Statistica Sinica **24** (2014), 957-969
doi:http://dx.doi.org/10.5705/ss.2012.161

EXACT MODERATE AND LARGE DEVIATIONS FOR LINEAR PROCESSES

Magda Peligrad[1], Hailin Sang[2], Yunda Zhong[3] and Wei Biao Wu[3]

[1] *University of Cincinnati,* [2] *University of Mississippi*
and [3] *University of Chicago*

Abstract: Large and moderate deviation probabilities play an important role in many applied areas, such as insurance and risk analysis. This paper studies the exact moderate, and large deviation asymptotics in non-logarithmic form for linear processes with independent innovations. The linear processes we analyze are general and they include the long memory case. We give an asymptotic representation for the probability of the tail of the normalized sums and specify the zones in which it can be approximated either by a standard normal distribution or by the marginal distribution of the innovation process. The results are then applied to regression estimates, moving averages, fractionally integrated processes, linear processes with regularly varying exponents, and functions of linear processes. We also consider the computation of value at risk and expected shortfall, fundamental quantities in risk theory and finance.

Key words and phrases: Large deviation, linear process, long memory, moderate deviation, non-logarithmic asymptotics, zone of normal convergence.

1. Introduction and Notations

Let $(\xi_i)_{i\in\mathbb{Z}}$ be a sequence of independent and identically distributed centered random variables with finite second moment, and c_{ni} a sequence of constants. This paper focuses on the moderate and large deviations in non-logarithmic form for the linear process

$$S_n = \sum_{i=1}^{k_n} c_{ni}\xi_i. \tag{1.1}$$

This class of linear processes is versatile enough to help analyze regression estimates, moving averages that include long memory processes, linear processes with regularly varying coefficients and fractionally integrated processes.

Our goal is to find an asymptotic representation for the tail probabilities of the normalized sums defined by (1.1). Estimations of deviation probabilities occur in a natural way in many applied areas, including insurance and risk analysis.

We aim to find a function $N_n(x)$ such that, as $n \to \infty$,

$$\frac{\mathbb{P}(S_n \geq x\sigma_n)}{N_n(x)} = 1 + o(1), \text{ where } \sigma_n^2 = \|S_n\|_2^2 = \mathbb{E}\xi_1^2 \sum_{i=1}^{k_n} c_{ni}^2. \tag{1.2}$$

If $x \geq 0$ is fixed, then (1.2) is the central limit theorem by letting $N_n(x) = 1 - \Phi(x)$, where $\Phi(x)$ is the standard normal distribution function. We call $\mathbb{P}(S_n/\sigma_n \geq x)$ the *moderate* or *large deviation* probabilities depending on the speed of convergence $x = x_n \to \infty$. These tail probabilities of rare events can be very small. Here we call (1.2) the *exact approximation*, which is more accurate than the logarithmic version

$$\frac{\log \mathbb{P}(S_n/\sigma_n \geq x)}{\log N_n(x)} = 1 + o(1), \tag{1.3}$$

which is often used in the literature in the context of large or moderate deviation. For example, if $\mathbb{P}(S_n/\sigma_n \geq x) = 10^{-4}$ and $N_n(x) = 10^{-5}$, then their logarithmic ratio is 0.8, not very different from 1, while the ratio for the exact version (1.2) is as big as 10. A multiplicative factor of this order can cause substantially different industrial standards in designing projects that can survive natural disasters.

As early as 1929, Khinchin considered the problem of moderate and large deviation probabilities in non-logarithmic form for independent Bernoulli random variables. The first large deviation probability result appeared in Nagaev (1965). Nagaev (1969) studied large deviation probabilities of i.i.d. random variables with regularly varying tails. Mikosch and Nagaev (1998) applied the large deviation probabilities for heavy-tailed random variables to insurance mathematics. The review work on this topic can be found in Nagaev (1979) and Rozovski (1993). Rubin and Sethuraman (1965), Slastnikov (1978), and Frolov (2005) considered the moderate or large deviations for arrays of independent random variables. Nagaev (1979) presented a useful result: in (1.1) assume $k_n = n$, $c_{ni} \equiv 1$, and that ξ_i has a regularly varying right tail,

$$\mathbb{P}(\xi_0 \geq x) = \frac{h(x)}{x^t} \text{ as } x \to \infty \text{ for some } t > 2, \tag{1.4}$$

where $h(x)$ is a slowly varying function (Bingham, Goldie, and Teugels (1987)). Here $\lim_{x\to\infty} h(\lambda x)/h(x) = 1$ for all $\lambda > 0$. If, in addition, for some $p > 2$ ξ_0 has absolute moment of order p, then

$$\mathbb{P}(\sum_{i=1}^{n} \xi_i \geq x\sigma_n) = (1 - \Phi(x))(1 + o(1)) + n\mathbb{P}(\xi_0 \geq x\sigma_n)(1 + o(1)) \tag{1.5}$$

for $n \to \infty$ and $x \geq 1$. Note that (1.5) implies (1.2) with

$$N_n(x) = (1 - \Phi(x)) + n\mathbb{P}(\xi_0 \geq x\sigma_n). \tag{1.6}$$

Hence if $1-\Phi(x)=o[n\mathbb{P}(\xi_0\geq x\sigma_n)]$ (resp. $n\mathbb{P}(\xi_0\geq x\sigma_n)=o(1-\Phi(x))$), then in (1.2) we can also choose $N_n(x)=1-\Phi(x)$ (resp. $N_n(x)=n\mathbb{P}(\xi_0\geq x\sigma_n)$).

The study of moderate and large deviation probabilities in non-logarithmic form for dependent random variables is still in its initial stage. Ghosh (1974) considered moderate deviations for m-dependent random variables. Chen (2001) obtained a moderate deviation result for Markov processes. Grama (1997) and Grama and Haeusler (2006) investigated the martingale case. Mikosch and Samorodnitsky (2000) obtained the limit $\lim_{x\to\infty}\mathbb{P}(X_k>x)/\mathbb{P}(|\xi_0|\geq x)$, where $X_k=\sum_{j=-\infty}^{\infty}a_{k-j}\xi_j$, ξ_j are i.i.d. with mean 0 satisfying the regular variation and tail balance conditions for index $t>1$ and coefficients a_j satisfying $\sum_{j=-\infty}^{\infty}|ja_j|<\infty$. Wu and Zhao (2008) studied moderate deviations for stationary processes which applies to many time series models. However the results in the latter two papers can only be applied to linear processes with short memory and/or their transformations.

For analyzing linear processes with long memory and for obtaining other interesting applications, we study processes of type (1.1). Under mild conditions on the coefficients, we point out the zones in which the deviation probabilities can be approximated either by a standard normal distribution or by using the distribution of ξ_0. Our main result is that (1.5) holds in our case with

$$N_n(x)=(1-\Phi(x))+\sum_{i=1}^{k_n}\mathbb{P}(c_{ni}\xi_0\geq x\sigma_n).$$

The paper has the following structure. Section 2 presents a general moderate and large deviation result and various applications. Section 3 illustrates the results of a numerical study. The proofs are given in the supplementary Material of this paper (Peligrad et al. (2013)).

We introduce here the notation that will be used throughout this paper: $a_n\sim b_n$ means that $\lim_{n\to\infty}a_n/b_n=1$, $a_n=O(b_n)$ and also $a_n\ll b_n$ mean $\limsup_{n\to\infty}a_n/b_n<\infty$; $a_n=o(b_n)$ if $\lim_{n\to\infty}a_n/b_n=0$. By $\|X\|_p$ we denote $(\mathbb{E}|X|^p)^{1/p}$. The notation $l(\cdot)$, $h(\cdot)$, and $\ell(\cdot)$ denote slowly varying functions. By convention, $0/0$ is interpreted as 0.

2. Main Results

Throughout, we assume the following,

Condition A. $(\xi_i)_{i\in\mathbb{Z}}$, are i.i.d. centered random variables with finite second moment, $\sigma^2=\mathbb{E}\xi_0^2$.

2.1. General linear processes

Our first results apply to general linear processes of type (1.1) with i.i.d. innovations. For $c_{ni}>0$ and $t>0$, let

$$B_{nt} = \sum_{i=1}^{k_n} c_{ni}^t, \tag{2.1}$$

$$\sigma_n^2 = var(S_n) = B_{n2}\mathbb{E}\xi_0^2, \tag{2.2}$$

$$D_{nt} = B_{n2}^{-t/2} B_{nt}. \tag{2.3}$$

Our basic assumption is the uniform asymptotic negligibility of the variance of individual summands,

$$\max_{1\le i\le k_n} c_{ni}^2/\sigma_n^2 \to 0. \tag{2.4}$$

Our first theorem extends Nagaev's result in (1.5) to general linear processes.

Theorem 1. *Assume that $(\xi_i)_{i\in\mathbb{Z}}$ satisfies Condition A and, for a certain $t > 2$, the right tail condition (1.4). Suppose for a certain $p > 2$, $\|\xi_0\|_p < \infty$, that $c_{ni} > 0$, and (2.4) is satisfied. Let $(x_n)_{n\ge1}$ be any sequence such that for some $c > 0$ we have $x_n \ge c$ for all n. Then, as $n \to \infty$,*

$$\mathbb{P}\left(S_n \ge x_n\sigma_n\right) = (1 + o(1)) \sum_{i=1}^{k_n} \mathbb{P}(c_{ni}\xi_0 \ge x_n\sigma_n) + (1 - \Phi(x_n))(1 + o(1)). \tag{2.5}$$

Remark 1. In (2.5), as well as in (2.6) and (2.7) below, by $o(1)$ we understand a function, that depends on x_n and on the underlying distribution, with the property that its limit as $n \to \infty$ is zero. The sequence $(x_n)_{n\ge1}$ may be bounded or may converge to infinity.

Corollary 1. *Under the conditions of Theorem 1, for $x_n \ge a(\ln D_{nt}^{-1})^{1/2}$ with $a > 2^{1/2}$ we have*

$$\mathbb{P}(S_n \ge x_n\sigma_n) = (1 + o(1)) \sum_{i=1}^{k_n} \mathbb{P}(c_{ni}\xi_0 \ge x_n\sigma_n) \text{ as } n \to \infty. \tag{2.6}$$

If $0 < x_n \le b(\ln D_{nt}^{-1})^{1/2}$ with $b < 2^{1/2}$, we have

$$\mathbb{P}\left(S_n \ge x_n\sigma_n\right) = (1 - \Phi(x_n))(1 + o(1)) \text{ as } n \to \infty. \tag{2.7}$$

Remark 2. Here (2.6) and (2.7) assert different approximations for the tail probability $\mathbb{P}(S_n \ge x\sigma_n)$: moderate behavior for $x = x_n$ smaller than a threshold; large deviation type of behavior for x larger than another threshold. The behavior at the boundary $\sqrt{2}(\ln D_{nt}^{-1})^{1/2}$ is more subtle and depends on the slowly varying function $h(\cdot)$. For the special case $\lim_{x\to\infty} h(x) \to h_0 > 0$, we have

$$\frac{\mathbb{P}(S_n \ge x\sigma_n)}{N_n(x)} = 1 + o(1), \text{ where } N_n(x) = (1 - \Phi(x)) + \frac{h_0}{(\sigma x)^t} D_{nt}. \tag{2.8}$$

If $x \ge a(\ln D_{nt}^{-1})^{1/2}$ with $a > 2^{1/2}$, then $N_n(x) \sim h_0 D_{nt}/(\sigma x)^t$.

The proofs of these results are of independent interest. We shall see in the next two theorems that a result similar to (2.6) holds without the assumption of the finite moment of order $p > 2$ while the moderate deviation (2.7) does not require a regularly varying right tail.

Theorem 2. *Assume that $(\xi_i)_{i\in\mathbb{Z}}$ satisfies Condition A and, for a certain $t > 2$,* (1.4)*. If $c_{ni} > 0$ is a sequence of constants satisfying* (2.4)*, then for any sequence $x_n \geq C_t(\ln D_{nt}^{-1})^{1/2}$ with $C_t > e^{t/2}(t+2)/\sqrt{2}$, (2.6) holds.*

Theorem 3. *Assume that $(\xi_i)_{i\in\mathbb{Z}}$ satisfies Condition A and, for a certain $p > 2$, $\|\xi_0\|_p < \infty$. If* (2.4) *is satisfied and $x_n^2 \leq 2\ln(D_{np}^{-1})$, then* (2.7) *holds.*

2.2. Applications to linear regression estimates

Many statistical procedures, such as estimation of regression coefficients, produce linear statistics of type (1.1). See for instance Chapter 9 in Beran (1994), for the case of parametric regression, or the paper by Robinson (1997), where kernel estimators are used for nonparametric regression. Here we consider the simple parametric regression model $Y_i = \beta\alpha_i + \xi_i$, where ξ_i are i.i.d. centered errors with $\mathbb{E}\xi_1^2 = \sigma^2$, (α_i) is a sequence of positive real numbers, and β is the parameter of interest. The least squares estimator $\hat{\beta}_n$ of β, based on a sample of size n, satisfies

$$S_n := \hat{\beta}_n - \beta = \frac{1}{\sum_{i=1}^n \alpha_i^2} \sum_{i=1}^n \alpha_i \xi_i, \tag{2.9}$$

so the representation (1.1) holds with $c_{ni} = \alpha_i / (\sum_{i=1}^n \alpha_i^2)$. Let $A_{nt} = \sum_{i=1}^n \alpha_i^t$. Notice that $var(S_n) = \sigma^2 / A_{n2}$ and assume that

$$\lim_{n\to\infty} A_{n2}^{-1} \max_{1\leq i\leq n} \alpha_i^2 = 0. \tag{2.10}$$

Corollary 2. (i) *Assume that $(\xi_i)_{i\in\mathbb{Z}}$ and $x = x_n$ satisfy the conditions in Theorem* 1*. Under* (2.10) *we have*

$$\mathbb{P}\Big(\hat{\beta}_n - \beta \geq \frac{x\sigma}{A_{n2}^{1/2}}\Big) = (1+o(1)) \sum_{i=1}^n \mathbb{P}\Big(\xi_i \geq \frac{x\sigma A_{n2}^{1/2}}{\alpha_i}\Big) + (1+o(1))(1-\Phi(x)).$$

(ii) *If $x > 0$ and $x^2 \leq 2\ln(A_{n2}^{t/2}/A_{nt})$, under the conditions in Theorem* 1 *we have*

$$\mathbb{P}\Big(\hat{\beta}_n - \beta \geq \frac{x\sigma}{A_{n2}^{1/2}}\Big) = (1+o(1))(1-\Phi(x)).$$

(iii) *If $x > 0$ and $x^2 \geq C_t^2 \ln(A_{n2}^{t/2}/A_{nt})$ with $C_t^2 > 2$, under the conditions in Theorem* 1*,*

$$\mathbb{P}\Big(\hat{\beta}_n - \beta \geq \frac{x\sigma}{A_{n2}^{1/2}}\Big) = (1+o(1))\sum_{i=1}^{n} \mathbb{P}\Big(\xi_i \geq \frac{x\sigma A_{n2}^{1/2}}{\alpha_i}\Big).$$

Similar results as in Theorems 2 and 3 can also be easily formulated.

Theorems 1, 2 and 3 are applicable to the nonlinear regression model $y_i = g(x_i) + \xi_i$, $1 \leq i \leq n$, where $g(x)$ is an unknown function and ξ_i is the noise. Let x_i be the deterministic design points. Then the Nadaraya-Watson estimate $\hat{g}_n$ satisfies

$$\hat{g}_n(x) - \mathbb{E}\hat{g}_n(x) = \sum_{i=1}^{n} c_{ni}(x)\xi_i$$

where, letting K be a kernel function and h_n be bandwidths,

$$c_{ni}(x) = \frac{K\left((x_i - x)/h_n\right)}{\sum_{i=1}^{n} K\left((x_i - x)/h_n\right)}.$$

Therefore it is of the type (1.1).

2.3. Application to moving averages

We consider the sum $S_n = \sum_{k=1}^{n} X_k$, where

$$X_k = \sum_{j=-\infty}^{\infty} a_{k-j}\xi_j. \tag{2.11}$$

We assume that $\sum_{i\in\mathbb{Z}} a_i^2 < \infty$, the necessary and sufficient condition for the existence of X_1. Now $S_n = \sum_{i=-\infty}^{\infty} b_{ni}\xi_i$ is of form (1.1) with

$$b_{ni} = a_{1-i} + \cdots + a_{n-i} \tag{2.12}$$

and $k_n = \infty$. Assume $b_{ni} > 0$ for all i and let

$$U_{nt} = \Big(\sum_i b_{ni}^2\Big)^{-t/2} \sum_i b_{ni}^t. \tag{2.13}$$

Set $\sigma_n^2 = \mathbb{E}\xi_0^2 \sum_i b_{ni}^2$. We know from Peligrad and Utev (1997) that under the assumption $\sigma_n^2 \to \infty$ we have

$$\sigma_n^{-2} \sup_i b_{ni}^2 \to 0 \text{ as } n \to \infty. \tag{2.14}$$

Therefore (2.4) is automatically satisfied.

Corollary 3. *Assume that* $(X_n)_{n\geq 1}$ *is defined by* (2.11) *and* $\sigma_n^2 \to \infty$.

(i) *If* $(\xi_i)_{i\in\mathbb{Z}}$ *and* x_n *satisfy the conditions of Theorem 1 and* $b_{ni} > 0$*, then* (2.5) *holds.*

(ii) *Let* $(\xi_i)_{i\in\mathbb{Z}}$ *be as in Theorem* 2*. Assume* $b_{ni} > 0$*. Then* (2.6) *holds for the sequence* $x_n \geq C_t(\ln U_{nt}^{-1})^{1/2}$ *with* $C_t > e^{t/2}(t+2)/\sqrt{2}$*.*

(iii) *If* $(\xi_i)_{i\in\mathbb{Z}}$ *is as in Theorem* 3*, then* (2.7) *holds for* $x_n^2 \leq 2\ln(U_{np}^{-1})$*.*

This corollary applies to general linear processes including the long memory processes with $\sum_i |a_i| = \infty$. Asymptotic properties for long memory processes can be quite different from those of processes with short memory, partially because the variance of the partial sum goes to infinity at an order different than n; see for example, Ho and Hsing (1997), Robinson (2003), Doukhan, Oppenheim, and Taqqu (2003), among others. Hall (1992) gave a Berry-Esseen bound for the convergence rate in the central limit theorem.

We apply the corollary to the important case of causal long-memory processes with

$$a_i = l(i+1)(1+i)^{-r},\ i \geq 0,\ \text{with } 1/2 < r < 1,\ \text{and } a_i = 0 \text{ otherwise.} \tag{2.15}$$

Here $l(\cdot)$ is a slowly varying function so the results can be given in a more precise form. In this case,

$$X_k = \sum_{j=-\infty}^{k} a_{k-j}\xi_j. \tag{2.16}$$

Let $a_0 = 1$. Long memory linear processes cover the fractional ARIMA processes (cf., Granger and Joyeux (1980); Hosking (1981)), which play an important role in financial time series modeling and application. As a special case, let $0 < d < 1/2$ and B be the backward shift operator with $B\varepsilon_k = \varepsilon_{k-1}$, and consider

$$X_k = (1-B)^{-d}\xi_k = \sum_{i\geq 0} a_i\xi_{k-i},\ \text{where } a_i = \frac{\Gamma(i+d)}{\Gamma(d)\Gamma(i+1)}.$$

Here $\lim_{n\to\infty} a_n/n^{d-1} = 1/\Gamma(d)$. These processes have long memory because $\sum_{j\geq 0} |a_j| = \infty$.

Corollary 4. *Assume* (2.15)*. If* $(\xi_i)_{i\in\mathbb{Z}}$ *satisfies the conditions of Theorem 1 then* (2.5) *holds. In particular* (2.6) *holds for* $x_n \geq c_1(\ln n)^{1/2}$ *with* $c_1 > (t-2)^{1/2}$*, while* (2.7) *holds if* $0 < x_n \leq c_2(\ln n)^{1/2}$ *with* $c_2 < (t-2)^{1/2}$*.*

Corollary 5. (i) *Let* $(\xi_i)_{i\in\mathbb{Z}}$ *be as in Theorem* 2*. Then* (2.6) *holds for* $x_n > c_1(\ln n)^{1/2}$ *with* $c_1 > (t-2)^{1/2}e^{t/2}(t+2)/2$*.*

(ii) *Let* $(\xi_i)_{i\in\mathbb{Z}}$ *be as in Theorem* 3*. Then* (2.7) *holds if* $x_n^2 \leq (p-2)\ln n$*.*

2.4. Application to risk measures

In risk theory and finance, value at risk (VaR) and expected shortfall (ES) play a fundamental role; see Jorion (2006), Holton (2003), McNeil, Frey, and Embrechts (2005), Acerbi and Tasche (2002), among others. Mathematically, they are equivalent to quantiles and tail conditional expectations. In practice one is most interested in their extremal behavior which corresponds to tail quantiles. Despite their importance, however, their computation can be quite difficult and the related asymptotic justification is far from trivial.

Here we apply Theorem 1 and provide approximate formulae for extremal quantiles and tail conditional expectations for S_n at (1.1). If $\lim_{x\to\infty} h(x) \to h_0 > 0$, by (2.8) and Theorem 1,

$$\mathbb{P}\left(S_n \geq x\sigma_n\right) = (1+o(1))\frac{h_0}{(\sigma x)^t}D_{nt} + (1-\Phi(x))(1+o(1)).$$

Given $\alpha \in (0,1)$, let $q_{\alpha,n}$ satisfy $\mathbb{P}(S_n \geq q_{\alpha,n}) = \alpha$. Elementary calculations show that $q_{\alpha,n}$ can be approximated by $x_\alpha \sigma_n$ in the sense that $\lim_{n\to\infty} x_\alpha\sigma_n/q_{\alpha,n} = 1$, where $x = x_\alpha$ is the solution to the equation

$$\frac{h_0}{(\sigma x)^t}D_{nt} + (1-\Phi(x)) = \alpha.$$

In particular, if $\alpha \leq h_0 D_{nt}((a\sigma)^2 \ln D_{nt}^{-1})^{-t/2}$ with $a > 2^{1/2}$, then, by Corollary 1, we can approximate $q_{\alpha,n}$ by $\sigma^{-1}(h_0 D_{nt}/\alpha)^{1/t}\sigma_n = \sigma^{-1}(B_{nt}h_0/\alpha)^{1/t}$. The approximation is understood in the sense that $\sigma^{-1}(B_{nt}h_0/\alpha)^{1/t}/q_{\alpha,n} \to 1$ as $n \to \infty$, and the tail conditional expectation or expected shortfall is computed as

$$\begin{aligned}\mathbb{E}(S_n|S_n \geq q_{\alpha,n}) &= \frac{q_{\alpha,n}\mathbb{P}(S_n \geq q_{\alpha,n}) + \int_{q_{\alpha,n}}^{\infty}\mathbb{P}(S_n \geq w)dw}{\mathbb{P}(S_n \geq q_{\alpha,n})} \\ &\sim q_{\alpha,n} + \frac{q_{\alpha,n}}{t-1} = \frac{tq_{\alpha,n}}{t-1} \sim \sigma^{-1}B_{nt}^{1/t}\frac{t(h_0/\alpha)^{1/t}}{t-1}.\end{aligned}$$

Without the exact moderate deviation principle in Corollary 1, the validity of this equivalence cannot be guaranteed. To the best of our knowledge, our example might be the only case where one can obtain explicit asymptotic expressions for VaR and ES for sums of dependent random variables.

2.5. Functionals of linear processes

In this subsection we use the result from (ii) of Corollary 5 to study the moderate deviation for nonlinear transformations of linear processes. Let K be a measurable transformation with $\mathbb{E}K(X_0) = 0$. Let

$$H_n = \sum_{i=1}^{n} K(X_i), \text{ where } X_i \text{ is defined by (2.16).}$$

Thus, if $K(X_0) = I(X_0 \leq \tau) - \mathbb{P}(X_0 \leq \tau)$, then H_n/n is the empirical process. If X_i is short memory, a_i absolutely summable, then we can apply the moderate deviation principle in Wu and Zhao (2008), but it does not apply to long-range dependent processes. The problem of moderate deviation under strong dependence has been rarely studied in the literature.

Here we establish such a principle in the context of nonlinear transforms of linear processes. Let $\mathcal{F}_n = (\cdots, \xi_{n-1}, \xi_n)$ be the shift process and define the projection operator $\mathcal{P}_i \cdot = \mathbb{E}(\cdot|\mathcal{F}_i) - \mathbb{E}(\cdot|\mathcal{F}_{i-1})$. Consider the truncated processes $X_{n,k} = \mathbb{E}(X_n|\mathcal{F}_k)$ and take $K_n(w) = \mathbb{E}[K(w+X_n-X_{n,0})]$ and $K_\infty(w) = \mathbb{E}[K(w+X_n)]$. We consider transformations K with $\kappa := K'_\infty(0) \neq 0$. Define

$$S_{n,1} = \sum_{i=1}^{n}[K(X_i) - \kappa X_i] = H_n - \kappa S_n, \text{ where } S_n = \sum_{i=1}^{n} X_i.$$

Then $H_n = \kappa S_n + S_{n,1}$. For a function g, let $g(w;\lambda) = \sup_{|y|\leq\lambda} |g(w+y)|$ be the local maximal function. Denote the collection of functions with second order partial derivatives by $\mathbb{C}^2(\mathbb{R})$.

Condition B. For $2 \leq q < p \leq 2q$, $\|\xi_0\|_p < \infty$. With $K_n \in C^2(\mathbb{R})$ for all large n, for some $\lambda > 0$,

$$\sum_{i=0}^{2} \|K_{n-1}^{(i)}(X_{n,0};\lambda)\|_q + \||\xi_1|^{p/q} K_{n-1}(X_{n,1})\|_q + \|\xi_1 K'_{n-1}(X_{n,1})\|_q = O(1).$$

A version of Condition B with $q = 2$ is used in Wu (2006). We establish a moderate deviation result. For $1/2 < r < 1$ and $1/2 \leq v < 1$ define

$$\chi(v,r) = v \max(r - \frac{r}{v}, \frac{1}{2} - r, r - 1),$$
$$\omega(r) = \operatorname*{argmin}_{1/2 \leq v < 1} \chi(v,r) \text{ and } \rho(r) = -\chi(\omega(r), r).$$

Theorem 4. *Assume that Condition B holds with $q = p\omega(r)$ and that the conditions of Corollary* 5 (ii) *are satisfied. Let c be such that $0 < c \leq p - 2$ and $c < 2p\rho(r)$. Then if $x \leq c \ln n$, we have*

$$\mathbb{P}(H_n \geq |\kappa|\sigma_n x) = (1 - \Phi(x))(1 + o(1)) \text{ as } n \to \infty. \tag{2.17}$$

Remark 3. As mentioned in the proof of Theorem 4 in the Supplementary Material of this paper (Peligrad et al. (2013)), (2.17) is still valid if the normalizing constant $|\kappa|\sigma_n$ is replaced by $\sqrt{var(H_n)}$.

Remark 4. Theorem 4 only asserts a moderate deviation with the Gaussian range. It is unclear whether the approximation (2.6) holds. We pose it as an open problem.

Remark 5. An explicit form for $\omega(r)$ can be obtained. If $r \geq 3/4$, then $\omega(r) = r$. If $r < 3/4$, then $\omega(r) = r/(2r - 1/2)$. If $2p\rho(r) \geq p - 2$, then the moderate deviation in (2.17) has the same range as for S_n. The latter happens, for example, if $r = 3/4$ and $2 < p < 16/5$, since in this case $2p\rho(3/4) \geq p - 2$.

Example 1. As an application to empirical processes, let $K(X) = I(X \leq \tau) - \mathbb{P}(X \leq \tau)$, where $\tau \in \mathbb{R}$ is fixed. Let $X_n = \xi_n + \sum_{i=1}^{\infty} a_i \xi_{n-i} =: \xi_n + Y_{n-1}$, where $\|\xi_0\|_p < \infty$, $p > 2$, and its density function f_ξ satisfies

$$\sup_u [f_\xi(u) + |f'_\xi(u)|] < \infty. \tag{2.18}$$

Then $K_1(w) = F_\xi(\tau - w) - F_X(\tau)$, where F_ξ is the distribution function of ξ_i. Under (2.18), we clearly have $\sup_w[|K'_1(w)| + |K''_1(w)|] < \infty$. Observe that we have the identity: for $n \geq 1$,

$$K_n(w) = \mathbb{E}K_1(w + a_1\xi_{n-1} + a_2\xi_{n-2} + \ldots + a_{n-1}\xi_1).$$

Hence $\sup_n \sup_w[|K'_n(w)| + |K''_n(w)|] < \infty$. So Condition B holds for any λ since $\xi_n \in L^p$, $p > 2$.

3. A Numerical Study

In this section we report on a numerical study of the accuracy of the large deviation (2.6), normal approximation (2.7), and the estimate (2.5). In particular, we studied the accuracy of the approximations in Corollary 4. In general it is time-consuming to calculate tail probabilities by Monte-Carlo simulation, especially if they are small. Here we approach the problem from a different angle.

Let $X_j = \sum_{i=1}^{\infty} a_i \xi_{j-i}$, where ξ_i, $i \in \mathbb{Z}$, have Student's t-distribution with degree of freedom $\nu = 3$, and $a_i = i^{-0.9}$. Let $S_n = \sum_{i=1}^{n} X_i$ with $n = 300$. The characteristic function of ξ_i is

$$\varphi(t) = \frac{(\sqrt{\nu}|t|)^{\nu/2} K_{\nu/2}(\sqrt{\nu}|t|)}{\Gamma(\nu/2) 2^{\nu/2-1}}, \tag{3.1}$$

where $K_{\nu/2}$ is the Bessel function (see Hurst (1995)). Then the characteristic function of S_n is

$$\varphi_{S_n}(t) = \prod_{j \in \mathbb{Z}} \varphi(b_{nj} t)$$

and, by the inversion formula,

$$\mathbb{P}(S_n \leq x) - \mathbb{P}(S_n \leq x') = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{\sqrt{-1}yx} - e^{\sqrt{-1}yx'}}{\sqrt{-1}y} \varphi_{S_n}(y) dy.$$

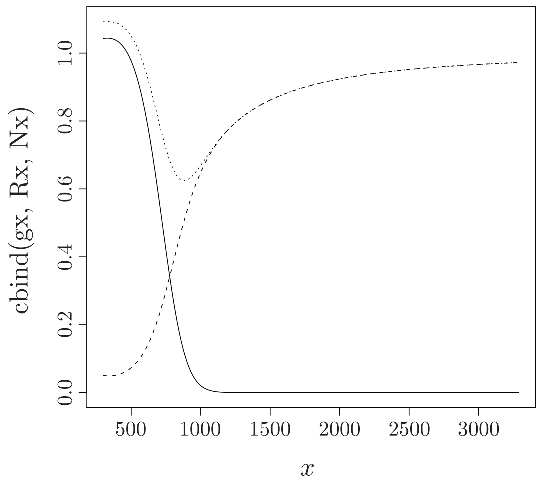

Figure 1. Tail approximation $R(x)$ (dashed curve), Gaussian approximation $g(x)$ (solid curve) and their sum (dotted curve) for long-memory processes with Student $t(3)$ innovations.

Take $x' = 0$. Since ξ_j is symmetric, $\mathbb{P}(S_n \le 0) = 1/2$. In our numerical study we use (3.1) to compute the probability $\mathbb{P}(S_n > x)$.

In Figure 1 we report the ratios $R(x) := \sum_i \mathbb{P}(b_{ni}\xi_0 \ge x)/\mathbb{P}(S_n > x)$ and $g(x) := (1 - \Phi(x/\sigma_n))/\mathbb{P}(S_n > x)$; see (2.6) with $c_{ni} = b_{ni}$. We can interpret $R(x)$ (resp. $g(x)$) as a tail (resp. Gaussian) approximation. As expected from Corollary 4, the Gaussian approximation is better if x is small, while the tail probability $R(x)$ approximation is better when x is large. In the intermediate region we approximate by their sum.

Acknowledgement

The authors would like to thank the referees for carefully reading the manuscript and for numerous suggestions that improved the presentation. This work was supported in part by the Taft Research Center at the University in Cincinnati. In addition, Magda Peligrad was supported in part by the NSA grant H98230-11-1-0135 and NSF DMS-1208237. Wei Biao Wu by NSF grants DMS-0906073 and DMS-1106790.

References

Acerbi, C. and Tasche, D. (2002). On the coherence of expected shortfall. *J. Banking and Finance* **26**, 1487-1503.

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

Bingham, N. H., Goldie, C. M. and Teugels, J. L. (1987). *Regular Variation*. Cambridge University Press, Cambridge.

Chen, X. (2001). Moderate deviations for Markovian occupation times. *Stochastic Process. Appl.* **94**, 51-70.

de la Peña, V. and Giné, E. (1999). *Decoupling. From Dependence to Independence.* Springer.

Doukhan, P., Oppenheim, G. and Taqqu, M. S. (editors) (2003). *Theory and Applications of Long-Range Dependence*, Birkhäuser, Boston.

Frolov, A. N. (2005). On probabilities of moderate deviations of sums for independent random variables. *J. Math. Sci.* **127**, 1787-1796.

Ghosh, M. (1974). Probabilities of moderate deviations under m-dependence. *Canad. J. Statist.* **2**, 157-168.

Grama, I. G. (1997). On moderate deviations for martingales. *Ann. Probab.* **25**, 152-183.

Grama, I. G. and Haeusler, E. (2006). An asymptotic expansion for probabilities of moderate deviations for multivariate martingales. *J. Theoret. Probab* **19**, 1-44.

Granger, C. W. and Joyeux, R. (1980). An introduction to long-memory time series models and fractional differencing. *J. Time Ser. Anal* **1**, 15-29.

Hall, P. (1992). Convergence rates in the central limit theorem for means of autoregressive and moving average sequences. *Stochastic Process. Appl.* **43**, 115-131.

Ho, H. C. and Hsing, T. (1997). Limit theorems for functionals of moving average. *Ann. Probab.* **25**, 1636-1669.

Holton, G. (2003). *Value-at-Risk: Theory and Practice.* Academic Press.

Hosking, J. R. M. (1981). Fractional differencing. *Biometrika* **68**, 165-176.

Hurst, S. (1995). The Characteristic Function of the Student-t Distribution. Financial Mathematics Research Report No. FMRR006-95, Statistics Research Report No. SRR044-95

Jorion, P. (2006). *Value at Risk: The New Benchmark for Managing Financial Risk.* 3rd edition. McGraw-Hill.

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

Mikosch, T. and Nagaev, A. V. (1998). Large deviations of heavy-tailed sums with applications in insurance. *Extremes* **1**, 81-110.

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

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

Nagaev, S. V. (1965). Some limit theorems for large deviations. *Teor. Veroyatn. Primen.* **10**, 231-254.

Nagaev, S. V. (1979). Large deviations of sums of independent random variables. *Ann. Probab.* **7**, 745-789.

Peligrad, M. and Sang, H. (2012). Asymptotic properties of self-normalized linear processes with long memory. *Econometric Theory* **28**, 548-569.

Peligrad, M., Sang, H., Zhong, Y. and Wu, W. B. (2013). Supplementary material for the paper "Exact Moderate and Large Deviations for Linear Processes".

Peligrad, M. and Utev, S. (1997). Central limit theorem for linear processes. *Ann. Probab.* **25** 443-456.

Robinson, P. M. (1997). Large-sample inference for non parametric regression with dependent errors. *Ann. Statist.* **25**, 2054-2083.

Robinson, P. M. (2003). *Time Series with Long Memory*, Oxford University Press.

Rozovski, L. V. (1993). Probabilities of large deviations on the whole axis. *Theory Probab. Appl.* **38**, 53-79.

Rubin, H. and Sethuraman, J. (1965). Probabilities of moderate deviations. *Sankhyā Ser. A* **27**, 325-346.

Slastnikov, A. D. (1978). Limit theorems for probabilities of moderate deviations *Teor. Veroyatn. Primen.* **24**, 340-357

Wu, W. B. (2006). Unit root testing for functionals of linear processes. *Econometric Theory* **22**, 1-14.

Wu, W. B. (2007). Strong invariance principles for dependent random variables. *Ann. Probab.* **35**, 2294-2320.

Wu, W. B. and Min, W. (2005). On Linear Processes with Dependent Innovations. *Stochastic Process. Appl.* **115**, 939-958.

Wu, W. B. and Zhao, Z. (2008). Moderate deviations for stationary processes. *Statist. Sinica* **18**, 769-782.

Department of Mathematical Sciences, University of Cincinnati, PO Box 210025, Cincinnati, OH 45221-0025, USA.

E-mail: peligrm@ucmail.uc.edu

Department of Mathematics, University of Mississippi, University, MS 38677-7071, USA.

E-mail: sang@olemiss.edu

Department of Statistics, University of Chicago, Chicago, IL 60637, USA.

E-mail: ydzhong@galton.uchicago.edu

Department of Statistics, University of Chicago, Chicago, IL 60637, USA.

E-mail: wbwu@galton.uchicago.edu

(Received May 2012; accepted April 2013)