*The Annals of Statistics*
2016, Vol. 44, No. 4, 1593–1617
DOI: 10.1214/15-AOS1429


# SELF-NORMALIZED CRAMÉR-TYPE MODERATE DEVIATIONS UNDER DEPENDENCE

BY XIAOHONG CHEN[1,*], QI-MAN SHAO[2,†], WEI BIAO WU[3,‡] AND LIHU XU[4,§]

*Yale University**, *Chinese University of Hong Kong*†, *University of Chicago*‡ *and University of Macau*§

We establish a Cramér-type moderate deviation result for self-normalized sums of weakly dependent random variables, where the moment requirement is much weaker than the non-self-normalized counterpart. The range of the moderate deviation is shown to depend on the moment condition and the degree of dependence of the underlying processes. We consider three types of self-normalization: the equal-block scheme, the big-block-small-block scheme and the interlacing scheme. Simulation study shows that the latter can have a better finite-sample performance. Our result is applied to multiple testing and construction of simultaneous confidence intervals for ultra-high dimensional time series mean vectors.

**1. Introduction.** Cramér-type moderate deviation principles for self-normalized sums have attracted considerable attention recently. In comparison with their non-self-normalized counterpart, the range of Gaussian approximation can be much wider under same polynomial moment conditions. This explains why self-normalized Cramér-type moderate deviation for independent data has been applied to multiple testing and simultaneous confidence sets construction problems with *ultra-high* dimensions (i.e., the high dimensional problems in which the dimension of the unknown parameters, $p$, could be exponentially larger than the sample size $n$). See, for example, Fan, Hall and Yao [24] and Liu and Shao [31]. Let $X_i$, $1 \le i \le n$, be independent mean zero random variables and $S_n = \sum_{i=1}^n X_i$. Define the self-normalized sum

$$T_n = \frac{S_n}{V_n} \qquad \text{where } V_n^2 = \sum_{i=1}^n X_i^2. \tag{1.1}$$

Received April 2015; revised December 2015.

[1]Supported by Cowles Foundation for Research in Economics.

[2]Supported in part by Hong Kong RGC GRF 403513 and 603710.

[3]Supported in part by NSF Grant DMS-11-06790.

[4]Supported by Macao S.A.R FDCT 049/2014/A1, MYRG2015-00021-FST.

*MSC2010 subject classifications.* 62E20, 60F10.

*Key words and phrases.* Cramér-type moderate deviation, absolutely regular, functional dependence measures, ultra-high dimensional time series.

Let

$$d_{n,\delta} = \left(\sum_{i=1}^{n} EX_i^2\right)^{1/2} \Big/ \left(\sum_{i=1}^{n} E|X_i|^{2+\delta}\right)^{1/(2+\delta)}.$$

The following Cramér-type moderate deviation result is a version of Theorems 2.1 and 2.3 of Jing, Shao and Wang [26]:

THEOREM 1.1. *Let $X_i$, $1 \le i \le n$, be independent with $EX_i = 0$, $E|X_i|^2 > 0$ and $E|X_i|^{2+\delta} < \infty$ for $0 < \delta \le 1$ and all $i$. Then there exists an absolute constant $A$ such that*

$$\left|\frac{P(T_n \ge x)}{1 - \Phi(x)} - 1\right| \le A \frac{(1+x)^{2+\delta}}{d_{n,\delta}^{2+\delta}} \tag{1.2}$$

*holds for all $0 \le x \le d_{n,\delta}$. If in addition*

$$\left(\sum_{i=1}^{n} EX_i^2\right) \max_{1 \le j \le n} \big(E(X_j^2)\big)^{\delta/2} \le A \sum_{i=1}^{n} E|X_i|^{2+\delta} \tag{1.3}$$

*then for all $0 \le x \le (d_{n,\delta}^{2+\delta}/A)^{1/\delta}$,*

$$\frac{P(T_n \ge x)}{1 - \Phi(x)} = \exp\left(O(1) \frac{(1+x)^{2+\delta}}{d_{n,\delta}^{2+\delta}}\right), \tag{1.4}$$

*where the constant $|O(1)| \le A$. If $E|X_i|^2 \ge c > 0$ and $E|X_i|^{2+\delta} \le c' < \infty$ for $0 < \delta \le 1$ and all $i$, then condition (1.3) is automatically satisfied, and equation (1.4) holds with $d_{n,\delta} \asymp n^{\delta/(4+2\delta)}$ for all $0 \le x \le n^{1/2}/A$.*

If in (1.1), (1.4) and (1.2), we use the non-self-normalized version with $T_n' = S_n/(E(V_n^2))^{1/2}$, then the range of $x$ such that (1.4) [or (1.2)] holds can be much narrower. The moderate deviation result of type (1.4) [or (1.2)] plays an important role in statistical inference of means since in practice one usually does not know the variance $\mathrm{var}(S_n) = E(V_n^2)$. Even if the latter is known, it is still advisable to use $T_n$, due to its wider range of Gaussian approximation.

There are many variations of asymptotic theories on self-normalized sums in the literature, and some allow for dependent data (see, e.g., [5, 27]). See [21, 39] for a comprehensive study and a recent review. Among the existing theories, perhaps the Cramér-type moderate deviations of self-normalized sums (Theorem 1.1) are the most useful ones for simultaneous confidence sets construction for ultra-high dimensional statistics. However, it has been an open question whether Theorem 1.1 could be generalized to dependent random variables. Such a generalization will be very useful for ultra-high dimensional statistical inference on dependent data with fat-tailed marginal distributions.

In this paper, we shall show that in general result (1.2) is *not valid* for the range of type $0 \le x \le n^{\rho}$ with any $\rho > 0$ if the dependence of the underlying process $\{X_t\}$ *decays algebraically*. In the latter case, one can only allow a much narrower range $0 \le x \le (\kappa \log n)^{1/2}$ with some constant $\kappa > 0$ (see Section 3 for details). On the positive side, by using block versions of $V_n$, we do establish Cramér-type moderate deviation results for self-normalized sums of weakly dependent processes with geometrically decaying dependence, under mild polynomial moment conditions. In particular, we introduce three types of self-normalized sums based on the big-block-small-block scheme, the equal-block and the interlacing scheme, respectively, and establish their associated Cramér-type moderate deviation theory. In the context of resampling theory for weakly dependent processes, block bootstrap procedures were proposed to adjust for dependence; see [7, 37] and [28], for example. The blocking technique has also been used in some recent work on time series models with slowly increasing dimension (see [10] and the reference therein). However, the accuracy of tail Gaussian approximation [of the type (1.2) or (1.4)] has not been studied before for dependent data. We show that, due to the dependence, the range of Gaussian approximation is narrower than their independent counterparts, but is still wider than their non-self-normalized ones under same polynomial moment conditions. We also present a time series two-sample moderate deviation extension. Our results are very useful for ultra-high dimensional statistical inference on dependent data with fat-tailed marginal distributions, such as multiple hypothesis testing of mean vectors of ultra-high dimensional time series models in one or two samples (see [14, 24, 31] and others for the results for independent data).

Although we focus on establishing Cramér-type moderate deviation results for weakly dependent data, our proof technique could be used to extend additional self-normalized limit theorems in [26, 31] and others surveyed in [39] from independent data to weakly dependent data with finite polynomial moments.

The rest of this paper is structured as follows. Section 2 introduces weakly dependent processes in terms of $\beta$-mixing coefficients and functional dependence measures. These notions of dependence are not nested. Together they cover a large class of widely used linear and nonlinear time series models. To ensure the validity of the Cramér-type moderate deviation with range of type $0 \le x \le n^{\rho}$ with some $\rho > 0$, we require that the dependence measures decay geometrically quickly. Section 3 provides a linear process example and shows that (1.2) is not valid for the range $0 \le x \le n^{\rho}$ with any $\rho > 0$. Section 4 provides three types of self-normalized sums for dependent data and derive their moderate deviation theorems. It also presents a two-sample moderate deviation extension. Section 5 gives an application to multiple test for ultra-high dimensional time series mean vectors, where the tests in [24] and [31] are generalized to the dependence setting. Section 6 presents a simulation study, which indicates that the self-normalized sums based on the interlacing scheme performs very well in finite samples. All the proofs are given in Section 7.

**2. Processes with geometrically decaying dependence.** There are many different notions of temporal dependence for general (nonlinear) time series; see [6] and [19] for recent reviews. In this paper, we focus on two measures of dependence that have been shown to cover a large class of time series models commonly used in statistics, econometrics, finance and engineering.

2.1. *$\beta$-mixing.* For a random process $\{X_t\}$ that may be nonstationary; let $\mathcal{I}_{-\infty}^t$ and $\mathcal{I}_{t+j}^{\infty}$ be $\sigma$-fields generated respectively by $(X_i, i \le t)$ and $(X_i, i \ge t + j)$. We say that $\{X_t\}$ is *$\beta$-mixing* (or absolutely regular) if

$$\beta(n) \equiv \sup_t E \sup\{|P(B|\mathcal{I}_{-\infty}^t) - P(B)| : B \in \mathcal{I}_{t+n}^{\infty}\} \to 0 \qquad \text{as } n \to \infty.$$

Assume that there exist positive numbers $a_1, a_2$ and $\tau$ such that

$$\beta(n) \le a_1 e^{-a_2 n^{\tau}}. \tag{2.1}$$

If $\{X_t\}$ is a strictly stationary Markov process on a set $\Omega \subseteq \mathcal{R}^d$, let $\|\phi\|_p^p = \int_\Omega |\phi(y)|^p \, dQ(y)$ and $\mathcal{T}_t\phi(x) = E[\phi(X_{t+1})|X_1 = x]$, then

$$\beta(t) = \int \sup_{0 \le \phi \le 1} \left| \mathcal{T}_t \phi(x) - \int \phi \, dQ \right| dQ.$$

The notion of $\beta$-mixing for a Markov process is closely related to the concept called $V$-ergodicity (Meyn and Tweedie [35]). Given a measurable function $V \ge 1$, the Markov process $\{X_t\}$ is $V$-uniformly ergodic if for all $t \ge 0$,

$$\sup_{0 \le \phi \le V} \left| \mathcal{T}_t \phi(x) - \int \phi \, dQ \right| \le c V(x) \exp(-\delta t)$$

for positive constants $c$ and $\delta$. If $E[V(X_t)] < \infty$, then the $V$-uniform ergodicity implies $\beta$-mixing with an exponential decay rate. This connection is valuable because one can show that a Markov process is $\beta$-mixing by applying the famous drift criterion (for ergodicity): There are constants $\lambda \in (0, 1)$ and $d \in (0, \infty)$, a norm-like function $\Gamma(\cdot) \ge 1$ and a small set $\mathbf{K}$ such that

$$E[\Gamma(X_t)|X_{t-1}] \le \lambda \Gamma(X_{t-1}) + d \times 1\{X_{t-1} \in \mathbf{K}\}.$$

In this case, $\{X_t\}$ is geometric ergodic and $\beta$-mixing (2.1) with $\tau = 1$.

Many nonlinear time series models are shown to be $\beta$-mixing via Tweedie's drift criterion approach. See, for example, Tong [41] for threshold models, Chen and Tsay [12, 13] for functional coefficient autoregressive models and nonlinear additive ARX models, Masry and Tjøstheim [33] for nonlinear ARCH, Carrasco and Chen [9] for GARCH, stochastic volatility and autoregressive conditional duration, Chen, Hansen and Carrasco [16] for diffusions, Chen, Wu and Yi [17] and Beare [3] for copula-based Markov models, Douc, Moulines, Olsson and van Handel [22] for a large class of generalized hidden Markov models. See Tong [41], Fan and Yao [25] and Chen [15] for additional example models and references.

2.2. *Functional dependence measures*. Another useful dependence measure for (nonlinear) time series is the so-called *functional dependence measure*; see, for example, Wu [44, 45]. Let $\varepsilon_t$, $t \in \mathbb{Z}$, be independent and identically distributed (i.i.d.) random variables. Suppose that $(X_t)$ is a causal process that can be represented as

$$X_t = G_t(\mathcal{F}_t), \tag{2.2}$$

where $G_t(\cdot)$ is a measurable function such that $X_t$ is well-defined, and $\mathcal{F}_t = \sigma(\dots, \varepsilon_{t-1}, \varepsilon_t)$. Let $(\varepsilon_i^*)_{i\in\mathbb{Z}}$ be an i.i.d. copy of $(\varepsilon_i)_{i\in\mathbb{Z}}$, and $\mathcal{F}_i^* = \sigma(\dots, \varepsilon_{i-1}^*, \varepsilon_i^*)$. Hence, $\varepsilon_i^*, \varepsilon_j, i, j \in \mathbb{Z}$, are i.i.d. Assume that, for all $t$, $X_t$ has finite $r$th moment, $r > 2$. Define the functional dependence measures as

$$\theta_r(m) = \sup_i \bigl\| X_i - G_i\bigl(\dots, \varepsilon_{i-m-2}, \varepsilon_{i-m-1}, \varepsilon_{i-m}^*, \varepsilon_{i-m+1}, \dots, \varepsilon_i\bigr)\bigr\|_r \tag{2.3}$$

and

$$\Delta_r(m) = \sup_i \bigl\| X_i - G_i\bigl(\mathcal{F}_{i-m}^*, \varepsilon_{i-m+1}, \dots, \varepsilon_i\bigr)\bigr\|_r. \tag{2.4}$$

Note that, if $(X_t)$ is a stationary linear process, then $\theta_r(m)$ is the absolute value of the impulse response function. Hence, we can interpret $\theta_r(m)$ as a nonlinear generalization of impulse response functions. We say that $(X_t)$ is *geometric moment contraction* (GMC; see Wu and Shao [46]) if there exist $\rho \in (0, 1)$, $a_1 > 0$, and $0 < \tau \le 1$ such that

$$\Delta_r(m) \le a_1 \rho^{m^\tau} = a_1 e^{-a_2 m^\tau} \qquad \text{with } a_2 = -\log\rho. \tag{2.5}$$

It is easily seen that (2.5) is equivalent to $\theta_r(m) = O(\rho_1^{m^\tau})$ for some $\rho_1 \in (0, 1)$. We emphasize that GMC does not imply geometric $\beta$-mixing. Andrews [1] gave a simple AR(1) example: $X_t = (X_{t-1} + \varepsilon_t)/2$, where $\varepsilon_t$ are i.i.d. Bernoulli(1/2). This process is not $\alpha$-mixing (and hence not $\beta$-mixing), however, it satisfies GMC (2.5) with $\rho = 1/2$ (or $a_2 = \log 2$).

*Examples of GMC.* Consider the infinite order autoregressive process

$$X_{k+1} = R(\varepsilon_{k+1}; X_k, X_{k-1}, \dots), \tag{2.6}$$

where $\varepsilon_k$ are i.i.d. and $R$ is a measurable function; see Wu [45] and Doukhan and Wintenberger [23]. Assume there exists a nonnegative sequence $(w_j)_{j\ge1}$ with $w_* = \sum_{j=1}^\infty w_j < 1$ such that

$$\bigl\| R(\varepsilon_0; x_{-1}, x_{-2}, \dots) - R\bigl(\varepsilon_0; x'_{-1}, x'_{-2}, \dots\bigr)\bigr\|_r \le \sum_{j=1}^\infty w_j \bigl|x_{-j} - x'_{-j}\bigr|.$$

By equations (27) and (28) in Wu [45], since $\sum_{j=1}^\infty w_j < 1$, (2.6) has a strictly stationary solution of the form

$$X_i = G(\varepsilon_i, \varepsilon_{i-1}, \dots),$$

whose functional dependence measures $(\theta_k)_{k\ge 0}$ satisfies

$$\theta_{k+1} \le \sum_{i=1}^{k+1} w_i \theta_{k+1-i}.$$

To obtain a bound for $\theta_k$, we define the sequence $a_k$ with $a_0 = \delta_0$, and

$$a_{k+1} = \sum_{i=1}^{k+1} w_i a_{k+1-i}. \tag{2.7}$$

If $w_j$ decays sub-geometrically in the sense that, for some $\rho \in (0, 1)$, $\tau \in (0, 1)$ and $C_0 > 0$, as $j \to \infty$,

$$w_j \sim C_0 \rho^{j^\tau}. \tag{2.8}$$

Then by elementary calculations, the recursion (2.7) has the asymptotic relation

$$a_k \sim \frac{a_0}{(1 - w_*)^2} C_0 \rho^{k^\tau}, \tag{2.9}$$

which entails GMC condition (2.5). If in (2.8) $\tau = 1$, then for some $\rho_0 \in (0, 1)$, we have

$$a_k \sim \frac{a_0}{(1 - w_*)^2} C_2 \rho_0^k. \tag{2.10}$$

**3. Cramér-type moderate deviations for linear processes with algebraically decaying coefficients.** In this section, we shall construct a linear process example and show that, if the dependence measure decays only algebraically slowly, then the Cramér-type moderate deviation is not valid at $x = (K \log n)^{1/2}$, where $K$ is a sufficiently large constant independent of $n$.

Let $\varepsilon_i$ be i.i.d. Student $t_\nu$ random variables with degrees of freedom $\nu > 4$; let $a_0 = 1$, $a_m \asymp m^{-\beta}$, $\beta > 1$ and define

$$X_j = \sum_{i=0}^{\infty} a_i \varepsilon_{j-i}. \tag{3.1}$$

For this process, the functional dependence measure $\theta_r(m) \asymp m^{-\beta}$ for $r \in (0, \nu)$. The process is also $\beta$-mixing if $\beta > 2 + \nu^{-1}$. To this end, we let $2 < r < \nu$ be such that $2 + r^{-1} < \beta$. By Theorem 2.1 in [36], its $\beta$-mixing coefficient $\beta(n)$ is of order $O(n(n^{1-\beta})^{r/(1+r)}) = O(n^{1+r(1-\beta)/(1+r)})$ since $\sum_{i=m}^{\infty} a_i = O(m^{1-\beta})$. By [18], we have $\operatorname{cov}(X_0, X_n) = O(\beta(n)^{1/p})$, where $p = r/(r-2)$. Note that $\operatorname{cov}(X_0, X_n) \asymp n^{-\beta}$. Hence, we have the lower bound for $\beta(n)$: $n^{-\beta r/(r-2)} = O(\beta(n))$.

For $S_n = \sum_{i=1}^{n} X_i$, we consider the self-normalized sum of the form

$$T_n = \frac{S_n}{\hat\sigma_n} \qquad \text{where } \hat\sigma_n^2 = \sum_{j,k \le n} X_j X_k w_{j,k}. \tag{3.2}$$

Here, $w_{j,k}$ are weights such that $|w_{j,k}| \le 1$ and $w_{j,k} = 0$ if $|j-k| \ge B$, where $B = B_n$ is the window size parameter. For example, we can choose the triangle kernel $w_{j,k} = \max(1 - |j-k|/B, 0)$, the rectangle kernel $w_{j,k} = 1_{|j-k|<B}$, or, with $n = Bk$ and $B$ being the block size,

$$\hat{\sigma}_n^2 = \sum_{j=1}^{k} \left( \sum_{i=1+(j-1)B}^{jB} X_i \right)^2 = \sum_{f,g \le n} X_f X_g w_{f,g},$$

where $w_{f,g} = 1$ if there exists $j$ such that $(j-1)B < f, g \le jB$; see also the block normalized sum (4.3). We can view $n^{-1}\hat{\sigma}_n^2$ as a lag-window estimate of the long-run variance $\sigma_\infty^2 = \sum_{j \in \mathbf{Z}} \gamma_j$, where $\gamma_j = E(X_0 X_j)$ is the covariance function. In comparison with (1.1), the cross-product terms $X_j X_k$ with $j \ne k$ in the expression of $\hat{\sigma}_n^2$ are introduced to adjust for the dependence. Assuming $B_n \to \infty$ and $B_n/n \to 0$, under suitable conditions of the weights $(w_{j,k})$, $n^{-1}\hat{\sigma}_n^2$ is a consistent estimate of $\sigma_\infty^2$.

THEOREM 3.1. *Assume* (3.1) *and that the lag* $B \asymp n^\theta$, $0 < \theta < 1$. *Let* $y_n = n^\alpha$, $0 < \alpha < (1-\theta)/2$. *Then there exists a constant* $c_1 > 0$, *independent of* $n$ *and the weights* $(w_{j,k})$, *such that for all sufficiently large* $n$, *we have*

$$P(S_n/\hat{\sigma}_n > y_n) \ge c_1 n^{-\beta\nu}. \tag{3.3}$$

PROOF. We shall compare the coefficients of $\varepsilon_{-n}$ in $S_n$ and $\hat{\sigma}_n^2$. Let $X_i^\circ = X_i - a_{n+i}\varepsilon_{-n}$ and $S_n^\circ = \sum_{i=1}^n X_i^\circ$. Then $S_n = S_n^\circ + A_n \varepsilon_{-n}$, where $A_n := \sum_{i=1}^n a_{n+i} \asymp n^{1-\beta}$, by the condition on $a_n$. We also write

$$\begin{aligned} \hat{\sigma}_n^2 &= \sum_{j,k \le n} X_j^\circ X_k^\circ w_{j,k} + \sum_{j,k \le n} (X_j^\circ a_{n+k} + a_{n+j} X_k^\circ) w_{j,k} \varepsilon_{-n} \\ &\quad + \sum_{j,k \le n} a_{n+j} a_{n+k} w_{j,k} \varepsilon_{-n}^2 \\ &=: Q_n + L_n \varepsilon_{-n} + f_n \varepsilon_{-n}^2, \end{aligned} \tag{3.4}$$

where $f_n = \sum_{j,k \le n} a_{n+j} a_{n+k} w_{j,k}$ satisfying $|f_n| \le C_1 n^{-2\beta} n B \le C_2 n^{1-2\beta+\theta}$. In the proof of this theorem, the constants $C_1, C_2, \ldots,$ are all independent of $n$ and the weights $(w_{j,k})$. Then

$$\begin{aligned} S_n^2 - y_n^2 \hat{\sigma}_n^2 &= (A_n^2 - y_n^2 f_n)\varepsilon_{-n}^2 + \bigl(2A_n S_n^\circ - y_n^2 L_n\bigr)\varepsilon_{-n} \\ &\quad + (S_n^\circ)^2 - y_n^2 Q_n. \end{aligned} \tag{3.5}$$

Let $g_n = A_n^2 - y_n^2 f_n$, $M_n = 2A_n S_n^\circ - y_n^2 L_n$ and $R_n = (S_n^\circ)^2 - y_n^2 Q_n$. By the moment inequality in Theorem 2 in [44], $E(S_n^\circ)^4 \le C_3 n^2$, $E(Q_n^2) \le C_4 n^2$, $\|L_n\| \le C_5 a_n B \sqrt{n}$. Hence, $\|M_n\| \le C_6 n^{3/2-\beta}$ and $\|R_n\| \le C_7 y_n^2 n$. Note that $y_n^2 f_n \le n^{2\alpha} C_2 n^{1-2\beta+\theta}$, we have $g_n \asymp n^{2-2\beta}$. Let $h_n = n^\beta$. Then

$$P\bigl[h_n \ge |M_n|/g_n + \bigl(|R_n|/g_n\bigr)^{1/2}\bigr] \to 1 \qquad \text{as } n \to \infty \tag{3.6}$$

since $n^{3/2-\beta}/g_n = o(h_n)$ and $y_n^2 n/g_n = o(h_n^2)$. By the independence of $\varepsilon_{-n}$ and $(M_n, R_n)$, we obtain

$$
\begin{aligned}
P(S_n^2 \geq y_n^2 \hat{\sigma}_n^2) &\geq P\bigl[\varepsilon_{-n} \geq |M_n|/g_n + \bigl(|R_n|/g_n\bigr)^{1/2}\bigr] \\
&\geq P\bigl[\varepsilon_{-n} \geq h_n, h_n \geq |M_n|/g_n + \bigl(|R_n|/g_n\bigr)^{1/2}\bigr] \\
&= P(\varepsilon_{-n} \geq h_n) P\bigl[h_n \geq |M_n|/g_n + \bigl(|R_n|/g_n\bigr)^{1/2}\bigr].
\end{aligned} \tag{3.7}
$$

Since $\varepsilon_i$ is $t_\nu$, $P(\varepsilon_{-n} \geq h_n) \sim c_\nu h_n^{-\nu}$ for some constant $c_\nu > 0$, (3.3) follows in view of the symmetry $P(S_n/\hat{\sigma}_n > y_n) = P(S_n/\hat{\sigma}_n < -y_n)$. $\square$

Note that, as $x \to \infty$, $1 - \Phi(x) \sim x^{-1}(2\pi)^{-1/2} \exp(-x^2/2)$. Theorem 3.1 implies that, if the constant $K > 2\beta\nu$, we have

$$
\frac{P(S_n/\hat{\sigma}_n > (K \log n)^{1/2})}{1 - \Phi((K \log n)^{1/2})} \to \infty. \tag{3.8}
$$

It suggests that, with only finite polynomial moment condition and when the dependence decays algebraically, then the range for the Cramér-type moderate deviation can be very narrow: it cannot go beyond $(K \log n)^{1/2}$ with some constant $K$. To ensure a wider range of type $0 \leq x \leq n^\rho$ with some $\rho > 0$, we need to impose a stronger condition on the weakness of the dependence. In the next section, we consider moderate deviations for processes that have geometrically decaying dependence.

## 4. Main results.

Let $\{X_i, i \geq 1\}$ be a sequence of random variables satisfying

$$
E(X_i) = \mu = 0, \qquad E|X_i|^r \leq c_1^r \qquad \text{for all } i \tag{4.1}
$$

for $r > 2$ and $c_1 < \infty$. Write $S_{k,m} = \sum_{i=k+1}^{k+m} X_i$ and $S_n = S_{0,n}$. Assume that there exists a positive number $c_2$ such that

$$
E(S_{k,m}^2) \geq c_2^2 m \qquad \text{for all } k \geq 0, m \geq 1. \tag{4.2}
$$

We shall assume that $\{X_i\}$ is weakly dependent which can be either geometric $\beta$-mixing or geometric moment contracting (GMC); see Sections 4.4 and 4.5, respectively. The condition of geometric decaying of dependence in general cannot be relaxed; see Section 3.

For independent random variables, (1.1) is the natural form for normalized sum. The situation is quite different when dependence is present. There are a few ways to account for dependence. The block normalized sum, big-block-small-block normalized sum, the interlacing normalized sum are introduced in Sections 4.1, 4.2 and 4.3, respectively. For all schemes, we can establish their moderate deviations for either geometric $\beta$-mixing or GMC processes. Blocking technique is a common way to weaken dependence; see, for example, Lin and Foster [30].

4.1. *Block normalized sum.* As a natural way to account for dependence, we can modify $V_n$ in (1.1) by using block sums. Assume at the outset that $\mu = 0$. Let $m = \lfloor n^\alpha \rfloor$, $0 < \alpha < 1$, $k = \lfloor n/m \rfloor$,

$$H_j^\circ = \{i : m(j-1) + 1 \le i \le m(j-1) + m\}, \qquad 1 \le j \le k, \tag{4.3}$$

and the block sums $Y_j^\circ = \sum_{l \in H_j^\circ} X_l$. Define

$$T_n^\circ = \frac{\sum_{j=1}^k Y_j^\circ}{V_k^\circ} \qquad \text{where } (V_k^\circ)^2 = \sum_{j=1}^k (Y_j^\circ)^2. \tag{4.4}$$

Note that, if $E(X_i) = 0$ and $(X_i)$ is a stationary process, then $(V^\circ)^2/(mk)$ is the classical nonoverlapped batched mean estimate of the long-run variance $\sigma_\infty^2 = \sum_{j \in \mathbf{Z}} \operatorname{cov}(X_0, X_j)$. See Politis, Romano and Wolf [37] and Bühlmann [7]. In general, when $\mu$ is unknown, we let $\bar{Y}^\circ = k^{-1} \sum_{j=1}^k Y_j^\circ$,

$$T_n^\dagger = \frac{\sum_{j=1}^k (Y_j^\circ - m\mu)}{V_k^\dagger} \qquad \text{where } (V_k^\dagger)^2 = \sum_{j=1}^k (Y_j^\circ - \bar{Y}^\circ)^2. \tag{4.5}$$

4.2. *Big-block-small-block normalized sum.* In (4.4), we use consecutive blocks with equal size. As a slightly modified version, we can adopt a big-block-small-block scheme and only use big blocks. Partition $\{X_i, 1 \le i \le n\}$ into consecutive big blocks and small blocks. Let $m_1 = \lfloor n^{\alpha_1} \rfloor$, $m_2 = \lfloor n^{\alpha_2} \rfloor$, where $1 > \alpha_1 > \alpha_2 > 0$, $m_* = m_1 + m_2$, $k = \lfloor n/m_* \rfloor$ and, for $1 \le j \le k$, put

$$\begin{aligned} H_{j,1} &= \{i : (j-1)m_* + 1 \le i \le (j-1)m_* + m_1\}, \\ H_{j,2} &= \{i : (j-1)m_* + m_1 + 1 \le i \le j m_*\}, \end{aligned}$$

where $H_{j,1}$ (resp., $H_{j,2}$) are large (resp., small) blocks, and the corresponding block sums

$$\begin{cases} Y_{j,1} = \sum_{i \in H_{j,1}} X_i, & Y_{j,2} = \sum_{i \in H_{j,2}} X_i, \\ S_{n,1} = \sum_{j=1}^k Y_{j,1}, & S_{n,2} = \sum_{j=1}^k Y_{j,2}, \\ V_{n,1}^2 = \sum_{j=1}^k Y_{j,1}^2, & V_{n,2}^2 = \sum_{j=1}^k Y_{j,2}^2. \end{cases} \tag{4.6}$$

Consider the self-normalized big-block sum

$$W_n = \frac{S_{n,1}}{V_{n,1}}. \tag{4.7}$$

Under geometric $\beta$-mixing condition (2.1) or GMC condition (2.5), one can easily prove that $W_n \xrightarrow{d.} N(0,1)$.

If the mean is common but unknown, that is, $E(X_i) = \mu$ for all $i$ with $\mu$ unknown, similarly as (4.5), we consider the Student $t$-statistic

$$W_n^* = \frac{S_{n,1}^*}{V_{n,1}^*} = \frac{\sum_{j=1}^k (Y_{j,1} - m_1\mu)}{\sqrt{\sum_{j=1}^k (Y_{j,1} - \bar{Y}_1)^2}}, \tag{4.8}$$

where $\bar{Y}_1 = k^{-1} \sum_{j=1}^k Y_{j,1}$.

4.3. *Interlacing normalized sum.* A particularly interesting case for $W_n$ in (4.7) is $\alpha_1 = \alpha_2 = \alpha \in (0,1)$. Let $m = \lfloor n^\alpha \rfloor$, $k := \lfloor n/(2m) \rfloor$ and

$$H_j = \{i : 2m(j-1) + 1 \le i \le 2m(j-1) + m\}, \qquad 1 \le j \le k. \tag{4.9}$$

Note that $H_j = H_{j,1}$. Let $Y_j = \sum_{l \in H_j} X_l$, $V^2 = \sum_{j=1}^k Y_j^2$ and

$$I_n = \frac{\sum_{j=1}^k Y_j}{V} = \frac{\sum_{j=1}^k Y_j}{\sqrt{\sum_{j=1}^k Y_j^2}}, \tag{4.10}$$

which is $W_n$ in (4.7). Denote by $I_n^*$ the interlaced version of $W_n^*$ in (4.8):

$$I_n^* = \frac{\sum_{j=1}^k (Y_j - m\mu)}{\sqrt{\sum_{j=1}^k (Y_j - \bar{Y})^2}} \qquad \text{where } \bar{Y} = k^{-1} \sum_{j=1}^k Y_j. \tag{4.11}$$

4.4. *Moderate deviation under geometric $\beta$-mixing.*

THEOREM 4.1. *Assume conditions* (4.1), (4.2) *and* (2.1). *Let* $0 < \alpha_2 \le \alpha_1 < 1$ *and* $0 < \delta \le 1$, $\delta < r - 2$. *Then there exist finite constants* $c_0$, $A$ *depending only on* $c_1/c_2, a_1, a_2, r$ *and* $\tau$ *such that*

$$\frac{P(W_n \ge x)}{1 - \Phi(x)} = \exp\bigl(O(1)(1+x)^{2+\delta} n^{-(1-\alpha_1)\delta/2}\bigr) \tag{4.12}$$

*uniformly in* $0 \le x \le c_0 \min(n^{(1-\alpha_1)/2}, n^{\alpha_2\tau/2})$, *and* $|O(1)| \le A$. *In particular, we have*

$$\frac{P(W_n \ge x)}{1 - \Phi(x)} = 1 + O(1)(1+x)^{2+\delta} n^{-(1-\alpha_1)\delta/2}, \tag{4.13}$$

*for all* $0 \le x \le c_0 \min(n^{(1-\alpha_1)\delta/(4+2\delta)}, n^{\alpha_2\tau/2})$ *and* $|O(1)| \le A$.

If $\tau = 1 = \delta$ and we choose $\alpha_1 = \alpha_2 = 1/2$, then (4.12) yields

$$\ln P(W_n \ge x_n) \sim -x_n^2/2 \tag{4.14}$$

as $x_n \to \infty$ and $x_n = o(n^{1/4})$. Note that when $X_i$ are independent with bounded 3rd moments, Theorem 1.1 gives a wider range of $x_n = o(n^{1/2})$.

If $\tau = 1 = \delta$ and we choose $\alpha_1 = \alpha_2 = 1/4$, then (4.13) implies

$$\frac{P(W_n \geq x)}{1 - \Phi(x)} = 1 + O(1)(1+x)^3 n^{-3/8} \to 1 \tag{4.15}$$

uniformly in $0 \leq x \leq o(n^{1/8})$. Again when $X_i$ are independent, Theorem 1.1 gives a wider range of $0 \leq x \leq o(n^{1/6})$.

In practice, it is more common to use the Student $t$-statistic $W_n^*$ rather than the self-normalized $W_n$. We have the same result for $W_n^*$ in (4.8). It readily follows from Theorem 4.1.

COROLLARY 4.1. *Let conditions* (4.1) (*with unknown mean* $\mu$), (4.2) *and* (2.1) *hold. Then Results* (4.12) *and* (4.13) *also hold for* $W_n^*$.

For the block normalized sums $T_n^\circ$ and $T_n^\dagger$ of (4.4), we have the following Theorem 4.2. Corollary 4.2 follows from Theorem 4.2.

THEOREM 4.2. *Assume that* (4.1), (4.2) *and* (2.1) *hold. Let* $\alpha \in (0, 1)$ *and* $0 < \delta \leq 1$, $\delta < r - 2$. *Then there exist some finite positive numbers* $c_0$ *and* $A$ *depending on* $\alpha, \tau, a_1, a_2, r, c_1$ *and* $c_2$ *such that*

$$\frac{P(T_n^\circ \geq x)}{1 - \Phi(x)} = 1 + O(1)\big[(1+x)^{4+\delta} n^{-\delta(1-\alpha)/2}\big]^{1/4} \tag{4.16}$$

*uniformly for* $0 \leq x \leq c_0 \min((\log n)^{-4/(4+\delta)} n^{(1-\alpha)\delta/(2(4+\delta))}, n^{\alpha\tau/2})$ *with* $|O(1)| \leq A$.

The proof of Theorem 4.2 is much more complicated and we give details in the supplemental article [11].

COROLLARY 4.2. *Let* (4.1) (*with unknown mean* $\mu$), (4.2) *and* (2.1) *hold. Let* $\alpha \in (0, 1)$ *and* $0 < \delta \leq 1$, $\delta < r - 2$. *Then there exist some finite positive numbers* $c_0$ *and* $A$ *depending on* $\alpha, \tau, a_1, a_2, r, c_1$ *and* $c_2$ *such that Result* (4.16) *also holds for* $T_n^\dagger$.

4.5. *Moderate deviation under geometric moment contraction.* In this section, we consider time series models that satisfy the GMC condition (2.5).

THEOREM 4.3. (i) *Assume* (4.1), (4.2) *and* (2.5). *Let* $0 < \alpha < 1$ *and* $2 < r \leq 3$. *Then there exist constants* $c_0, A > 0$, *depending only on* $c_1/c_2, a_1, a_2, \alpha, r$ *and* $\tau$ *such that* $I_n$ *in* (4.10) *satisfies the moderate deviation*

$$\frac{P(I_n \geq x)}{1 - \Phi(x)} = \exp\big(O(1)(1+x)^r n^{(1-r/2)(1-\alpha)}\big) \tag{4.17}$$

*for all* $0 \le x \le c_0 \min(n^{(1-\alpha)/2}, n^{\alpha\tau/2})$ *and* $|O(1)| \le A$. *In particular, we have*

$$\frac{P(I_n \ge x)}{1 - \Phi(x)} = 1 + O(1)(1 + x)^r n^{(1-r/2)(1-\alpha)} \tag{4.18}$$

*for all* $0 \le x \le c_0 \min(n^{(1-\alpha)(r-2)/2r}, n^{\alpha\tau/2})$ *and* $|O(1)| \le A$.

(ii) *If condition* (4.1) *holds with unknown* $\mu$ *in part* (1), *then results* (4.17) *and* (4.18) *hold with* $I_n^*$ *in* (4.11).

If we increase $\tau$ or $r$, then the range for $x$ can be wider. Let $\tau = 1$, $r = 3$ and $\alpha = 1/4$. Then the moderate deviation (4.18) implies (4.15) uniformly in the range $0 \le x \le o(n^{1/8})$. In comparison, if $\alpha_1 > \alpha_2$, then the big-block-small-block self-normalized sum (4.13) has a moderate deviation with a slightly narrower range since $\delta < 1$. Similarly, the range for the block normalized sum $T_n^\circ$ is also slightly narrower.

Similarly, as Theorem 4.2, using the GMC condition (2.5), we have the following.

COROLLARY 4.3. *Assume that* (4.1) (4.2) *and* (2.5) *hold. Let* $\alpha \in (0, 1)$, $2 < r \le 3$ *and* $\delta = r - 2$. *Then there exist some finite positive numbers* $c_0$ *and* $A$ *depending on* $\alpha, \tau, a_1, a_2, r, c_1$ *and* $c_2$ *such that* (4.16) *holds.*

4.6. *Moderate deviation for two-sample statistic*. The results in Sections 4.4 and 4.5 can be easily extended to the two-sample case. Let $\{X_i^{(1)}, i \ge 1\}$ and $\{X_i^{(2)}, i \ge 1\}$ be two independent sequences of random variables, both of them satisfying (2.1) or (2.5). Assume that

$$E(X_i^{(l)}) = 0, \qquad E|X_i^{(l)}|^r \le c_1^r, \qquad l = 1, 2 \text{ for all } i \tag{4.19}$$

for $r > 2$ and $c_1 < \infty$. Set $S_{k,m}^{(l)} = \sum_{i=k+1}^{k+m} X_i^{(l)}$ and $S_n^{(l)} = S_{0,n}^{(l)}$. Assume that there exist positive numbers $c_2, a_1, a_2$ and $\tau$ such that

$$E([S_{k,m}^{(l)}]^2) \ge c_2^2 m \qquad \text{for all } k \ge 0, m \ge 1, l = 1, 2; \tag{4.20}$$

and (2.1) holds for both processes.

Assume $n_1 \asymp n_2 \asymp n$. For $l = 1, 2$, we partition $\{X_i^{(l)}, 1 \le i \le n_l\}$ into big blocks and small blocks. Let $m_1 = \lfloor (n_1 + n_2)^{\alpha_1} \rfloor$, $m_2 = \lfloor (n_1 + n_2)^{\alpha_2} \rfloor$, where $1 > \alpha_1 \ge \alpha_2 > 0$, $m_* = m_1 + m_2$, $k_l = \lfloor n_l / m_* \rfloor$ for $l = 1, 2$, and for $1 \le j \le \max(k_1, k_2)$, put

$$H_{l;j,1} = \{i : (j-1)m_* + 1 \le i \le \min(n_l, (j-1)m_* + m_1)\},$$
$$H_{l;j,2} = \{i : (j-1)m_* + m_1 + 1 \le i \le \min(n_l, jm_*)\}.$$

For $l = 1, 2$,

$$Y_{j,1}^{(l)} = \sum_{i \in H_{l;j,1}} X_i^{(l)}, \qquad Y_{j,2}^{(l)} = \sum_{i \in H_{l;j,2}} X_i^{(l)},$$

$$S_{n,1}^{(l)} = \sum_{j=1}^{k_l} Y_{j,1}^{(l)}, \qquad S_{n,2}^{(l)} = \sum_{j=1}^{k_l} Y_{j,2}^{(l)},$$

$$V_{n,1}^{(l)2} = \sum_{j=1}^{k_l} [Y_{j,1}^{(l)}]^2, \qquad V_{n,2}^{(l)2} = \sum_{j=1}^{k_l} [Y_{j,2}^{(l)}]^2.$$

Consider

$$\hat{W}_n = \frac{k_1^{-1} S_{n,1}^{(1)} - k_2^{-1} S_{n,1}^{(2)}}{(k_1^{-2} V_{n,1}^{(1)2} + k_2^{-2} V_{n,1}^{(2)2})^{1/2}}.$$

THEOREM 4.4. *Assume* (4.19) *and* (4.20). *Then* (i) *under* (2.1), *results* (4.12) *and* (4.13) *remain valid for* $\hat{W}_n$; *and* (ii) *under* (2.5), *the moderate deviations* (4.17) *and* (4.18) *hold for* $\hat{W}_n$ *with* $\alpha_1 = \alpha_2$.

As Theorem 4.2, the block normalized version of Theorem 4.4 can be similarly formulated. Details are omitted.

4.7. *Small sample corrections.* In our interlacing normalized sum $I_n$ in (4.10), if $Y_j$ are i.i.d. standard normal, then $I_n \sim t_k$, a $t$-distribution with degrees of freedom $k$. Note that $k \sim n^{1-\alpha}/2$, which is much smaller than $n$. In actual application of Theorem 4.3, instead of the normal distribution function $\Phi$, we suggest using the $t_k$ distribution. Similar claims can be made for $I_n^*$, $W_n$, $W_n^*$, $T_n^\circ$ and $T_n^\dagger$ as well. See [20, 43] and others for similar suggestions.

**5. Applications.** As the result of Jing, Shao and Wang [26] has been widely applied in statistics and econometrics for independent data, our results are very useful in similar applications with spatially dependent data and time series observations. As an illustrative yet important application, in this section we apply our theory to a time series extension of multiple tests of Fan, Hall and Yao [24] and Liu and Shao [31].

Consider the problem of constructing simultaneous confidence intervals for the mean vector $\boldsymbol{\mu} = (\mu_1, \dots, \mu_p)'$ of the stationary $p$-dimensional process

$$\mathbf{Z}_i = (Z_{i1}, \dots, Z_{ip})' = \mathbf{G}(\mathcal{F}_i), \tag{5.1}$$

where $\mathcal{F}_i = (\dots, \varepsilon_{i-1}, \varepsilon_i)$, $\varepsilon_i$ are i.i.d. and $\mathbf{G}(\cdot) = (G^{(1)}(\cdot), \dots, G^{(p)}(\cdot))'$ is a function such that $\mathbf{Z}_i$ is well-defined. Assume that the long-run covariance matrix

$$\Sigma_\infty = \sum_{i=-\infty}^{\infty} \operatorname{cov}(\mathbf{Z}_0, \mathbf{Z}_i) = (\omega_{jl})_{j,l \le p} \tag{5.2}$$

exists and satisfies the following condition.

ASSUMPTION (R0). There exists a constant $\zeta > 0$ such that the long-run variance $\omega_{jj} \geq \zeta$ holds for all $j \leq p$.

We shall impose the following uniform geometric moment contraction condition, which is a uniform version of (2.5) on the component processes $Z_{il} = G^{(l)}(\mathcal{F}_i)$, $1 \leq l \leq p$. Here, for the sake of conciseness we only deal with the case $\tau = 1$. Similar results can be derived when $\tau < 1$, or when one uses the uniform version of the geometric $\beta$-mixing (2.1).

ASSUMPTION (G). Let $\Delta_r^{(l)}(m) = \|G^{(l)}(\mathcal{F}_i) - G^{(l)}(\mathcal{F}_{i-m}^*, \varepsilon_{i-m+1}, \ldots, \varepsilon_i)\|_r$ [cf. (2.4)] be the functional dependence measure for the component process $Z_{il} = G^{(l)}(\mathcal{F}_i)$, $1 \leq l \leq p$. Assume that there exist $a_1, a_2 > 0$ such that $\max_{l \leq p} \Delta_r^{(l)}(m) \leq a_1 e^{-a_2 m}$ holds for all $m \geq 0$.

5.1. *Construction of conservative simultaneous confidence intervals.* Here, for illustration purposes, we use the interlacing normalized sum. Other versions are similar. Given the data $\mathbf{Z}_1, \ldots, \mathbf{Z}_n$, let

$$T_{nl} = \frac{\sum_{j=1}^k (Y_{jl} - m\mu_l)}{\sqrt{\sum_{j=1}^k (Y_{jl} - \bar{Y}_l)^2}}, \qquad l = 1, \ldots, p, \tag{5.3}$$

where $m \asymp n^{1/4}$, $k = \lfloor n/(2m) \rfloor$, $Y_{jl} = \sum_{g \in H_j} Z_{gl}$ and $H_j$ is given in (4.9), and $\bar{Y}_l = k^{-1} \sum_{j=1}^k Y_{jl}$.

COROLLARY 5.1. *Let Assumption (R0) be satisfied. Assume that the dimension $p$ satisfies*

$$\log p = o(n^{1/4}). \tag{5.4}$$

*Let Assumption (G) be satisfied with $r = 3$. Let $\alpha \in (0, 1)$. Then*

$$\frac{\bar{Y}_l}{m} \pm \frac{\Phi^{-1}(1 - \alpha/(2p))}{km} \sqrt{\sum_{j=1}^k (Y_{jl} - \bar{Y}_l)^2}, \tag{5.5}$$

*are $1 - \alpha$ (conservative) simultaneous confidence intervals for $(\mu_l)_{l=1}^p$.*

PROOF. Note that the upper $(\alpha/(2p))$th quantile of a standard normal distribution $\Phi^{-1}(1 - \alpha/(2p)) = O((\log p)^{1/2})$, which by (5.4) is of order $o(n^{1/8})$. Assumption (R0) implies that (4.2) holds for all component processes $(Z_{il})_i$. Then the corollary follows from applying Theorem 4.3 to $T_{nl}$ via the Bonferroni procedure. $\square$

As discussed in Section 4.7, the finite-sample performance can be improved if in (5.5) we use quantiles of $t$ distributions. Namely, the quantity $\Phi^{-1}(1-\alpha/(2p))$ is replaced by $u = u(\alpha, p, k)$ for which $P(|t_{k-1}| \geq u) = \alpha/p$, where $t_{k-1}$ follows $t$-distribution with degrees of freedom $k-1$.

5.2. *Simultaneous confidence intervals with asymptotically correct coverage probabilities*. Due to the Bonferroni correction, the confidence intervals (5.5) can be overly wider. To construct simultaneous confidence intervals with asymptotically correct coverage probabilities, we shall make an additional assumption on the long-run correlation matrix

$$R = \left(\frac{\omega_{jl}}{\omega_{jj}^{1/2}\omega_{ll}^{1/2}}\right)_{j,l \leq p}, \tag{5.6}$$

where $\Sigma_\infty$ is the long-run covariance matrix of the process $(\mathbf{Z}_i)$; cf. (5.2).

ASSUMPTION (R). (i) There exists a constant $\zeta > 0$ such that the long-run variance $\omega_{jj} \geq \zeta$ holds for all $j \leq p$; (ii) for some $\gamma > 0$,

$$\max_{j \leq p} \#\{l \leq p : |R_{lj}| \geq (\log p)^{-1-\gamma}\} = O(p^\rho) \tag{5.7}$$

holds for all $\rho > 0$.

In Assumption (R), requirement (ii) indicates that, the long-run correlation between the component processes $(Z_{il})_{i \in \mathbb{Z}}$ and $(Z_{ij})_{i \in \mathbb{Z}}$ are sufficiently weak. This condition is mild enough. Liu and Shao [31] proposed a similar condition for i.i.d. random vectors.

THEOREM 5.1. *Let Assumptions (R) and (G) be satisfied with some $r > 3$. Assume that for some $\chi > 0$,*

$$(\log p)^{1+\chi} = o(n^{1/4}). \tag{5.8}$$

*Let $\mathcal{G}$ follow the Gumbel distribution $P(\mathcal{G} \leq y) = \exp(-e^{-y/2}/\pi^{1/2})$. Then*

$$\max_{l \leq p} T_{nl}^2 - 2\log p + \log\log p \Rightarrow \mathcal{G}. \tag{5.9}$$

The convergence in (5.9) can be quite slow. In practice with relatively small sample sizes, we can apply $t$-distribution calibration for an improvement. Choose $\lambda = \lambda(k, \alpha, p)$ such that $P(|t_{k-1}| \leq \lambda) = (1-\alpha)^{1/p}$. Then the asymptotically correct $1-\alpha$ simultaneous confidence intervals for $(\mu_l)_{l=1}^p$ can be constructed in the form of (5.5) with the cutoff value $\Phi^{-1}(1-\alpha/(2p))$ therein replaced by $\lambda$. The latter simultaneous confidence intervals can be used for testing the hypothesis $H_0 : \mu = \mu^\circ$, namely $\mu_1 = \mu_1^\circ, \ldots, \mu_p = \mu_p^\circ$. We reject the null hypothesis at level $\alpha$ if there exists one of the intervals that does not include the corresponding $\mu_l^\circ$.

**6. A simulation study.** In this section, we shall study the finite-sample approximation accuracies for various normalized sums in Theorems 4.1, 4.2 and 4.3. Consider the AR-GARCH process

$$X_i = \rho X_{i-1} + \varepsilon_i, \tag{6.1}$$

where $|\rho| < 1$ and $(\varepsilon_i)$ is the GARCH(1, 1) process with

$$\varepsilon_i = \eta_i \sigma_i, \qquad \sigma_i^2 = \alpha_0^2 + \alpha_1^2 \eta_{i-1}^2 + \beta_1^2 \sigma_{i-1}^2, \tag{6.2}$$

where $\eta_i$ are i.i.d. random variables with mean 0 and $\alpha_0, \alpha_1, \beta_1$ are real parameters. The AR-GARCH process has been widely used to study heavy-tailed financial time series; see, for example, [29, 34] among others. In our simulation study, we let $\eta_i \sim 0.75^{1/2} t_8$ so that it has variance 1. By Basrak, Davis and Mikosch [2], let $p = p(\beta_1)$ be such that $E(|\beta_1 \eta_0|^p) = 1$. Then $\varepsilon_i$ has finite $r$th moment with $r \in (0, p)$, but $E(|\varepsilon_i|^p) = \infty$.

We let $\alpha_0 = 1, \alpha_1 = 0.4, \beta_1 = 0.4$, $n = 200$, and choose 10 levels of $\rho$: $\rho = 0, 0.1, \ldots, 0.9$. In the block normalized sum $T_n^\dagger$ in (4.5), we let $m = 10$ and $b = n/m = 20$. For $W_n^*$ of (4.8), we let $m_2 = \lfloor m^{1/2} \rfloor$ and $m_1 = m - m_2$. In the interlacing version $I_n^*$ of (4.11) and in Corollaries 4.1, 4.2 and Theorem 4.3, we consider 11 levels of $x$: $x = 2.0, 2.2, \ldots, 4.0$. As discussed in Section 4.7, instead of the Gaussian approximation, more accuracy can be gained if we use the Student $t$-distribution. If $X_i$ were i.i.d. standard normal, then $I_n^*$ has $t$ distribution with degrees of freedom $n/(2m) - 1 = 9$, $T_n^\dagger \sim t_{n/m-1} = t_{19}$ and $W_n^* \sim t_{k-1} = t_{19}$. When $x$ becomes larger, as expected, the Gaussian approximation becomes worse. For example, $(1 - P(t_9 \geq 4))/(1 - \Phi(4)) = 49.1$.

Tables 1, 2 and 3 show the ratios $P(T_n^\dagger \geq x)/(1 - P(t_{b-1} \geq x))$, $P(I_n^* \geq x)/(1 - P(t_{b/2-1} \geq x))$ and $P(W_n^* \geq x)/(1 - P(t_{b-1} \geq x))$, where the probabilities $P(T_n^\dagger \geq x)$, $P(I_n^* \geq x)$ and $P(W_n^* \geq x)$ are approximated by simulating

TABLE 1
*Moderate deviation ratios $P(T_n^\dagger \geq x)/(1 - P(t_{b/2-1} \geq x))$ for the* AR(1)–GARCH(1, 1) *process* (6.1) *with $x = 2, 2.2, \ldots, 4.0$ and $\rho = 0, \ldots, 0.9$*

| $x$ | $\rho = 0$ | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 |
|---|---|---|---|---|---|---|---|---|---|---|
| 2 | 1.01 | 1.05 | 1.10 | 1.15 | 1.23 | 1.33 | 1.49 | 1.77 | 2.41 | 4.22 |
| 2.2 | 1.00 | 1.05 | 1.11 | 1.17 | 1.26 | 1.37 | 1.56 | 1.92 | 2.73 | 5.22 |
| 2.4 | 0.99 | 1.05 | 1.11 | 1.18 | 1.29 | 1.42 | 1.65 | 2.08 | 3.12 | 6.50 |
| 2.6 | 0.99 | 1.05 | 1.12 | 1.19 | 1.32 | 1.47 | 1.73 | 2.25 | 3.56 | 8.14 |
| 2.8 | 0.97 | 1.06 | 1.12 | 1.21 | 1.35 | 1.53 | 1.81 | 2.44 | 4.05 | 10.22 |
| 3 | 0.96 | 1.05 | 1.13 | 1.23 | 1.38 | 1.57 | 1.90 | 2.63 | 4.62 | 12.88 |
| 3.2 | 0.95 | 1.04 | 1.13 | 1.24 | 1.40 | 1.61 | 2.01 | 2.84 | 5.28 | 16.23 |
| 3.4 | 0.95 | 1.04 | 1.13 | 1.24 | 1.43 | 1.66 | 2.11 | 3.07 | 5.99 | 20.45 |
| 3.6 | 0.93 | 1.05 | 1.13 | 1.24 | 1.45 | 1.70 | 2.20 | 3.32 | 6.77 | 25.74 |
| 3.8 | 0.90 | 1.04 | 1.15 | 1.23 | 1.45 | 1.74 | 2.33 | 3.57 | 7.68 | 32.35 |
| 4 | 0.88 | 1.03 | 1.17 | 1.22 | 1.47 | 1.77 | 2.45 | 3.82 | 8.59 | 40.48 |

TABLE 2
*Moderate deviation ratios* $P(I_n^* \geq x)/(1 - P(t_{b/2-1} \geq x))$ *for the* AR(1)–GARCH(1, 1) *process* (6.1) *with* $x = 2, 2.2, \ldots, 4.0$ *and* $\rho = 0, \ldots, 0.9$

| $x$ | $\rho = 0$ | **0.1** | **0.2** | **0.3** | **0.4** | **0.5** | **0.6** | **0.7** | **0.8** | **0.9** |
|---|---|---|---|---|---|---|---|---|---|---|
| 2 | 0.98 | 0.98 | 0.98 | 0.98 | 0.98 | 0.98 | 0.98 | 1.00 | 1.09 | 1.71 |
| 2.2 | 0.97 | 0.97 | 0.97 | 0.97 | 0.97 | 0.98 | 0.97 | 0.99 | 1.10 | 1.83 |
| 2.4 | 0.97 | 0.96 | 0.96 | 0.96 | 0.96 | 0.97 | 0.97 | 0.99 | 1.10 | 1.93 |
| 2.6 | 0.96 | 0.95 | 0.95 | 0.95 | 0.96 | 0.96 | 0.95 | 0.98 | 1.10 | 2.04 |
| 2.8 | 0.95 | 0.94 | 0.94 | 0.94 | 0.94 | 0.96 | 0.94 | 0.97 | 1.10 | 2.15 |
| 3 | 0.94 | 0.93 | 0.93 | 0.93 | 0.94 | 0.94 | 0.93 | 0.96 | 1.11 | 2.26 |
| 3.2 | 0.93 | 0.92 | 0.92 | 0.91 | 0.93 | 0.94 | 0.92 | 0.95 | 1.10 | 2.37 |
| 3.4 | 0.92 | 0.91 | 0.91 | 0.90 | 0.91 | 0.93 | 0.91 | 0.95 | 1.11 | 2.48 |
| 3.6 | 0.90 | 0.91 | 0.90 | 0.89 | 0.90 | 0.92 | 0.90 | 0.94 | 1.11 | 2.57 |
| 3.8 | 0.89 | 0.90 | 0.88 | 0.89 | 0.88 | 0.91 | 0.90 | 0.93 | 1.11 | 2.66 |
| 4 | 0.88 | 0.89 | 0.87 | 0.88 | 0.88 | 0.91 | 0.89 | 0.92 | 1.11 | 2.76 |

$5 \times 10^6$ realizations of the AR(1)–GARCH(1, 1) process (6.1). When $\rho$ is small, all the three normalized sums have comparable performance. As the dependence becomes stronger, namely $\rho$ is bigger, or $x$ moves away from 0, the moderate deviation approximations for $T_n^\dagger$ and $W_n^*$ become worse with the latter being slightly better, while the interlacing normalized sum $I_n^*$ has a relatively consistent good performance. The better finite-sample performance of the interlacing normalized sum can be intuitively explained by the fact that, due to the dependence, for two consecutive blocks, the second block does not add too much new information. This is especially so when the dependence is strong. In practice, we suggest using $I_n^*$.

TABLE 3
*Moderate deviation ratios* $P(W_n^* \geq x)/(1 - P(t_{b-1} \geq x))$ *for the* AR(1)–GARCH(1, 1) *process* (6.1) *with* $x = 2, 2.2, \ldots, 4.0$ *and* $\rho = 0, \ldots, 0.9$

| $x$ | $\rho = 0$ | **0.1** | **0.2** | **0.3** | **0.4** | **0.5** | **0.6** | **0.7** | **0.8** | **0.9** |
|---|---|---|---|---|---|---|---|---|---|---|
| 2 | 0.99 | 0.99 | 0.99 | 1.00 | 1.02 | 1.06 | 1.16 | 1.39 | 1.99 | 3.86 |
| 2.2 | 0.98 | 0.98 | 0.98 | 0.99 | 1.02 | 1.06 | 1.18 | 1.45 | 2.19 | 4.71 |
| 2.4 | 0.97 | 0.97 | 0.97 | 0.98 | 1.01 | 1.06 | 1.20 | 1.51 | 2.42 | 5.78 |
| 2.6 | 0.95 | 0.96 | 0.96 | 0.97 | 1.00 | 1.06 | 1.21 | 1.57 | 2.67 | 7.12 |
| 2.8 | 0.94 | 0.95 | 0.95 | 0.96 | 1.00 | 1.05 | 1.22 | 1.62 | 2.94 | 8.79 |
| 3 | 0.93 | 0.94 | 0.94 | 0.95 | 0.98 | 1.04 | 1.23 | 1.68 | 3.25 | 10.88 |
| 3.2 | 0.93 | 0.92 | 0.92 | 0.94 | 0.97 | 1.03 | 1.24 | 1.75 | 3.57 | 13.45 |
| 3.4 | 0.92 | 0.90 | 0.91 | 0.93 | 0.96 | 1.03 | 1.27 | 1.83 | 3.90 | 16.65 |
| 3.6 | 0.90 | 0.87 | 0.90 | 0.91 | 0.94 | 1.04 | 1.28 | 1.89 | 4.27 | 20.53 |
| 3.8 | 0.91 | 0.85 | 0.86 | 0.88 | 0.94 | 1.03 | 1.29 | 1.94 | 4.67 | 25.29 |
| 4 | 0.87 | 0.84 | 0.85 | 0.86 | 0.93 | 1.01 | 1.31 | 2.01 | 5.13 | 31.07 |

**7. Proofs.** The main idea of the proofs of Theorems 4.1 and 4.3 is to use $m$-dependence approximation. For $\beta$-mixing variables, we can apply Berbee's [4] theorem and convert them to independent variables. For GMC processes, we can also use $m$-dependence approximation. Then we apply the moderate deviation of [26] for independent random variables. The proof of Theorem 4.2 involves a much more delicate argument, as it requires an interesting recursive use of a tail probability inequality of self-normalized sums of dependent random variables.

7.1. *Proof of Theorem* 4.1. Before we prove Theorem 4.1, we first collect some preliminary lemmas.

LEMMA 7.1. *Let $\xi_i, 1 \le i \le n$ be a sequence of random variables on the same probability space and define $\beta^{(i)} = \beta(\xi_i, (\xi_{i+1}, \ldots, \xi_n))$. Then the probability space can be extended with random variables $\tilde{\xi}_i$ distributed as $\xi_i$ such that $\tilde{\xi}_i, 1 \le i \le n$ are independent and*

$$P(\xi_i \ne \tilde{\xi}_i \text{ for some } 1 \le i \le n) \le \beta^{(1)} + \cdots + \beta^{(n-1)}.$$

This is Lemma 2.1 of Berbee [4]. By Theorem 4.1 in Shao and Yu [40], we have the following.

LEMMA 7.2. *Under Assumptions* (4.1) *and* (2.1), *the following holds*:

$$E|S_{k,m}|^{r'} \le c_0 m^{r'/2} c_1^{r'}, \tag{7.1}$$

*for any $2 \le r' < r$, $m \ge 1$, $k \ge 0$, where $c_0$ is a constant depending only on $r', r, a_1, a_2$ and $\tau$.*

PROOF OF THEOREM 4.1. Clearly, (7.1) and (4.2) yield

$$\frac{\sum_{j=1}^k E|Y_{j,1}|^{2+\delta}}{(\sum_{j=1}^k EY_{j,1}^2)^{(2+\delta)/2}} \le 2c_0 n^{-(1-\alpha_1)\delta/2}(c_1/c_2)^{2+\delta}. \tag{7.2}$$

Let $\tilde{Y}_j, 1 \le j \le k$ be independent random variables such that $\tilde{Y}_j$ and $Y_{j,1}$ have the same distribution for each $1 \le j \le k$. Set

$$\tilde{W}_n = \frac{\sum_{j=1}^k \tilde{Y}_j}{(\sum_{j=1}^k \tilde{Y}_j^2)^{1/2}}.$$

By Lemma 7.1 and $k \le n/(2m_2)$, we have

$$\big|P(W_n \ge x) - P(\tilde{W}_n \ge x)\big| \le k\beta(m_2) \le a_1 \exp(-0.5 a_2 n^{\tau\alpha_2}). \tag{7.3}$$

We next apply Theorem 1.1 to $\tilde{W}_n$. It follows from (1.4) that

$$\frac{P(\tilde{W}_n \ge x)}{1 - \Phi(x)} = \exp\big(O(1)(1+x)^{2+\delta} n^{-(1-\alpha_1)\delta/2}\big) \tag{7.4}$$

for all $0 \le x \le O(1)n^{(1-\alpha_1)/2}$.

This and (7.3) imply that there exist finite constants $c_0$, $A$ depending only on $c_1/c_2, a_1, a_2, r$ and $\tau$ such that

$$\begin{aligned}\frac{P(W_n \ge x)}{1-\Phi(x)} &= \exp\bigl(O(1)(1+x)^{2+\delta} n^{-(1-\alpha_1)\delta/2}\bigr) \\ &\quad + O(1)\frac{\exp(-0.5a_2 n^{\tau\alpha_2})}{1-\Phi(x)}\end{aligned} \tag{7.5}$$

uniformly in $0 \le x \le c_0 n^{(1-\alpha_1)/2}$, and $|O(1)| \le A$. This proves (4.12).

It also follows from (1.2) that

$$\frac{P(\tilde{W}_n \ge x)}{1-\Phi(x)} = 1 + O(1)(1+x)^{2+\delta} n^{-(1-\alpha_1)\delta/2} \tag{7.6}$$

for all $0 \le x \le O(1)n^{(1-\alpha_1)\delta/(2(2+\delta))}$. This and (7.3) imply (4.13). $\square$

7.2. *Proof of Theorem* 4.3. For the proof of Theorem 4.3, we need to use the following lemma.

LEMMA 7.3. *Let $\zeta_i$, $1 \le i \le n$ be independent nonnegative random variables with $E\zeta_i^p < \infty$, where $1 < p \le 2$. Then for any $0 < y < \sum_{i=1}^n E\zeta_i$*

$$\begin{aligned}&P\left(\sum_{i=1}^n \zeta_i \le \sum_{i=1}^n E\zeta_i - y\right) \\ &\quad \le \exp\left(-\frac{(p-1)}{4}\frac{y^{p/(p-1)}}{(\sum_{i=1}^n E\zeta_i^p)^{1/(p-1)}}\right).\end{aligned} \tag{7.7}$$

PROOF. When $p = 2$, (7.7) is Theorem 2.19 in [21] with a constant $1/2$. For $1 < p \le 2$, observing that

$$e^{-x} \le 1 - x + x^p \qquad \text{for } x \ge 0,$$

we have for $t > 0$

$$\begin{aligned}&P\left(\sum_{i=1}^n \zeta_i \le \sum_{i=1}^n E\zeta_i - y\right) \\ &\quad \le e^{-ty + t\sum_{i=1}^n E\zeta_i} Ee^{-t\sum_{i=1}^n \zeta_i} \\ &\quad \le e^{-ty + t\sum_{i=1}^n E\zeta_i} \prod_{i=1}^n (1 - tE\zeta_i + t^p E\zeta_i^p) \\ &\quad \le \exp\left(-ty + t^p \sum_{i=1}^n E\zeta_i^p\right).\end{aligned}$$

Letting

$$t = \left(\frac{y^p}{p\sum_{i=1}^n E\zeta_i^p}\right)^{1/(p-1)}$$

yields

$$\begin{aligned}
P&\left(\sum_{i=1}^n \zeta_i \le \sum_{i=1}^n E\zeta_i - y\right)\\
&\le \exp\left(-\frac{(p-1)y^{p/(p-1)}}{p^{p/(p-1)}(\sum_{i=1}^n E\zeta_i^p)^{1/(p-1)}}\right)\\
&\le \exp\left(-\frac{(p-1)y^{p/(p-1)}}{4(\sum_{i=1}^n E\zeta_i^p)^{1/(p-1)}}\right),
\end{aligned}$$

as desired. □

PROOF OF THEOREM 4.3. Recall (4.10) for $Y_j$, $1 \le j \le k$. Let

$$\tilde{Y}_j = E(Y_j | \varepsilon_l, 2mj - 3m + 1 \le l \le 2mj - m),$$

and

$$\tilde{I}_n = \frac{\sum_{j=1}^k \tilde{Y}_j}{\tilde{V}} \qquad \text{where } \tilde{V}^2 = \sum_{j=1}^k \tilde{Y}_j^2.$$

Note that $\tilde{Y}_j$ are independent, and by (2.5),

$$\|Y_j - \tilde{Y}_j\|_r \le m a_1 e^{-a_2 m^\tau}. \tag{7.8}$$

Under (2.5), since $X_l = \sum_{i=0}^\infty \mathcal{P}_{l-i} X_l$, where $\mathcal{P}_k \cdot = E(\cdot|\mathcal{F}_k) - E(\cdot|\mathcal{F}_{k-1})$, we have by Burkholder's martingale inequality (cf. [8]) that

$$\begin{aligned}
\|Y_j\|_r &= \left\|\sum_{i=0}^\infty \sum_{l\in H_j} \mathcal{P}_{l-i} X_l\right\|_r\\
&\le \sum_{i=0}^\infty \left\|\sum_{l\in H_j} \mathcal{P}_{l-i} X_l\right\|_r\\
&\le \sum_{i=0}^\infty (r-1)^{1/2}\left(\sum_{l\in H_j} \|\mathcal{P}_{l-i} X_l\|_r^2\right)^{1/2}\\
&\le (r-1)^{1/2}\sum_{i=0}^\infty \big(m\theta_r^2(i)\big)^{1/2} = c_3 m^{1/2},
\end{aligned}$$

where $c_3 = (r-1)^{1/2} \sum_{i=0}^{\infty} \theta_r(i) < \infty$. By condition (4.2) and (7.8), there exists a constant $c_5 > 0$ such that $E\tilde{V}^2 \geq c_5 n$. By Lemma 7.3 with $p = r/2$, $\zeta_j = \tilde{Y}_j^2$ and $y = c_5 n/2$, we have by elementary calculations that

$$P(\tilde{V}^2 \geq c_5 n/2) \geq 1 - \exp(-c_6 k) \geq 1 - \exp(-c_6' n^{1-\alpha}) \tag{7.9}$$

for some constants $c_6, c_6' > 0$. Also (7.8) and $m \asymp n^\alpha$ imply

$$P(|Y_j - \tilde{Y}_j| \geq n^{-9}) \leq n^{9r} m^r a_1^r e^{-ra_2 m^\tau} = O(1)\exp(-ra_2 n^{\tau\alpha}/2).$$

Hence, there exist $c_7, c_8 > 0$ such that

$$\begin{aligned} P(|I_n - \tilde{I}_n| \geq n^{-2}, \tilde{V}^2 \geq c_5 n) &\leq c_7 n^{c_8} e^{-ra_2 m^\tau} \\ &= O(1)\exp(-ra_2 n^{\tau\alpha}/2). \end{aligned} \tag{7.10}$$

Observe that

$$\max_{0 \leq x \leq n} \left| \frac{1 - \Phi(x)}{1 - \Phi(x \pm n^{-2})} - 1 \right| = O(n^{-1}). \tag{7.11}$$

For $0 \leq x \leq c_0 n^{\min((1-\alpha), \tau\alpha)/2}$ with a small constant $c_0 > 0$, it is easy to see that

$$\exp(-c_6' n^{1-\alpha}) + \exp(-ra_2 n^{\tau\alpha}/2) = o(1)(1 - \Phi(x)) \exp\left(O(1)\frac{(1+x)^r}{n^{(1-\alpha)\delta/2}}\right).$$

Applying Theorem 1.1 to $\tilde{I}_n$, we have, for some constant $c_4 > 0$, that

$$\frac{P(\tilde{I}_n \geq x)}{1 - \Phi(x)} = \exp\left(O(1)\frac{(1+x)^r}{k^{r/2-1}}\right) \tag{7.12}$$

for $0 \leq x \leq c_4 k^{1/2}$. Hence, (4.17) follows from (7.12), (7.9), (7.10) and (7.11) with elementary calculations. (4.18) follows similarly. □

PROOF OF COROLLARY 4.3. Using the arguments in the proof of Theorem 4.2 and the $L^r$ moment inequalities in the proof of Theorem 4.3, Corollary 4.3 readily follows. Details are omitted. □

7.3. *Proof of Theorem* 5.1. Assume without loss of generality that $\mu_l = 0$ for all $l \leq p$. As in the proof of Theorem 4.3 let

$$\tilde{Y}_{il} = E(Y_{il} | \varepsilon_h, 2mi - 3m + 1 \leq h \leq 2mi - m), \qquad 1 \leq l \leq p.$$

Let $\tilde{X}_{il} = \tilde{Y}_{il}/\sqrt{m}$, $\bar{Y}_l^\diamond = k^{-1}\sum_{j=1}^k \tilde{Y}_{jl}$, $\bar{X}_l^\diamond = k^{-1}\sum_{j=1}^k \tilde{X}_{jl}$ and

$$\tilde{T}_{nl} = \frac{\sum_{j=1}^k \tilde{Y}_{jl}}{\sqrt{\sum_{j=1}^k (\tilde{Y}_{jl} - \bar{Y}_l^\diamond)^2}} = \frac{\sum_{j=1}^k \tilde{X}_{jl}}{\sqrt{\sum_{j=1}^k (\tilde{X}_{jl} - \bar{X}_l^\diamond)^2}}. \tag{7.13}$$

By Theorem 2 in [44], there exists a constant $c_r$, only depending on $r$, such that $\|\tilde{X}_{il}\|_r \leq c_r \sum_{h=0}^\infty \theta_r(h)$. Then by Assumption (G), $\|\tilde{X}_{il}\|_r \leq c_r \sum_{h=0}^\infty a_1 e^{-a_2 h} =$

$c_r a_1/(1-e^{-a_2})$. Let $\tilde{\omega}_{jl} = E(\tilde{X}_{ij}\tilde{X}_{il})$, $\tilde{R}_{jl} = \tilde{\omega}_{jl}/(\tilde{\omega}_{jj}\tilde{\omega}_{ll})^{1/2}$ and $\tilde{R} = (\tilde{R}_{jl})_{j,l\le p}$. By Assumption (G) and (7.8), $\|Y_{1l} - \tilde{Y}_{1l}\| \le m a_1 e^{-a_2 m}$. Here the $L^2$ norm $\|\cdot\| = \|\cdot\|_2$. Again by Assumption (G),

$$
\begin{aligned}
\left|E(Y_{1l}Y_{1j}) - m\omega_{lj}\right| &\le \sum_{i\in\mathbb{Z}} \left|iE(Z_{0l}Z_{ij})\right| \\
&= 2\sum_{i=1}^{\infty} i\left|E\big(Z_{0l}E(Z_{ij}|\mathcal{F}_0)\big)\right| \\
&\le 2\sum_{i=1}^{\infty} i\|Z_{0l}\|\left\|E(Z_{ij}|\mathcal{F}_0)\right\| \qquad (7.14)\\
&\le 2\sum_{i=1}^{\infty} i\|Z_{0l}\|\Delta_2^{(j)}(i) \\
&\le 2\sum_{i=1}^{\infty} i\|Z_{0l}\| a_1 e^{-a_2 i} \le C_1.
\end{aligned}
$$

Note that $E(\tilde{X}_{1l}\tilde{X}_{1j}) = E(\tilde{Y}_{1l}\tilde{Y}_{1j})/m$. Then we have uniformly over $j, l \le p$ that $|\tilde{\omega}_{jl} - \omega_{jl}| = O(m^{-1})$, which implies that $\max_{j,k\le p} |\tilde{R}_{jl} - R_{jl}| = O(m^{-1})$ in view of Assumption (R)(i). Let $\gamma = \chi/2$. By (5.8), since $m \asymp n^{1/4}$, we have $m^{-1} = o((\log p)^{-1-\gamma})$. Hence, Assumption (R) is also valid with $R_{jk}$ therein replaced by $\tilde{R}_{jl}$. For any constant $u > 0$, we have by (5.7) that $\#\{j \le p : |R_{jl}| \ge u \text{ for some } j \ne l\} = o(p)$. Consequently, by Theorem 3.1 in [31], we have

$$
\max_{l\le p} \tilde{T}_{nl}^2 - 2\log p + \log\log p \Rightarrow \mathcal{G}. \qquad (7.15)
$$

We remark that the original form of Theorem 3.1 in [31] is for two-sample mean comparisons. However, a careful check of its proof indicates that the argument works for the one sample case as well. Let

$$
\tilde{\sigma}_l^2 = k^{-1}\sum_{j=1}^{k} (\tilde{X}_{jl} - \bar{X}_l^\diamond)^2.
$$

By the $m$-dependence approximation arguments in (7.8)–(7.10), we obtain

$$
P\big(|T_{nl} - \tilde{T}_{nl}| \ge n^{-2}, \tilde{\sigma}_l^2 \ge c_1, \text{holds for all } l \le p\big) = O\big(pn^{c_2}e^{-ra_2 m}\big)
$$

for some constants $c_1, c_2 > 0$. Note that, by (5.8) and $m \asymp n^{1/4}$, we have $pn^{c_2}e^{-ra_2 m} \to 0$, which by (7.15) implies (5.9) via elementary manipulations.

**Acknowledgments.** We are grateful to the Editor, an Associate Editor and two anonymous referees for their helpful comments and suggestions.

## SUPPLEMENTARY MATERIAL

**Supplement to "Self-normalized Cramér type moderate deviations under dependence"** (DOI: 10.1214/15-AOS1429SUPP; .pdf). The supplement gives the detailed proof for Theorem 4.2.

## REFERENCES


[1] ANDREWS, D. W. K. (1984). Nonstrong mixing autoregressive processes. *J. Appl. Probab.* **21** 930–934. MR0766830

[2] BASRAK, B., DAVIS, R. A. and MIKOSCH, T. (2002). Regular variation of GARCH processes. *Stochastic Process. Appl.* **99** 95–115. MR1894253

[3] BEARE, B. K. (2010). Copulas and temporal dependence. *Econometrica* **78** 395–410. MR2642867

[4] BERBEE, H. (1987). Convergence rates in the strong law for bounded mixing sequences. *Probab. Theory Related Fields* **74** 255–270. MR0871254

[5] BERCU, B. and TOUATI, A. (2008). Exponential inequalities for self-normalized martingales with applications. *Ann. Appl. Probab.* **18** 1848–1869. MR2462551

[6] BRADLEY, R. (2007). *Introduction to Strong Mixing Conditions*. Kendrick Press, Heber City, UT.

[7] BÜHLMANN, P. (2002). Bootstraps for time series. *Statist. Sci.* **17** 52–72. MR1910074

[8] BURKHOLDER, D. L. (1988). Sharp inequalities for martingales and stochastic integrals. *Astérisque* **157–158** 75–94. MR0976214

[9] CARRASCO, M. and CHEN, X. (2002). Mixing and moment properties of various GARCH and stochastic volatility models. *Econometric Theory* **18** 17–39. MR1885348

[10] CHANG, J., CHEN, S. X. and CHEN, X. (2015). High dimensional generalized empirical likelihood for moment restrictions with dependent data. *J. Econometrics* **185** 283–304. MR3300347

[11] CHEN, X., SHAO, Q., WU, W. and XU, L. (2016). Supplement to "Self-normalized Cramér-type moderate deviations under dependence." DOI:10.1214/15-AOS1429SUPP.

[12] CHEN, R. and TSAY, R. (1993). Nonlinear additive ARX models. *J. Amer. Statist. Assoc.* **88** 955–967.

[13] CHEN, R. and TSAY, R. S. (1993). Functional-coefficient autoregressive models. *J. Amer. Statist. Assoc.* **88** 298–308. MR1212492

[14] CHEN, S. X. and QIN, Y. (2010). A two-sample test for high-dimensional data with applications to gene-set testing. *Ann. Statist.* **38** 808–835. MR2604697

[15] CHEN, X. (2012). Penalized sieve estimation and inference of semi-nonparametric dynamic models: A selective review. In *Advances in Economics and Econometrics*, 2010 *World Congress of the Econometric Society Book Volumes* Cambridge Univ. Press, Cambridge.

[16] CHEN, X., HANSEN, L. P. and CARRASCO, M. (2010). Nonlinearity and temporal dependence. *J. Econometrics* **155** 155–169. MR2607192

[17] CHEN, X., WU, W. B. and YI, Y. (2009). Efficient estimation of copula-based semiparametric Markov models. *Ann. Statist.* **37** 4214–4253. MR2572458

[18] DAVYDOV, Y. A. (1968). Convergence of distributions generated by stationary stochastic processes. *Theory Probab. Appl.* **13** 691–696.

[19] DEDECKER, J., DOUKHAN, P., LANG, G., LEÓN, J. R., LOUHICHI, S. and PRIEUR, C. (2007). *Weak Dependence: With Examples and Applications*. *Lecture Notes in Statistics* **190**. Springer, New York. MR2338725

[20] DELAIGLE, A., HALL, P. and JIN, J. (2011). Robustness and accuracy of methods for high dimensional data analysis based on Student's $t$-statistic. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **73** 283–301. MR2815777

[21] DE LA PEÑA, V. H., LAI, T. L. and SHAO, Q. (2009). *Self-Normalized Processes*: *Limit Theory and Statistical Applications*. Springer, Berlin. MR2488094

[22] DOUC, R., MOULINES, E., OLSSON, J. and VAN HANDEL, R. (2011). Consistency of the maximum likelihood estimator for general hidden Markov models. *Ann*. *Statist*. **39** 474–513. MR2797854

[23] DOUKHAN, P. and WINTENBERGER, O. (2008). Weakly dependent chains with infinite memory. *Stochastic Process*. *Appl*. **118** 1997–2013. MR2462284

[24] FAN, J., HALL, P. and YAO, Q. (2007). To how many simultaneous hypothesis tests can normal, Student's $t$ or bootstrap calibration be applied? *J*. *Amer*. *Statist*. *Assoc*. **102** 1282–1288. MR2372536

[25] FAN, J. and YAO, Q. (2003). *Nonlinear Time Series*: *Nonparametric and Parametric Methods*. Springer, New York. MR1964455

[26] JING, B., SHAO, Q. and WANG, Q. (2003). Self-normalized Cramér-type large deviations for independent random variables. *Ann*. *Probab*. **31** 2167–2215. MR2016616

[27] JUODIS, M. and RAČKAUSKAS, A. (2005). A remark on self-normalization for dependent random variables. *Lith*. *Math*. *J*. **45** 142–151.

[28] LAHIRI, S. N. (2003). *Resampling Methods for Dependent Data*. Springer, New York. MR2001447

[29] LANGE, T. (2011). Tail behavior and OLS estimation in AR–GARCH models. *Statist*. *Sinica* **21** 1191–1200. MR2827520

[30] LIN, D. and FOSTER, D. (2014). The power of a few large blocks: A credible assumption with incredible efficiency. Working paper.

[31] LIU, W. and SHAO, Q. (2013). A Carmér moderate deviation theorem for Hotelling's $T^2$-statistic with applications to global tests. *Ann*. *Statist*. **41** 296–322. MR3059419

[32] LIU, W., SHAO, Q. and WANG, Q. (2013). Self-normalized Cramér type moderate deviations for the maximum of sums. *Bernoulli* **19** 1006–1027. MR3079304

[33] MASRY, E. and TJØSTHEIM, D. (1995). Nonparametric estimation and identification of nonlinear ARCH time series. *Econometric Theory* **11** 258–289. MR1341250

[34] MEITZ, M. and SAIKKONEN, P. (2011). Parameter estimation in nonlinear AR–GARCH models. *Econometric Theory* **27** 1236–1278. MR2868839

[35] MEYN, S. P. and TWEEDIE, R. L. (1993). *Markov Chains and Stochastic Stability*. Springer, London. MR1287609

[36] PHAM, T. D. and TRAN, L. T. (1985). Some mixing properties of time series models. *Stochastic Process*. *Appl*. **19** 297–303. MR0787587

[37] POLITIS, D. N., ROMANO, J. P. and WOLF, M. (1999). *Subsampling*. Springer, New York. MR1707286

[38] SAKHANENKO, A. I. (1991). Estimates of Berry–Esseen type for the probabilities of large deviations. *Siberian Math*. *J*. **32** 647–656.

[39] SHAO, Q. and WANG, Q. (2013). Self-normalized limit theorems: A survey. *Probab*. *Surv*. **10** 69–93. MR3161676

[40] SHAO, Q. and YU, H. (1996). Weak convergence for weighted empirical processes of dependent sequences. *Ann*. *Probab*. **24** 2098–2127. MR1415243

[41] TONG, H. (1990). *Nonlinear Time Series*: *A Dynamical System Approach*. *Oxford Statistical Science Series* **6**. Clarendon Press, Oxford. MR1079320

[42] VAN DER VAART, A. W. (1998). *Asymptotic Statistics*. *Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge Univ. Press, Cambridge. MR1652247

[43] WANG, Q. and HALL, P. (2009). Relative errors in central limit theorems for Student's $t$ statistic, with applications. *Statist*. *Sinica* **19** 343–354. MR2487894

[44] WU, W. B. (2005). Nonlinear system theory: Another look at dependence. *Proc*. *Natl*. *Acad*. *Sci*. *USA* **102** 14150–14154 (electronic). MR2172215

[45] WU, W. B. (2011). Asymptotic theory for stationary processes. *Stat. Interface* **4** 207–226. MR2812816

[46] WU, W. B. and SHAO, X. (2004). Limit theorems for iterated random functions. *J. Appl. Probab.* **41** 425–436. MR2052582

X. CHEN
DEPARTMENT OF ECONOMICS
YALE UNIVERSITY
30 HILLHOUSE AVE., BOX 208281
NEW HAVEN, CONNECTICUT 06520
USA
E-MAIL: xiaohong.chen@yale.edu

Q.-M. SHAO
DEPARTMENT OF STATISTICS
CHINESE UNIVERSITY OF HONG KONG
SHATIN, N.T.
HONG KONG
E-MAIL: qmshao@sta.cuhk.edu.hk

W. B. WU
DEPARTMENT OF STATISTICS
UNIVERSITY OF CHICAGO
5734 S UNIVERSITY AVE
CHICAGO, ILLINOIS 60637
USA
E-MAIL: wbwu@galton.uchicago.edu

L. XU
FACULTY OF SCIENCE AND TECHNOLOGY
UNIVERSITY OF MACAU
AVENIDA DA UNIVERSIDADE, TAIPA
MACAU
CHINA
E-MAIL: lihuxu@umac.mo