big| \le \rho_n \end{aligned}$$

for some $\rho_n \to 0$ as $n \to \infty$ that does not depend on $j$. Therefore, for every $d = -K, \dots, K$,

$$\begin{aligned}
P\left(\frac{1}{k_0}\sum_{j=d-k_0+1}^{d}\widetilde{Z}_{-j+i_0+1} \le -\frac{\varepsilon\theta}{8K\max(|\varphi_j|)}\right)
&\le P\left(\frac{1}{k_0}\sum_{j=1}^{k_0} Z_j \le -\frac{\varepsilon\theta}{16K\max(|\varphi_j|)}\right)\\
&\quad + P\left(\frac{1}{k_0}\sum_{j=d-k_0+1}^{d}|\widetilde{Z}_{-j+i_0+1} - Z^*_{-j+i_0+1}| > \frac{\varepsilon\theta}{16K\max(|\varphi_j|)}\right).
\end{aligned}$$

The first probability in the right-hand side above goes to zero by the law of large numbers, whereas the second probability is bounded from above by

$$\begin{aligned}
&\left(\frac{\varepsilon\theta}{16K\max(|\varphi_j|)}\right)^{-1}\frac{1}{k_0}\sum_{j=d-k_0+1}^{d} E|\widetilde{Z}_{-j+i_0+1} - Z^*_{-j+i_0+1}|\\
&\qquad \le \rho_n\left(\frac{\varepsilon\theta}{16K\max(|\varphi_j|)}\right)^{-1}.
\end{aligned}$$

Therefore,

$$\lim_{n\to\infty}\sup_{j_0} P_{122,\,j_0}(A_2) = 0 \tag{3.16}$$

for every $K \ge 1$. It follows from (3.14)–(3.16) that

$$\lim_{n\to\infty}\sup_{j_0} P_{12,\,j_0}(A_2) = 0, \tag{3.17}$$

and so (3.6) follows by (3.12) and (3.17) and, hence, (3.5) has been established as well.

The proof of

$$\lim_{n\to\infty} P(A_3) = 0 \tag{3.18}$$

is similar to the proof of (3.5), but is quite a bit simpler, hence omitted.

It follows from (3.2), (3.3), (3.5) and (3.18) that

$$\begin{aligned}
&\limsup_{n\to\infty} P\Big(a_n^{-1}R_n(\theta) \le x,\, a_n^{-1}M_n \le y\Big)\\
&\quad \le \exp\Big\{-(1+\varepsilon)^{-\alpha}\Big(p\max\Big(M_+(\varphi)^\alpha\theta^{-\alpha}x^{-\alpha},\, m_+(\varphi)^\alpha y^{-\alpha}\Big)\\
&\qquad\qquad - q\max\Big(M_-(\varphi)^\alpha\theta^{-\alpha}x^{-\alpha},\, m_-(\varphi)^\alpha y^{-\alpha}\Big)\Big)\Big\},
\end{aligned}$$

and, since $\varepsilon$ can be taken as small as we wish, we conclude that

$$\begin{aligned} &\limsup_{n\to\infty} P\Big(a_n^{-1}R_n(\theta) \le x, a_n^{-1}M_n \le y\Big) \\ &\quad \le \exp\Big\{-\Big(p\max\big(M_+(\varphi)^\alpha\theta^{-\alpha}x^{-\alpha}, m_+(\varphi)^\alpha y^{-\alpha}\big) \\ &\qquad\qquad - q\max\big(M_-(\varphi)^\alpha\theta^{-\alpha}x^{-\alpha}, m_-(\varphi)^\alpha y^{-\alpha}\big)\Big)\Big\}. \end{aligned} \tag{3.19}$$

We now proceed to establish a matching lower bound. Fix $\varepsilon \in (0,1)$, $N > \max(1, x)$ and $\delta > 0$ and write

$$\begin{aligned} &P\big(a_n^{-1}R_n(\theta) \le x, a_n^{-1}M_n \le y\big) \\ &\quad \ge P\Big(\big\{a_n^{-1}R_n(\theta) \le x, a_n^{-1}M_n \le y\big\} \bigcap B_1 \bigcap B_2\Big), \end{aligned} \tag{3.20}$$

where

$$\begin{aligned} B_1 = \Bigg\{ &\text{for all } j = \dots, -1, 0, 1, \dots, \quad \Bigg(\sum_{d=i+1-j}^{i+k-j} \varphi_d\Bigg) Z_j \le k\theta(1-\varepsilon) \\ &\text{for all } xa_n \le k \le n \text{ and all } i = 0, \dots, n-k, \\ &\text{and for all } j = \dots, -1, 0, 1, \dots, \quad Z_j\varphi_{i-j} \le ya_n(1-\varepsilon) \\ &\text{for all } i = 1, \dots, n \Bigg\} \end{aligned}$$

and

$$\begin{aligned} B_2 = \Bigg\{ &\text{for each } xa_n \le k < Na_n \text{ and each } i = 0, \dots, n-k, \\ &\Bigg(\sum_{d=i+1-j}^{i+k-j} \varphi_d\Bigg) Z_j > \delta a_n \text{ for at most one } j = \dots, -1, 0, 1, \dots, \\ &\text{and for each } i = 1, \dots, n, \quad Z_j\varphi_{i-j} > \delta a_n \\ &\text{for at most one } j = \dots, -1, 0, 1, \dots \Bigg\}. \end{aligned}$$

Therefore,

$$\begin{aligned} &P(a_n^{-1}R_n(\theta) \le x,\ a_n^{-1}M_n \le y) \\ &\quad \ge P(B_1) - P(B_2^c) - P\Big(\Big\{a_n^{-1}R_n(\theta) > x\Big\} \bigcap B_1 \bigcap B_2\Big) \\ &\qquad - P\Big(\Big\{a_n^{-1}M_n > y\Big\} \bigcap B_1 \bigcap B_2\Big). \end{aligned} \tag{3.21}$$

We identify $P(B_1)$ as $P_M(\theta(1-\varepsilon), x, y(1-\varepsilon), n)$ in (2.22), and so another appeal to Lemma 4.2 gives us

$$\begin{aligned}\lim_{n\to\infty} P(B_1) = \exp\Big\{-(1-\varepsilon)^{-\alpha}\Big(p\max\Big(M_+(\varphi)^\alpha\theta^{-\alpha}x^{-\alpha}, m_+(\varphi)^\alpha y^{-\alpha}\Big)\\ - q\max\Big(M_-(\varphi)^\alpha\theta^{-\alpha}x^{-\alpha}, m_-(\varphi)^\alpha y^{-\alpha}\Big)\Big)\Big\}.\end{aligned} \tag{3.22}$$

Furthermore, an application of Lemma 4.4 shows that

$$\lim_{n\to\infty} P\big(B_2^c\big) = 0. \tag{3.23}$$

Next, write

$$\begin{aligned}&P\Big(\Big\{a_n^{-1}R_n(\theta) > x\Big\}\bigcap B_1\bigcap B_2\Big)\\ &\quad\le P(a_n^{-1}R_n(\theta)\ge N) + P\Big(\Big\{x < a_n^{-1}R_n(\theta) < N\Big\}\bigcap B_1\bigcap B_2\Big).\end{aligned} \tag{3.24}$$

It follows from Lemma 4.8 that

$$\lim_{N\to\infty}\limsup_{n\to\infty} P(a_n^{-1}R_n(\theta)\ge N) = 0 \tag{3.25}$$

and, further,

$$\begin{aligned}&P\Big(\Big\{x < a_n^{-1}R_n(\theta) < N\Big\}\bigcap B_1\bigcap B_2\Big)\\ &\quad\le \sum_{k=[xa_n]+1}^{[Na_n]}\sum_{i=0}^{n-k} P\left(\left\{\sum_{j=-\infty}^{\infty}\left(\sum_{d=i+1-j}^{i+k-j}\varphi_d\right)Z_j > k\theta\right\}\bigcap B_1\bigcap B_2\right)\\ &\quad\le n\sum_{k=[xa_n]+1}^{[Na_n]} p_k,\end{aligned} \tag{3.26}$$

where

$$\begin{aligned}p_k = P\Bigg(&\sum_{j=-\infty}^{\infty}\left(\sum_{d=j+1}^{i+k}\varphi_d\right)Z_j > k\theta, \left(\sum_{d=j+1}^{j+k}\varphi_d\right)Z_j \le k\theta(1-\varepsilon)\\ &\text{for all } j = \dots,-1,0,1,\dots,\\ &\text{and for at most one } j, \quad \left(\sum_{d=j+1}^{j+k}\varphi_d\right)Z_j > k\delta/x\Bigg).\end{aligned}$$

We proceed similarly to the argument in the proof of (3.5). Let

$$T := \inf\left\{ j = \dots, -1, 0, 1, \dots, \quad \left(\sum_{d=j+1}^{j+k} \varphi_d\right) Z_j > k\delta/x \right\},$$

and note that, by Corollary 4.3, $T$ can take the value $\infty$ but not $-\infty$. We conclude that

$$p_k \le P\left( \sum_{j=-\infty}^{\infty} \left(\sum_{d=j+1}^{i+k} \varphi_d\right) Z_j > k\theta, \left(\sum_{d=j+1}^{j+k} \varphi_d\right) Z_j \le k\delta/x \right.$$
$$\left. \text{for all } j = \dots, -1, 0, 1, \dots, \right)$$
$$(3.27) \qquad + \sum_{i=-\infty}^{\infty} P(T=i) P\left( \sum_{j\ne i} \left(\sum_{d=j+1}^{i+k} \varphi_d\right) Z_j > k\theta\varepsilon, \right.$$
$$\left(\sum_{d=j+1}^{j+k} \varphi_d\right) Z_j \le k\delta/x$$
$$\left. \text{for all } j \ne i \,\middle|\, \left(\sum_{d=j+1}^{j+k} \varphi_d\right) Z_j \le k\delta/x \text{ for all } j < i \right).$$

Note that, by Lemma 4.6, the first probability in the right-hand side of (3.27) is bounded from above by a constant times $k^{-(2+\alpha)}$ as long as $\delta$ is chosen small enough comparatively to $\theta x$. Furthermore, it follows from (4.2) and (4.3) that, for every $i$, the second factor under the sum in the right-hand side of (3.27) is

$$\frac{P\left(\sum_{j\ne i}\left(\sum_{d=j+1}^{i+k} \varphi_d\right) Z_j > k\theta\varepsilon, \left(\sum_{d=j+1}^{j+k} \varphi_d\right) Z_j \le k\delta/x \text{ for all } j \ne i\right)}{P\left(\left(\sum_{d=j+1}^{j+k} \varphi_d\right) Z_j \le k\delta/x \text{ for all } j < i\right)}$$
$$\le CP\left( \sum_{j\ne i} \left(\sum_{d=j+1}^{i+k} \varphi_d\right) Z_j > k\theta\varepsilon, \left(\sum_{d=j+1}^{j+k} \varphi_d\right) Z_j \le k\delta/x \text{ for all } j \ne i \right)$$

for some positive constant $C$ that may depend on $\delta$, $x$ and $N$ but not on $i$ or $k$ in its range. Denoting by $r > 0$ the smaller of $P(Z_1 \ge 0)$ and $P(Z_1 \le 0)$ and using the reflection principle, we see that, for all $k$ large enough (indeed, $n$ large enough assures that $k$ is large enough, independently of $i$), the latter expression is further bounded from above by

$$C(2/r) P\left( \sum_{j=-\infty}^{\infty} \left(\sum_{d=j+1}^{i+k} \varphi_d\right) Z_j > k\theta\varepsilon, \left(\sum_{d=j+1}^{j+k} \varphi_d\right) Z_j \le k\delta/x \right.$$
$$\left. \text{for all } j = \dots, -1, 0, 1, \dots \right),$$

and by Lemma 4.6 the latter expression is bounded from above by a constant times $k^{-(2+\alpha)}$ as long as $\delta$ is chosen small enough compared to $\theta\varepsilon x$. We conclude that

$$p_k \le Ck^{-(2+\alpha)} \qquad \text{for all } xa_n < k \le Na_n,$$

for some constant $C > 0$, and we conclude by (3.26) that, for every $N > \max(1, x)$, $\varepsilon \in (0, 1)$ and $\delta > 0$ small enough (again compared to $\theta\varepsilon x$), we have

$$\lim_{n\to\infty} P\Big(\{x < a_n^{-1} R_n(\theta) < N\} \bigcap B_1 \bigcap B_2\Big) = 0. \tag{3.28}$$

It remains to treat the last term in the right-hand side of (3.21). We have

$$\begin{aligned}
&P\Big(\{a_n^{-1} M_n > y\} \bigcap B_1 \bigcap B_2\Big) \\
&\quad\le \sum_{i=1}^{n} P\Bigg(\sum_{j=-\infty}^{\infty} \varphi_{i-j} Z_j > ya_n, \varphi_{i-j} Z_j \le (1-\varepsilon) ya_n \text{ for all } j \\
&\qquad\qquad\qquad \text{and } \varphi_{i-j} Z_j \le \delta a_n \text{ for all } j \text{ except, perhaps, one } j\Bigg) \\
&\quad := \sum_{i=1}^{n} q_i.
\end{aligned} \tag{3.29}$$

Proceeding as above, we define

$$T := \inf\{ j = \dots, -1, 0, 1, \dots, \ \varphi_{i-j} Z_j > \delta a_n \},$$

note that, once again, $T$ can take the value $\infty$ but not $-\infty$, and write

$$\begin{aligned}
q_i &\le P\Bigg(\sum_{j=-\infty}^{\infty} \varphi_{i-j} Z_j > ya_n \text{ and } \varphi_{i-j} Z_j \le \delta a_n \text{ for all } j\Bigg) \\
&\quad + \sum_{i=-\infty}^{\infty} P(T = i) P\Bigg(\sum_{j\ne i} \varphi_{i-j} Z_j > \varepsilon ya_n \text{ and } \varphi_{i-j} Z_j \le \delta a_n \text{ for all } j \ne i\Bigg).
\end{aligned}$$

Repeating the argument used to prove (3.28) and appealing to Lemma 4.7, we see that $q_i \le Cn^{-2}$ for some $C > 0$ that does not depend on $i = 1, \dots, n$ as long as $\delta$ is small enough in comparison with $\varepsilon y$. Therefore, it follows from (3.29) that, for all such $\delta$ and $\varepsilon$,

$$\lim_{n\to\infty} P\Big(\{a_n^{-1} M_n > y\} \bigcap B_1 \bigcap B_2\Big) = 0. \tag{3.30}$$

We conclude by (3.20), (3.22), (3.23), (3.25), (3.28) and (3.30) that

$$\begin{aligned}\liminf_{n\to\infty} P\big(a_n^{-1}R_n(\theta) \le x, a_n^{-1}M_n \le y\big) \\ \ge \exp\Big\{-(1-\varepsilon)^{-\alpha}\Big(p\max(M_+(\varphi)^\alpha\theta^{-\alpha}x^{-\alpha},\ m_+(\varphi)^\alpha y^{-\alpha}) \\ -\ q\max(M_-(\varphi)^\alpha\theta^{-\alpha}x^{-\alpha},\ m_-(\varphi)^\alpha y^{-\alpha})\Big)\Big\},\end{aligned}$$

and, since $\varepsilon$ can be taken arbitrarily close to zero, this establishes the required lower bound matching (3.19). Hence, the proof of the theorem is now complete. □

The section concludes with the proof of Corollary 2.6.

PROOF. Since we are assuming that not all the coefficients $(\varphi_j)$ are equal to zero, the limiting distribution in Theorem 2.1 is a bivariate extreme value distribution (2.9) whose coordinates are strictly positive with probability 1, and whose spectral measure is a two-point measure given by (2.10)–(2.13). Therefore, by the continuous mapping theorem, $W_n(\theta)$ converges weakly to the ratio of the coordinates of the limiting bivariate extreme value distribution from Theorem 2.1. The form of the distribution of the latter ratio of the coordinates follows from Lemma 4.9. □

**4. Lemmas.** We start with a simple lemma dealing with certain tail probabilities.

LEMMA 4.1. *Let $Z$ be a random variable. Let $b_+(j,n)$, $j = \dots,-1,0,1,\dots$, $n = 1,2,\dots$, and $b_-(j,n)$, $j = \dots,-1,0,1,\dots$, $n = 1,2,\dots$, be two arrays of nonnegative numbers such that*

$$\lim_{n\to\infty}\inf_j b_+(j,n) = \infty, \qquad \lim_{n\to\infty}\inf_j b_-(j,n) = \infty,$$

$$\lim_{n\to\infty}\sum_{j=-\infty}^{\infty} P(Z > b_+(j,n)) = B_+$$

*and*

$$\lim_{n\to\infty}\sum_{j=-\infty}^{\infty} P(Z < -b_-(j,n)) = B_-.$$

*Then*

$$\lim_{n\to\infty}\prod_{j=-\infty}^{\infty} P(-b_-(j,n) \le Z \le b_+(j,n)) = \exp(-(B_+ + B_-)).$$

PROOF. Since

$$\prod_{j=-\infty}^{\infty} P(-b_-(j,n) \le Z \le b_+(j,n))$$
$$= \prod_{j=-\infty}^{\infty} (1 - P(Z > b_+(j,n)) - P(Z < -b_-(j,n))),$$

the claim of the lemma follows from the fact that $\log(1-x) \sim x$ as $x \to 0$. $\square$

Our next lemma describes the behavior of the probability $P_M(x, y, n)$ of the main event responsible for long strange intervals and high maxima given in the right-hand side of (2.22).

LEMMA 4.2. *Assume* (1.6). *Then for every $x > 0$ and $y > 0$,*

$$\lim_{n\to\infty} P_M(\theta, x, y, n) = \exp\{-p\max(M_+(\varphi)^\alpha \theta^{-\alpha} x^{-\alpha},\ m_+(\varphi)^\alpha y^{-\alpha}) - q\max(M_-(\varphi)^\alpha \theta^{-\alpha} x^{-\alpha},\ m_-(\varphi)^\alpha y^{-\alpha})\}.$$

PROOF. We start by rewriting $P_M(\theta, x, y, n)$ in a form similar to (2.4):

$$\begin{aligned} P_M(\theta, x, y, n) = \prod_{j=-\infty}^{\infty} P\Bigg(&-\min\Bigg(\theta\Bigg(\sup_{xa_n \le k \le n} \frac{1}{k} \sup_{i=0,\dots,n-k}\Bigg(\sum_{d=i+1-j}^{i+k-j} \varphi_d\Bigg)_-\Bigg)^{-1}, \\ &ya_n\Big(\sup_{i=1,\dots,n} (\varphi_{i-j})_-\Big)^{-1}\Bigg) \le Z_1 \\ &\le \min\Bigg(\theta\Bigg(\sup_{xa_n \le k \le n} \frac{1}{k} \sup_{i=0,\dots,n-k}\Bigg(\sum_{d=i+1-j}^{i+k-j} \varphi_d\Bigg)_+\Bigg)^{-1}, \\ &ya_n\Big(\sup_{i=1,\dots,n} (\varphi_{i-j})_+\Big)^{-1}\Bigg)\Bigg) \\ := \prod_{j=-\infty}^{\infty} P(&-b_-(j,n) \le Z_1 \le b_+(j,n)). \end{aligned} \tag{4.1}$$

It follows immediately from Lemma 4.1 and (1.6) that the claim of the lemma will follow once we check that

$$\lim_{n\to\infty} \sum_{j=-\infty}^{\infty} P(Z_1 > b_+(j,n)) = p\max(M_+(\varphi)^\alpha \theta^{-\alpha} x^{-\alpha},\ m_+(\varphi)^\alpha y^{-\alpha}) \tag{4.2}$$

and

$$\lim_{n\to\infty} \sum_{j=-\infty}^{\infty} P(Z_1 < -b_-(j,n)) = q\max(M_-(\varphi)^\alpha \theta^{-\alpha} x^{-\alpha},\ m_-(\varphi)^\alpha y^{-\alpha}). \tag{4.3}$$

Since the two statements are, obviously, of the same nature, we will only check (4.2).

Recalling the definition (2.5) of $a_n$, it follows from Potter's bounds [see, e.g., Resnick (1987), Proposition 0.8] and (1.6) that (4.2) will follows once we show that, for all $\beta$ in some neighborhood $(\alpha - \varepsilon, \alpha + \varepsilon)$, we have

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=-\infty}^{\infty} \left(\frac{b_+(j,n)}{a_n}\right)^{-\beta} = \max(M_+(\varphi)^\beta \theta^{-\beta} x^{-\beta},\ m_+(\varphi)^\beta y^{-\beta}), \tag{4.4}$$

which is what we now proceed to do. Obviously, we may assume that $\beta > 1$. The first step is the following statement:

$$\lim_{M\to\infty} \limsup_{n\to\infty} \frac{1}{n} \sum_{j=-\infty}^{-M} \left(\frac{b_+(j,n)}{a_n}\right)^{-\beta} = 0. \tag{4.5}$$

To this end, it is enough to prove two things:

$$\lim_{M\to\infty} \limsup_{n\to\infty} \frac{1}{n} \sum_{j=-\infty}^{-M} \left(a_n \left(\sup_{xa_n \le k \le n} \frac{1}{k} \sup_{i=0,\dots,n-k} \left(\sum_{d=i+1-j}^{i+k-j} \varphi_d\right)_+\right)\right)^{\beta} = 0 \tag{4.6}$$

and

$$\lim_{M\to\infty} \limsup_{n\to\infty} \frac{1}{n} \sum_{j=-\infty}^{-M} \left(\sup_{i=1,\dots,n} (\varphi_{i-j})_+\right)^{\beta} = 0. \tag{4.7}$$

For every $M$ and any $n \ge 1$,

$$\begin{aligned}
&\frac{1}{n} \sum_{j=-\infty}^{-M} \left(a_n \left(\sup_{xa_n \le k \le n} \frac{1}{k} \sup_{i=0,\dots,n-k} \left(\sum_{d=i+1-j}^{i+k-j} \varphi_d\right)_+\right)\right)^{\beta} \\
&\quad \le x^{-\beta} \frac{1}{n} \sum_{j=-\infty}^{-M} \left(\sum_{d=1-j}^{n-j} |\varphi_d|\right)^{\beta} \\
&\quad \le x^{-\beta} \left(\sum_{j=M+1}^{\infty} |\varphi_j|\right)^{\beta-1} \frac{1}{n} \sum_{j=-\infty}^{-M} \sum_{d=1-j}^{n-j} |\varphi_d| \\
&\quad \le x^{-\beta} \left(\sum_{d=M+1}^{\infty} |\varphi_d|\right)^{\beta},
\end{aligned}$$

and so (4.6) follows, and (4.7) is similar, but easier. Therefore, we have established (4.5).

The next step is the statement

$$\lim_{M\to\infty} \limsup_{n\to\infty} \frac{1}{n} \sum_{j=M+n}^{\infty} \left(\frac{b_+(j,n)}{a_n}\right)^{-\beta} = 0, \tag{4.8}$$

whose proof is entirely similar to that of (4.5) and, hence, we do not repeat the argument.

Next we check that, for all $M_1$, $M_2$,

$$\limsup_{n\to\infty} \frac{1}{n} \sum_{j=M_1}^{M_2+n} \left(\frac{b_+(j,n)}{a_n}\right)^{-\beta} \le \max\left(M_+(\varphi)^\beta \theta^{-\beta} x^{-\beta},\ m_+(\varphi)^\beta y^{-\beta}\right). \tag{4.9}$$

Observe that

$$\limsup_{k\to\infty} \sup_{-\infty<i<\infty} \left(\sum_{d=i+1}^{i+k} \varphi_d\right)_+ \le M_+(\varphi). \tag{4.10}$$

Indeed,

$$\lim_{k\to\infty} \inf_{-\infty<i<\infty} \max(i+k, -(i+1)) = \infty$$

and, for every $\delta > 0$, there is an $M(\delta)$ such that

$$\sup_{i_1<i_2,\ i_2>M(\delta)} \left(\sum_{d=i+1}^{i+k} \varphi_d\right)_+ \le \delta + \sup_{-\infty<k<\infty} \left(\sum_{j=k}^{\infty} \varphi_j\right)_+$$

and

$$\sup_{i_1<i_2,\ i_2<-M(\delta)} \left(\sum_{d=i+1}^{i+k} \varphi_d\right)_+ \le \delta + \sup_{-\infty<k<\infty} \left(\sum_{j=-\infty}^{k} \varphi_j\right)_+.$$

Therefore,

$$\limsup_{k\to\infty} \sup_{-\infty<i<\infty} \left(\sum_{d=i+1}^{i+k} \varphi_d\right)_+ \le \delta + M_+(\varphi),$$

and, since $\delta > 0$ is arbitrary, (4.10) follows. As a matter of fact, it is easy to check that the complete limit in the left-hand side of (4.10) exists and is equal to its right-hand side, but we will not use this fact. We immediately conclude from (4.10) that, given a $\delta > 0$ for all $n$ large enough,

$$\begin{aligned} &\frac{1}{n} \sum_{j=M_1}^{M_2+n} \left(\frac{b_+(j,n)}{a_n}\right)^{-\beta} \\ &\qquad \le \frac{M_2 - M_1 + 1 + n}{n} \max\left((\delta + M_+(\varphi))^\beta \theta^{-\beta} x^{-\beta},\ m_+(\varphi)^\beta y^{-\beta}\right), \end{aligned}$$

and (4.9) follows by letting first $n \to 0$ and then $\delta \to \infty$. It is now clear from (4.5), (4.8) and (4.9) that

$$\limsup_{n\to\infty} \frac{1}{n} \sum_{j=-\infty}^{\infty} \left(\frac{b_+(j,n)}{a_n}\right)^{-\beta} \le \max\left(M_+(\varphi)^\beta \theta^{-\beta} x^{-\beta},\ m_+(\varphi)^\beta y^{-\beta}\right). \tag{4.11}$$

We now get a matching lower bound. It is, obviously, enough to prove that

$$\liminf_{n\to\infty} \frac{1}{n} \sum_{j=-\infty}^{\infty} \left(\frac{b_+(j,n)}{a_n}\right)^{-\beta} \ge M_+(\varphi)^\beta \theta^{-\beta} x^{-\beta} \tag{4.12}$$

and

$$\liminf_{n\to\infty} \frac{1}{n} \sum_{j=-\infty}^{\infty} \left(\frac{b_+(j,n)}{a_n}\right)^{-\beta} \geq m_+(\varphi)^\beta y^{-\beta}. \tag{4.13}$$

Let us start with (4.12). We need to prove that, for every $-\infty < k < \infty$,

$$\liminf_{n\to\infty} \frac{1}{n} \sum_{j=-\infty}^{\infty} \left(\frac{b_+(j,n)}{a_n}\right)^{-\beta} \geq \left(\sum_{j=-\infty}^{k} \varphi_j\right)_+^\beta \theta^{-\beta} x^{-\beta} \tag{4.14}$$

and

$$\liminf_{n\to\infty} \frac{1}{n} \sum_{j=-\infty}^{\infty} \left(\frac{b_+(j,n)}{a_n}\right)^{-\beta} \geq \left(\sum_{j=k}^{\infty} \varphi_j\right)_+^\beta \theta^{-\beta} x^{-\beta}. \tag{4.15}$$

For (4.14), if $(\sum_{j=-\infty}^{k} \varphi_j)_+ = 0$, then there is nothing to prove, so we will assume that $(\sum_{j=-\infty}^{k} \varphi_j)_+ > 0$.

Given an $\varepsilon > 0$ small enough, choose a $J(\varepsilon)$ such that, for all $i \geq J(\varepsilon)$,

$$\sum_{j=-i}^{k} \varphi_j \geq (1-\varepsilon)\left(\sum_{j=-\infty}^{k} \varphi_j\right)_+.$$

We have

$$\begin{aligned} \frac{1}{n} \sum_{j=-\infty}^{\infty} \left(\frac{b_+(j,n)}{a_n}\right)^{-\beta} &\geq (1-o(1))\theta^{-\beta} x^{-\beta} \\ &\quad \times \frac{1}{n} \sum_{j=-\infty}^{\infty} \left(\sup_{i=0,\dots,n-\lceil x a_n\rceil} \left(\sum_{d=i+1-j}^{i+\lceil x a_n\rceil - j} \varphi_d\right)_+\right)^\beta \\ &\geq (1-o(1))\theta^{-\beta} x^{-\beta} \\ &\quad \times \frac{1}{n} \sum_{j=-k+\lceil x a_n\rceil}^{-k+n} \left(\sup_{i=0,\dots,n-\lceil x a_n\rceil} \left(\sum_{d=i+1-j}^{i+\lceil x a_n\rceil - j} \varphi_d\right)_+\right)^\beta \\ &\geq (1-o(1))\theta^{-\beta} x^{-\beta} \\ &\quad \times \frac{1}{n} \sum_{j=-k+\lceil x a_n\rceil}^{-k+n} \left(\sum_{d=k+1-\lceil x a_n\rceil}^{k} \varphi_d\right)^\beta + . \end{aligned} \tag{4.16}$$

Here, as usual, $\lceil a\rceil$ is the smallest integer greater than or equal to $a$. Now, for all $n$ so large that

$$k+1-\lceil x a_n\rceil \leq -J(\varepsilon),$$

the latter expression is at least

$$(1-o(1))\theta^{-\beta} x^{-\beta} \frac{n-\lceil x a_n\rceil + 1}{n} (1-\varepsilon)\left(\sum_{j=-\infty}^{k} \varphi_j\right)_+^\beta,$$

and letting first $n \to \infty$ and then $\varepsilon \to 0$, we obtain (4.14).

For (4.15), the argument is similar. If $(\sum_{j=k}^{\infty} \varphi_j)_+ = 0$, then there is nothing to prove, so we will assume that $(\sum_{j=k}^{\infty} \varphi_j)_+ > 0$. Given and $\varepsilon > 0$ small enough, choose a $J(\varepsilon)$ such that, for all $i \geq J(\varepsilon)$,

$$\sum_{j=k}^{i} \varphi_j \geq (1-\varepsilon)\Bigl(\sum_{j=k}^{\infty} \varphi_j\Bigr)_+.$$

We have, continuing after the first line in (4.16),

$$\begin{aligned}
\frac{1}{n}\sum_{j=-\infty}^{\infty}\left(\frac{b_+(j,n)}{a_n}\right)^{-\beta} &\geq (1-o(1))\theta^{-\beta}x^{-\beta} \\
&\quad\times\frac{1}{n}\sum_{j=-k+1}^{-k+1+n-\lceil xa_n\rceil}\left(\sup_{i=0,\dots,n-\lceil xa_n\rceil}\left(\sum_{d=i+1-j}^{i+\lceil xa_n\rceil-j}\varphi_d\right)_+\right)^{\beta} \\
&\geq (1-o(1))\theta^{-\beta}x^{-\beta} \\
&\quad\times\frac{1}{n}\sum_{j=-k+1}^{-k+1+n-\lceil xa_n\rceil}\left(\sum_{d=k}^{k-1+\lceil xa_n\rceil}\varphi_d\right)_+^{\beta}.
\end{aligned}$$

Once again, for all $n$ so large that

$$k-1+\lceil xa_n\rceil \geq J(\varepsilon),$$

the latter expression is at least

$$(1-o(1))\theta^{-\beta}x^{-\beta}\frac{n-\lceil xa_n\rceil+1}{n}(1-\varepsilon)\Bigl(\sum_{j=k}^{\infty}\varphi_j\Bigr)_+,$$

and (4.15) follows by letting first $n \to \infty$ and then $\varepsilon \to 0$.

This proves (4.12). Since the proof of (4.13) is similar, but easier, we omit it. □

The following is an immediate conclusion of the preceding lemma.

Corollary 4.3. *Assume* (1.6). *Then, for every $x > 0$ and $y > 0$,*

$$\begin{aligned}
P\Biggl(&\left(\sum_{d=i_+(j)+1-j}^{i_+(j)+k_+(j)-j}\varphi_d\right)_+ Z_j > k_+(j)\theta \\
&\textit{or}\ \left(\sum_{d=i_-(j)+1-j}^{i_-(j)+k_-(j)-j}\varphi_d\right)_- Z_j < -k_-(j)\theta \\
&\textit{for infinitely many}\ j=\dots,-1,0,1,\dots\Biggr) = 0,
\end{aligned} \tag{4.17}$$

*at least for all $n$ large enough, where $k_+(j)$, $k_-(j)$, $i_+(j)$ and $i_-(j)$ are defined in* (3.1).

PROOF. By the Kolmogorov 0–1 law, the probability in the right-hand side of (4.17) is equal to 0 or 1. However, by Lemma 4.2, this probability is strictly less than 1, at least for large $n$. Hence the conclusion. □

The next lemma makes precise in the present general framework the intuition about separation of large values of the noise discussed following (2.3).

LEMMA 4.4. *Assume* (1.6). *For all* $M \geq 1$ *and* $\delta > 0$,

$$\lim_{n\to\infty} P\Bigg(\text{for some } k = 1, \dots, [Ma_n] \text{ and some } i = 0, \dots, n-k, \quad \Bigg(\sum_{d=i+1-j}^{i+k-j} \varphi_d\Bigg) Z_j > \delta a_n \text{ for at least two different } j = \dots, -1, 0, 1, \dots\Bigg) = 0 \tag{4.18}$$

*and*

$$\lim_{n\to\infty} P\Big(\text{for some} \quad i = 1, \dots, n,\ Z_j \varphi_{i-j} > \delta a_n \text{ for at least two different } j = \dots, -1, 0, 1, \dots\Big) = 0. \tag{4.19}$$

PROOF. As usual, the proofs of (4.18) and (4.19) are similar, and the proof of the latter is quite a bit easier than that of the former. Nonetheless, for a demonstration, we will give an argument for both. We will start with an easier statement. Observe that the probability in the left-hand side of (4.19) is bounded from above by

$$\sum_{i=1}^{n} \sum_{j_1=-\infty}^{\infty} \sum_{j_2=-\infty}^{\infty} P\bigg(|Z_1| > \frac{\delta a_n}{|\varphi_{i-j_1}|}\bigg) P\bigg(|Z_1| > \frac{\delta a_n}{|\varphi_{i-j_2}|}\bigg).$$

Choose an $\varepsilon \in (0, \alpha - 1)$ and use Potter's bounds once again to conclude that the latter quantity is further bounded from above by

$$C(\varepsilon)\frac{1}{n^2} \sum_{i=1}^{n} \sum_{j_1=-\infty}^{\infty} \sum_{j_2=-\infty}^{\infty} |\varphi_{i-j_1}|^{\alpha-\varepsilon} |\varphi_{i-j_2}|^{\alpha-\varepsilon} = \frac{C(\varepsilon)}{n} \Bigg(\sum_{j=-\infty}^{\infty} |\varphi_j|^{\alpha-\varepsilon}\Bigg)^2 \to 0$$

as $n \to \infty$. Here $C(\varepsilon)$ is a finite positive constant. This proves (4.19).

For (4.18), we first truncate the range of the coefficients $(\varphi_j)$, and for a $K \geq 1$, define $(\tilde{\varphi}_j)$ as in (3.9). The probability in the left-hand side of (4.18) is bounded from above by

$$
\begin{aligned}
&P\Bigg(\text{for some } k=1,\dots,[Ma_n] \text{ and some } i=0,\dots,n-k,\\
&\qquad \Bigg(\sum_{d=i+1-j}^{i+k-j}\tilde{\varphi}_d\Bigg)Z_j>\delta a_n/2 \text{ for at least}\\
&\qquad \text{two different } j=\dots,-1,0,1,\dots\Bigg) \qquad (4.20)\\
&+P\Bigg(\text{for some } k=1,\dots,[Ma_n],\ \text{some } i=0,\dots,n-k\\
&\qquad \text{and some } j=\dots,-1,0,1,\dots,\Bigg(\sum_{d=i+1-j}^{i+k-j}(\varphi_d-\tilde{\varphi}_d)\Bigg)Z_j>\delta a_n/2\Bigg).
\end{aligned}
$$

The first probability in (4.20) is bounded from above by

$$
\begin{aligned}
\sum_{j_1=-\infty}^{\infty}\sum_{j_1=-\infty}^{\infty}P\Bigg(&\Bigg(\sum_{d=i+1-j_1}^{i+k-j_1}\tilde{\varphi}_d\Bigg)Z_1>\delta a_n/2,\\
&\Bigg(\sum_{d=i+1-j_2}^{i+k-j_2}\tilde{\varphi}_d\Bigg)Z_2>\delta a_n/2 \text{ for some} \qquad (4.21)\\
&k=1,\dots,[Ma_n] \text{ and some } i=0,\dots,n-k\Bigg).
\end{aligned}
$$

Now, in order for the probability under the sum in (4.21) to be different from zero, one must have $i+1-j_1\le K$ for some $i=0,\dots,n-1$, which requires $j_1\ge -K$. Similarly, one must have $i+[Ma_n]-j_1\ge -K$ for some $i=0,\dots,$ $n-1$, which requires $j_1\le n+Ma_n+K$. The same argument shows that we must have $-K\le j_2\le n+Ma_n+K$. Moreover, since for some $i=0,\dots,n-1$, we must have both $i+1-j_1\le K$ and $i+[Ma_n]-j_2\ge -K$, it follows that $j_1-j_2\le 2K+Ma_n$ and, since the roles of $j_1$ and $j_2$ are interchangeable, we have $|j_1-j_2|\le 2K+Ma_n$.

The above discussion shows that the expression in (4.21) is bounded from above by

$$
\begin{aligned}
&\sum_{j_1=-K}^{n+[Ma_n]+K}\ \sum_{j_2=-K}^{n+[Ma_n]+K}\mathbf{1}(|j_1-j_2|\le 2K+Ma_n)\ \Bigg(P\Bigg(\Bigg(\sum_{j=-\infty}^{\infty}|\varphi_j|\Bigg)|Z_1|>\delta a_n/2\Bigg)\Bigg)^2\\
&\quad\le\frac{C}{n^2}\sum_{j_1=-K}^{n+[Ma_n]+K}\ \sum_{j_2=-K}^{n+[Ma_n]+K}\mathbf{1}(|j_1-j_2|\le 2K+Ma_n)\\
&\quad\le\frac{C}{n^2}(n+2K+Ma_n)(2K+Ma_n)\to 0,
\end{aligned}
$$

as $n \to \infty$ because $a_n = o(n)$. Here $C$ is a finite positive constant that changes from place to place. Therefore, to complete the proof of the lemma, we only need to show that the upper limit of the second probability in (4.20) can be made arbitrarily small by choosing $K$ large enough. However, choosing an $\varepsilon \in (0, \alpha - 1)$ and using Potter's bounds, the latter probability can be bounded from above by

$$\begin{aligned}
&\sum_{j=-\infty}^{\infty} P\Bigg(\sup_{k=1,\dots,[Ma_n]} \sup_{i=0,\dots,n-k} \sum_{d=i+1-j}^{i+k-j} |\varphi_d - \tilde{\varphi}_d||Z_1| > \delta a_n/2\Bigg) \\
&\qquad \le \frac{C(\varepsilon)}{n} \sum_{j=-\infty}^{\infty} \Bigg(\sup_{k=1,\dots,[Ma_n]} \sup_{i=0,\dots,n-k} \sum_{d=i+1-j}^{i+k-j} |\varphi_d - \tilde{\varphi}_d|\Bigg)^{\alpha-\varepsilon}.
\end{aligned} \tag{4.22}$$

Now, for every $j$, we have

$$\sup_{k=1,\dots,[Ma_n]} \sup_{i=0,\dots,n-k} \sum_{d=i+1-j}^{i+k-j} |\varphi_d - \tilde{\varphi}_d| \le \sum_{d=-\infty}^{\infty} |\varphi_d - \tilde{\varphi}_d| = \sum_{|d|>K} |\varphi_d|,$$

and the same argument as above shows that the terms in the sum in the right-hand side of (4.22) are equal to zero for $j$ not in the interval $[1-K, n+[Ma_n]+K]$. Therefore, the expression in the right-hand side of (4.22) can be bounded from above by

$$C(\varepsilon)\frac{n+[Ma_n]+2K}{n} \sum_{|d|>K} |\varphi_d| \to C(\varepsilon) \sum_{|d|>K} |\varphi_d|$$

as $n \to \infty$, and the latter expression can be made arbitrarily small by choosing $K$ large enough. This completes the proof of the lemma. $\square$

Our next lemma provides bounds on the Laplace transform of certain random variables. The result is probably already known, but we were unable to find an appropriate reference.

LEMMA 4.5. *Let $X$ be a zero mean random variable such that, for some $B \in \mathbb{R}$, $C > 0$ and $\alpha > 1$, $X \ge B$ a.s. and, for every $x > 0$,*

$$P(X > x) \le CX^{-\alpha}.$$

*Then there is $D > 0$ that depends only on $B, C$ and $\alpha$ such that, for all $0 < \gamma \le 1$,*

$$Ee^{-\gamma X} \le e^{D\gamma^\alpha} \qquad \textit{if } 1 < \alpha < 2, \tag{4.23}$$

$$Ee^{-\gamma X} \le e^{D\gamma^2} \qquad \textit{if } \alpha > 2 \tag{4.24}$$

*and*

$$Ee^{-\gamma X} \le e^{D\gamma^2 \log(e+1/\gamma)} \qquad \textit{if } \alpha = 2. \tag{4.25}$$

PROOF. The proof is standard. Let $F$ denote the law of $X$. We have

$$\begin{aligned} Ee^{-\gamma X} &= 1 + \int_B^\infty (e^{-\gamma x} - 1 + \gamma x) F(dx) \\ &= 1 + \int_B^{1/\gamma} (\ ) + \int_{1/\gamma}^\infty (\ ) \colon 1 + I_1(\gamma) + I_2(\gamma). \end{aligned}$$

Now,

$$|e^{-x} - 1 + x| \le cx^2 \qquad \text{for all } x \ge B,$$

where $c$ is positive constant that depends only on $B$ (and in the sequel such a constant will be allowed to change). Therefore, for all $0 < \gamma \le 1$,

$$|I_1(\gamma)| \le c\gamma^2 \left(1 + \int_0^{1/\gamma} xP(X > x)\,dx\right).$$

Let, say, $1 < \alpha < 2$. We immediately conclude that, for all $\gamma \le 1$,

$$|I_1(\gamma)| \le c\gamma^\alpha,$$

where this time $c$ is a positive constant that may depend on $B$, $C$ and $\alpha$. Since, for all $x > 0$, $|e^{-x} - 1| \le x$, we conclude that, for all $\gamma \le 1$,

$$|I_2(\gamma)| \le 2\gamma \int_{1/\gamma}^\infty xF(dx) \le c\gamma^\alpha$$

if $1 < \alpha < 2$. Therefore,

$$Ee^{-\gamma X} \le 1 + D\gamma^\alpha \le e^{D\gamma^\alpha}$$

for all $\gamma \le 1$.

The case $\alpha \ge 2$ is similar. $\square$

The next lemma shows that it is very unlikely to have a long strange interval without a significant contribution from a single noise variable. Note that its conclusion does not require the full strength of the assumption (1.6).

LEMMA 4.6. *Assume that the coefficients* $(\varphi_j)$ *satisfy the assumption*

$$\sum_{j=-\infty}^{\infty} |\varphi_j|^{\alpha'} < \infty \tag{4.26}$$

*for some* $\alpha' < \min(2, \alpha)$. *Then, for very* $\theta > 0$ *and* $\varepsilon > 0$,

$$\begin{aligned} \lim_{k\to\infty} k^p \sup_{-\infty<i<\infty} P\Bigg( & X_{i+1} + \cdots + X_{i+k} \ge k\theta, \\ & \left(\sum_{d=i+1-j}^{i+k-j} \varphi_d\right) Z_j \le \varepsilon k\theta \text{ for all } j = \ldots, -1, 0, 1 \ldots \Bigg) = 0, \end{aligned} \tag{4.27}$$

*for every*

$$p < \frac{1}{2\varepsilon}\left(\frac{\min(2,\alpha)}{\alpha'} - 1\right). \tag{4.28}$$

PROOF. Denote

$$b_{i,k} = \sum_{d=i+1}^{i+k} \varphi_d$$

and observe that, for every $i$,

$$\begin{aligned}
&P\Big(X_{i+1} + \cdots + X_{i+k} \geq k\theta,\ b_{i-j,k}Z_j \leq \varepsilon k\theta \text{ for all } j = \ldots, -1, 0, 1, \ldots\Big)\\
&\quad = P\Big(\sum_{j=-\infty}^{\infty} b_{i-j,k}Z_j \geq k\theta, b_{i-j,k}Z_j \leq \varepsilon k\theta \text{ for all } j = \ldots, -1, 0, 1, \ldots\Big)\\
&\quad \leq P\Big(\sum_{j\in J^+} b_{i-j,k}Z_j \geq k\theta/2,\ b_{i-j,k}Z_j \leq \varepsilon k\theta \text{ for all } j \in J^+\Big)\\
&\qquad + P\Big(\sum_{j\in J^-} b_{i-j,k}Z_j \geq k\theta/2,\ b_{i-j,k}Z_j \leq \varepsilon k\theta \text{ for all } j \in J^-\Big)\\
&\quad := P_+(i,k) + P_-(i,k),
\end{aligned}$$

where $J^+$ is the collection of $j$ for which $b_{i-j,k} > 0$, and $J^-$ is the collection of $j$ for which $b_{i-j,k} < 0$. We will prove that

$$\lim_{k\to\infty} k^p \sup_{-\infty<i<\infty} P_+(i,k) = 0 \tag{4.29}$$

for every $p$ as in (4.28). Since the corresponding statement for $P_-(i,k)$ can be proved in a similar manner, the result of the lemma will then follow.

Fix for a moment $\delta \in (0,1)$ and write

$$\begin{aligned}
P_+(i,k) \leq\ & P\Big(\sum_{j\in J^+} (-b_{i-j,k}(Z_j)_- \mathbf{1}(b_{i-j,k}Z_j \leq \varepsilon k\theta)\\
&\qquad + b_{i-j,k}E((Z_j)_- \mathbf{1}(b_{i-j,k}Z_j \leq \varepsilon k\theta)) \geq k\delta\theta/2\Big)\\
&+ P\Big(\sum_{j\in J^+} \Big(b_{i-j,k}(Z_j)_+ \mathbf{1}(b_{i-j,k}Z_j \leq \varepsilon k\theta)\\
&\qquad - b_{i-j,k}E((Z_j)_+ \mathbf{1}(b_{i-j,k}Z_j \leq \varepsilon k\theta))\Big) \geq k(1-\delta)\theta/2\Big)\\
:=\ & P_{1,+}(i,k) + P_{2,+}(i,k).
\end{aligned} \tag{4.30}$$

Now, for any $\gamma > 0$, by Markov's inequality,

$$\begin{aligned} P_{1,+}(i,k) &= P\bigg(\sum_{j\in J^+}\big(-b_{i-j,k}(Z_j)_- + b_{i-j,k}E\big((Z_j)_-\big)\big) \geq k\delta\theta/2\bigg) \\ &\leq \exp\bigg\{-\gamma\frac{k\delta\theta}{2}\bigg\}\prod_{j\in J^+} E\exp\big(\gamma b_{i-j,k}(E(Z_1)_- - (Z_1)_-)\big). \end{aligned} \tag{4.31}$$

Observe that, for every $i$ and $k$,

$$\begin{aligned} |b_{i,k}| &\leq k^{1-1/\alpha'}\bigg(\sum_{d=i+1}^{i+k}|\varphi_d|^{\alpha'}\bigg)^{1/\alpha'} \\ &\leq k^{1-1/\alpha'}\bigg(\sum_{d=-\infty}^{\infty}|\varphi_d|^{\alpha'}\bigg)^{1/\alpha'}. \end{aligned} \tag{4.32}$$

Suppose, for example, that $\alpha \leq 2$. Take $\beta \in (\alpha', \alpha)$, and select $\gamma = k^{-h}$ with

$$\frac{\beta/(\beta-1)}{\alpha'/(\alpha'-1)} < h < 1.$$

Note that then $h > 1-1/\alpha'$, and so we may apply Lemma 4.5 to conclude that, for all $k$ large enough, all $j \in J^+$, and all $i$,

$$E\exp\big(\gamma b_{i-j,k}(E(Z_1)_- - (Z_1)_-)\big) \leq \exp\{C\gamma^\beta|b_{i-j,k}|^\beta\}$$

for some absolute constant $C > 0$, that may, in the sequel, change. We conclude by (4.31) that

$$P_{1,+}(i,k) \leq \exp\bigg\{-\gamma\frac{k\delta\theta}{2} + C\gamma^\beta\sum_{j\in J^+}|b_{i-j,k}|^\beta\bigg\}.$$

Now, by (4.32),

$$\begin{aligned} \sum_{j\in J^+}|b_{i-j,k}|^\beta &\leq k^{\beta(1-1/\alpha')}\sum_{j=-\infty}^{\infty}\bigg(\sum_{d=i+1-j}^{i+k-j}|\varphi_d|^{\alpha'}\bigg)^{\beta/\alpha'} \\ &\leq Ck^{\beta(1-1/\alpha')}\sum_{j=-\infty}^{\infty}\bigg(\sum_{d=i+1-j}^{i+k-j}|\varphi_d|^{\alpha'}\bigg) \\ &= Ck^{1+\beta(1-1/\alpha')}, \end{aligned} \tag{4.33}$$

and so

$$P_{1,+}(i,k) \leq \exp\bigg\{-k^{1-h}\frac{\delta\theta}{2} + Ck^{1+\beta(1-1/\alpha')-\beta h}\bigg\}.$$

However, by the choice of $h$,

$$1-h > 1+\beta(1-1/\alpha') - \beta h,$$

and so for all $k$ large enough, all $i$,

$$P_{1,+}(i,k) \le \exp\left\{-k^{1-h}\frac{\delta\theta}{3}\right\}.$$

That is,

$$\lim_{k\to\infty} k^p \sup_{-\infty<i<\infty} P_{1,+}(i,k) = 0 \tag{4.34}$$

for every $p$ as in (4.28).

In the case $\alpha > 2$, (4.34) follows from Lemma 4.5 in the same way.

We now consider $P_{2,+}(i,k)$ in (4.30). Denote

$$\begin{aligned} W_j = {} & b_{i-j,k}(Z_j)_+\mathbf{1}(b_{i-j,k}Z_j \le \varepsilon k\theta) \\ & -b_{i-j,k}E\Big((Z_j)_+\mathbf{1}(b_{i-j,k}Z_j \le \varepsilon k\theta)\Big), \qquad j \in J^+, \end{aligned}$$

and observe that $(W_j)$ are independent zero mean random variables such that $|W_j| \le \varepsilon k\theta$ for all $j$. We may, therefore, apply an inequality by Petrov [see Petrov (1995), 2.6.1 on page 77, or Mikosch and Samorodnitsky (2000), Lemma A3.6] to conclude that

$$\begin{aligned} P_{2,+}(i,k) &= P\bigg(\sum_{j\in J^+} W_j \ge k(1-\delta)\theta/2\bigg) \\ &\le \exp\left\{-\frac{1-\delta}{2\varepsilon}\operatorname{arcsinh}\left(\frac{k^2(1-\delta)\varepsilon\theta^2}{4\operatorname{Var}\big(\sum_{j\in J^+} W_j\big)}\right)\right\}. \end{aligned} \tag{4.35}$$

Suppose that $1 < \alpha < 2$. We have, for a $\beta \in (\alpha', \alpha)$,

$$\operatorname{Var} W_j \le b^2_{i-j,k}E\big((Z_1)^2_+\mathbf{1}(b_{i-j,k}Z_1 \le \varepsilon k\theta)\big) \le Ck^{2-\beta}|b_{i-j,k}|^\beta$$

for some $C > 0$ that depends only on $\beta$, $\varepsilon$ and $\theta$ but not on $i$, $j$ or $k$, and that is allowed to change from time to time. Hence

$$\begin{aligned} \operatorname{Var}\bigg(\sum_{j\in J^+} W_j\bigg) &= \sum_{j\in J^+}\operatorname{Var} W_j \\ &\le Ck^{2-\beta}\sum_{j\in J^+}|b_{i-j,k}|^\beta \\ &\le Ck^{2-\beta}k^{1+\beta(1-1/\alpha')} \\ &= Ck^{3-\beta/\alpha'} \end{aligned}$$

by (4.33). Therefore,

$$\begin{aligned} P_{2,+}(i,k) &\le \exp\left\{-\frac{1-\delta}{2\varepsilon}\operatorname{arcsinh}\big(Ck^{\beta/\alpha'}\big)\right\} \\ &\le C\exp\Big\{-(\beta/\alpha'-1)(1-\delta)(2\varepsilon)^{-1}\log k\Big\} \\ &= Ck^{-(\beta/\alpha'-1)(1-\delta)(2\varepsilon)^{-1}}, \end{aligned}$$

and since $\beta$ can be taken arbitrarily close to $\alpha$ and $\delta$ can be taken arbitrarily close to zero, we conclude that

$$\lim_{k\to\infty} k^p \sup_{-\infty<i<\infty} P_{2,+}(i,k) = 0 \tag{4.36}$$

for every $p$ as in (4.28). The case $\alpha \geq 2$ is similar.

Now (4.29) follows from (4.34) and (4.36). □

An identical argument establishes a similar statement for the maximum of the process. We do not repeat the proof.

LEMMA 4.7. *Assume that the coefficients* $(\varphi_j)$ *satisfy* (4.26). *Then, for every* $y > 0$ *and* $\varepsilon > 0$,

$$\begin{aligned}\lim_{k\to\infty} k^p P\Big(\max_{i=1,\dots,n} X_i > ya_n \text{ and } Z_j\varphi_{i-j} \leq \varepsilon y a_n \text{ for all } i=1,\dots,n \\ \text{and all } j=\dots,-1,0,1,\dots\Big) = 0\end{aligned} \tag{4.37}$$

*for every*

$$p < \frac{\min(2,\alpha)}{2\varepsilon}. \tag{4.38}$$

Our next result shows that the length of the longest strange interval cannot grow at the rate higher than $a_n$ in (2.5).

LEMMA 4.8. *Assume* (1.6). *Then*

$$\lim_{M\to\infty} \limsup_{n\to\infty} P(R_n(\theta) > Ma_n) = 0. \tag{4.39}$$

PROOF. Fix for a moment $\varepsilon > 0$ and observe that, for all $M \geq 1$,

$$\begin{aligned} P(R_n(\theta) > Ma_n) &\leq P\Bigg(\sup_{-\infty<j<\infty} \sup_{Ma_n<k\leq n} \frac{1}{k} \sup_{i=0,\dots,n-k} \Bigg(\sum_{d=i+1-j}^{i+k-j} \varphi_d\Bigg) Z_j > \varepsilon\Bigg) \\ &\quad + P\Bigg(R_n(\theta) > Ma_n \text{ and } \Bigg(\sum_{d=i+1-j}^{i+k-j} \varphi_d\Bigg) Z_j \leq \varepsilon k \text{ for all} \\ &\qquad j = \dots, -1, 0, 1, \dots, \quad \text{for all } Ma_n < k \leq n \\ &\qquad \text{and} \quad i = 0, \dots, n-k\Bigg). \end{aligned} \tag{4.40}$$

It follows immediately from Lemma 4.2 that

$$\lim_{M\to\infty} \lim_{n\to\infty} P\Bigg(\sup_{-\infty<j<\infty} \sup_{Ma_n<k\leq n} \frac{1}{k} \sup_{i=0,\dots,n-k} \Bigg(\sum_{d=i+1-j}^{i+k-j} \varphi_d\Bigg) Z_j > \varepsilon\Bigg) = 0,$$

and so it only remains to consider the second term in the right-hand side of (4.40). Notice that the latter term is bounded from above by

$$\sum_{k=\lceil a_n \rceil}^{n} \sum_{i=0}^{n-k} P\Bigg( X_{i+1} + \cdots + X_{i+k} \geq k\theta, \Bigg( \sum_{d=i+1-j}^{i+k-j} \varphi_d \Bigg) Z_j \leq \varepsilon k \quad \text{for all } j = \ldots, -1, 0, 1, \ldots \Bigg),$$

and, applying Lemma 4.6, we see that the above quantity can be further bounded from above by

$$n \sum_{k=\lceil a_n \rceil}^{n} k^{-C/\varepsilon},$$

where $C$ is a finite positive constant. Therefore, if $\varepsilon$ is small enough, then the second term in the right-hand side of (4.40) goes to zero as well, and we are done. □

Our final lemma deals with the distribution of the ratio of the coordinates of a random vector with a particular bivariate extreme value distribution.

LEMMA 4.9. *Let* $(X, Y)$ *be a random vector with a bivariate extreme value distribution* (2.9), *for which the spectral measure* $m$ *is a two-point measure*

$$m = w\beta\delta_{\mathbf{h}^{(1)}} + w(1-\beta)\delta_{\mathbf{h}^{(2)}}, \tag{4.41}$$

*where* $w > 0$, $0 < \beta \leq 1$, *and* $\mathbf{h}^{(1)} = (h_{11}, h_{12})$ *and* $\mathbf{h}^{(2)} = (h_{21}, h_{22})$ *belong to* $B_2^+$ *and have strictly positive coordinates. Let* $W = X/Y$.

(i) *Let* $0 < \beta < 1$. *If*

$$\frac{h_{11}}{h_{12}} \geq \frac{h_{21}}{h_{22}}, \tag{4.42}$$

*then the distribution of* $W$ *is concentrated on the interval* $[h_{21}/h_{22}, h_{11}/h_{12}]$, *and*

$$\begin{aligned} P\bigg(W = \frac{h_{21}}{h_{22}}\bigg) &= \frac{1}{1 + (\beta/(1-\beta))(h_{11}/h_{21})^{\alpha}}, \\ P\bigg(W = \frac{h_{11}}{h_{12}}\bigg) &= \frac{1}{1 + ((1-\beta)/\beta)(h_{22}/h_{12})^{\alpha}}, \end{aligned} \tag{4.43}$$

*while on the interval* $(h_{21}/h_{22}, h_{11}/h_{12})$ *the law of* $W$ *is absolutely continuous with the density*

$$f_W(w) = \frac{((1-\beta)/\beta)(h_{22}/h_{11})^{\alpha}\alpha w^{\alpha-1}}{\big(1 + ((1-\beta)/\beta)(h_{22}/h_{11})^{\alpha} w^{\alpha}\big)^2}. \tag{4.44}$$

*If, on the other hand,*

$$\frac{h_{11}}{h_{12}} < \frac{h_{21}}{h_{22}}, \tag{4.45}$$

*then the distribution of* $W$ *is concentrated on the interval* $[h_{11}/h_{12}, h_{21}/h_{22}]$, *and*

$$\begin{aligned} P\left(W = \frac{h_{11}}{h_{12}}\right) &= \frac{1}{1 + ((1-\beta)/\beta)(h_{21}/h_{11})^{\alpha}}, \\ P\left(W = \frac{h_{21}}{h_{22}}\right) &= \frac{1}{1 + (\beta/(1-\beta))(h_{12}/h_{22})^{\alpha}}, \end{aligned} \tag{4.46}$$

*while on the interval* $(h_{11}/h_{12}, h_{21}/h_{22})$ *the law of* $W$ *is absolutely continuous with the density*

$$f_W(w) = \frac{(\beta/(1-\beta))(h_{12}/h_{21})^{\alpha} \alpha w^{\alpha-1}}{\left(1 + (\beta/(1-\beta))(h_{12}/h_{21})^{\alpha} w^{\alpha}\right)^2}. \tag{4.47}$$

(ii) *Let* $\beta = 1$. *Then* $W = h_{11}/h_{12}$ *with probability* 1.

PROOF. Let $E_1$ and $E_2$ be two independent standard exponential random variables. It is straightforward to check by a direct computation that the vector $(X, Y)$ has the same distribution as the vector

$$\max\left(\beta^{1/\alpha} E_1^{-1/\alpha} \mathbf{h}^{(1)}, (1-\beta)^{1/\alpha} E_2^{-1/\alpha} \mathbf{h}^{(2)}\right),$$

where the maximum is taken component by component. The immediate conclusion is that the law of $W = X/Y$ is concentrated between the smaller and the bigger of $h_{11}/h_{12}$ and $h_{21}/h_{22}$ and, if $\beta = 1$, then $W = h_{11}/h_{12}$ with probability 1.

Suppose now that $0 < \beta < 1$. We have

$$\begin{aligned} P\left(W = \frac{h_{11}}{h_{12}}\right) &= P\Bigg(\beta^{1/\alpha} E_1^{-1/\alpha} h_{11} > (1-\beta)^{1/\alpha} E_2^{-1/\alpha} h_{21}, \\ &\qquad\qquad \beta^{1/\alpha} E_1^{-1/\alpha} h_{12} > (1-\beta)^{1/\alpha} E_2^{-1/\alpha} h_{22}\Bigg) \\ &= P\left(\frac{E_2}{E_1} > \frac{1-\beta}{\beta} \max\left(\frac{h_{21}}{h_{11}}, \frac{h_{22}}{h_{12}}\right)^{\alpha}\right) \\ &= \left(1 + \frac{1-\beta}{\beta} \max\left(\frac{h_{21}}{h_{11}}, \frac{h_{22}}{h_{12}}\right)^{\alpha}\right)^{-1}, \end{aligned}$$

while

$$P\left(W=\frac{h_{21}}{h_{22}}\right)=P\left(\frac{E_1}{E_2}>\frac{\beta}{1-\beta}\max\left(\frac{h_{11}}{h_{21}},\frac{h_{12}}{h_{22}}\right)^{\alpha}\right)$$
$$=\left(1+\frac{\beta}{1-\beta}\max\left(\frac{h_{11}}{h_{21}},\frac{h_{12}}{h_{22}}\right)^{\alpha}\right)^{-1}.$$

In particular, under (4.42) we obtain (4.43), and under the assumption (4.45) we obtain (4.46).

Under the assumption (4.42) for any $h_{21}/h_{22} < a < b < h_{11}/h_{12}$, we have

$$P(a < W \le b) = P\left(a < \left(\frac{\beta}{1-\beta}\right)^{1/\alpha}\frac{h_{11}}{h_{22}}\left(\frac{E_2}{E_1}\right)^{1/\alpha} \le b\right),$$

from which we can read off the density of $W$ in the interior of the interval $(h_{21}/h_{22}, h_{11}/h_{12})$ as (4.44). Similarly, under the assumption (4.45) we obtain (4.47). This completes the proof. □

**5. Several data sets.** In this section we look at two financial data sets. Our analysis is, however, applicable to data of a different origin, in particular to data sets coming from communication networks applications.

The financial data sets we are using were provided by Olsen and Associates at the first HFDF conference, March 1995 in Zürich. The data contains foreign exchange returns for the period of October 1, 1992 through September 29, 1993. The first data set contains 51414 10-minute returns on US dollar against German mark, while the second data set contains 52376 returns on Japanese yen against German mark. Both return series are given in the $\theta$-time designed by the researchers at Olsen and Associates to deseasonalize the data; see Dacorogna et al. (1993).

The returns on US dollar against German mark are presented in Figure 5.1. Note that the Hill estimator of the tail index $\alpha$ [see, e.g., Resnick and Stărică (1995)] is between 2.5 and 3.0. We have chosen to plot the statistic $\theta W_n(\theta)$ for $W_n(\theta)$ defined in (2.6) against $\theta$, because by Theorem 2.1 the limiting distribution of this statistic is independent of $\theta$. Selecting the right range of $\theta$ is an important and difficult issue; it is somewhat similar to the issue of selecting the number of upper-order statistics while computing the Hill estimator [see, e.g., Drees, De Haan and Resnick (2000)]. The usual technique in the case of the Hill estimator is to look for the range of the number of upper-order statistics where the value of the estimator stabilizes. We do likewise with the estimator $\theta W_n(\theta)$, and this is how the range of $\theta$ is selected. It is interesting to try to judge the significance of the obtained results. Using the procedure in Example 2.7, we see that, at the significance level of .05, we should conclude that long-range dependence may be present once the value of our statistic is at least 19. Based on that approach, the absolute returns clearly demonstrate presence of long-range dependence.

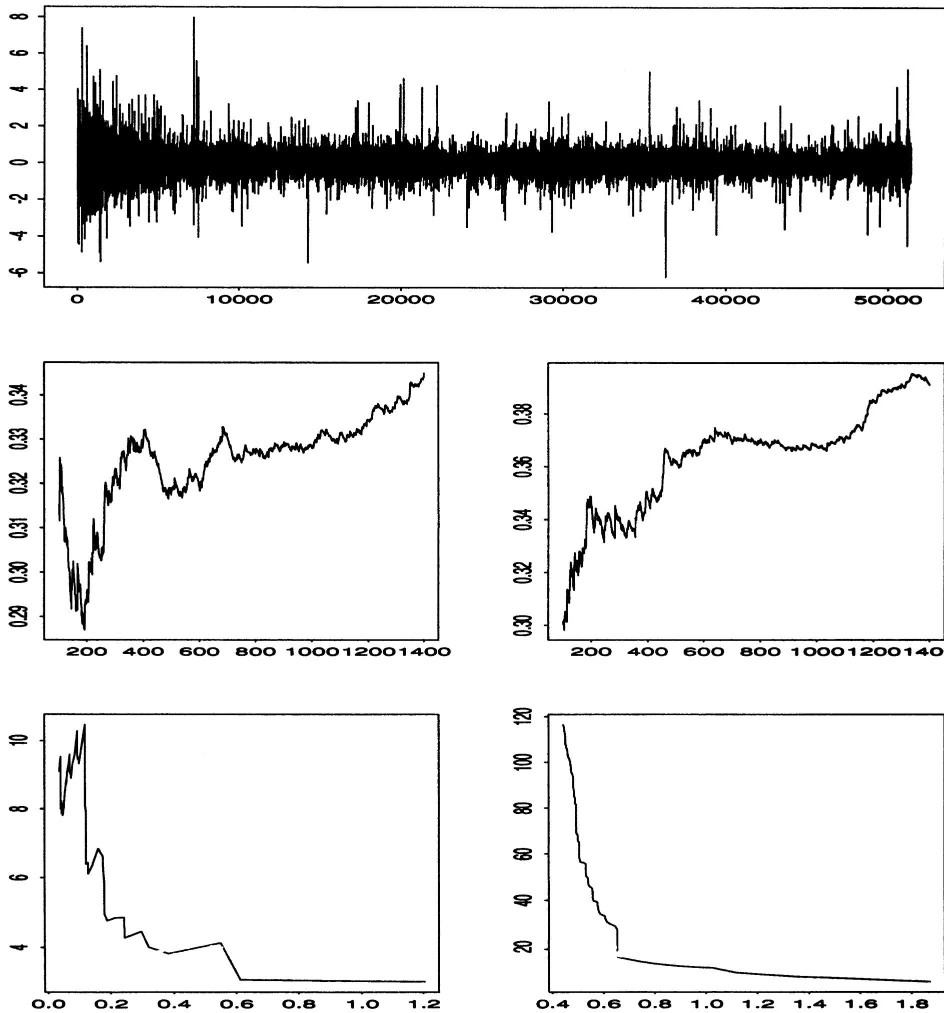


FIG. 5.1. *Ten-minute returns on US dollar versus German mark. The top plot is the time series plot of the returns. The two plots in the second row are the Hill plots (of $1/\alpha$) for the returns (the left plot) and absolute returns (the right plot) as a function of the number of upper-order statistics. The two plots in the third row are the values of the statistic $\theta W_n(\theta)$ over a range of $\theta$ for the returns (the left plot) and absolute returns (the right plot).*

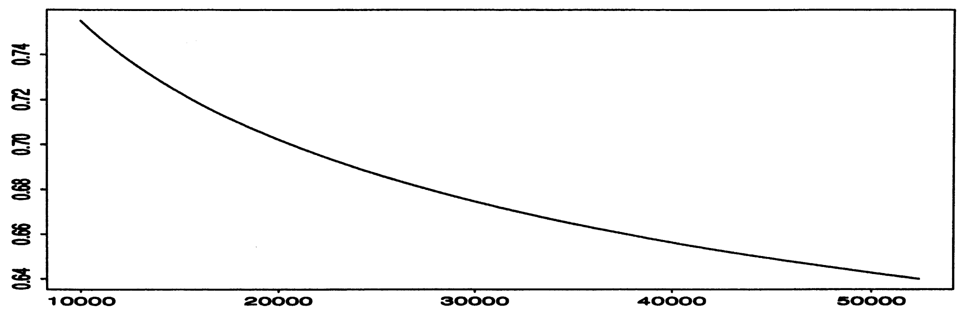


FIG. 5.2. *The values of* $\log R_n(\theta)/\log n$ *for* $\theta = 0.5$ *computed as a function of the sample size* $n$ *for* 10-*minute returns on US dollar versus German mark.*

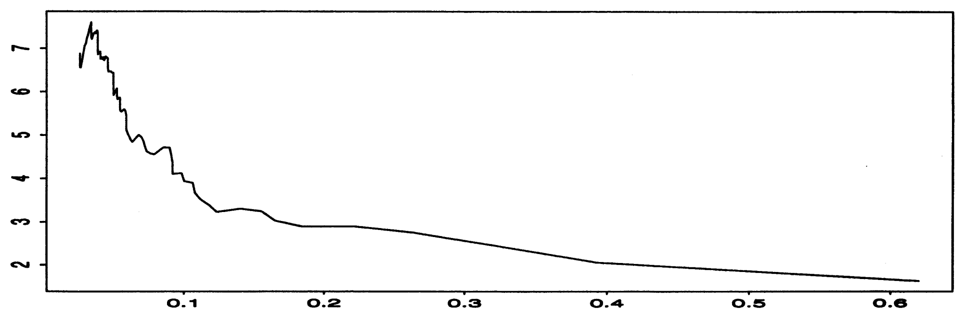


FIG. 5.3. *The values of the statistic* $\theta W_n(\theta)$ *over a range of* $\theta$ *for the randomized absolute* 10-*minute returns on US dollar versus German mark.*

The plot in the left column of the third row in Figure 5.1 does not provide a clear indication that long strange intervals for the raw returns on US dollar versus German mark grow faster than expected under short-range dependence. For illustration purposes, we include in Figure 5.2 the plot of $\log R_n(\theta)/\log n$ computed as a function of the sample size $n$ for $\theta = 0.5$. That is, we compute the statistic $R_n(\theta)$ based on the first $n$ observations in the data set. Observe that the range of values on this plot is quite a bit higher than the range of values of $1/\alpha$ on the Hill plot in the left column of the second row in Figure 5.1. According to the discussion before the statement of Theorem 2.1, this can be taken as informal evidence for presence of long-range dependence.

To check how much the significance of the observed values of the statistic $\theta W_n(\theta)$ owes, in practice, to dependence as opposed to, say, the tails, we took a random permutation of the absolute 10-minute returns on US dollar versus German mark and computed the statistic $\theta W_n(\theta)$ over a range of $\theta$ for the randomized absolute returns. The result is presented in Figure 5.3. Observe how low the values of the statistic are in comparison to the plot in the third row, right column, of Figure 5.1.

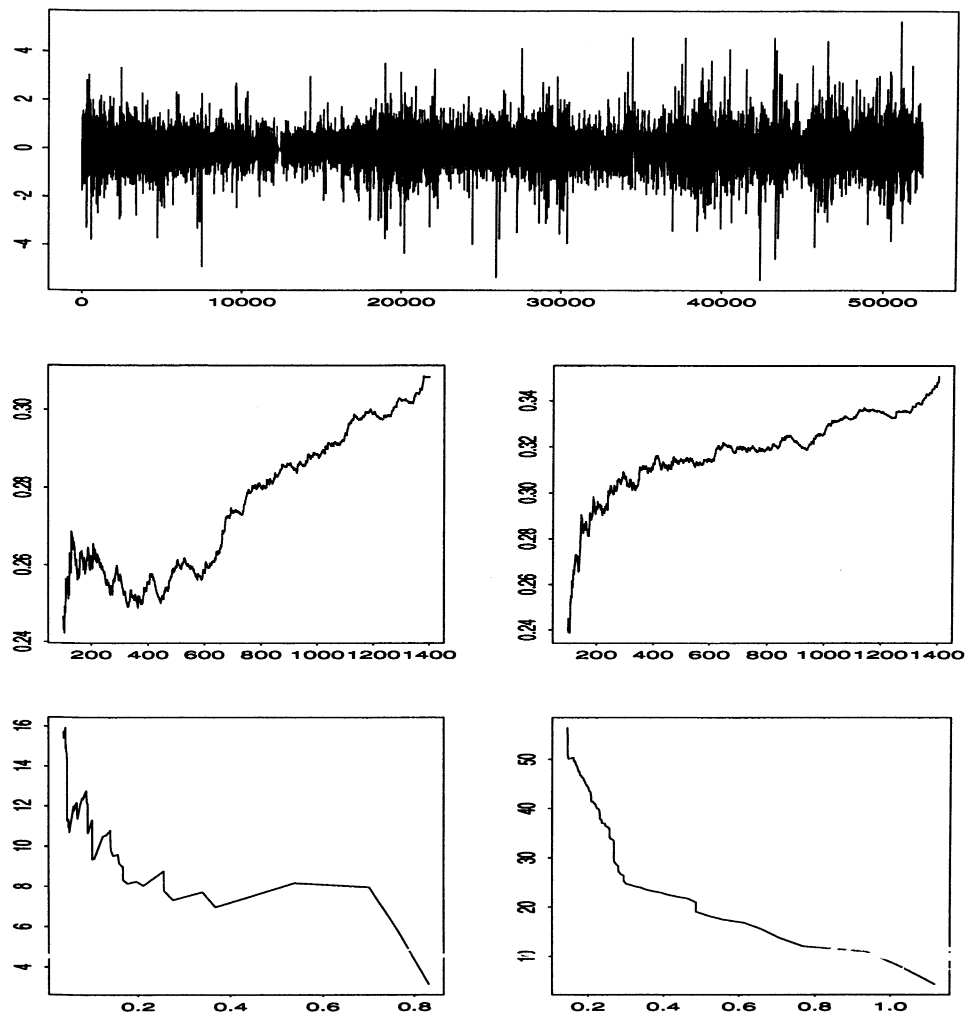


FIG. 5.4. *Ten-minute returns on Japanese yen versus German mark. The top plot is the time series plot of the returns. The two plots in the second row are the Hill plots (of $1/\alpha$) for the returns (the left plot) and absolute returns (the right plot) as a function of the number of upper-order statistics. The two plots in the third row are the values of the statistic $\theta W_n(\theta)$ over a range of $\theta$ for the returns (the left plot) and absolute returns (the right plot).*

We get a similar conclusion from the analysis of the returns on Japanese yen against German mark as presented in Figure 5.4. The Hill estimator of the tail index $\alpha$ is now between 3.5 and 4.0, and the absolute returns clearly demonstrate presence of long-range dependence.

## REFERENCES

BERAN, J. (1994). *Statistics for Long-Memory Processes*. Chapman and Hall, New York.

BERAN, J., SHERMAN, R., WILLINGER, W. and TAQQU, M. (1995). Long-range dependence in variable-bit-rate video traffic. *IEEE Trans. Comm.* **43** 1566–1579.

BHATTACHARYA, R., GUPTA, V. and WAYMIRE, E. (1983). The Hurst effect under trends. *J. Appl. Probab.* **20** 649–662.

BOLLERSLEV, T. and MIKKELSEN, H. (1996). Modeling and pricing long memory in stock market volatility. *J. Econometrics* **73** 151–184.

BREIDT, F., CRATO, N. and DE LIMA, P. (1996). The detection and estimation of long memory in stochastic volatility. *J. Econometrics* **83** 325–348.

BROCKWELL, P. and DAVIS, R. (1991). *Time Series: Theory and Methods*, 2nd ed. Springer, New York.

CROVELLA, M. and BESTAVROS, A. (1996). Self-similarity in World Wide Web traffic: evidence and possible causes. *Performance Evaluation Rev.* **24** 160–169.

DACOROGNA, M., MÜLLER, U., NAGLER, R., OLSEN, R. and PICTET, O. (1993) A geographical model for the daily and weekly seasonal volatility in the FX market. *J. Internat. Money Finance* **12** 413–438.

DEMBO, A. and ZEITOUNI, O. (1993). *Large Deviations Techniques and Applications*. Jones and Bartlett, Boston.

DING, Z., GRANGER, C. and ENGLE, R. (1993). A long memory property of stock market returns and a new model. *J. Empir. Finance* **1** 83–106.

DREES, H., DE HAAN, L. and RESNICK, S. (2000). How to make a Hill plot. *Ann. Statist.* **28** 254–274.

EMBRECHTS, P., KLÜPPELBERG, C. and MIKOSCH, T. (1997). *Modelling Extremal Events for Insurance and Finance*. Springer, Berlin.

HURST, H. (1951). Long-term storage capacity of reservoirs. *Trans. Amer. Soc. Civil Eng.* **116** 770–808.

LEADBETTER, M., LINDGREN, G. and ROOTZÉN, H. (1983). *Extremes and Related Properties of Random Sequences and Processes*. Springer, New York.

MANDELBROT, B. and VAN NESS, J. (1968). Fractional Brownian motions, fractional noises and applications. *SIAM Rev.* **10** 422–437.

MANDELBROT, B. and WALLIS, J. (1968). Noah, Joseph and operational hydrology. *Water Resources and Resevoirs* **4** 909–918.

MANDELBROT, B. and WALLIS, J. (1969a). Computer experiments with fractional Gaussian noises, Parts 1, 2, 3. *Water Resources and Resevoirs* **5** 228–267.

MANDELBROT, B. and WALLIS, J. (1969b). Robustness of the rescaled range R/S in the measurement of noncyclic long-run statistical dependence. *Water Resources and Resevoirs* **5** 967–988.

MANDELBROT, B. and WALLIS, J. (1969c). Some long-run properties of geophysical records. *Water Resources and Resevoirs* **5** 321–340.

MIKOSCH, T. and SAMORODNITSKY, G. (2000). The supremum of a negative drift random walk with dependent heavy-tailed steps. *Ann. Appl. Probab.* **10** 1025–1064.

MIKOSCH, T. and STĂRIČA, C. (1999). Change of structure in financial time series, long range dependence and the GARCH model. IWI Preprint 99-5-06, Univ. Groningen.

PETROV, V. (1995). *Limit Theorems of Probability Theory*. Oxford Univ. Press.

RESNICK, S. (1987). *Extreme Values, Regular Variation and Point Processes*. Springer, New York.

RESNICK, S. (1997). Why non-linearities can ruin the heavy tailed modeler's day. In *A Practical Guide to Heavy Tails: Statistical Techniques for Analysing Heavy Tailed Distributions* (R. Adler, R. Feldman and M. Taqqu, eds.) 219–240. Birkhäuser, Boston.

RESNICK, S. and Samorodnitsky, G. (1999). Activity periods of an infinite server queue and performance of certain heavy tailed fluid queues. *Questa* **33** 43–71.

RESNICK, S. and STĂRICĂ, C. (1995). Consistency of Hill estimator for dependent data. *J. Appl. Probab.* **32** 139–167.

TAQQU, M. (1987). Random processes with long-range dependence and high variability. *J. Geophys. Res.* **92** 9683–9686.

TAQQU, M. and TEVEROVSKY, V. (1998). On estimating the intensity of long-range dependence in finite and infinite variance time series. In *A Practical Guide to Heavy Tails: Statistical Techniques for Analysing Heavy Tailed Distributions* (R. Adler, R. Feldman and M. Taqqu, eds.) 177–218. Birkhäuser, Boston.

WILLINGER, W., TAQQU, M., LELAND, M. and WILSON, D. (1995). Self-similarity through high variability: statistical analysis of ethernet LAN traffic at the source level. *Computer Comm. Rev.* **25** 100–113.

SCHOOL OF ACCOUNTING AND FINANCE
UNIVERSITY OF TASMANIA
GPO BOX 252-86
HOBART, TASMANIA 7001
AUSTRALIA
AND
THEORY CENTER
628 RHODES HALL
CORNELL UNIVERSITY
ITHACA, NEW YORK 14853
E-MAIL: peter.mansfield@utas.edu.au,
peterm@tc.cornell.edu

INSTITUTE OF STATISTICS AND
MATHEMATICAL ECONOMICS
KOLLEGIUM AM SCHLOSS
UNIVERSITY OF KARLSRUHE
POSTFACH 6980
D-76128
KARLSRUHE
GERMANY
E-MAIL: zari.rachev@wiwi.uni-karlsruhe.de

SCHOOL OF OPERATIONS RESEARCH AND
INDUSTRIAL ENGINEERING AND
DEPARTMENT OF STATISTICAL SCIENCE
CORNELL UNIVERSITY
ITHACA, NEW YORK 14853
E-MAIL: gennady@orie.cornell.edu