The supremum of a negative drift random walk with dependent heavy-tailed steps

Mikosch, T; Smorodnitsky, G

Published in:
Annals of applied probability

The Annals of Applied Probability
2000, Vol. 10, No. 3, 1025–1064

THE SUPREMUM OF A NEGATIVE DRIFT RANDOM WALK WITH DEPENDENT HEAVY-TAILED STEPS

BY THOMAS MIKOSCH[1] AND GENNADY SAMORODNITSKY[1, 2]

University of Groningen and Cornell University

Many important probabilistic models in queuing theory, insurance and finance deal with partial sums of a negative mean stationary process (a negative drift random walk), and the law of the supremum of such a process is used to calculate, depending on the context, the ruin probability, the steady state distribution of the number of customers in the system or the value at risk. When the stationary process is heavy-tailed, the corresponding ruin probabilities are high and the stationary distributions are heavy-tailed as well. If the steps of the random walk are independent, then the exact asymptotic behavior of such probability tails was described by Embrechts and Veraverbeke. We show that this asymptotic behavior may be different if the steps of the random walk are not independent, and the dependence affects the joint probability tails of the stationary process. Such type of dependence can be modeled, for example, by a linear process.

1. Introduction. In various applied fields, such as insurance mathematics, queuing theory, finance and time series analysis among others, the model of a random walk with negative drift occurs in a natural way. For example, the probability of ruin in a homogeneous insurance portfolio can be written in terms of the distribution of the supremum of such a random walk; see Embrechts, Klüppelberg and Mikosch (1997) (Hereafter EKM), Chapter 1. The tail probability of solutions to stochastic recurrence equations, including the tails of ARCH and GARCH processes, can be obtained in a similar way; see EKM (1997), Section 8.4, and the references therein. The solution to the most important random recursion in queuing theory, the Lindley equation, is of the same form; see for instance Baccelli and Brémaud (1994). In the latter case the tail distribution of the stationary solution is often viewed as an overflow probability.

There exists extensive literature on the asymptotic behavior of the ruin probability and the tails of the stationary solutions to random recursions. Both the cases of light-tailed step distributions and heavy-tailed step distributions have been considered. Most of this literature deals with the "usual" random walk, which means iid steps. We refer the reader to EKM [(1997), Chapter 1] for the most important results and additional references. The basic result for heavy-tailed random walks with iid steps is due to Embrechts and

Received July 1998; revised August 1999.
[1]Research supported by NATO Collaborative Research Grant CRG 971579.
[2]Supported in part by NSF Grant DMS-97-04982 and NSA Grant MDA904-98-1-0041 at Cornell University.
AMS 1991 *subject classifications*. Primary 60F10, 60K30; Secondary 60G10, 60K25.
Key words and phrases. Random walk, stationary process, ruin probability, heavy tails, large deviations, random recursion, stationary queue, risk.

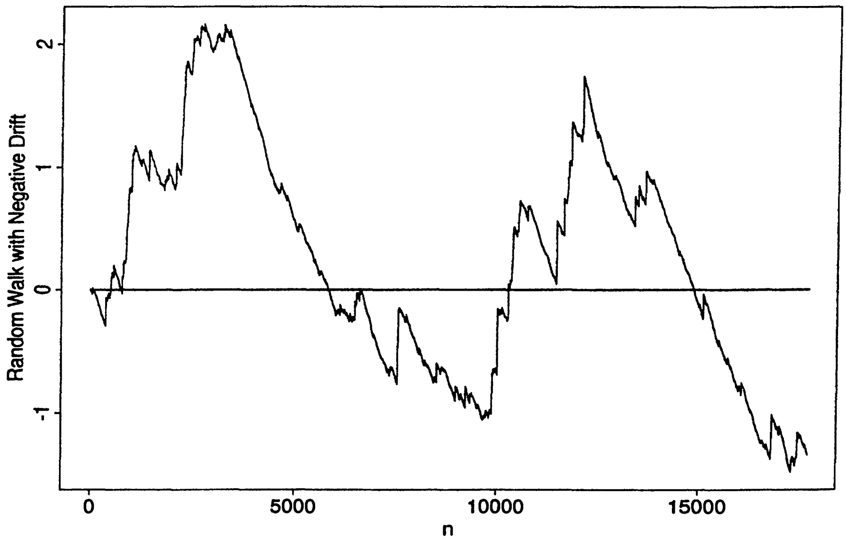

FIG. 1. *The dependent step random walk generated by teletraffic data; see Section* 3 *for a precise description of this set. These data are extremely heavy-tailed and dependent. The above computer graph shows the random walk* (S_n) *with mean* $-(1+0.05)\mu n$, *where* μ *is the estimated value of the expectation of the (positive) teletraffic data. The unit on the y-axis is* 10 *millions.*

Veraverbeke (1982); compare Theorem 1.3.6 in EKM. Let X_n, $n \in \mathbb{Z}$, be iid subexponential random variables [that is, $P(X_1 + X_2 > \lambda) \sim 2P(X_1 > \lambda)$ as $\lambda \to \infty$; see Chistyakov (1964)]. They generate the random walk

$$S_0 = 0, \qquad S_n = X_1 + \cdots + X_n, \qquad n \geq 1. \tag{1.1}$$

Let F denote the common law of the X_n's, and $-\mu < 0$ be the common negative mean. Then

$$P\Big(\sup_{n\geq 0} S_n > \lambda\Big) \sim \frac{1}{\mu}\int_\lambda^\infty (1 - F(x))\,dx \quad \text{as } \lambda \to \infty. \tag{1.2}$$

In most applications (except, perhaps, insurance) the assumption of independent step sizes is, clearly, unrealistic. For example, in the queuing context a typical model has steps distributed as the difference between service times and interarrival times of successive customers, and the independence assumption is universally believed not to hold. Rather, one hopes that the dependence existing in the data does not matter as far as quantities of interest, such as the ruin probability or the overflow probability, are concerned. Certain results available to date confirm this hope. For example, Asmussen, Schmidli and Schmidt (1999) show that the Embrechts and Veraverbeke result (1.2) remains valid (in the queuing context) under fairly general dependence structure of the interarrival times if the service times are still independent.

An important type of dependence is that of clustering of exceedances of high thresholds. This is a well-known phenomenon in econometric modeling where ARCH and GARCH types of models are commonly used for precisely that feature: data exhibit periods of high activity and low activity. We will show in this paper that this kind of dependence can result in a situation where the tail equivalence (1.2) is no longer valid; see the statement of Theorem 2.1 below and see relation (1.12).

In this paper we choose to model the steps X_n, $n \in \mathbb{Z}$, of the random walk as a two-sided linear process,

$$X_n = -\mu + \sum_{j=-\infty}^{\infty} \varphi_{n-j}\, \varepsilon_j, \qquad n \in \mathbb{Z}, \tag{1.3}$$

where $(\varepsilon_n)_{n\in\mathbb{Z}}$ is a sequence of zero mean iid random variables and $\mu > 0$ is a constant. Note that it is, actually, abuse of terminology to call the process $(S_n)_{n\geq 0}$ in (1.1) a random walk if the step sizes are not iid. We choose, however, to use this name because of its clear intuitive meaning, and we believe that no confusion will result. Notice that ARMA and fractional ARIMA processes have representation as one-sided, that is, causal, linear processes (i.e., $\varphi_n = 0$ for $n < 0$); see for example Brockwell and Davis (1991).

In this paper we assume that $\varepsilon = \varepsilon_0$ satifies the following regular variation and tail balance conditions:

$$\begin{cases} P(|\varepsilon| > \lambda) = L(\lambda)\lambda^{-\alpha}, \\ \lim\limits_{\lambda\to\infty} \dfrac{P(\varepsilon > \lambda)}{P(|\varepsilon| > \lambda)} = p, \qquad \lim\limits_{\lambda\to\infty} \dfrac{P(\varepsilon < -\lambda)}{P(|\varepsilon| > \lambda)} = q, \end{cases} \tag{1.4}$$

as $\lambda \to \infty$, for some $\alpha > 1$ and $0 < p \leq 1$. Here L is a slowly varying (at infinity) function. The coefficients φ_j, not all of which are equal to zero, are assumed to satisfy the following condition:

$$\sum_{j=-\infty}^{\infty} |j\, \varphi_j| < \infty. \tag{1.5}$$

A few remarks are, obviously, in order.

REMARK 1.1. Condition (1.5) excludes linear processes with *long-range dependence* which condition can be defined via $\sum_j |\varphi_j| = \infty$. Such a condition is fulfilled, for example for finite variance FARIMA(p, d, q) processes with $d \in (0, 0.5)$; see Brockwell and Davis [(1991), Section 13.2]. However, the weak dependence condition (1.5) is not uncommon in many results of time series analysis and trivially satisfied for causal invertible ARMA(p, q) processes [Brockwell and Davis (1991), Chapter 3]. If one departs from (1.5), proofs might become even more technical because (φ_j) can be close to being not absolutely summable. See for example Kokoszka and Taqqu (1996) in order to get some flavor of the difficulties one has to face. Moreover, a condition of type (1.5) is needed for our main result. To be more specific, the formulation of our main result [see (1.12)] involves the infinite series $\sum_j \varphi_j^{\pm}$ which have to be finite. Thus Theorem 2.1 is not applicable to FARIMA(p, d, q) processes of order $d \in (0, 0.5)$. Whether the additional $|j|$ in (1.5) can be avoided, and Theorem 2.1 be applied to, for example, FARIMA(p, d, q) processes with $d \in (-0.5, 0)$, is an open question. Finally, the reader will observe that, to the best of our knowledge, even results on the tail of a series with independent terms of the type presented in Lemma A.3 require, in general, conditions stronger than those needed for convergence of the series. By not striving to achieve the weakest possible conditions, one gains somewhat in the transparency of the results.

REMARK 1.2. There is well-founded skepticism about using heavy-tailed linear processes for probabilistic modeling. Indeed, in classical time series analysis the main attraction of using linear processes is the fact that their correlations (or spectra) are flexible enough to approximate the correlations (or spectrum) of an arbitrary second-order stationary process. However, correlations and spectra, even when defined, are not natural to concentrate on in the heavy-tailed case. In fact, sample autocorrelations of heavy-tailed linear processes can behave very differently from those of other important classes of heavy-tailed processes, and the autocorrelations in available data often do not support the assumption of a linear model. See, for example, Resnick (1997) and Resnick, Samorodnitsky and Xue (1999). However, we are NOT interested in correlations. Rather, we are interested in the tails, which is exactly the reason why heavy-tailed processes are important in the first place. Linear processes are well suited to model a great variety of dependence in the tails of stationary heavy-tailed processes. This means that heavy-tailed linear processes can be used to model the clusters of high-threshold exceedances by a dependent stationary sequence in terms of limiting compound Poisson processes. The description of the clustering behavior of dependent sequences is one of the keys to the understanding of their extremal behavior and related topics. See the discussion and references in EKM (1997), Sections 5.5 and 8.1.

REMARK 1.3. Random variables with regularly varying tails are also subexponential. We do not know if an appropriate analogue of our results holds when the ε_j's have a subexponential distribution. The argument of Embrechts and Veraverbeke (1982) for the supremum tail (1.2) in the case of iid steps requires Wiener–Hopf factorization and Markov property. We conjecture that the result holds in some form in the subexponential case. The argument we use is relatively easy to extend to bigger subclasses of the subexponential class of distributions (e.g., the distributions with so called intermediate regularly varying tails).

REMARK 1.4. Conditions (1.4), (1.5) and $E\varepsilon = 0$ imply that the infinite series in (1.3) converges absolutely with probability 1 and that $X = X_0$ has expectation $-\mu$. Furthermore, by Lemma A.3 in the Appendix,

$$\frac{P(X > \lambda)}{P(|\varepsilon| > \lambda)} \sim \sum_{j=-\infty}^{\infty} |\varphi_j|^\alpha (pI_{\{\varphi_j > 0\}} + qI_{\{\varphi_j < 0\}}) =: \|\varphi\|_\alpha^\alpha \qquad \text{as } \lambda \to \infty. \tag{1.6}$$

Observe that the (dependent step) random walk $(S_n)_{n\geq 0}$ has negative drift. Since (X_n) is mixing [see Rosenblatt (1962), page 112] this implies that $S_n/n \to -\mu$ a.s. In particular, $\sup_{n\geq 1} S_n < \infty$ a.s., and we will concentrate on

$$\psi(\lambda) = P\left(\sup_{n\geq 0} S_n > \lambda\right)$$

as $\lambda \to \infty$. If $(S_n)_{n\geq 0}$ had iid steps with the same marginal distribution [or even only the same negative mean and the same tail behavior as X has in (1.6)], then the Embrechts and Veraverbeke result (1.2) and Karamata's theorem [see Theorem 1.5.11 in Bingham, Goldie and Teugels (1987)] would show that

$$\psi_{\text{ind}}(\lambda) \sim \frac{1}{\mu(\alpha-1)}\lambda P(X > \lambda) \sim \frac{\|\varphi\|_\alpha^\alpha}{\alpha-1}\frac{1}{\mu}\lambda P(|\varepsilon| > \lambda) \tag{1.7}$$

as $\lambda \to \infty$. (We use the notation ψ_{ind} to remind us that we are dealing with iid steps.) We will see that in the case of dependent steps (1.7) is, in general, false.

The following heuristics give us a taste of what the true behavior of the tail $\psi(\lambda)$ may be. It also provides us with a road map of the proof in the next section. However, heuristics cannot replace the very technical proof; only its study will enable one to understand the complicated mechanism which causes the asymptotic behavior of the ruin probabilities to deviate from the iid case.

Because of the heavy tails, we expect the event $\{\sup_n S_n > \lambda\}$ for large λ to occur because of a single very large positive or very small negative value of the noise ε_n. The largest ever contribution of the "important" noise variables ε_j to the state of the random walk can be seen from the expression

$$\begin{aligned} S_n &= -n\mu + \sum_{k=1}^{n}\sum_{j=-\infty}^{\infty} \varepsilon_{k-j}\varepsilon_j \\ &= -n\mu + \sum_{j=-\infty}^{\infty} \varepsilon_j \sum_{k=1-j}^{n-j} \varphi_k. \end{aligned} \tag{1.8}$$

Let us concentrate first on the large positive values of the noise. A potentially large positive contribution of ε_j^+ to S_n is multiplied by $\sum_{k=1-j}^{n-j}\varphi_k$. Here, as usual, for any real number x,

$$x^+ = \max(0, x) \quad \text{and} \quad x^- = -\min(0, x).$$

When j is a very small negative number, this factor is by (1.5) small, uniformly (in n). We do not expect each individual ε_j^+ to make a sizable contribution to the tail of the process. Indeed, the tail $P(\varepsilon > \lambda)$ of each individual ε_j^+ is of a smaller order than that predicted either by the Embrechts and Veraverbeke result (1.7) or what we expect in (1.12) below. Furthermore, because of the negative drift, the contribution of each noise variable dissipates with time. It is now easy to convince ourselves that, if very small negative j's do not play an important role, and neither is this role played by any individual value of j, then the "important" noise variables ε_j are those with high j's, in which case the multiplicative factor of ε_j^+ becomes about $\sum_{k=-\infty}^{n-j} \varphi_k$, and the largest this factor can ever get over all possible n's (i.e., over all positions of the random walk) is

$$m_\varphi^+ = \sup_{-\infty<n<\infty} \sum_{k=-\infty}^{n} \varphi_k. \tag{1.9}$$

Clearly, the values of S_n in which ε_j^+ gets multiplied by this factor are those with n being about equal to j (simply because we choose n such that $n-j$ lies in a particular region), and because of ergodicity of the step sizes the random walk is at that time at about the level $-j\mu$. If we apply the same reasoning to the small negative values of the noise variables ε_j and use the notation

$$m_\varphi^- = \sup_{-\infty<n<\infty} \sum_{k=-\infty}^{n} (-\varphi_k), \tag{1.10}$$

we expect that the following asymptotic relation holds:

$$\begin{aligned}\psi(\lambda) &\sim \sum_{j=1}^{\infty} \Big(P(m_\varphi^+ \varepsilon_j^+ > \lambda + j\mu) + P(m_\varphi^- \varepsilon_j^- > \lambda + j\mu) \Big)\\ &\sim \int_1^\infty P(m_\varphi^+ \varepsilon^+ > \lambda + y\mu)\, dy + \int_1^\infty P(m_\varphi^- \varepsilon^- > \lambda + y\mu)\, dy\\ &\sim \frac{m_\varphi^+}{\mu} \int_{\lambda/m_\varphi^+}^\infty P(\varepsilon > y)\, dy + \frac{m_\varphi^-}{\mu} \int_{\lambda/m_\varphi^-}^\infty P(\varepsilon < -y)\, dy.\end{aligned} \tag{1.11}$$

Of course, the reason for adding up the probabilities above is that we do not expect more than one event in question to occur. Furthermore, because of our conclusion that individual values of the noise variables ε_j do not play an important role in the asymptotic behavior of $\psi(\lambda)$, we can start the summation at any given place; indeed, it is simple to check that the asymptotic behavior derived above does not depend on the position of the first term in the sum.

Applying once again Karamata's theorem, we expect then to have

$$\begin{aligned}\psi(\lambda) &\sim \frac{[p(m_\varphi^+)^\alpha + q(m_\varphi^-)^\alpha]}{\alpha - 1} \frac{1}{\mu} \lambda P(|\varepsilon| > \lambda)\\ &\sim \left[\frac{[p(m_\varphi^+)^\alpha + q(m_\varphi^-)^\alpha]}{\|\varphi\|_\alpha^\alpha} \right] \frac{1}{\mu(\alpha - 1)} \lambda P(X > \lambda),\end{aligned} \tag{1.12}$$

as $\lambda \to \infty$. Here we interpret the right-hand side as $o(\lambda P(X > \lambda))$ if the bracket vanishes. Notice that the bracket equals one if (X_n) is an iid sequence; compare (1.7). It is impossible to interpret the bracket in terms of known measures of dependence such as the autocorrelation function or the extremal index of linear processes; see, for example, EKM (1997), Sections 7.3 and 8.1. However, the bracket can be thought of as a means of describing the strength of the tail dependence in the random walk. Intuitively, the higher is, say, m_φ^+, the higher can be the contribution of a positive value of a given noise variable to the position of the random walk. Compare this to an example from a different universe: the variance of a sum of terms with equal variance each may be thought of as a measure of the dependence between the terms. Similarly, the higher m_φ^- is, the higher can be the contribution of a negative value of a given noise variable to the position of the random walk.

The limiting relation (1.12) will be proved in the next section. It is important to note that in the case $p = 1$ and $m_\varphi^+ = 0$ the tail of $\psi(\lambda)$ is of a smaller order than that promised by (1.7). This will be the case, for example, when $\varphi_0 = 1, \varphi_{-1} = -1$ and $\varphi_j = 0$ for $j \neq -1, 0$ (and $p = 1$). The true order of magnitude of the tail of $\psi(\lambda)$ will depend, in that case, on the relationship between the left and the right tails of the noise variables (that is, one needs information more precise than just $p = 1$). This point, however, is not pursued in this paper. We note that in this example (with $\varphi_0 = 1, \varphi_{-1} = -1$ and $\varphi_j = 0$ for $j \neq -1, 0$) the tail of $\psi(\lambda)$ is still of the same order as in (1.7) if $0 \le p < 1$. Intuitively, this happens because very small negative values of the noise variables ε_j get a chance to affect the position of the random walk before they get cancelled on the next step. This is not possible if $p = 1$, because then the noise variables are not as likely to take very small negative values.

On the other hand, in the case of a causal (i.e., one-sided) linear process for which $\varphi_j = 0$ for $j < 0$ and $\varphi_0 > 0$ we have

$$m_\varphi^+ \ge \varphi_0 > 0.$$

This implies (for $p > 0$) that the order of magnitude of $\psi(\lambda)$ cannot be smaller than in the iid case.

This paper is organized as follows. In the next section we prove (1.12), which is the main result of this paper. In Section 3 we perform an exploratory statistical analysis of a data set with file sizes requested via Internet. It is our intention to emphasize that this data set has heavy-tailed marginal distributions and lacks independence. Finally, in the Appendix we collect and prove some related results, dealing with the tail behavior of an infinite linear combination of random variables with regularly varying tails, and with large deviations of the partial sums of the infinite moving average (1.3). Although not all of these results are needed for the proof of (1.12), they provide additional information about sums and maxima of a linear process with regularly varying tails and might also be of independent interest.

2. The asymptotics of the ruin probability.

The following is the main result of this paper.

THEOREM 2.1. *Let (X_n) be a linear process* (1.3) *with a negative mean $-\mu$ and assume that the iid mean-zero noise sequence (ε_n) satisfies the regular variation and tail balance conditions* (1.4) *for some $\alpha > 1$. Moreover, suppose that the real coefficients φ_n satisfy* (1.5). *Then* (1.12) *holds.*

PROOF. Our argument frequently uses the notation

$$\beta_{nj} = \sum_{i=1-j}^{n-j} \varphi_i, \tag{2.1}$$

with which we can rewrite the representation (1.8) of the random walk in the form

$$S_n = -n\mu + \sum_{j=-\infty}^{\infty} \varepsilon_j \beta_{nj}, \qquad n \ge 0. \tag{2.2}$$

In what follows we prove several lemmas which involve *any* value of $\mu > 0$, and some of the lemmas can be formulated even for $\mu = 0$; but we omit further details. For the proof of the theorem, we will apply these lemmas *not necessarily* for the value μ in the formulation of the theorem; it will, however, become clear from the context *which* value of μ will be utilized in which lemma.

The result is proved by a series of technical lemmas. Before we start with a detailed analysis we give a short outline of the main ideas of the proof.

A. In Section 2.1 we start with proving Theorem 2.1 in the case $p = 1$, that is, when the right tail of the noise variables is fatter than the left tail. We further assume $m_\varphi^+ > 0$. Since we want to study the tail probability $P(\sup_{n\ge 0} S_n > \lambda)$, we have to find out for which values of n the random walk S_n is closest to its supremum and which summands $\varepsilon_j\beta_{nj}$ in (2.2) make a main contribution to S_n.

1. We show that the contributions of the following terms are asymptotically negligible as $\lambda \to \infty$ when compared with $\psi(\lambda)$:
   (a) The values of $-n\mu + \sum_{j=-\infty}^{k} \varepsilon_j\beta_{nj}$ for *all* n and *fixed* k (Lemma 2.2).
   (b) The values of $-n\mu + \sum_{j=n+k}^{\infty} \varepsilon_j\beta_{nj}$ for *all* n and *large* k (Lemma 2.3).
2. Thus it suffices to study the asymptotic behavior of
$$P(\sup_{n\ge 0}(-n\mu + \sum_{j=\tilde k}^{n+k} \varepsilon_j\beta_{nj} > \lambda)$$
   for *large* λ, *large* (*but fixed*) $\tilde k$ and *large* (*but fixed*) k.
3. We proceed by splitting the supremum into different parts.
4. We show that the contribution of the following probabilities to $\psi(\lambda)$ is asymptotically negligible, first letting $\lambda \to \infty$, then $M \to \infty$:
$$P\left(\sup_{n\le \lambda/M}\left(-n\mu + \sum_{j=\tilde k}^{n+k} \varepsilon_j\beta_{nj} > \lambda\right)\right) \quad \text{for } \textit{fixed } k, \tilde k \ \text{(Lemma 2.4)},$$
$$P\left(\sup_{n\ge 0}\left(-n\mu + \sum_{j=\tilde k}^{\lambda/M} \varepsilon_j\beta_{nj}\right) > \lambda\right) \quad \text{for } \textit{fixed } \tilde k \ \text{(Lemma 2.5)}.$$
5. Thus it suffices to study the probability $P(\sup_{n\ge\lambda/M}(-n\mu+\sum_{j=[\lambda/M]}^{n+k} \varepsilon_j\beta_{nj} > \lambda)$ for *large* λ, M and *large* (*but fixed*) k.
6. We show that the latter probability is of the same asymptotic order as
$$P\left(\sup_{n\ge\lambda/M}\left(-n\mu + \sum_{j=[\lambda/M]}^{n+k} \varepsilon_j\beta_{nj}^{+}\right) > \lambda\right) \tag{2.3}$$

and that the latter probability is of the same order as

$$P\left(\varepsilon_j > \frac{j\mu+\lambda}{m_\varphi^+} \text{ for some } j \geq 1\right) \sim \frac{(m_\varphi^+)^\alpha}{(\alpha-1)\mu}\lambda P(\varepsilon > \lambda). \tag{2.4}$$

This goal is achieved by a series of lemmas.

(a) Lemmas 2.6–2.8 are, essentially, large deviations results that establish that exactly one noise variable is responsible for the high value of the supremum of the random walk. They are used to derive the right upper bound in (2.4) for the probability (2.3) (Lemma 2.9).
(b) We show that the probability (2.3) with β_{nj}^+ replaced by $-\beta_{nj}^-$ does not contribute to the asymptotic order in (2.4) (Lemma 2.10).
(c) We establish the matching lower bound in (2.4) by utilizing the same preliminary estimates (Lemma 2.12).

This proves the theorem for $p = 1$ and $m_\varphi^+ > 0$.

B. In Section 2.2 we proceed with the case $p = 1$ and $m_\varphi^+ = 0$.
C. In Section 2.3 we treat the case $0 < p < 1$.

2.1. *The case* $p = 1$ *and* $m_\varphi^+ > 0$. We start by truncating the infinite series in (2.2) from below.

LEMMA 2.2. *For every* $k > -\infty$,

$$\lim_{\lambda\to\infty} \frac{P\left(\sup_{n\geq 1}\left(-n\mu + \sum_{j=-\infty}^{k} \varepsilon_j\beta_{nj}\right) > \lambda\right)}{\lambda P(\varepsilon > \lambda)} = 0. \tag{2.5}$$

PROOF. Obviously,

$$p_k := P\left(\sup_{n\geq 1}\left(-n\mu + \sum_{j=-\infty}^{k} \varepsilon_j\beta_{nj}\right) > \lambda\right) \leq P\left(\sum_{j=-\infty}^{k} |\varepsilon_j| \sum_{i=1-j}^{\infty} |\varphi_i| > \lambda\right).$$

Write $\tilde{\varphi} := \sum_{j=-\infty}^{\infty} |\varphi_j| < \infty$. Since (1.5) holds, Lemma A3.7 applies. Therefore and since $\alpha > 1$,

$$\limsup_{\lambda\to\infty} \frac{p_k}{P(\varepsilon > \lambda)} \leq \sum_{j=-\infty}^{k}\left(\sum_{i=1-j}^{\infty} |\varphi_i|\right)^\alpha \leq \tilde{\varphi}^\alpha \sum_{j=-\infty}^{k}\sum_{i=1-j}^{\infty} [|\varphi_j|/\tilde{\varphi}].$$

The right-hand side is finite by virtue of (1.5). This proves the lemma. □

The next step consists of truncating the infinite series (2.2) from above.

LEMMA 2.3.

$$\lim_{k\to\infty}\limsup_{\lambda\to\infty} \frac{P\left(\sup_{n\geq 1}\left(-n\mu + \sum_{j=n+k}^{\infty} \varepsilon_j\beta_{nj}\right) > \lambda\right)}{\lambda P(\varepsilon > \lambda)} = 0. \tag{2.6}$$

PROOF. Notice that

$$\begin{aligned} q_k &:= P\left(\sup_{n\ge 1}\left(-n\mu + \sum_{j=n+k}^{\infty} \varepsilon_j \beta_{nj}\right) > \lambda\right) \\ &\le \sum_{n=1}^{\infty} P\left(\sum_{j=n+k}^{\infty} |\varepsilon_j| \sum_{i=-\infty}^{n-j} |\varphi_i| > \lambda + n\mu\right) \\ &= \sum_{n=1}^{\infty} P\left(\sum_{j=k}^{\infty} |\varepsilon_j| \sum_{i=-\infty}^{-j} |\varphi_i| > \lambda + n\mu\right) \end{aligned}$$

and by (1.5) there exists a constant $c > 0$ such that

$$q_k \le \sum_{n=1}^{\infty} P\left(\sum_{j=k}^{\infty} [|\varepsilon_j| - E|\varepsilon_1|] \sum_{i=-\infty}^{-j} |\varphi_i| > \lambda + n\mu - c\right).$$

An application of (1.5), Lemma A.3 and Karamata's theorem yield that

$$\limsup_{\lambda\to\infty} \frac{q_k}{\lambda P(\varepsilon > \lambda)} \le \text{const} \sum_{j=-\infty}^{-k} |j\varphi_j|.$$

Now let $k \to \infty$. This proves the lemma. □

Next we consider the main part of the exceedance probability $\psi(\lambda)$. For fixed $k, \tilde{k} \ge 1$ we study the behavior of

$$P\left(\sup_{n\ge 1}\left(-n\mu + \sum_{j=\tilde{k}}^{n+k} \varepsilon_j \beta_{nj}\right) > \lambda\right) \tag{2.7}$$

as $\lambda \to \infty$. Later the integers $k, \tilde{k}$ will be chosen sufficiently large. We split the supremum in (2.7) into separate parts. We start by showing that the values of n much smaller than λ do not matter asymptotically; a large deviation result for sums of iid heavy-tailed random variables indicates why this is expected; see Lemma A.1.

LEMMA 2.4. *For every fixed* $k, \tilde{k} \ge 1$,

$$\lim_{M\to\infty} \limsup_{\lambda\to\infty} \frac{P\left(\sup_{n\le\lambda/M}\left(-n\mu + \sum_{j=\tilde{k}}^{n+k} \varepsilon_j \beta_{nj}\right) > \lambda\right)}{\lambda P(\varepsilon > \lambda)} = 0. \tag{2.8}$$

PROOF. The following elementary inequality holds:

$$\begin{aligned} &P\left(\sup_{n\le\lambda/M}\left(-n\mu + \sum_{j=\tilde{k}}^{n+k} \varepsilon_j \beta_{nj}\right) > \lambda\right) \\ &\quad \le P\left(\sup_{n\le\lambda/M} \sum_{j=\tilde{k}}^{n+k} \varepsilon_j \beta_{nj} > \lambda\right) \le P\left(m_{|\varphi|} \sum_{\tilde{k}\le j\le\lambda/M+k} |\varepsilon_j| > \lambda\right), \end{aligned} \tag{2.9}$$

where

$$m_{|\varphi|} = \sum_{j=-\infty}^{\infty} |\varphi_j|. \tag{2.10}$$

For $M > m_{|\varphi|}E|\varepsilon|$, a large deviation result for sums of iid mean-zero random variables with regularly varying tails (see Lemma A.1 in the Appendix) implies that the probability in (2.9) is asymptotically of the order

$$\frac{\lambda}{M}P\biggl(\varepsilon > \lambda\biggl(\frac{1}{m_{|\varphi|}} - \frac{E|\varepsilon|}{M}\biggr)\biggr) \quad \text{as } \lambda \to \infty.$$

This and the fact that $P(\varepsilon > \lambda)$ is regularly varying with index $\alpha > 1$ prove the lemma. It is the factor $1/M$ that yields the result in the limit. □

In what follows, we assume for ease of representation that λ/M is an integer.

Our next step is to show that the noise variables ε_j with j much smaller than λ do not contribute to the order of magnitude of $\psi(\lambda)$.

LEMMA 2.5. *For every $\tilde{k} \geq 1$,*

$$\lim_{M\to\infty} \limsup_{\lambda\to\infty} \frac{P\Bigl(\sup_{n\geq 0}\Bigl(-n\mu + \sum_{j=\tilde{k}}^{\lambda/M} \varepsilon_j\beta_{nj}\Bigr) > \lambda\Bigr)}{\lambda P(\varepsilon > \lambda)} = 0.$$

PROOF. We have

$$P\Biggl(\sup_{n\geq 0}\Biggl(-n\mu + \sum_{j=\tilde{k}}^{\lambda/M} \varepsilon_j\beta_{nj}\Biggr) > \lambda\Biggr) \leq P\Biggl(m_{|\varphi|}\sum_{j=1}^{\lambda/M}(|\varepsilon_j| - E|\varepsilon|) > \lambda - m_{|\varphi|}E|\varepsilon|\frac{\lambda}{M}\Biggr).$$

The right-hand side probability can be estimated in the same way as in the proof of Lemma 2.4. This proves the lemma. □

The next few lemmas treat the supremum in the probability (2.7) for the values of n of the order λ (or higher). Our task is to formalize the statement that the event $\{\sup_n S_n > \lambda\}$ for large λ occurs due to a single large jump in the noise. We show first that, asymptotically, we cannot have this event occurring without observing a value of ε_j of the order λ. To make it easier to see the effect of positive values of ε_j's we look first at the positive parts β_{nj}^+ of the coefficients β_{nj}. Notice that the statements of Lemmas 2.2, 2.3, 2.4 and 2.5 remain valid if we similarly replace the β_{nj}'s with their positive parts in the corresponding statements.

LEMMA 2.6. *For every $M > 0$, there exists a small $\theta > 0$ such that for all $k \geq 1$,*

$$\lim_{\lambda\to\infty} \frac{P(\bigcup_{n\geq\lambda/M}\{-n\mu+\sum_{l=\lambda/M}^{n+k}\varepsilon_l\beta_{nl}^+>\lambda, \varepsilon_j\leq\theta(j+\lambda) \text{ for } j=\lambda/M,\dots,n+k\})}{\lambda P(\varepsilon>\lambda)} = 0. \tag{2.11}$$

PROOF. Our first observation is that it is enough to prove the lemma in the case when the noise variables ε_j have a continuous distribution. Indeed, let $(U_n)_{n\in\mathbb{Z}}$ be a sequence of iid random variables uniformly distributed in $(-1, 1)$ and independent of the noise sequence $(\varepsilon_n)_{n\in\mathbb{Z}}$. Let $\varepsilon'_n = \varepsilon_n + U_n,\ n \in \mathbb{Z}$. Observe that the sequence $(\varepsilon'_n)_{n\in\mathbb{Z}}$ satisfies all the requirements we placed on the original sequence $(\varepsilon_n)_{n\in\mathbb{Z}}$. Moreover, it has a continuous distribution. Furthermore, for every $M > 0$, $\theta > 0$ and $k \geq 1$ we have, by symmetry, for all $\lambda > 2/\theta$,

$$P\left(\bigcup_{n\geq\lambda/M}\left\{-n\mu+\sum_{l=\lambda/M}^{n+k}\varepsilon_l\beta_{nl}^+>\lambda,\ \varepsilon_j\leq\frac{\theta}{2}(j+\lambda)\text{ for } j=\lambda/M,\dots,n+k\right\}\right)$$
$$\leq 2P\left(\bigcup_{n\geq\lambda/M}\left\{-n\mu+\sum_{l=\lambda/M}^{n+k}\varepsilon'_l\beta_{nl}^+>\lambda,\ \varepsilon'_j\leq\theta(j+\lambda)\text{ for } j=\lambda/M,\dots,n+k\right\}\right).$$

That is, once one proves the statement of the lemma for the sequence $(\varepsilon'_n)_{n\in\mathbb{Z}}$ and halves the value of θ, the statement of the lemma for the sequence $(\varepsilon_n)_{n\in\mathbb{Z}}$ follows.

We proceed, therefore, to prove the lemma under the assumption of continuity of the distribution of the noise variables. Observe that for any $\theta > 0$,

$$\begin{aligned}(2.12)\qquad & P\left(\bigcup_{n\geq\lambda/M}\left\{-n\mu+\sum_{l=\lambda/M}^{n+k}\varepsilon_l\beta_{nl}^+>\lambda,\ \varepsilon_j\leq\theta(j+\lambda)\right.\right.\\ &\qquad\qquad\left.\left.\text{for } j=\lambda/M,\dots,n+k\right\}\right)\\ &\leq\sum_{n\geq\lambda/M}P\left(\sum_{l=\lambda/M}^{n+k}\varepsilon_l\beta_{nl}^+>n\mu,\ \varepsilon_j\leq\theta_{nk},\, j=\lambda/M,\dots,n+k\right),\end{aligned}$$

where

$$\theta_{nk} = \theta(1+M)(n+k).$$

For any $a > 0$, since $E\varepsilon I_{(-\infty,a]} < 0$, it is possible to define

$$a_* = \inf\{b > 0\colon E\varepsilon I_{[-b,a]}(\varepsilon) \leq 0\}. \tag{2.13}$$

Since $E\varepsilon = 0$ we have

$$E\varepsilon I_{[-a_*,a]}(\varepsilon) = 0, \tag{2.14}$$

because of the continuity of the distribution of ε.

Furthermore, in view of condition (1.4) on the tails, together with the current assumption that $p = 1$, we also have

$$a_* \leq a \quad \text{for all sufficiently large } a. \tag{2.15}$$

Indeed, by (2.14),

$$-E\varepsilon I_{(-\infty,-a_*)}(\varepsilon) = E\varepsilon I_{(a,\infty)}(\varepsilon)$$

and now (2.15) follows from

$$-E\varepsilon I_{(-\infty,\,-a_*)}(\varepsilon) = a_* P(\varepsilon < -a_*) + \int_{-\infty}^{-a_*} P(\varepsilon \le u)\,du,$$

$$E\varepsilon I_{(a,\,\infty)}(\varepsilon) = aP(\varepsilon > a) + \int_a^{\infty} P(\varepsilon > u)\,du.$$

From now on for any θ we consider λ so large that (2.15) *holds for $a = \theta_{nk}$ and all $n \ge \lambda/M$.*

Let

$$\tilde{\varepsilon}_j = \varepsilon_j I_{[-(\theta_{nk})_*,\theta_{nk}]}(\varepsilon_j), \qquad j = 1,\dots,n. \tag{2.16}$$

Then $E\tilde{\varepsilon}_j = 0$ and $|\tilde{\varepsilon}_j| \le \theta_{nk}$ for all j, and we observe that

$$P\left(\sum_{l=\lambda/M}^{n+k} \varepsilon_l \beta_{nl}^+ > n\mu,\ \varepsilon_j \le \theta_{nk},\ j=\lambda/M,\dots,n+k\right) \le P\left(\sum_{l=\lambda/M}^{n+k} \tilde{\varepsilon}_l \beta_{nl}^+ > n\mu\right) =: p_n.$$

Using Lemma A.2 in the Appendix, we conclude that

$$p_n \le \exp\left\{-\frac{n\mu}{2\theta_{nk}\, m_{|\varphi|}} \operatorname{arsinh} \frac{\theta_{nk} n\mu}{2(n+k-\lambda/M+1)m_{|\varphi|}\operatorname{var}(\tilde{\varepsilon}_1)}\right\}.$$

Here we make use of the fact that arsinh $y > \ln y$ for $y > 1$. Since $\alpha > 1$, there are constants $\beta < 2$ and $c_1 > 0$ such that $n \operatorname{var}(\tilde{\varepsilon}_1) \le c_1 n^{\beta}$ for all n and hence, for some constant $c_2 = c_2(k) > 0$,

$$p_n \le \exp\left\{-\frac{\mu}{2(1+M)m_{|\varphi|}} \frac{\ln(c_2 \theta n^{2-\beta})}{\theta}\right\}. \tag{2.17}$$

Now choose θ so small that

$$\frac{(2-\beta)\mu}{2(1+M)m_{|\varphi|}} \frac{1}{\theta} > \alpha + 1.$$

We then have by (2.17) that for all $n \ge \lambda/M$ and λ sufficiently large,

$$p_n \le \text{ const } n^{-(\alpha+1)},$$

which, together with (2.17), implies that the right-hand side of (2.12) is bounded by const $\lambda^{-\alpha}$. This concludes the proof of the lemma. □

The following result tells us that it is very unlikely to have two different noise variables ε_j that are large enough to contribute to very high values of $\sup_{n\ge 0} S_n$.

LEMMA 2.7. *For every $M > 0$ and $\theta > 0$,*

$$\lim_{\lambda\to\infty} \frac{P(\varepsilon_j > \theta j \text{ for at least two } j \ge \lambda/M)}{\lambda P(\varepsilon > \lambda)} = 0. \tag{2.18}$$

PROOF. Indeed, let

$$N = \inf\{j \geq \lambda/M\colon\ \varepsilon_j > \theta j\}.$$

Then

$$\begin{aligned} &P(\varepsilon_j > \theta j \text{ for at least two } j \geq \lambda/M) \\ &\quad = \sum_{l \geq \lambda/M} P(N = l,\ \varepsilon_j > \theta j \text{ for at least one } j > l) \qquad (2.19) \\ &\quad \leq \sum_{l \geq \lambda/M} P(N = l) P(N < \infty) = [P(N < \infty)]^2. \end{aligned}$$

But for large λ and a constant depending on M and θ, an application of Karamata's theorem yields

$$P(N < \infty) \leq \sum_{j \geq \lambda/M} P(\varepsilon > \theta j) \leq \operatorname{const} \lambda P(\varepsilon > \lambda).$$

The latter relation together with (2.19) proves (2.18). □

The following lemma is the key to the upper bound on $\psi(\lambda)$. It is a refined version of Lemma 2.6. Not only the event $\{\sup_n S_n > \lambda\}$ for large λ requires a noise variable ε_j not much smaller than $j + \lambda$, but this large noise variable has to take us almost all the way across the level λ.

LEMMA 2.8. *For every $M > 0$, $\delta \in (0, 1)$ and $k \geq 1$,*

$$\lim_{\lambda \to \infty} \frac{P\big(\sup_{n \geq \lambda/M}\big(-n\mu + \sum_{l=\lambda/M}^{n+k} \varepsilon_l \beta_{nl}^+\big) > \lambda,\ \varepsilon_j \leq (1-\delta)(j\mu + \lambda)/m_\varphi^+ \ all\ j \geq \lambda/M\big)}{\lambda P(\varepsilon > \lambda)}$$

$$= 0. \qquad (2.20)$$

PROOF. Write for any $\theta > 0$,

$$\begin{aligned} \psi_1(\lambda) := P\Bigg(&\sup_{n \geq \lambda/M} \Bigg(-n\mu + \sum_{r=\lambda/M}^{n+k} \varepsilon_r \beta_{nr}^+ \Bigg) > \lambda,\ \varepsilon_j \leq (1 - \delta)(j\mu + \lambda)/m_\varphi^+ \\ &\text{for all } j \geq \lambda/M, \text{for exactly one } l \geq \lambda/M \text{ we have } \varepsilon_l > \theta(l + \lambda) \Bigg). \end{aligned}$$

By Lemmas 2.7 and 2.6, there exists a θ sufficiently small such that

$$\begin{aligned} &P\Bigg(\sup_{n \geq \lambda/M} \Bigg(-n\mu + \sum_{l=\lambda/M}^{n+k} \varepsilon_l \beta_{nl}^+ \Bigg) > \lambda,\ \varepsilon_j \leq (1-\delta)(j\mu + \lambda)/m_\varphi^+ \text{ for all } j \geq \lambda/M \Bigg) \\ &\quad = \psi_1(\lambda) + o(\lambda P(\varepsilon > \lambda)) \quad \text{as } \lambda \to \infty. \end{aligned}$$

Let

$$N_1 = \inf\{j \geq \lambda/M\colon\ \varepsilon_j > \theta(j + \lambda)\}.$$

Choose λ so large that for $j > \lambda/M$,

$$\sum_{i=-\infty}^{1-j} \varphi_i \geq -\frac{\delta/2}{1-\delta} m_\varphi^+.$$

Observe that with this choice of λ,

$$\beta_{nj} = \sum_{i=-\infty}^{n-j} \varphi_i - \sum_{i=-\infty}^{-j} \varphi_i \leq m_\varphi^+ - \frac{\delta/2}{1-\delta} m_\varphi^+ = \frac{1-\delta/2}{1-\delta} m_\varphi^+.$$

Then we have

$$\psi_1(\lambda) \leq \sum_{l \geq \lambda/M} P(N_1 = l, A_l^{(1)}) + \sum_{l \geq \lambda/M} P(N_1 = l,\ A_l^{(2)}) =: \psi_2(\lambda) + \psi_3(\lambda),$$

where

$$A_l^{(1)} = \bigcup_{\lambda/M \leq n < l-k} \left\{ -n\mu + \sum_{i=\lambda/M}^{n+k} \varepsilon_i \beta_{ni}^+ > \lambda,\ \varepsilon_j \leq \theta(j+\lambda),\ j = \lambda/M, \ldots, l-1 \right\},$$

$$A_l^{(2)} = \bigcup_{n \geq l-k} \Bigg\{ -n\mu\delta/2 + \sum_{i=\lambda/M, i \neq l}^{n+k} \varepsilon_i \beta_{ni}^+ > \delta\lambda/2 - k\mu(1-\delta/2),\ \varepsilon_j \leq \theta(j+\lambda),$$

$$j = \lambda/M, \ldots, n+k,\ j \neq l \Bigg\}.$$

By Lemma 2.6 we have that for $\theta > 0$ small enough,

$$\psi_2(\lambda) = o(\lambda P(\varepsilon > \lambda)), \qquad \lambda \to \infty.$$

Since $E\varepsilon = 0$, $P(\varepsilon > 0) > 0$. Let $(\hat{\varepsilon}_n)$ be an independent copy of (ε_n). Then

$$\psi_3(\lambda) \leq \sum_{l \geq \lambda/M} \frac{1}{P(0 < \hat{\varepsilon}_l \leq \theta(l+\lambda))}$$

$$\times P\Bigg(\bigcup_{n \geq l-k} \Bigg\{ N_1 = l,\ -n\mu\delta/2 + \hat{\varepsilon}_l \beta_{nl}^+ + \sum_{i=\lambda/M, i \neq l}^{n+k} \varepsilon_i \beta_{ni}^+ > \delta\lambda/2 - k\mu(1-\delta/2),$$

$$\varepsilon_j \leq \theta(j+\lambda),\ j = \lambda/M, \ldots, n+k,\ j \neq l,\ 0 < \hat{\varepsilon}_l \leq \theta(l+\lambda) \Bigg\} \Bigg).$$

Note that $\{N_1 = l\} \subset \{\varepsilon_l > \theta(l+\lambda)\}$ which event is independent of $(\varepsilon_i)_{i \neq l}$. Hence the probability $P(\varepsilon_l > \theta(l+\lambda))$ can be factored out and, by observing that $P(0 < \hat{\varepsilon}_l \leq \theta(l+\lambda)) \geq 0.5 P(\varepsilon > 0)$ for sufficiently large λ, we have

$$\psi_3(\lambda) \leq \frac{2}{P(\varepsilon > 0)} \sum_{l \geq \lambda/M} P\Big(\varepsilon > \theta(l+\lambda)\Big)$$

$$\times P\Bigg(\bigcup_{n \geq l-k} \Bigg\{ -n\mu\delta/2 + \sum_{i=\lambda/M}^{n+k} \varepsilon_i \beta_{ni}^+ > \delta\lambda/2 - k\mu(1-\delta/2),$$

$$\varepsilon_j \leq \theta(j+\lambda),\ j = \lambda/M, \ldots, n+k \Bigg\} \Bigg)$$

$$\leq \frac{2}{P(\varepsilon > 0)} \sum_{l \geq \lambda/M} P\Big(\varepsilon > \theta(l+\lambda)\Big)$$
$$\times P\left(\bigcup_{n \geq (\lambda/M)-k} \left\{ -n\mu\delta/2 + \sum_{i=\lambda/M}^{n+k} \varepsilon_i \beta_{ni}^+ > \delta\lambda/2 - k\mu(1-\delta/2), \right.\right.$$
$$\left.\left. \varepsilon_j \leq \theta(j+\lambda),\ j = \lambda/M, \ldots, n+k \right\} \right).$$

By Lemma 2.6 the right-hand side of the latter relation is $o(\lambda P(\varepsilon > \lambda))$ provided θ is chosen small enough in comparison with δ. This concludes the proof of (2.20). □

Now we are ready to derive an upper bound for $\psi(\lambda)$. Observe that we are still treating the case when $p = 1$ and $m_\varphi^+ > 0$. Let

$$\psi^+(\lambda) = P\left(\sup_{n \geq 1} \left(-n\mu + \sum_{l=-\infty}^{\infty} \varepsilon_l \beta_{nl}^+ \right) > \lambda \right).$$

LEMMA 2.9. *The following relation holds*:

$$\limsup_{\lambda \to \infty} \frac{\psi^+(\lambda)}{\lambda P(\varepsilon > \lambda)} \leq \frac{(m_\varphi^+)^\alpha}{\alpha - 1} \frac{1}{\mu}. \tag{2.21}$$

PROOF. A straightforward argument [similar to (2.25) below] in combination with Lemmas 2.2–2.5 and 2.8 gives for any $\delta \in (0, 1)$,

$$\limsup_{\lambda \to \infty} \frac{\psi^+(\lambda)}{\lambda P(\varepsilon > \lambda)} \leq \limsup_{\lambda \to \infty} \frac{P\Big(\varepsilon_j > ((1-\delta)/m_\varphi^+)(j\mu + \lambda) \text{ for some } j \geq 1\Big)}{\lambda P(\varepsilon > \lambda)}.$$

However, by Karamata's theorem,

$$P\left(\varepsilon_j > \frac{1-\delta}{m_\varphi^+}(j\mu + \lambda) \text{ for some } j \geq 1 \right)$$
$$\leq \sum_{j=1}^{\infty} P\left(\varepsilon > \frac{1-\delta}{m_\varphi^+}(j\mu + \lambda) \right) \sim \frac{(m_\varphi^+)^\alpha}{(1-\alpha)(1-\delta)^\alpha} \frac{1}{\mu} \lambda P(\varepsilon > \lambda).$$

Since we may choose δ as close to zero as we wish, we conclude that (2.21) holds. □

What happens if one replaces the positive parts β_{nj}^+ of the coefficients β_{nj} with $-\beta_{nj}^-$? The following lemma provides the answer. Recall that we still consider the case $p = 1$.

LEMMA 2.10. *For any* $\mu > 0$,

$$\lim_{\lambda \to \infty} \frac{P\Big(\sup_{n \geq 1} \big(-n\mu - \sum_{j=1}^n \varepsilon_j \beta_{nj}^- \big) > \lambda \Big)}{\lambda P(\varepsilon > \lambda)} = 0. \tag{2.22}$$

PROOF. Choose K so large that

$$r_K := E\varepsilon I_{(-\infty,\,K]}(\varepsilon) \geq -\mu/(2m_{|\varphi|}), \tag{2.23}$$

where $m_{|\varphi|}$ is defined in (2.10). Write

$$\tilde{\varepsilon}_j^{(K)} = \varepsilon_j I_{(-\infty,\,K]}(\varepsilon_j), \qquad \hat{\varepsilon}_j^{(K)} = -[\tilde{\varepsilon}_j^{(K)} - r_K], \qquad j = 1, 2, \dots .$$

We have by (2.23),

$$\begin{aligned}\tilde{p} := P\left(\sup_{n\geq 1}\left(-n\mu - \sum_{j=1}^{n}\varepsilon_j\beta_{nj}^-\right) > \lambda\right) &\leq P\left(\sup_{n\geq 1}\left(-n\mu - \sum_{j=1}^{n}\tilde{\varepsilon}_j^{(K)}\beta_{nj}^-\right) > \lambda\right)\\ &\leq P\left(\sup_{n\geq 1}\left(-n\mu/2 + \sum_{j=1}^{n}\hat{\varepsilon}_j^{(K)}\beta_{nj}^-\right) > \lambda\right).\end{aligned} \tag{2.24}$$

The random variables $\hat{\varepsilon}_j^{(K)}$ are iid, have mean zero and are bounded from below. In view of the tail balancing condition (1.4) and the current assumption $p = 1$, for any $\rho > 0$ we can find a sequence (η_j) of iid random variables such that:

1. $\eta_1 \overset{\text{st}}{\geq} \hat{\varepsilon}_1^{(K)}$, where $\overset{\text{st}}{\geq}$ stands for stochastic domination, that is, $P(\eta_1 > x) \geq P(\hat{\varepsilon}_1^{(K)} > x)$ for all x.
2. $E\eta_1 = 0$.
3. η_1 is bounded from below.
4. $\lim_{\lambda\to\infty} \frac{P(\eta_1>\lambda)}{P(\varepsilon>\lambda)} = \rho$.

Hence, the sequence (η_j) satisfies all assumptions imposed on (ε_j), and so we may utilize all the results proved so far with (ε_j) replaced with (η_j). Recall that stochastic domination of $\hat{\varepsilon}_i^{(K)}$ by η_j implies that the sequence (η_j) has a higher probability to belong to any measurable increasing set in $\mathbb{R}^\infty$ than the sequence $(\hat{\varepsilon}_j^{(K)})$ does [see, e.g., Strassen (1965)]. Therefore, (2.24), Lemma 2.9 and stochastic domination imply that

$$\begin{aligned}\limsup_{\lambda\to\infty}\frac{\tilde{p}}{\lambda P(\varepsilon>\lambda)} &\leq \limsup_{\lambda\to\infty}\frac{P\left(\sup_{n\geq 1}\left(-n\mu/2+\sum_{j=1}^{n}\hat{\varepsilon}_j^{(K)}\beta_{nj}^-\right)>\lambda\right)}{\lambda P(\varepsilon>\lambda)}\\ &\leq \limsup_{\lambda\to\infty}\frac{P(\eta_1>\lambda)}{P(\varepsilon>\lambda)}\,\frac{P\left(\sup_{n\geq 1}\left(-n\mu/2+\sum_{j=1}^{n}\eta_j\beta_{nj}^-\right)>\lambda\right)}{\lambda P(\eta_1>\lambda)}\\ &\leq \rho\frac{m_{|\varphi|}^{\alpha}}{\alpha-1}\frac{2}{\mu}.\end{aligned}$$

Now let $\rho \to 0$. This establishes (2.22). $\square$

We can now put the pieces together and bound the probability $\psi(\lambda)$ from above. From its definition, for $\delta \in (0, 1)$,

$$\psi(\lambda) \le P\left(\sup_{n\ge 1}\left(-n\mu(1-\delta)+\sum_{j=1}^{n}\beta_{nj}^{+}\varepsilon_j\right)+\sup_{n\ge 1}\left(-n\mu\delta-\sum_{j=1}^{n}\beta_{nj}^{-}\varepsilon_j\right)>\lambda\right). \tag{2.25}$$

Combining (2.25) with Lemmas 2.9 and 2.10, we immediately conclude the lemma.

LEMMA 2.11.

$$\limsup_{\lambda\to\infty}\frac{\psi(\lambda)}{\lambda P(\varepsilon>\lambda)} \le \frac{(m_\varphi^+)^\alpha}{\alpha-1}\frac{1}{\mu}.$$

This yields the upper bound in (1.12). It remains to show the lower bound.

LEMMA 2.12.

$$\liminf_{\lambda\to\infty}\frac{\psi(\lambda)}{\lambda P(\varepsilon>\lambda)} \ge \frac{(m_\varphi^+)^\alpha}{\alpha-1}\frac{1}{\mu}. \tag{2.26}$$

PROOF. Recall that by Lemma 2.2 for any $K_1 > -\infty$,

$$\liminf_{\lambda\to\infty}\frac{\psi(\lambda)}{\lambda P(\varepsilon>\lambda)} \ge \liminf_{\lambda\to\infty}\frac{P\left(\sup_{n\ge 1}\left(-n\mu+\sum_{j=K_1+1}^{\infty}\varepsilon_j\beta_{nj}\right)>\lambda\right)}{\lambda P(\varepsilon>\lambda)}. \tag{2.27}$$

Fix $\delta \in (0, 0.5)$ and choose K_1 so large that

$$\sum_{l=-\infty}^{j}\varphi_l \le \delta m_\varphi^+$$

for all $j \le -K_1$. Choose a fixed $i = i(\delta)$ such that

$$\sum_{l=-\infty}^{i}\varphi_l \ge (1-\delta)m_\varphi^+.$$

The above inequalities imply that for all $n \ge K_1$,

$$\left\{m_\varphi^+(1-2\delta)\varepsilon_n>(1+\delta)(n\mu+\lambda), -i\mu+\delta n\mu+\sum_{j=K_1+1,\, j\ne n}^{\infty}\varepsilon_j\beta_{n+i,j}>-\delta\lambda\right\}$$

$$\subset\left\{-(n+i)\mu+\sum_{j=K_1+1}^{\infty}\varepsilon_j\beta_{n+i,j}>\lambda\right\}$$

and so

$$P\left(\sup_{n\geq 1}\left(-n\mu+\sum_{j=K_1+1}^{\infty}\varepsilon_j\beta_{nj}\right)>\lambda\right)$$
$$\geq P\left(\sup_{n\geq K_1+i}\left(-n\mu+\sum_{j=K_1+1}^{\infty}\varepsilon_j\beta_{nj}\right)>\lambda\right)$$
$$\geq P\left(\bigcup_{n\geq K_1}\Big\{m_\varphi^+(1-2\delta)\varepsilon_n>(1+\delta)(\lambda+n\mu),\right.$$
$$\left.-i\mu+\delta n\mu+\sum_{j=K_1+1,\, j\neq n}^{\infty}\varepsilon_j\beta_{n+i,\,j}>-\delta\lambda\Big\}\right).$$

Furthermore, for a $\theta>0$ the latter probability cannot be smaller than

$$P\left(\bigcup_{n\geq K_1}\Big\{m_\varphi^+(1-2\delta)\varepsilon_n>(1+\delta)(\lambda+n\mu)\Big\}\right)$$
$$-P\left(\bigcup_{n\geq K_1}\Big\{-i\mu+\delta n\mu+\sum_{j=K_1+1}^{\infty}\varepsilon_j\beta_{n+i,j}\leq-\delta\lambda,\right.$$
$$\left.|\varepsilon_j|\leq\theta(j+\lambda),\ j=1,\dots,n-1\Big\}\right)$$
$$-P\left(\bigcup_{n\geq K_1}\Big\{m_\varphi^+(1-2\delta)\varepsilon_n>(1+\delta)(\lambda+n\mu),\ -i\mu+\delta n\mu\right.$$
$$\left.+\sum_{j=K_1+1}^{\infty}\varepsilon_j\beta_{n+i,j}\leq-\delta\lambda,\ |\varepsilon_j|>\theta(j+\lambda)\text{ for some }\ j=1,\dots,n-1\Big\}\right).$$

For θ small enough relative to δ, the first probability being subtracted above is of a smaller order than $\lambda P(\varepsilon>\lambda)$ as $\lambda\to\infty$ by Lemma 2.6 (notice that the lemmas preceding Lemma 2.6 show that taking the supremum over a large set and different bounds of summation contribute only terms of a smaller order as well). Here (and in the sequel) "small enough relative to δ" means that θ/δ are small enough for the requirements of Lemma 2.6. Similarly, Lemma 2.7 and the remark just made show that for any $\theta>0$ the second probability being subtracted above is of a smaller order than $\lambda P(\varepsilon>\lambda)$ as $\lambda\to\infty$. Therefore, we conclude that

$$P\left(\sup_{n\geq 1}\left(-n\mu+\sum_{j=K_1+1}^{\infty}\varepsilon_j\beta_{nj}\right)>\lambda\right)$$
$$\geq P\left(\bigcup_{n\geq K_1}\Big\{m_\varphi^+(1-2\delta)\varepsilon_n>(1+\delta)(\lambda+n\mu)\Big\}\right)-o(\lambda P(\varepsilon>\lambda)).$$

Yet another application of Lemma 2.7 shows now that the right-hand side above is

$$\sum_{n=K_1}^{\infty} P\Big(m_\varphi^+(1-2\delta)\varepsilon_n > (1+\delta)(\lambda+n\mu)\Big) - o(\lambda P(\varepsilon>\lambda)).$$

Therefore,

$$\begin{aligned}\liminf_{\lambda\to\infty}\frac{\psi(\lambda)}{\lambda P(\varepsilon>\lambda)} &\geq \liminf_{\lambda\to\infty}\frac{\sum_{n=K_1}^{\infty} P\Big(m_\varphi^+(1-2\delta)\varepsilon_n > (1+\delta)(\lambda+n\mu)\Big)}{\lambda P(\varepsilon>\lambda)}\\ &= \left(\frac{1-2\delta}{1+\delta}\right)^{\alpha}\frac{(m_\varphi^+)^\alpha}{\alpha-1}\frac{1}{\mu}.\end{aligned}$$

Letting $\delta\to 0$, we finally arrive at the lower bound (2.26). This proves the lemma. □

That is, we have proved Theorem 2.1 in the case $p=1$ and $m_\varphi^+>0$.

2.2. *The case $p=1$ and $m_\varphi^+=0$.* Pick a $\theta>0$ and choose an $i=i(\theta)$ that has the following property: if one defines

$$\tilde\varphi_j = \begin{cases}\varphi_j, & \text{if } j\neq -i,\\ \theta+\varphi_{-i}, & \text{if } j=-i,\end{cases} \tag{2.28}$$

then for the new set of coefficients one has

$$m_\varphi^+(\theta) := \sup_{-\infty<n<\infty}\sum_{k=-\infty}^{n}\tilde\varphi_k > 0.$$

Clearly,

$$m_\varphi^+(\theta)\leq\theta. \tag{2.29}$$

For a fixed $0<\delta<\min(\mu,1)$ let

$$X_n^{(\theta)} = -(\mu-\delta)+\sum_{j=-\infty}^{\infty}\tilde\varphi_{n-j}\varepsilon_j,\qquad n\in\mathbb{Z},$$

and consider the dependent step random walk $(S_n^{(\theta)})_{n\geq 0}$,

$$S_0^{(\theta)}=0,\qquad S_n^{(\theta)}=X_1^{(\theta)}+\cdots+X_n^{(\theta)},\qquad n\geq 1.$$

Since we have already proved the theorem in the case $m_\varphi^+>0$, it follows from (2.29) that

$$\lim_{\lambda\to\infty}\frac{P\Big(\sup_{n\geq 0} S_n^{(\theta)}>\lambda\Big)}{\lambda P(\varepsilon>\lambda)} = \frac{[m_\varphi^+(\theta)]^\alpha}{\alpha-1}\frac{1}{\mu-\delta}\leq\frac{\theta^\alpha}{\alpha-1}\frac{1}{\mu-\delta}. \tag{2.30}$$

Observe that

$$S_n^{(\theta)} = S_n+\theta\sum_{k=1}^{n}\varepsilon_{k+i}+\delta n,\qquad n\geq 1.$$

Therefore,

$$P\left(\sup_{n\geq 0} S_n^{(\theta)} > \lambda(1-\delta)\right) \geq P\left(\sup_{n\geq 0} S_n > \lambda, \inf_{n\geq 0}\left(\theta \sum_{k=1}^{n} \varepsilon_{k+i} + \delta n\right) \geq -\delta\lambda\right)$$

$$\geq P\left(\sup_{n\geq 0} S_n > \lambda\right) - P\left(\inf_{n\geq 0}\left(\theta \sum_{k=1}^{n} \varepsilon_{k+i} + \delta n\right) < -\delta\lambda\right).$$

Since we are still assuming that $p = 1$, it follows by the Embrechts and Veraverbeke result (1.2) that

$$P\left(\inf_{n\geq 0}\left(\theta \sum_{k=1}^{n} \varepsilon_{k+i} + \delta n\right) < -\delta\lambda\right) = o(\lambda P(\varepsilon > \lambda)) \qquad \text{as } \lambda \to \infty.$$

Therefore, it follows from (2.30) that

$$\limsup_{\lambda\to\infty} \frac{P\left(\sup_{n\geq 0} S_n > \lambda\right)}{\lambda P(\varepsilon > \lambda)} \leq \frac{(1-\delta)^{1-\alpha}\theta^{\alpha}}{\alpha - 1}\frac{1}{\mu - \delta}.$$

Since θ can be taken arbitrarily small, the statement of the theorem in the case $p = 1$ and $m_{\varphi}^{+} = 0$ follows.

2.3. *The case* $0 < p < 1$. Denote

$$\hat{\varepsilon}_j^{\pm} = \varepsilon_j^{\pm} - E\varepsilon_j^{\pm}, \qquad j \in \mathbb{Z}.$$

Observe that $(\hat{\varepsilon}_j^{+})_{j\in\mathbb{Z}}$ and $(\hat{\varepsilon}_j^{-})_{j\in\mathbb{Z}}$ are two (nonindependent) sequences of iid zero-mean random variables that satisfy the regular variation and tail balance conditions (1.4) (corresponding to the parameter $p = 1$). Let

$$Q_n^{\pm} = \sum_{j=1}^{\infty} \hat{\varepsilon}_j^{\pm}\beta_{nj}, \qquad n \in \mathbb{Z}.$$

We will now study the contributions of $\{Q_n^{+}\}$ and $\{Q_n^{-}\}$ to the overall ruin probability. Lemma 2.2 allows us to disregard the influence of the noise variables ε_j with $j \leq 0$, which turns out to be useful in the sequel. We have

$$\begin{aligned}\psi(\lambda) &= P\left(\sup_{n\geq 0}\left(-n\mu + \sum_{j=1}^{\infty} \varepsilon_j\beta_{nj}\right) > \lambda\right) + o(\lambda P(\varepsilon > \lambda))\\ &= P\left(\sup_{n\geq 0}(-n\mu + Q_n^{+} - Q_n^{-}) > \lambda\right) + o(\lambda P(\varepsilon > \lambda))\\ &=: P(A) + o(\lambda P(\varepsilon > \lambda)).\end{aligned}$$

An application of Lemmas 2.6 and 2.7 shows that for any $\theta > 0$ small enough (compared to μ),

$$\begin{aligned} P(A) &= P\Big(A,\ |\varepsilon_j| > \theta(j+\lambda) \text{ for exactly one } j \geq 1\Big) + o(\lambda P(\varepsilon > \lambda)) \\ &= P\Big(A,\ \varepsilon_j < -\theta(j+\lambda) \text{ for exactly one } j \geq 1,\ \varepsilon_j \leq \theta(j+\lambda) \text{ for all } j \geq 1\Big) \\ &\quad + P\Big(A,\ \varepsilon_j > \theta(j+\lambda) \text{ for exactly one } j \geq 1, \\ &\qquad\qquad \varepsilon_j \geq -\theta(j+\lambda) \text{ for all } j \geq 1\Big) + o(\lambda P(\varepsilon > \lambda)) \\ &=: P(A^{(1)}) + P(A^{(2)}) + o(\lambda P(\varepsilon > \lambda)). \end{aligned}$$

Hence

$$\lim_{\lambda\to\infty} \frac{\psi(\lambda)}{\lambda P(\varepsilon > \lambda)} = \lim_{\lambda\to\infty} \frac{P(A^{(1)})}{\lambda P(\varepsilon > \lambda)} + \lim_{\lambda\to\infty} \frac{P(A^{(2)})}{\lambda P(\varepsilon > \lambda)}. \tag{2.31}$$

We have for any $\delta \in (0, 1)$,

$$\begin{aligned} P(A^{(1)}) \leq\ & P\Big(\sup_{n\geq 0}(-n\mu(1-\delta) - Q_n^-) > \lambda(1-\delta)\Big) \\ &+ P\Big(\sup_{n\geq 0}(-n\mu\delta + Q_n^+) > \lambda\delta,\ \hat{\varepsilon}_j^+ \leq \theta(j+\lambda) \text{ for all } j \geq 1\Big) \\ =:\ & P\Big(A^{(1,1)}\Big) + P\Big(A^{(1,2)}\Big). \end{aligned}$$

If θ is small (compared to δ), we conclude by the same arguments as in Section 2.1 (where Lemmas 2.6 and 2.10 play a key role) that

$$\lim_{\lambda\to\infty} \frac{P(A^{(1,2)})}{\lambda P(\varepsilon > \lambda)} = 0.$$

On the other hand, since the theorem has already been proved in the case $p = 1$, we immediately conclude that

$$\lim_{\lambda\to\infty} \frac{P(A^{(1,1)})}{\lambda P(\varepsilon > \lambda)} = \lim_{\lambda\to\infty} \frac{P(A^{(1,1)})}{\lambda P(\hat{\varepsilon}^- > \lambda)} \frac{P(\hat{\varepsilon}^- > \lambda)}{P(\varepsilon > \lambda)} = q \frac{1}{(1-\delta)^\alpha} \frac{(m_\varphi^-)^\alpha}{\alpha - 1} \frac{1}{\mu}.$$

Since δ can be taken arbitrarily small, we have

$$\limsup_{\lambda\to\infty} \frac{P(A^{(1)})}{\lambda P(\varepsilon > \lambda)} \leq q \frac{(m_\varphi^-)^\alpha}{\alpha - 1} \frac{1}{\mu}. \tag{2.32}$$

On the other hand, for any $\delta > 0$ one has by Lemmas 2.6 and 2.7,

$$\begin{aligned}P(A^{(1)}) &\geq P\Big(\sup_{n\geq 0}\big(-n\mu(1+\delta)Q_n^-\big) > \lambda(1+\delta)\Big)\\&\quad - P\Big(\sup_{n\geq 0}(n\mu\delta + Q_n^+) \leq -\lambda\delta,\ \hat{\varepsilon}_j^+ \leq \theta(j+\lambda) \text{ for all } j \geq 1\Big)\\&\quad - o(\lambda P(\varepsilon > \lambda))\\&=: P(A^{(1,-1)}) - P(A^{(1,-2)}) - o(\lambda P(\varepsilon > \lambda)).\end{aligned}$$

The same argument as above shows that, if θ is small (compared to δ), then

$$\lim_{\lambda\to\infty}\frac{P(A^{(1,-2)})}{\lambda P(\varepsilon > \lambda)} = 0 \quad\text{and}\quad \lim_{\lambda\to\infty}\frac{P(A^{(1,-1)})}{\lambda P(\varepsilon > \lambda)} = q\frac{1}{(1+\delta)^\alpha}\frac{(m_\varphi^-)^\alpha}{\alpha-1}\frac{1}{\mu}.$$

Since δ can be taken arbitrarily small, we have

$$\liminf_{\lambda\to\infty}\frac{P(A^{(1)})}{\lambda P(\varepsilon > \lambda)} \geq q\frac{(m_\varphi^-)^\alpha}{\alpha-1}\frac{1}{\mu},$$

which together with (2.32) shows that

$$\lim_{\lambda\to\infty}\frac{P(A^{(1)})}{\lambda P(\varepsilon > \lambda)} = q\frac{(m_\varphi^-)^\alpha}{\alpha-1}\frac{1}{\mu}. \tag{2.33}$$

An identical argument shows that

$$\lim_{\lambda\to\infty}\frac{P(A^{(2)})}{\lambda P(\varepsilon > \lambda)} = p\frac{(m_\varphi^+)^\alpha}{\alpha-1}\frac{1}{\mu}, \tag{2.34}$$

and combining (2.31), (2.33) and (2.34) we obtain the statement of the theorem in its full generality. This completes the proof of the theorem. □

REMARK 2.13. A careful analysis of the proof shows that Theorem 2.1 remains valid if the step sequence (X_n) of the random walk (S_n) is replaced with $(X_n + Y_n)$ where (Y_n) is an iid sequence independent of (X_n) such that $P(Y_1 > x) = o(P(X_1 > x))$ as $x \to \infty$ and $-\infty < E(X_1 + Y_1) < 0$. However, one has to replace μ in (1.12) with $-E(X_1 + Y_1)$. A special case occurs when one considers the ruin probability

$$P\Big(\sup_{t\geq 0}(S_{N(t)} - ct) > \lambda\Big) = P\left(\sup_{n\geq 1}\sum_{i=1}^{n}(X_i - cZ_i) > \lambda\right),$$

where (Z_i) is a sequence of iid nonnegative random variables with positive mean, independent of (X_i),

$$N(t) = \#\{i: Z_1 + \cdots + Z_i \leq t\}, \qquad t \geq 0$$

is the corresponding renewal counting process and $c > 0$ is positive constant. In the latter case no assumptions on the distribution of the Z_i's are necessary, apart from finiteness of the mean.

3. A teletraffic data set. Many important queuing systems of today are found in computer communication networks. A data set we consider is a part of a larger data set collected and reported by Cunha, Bestavaros and Crovella (1995). It consists of traces of WWW sessions run from 32 workstations in an undergraduate computer lab in Boston University from November 1994 through February 1995. We use only the data for January 1995. The traces of the sessions come with the sizes of the files that a user requested and with the time stamp of the request. We have combined the file sizes for the month of January in a single time series ordered according to the time of the request. Those requests that could be filled using cached files did not require network transmission and, hence, were deleted from the time series. The remaining 17,675 requests must be fed to a communication link, and then they are responsible for the right tail of the steps in the Lindley equation that describes the behavior of that link. It is not our goal here to fit any particular model to this time series. Rather, we would like to show that this data set exhibits the characteristics that led us to the present study in the first place: it is heavy-tailed, and there is obvious dependence in the right tails of the observations. Figure 1 shows the graph of a negative drift random walk generated by this data.

We start by estimating the thickness of the tail. An exploratory means is to consider the asymptotic behavior of the ratio

$$T_n(p) = \frac{\max_{i=1,\dots,n} X_i^p}{\sum_{i=1}^n X_i^p} \tag{3.1}$$

for some $p > 0$; see EKM (1997), Section 6.2.6. Indeed, for a stationary ergodic sequence, if $EX^p < \infty$, the ergodic theorem implies that $T_n(p) \to 0$ a.s. as $n \to \infty$. A glance at the left part of Figure 3 convinces one that this is hardly the case for $p = 1.3$ and so we may guess that the 1.3th moment is infinite. Further confirmation of this fact comes from considering the *Hill estimator*,

$$H_n^{(m)} = \left(\frac{1}{m} \sum_{i=1}^{m} \ln X_{(i)} - \ln X_{(m)} \right)^{-1},$$

where $X_{(n)} \le \cdots \le X_{(1)}$ are the order statistics of the sample $X_1, \dots, X_n$. The statistic $H_n^{(m)}$ is a consistent estimator for the parameter α of the tail $P(X > x) = L(x)x^{-\alpha}$ for some $\alpha > 0$ and a slowly varying function L, provided that

$$m = m_n \to \infty \quad \text{and} \quad m/n \to 0. \tag{3.2}$$

It is also asymptotically normal for a weakly dependent sequence and under further conditions on L. For an extensive discussion of the Hill and related tail parameter estimators, see EKM [(1997), Section 6.4]. The right part of Figure 3 shows a *Hill-plot*,

$$(m, H_n^{(m)}) \tag{3.3}$$

with asymptotic confidence bands corresponding to an iid sequence (X_i) with tail $P(X > x) \sim \text{const } x^{-\alpha}$. We may conclude that the Hill-plot gives an estimate of the value 1.3 in the m-region $(50, 400)$, say.

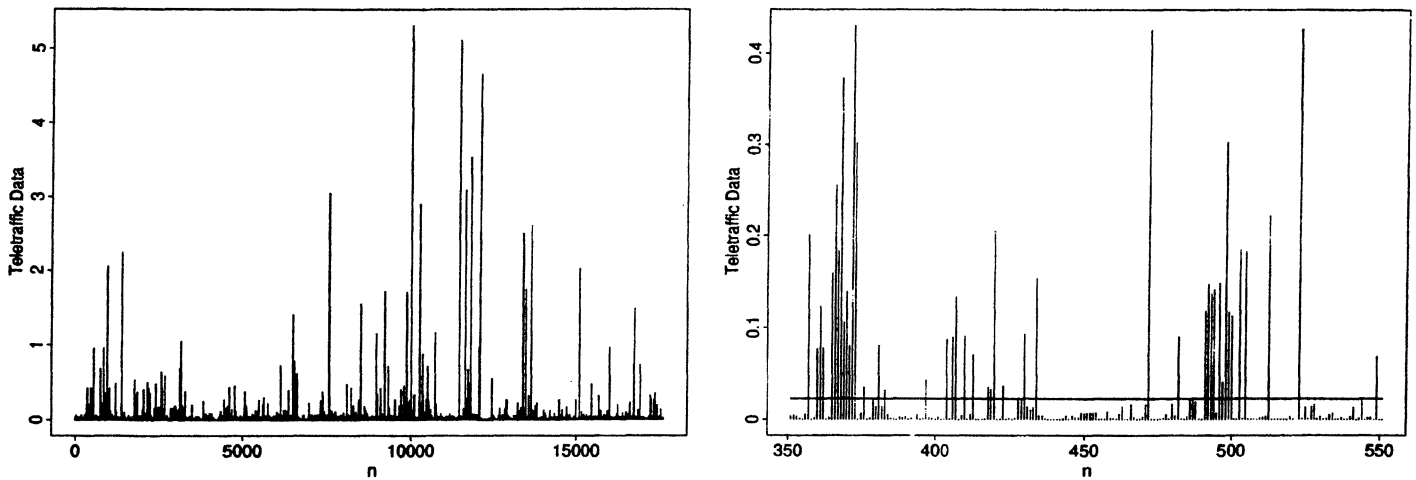

FIG. 2. *Left: the entire teletraffic data set,* $n = 17,675$. *Right: the teletraffic data* $X_{350}, \dots, X_{550}$. *The solid line represents the threshold* $u = 23,234$ *corresponding to the* 90% *quantile of the data. Exceedances of this high threshold occur in clusters. This indicates that there is dependence in the tails in the data. The unit on the y-axis is a million.*

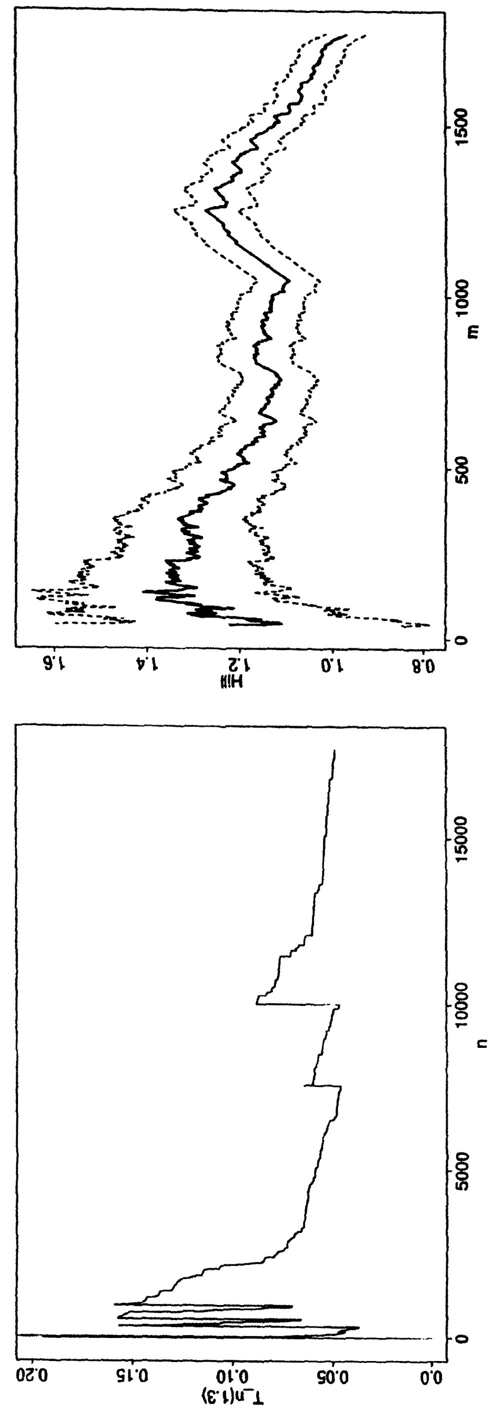

FIG. 3. *Left: the statistic* $T_n(1.3)$ *as a function of n; see* (3.1). *Right: the Hill plot for the teletraffic data; see* (3.3).

We now look at the dependence in the time series (X_i). To this end we use various tools. The most common one is the *sample autocorrelation function.* It is given in the left part of Figure 4 for the first 1500 lags. We omitted the asymptotic $\pm 1.96/n^{1/2}$ confidence bands which correspond to an iid Gaussian sequence. In view of the extremely heavy tails of X (the second moment does not exist) it is not clear what the sample autocorrelation function actually represents. Work by Davis and Resnick (1985, 1986) [see also Section 13.3 in Brockwell and Davis (1991)], shows that the sample autocorrelation at lag h estimates the quantity $\sum_j \varphi_j \varphi_{j+h} / \sum_j \varphi_j^2$ which can be interpreted as the autocorrelation at lag h of a linear process (1.3) with an iid standard Gaussian sequence (ε_j). However, recent work by Davis and Resnick (1996), Resnick, Samorodnitsky and Xue (1999) and Davis and Mikosch (1998) shows that the sample autocorrelations of nonlinear stationary sequences can be extremely unreliable in the sense that the convergence rate can be very slow or that the sample autocorrelations can have nondegenerate weak limits. Therefore, we prefer here to consider some alternative methods to detect dependence in a time series.

Consider a random walk $\widetilde{S}_n = Y_1 + \cdots + Y_n$ for a stationary sequence (Y_i) of random variables assuming values 0 and 1, where $P(Y_1 = 1) = p \in (0, 1)$. We may assume that the sequence (Y_i) is generated from a stationary sequence (X_i) as follows:

$$Y_i = I_{(u,\infty)}(X_i), \qquad i = 1, 2, \dots,$$

for some given threshold $u > 0$. If the sequence (X_i) is iid, a well-established theory exists for the longest run of 1's. A *run of length j in $Y_1, \dots, Y_n$* is defined as a subsequence $(Y_{i+1}, \dots, Y_{i+j})$ of $(Y_1, \dots, Y_n)$ such that

$$Y_i = 0, \qquad Y_{i+1} = \cdots = Y_{i+j} = 1, \qquad Y_{i+j+1} = 0,$$

where we formally set $Y_0 = Y_{n+1} = 0$. Some theory about the asymptotic behavior of the longest run Z_n of 1's in an iid sequence $X_1, \dots, X_n$ is provided in EKM (1997), Section 8.5. Corollary 8.5.10 in the latter reference states that the longest run Z_n, with probability 1, falls for large n in the interval $[\alpha_n, \beta_n]$, where

$$\alpha_n = \left[\frac{\ln(nq) - \ln_3(nq) - 0.001}{-\ln p} \right] - 1$$

and

$$\beta_n = \left[\frac{\ln(nq) + \ln_2(nq) + 1.001 \ln_3(nq)}{-\ln p} \right],$$

where $[x]$ denotes the integer part of x, $q = 1 - p$, $\ln_2 x = \ln\ln x$ and $\ln_3 x = \ln\ln\ln x$. The right part of Figure 4 shows the graphs of α_n, β_n together with the longest run of 1's for an iid sequence $(I_{(u,\infty)}(X_i))$ with the property that $P(X > u) = 0.1$ for an appropriately chosen threshold u (this curve lies nicely between α_n and β_n) and for the teletraffic data (X_n). In this case it is obvious that the longest runs of 1's of the indicators $(I_{(u,\infty)}(X_i))$ are significantly longer than for an iid sequence. This implies that there is dependence

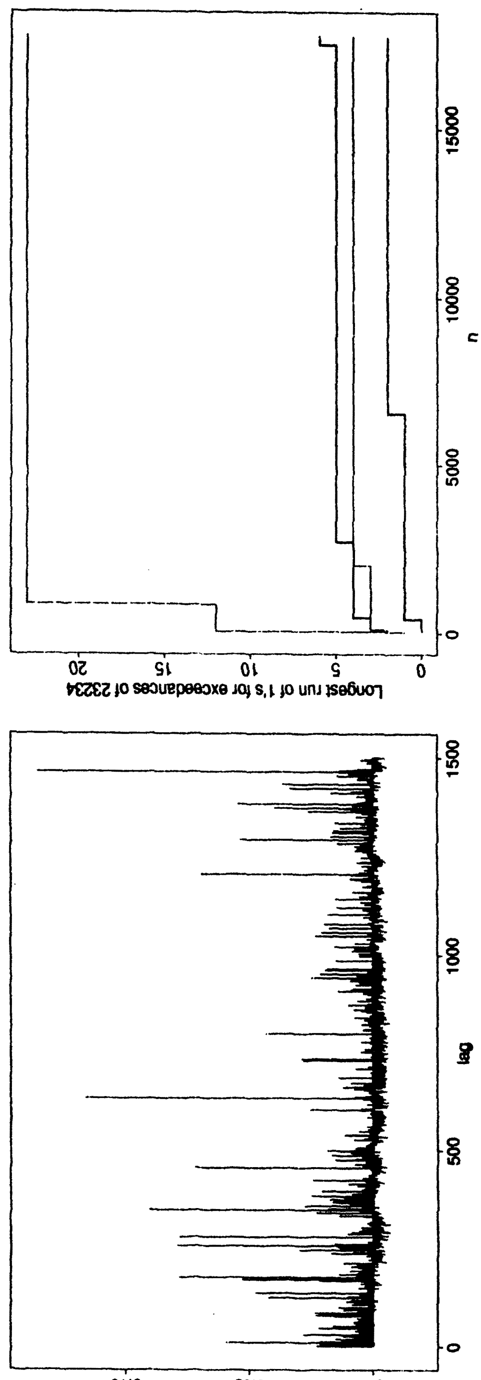

FIG. 4. *Left: the sample autocorrelation function of the teletraffic data for the first 1500 lags. Right: the longest runs of 1's (top curve) of the sequence* $I_{(23234,\infty)}(X_i)$. *The threshold* $u = 23234$ *corresponds to the 90% quantile of the data. The following curves (from above) are* β_n, *the longest runs of 1's of an iid sequence with* $p = 0.1$ *and* α_n.

in the teletraffic sequence (X_n) and that exceedances of the high threshold u occur in clusters.

Another tool for detecting dependence and clusters in data is the *extremal index* θ. For a stationary sequence (X_n) the quantity $\theta \in [0, 1]$ satisfies the relation

$$\lim_{n\to\infty} P\left(\max_{i=1,\dots,n} X_i \le u_n \right) = e^{-\tau\theta}$$

for (u_n) with $\lim_{n\to\infty} nP(X > u_n) = \tau > 0$. See EKM (1997), Section 8.1, for the definition, interpretation and statistical estimation of θ. It has been mentioned, for example in Hsing, Hüsler and Leadbetter (1988), that θ can be interpreted as the reciprocal of the mean cluster size $E\xi_i$ of the weak limit of the point processes of exceedances

$$\sum_{i=1}^{n} \delta_{n^{-1}i} I_{(u_n,\infty)}(X_i) \Rightarrow \sum_{i=1}^{\infty} \xi_i \delta_{\Gamma_i},$$

where δ_x is the Dirac measure at x, Γ_i are the points of a homogeneous unit rate Poisson process and (ξ_i) is the iid sequence of the cluster sizes, independent of (Γ_i). Clearly, for iid data, $\theta = 1$.

Natural estimators of θ are

$$\theta_n^{(1)} = \frac{k}{n} \frac{\ln(1 - K/k)}{\ln(1 - N/n)} \quad \text{and} \quad \theta_n^{(2)} = \frac{K}{N}, \tag{3.4}$$

where N is the number of exceedances of u_n by $X_1, \dots, X_n$, K is the number of blocks of length r:

$$X_{lr+1}, \dots, X_{(l+1)r}, \qquad l = 0, \dots, k-1,$$

in which at least one of the observations exceeds u_n. Further, $k = [n/r]$, $r = r_n \to \infty$, $r/n \to 0$, and the threshold sequence (u_n) is such that

$$\lim_{n\to\infty} nP(X > u_n) = \tau \quad \text{for some } \tau > 0.$$

For obvious reasons, this method of estimation is called *the blocks method*. In Figure 5 the behavior of $\theta_n^{(1)}$ and $\theta_n^{(2)}$ is illustrated as a function of the threshold $u = u_n$. Both estimator indicate that θ is about 0.9. This makes it clear that the observations (X_i) exhibit significant dependence in the tails.

APPENDIX

In this section we collect several results, some of which are needed for the proof of the main result of the paper in Section 2. Further results here describe additional extremal features of the dependent step random walk with steps (1.3).

In what follows, (Y_n) is a sequence of mean-zero random variables and

$$\widetilde{S}_n = Y_1 + \cdots + Y_n, \qquad n = 1, 2, \dots .$$

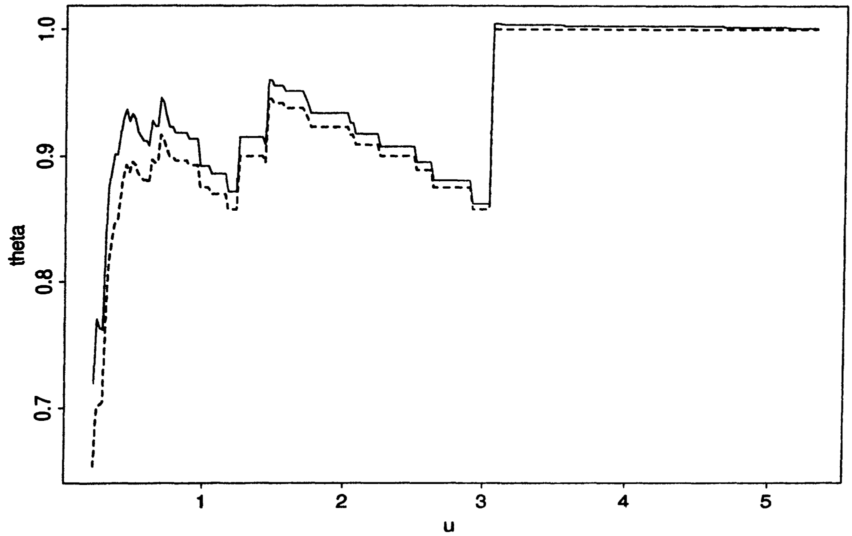

FIG. 5. *The estimators* $\theta_n^{(1)}$ *(solid line) and* $\theta_n^{(2)}$ *of the extremal index* θ *as a function of the threshold* u; *see* (3.4). *The unit on the u-axis is one million. Above the threshold* $u = 3$ *millions only seven values were observed*; *therefore the estimate of* 1 *for* θ *is not meaningful.*

Large deviations for sums of iid random variables with regularly varying tails. The following large deviation result for sums of iid random variables with regularly varying tails can be found in Nagaev (1969a, b) in the case $\alpha > 2$ and for $\alpha > 1$ in Cline and Hsing (1991).

LEMMA A.1. *Let* (Y_n) *be an iid sequence such that* $P(Y_1 > \lambda) = L(\lambda)\lambda^{-\alpha}$ *for some* $\alpha > 1$ *and a regularly varying function L. Then for every* $\delta > 0$,

$$\sup_{\lambda \geq \delta n} \left| \frac{P(\widetilde{S}_n > \lambda)}{nP(Y_1 > \lambda)} - 1 \right| \to 0, \qquad n \to \infty.$$

Tail estimate for sums of independent random variables. The following inequality is due to Prokhorov (1959); compare Petrov (1995), 2.6.1 on page 77.

LEMMA A.2. *Let* (Y_n) *be such that* $|Y_n| \leq c$ *for some* $c > 0$. *Then*

$$P(\widetilde{S}_n > \lambda) \leq \exp\left\{ -\frac{\lambda}{2c} \operatorname{arsinh} \frac{c\lambda}{2\operatorname{var}(\widetilde{S}_n)} \right\}, \qquad \lambda > 0.$$

The tail of an infinite series of independent random variables. In this subsection we consider the right tail of an infinite series

$$X = \sum_{j=-\infty}^{\infty} \varphi_j \varepsilon_j. \tag{A.1}$$

Here $(\varepsilon_n)_{n\in\mathbb{Z}}$ is a sequence of iid random variables satisfying the regular variation and tail balance conditions (1.4) with any $\alpha > 0$ (and not only $\alpha > 1$ as in the first part of this paper), and the coefficients φ_j are such that the infinite series (A.1) converges. It is a part of the folklore that under *some* conditions one has

$$\frac{P(X > x)}{P(|\varepsilon| > x)} \sim \sum_{j=-\infty}^{\infty} |\varphi_j|^\alpha [pI_{\{\varphi_j>0\}} + qI_{\{\varphi_j<0\}}] =: ||\varphi||_\alpha^\alpha. \tag{A.2}$$

We are aware of a large number of publications where such results are proved or referred to (and, undoubtedly, there are many publications that we are not aware of that deal with such results). However, these results are usually proved for particular cases, under generally more stringent conditions on (φ_n) than necessary, and are sometimes misquoted. We prove here (A.2) in all cases and under conditions that are close to being necessary. Observe that the very statement of (A.2) requires the condition $||\varphi||_\alpha < \infty$ which, in general, is not sufficient for a.s. convergence in (A.1). [This is just the three-series theorem; see, for example, Petrov (1995), Theorem 6.1 on page 205; it is easy to construct an example with $\alpha \in (0, 2]$ in which $||\varphi||_\alpha < \infty$ but the series does not converge.] We introduce the following conditions on (φ_n) which are more restrictive than $||\varphi||_\alpha < \infty$:

$$\begin{cases} \sum_{j=-\infty}^{\infty} \varphi_j^2 < \infty, & \text{for } \alpha > 2, \\ \sum_{j=-\infty}^{\infty} |\varphi_j|^{\alpha-\epsilon} < \infty, \quad \text{for some } \varepsilon > 0 & \text{for } \alpha \le 2. \end{cases} \tag{A.3}$$

LEMMA A.3. *Let the iid sequence $(\varepsilon_n)_{n\in\mathbb{Z}}$ satisfy the regular variation and tail balance conditions* (1.4) *with an $\alpha > 0$. If $\alpha > 1$, assume that $E\varepsilon = 0$. If the coefficients φ_n satisfy condition* (A.3), *then the infinite series* (A.1) *converges a.s. and* (A.2) *holds.*

The statement of Lemma A.3 coincides with the one of Lemma 4.24 in Resnick (1987) [attributed to Cline (1983a, b)] if $\alpha \le 1$, and with the one of Theorem 2.2 in Kokoszka and Taqqu (1996) if $\alpha \in (1, 2)$. For $\alpha > 2$ the conditions in (A.3) are the weakest possible since they are necessary for a.s. convergence of the series (A.1). For $0 < \alpha \le 2$ we will show in the sequel that the $\alpha - \epsilon$ power in (A.3) can be replaced by α under certain conditions on the slowly varying function L in (1.4).

PROOF. Convergence with probability 1 of the infinite series (A.1) follows from the three-series theorem [see, e.g., Petrov (1995), Theorem 6.1 on page 205], so we concentrate on the tails. For simplicity of representation we only consider one-sided processes $X = \sum_{j=0}^{\infty} \varphi_j \varepsilon_j$; the two-sided case is completely analogous. Write $X = X^{(K)} + Y^{(K)}$, where

$$X^{(K)} = \sum_{j=0}^{K} \varphi_j \varepsilon_j, \qquad K = 0, 1, \dots.$$

Then for $\delta \in (0, 1)$,

$$\begin{aligned} P\big(X^{(K)} > \lambda(1+\delta)\big) - P\big(Y^{(K)} \le -\delta\lambda\big) &\le P\big(X^{(K)} > \lambda(1+\delta), Y^{(K)} > -\delta\lambda\big) \\ &\le P(X > \lambda) \\ &\le P\big(X^{(K)} > \lambda(1-\delta)\big) + P\big(Y^{(K)} > \delta\lambda\big). \end{aligned} \tag{A.4}$$

Using standard results for convolutions of distributions with regularly varying tails [e.g., EKM (1997), Lemmas A3.26 or 1.3.1], it is not difficult to see that

$$\frac{P\big(X^{(K)} > \lambda\big)}{P(|\varepsilon| > \lambda)} \sim \sum_{j=1}^{K} |\varphi_j|^{\alpha} [pI_{\{\varphi_j > 0\}} + qI_{\{\varphi_j < 0\}}], \qquad \lambda \to \infty.$$

From this relation and (A.4) it follows that it suffices to prove that

$$\lim_{K\to\infty} \limsup_{\lambda\to\infty} \frac{P\big(|Y^{(K)}| > \lambda\big)}{P(|\varepsilon| > \lambda)} = 0. \tag{A.5}$$

We will show (A.5) with $P\big(|Y^{(K)}| > \lambda\big)$ replaced by $P\big(Y^{(K)} > \lambda\big)$; the case of $P\big(Y^{(K)} < -\lambda\big)$ is analogous.

It follows from Lemma 4.24 in Resnick (1987) that (A.5) holds for $\alpha \le 1$. Now assume that $\alpha \in (1, 2]$. Without loss of generality we may assume that the random variables ε_n are symmetric: indeed, since the sequence $\big(Y^{(K)}\big)_{K\ge1}$ is tight, we may choose an $M = M_{\varphi}$ independently of K so large that $P\big(Y^{(K)} \le M\big) \ge 0.5$ for all $K \ge 1$. Then for an independent copy $\widetilde{Y}^{(K)}$ of $Y^{(K)}$,

$$P\big(\widetilde{Y}^{(K)} - Y^{(K)} > \lambda - M\big) \ge P\big(\widetilde{Y}^{(K)} > \lambda, Y^{(K)} \le M\big) \ge \tfrac{1}{2} P\big(Y^{(K)} > \lambda\big),$$

and so if (A.5) is established for the sequence of symmetric sums $(\widetilde{Y}^{(K)} - Y^{(K)})_{K\ge1}$, then it will follow for the original sequence $\big(Y^{(K)}\big)_{K\ge1}$ as well.

Let (N_n) be a sequence of iid standard normal random variables, independent of (ε_n). Then, using a strong domination inequality [see, e.g., Theorem 3.2.1 in Kwapień and Woyczyński (1992)],

$$\begin{aligned} P\big(Y^{(K)} > \lambda\big) &\le c_1 P\Bigg(\sum_{j=K+1}^{\infty} \varphi_j N_j \varepsilon_j > c_2 \lambda \Bigg) \\ &= c_1 P\Bigg(N_1 \Bigg(\sum_{j=K+1}^{\infty} \varphi_j^2 \varepsilon_j^2 \Bigg)^{1/2} > c_2 \lambda \Bigg) \end{aligned} \tag{A.6}$$

for positive constants c_1 and c_2. Applying the result for the case $\alpha \le 1$ it follows that the tail of $\sum_{j=K+1}^{\infty} \varphi_j^2 \varepsilon_j^2$ is regularly varying with index $-\alpha/2 \in$

$[-1, 0)$. Since N_1 is independent of (ε_n), we conclude that the right-hand side expression in (A.6) is asymptotically of the order

$$c_3 \sum_{j=K+1}^{\infty} |\varphi_j|^{\alpha} P(\varepsilon > \lambda) \quad \text{as } \lambda \to \infty,$$

for some constant $c_3 > 0$ independent of K. This proves the lemma for $\alpha \in (1, 2]$.

In the general case $\alpha \in (2k-1, 2k]$ for some integer $k > 1$ one can follow the steps of the proof above: first symmetrize $Y^{(K)}$, then replace the Rademacher sequence by a Gaussian sequence and reduce the problem of bounding the tail to a corresponding task for $\sum_{j=K+1}^{\infty} \varphi_j^2 \varepsilon_j^2$. By doing so one reduces the index of regular variation to $\alpha/2 \in [k - 0.5, k]$, and one can use an obvious inductive procedure. □

For $\alpha \in (0, 2]$, the assumptions (A.3) on the coefficients φ_n can be relaxed provided the slowly varying function L in (1.4) satisfies certain additional assumptions. We consider two such possible assumptions:

$$L(\lambda_2) \le cL(\lambda_1) \quad \text{for } \lambda_0 < \lambda_1 < \lambda_2, \text{ some constants } c, \lambda_0 > 0, \tag{A.7}$$

$$L(\lambda_1\lambda_2) \le cL(\lambda_1)L(\lambda_2) \quad \text{for } \lambda_1, \lambda_2 \ge \lambda_0 > 0, \text{ some constants } c, \lambda_0 > 0. \tag{A.8}$$

LEMMA A.4. *Assume that the regular variation and tail balance condition* (1.4) *holds for some* $\alpha \in (0, 2]$, *that the infinite series* (A.1) *converges a.s.*,

$$\sum_{j=-\infty}^{\infty} |\varphi_j|^{\alpha} < \infty, \tag{A.9}$$

and one of the conditions (A.8) *or* (A.7) *is satisfied. Then relation* (A.2) *holds.*

Thus (A.2) holds not only under the condition (A.3), but also under (A.9) provided (A.8) or (A.7) hold. Notice that (A.8) holds for Pareto-like tails $P(\varepsilon > \lambda) \sim c\lambda^{-\alpha}$ and in particular for α-stable random variables ε. Moreover, (A.8) is satisfied for slowly varying functions $L(\lambda) = (\ln_k(\lambda))^{\beta}$ for any real β, where $\ln_k \lambda$ is the k times iterated logarithm of λ.

PROOF. Following the steps in the proof of Lemma A.3, it suffices to show that (A.5) holds for $\alpha < 1$. In the latter proof we used Lemma 4.24 in Resnick (1987); for its application we needed condition (A.3) for some $\varepsilon > 0$. A careful study of pages 228 and 229 in Resnick (1987) shows that this condition is only needed for proving

$$\lim_{\lambda \to \infty} \sum_{j=-\infty}^{\infty} \frac{P(|\varphi_j \varepsilon| > \lambda)}{P(\varepsilon > \lambda)} = \sum_{j=-\infty}^{\infty} |\varphi_j|^{\alpha}, \tag{A.10}$$

in which case the relation

$$\frac{P(|\varphi_j \varepsilon| > \lambda)}{P(|\varepsilon| > \lambda)} \le \text{const} |\varphi_j|^{\alpha - \varepsilon}$$

allows one to apply Lebesgue dominated convergence in (A.10) when interchanging the sum and the limit as $\lambda \to \infty$.

Now assume that (A.8) holds. Then we have for large λ,

$$\frac{P(|\varphi_j \varepsilon| > \lambda)}{P(\varepsilon > \lambda)} = \frac{|\varphi_j|^\alpha L\left(|\varphi_j|^{-1}\lambda\right)}{L(\lambda)} \le c|\varphi_j|^\alpha L\left(|\varphi_j|^{-1}\right) = cP(|\varphi_j \varepsilon| > 1).$$

Since series (A.1) converges a.s.,

$$\sum_{j=-\infty}^{\infty} P(|\varphi_j \varepsilon| > 1) < \infty,$$

by virtue of the three-series theorem. Using the latter bound, we may apply Lebesgue dominated convergence in (A.10), and so Lemma 4.24 in Resnick (1987) remains valid under assumption (A.8).

Now assume that (A.7) holds. Then for large λ,

$$\frac{P(|\varphi_j \varepsilon| > \lambda)}{P(|\varepsilon| > \lambda)} = \frac{|\varphi_j|^\alpha L\left(|\varphi_j|^{-1}\lambda\right)}{L(\lambda)} \le c|\varphi_j|^\alpha.$$

By virtue of (A.9) and the latter bound, one may apply Lebesgue dominated convergence to obtain (A.10). □

Large deviations for sums of linear processes with a regularly varying tail. In what follows, we extend the large deviation result of Lemma A.1 for sums of iid random variables to sums of linear processes. As before, (X_n) denotes a two-sided linear process (1.3) with iid noise variables ε_n with $E\varepsilon = 0$ satisfying the regular variation condition (1.4) for some $\alpha > 1$ and coefficients φ_n satisfying (1.5). We also assume that $\mu = 0$ in (1.3).

LEMMA A.5. *Let*

$$m_\varphi^{(0)} := \sum_{j=-\infty}^{\infty} \varphi_j.$$

If $m_\varphi^{(0)} < 0$ *we also assume that* $0 < p < 1$. *Then the relation*

$$\text{(A.11)} \quad \sup_{\lambda \ge \lambda_n} \left| \frac{P(S_n > \lambda)}{nP(|\varepsilon| > \lambda)} - \left(\left[\left(m_\varphi^{(0)}\right)^+ \right]^\alpha p + \left[\left(m_\varphi^{(0)}\right)^- \right]^\alpha q \right) \right| \to 0, \qquad n \to \infty$$

holds for every sequence (λ_n) *of positive numbers converging to infinity such that*

$$\text{(A.12)} \quad \sup_{\lambda \ge \lambda_n} \left| \frac{P\left(\sum_{j=1}^n \varepsilon_j > c\lambda\right)}{nP(\varepsilon > \lambda)} - c^{-\alpha} \right| \to 0, \qquad n \to \infty \textit{ for every fixed } c > 0,$$

if $m_\varphi^{(0)} > 0$;

$$\text{(A.13)} \quad \sup_{\lambda \ge \lambda_n} \left| \frac{P\left(\sum_{j=1}^n \varepsilon_j \le -c\lambda\right)}{nP(\varepsilon \le -\lambda)} - c^{-\alpha} \right| \to 0, \qquad n \to \infty \textit{ for every fixed } c > 0,$$

if $m_\varphi^{(0)} < 0$ *and* $0 < p < 1$.

REMARK A.6. If $\alpha > 1$, (A.12) holds for $\lambda_n = \delta n$ for any $\delta > 0$; see Lemma A.1. If $\alpha > 2$ then (A.12) holds for $\lambda_n = (a_n n \ln n)^{1/2}$, where (a_n) is any sequence of real numbers $a_n \to \infty$; see Nagaev (1979).

PROOF. We will prove the lemma in the case $m_\varphi^{(0)} > 0$. All other cases are similar. We have

$$\begin{aligned} S_n = \sum_{j=-\infty}^{\infty} \varepsilon_j \beta_{nj} &= \sum_{j=-\infty}^{0} \varepsilon_j \beta_{nj} + \sum_{j=n}^{\infty} \varepsilon_j \beta_{nj} + \sum_{j=1}^{n} \varepsilon_j \beta_{nj} \\ &= S_n^{(1)} + S_n^{(2)} + S_n^{(3)}. \end{aligned}$$

By virtue of Lemma A.3,

$$\begin{aligned} P\left(\left|S_n^{(1)} + S_n^{(2)}\right| > \lambda\right) &\le P\left(\sum_{j=-\infty}^{0} |\varepsilon_j| \sum_{i=1-j}^{\infty} |\varphi_i| + \sum_{j=n}^{\infty} |\varepsilon_j| \sum_{i=-\infty}^{n-j} |\varphi_i| > \lambda\right) \\ &\le \operatorname{const} P(|\varepsilon| > \lambda) \left[\sum_{j=-\infty}^{0} \left(\sum_{i=1-j}^{\infty} |\varphi_i|\right)^{\alpha} + \sum_{j=0}^{\infty} \left(\sum_{i=-\infty}^{-j} |\varphi_i|\right)^{\alpha}\right]. \end{aligned}$$

Since both λ and n converge to ∞, it suffices to show that

$$\sup_{\lambda \ge \lambda_n} \left| \frac{P(S_n^{(3)} > \lambda)}{n P(\varepsilon > \lambda)} - \left[m_\varphi^{(0)}\right]^{\alpha} \right| \to 0, \qquad n \to \infty.$$

We have

$$\begin{aligned} S_n^{(3)} &= m_\varphi^{(0)} \sum_{j=1}^{n} \varepsilon_j - \sum_{j=1}^{n} \varepsilon_j \left(\sum_{i=-\infty}^{1-j-1} \varphi_i + \sum_{i=n-j+1}^{\infty} \varphi_i \right) \\ &= S_n^{(3,1)} + S_n^{(3,2)}. \end{aligned}$$

By virtue of assumption (A.12) we have

$$\sup_{\lambda \ge \lambda_n} \left| \frac{P(S_n^{(3,1)} > \lambda)}{n P(\varepsilon > \lambda)} - \left[m_\varphi^{(0)}\right]^{\alpha} \right| \to 0, \qquad n \to \infty.$$

Thus it remains to show that the contribution of $S_n^{(3,2)}$ to the large deviations is negligible. Again by Lemma A.3,

$$\begin{aligned} P\left(\left|S_n^{(3,2)}\right| > \lambda\right) &\le P\left(\sum_{j=1}^{n} |\varepsilon_j| \sum_{i=n-j+1}^{\infty} |\varphi_i| > \lambda/2\right) + P\left(\sum_{j=1}^{n} |\varepsilon_j| \sum_{i=-\infty}^{1-j-1} |\varphi_i| > \lambda/2\right) \\ &\le \operatorname{const} P(|\varepsilon| > \lambda) \left[\sum_{j=0}^{\infty} \left(\sum_{i=j}^{\infty} |\varphi_i|\right)^{\alpha} + \sum_{j=1}^{\infty} \left(\sum_{i=-\infty}^{1-j-1} |\varphi_i|\right)^{\alpha}\right]. \end{aligned}$$

This concludes the proof of the lemma. $\square$

For iid random variables Y_i with regularly varying tail a simple computation shows that one also has, in addition to the conclusion of Lemma A.1 that

$$\sup_{\lambda \geq \delta n} \left| \frac{P(\max_{i=1,\ldots,n} Y_i > \lambda)}{nP(Y > \lambda)} - 1 \right| \to 0, \qquad n \to \infty,$$

for every $\delta > 0$, and so, in particular, for every such δ one has also

$$\sup_{\lambda \geq \delta n} \left| \frac{P(\widetilde{S}_n > \lambda)}{P(\max_{i=1,\ldots,n} Y_i > \lambda)} - 1 \right| \to 0, \qquad n \to \infty.$$

The following lemma shows that this result does in general not remain valid for the linear process (1.3).

LEMMA A.7. *Let*

$$\left[m_\varphi^{(1)}\right]^\alpha = p\varphi_+^\alpha + q\varphi_-^\alpha \quad \textit{and} \quad \varphi_+ = \sup_n \varphi_n^+, \qquad \varphi_- = \sup_n \varphi_n^-.$$

Under the assumptions of Lemma A.5 *the following relation holds for every* $\delta > 0$:

$$\text{(A.14)} \qquad \sup_{\lambda \geq \delta n} \left| \frac{P(\max_{i=1,\ldots,n} X_i > \lambda)}{nP(X > \lambda)} - \frac{\left[m_\varphi^{(1)}\right]^\alpha}{\|\varphi\|_\alpha^\alpha} \right| \to 0, \qquad n \to \infty,$$

where $\|\varphi\|_\alpha$ *is defined in* (1.6). *In particular,*

$$\text{(A.15)} \qquad \sup_{\lambda \geq \delta n} \left| \frac{P(S_n > \lambda)}{P(\max_{i=1,\ldots,n} X_i > \lambda)} - \frac{\left[\left(m_\varphi^{(0)}\right)^+\right]^\alpha p + \left[\left(m_\varphi^{(0)}\right)^-\right]^\alpha q}{\left[m_\varphi^{(1)}\right]^\alpha} \right| \to 0, \quad n \to \infty.$$

Observe that $m_\varphi^{(1)} > 0$ under the assumptions of Lemma A.5.

PROOF. It follows from Lemmas A.3 and A.5 that both claims will follow once we prove that for every $\delta > 0$,

$$\text{(A.16)} \qquad \sup_{\lambda \geq \delta n} \left| \frac{P(\max_{i=1,\ldots,n} X_i > \lambda)}{nP(|\varepsilon| > \lambda)} - \left[m_\varphi^{(1)}\right]^\alpha \right| \to 0, \qquad n \to \infty.$$

For $k > 1$ write

$$X_i = \sum_{j=-\infty}^{0} \varepsilon_j \varphi_{i-j} + \sum_{j=1}^{n+k} \varepsilon_j \varphi_{i-j} + \sum_{j=n+k+1}^{\infty} \varepsilon_j \varphi_{i-j} =: X_i^{(1)} + X_{i,n}^{(2)} + X_{i,n}^{(3)}.$$

Exactly as in the proof of Theorem 2.1 it is enough to consider the case $p = 1$ and $\varphi_+ > 0$. It follows from Lemma A.3 that

$$\sup_{\lambda \geq \delta n} \frac{P\left(\max_{i=1,\ldots,n} X_i^{(1)} > \lambda\right)}{nP(|\varepsilon| > \lambda)} = \sup_{\lambda \geq \delta n} \frac{P\left(X_1^{(1)} > \lambda\right)}{nP(|\varepsilon| > \lambda)} \to 0, \qquad n \to \infty,$$

and that

$$\lim_{k\to\infty}\limsup_{n\to\infty}\sup_{\lambda\ge\delta n}\frac{P\left(\max_{i=1,\dots,n}X_{i,n}^{(3)}>\lambda\right)}{nP(|\varepsilon|>\lambda)} \le \lim_{k\to\infty}\limsup_{n\to\infty}\sup_{\lambda\ge\delta n}\frac{\sum_{i=1}^{n}P\left(X_{i,n}^{(3)}>\lambda\right)}{nP(|\varepsilon|>\lambda)}=0.$$

Therefore, (A.16) will follow once we prove that for every $k>1$,

$$\text{(A.17)}\qquad \sup_{\lambda\ge\delta n}\left|\frac{P\left(\max_{i=1,\dots,n}X_{i,n}^{(2)}>\lambda\right)}{nP(|\varepsilon|>\lambda)}-\left[m_\varphi^{(1)}\right]^\alpha\right|\to 0,\qquad n\to\infty.$$

Repeating the argument of Lemma 2.6 one sees that there is a $\theta>0$ small enough compared to δ such that

$$\text{(A.18)}\qquad \sup_{\lambda\ge\delta n}\left|\frac{P\left(\max_{i=1,\dots,n}X_{i,n}^{(2)}>\lambda,\ \varepsilon_j\le\theta\lambda \text{ for } j=1,\dots,n+k\right)}{nP(|\varepsilon|>\lambda)}\right| \to 0,\quad n\to\infty.$$

Fix a $\tau\in(0,1)$. It follows from (A.18) and the argument of Lemma 2.8 that

$$\sup_{\lambda\ge\delta n}\left(\frac{P\left(\max_{i=1,\dots,n}X_{i,n}^{(2)}>\lambda\right)}{nP(|\varepsilon|>\lambda)} - \frac{P\left(\max_{i=1,\dots,n}X_{i,n}^{(2)}>\lambda,\ \varepsilon_j>(1-\tau)(\varphi_+)^{-1}\lambda \text{ for some } 1\le l\le n+k\right)}{nP(|\varepsilon|>\lambda)}\right) \to 0,\qquad n\to\infty.$$

Since

$$\begin{aligned}
&P\left(\max_{i=1,\dots,n}X_{i,n}^{(2)}>\lambda,\ \varepsilon_j>(1-\tau)(\varphi_+)^{-1}\lambda \text{ for some } 1\le l\le n+k\right)\\
&\quad\le P\left(\varepsilon_j>(1-\tau)(\varphi_+)^{-1}\lambda \text{ for some } 1\le l\le n+k\right)\\
&\quad\le (n+k)P\left(\varepsilon>(1-\tau)(\varphi_+)^{-1}\lambda\right),
\end{aligned}$$

we immediately conclude that

$$\limsup_{n\to\infty}\sup_{\lambda\ge\delta n}\left(\frac{P\left(\max_{i=1,\dots,n}X_{i,n}^{(2)}>\lambda\right)}{nP(|\varepsilon|>\lambda)}-\left[m_\varphi^{(1)}\right]^\alpha\right)\le\left((1-\tau)^{-\alpha}-1\right)\left[m_\varphi^{(1)}\right]^\alpha,$$

and letting $\tau\to 0$ we obtain

$$\text{(A.19)}\qquad \limsup_{n\to\infty}\sup_{\lambda\ge\delta n}\left(\frac{P\left(\max_{i=1,\dots,n}X_{i,n}^{(2)}>\lambda\right)}{nP(|\varepsilon|>\lambda)}-\left[m_\varphi^{(1)}\right]^\alpha\right)\le 0.$$

Furthermore, define for a $K \geq 1$,

$$X^{(2)}_{i,n,K} = \sum_{j=K+1}^{n-K} \varepsilon_j \varphi_{i-j}.$$

Then

$$\text{(A.20)}\qquad \sup_{\lambda \geq \delta n} \left| \frac{P\big(\max_{i=1,\dots,n} X^{(2)}_{i,n} > \lambda\big)}{nP(|\varepsilon| > \lambda)} - \frac{P\big(\max_{i=1,\dots,n} X^{(2)}_{i,n,K} > \lambda\big)}{nP(|\varepsilon| > \lambda)} \right| \to 0, \qquad n \to \infty.$$

Let, once again, τ be a number in $(0, 1)$. Choose a K so large that

$$\sup_{1-K \leq j \leq K} \varphi_j^+ \geq (1+\tau)^{-1} \varphi_+.$$

We have by the choice of K, for any $\theta > 0$ small enough, as in Lemma 2.6,

$$\begin{aligned}
&P\Big(\max_{i=1,\dots,n} X^{(2)}_{i,n,K} > \lambda\Big)\\
&\quad \geq P\Big(\max_{i=1,\dots,n} X^{(2)}_{i,n,K} > \lambda,\ \varepsilon_j \geq (1+\tau)^2 (\varphi_+)^{-1}\lambda\\
&\qquad\quad \text{for some } K+1 \leq l \leq n-K,\ \varepsilon_l \leq \theta\lambda \text{ for all other } K+1 \leq l \leq n-K\Big)\\
&\quad = \sum_{l=K+1}^{n-K} P\Big(\max_{i=1,\dots,n} X^{(2)}_{i,n,K} > \lambda, \varepsilon_l > (1+\tau)^2 (\varphi_+)^{-1}\lambda,\\
&\qquad\qquad\qquad \varepsilon_j \leq \theta\lambda,\ j = K+1, \dots, n-K,\ j \neq l\Big)\\
&\quad \geq \sum_{l=K+1}^{n-K} P\big(\varepsilon_l > (1+\tau)^2 (\varphi_+)^{-1}\lambda\big) - h_n(\lambda),
\end{aligned}$$

with h_n satisfying

$$\sup_{\lambda \geq \delta n} \left| \frac{h_n(\lambda)}{nP(|\varepsilon| > \lambda)} \right| \to 0, \qquad n \to \infty.$$

We immediately conclude by (A.20) that

$$\limsup_{n\to\infty} \sup_{\lambda \geq \delta n} \left(\left[m_\varphi^{(1)}\right]^\alpha - \frac{P\big(\max_{i=1,\dots,n} X^{(2)}_{i,n} > \lambda\big)}{nP(|\varepsilon| > \lambda)} \right) \leq \big(1 - (1+\tau)^{-2\alpha}\big)\left[m_\varphi^{(1)}\right]^\alpha,$$

and letting $\tau \to 0$ we obtain

$$\text{(A.21)}\qquad \limsup_{n\to\infty} \sup_{\lambda \geq \delta n} \left(\left[m_\varphi^{(1)}\right]^\alpha - \frac{P\big(\max_{i=1,\dots,n} X^{(2)}_{i,n} > \lambda\big)}{nP(|\varepsilon| > \lambda)} \right) \leq 0.$$

The claim of the lemma now follows from (A.19) and (A.21). $\square$

Acknowledgments. This research has been conducted during mutual visits of each of the authors to the home institution of the other author. Both authors thank Cornell University and the University of Groningen for the hospitality. The authors take pleasure in thanking the anonymous referee for the extraordinarily detailed report that contributed significantly to improvement of the way the material is presented in this paper.

REFERENCES

ASMUSSEN, S., SCHMIDLI, H. and SCHMIDT, V. (1999). Tail probabilities for nonstandard risk and queueing processes with subexponential jumps. *Adv. Appl. Probab.* **31** 442–447.

BACCELLI, F. and BRÉMAUD, P. (1994). *Elements of Queueing Theory. Palm-Martingale Calculus and Stochastic Recurrences*. Springer, Berlin.

BINGHAM, N., GOLDIE, C. and TEUGELS, J. (1987). *Regular Variation*. Cambridge Univ. Press.

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

CHISTYAKOV, V. (1964). A theorem on sums of independent random variables and its applications to branching random processes. *Theory Probab. Appl.* **9** 640–648.

CLINE, D. (1983a). Estimation and linear prediction for regression, autoregression and ARMA with infinite variance data. Ph.D. dissertation, Colorado State Univ.

CLINE, D. (1983b). Infinite series of random variables with regularly varying tails. Technical Report 83-24, Institute Applied Math. Statist., Univ. British Columbia, Vancouver, B.C.

CLINE, D. and HSING, T. (1991). Large deviation probabilities for sums and maxima of random variables with heavy or subexponential tails. Preprint, Texas A&M Univ.

CUNHA, C., BESTAVROS, A. and CROVELLA, M. (1995). Characteristics of WWW client-based traces. Preprint. Available as BU-CS-95-010 from {crovella, best}cs.bu.edu.

DAVIS, R. and MIKOSCH, T. (1998). The sample autocorrelations of heavy-tailed stationary processes with applications to ARCH. *Ann. Statist.* **26** 2049–2080.

DAVIS, R. and RESNICK, S. (1985). Limit theory for moving averages of random variables with regularly varying tail probabilities. *Ann. Probab.* **13** 179–195.

DAVIS, R. and RESNICK, S. (1986). Limit theory for the sample covariance and correlation functions of moving averages. *Ann. Statist.* **14** 533–558.

DAVIS, R. and RESNICK, S. (1996). Limit theory for bilinear processes with heavy-tailed noise. *Ann. Appl. Probab.* **6** 1191–1210.

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

EMBRECHTS, P. and VERAVERBEKE, N. (1982). Estimates for the probability of ruin with special emphasis on the possibility of large claims. *Insurance Math. Econom.* **1** 55–72.

HSING, T., HÜSLER, J. and LEADBETTER, M. (1988). On the exceedance point process for a stationary sequence. *Probab. Theory Related Fields* **78** 97–112.

KOKOSZKA, P. and TAQQU, M. (1996). Parameter estimation for infinite variance fractional ARIMA. *Ann. Statist.* **24** 1880–1913.

KWAPIEŃ, S. and WOYCZYŃSKI, N. (1992). *Random Series and Stochastic Integrals: Single and Multiple*. Birkhäuser, Boston.

NAGAEV, A. (1969a). Integral limit theorems for large deviations when Cramér's condition is not fulfilled I, II. *Theory Probab. Appl.* **14** 51–64, 193–208.

NAGAEV, A. (1969b). Limit theorems for large deviations where Cramér's conditions are violated. *Izv. Akad. Nauk UzSSR Ser. Fiz.-Mat. Nauk* **6** 17–22. (In Russian.)

NAGAEV, S. (1979). Large deviations of sums of independent random variables. *Ann. Probab.* **7** 745–789.

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

PROKHOROV, Y. (1959). An extremal problem in probability theory. *Theory Probab. Appl.* **4** 201–204.

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 Practial Guide to Heavy Tails: Statistical Techniques for Analysing Heavy Tailed Distributions* 219–240. Birkhäuser, Boston.

RESNICK, S., SAMORODNITSKY, G. and XUE, F. (1999). How misleading can sample ACF's of stable MA's be? (Very!). *Ann. Appl. Probab.* **9** 797–817.

ROSENBLATT, M. (1962). *Random Processes*. Oxford Univ. Press.

STRASSEN, V. (1965). The existence of probability measures with given marginals. *Ann. Math. Statist.* **36** 423–439.

DEPARTMENT OF MATHEMATICS
P.O. BOX 800
UNIVERSITY OF GRONINGEN
NL-9700 AV GRONINGEN
THE NETHERLANDS
E-MAIL: mikosch@math.rug.nl

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