Let \( \{(X_i, Y_i)\}_{i=1}^n \) be a sequence of independent bivariate random vectors. In this paper, we establish a refined Cramér type moderate deviation theorem for the general self-normalized sum \( \frac{\sum_{i=1}^n X_i}{(\sum_{i=1}^n Y_i^2)^{1/2}} \), which unifies and extends the classical Cramér (1938) theorem and the self-normalized Cramér type moderate deviation theorems by Jing, Shao and Wang (2003) as well as the further refined version by Wang (2011). The advantage of our result is evidenced through successful applications to weakly dependent random variables and self-normalized winsorized mean. Specifically, by applying our new framework on general self-normalized sum, we significantly improve Cramér type moderate deviation theorems for one-dependent random variables, geometrically \( \beta \)-mixing random variables and causal processes under geometrical moment contraction. As an additional application, we also derive the Cramér type moderate deviation theorems for self-normalized winsorized mean.

1. Introduction. Let \( X_1, X_2, \ldots, X_n \) be independent random variables with \( \mathbb{E}X_i = 0 \) and \( \mathbb{E}X_i^2 < \infty \) for \( i \geq 1 \). Set \( B_n^2 = \sum_{i=1}^n \mathbb{E}X_i^2 \),

\[
S_n = \sum_{i=1}^n X_i, \quad V_n^2 = \sum_{i=1}^n X_i^2, \quad \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i
\]

and \( \hat{\sigma}_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 \).

The self-normalized sum is defined by \( S_n/V_n \) and is closely related to the widely-used Student’s \( t \) statistic \( t_n = S_n/(\sqrt{n} \hat{\sigma}_n) \) in the sense that

\[
\mathbb{P}(t_n \geq x) = \mathbb{P}(S_n/V_n \geq x[n/(n + x^2 - 1)]^{1/2}).
\]

Therefore, to investigate the distribution of Student’s \( t \) statistic is equivalent to consider that of the less complex self-normalized statistic.

The past three decades have witnessed the flourishing development of asymptotic theory for self-normalized sums of independent random variables. Regarding the sufficient and necessary conditions for the self-normalized central limit theorem, we refer to Giné, Götze and Mason (1997) and Shao (2018) for independent and identically distributed (i.i.d.) random variables and general non-i.i.d. random variables, respectively. Specifically, for the i.i.d.

\[†\] The research is partially supported by National Nature Science Foundation of China NSFC 12031005, Shenzhen Outstanding Talents Training Fund, Hong Kong Research Grants Council GRF-14302515 and 14304917. MSC2020 subject classifications: Primary 62E17, 62E20; secondary 62J07.

Keywords and phrases: Moderate deviation, Self-normalized sums, Dependent random variables, High-dimensional, Winsorized mean.
In addition, Wang (2011) corrected the skewness in normal approximation and proved that if \(|X_i|^3 < \infty\) for \(i \geq 1\), then there exist positive constants \(A_0\) and \(C_0\) such that

\[
\frac{\mathbb{P}(S_n > xV_n)}{1 - \Phi(x)} = e^{O_1 \Delta_{n,x}} \left[ 1 + O_2 \left( (1 + x)^3 \sum_{i=1}^{n} \mathbb{E}|X_i|^3 / B_n^3 \right) \right],
\]

holds uniformly for \(|c| \leq x/5\) and for all \(0 < x \leq \frac{1}{3} B_n (\max_{i} \mathbb{E}|X_i|^3)^{-1/3}\) and \(x \leq C_0 B_n^3 / \sum_{i=1}^{n} \mathbb{E}|X_i|^3\), where \(|O_1| \leq A_0\), \(|O_2| \leq A_0\) and

\[
\psi_x = \exp \left[ \gamma^2 \left( \frac{4\gamma}{3} - 2 \right) x^3 \sum_{i=1}^{n} \mathbb{E} X_i^3 / B_n^3 \right],
\]

\[
\Delta_{n,x} = (1 + x)^3 B_n^{-3} \sum_{i=1}^{n} \mathbb{E} |X_i|^3 1((1 + x)|X_i| \geq B_n) \]

\[
\quad + (1 + x)^4 B_n^{-4} \sum_{i=1}^{n} \mathbb{E} |X_i|^4 1((1 + x)|X_i| \leq B_n),
\]

with \(\gamma = \frac{1}{2}(1 + c/x)\). Especially, if \(X_1, \ldots, X_n\) are i.i.d. random variables with \(\mathbb{E}X_i^4 < \infty\), then (1.2) implies there exist positive constants \(A_0\) and \(C_0\) depending on \(\mathbb{E}X_i^2\) and \(\mathbb{E}X_i^4\) such that

\[
\frac{\mathbb{P}(S_n > xV_n)}{1 - \Phi(x)} = \exp \left\{ - \frac{x^3 \mathbb{E}X_i^3}{3 \sqrt{n} (\mathbb{E}X_i^2)^{3/2}} \right\} \left[ 1 + O_1 \left( \frac{1 + x}{\sqrt{n}} + \frac{(1 + x)^4}{n} \right) \right],
\]

uniformly in \(0 < x \leq C_0 n^{1/4}\), where \(|O_1| \leq A_0\). Observe that in the i.i.d. case, the classical self-normalized Cramér type moderate deviation presented in (1.1) gives a convergence rate
of \((1 + x)^3 / \sqrt{n}\) and the corresponding range of convergence \(x = o(n^{1/6})\). Thus, by specifying
the skewness correction term \(\Psi_x\), (1.2) can improve the result of Jing, Shao and Wang (2003) in terms of both the convergence rate and the range of convergence when the higher
fourth moments exist.

It is worth mentioning that the moment conditions for self-normalized Cramér type moderate
deivation theorems are much weaker than those in the classical theorems for standardized
sums. As a result, to account for robustness against heavy-tailed data, the self-normalized
sum would be recommended in real-world applications. We refer to de la Peña, Lai and
Shao (2009) for a systematic introduction to the theory and statistical applications of self-
normalized statistics.

Due to its rigorous control on the ratio of tail probabilities, the self-normalized Cramér


type moderate deviation has been successfully applied in high-dimensional statistical analysis,
including large-scale multiple testing (Fan, Hall and Yao, 2007; Liu and Shao, 2010, 2013), signal detection (Delaigle and Hall, 2009), classification (Fan and Fan, 2008) and
feature screening (Chang, Tang and Wu, 2016) among others.

Most of the existing works have focused on the classical self-normalized sum, that is,
\[ \sum_{i=1}^{n} X_i / \left( \sum_{i=1}^{n} X_i^2 \right)^{1/2} \]
for independent random variables \(\{X_i\}_{i=1}^{n}\). Yet, in some scenarios, the sequence used for normalizing in the denominator could be different from the numerator, which occurs for a variety of commonly used studentized nonlinear statistics such as the
studentized U-statistic and the studentized L-statistics. Therefore, investigations into general
self-normalized processes beyond the classical form are imperative. Shao and Zhou (2016)
attempted to extend the Cramér type moderate deviation theorem to a more general setting,
that is, \( \left( \sum_{i=1}^{n} X_i + D_{1n} \right) / \left( \left( \sum_{i=1}^{n} X_i^2 \right)(1 + D_{2n}) \right)^{1/2} \), where the remainders \(D_{1n}\) and \(D_{2n}\)
are measurable functions of \(\{X_i\}_{i=1}^{n}\) but negligible. Our present work will establish a fundamental framework in Theorem 2.1 on the Cramér type moderate deviation for a more general
self-normalized form of \( \sum_{i=1}^{n} X_i / \left( \sum_{i=1}^{n} Y_i^2 \right)^{1/2} \), where \(\{(X_i, Y_i)\}_{i=1}^{n}\) is a sequence of independent bivariate random vectors and \(X_i\) and \(Y_i\) could be different from each other. It is
worthwhile to mention that our Theorem 2.1 can cover not only the classical Cramér type
moderate deviation for standardized sums by Cramér (1938), but also the self-normalized
counterparts by Jing, Shao and Wang (2003) and Wang (2011).

Our investigation into the general self-normalized sum is also motivated by seeking to
develop sharper self-normalized Cramér type moderate deviation results for weakly dependent
random variables. Though Cramér type moderate deviation theory has been well studied
for independent random variables, the theory for dependent data remains largely undeveloped.
The biggest challenge is that the classical theory for independent random variables
cannot be directly applied due to dependence. Chen et al. (2016) made the first attempt to
develop the theory for self-normalized sums of dependent random variables with geometrically
decaying dependence. However, their result can be further improved by applying our framework on the general self-normalized sum. The key observation is that after dividing
the weakly dependent random variables into consecutive big blocks and small blocks, its
self-normalized sum can be approximated by a general self-normalized sum of independent
bivariate random vectors. Therefore, Cramér type moderate deviation theorems for the self-
normalized sums of weakly dependent random variables can be established based on our
fundamental theory on the general self-normalized sum. More details will be presented in
Section 3.

The rest of the paper is organized as follows. Our framework on general self-normalized
Cramér type moderate deviation is presented in Section 2. Section 3 shows applications to
the self-normalized sums of weakly dependent random variables under one-dependence, ge-
ometrically \(\beta\)-mixing condition, and geometric moment contraction. Section 4 presents an
additional application to studentized winsorized mean that naturally takes the form of a gen-
eral self-normalized sum. Section 5 is devoted to proofs of the theorems in Sections 2–4.
Other technical proof details are included in the Supplementary Material.
2. Main Results. Let \((X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)\) be independent bivariate random vectors satisfying

\[
\begin{align*}
\mathbb{E}X_i &= 0 \quad \text{for } i \geq 1 \quad \text{and} \quad \sum_{i=1}^{n} \mathbb{E}X_i^2 = \sum_{i=1}^{n} \mathbb{E}Y_i^2.
\end{align*}
\]

We remark that for the convenience of presentation, \(\{X_i\}\) and \(\{Y_i\}\) are standardized so \(\sum_{i=1}^{n} \mathbb{E}X_i^2 = \sum_{i=1}^{n} \mathbb{E}Y_i^2\). In other words, one should think of \(X_i\) as \(X_{n,i}\) and similarly \(Y_i\) as \(Y_{n,i}\). Let

\[
\begin{align*}
S_n &= \sum_{i=1}^{n} X_i, \quad V_n^2 = \sum_{i=1}^{n} Y_i^2 \quad \text{and} \quad T_n = \frac{S_n}{V_n}.
\end{align*}
\]

We first propose an exponential moment condition as follows. Suppose there exists some constant \(c_0 \geq 0\) such that for \(x > 0\) satisfying (2.7),

\[
\mathbb{E}e^{\min\left\{ \frac{x^2}{Y_i^2 + c_0}, 2x \right\} \leq \infty.}
\]

The above moment condition links \(X_i\) with \(Y_i\) and shows how they interact with each other. In particular, it is automatically satisfied for the classical self-normalized sum with \(Y_i = X_i\). In other words, one should think of \(X_i\) as \(X_{n,i}\) and similarly \(Y_i\) as \(Y_{n,i}\).

The following notations will be used throughout the paper. Define

\[
\begin{align*}
L_{3,n} &= \sum_{i=1}^{n} (\mathbb{E}|X_i|^3 + \mathbb{E}|Y_i|^3), \\
\delta_{x,i} &= (1 + x)^3 \left( \mathbb{E}|X_i|^3 \mathbb{1}(|(1 + x)X_i| > 1) + \mathbb{E}|Y_i|^3 \mathbb{1}(|(1 + x)Y_i| > 1) \right) \\
&\quad + (1 + x)^2 \left( \mathbb{E}|X_i|^4 \mathbb{1}(|(1 + x)X_i| \leq 1) + \mathbb{E}|Y_i|^4 \mathbb{1}(|(1 + x)Y_i| \leq 1) \right), \\
r_{x,i} &= \mathbb{E} \left[ \exp \left( \min \left\{ \frac{X_i^2}{Y_i^2 + c_0}, 2x \right\} \right) \mathbb{1}(|(1 + x)X_i| > 1) \right], \\
R_{x,i} &= \delta_{x,i} + r_{x,i}, \quad \delta_x = \sum_{i=1}^{n} \delta_{x,i}, \quad r_x = \sum_{i=1}^{n} r_{x,i}, \quad \text{and} \quad R_x = \delta_x + r_x.
\end{align*}
\]

**Theorem 2.1.** Assume (2.1) and (2.3) are satisfied. In addition, \(\mathbb{E}|X_i|^3 < \infty\) and \(\mathbb{E}|Y_i|^3 < \infty\) for \(i \geq 1\). Then there exist absolute positive constants \(0 < c_1 \leq 1/4\) and \(A > 0\) such that

\[
\mathbb{P}(S_n \geq xV_n + c) = (1 - \Phi(x + c)) \Psi_x e^{O_1R_x (1 + O_2(1 + x)L_{3,n})},
\]

where

\[
\Psi_x = \exp \left\{ x^3 \left( \frac{1}{3} \frac{3n}{\gamma^3} \sum_{i=1}^{n} \mathbb{E}X_i^3 - 2\gamma^2 \sum_{i=1}^{n} \mathbb{E}[X_iY_i^2] \right) \right\} \quad \text{and} \quad \gamma = \frac{1}{2} \left( 1 + \frac{c}{x} \right),
\]

uniformly for \(|c| \leq x/5\) and for all \(x > 0\) satisfying

\[
(1 + x)L_{3,n} \leq c_1, \quad x^{-2}R_x \leq c_1,
\]

and

\[
x \leq \frac{1/4 \wedge \frac{1}{2\sqrt{c_0}}}{\left[ \max_i(\mathbb{E}|X_i|^3 + \mathbb{E}|Y_i|^3) \right]^{1/3}}, \quad \max_i r_{x,i} \leq c_1,
\]

where \(|O_1| \leq A\) and \(|O_2| \leq A\).
Theorem 2.1 unifies the classical standard and self-normalized Cramér type moderate deviation theorems as well as the refined version by Wang (2011). The assumption (2.3) is satisfied for a wide class of statistics, including the block sums of weakly dependent random variables and the self-normalized winsorized mean. More details will be provided in the proof of Theorems 3.1–3.3 and 4.1.

The following corollary is a straightforward application of Theorem 2.1 to the classical standardized sum of independent random variables. The proof will be given in Section A.18 in the Supplementary Material.

**Corollary 2.1.** Let $X_1, \ldots, X_n$ be i.i.d. random variables with $E[X_i] = 0$ and $E[X_i^2] = \sigma^2$. Denote $S_n = \sum_{i=1}^n X_i$. If there exists a positive constant $t_0$ such that $E[e^{tX_1}] < \infty$ for $|t| \leq t_0$, then we have

$$P(S_n > x\sigma \sqrt{n}) = \exp \left\{ \frac{x^3 E[X_1^3]}{6\sigma^3 \sqrt{n}} \left[ 1 + O\left( \frac{(1 + x)^4 n}{\sqrt{n}} \right) \right] \right\}$$

(2.8)

holds uniformly for $0 < x \leq O(n^{1/4})$.

3. Applications to Weakly Dependent Random Variables. The Cramér type moderate deviation theory has been well studied for independent random variables, yet there are few results available for dependent data. The novel work of Chen et al. (2016) made the first attempt to develop the theory for self-normalized sums of weakly dependent random variables satisfying the geometrically $\beta$-mixing condition or geometric moment contraction. In this section, we will further improve their results by applying our fundamental framework on the general self-normalized sum. Before that, we will start with a Cramér type moderate deviation theorem for self-normalized sums of one-dependent random variables. The reason to first investigate under one-dependence is two-fold. First, one-dependence is the simplest scenario of dependency and the result for one-dependent random variables can be applied to $m$-dependence, where $m$ could also go to infinity. Second, many weakly dependent random variables can be approximated by some one-dependent random variables. A typical example includes the block sums of random variables satisfying geometric moment contraction, which will be presented in Section 3.3. Therefore, the following Theorem 3.1 under one-dependence lays the foundation for Theorem 3.3 under geometric moment contraction.

3.1. Cramér type moderate deviation under one-dependence. Let $\xi_1, \xi_2, \ldots$ be one-dependent random variables, which means for $i, j \geq 1$, $\xi_i$ is independent of $\xi_j$ if $|j - i| \geq 2$. Put

$$S_n = \sum_{i=1}^n \xi_i, \quad V_n^2 = \sum_{i=1}^n \xi_i^2, \quad \rho_n = \frac{\sum_{i=1}^{n-1} E[\xi_i \xi_{i+1}]}{\sum_{i=1}^n E[\xi_i^2]}.$$

(3.1)

Note that $|\rho_n| \leq 1/2$. Moreover, in many applications where weakly dependent sequence can be approximated by some one-dependent random variables, the covariances $E[\xi_i \xi_{i+1}]$ are negligible compared to the variables $E[\xi_i^2]$ due to weak dependence, hence $\rho_n \to 0$ as $n \to \infty$. Therefore, $\rho_n$ can be moved to the remainders and the limiting distribution will still be standard normal. One can find more details in the proof of Theorem 3.3, which obtains Cramér type moderate deviation result under geometric moment contraction by applying Theorem 3.1.

Under existence of the fourth moment, we have the following theorem for self-normalized sums of one-dependent random variables.
THEOREM 3.1. Assume that \( \mathbb{E}\xi_i = 0, \mathbb{E}\xi_i^4 \leq a_i^4 \) and \( \mathbb{E}\xi_i^2 \geq a_i^2 \) for \( 1 \leq i \leq n \) and \( \rho_n \geq \rho \) for some \( \rho > -1/2 \). Denote \( a = a_1/a_2 \). Then there exist positive numbers \( a_0 \) and \( A(\rho) \) depending on \( \rho \) such that

\[
\mathbb{P}(S_n > xV_n) = \left[ 1 - \Phi\left( \frac{x}{\sqrt{1 + 2\rho_n}} \right) \right] \left( 1 + O_1 \frac{1 + x^2}{n^{1/4}} \right)
\]

holds uniformly for \( x \in (0, a_0a^{-2}n^{1/2}) \), where \( |O_1| \leq A(\rho) \).

The proof of Theorem 3.1 relies on the big-block-small-block technique and an application of Theorem 2.1. The main idea is to approximate the self-normalized sum of one-dependent random variables by a general self-normalized sum of independent random vectors based on the big blocks. In more details, let the length of big blocks be \( l = [n^\alpha] \) for \( 0 < \alpha < 1 \), where \([x]\) denotes the integer part of \( x \) for any \( x > 0 \), and the length of small blocks be only 1. Denote \( k = [n/(l+1)] \). For \( 1 \leq j \leq k \), we define the \( j \)-th big block by

\[
H_j = \{ i : (j-1)(l+1) + 1 \leq i \leq jl+1 \}
\]

and the sums over \( j \)-th big block by

\[
X_j = \sum_{i \in H_j} \xi_i \quad \text{and} \quad Y_j^2 = \sum_{i \in H_j} \xi_i^2.
\]

Observe that by this construction, the big-block sums \( \{X_j\}_{j=1}^k \) and \( \{Y_j\}_{j=1}^k \) are both sequences of independent random variables, because \( \{\xi_i\}_{i=1}^n \) are one-dependent and the adjacent big blocks \( H_j \) and \( H_{j+1} \) are separated by a random variable \( \xi_{j(l+1)} \). Since the big blocks contain \( (1-n^{-\alpha}) \) proportion of the random variables in \( \{\xi_i\}_{i=1}^n \), we can approximate the self-normalized sum \( S_n/V_n \) of \( \{\xi_i\}_{i=1}^n \) by the general self-normalized sum \( \sum_{j=1}^k X_j / (\sum_{j=1}^k Y_j^2)^{1/2} \). The crucial quantity \( r_{x,j} \) can be separated into two self-normalized sums of independent random variables due to one-dependence and thus can be bounded by using Lemma A.5 in the Supplementary Material. Therefore, Theorem 3.1 can be proved by applying Theorem 2.1 and calculating the error terms involved. More details of the proof will be provided in Section 5.2.

Compared with the classical result (1.1) for independent data, one-dependence results in a narrower zone of convergence and a slower convergence rate. Moreover, (3.2) can be easily extended to general \( m \)-dependent random variables, where \( m \) could depend on \( n \) and go to infinity. Indeed, if \( Z_1, \ldots, Z_n \) are \( m \)-dependent and suppose \( b = n/m \) is an integer for simplicity, we define \( \xi_j = \sum_{i=1+(j-1)m}^{jm} Z_i \) for \( 1 \leq j \leq b \), then \( \xi_1, \ldots, \xi_b \) are one-dependent random variables and Theorem 3.1 can be applied.

3.2. Cramér type moderate deviation under \( \beta \)-mixing. In time series, asymptotic independence conditions such as mixing conditions are usually proposed to replace independence, among which \( \beta \)-mixing is an important dependent structure and has been connected with a large class of time series models including ARMA models, GARCH models and certain Markov processes. This subsection provides a Cramér type moderate deviation theorem for block-normalized sums of geometrically \( \beta \)-mixing random variables, which improves the result by Chen et al. (2016).

Let \( \{X_i\}_{i=1}^n \) be a sequence of random variables. Let \( \sigma_{-\infty}^t \) and \( \sigma_{t+m}^\infty \) be \( \sigma \)-fields generated by \( \{X_i\}_{1 \leq i \leq t} \) and \( \{X_i\}_{i \geq t+m} \), respectively. The \( \beta \)-mixing coefficient is given by

\[
\beta(m) := \sup_t \mathbb{E}\sup_B \{|\mathbb{P}(B|\sigma_{-\infty}^t) - \mathbb{P}(B)| : B \in \sigma_{t+m}^\infty\}.
\]
We say \( \{X_i\}_{i \geq 1} \) is geometrically \( \beta \)-mixing if \( \beta(m) \) admits an exponentially decaying rate, that is, there exist positive numbers \( a_1, a_2 \) and \( \tau \) such that
\[
\beta(m) \leq a_1 e^{-a_2 m^\tau}.
\]
(3.6)

To account for dependence, the block technique is naturally used to estimate the variance of sums of dependent random variables (see Chen et al. (2016)). Set \( l = \lceil n^\alpha \rceil + 1 \) for \( 0 < \alpha < 1 \) and \( k = \lfloor n/l \rfloor \). For \( 1 \leq j \leq k \), define the \( j \)-th block and corresponding \( j \)-th block sum by
\[
H_j = \{i : l(j-1) + 1 \leq i \leq lj\} \quad \text{and} \quad Y_j = \sum_{i \in H_j} X_i,
\]
respectively. The block-normalized sum is then defined by
\[
T_k = \frac{\sum_{j=1}^k Y_j}{\sqrt{\sum_{j=1}^k Y_j^2}}.
\]

**Theorem 3.2.** Let \( \{X_i\}_{i=1}^n \) be a \( \beta \)-mixing sequence satisfying (3.6). Assume \( E[X_i] = 0 \) and there exist positive numbers \( \mu_1 \) and \( \mu_2 \) such that \( E|X_i|^r \leq \mu_1^r \) for \( r > 4 \) and \( E(\sum_{i=s}^{s+t} X_i)^2 \geq \mu_2^2 t \) for all \( i \geq 1, s \geq 0, t \geq 1 \). Then, for \( 0 < \alpha < 1 \) and \( \tau > 0 \), there exist positive numbers \( A \) and \( d_0 \) depending on \( a_1, a_2, \mu_1, \mu_2, \alpha, \tau \) such that
\[
\left| \frac{P(T_k \geq x)}{1 - \Phi(x)} - 1 \right| \leq A \left( \frac{(1+x)^2}{n^\alpha} + \frac{(1+x)^2 \log n}{n^{\min\{\alpha/2, (\alpha \tau)^2\}}} \right)
\]
uniformly in \( 0 \leq x \leq d_0 \min\{n^{\alpha/2}, (\log n)^{-1/2} n^{\min\{\alpha/2, \alpha \tau^2/4\}}\} \).

As for the choice of \( \alpha \), for any given \( \tau \), we can always choose \( \alpha \) such that \( 1 - \alpha \leq 2 \alpha \tau \), that is, \( \alpha \geq \frac{1}{1 + 2 \tau} \). To optimize the convergence rate and the range of \( x \) in (3.9),
(i) when \( \tau \geq 2 \), let \( \alpha = \frac{1}{5} \), then
\[
\left| \frac{P(T_k \geq x)}{1 - \Phi(x)} - 1 \right| = 1 + O\left( \frac{(1+x)^2 \log n}{n^{1/5}} \right)
\]
uniformly for \( x \in (0, d_1 (\log n)^{-1/2} n^{1/10}) \);
(ii) when \( \tau < 2 \), let \( \alpha = \frac{1}{1 + 2 \tau} \), then
\[
\left| \frac{P(T_k \geq x)}{1 - \Phi(x)} - 1 \right| = 1 + O\left( \frac{(1+x)^2 \log n}{n^{2/(1+2 \tau)}} \right)
\]
uniformly for \( x \in (0, d_1 (\log n)^{-1/2} n^{(1 - 2 \tau)/(1+2 \tau)}) \).

**Remark 3.1.** We now compare our result with Theorem 4.2 in Chen et al. (2016). They proved that given \( \mathbb{E}|X_i|^r < \infty \) for \( r > 3 \) and the same assumption (3.6),
\[
\left| \frac{P(T_k \geq x)}{1 - \Phi(x)} - 1 \right| \leq A \left( \frac{(1+x)^2}{n^\alpha} + \frac{(1+x)^{5/4}}{n^{(1-\alpha)/8}} \right)
\]
uniformly in \( 0 \leq x \leq d_0 \min\{\{\log n\}^{-4/5} n^{(1-\alpha)/10}, n^{\alpha \tau/2}, n^{\alpha/2}\} \). (We have to mention that the original version of Theorem 4.2 in Chen et al. (2016) missed the error term \( \frac{(1+x)^2}{n^\alpha} \) and the corresponding condition \( x \leq d_0 n^{\alpha/2} \).) When \( (1 - \alpha) > 2 \alpha \tau \), our results might be worse for some choices of \( \alpha \). However, when \( (1 - \alpha) \leq 2 \alpha \tau \), our results improve theirs in terms of both the convergence rate and the corresponding range of \( x \). The improvement is achieved by applying our framework for general self-normalized sum and correcting the bias to normal approximation by specifying the skewness term \( \Psi^*_x \).
The proof of Theorem 3.2 again builds on the big-block-small-block technique. Recall that for $1 \leq j \leq k$, $Y_j = \sum_{i \in H_j} X_i$ is a block sum defined in (3.7). We first apply big-block-small-block technique to separate the sequence $\{Y_j\}_{j=1}^k$ into consecutive big blocks and small blocks. Let the size of big-blocks be $m_1 = \lceil n^{\alpha_1} \rceil$ for $0 < \alpha_1 < 1 - \alpha$ and the size of small-blocks be only 1. Denote $k_1 = \lceil k/(m_1 + 1) \rceil$. For $1 \leq u \leq k_1$, define the $u$-th big block by

\begin{equation}
I_u = \{j : (m_1 + 1)(u - 1) + 1 \leq j \leq (m_1 + 1)u - 1\},
\end{equation}

and the sums over $u$-th big block by

\begin{equation}
\zeta_u = \sum_{j \in I_u} Y_j, \quad \eta_u^2 = \sum_{j \in I_u} Y_j^2.
\end{equation}

Then the self-normalized sum $T_k = \sum_{j=1}^k Y_j / (\sum_{j=1}^k Y_j^2)^{1/2}$ can be approximated by the general self-normalized sum $\sum_{u=1}^{k_1} \zeta_u / (\sum_{u=1}^{k_1} \eta_u^2)^{1/2}$ constructed on the big blocks. However, unlike the big-block sums under one-dependence, the big-block sums are not independent under $\beta$-mixing assumption in (3.6), but weakly dependent. Therefore, Theorem 2.1 can not be directly applied. Note that the adjacent random vectors $(\zeta_u, \eta_u)$ and $(\zeta_{u+1}, \eta_{u+1})$ depend on $\{Y_j\}_{j=(m_1+1)(u-1)+1}^{(m_1+1)u-1}$ and $\{Y_j\}_{j=(m_1+1)u+1}^{(m_1+1)u+1}$, respectively. Since $Y_j$ defined in (3.7) is a block sum of $\{X_i\}$ and by the $\beta$-mixing assumption on $\{X_i\}$, we can see the $\beta$-mixing dependence coefficient between $(\zeta_u, \eta_u)$ and $(\zeta_{u+1}, \eta_{u+1})$ is bounded by $O(e^{-a_2 n^{\alpha_1}})$, which converges to 0 as $n \to \infty$. According to Lemma 5.3 (Berbee, 1987) presented in Section 5.3, the weakly dependent random vectors $\{(\zeta_u, \eta_u)\}_{u=1}^{k_1}$ can be replaced with independent random vectors $\{(\zeta_u, \tilde{\eta}_u)\}_{u=1}^{k_1}$ that have the same marginal distributions, with probability $1 - O(k_1 e^{-a_2 n^{\alpha_1}})$. Moreover, the crucial quantity $r_{x,u}$ can be approximated by two self-normalized sums of independent random variables due to the $\beta$-mixing assumption (see Lemma 5.3) and the block technique. Consequently, our main result in Theorem 2.1 can be applied to the general self-normalized sum $\sum_{u=1}^{k_1} \tilde{\zeta}_u / (\sum_{u=1}^{k_1} \tilde{\eta}_u^2)^{1/2}$. Detailed proof will be given in Section 5.3.

3.3. Cramér type moderate deviation for causal processes under geometric moment contraction (GMC). The GMC (see Wu and Shao (2004), Hsing and Wu (2004) and Wu (2005, 2011)) is satisfied by many non-linear time series models including various GARCH models that are commonly used in statistics, econometrics and engineering. In this subsection, we present a Cramér type moderate deviation theorem for block normalized sums of random variables satisfying GMC.

Let $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ be i.i.d. random variables and define $\sigma$-fields $\mathcal{F}_t = \sigma(\ldots, \varepsilon_{t-1}, \varepsilon_t)$. Suppose that $\{X_i = G_i(\mathcal{F}_i)\}_{i \geq 1}$ is a causal process with $G_i(\cdot)$ being a measurable function such that $X_i$ is well-defined. Let $\{\varepsilon^*_i\}_{t \in \mathbb{Z}}$ be an independent copy of $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ and we similarly define $\mathcal{F}^*_t = \sigma(\ldots, \varepsilon^*_{t-1}, \varepsilon^*_t)$.

**Definition 3.1.** (GMC). Assume that $\mathbb{E}|X_i|^r < \infty$ for all $i \geq 1$ with $r > 2$. Define the functional dependence measure by

\begin{equation}
\Delta_r(n) = \sup_i \|X_i - G_i(\mathcal{F}^*_{i-n}, \varepsilon_{i-n+1}, \ldots, \varepsilon_i)\|_r,
\end{equation}

where $\| \cdot \|_r = (\mathbb{E}|\cdot|^r)^{1/r}$. We say $\{X_i\}_{i \geq 1}$ satisfies GMC if there exist positive constants $a_1$, $a_2$ and $0 < \tau \leq 1$ such that

\begin{equation}
\Delta_r(n) \leq a_1 e^{-a_2 n^{\tau}}.
\end{equation}
Note that the GMC property (3.16) implies \( \{X_i\}_{i \geq 1} \) forgets the past \( \mathcal{F}_0 = \sigma(\ldots, \varepsilon_{-1}, \varepsilon_0) \) geometrically fast.

**Remark 3.2.** Define another functional dependence measure as

\[
\theta_i(n) = \sup_i \|X_i - G_i(\ldots, \varepsilon_{i-n-2}, \varepsilon_{i-n-1}, \varepsilon_i^x, \varepsilon_{i-n}, \varepsilon_{i-n+1}, \ldots, \varepsilon_i)\|_r.
\]

The property (3.16) is equivalent to \( \theta_r(n) \leq a'_1 e^{-a'_2 n^r} \) for some positive constants \( a'_1 \) and \( a'_2 \).

Now we assume \( \{X_i\}_{i=1}^n \) is a sequence of random variables with

\[
\mathbb{E}X_i = 0, \quad \mathbb{E}|X_i|^4 < \infty,
\]

for all \( i \geq 1 \). Write \( S_{k,m} = \sum_{i=k+1}^{k+m} X_i \). Assume there exists a positive number \( \omega_1 \) such that for any \( k \geq 0, m \geq 1, \)

\[
\mathbb{E}(S_{k,m}^2) \geq \omega_1^2 m.
\]

As with the procedure for \( \beta \)-mixing random variables, we construct the block-normalized sum for random variables satisfying GMC. Let the block size \( m = \lfloor n^\alpha \rfloor \) for \( 0 < \alpha < 1 \) and \( k = \lfloor n/m \rfloor \). For \( 1 \leq j \leq k \), define the \( j \)-th block and the \( j \)-th block sum by

\[ H_j = \{ i : m(j-1) + 1 \leq i \leq mj \} \quad \text{and} \quad Y_j = \sum_{i \in H_j} X_i. \]

The block-normalized sum is then given by

\[
T_k = \frac{\sum_{j=1}^{k} Y_j}{\sqrt{\sum_{j=1}^{k} Y_j^2}}.
\]

**Theorem 3.3.** Assume \( \{X_i\}_{i=1}^n \) is a causal process satisfying (3.16), (3.18) and (3.19). Then we have for \( 0 < \alpha < 1 \) and \( \tau > 0 \), there exist positive numbers \( A \) and \( d_0 \) depending on \( a_1, a_2, \omega_1, \alpha \) and \( \tau \) such that

\[
\mathbb{P}(T_k \geq x) = \left[ 1 - \Phi(x) \right] \left( 1 + O_1\left( \frac{1+x^2}{n^{\alpha}} + \frac{1}{n^{(1-\alpha)/4}} \right) \right)
\]

uniformly in \( 0 \leq x \leq d_0 \min\{n^{\alpha/2}, n^{\alpha/2}, n^{(1-\alpha)/8}\} \), where \( |O_1| \leq A \).

Corollary 4.3 in Chen et al. (2016) stated that (3.12) also holds for the self-normalized block sum of GMC random variables. Compared with their result, our convergence rate and the associated converging range of \( x \) significantly improve theirs.

The main idea of our proof for Theorem 3.3 is to approximate \( \{Y_j\}_{j=1}^k \) by one-dependent random variables and then apply Theorem 3.1. Define

\[
\hat{Y}_j = \mathbb{E}(Y_j | \varepsilon_1, m(j-2) + 1 \leq l \leq mj)
\]

and

\[
\hat{T}_k = \frac{\sum_{j=1}^{k} \hat{Y}_j}{\left( \sum_{j=1}^{k} Y_j^2 \right)^{1/2}},
\]

where \( m = \lfloor n^\alpha \rfloor \). Since \( Y_j = \sum_{i \in H_j} X_i \) belongs to \( \mathcal{F}_{mj} \) and depends weakly on \( \mathcal{F}_{m(j-2)} \) by the GMC assumption (3.16), it is intuitive that \( \hat{Y}_j \) is close to \( Y_j \). In particular, we can prove \( \| \hat{Y}_j - Y_j \|_r \leq a_1 m e^{-a_2 m^r} \). Therefore, the self-normalized sum \( T_k \) of \( \{Y_j\}_{j=1}^k \) can be
well approximated by the self-normalized sum $\tilde{T}_k$ of $\{\tilde{Y}_j\}_{j=1}^k$. Moreover, since $\{\varepsilon_t\}_{t\in \mathbb{Z}}$ are i.i.d. random variables, it is easy to see $\{\tilde{Y}_j\}_{j=1}^k$ are one-dependent. Consequently, Theorem 3.3 can be proved by applying Theorem 3.1 and controlling the errors caused by the approximation by one-dependent random variables. We will present the detailed proof in Section 5.4.

4. Applications to self-normalized winsorized mean. Although the sample mean has always been a prominent unbiased estimator for a location parameter, it has the troubling disadvantage of being heavily influenced by gross outliers. Yet, robustness is often a desirable property, especially in real-world applications. Thus robust alternatives, typically including the trimmed mean (Rothenberg, Fisher and Tilanus (1964)), the winsorized mean (Dixon (1960), Huber (1964)), and the Huber estimator (Huber (1964, 1973)), are imperative to make more reliable statistical inference for unknown parameters. Suppose we have i.i.d. observations $Y_1, Y_2, \ldots, Y_n$ with common distribution $Y$ and

\[ \mu = \mathbb{E}[Y] \text{ and } \sigma^2 = \text{Var}(Y). \]

For a thresholding parameter $\tau > 0$ that determines the tradeoff between bias and robustness, the winsorized mean is defined by

\begin{equation}
\hat{\mu}_W = n^{-1} \sum_{i=1}^n f(Y_i),
\end{equation}

where

\begin{equation}
f(x) = x 1\{|x| \leq \tau\} + \tau 1\{x > \tau\} - \tau 1\{x < -\tau\}.
\end{equation}

The trimmed mean is defined by

\begin{equation}
\hat{\mu}_T = c_{\tau,n}^{-1} \sum_{i=1}^n Y_i 1\{|Y_i| \leq \tau\},
\end{equation}

where $c_{\tau,n} = \sum_{i=1}^n 1\{|Y_i| \leq \tau\}$. Moreover, the Huber loss (Huber (1964)) is given by

\begin{equation}
\ell_\tau(u) = \begin{cases} 
\frac{1}{2}u^2 & \text{if } |u| \leq \tau, \\
\tau|u| - \frac{1}{2}\tau^2 & \text{if } |u| > \tau,
\end{cases}
\end{equation}

which is a compromise between square loss and absolute loss. The Huber estimator is then defined as

\begin{equation}
\hat{\mu}_H = \arg\min_{\mu \in \mathbb{R}} \sum_{i=1}^n \ell_\tau(Y_i - \mu).
\end{equation}

These robust estimators are common in reducing the impact of outliers and are all asymptotically equivalent to the sample mean when the associated tuning parameter $\tau$ tends to infinity. Compared with the Huber estimator, the trimmed mean and the winsorized mean have explicit formulas and therefore are easier to be applied in real-world applications. It is well-known that these robust estimators are asymptotically normal under some regularity conditions. Recently Zhou et al. (2018) obtained a Cramér-type moderate deviation theorem for the Huber estimator when allowing the tuning parameter $\tau$ to diverge with the sample size $n$ in some regime, and they applied the result to establish theoretical guarantees for the false discovery rate in multiple testing procedure for population means. However, the statistic they investigated depends on the unknown variance, which needs to be well estimated in practice.
4.1. *Cramér-type moderate deviation for self-normalized winsorized mean and trimmed mean.* In this section, we will provide Cramér-type moderate deviation theorems for the self-normalized winsorized mean defined in (4.6) and self-normalized trimmed mean defined in (4.9), as an application of our main Theorem 2.1. The self-normalized winsorized mean and trimmed mean are asymptotically pivotal statistics in the sense that their asymptotic distributions do not depend on unknown parameters as \((n, \tau) \to (\infty, \infty)\), therefore they can be directly used in the multiple testing of population means with theoretical justification. In addition, we will see our results for self-normalized winsorized mean and trimmed mean outperform that for the Huber estimator established in Zhou et al. (2018).

Since the winsorized mean and trimmed mean have explicit expressions as presented in (4.1) and (4.3), we can easily construct the studentized counterparts by plugging in the sample variance. The studentized winsorized mean is given by

\[
S_{\tau,n} = \frac{\sum_{i=1}^{n} (f(Y_i) - \mu)}{\sqrt{\sum_{i=1}^{n} (f(Y_i) - \hat{\mu}_W)^2}}
\]

and the studentized trimmed mean is given by

\[
U_{\tau,n} = \frac{\sum_{i \in \mathcal{N}} (Y_i 1 \{ |Y_i| \leq \tau \} - \mu)}{\sqrt{\sum_{i \in \mathcal{N}} (Y_i 1 \{ |Y_i| \leq \tau \} - \hat{\mu}_T)^2}},
\]

where \(\mathcal{N} = \{1 \leq i \leq n : |Y_i| \leq \tau\}\). Observe that

\[
S_{\tau,n} = \frac{\sum_{i=1}^{n} (f(Y_i) - \mu)}{\sqrt{\sum_{i=1}^{n} (f(Y_i) - \mu + \mu - \hat{\mu}_W)^2}} = \frac{S_{\tau,n}^*}{\sqrt{1 - \frac{\sum_{i=1}^{n} (f(Y_i) - \mu)^2}{n (\sum_{i=1}^{n} (f(Y_i) - \mu))^2}}},
\]

where \(S_{\tau,n}^*\) is the self-normalized winsorized mean defined as

\[
S_{\tau,n}^* = \frac{\sum_{i=1}^{n} (f(Y_i) - \mu)}{\sqrt{\sum_{i=1}^{n} (f(Y_i) - \mu)^2}}
\]

Similarly, we have for the studentized trimmed mean that

\[
U_{\tau,n} = \frac{U_{\tau,n}^*}{\sqrt{1 - \frac{1}{c_{\tau,n}} (U_{\tau,n}^*)^2}},
\]

where \(U_{\tau,n}^*\) is the self-normalized trimmed mean defined as

\[
U_{\tau,n}^* = \frac{\sum_{i \in \mathcal{N}} (Y_i 1 \{ |Y_i| \leq \tau \} - \mu)}{\sqrt{\sum_{i \in \mathcal{N}} (Y_i 1 \{ |Y_i| \leq \tau \} - \mu)^2}}
\]

and \(c_{\tau,n} = \sum_{i=1}^{n} 1 \{ |Y_i| \leq \tau\}\). Since the function \(x/(1 - \frac{1}{n} x^2)^{1/2}\) is an increasing function for \(0 < x < n^{1/2}\), we have

\[
P(S_{\tau,n} > x) = P\left( S_{\tau,n}^* > \frac{x}{\sqrt{1 + \frac{x^2}{n}}} \right)
\]

and

\[
P(U_{\tau,n} > x) = P\left( U_{\tau,n}^* > \frac{x}{\sqrt{1 + \frac{x^2}{c_{\tau,n}}}} \right)
\]
Therefore, to investigate the limiting properties of \( S_{\tau,n}^* \) and \( U_{\tau,n}^* \) is equivalent to investigate that for the simpler self-normalized statistics \( S_{\tau,n}^* \) and \( U_{\tau,n}^* \), respectively.

Before stating our Cramér type moderate deviation results, let us first present how the self-normalized winsorized mean and trimmed mean connect with the general self-normalized sum investigated in our main Theorem 2.1. First for the self-normalized winsorized mean, though (4.8) presents the form of a self-normalized sum of independent random variables sum investigated in our main Theorem 2.1. First for the self-normalized winsorized mean, though (4.8) presents the form of a self-normalized sum of independent random variables for \( S_{\tau,n}^* \), the expectation of \( f(Y) - \mu \) is slightly deviated from 0 and needs to be calibrated. Denote

\[
\mu = \mathbb{E}f(Y), \quad \sigma_1^2 = \mathbb{E}(f(Y) - \mu)^2, \quad \sigma_2^2 = \mathbb{E}(f(Y) - \mu)^2.
\]

Then we can write

\[
S_{\tau,n}^* = \frac{\sum_{i=1}^{n} (f(Y_i) - \tilde{\mu}) - \sqrt{n}(\mu - \tilde{\mu})}{\sqrt{\sum_{i=1}^{n} (f(Y_i) - \mu)^2}} \cdot \frac{\sigma_1}{\sigma_2}.
\]

(4.10)

where \( S_n, V_n \) and \( c \) are denoted by

\[
S_n = \sum_{i=1}^{n} \frac{f(Y_i) - \tilde{\mu}}{\sqrt{n}\sigma_1}, \quad V_n^2 = \sum_{i=1}^{n} \frac{(f(Y_i) - \mu)^2}{n\sigma_2^2}
\]

and \( c = \frac{\sqrt{n}(\mu - \tilde{\mu})}{\sigma_1} \).

Therefore,

\[
\mathbb{P}(S_{\tau,n}^* > x) = \mathbb{P}\left( \frac{S_n - c}{V_n} > \frac{\sigma_2}{\sigma_1} x \right).
\]

Note that the random variables involved in \( S_n \) and \( V_n \) are different, which means the existing results for classical self-normalized sums cannot be directly applied.

As for the self-normalized trimmed mean, note that in the numerator, \( \sum_{i \in \mathcal{N}} (Y_i \mathbb{1}\{|Y_i| \leq \tau\} - \mu) \) in (4.9) is equal to \( \sum_{i=1}^{n} (Y_i - \mu) \mathbb{1}\{|Y_i| \leq \tau\} \). Similarly in the denominator, \( \sum_{i \in \mathcal{N}} (Y_i \mathbb{1}\{|Y_i| \leq \tau\})^2 \) is equal to \( \sum_{i=1}^{n} (Y_i - \mu)^2 \mathbb{1}\{|Y_i| \leq \tau\}^2 \). Thus we have

\[
U_{\tau,n}^* = \frac{\sum_{i=1}^{n} (Y_i - \mu) \mathbb{1}\{|Y_i| \leq \tau\}}{\sqrt{\sum_{i=1}^{n} (Y_i - \mu)^2 \mathbb{1}\{|Y_i| \leq \tau\}^2}}
\]

Denote

\[
\mu_0 = \mathbb{E}[(Y - \mu) \mathbb{1}\{|Y| \leq \tau\}], \quad \sigma_3^2 = \mathbb{E}[(Y - \mu)^2 \mathbb{1}\{|Y| \leq \tau\} - \mu_0]^2,
\]

and \( \sigma_4^2 = \mathbb{E}[(Y - \mu)^2 \mathbb{1}\{|Y| \leq \tau\}] \).

Similar to the self-normalized winsorized mean, we can obtain

\[
\mathbb{P}(U_{\tau,n}^* > x) = \mathbb{P}\left( \frac{S_n^o - \delta}{V_n^o} > \frac{\sigma_4}{\sigma_3} x \right),
\]

where

\[
S_n^o = \sum_{i=1}^{n} \frac{Y_i \mathbb{1}\{|Y_i| \leq \tau\} - \mu_0}{\sqrt{n}\sigma_3}, \quad (V_n^o)^2 = \sum_{i=1}^{n} \frac{(Y_i - \mu) \mathbb{1}\{|Y_i| \leq \tau\}^2}{n\sigma_4^2}
\]

and \( \delta = \frac{\sqrt{n}\mu_0}{\sigma_3} \).
Consequently, our result for general self-normalized sums in Theorem 2.1 can be directly applied to $S^*_{\tau,n}$ and $U^*_{\tau,n}$ to derive the following bias-corrected Cramér-type moderate deviation theorems for the self-normalized winsorized mean and trimmed mean under the fourth moment.

**Theorem 4.1.** Assume $\mathbb{E}[Y^4] < \infty$. Then there exist an absolute positive constant $c_1$ and positive constants $c_2$ and $A$ depending on $\sigma$, $\mathbb{E}[|Y|^3]$ and $\mathbb{E}[Y^4]$, such that for
\begin{equation}
\tau \geq c_1 n^{1/6} \max \{ (\mathbb{E}[Y^4])^{1/2}/\sigma, (\mathbb{E}[Y^4]/\sigma)^{1/3} \},
\end{equation}
it holds that
\begin{equation}
\mathbb{P}(S^*_{\tau,n} > x) = [1 - \Phi(x)] \exp \left\{ -\frac{x^3 \mathbb{E}(Y - \mu)^3}{3 \sqrt{n} \sigma^3} \right\} \\
\times \left[ 1 + O_1 \left( \frac{(1 + x^4)}{n} + \frac{(1 + x) \sqrt{n}}{\tau^3} + \frac{(1 + x)}{\sqrt{n}} \right) \right]
\end{equation}
uniformly for $x \in (0, c_2 \min\{n^{1/4}, \tau^3 n^{-1/2}\})$, where $O_1$ is a bounded quantity satisfying $|O_1| \leq A$. Similar result holds for $\mathbb{P}(S^*_{\tau,n} < -x)$.

**Theorem 4.2.** Under the conditions of Theorem 4.1, the same result as (4.12) holds for $U^*_{\tau,n}$.

Observe that under the fourth moment, the general framework Theorem 2.1 enables us to pin down the bias-corrected term $\exp\{ -\frac{x^3 \mathbb{E}(Y - \mu)^3}{3 \sqrt{n} \sigma^3} \}$ which depends on the skewness of the underlying distribution. After correcting this skewness in normal approximation, the convergence rate and the converging range significantly improve that given in Theorem 4.3, where only third moment is assumed.

The choice of $\tau$ should be determined by taking both convergence rate and robustness of estimator into account. We observe from (4.12) that the ratio $\mathbb{P}(S^*_{\tau,n} > x)/([1 - \Phi(x)] \exp\{ -\frac{x^3 \mathbb{E}(Y - \mu)^3}{3 \sqrt{n} \sigma^3} \})$ converges to 1 for $x \in (0, o(\min\{n^{1/4}, \tau^3 n^{-1/2}\}))$. The widest possible range $x \in (0, o(n^{1/4}))$ can be achieved by choosing $\tau \geq O(n^{1/4})$. When $\tau \leq O(n^{1/3})$, the larger $\tau$ is, the faster rate of convergence and wider range of $x$ can be obtained. Yet, once $\tau$ exceeds $O(n^{1/3})$, our result reduces to the bias-corrected Cramér-type moderate deviation for the classical self-normalized sample mean (see Theorem 1.1 in Wang (2011)), which is reasonable because the winsorized mean and trimmed mean are asymptotically equivalent to the sample mean as $\tau \to \infty$. It is worth mentioning that when $O(n^{1/6}) \leq \tau \leq O(n^{1/3})$, the ratio $\mathbb{P}(S^*_{\tau,n} > x)/([1 - \Phi(x)] \exp\{ -\frac{x^3 \mathbb{E}(Y - \mu)^3}{3 \sqrt{n} \sigma^3} \})$ converges to 1 at the rate of $O((1 + x)^4 n^{-1} + (1 + x) \sqrt{n} \tau^{-3})$ uniformly for $x \in o(\min\{n^{1/4}, \tau^3 n^{-1/2}\})$.

In this regime of $\tau$, though the convergence rate of winsorized mean and trimmed mean could be slightly slower than that of the classical self-normalized sample mean and the ranges of $x$ for convergence could be narrower, the winsorized mean and trimmed mean provide robust inference. We will provide the proof of Theorem 4.1 in Section A.11 and the proof of Theorem 4.2 in Section A.12.

Theorem 2.3 in Zhou et al. (2018) is closely related to ours. They established a Cramér-type moderate deviation result for Huber estimator $\hat{\mu}_H$ defined in (4.5) by using a Bahadur representation for the Huber estimator. Theorem 2.1 in their paper reveals that $\hat{\mu}_H - \mu$ is asymptotically close to $n^{-1} \sum_{i=1}^n f(Y_i - \mu)$, where $f(\cdot)$ is defined by (4.2). Therefore, it is easy to see that the Huber estimator $\hat{\mu}_H$ is close to the winsorized mean $\hat{\mu}_W = n^{-1} \sum_{i=1}^n f(Y_i)$ as $\tau \to \infty$. Theorem 2.3 in Zhou et al. (2018) for the Huber estimator can be restated as follows. The notation $a_n \ll b_n$ means $a_n = o(b_n)$ as $n \to \infty$.
Remark 4.1. Assume $\mathbb{E}[Y^3] < \infty$. Zhou et al. (2018) proved for $n^{1/4} \ll \tau \ll n^{1/2}$ that

$$
\frac{\mathbb{P}(\sqrt{n} \sigma^{-1} |\hat{\mu}_H - \mu| > x)}{2(1 - \Phi(x))} = 1 + O(1) \left\{ \left( \frac{\log n + x}{\sqrt{n}} \right)^3 + \frac{1 + x}{n^{3/10}} + \frac{(1 + x) \sqrt{n}}{\tau^2} + e^{-O(n^{1/10})} \right\}
$$

uniformly for $0 \leq x = o(\min\{\sqrt{n}/\tau, \tau^2/\sqrt{n}\})$.

Compared to their condition $n^{1/4} \ll \tau \ll n^{1/2}$, our condition $\tau \geq O(n^{1/6})$ is less restrictive. Moreover, when $\tau \gg n^{1/4}$, both of our convergence rate and the associated converging range of $x$ improve theirs. Our improvement mainly relies on the explicit formula of the self-normalized winsorized mean presented in (4.10) and our fundamental result for general self-normalized sum established in Theorem 2.1. In addition, since the higher moment $\mathbb{E}[Y^4] < \infty$ is assumed, after correcting the bias in normal approximation, the convergence rate and the associated range of $x$ could be significantly improved.

The common downside of our bias-corrected result in Theorem 4.1 and the normal approximation for Huber estimator by Zhou et al. (2018) in Remark 4.1 is that the limiting distributions depend on unknown parameters. In real-world applications, if reliable estimations for the unknown parameters are unavailable, we can directly use normal approximation for the self-normalized winsorized mean presented in the following theorem, where only third moment is required and the limiting distribution does not depend on any unknown parameters.

Theorem 4.3. Assume $\mathbb{E}[Y^3] < \infty$. Then there exist absolute positive constants $c_1, c_2$ and $A$ such that for

$$\tau \geq c_1 n^{1/4} \max\{\mathbb{E}[Y^3]/\sigma^2, (\mathbb{E}[Y^3]/\sigma)^{1/2}\},$$

it holds that

$$
\mathbb{P}(S_{\tau,n}^* > x) = 1 - \Phi(x) = 1 + \frac{1 + x^3 \mathbb{E}[Y^3]}{\sigma^3 \sqrt{n}} + \frac{(1 + x) \sqrt{n} \mathbb{E}[Y^3]}{\ln n} \left( 1 + O_1\left( \frac{\sqrt{n} \mathbb{E}[Y^3]}{\ln n} \right) \right),
$$

uniformly for $x \in (0, c_2 \min\{n^{1/6} \sigma^3/\mathbb{E}[Y^3], \tau^2 \mathbb{E}[Y^3]/(\sqrt{n} \mathbb{E}[Y^3])\})$, where $O_1$ is a bounded quantity satisfying $|O_1| \leq A$. Similar result holds for $\mathbb{P}(S_{\tau,n}^* < -x)$.

Theorem 4.4. Under the conditions of Theorem 4.3, the same result as (4.15) holds for $U_{\tau,n}^*$

It can be observed that the convergence rate and the converging range of $x$ also outperform the results of Zhou et al. (2018) shown in Remark 4.1, and our condition on $\tau$ is less restrictive. We relegate the proof of Theorem 4.3 to Section A.13 and the proof of Theorem 4.4 in Section A.14 in the Supplementary Material.

4.2. Simultaneous confidence intervals. Cramér type moderate deviation results are useful in providing theoretical guarantees for a wide spectrum of statistical applications, including the multiple testing procedure and multiple confidence intervals for ultra-high dimensional parameters. For an illustrative example, we will construct simultaneous confidence intervals for the means under the following ultra-high mean model by using the studentized winsorized mean estimator defined in (4.6). We consider

$$Z_i = \mu + \epsilon_i, \quad i = 1, \ldots, n,$$
where \( \{Z_1, \ldots, Z_n\} \) are i.i.d. observations, \( \mu = (\mu_1, \ldots, \mu_p)^T \in \mathbb{R}^p \), and \( \{e_1, \ldots, e_i\} \) are i.i.d. errors. Denote \( \Sigma = \text{Cov}(\epsilon_i) := (\Sigma_{ij})_{p \times p} \). Assume there exist constants \( C_1, C_2 \) such that \( \max_{1 \leq j \leq p} \mathbb{E}|Z_{ij}|^3 \leq C_1 \) and \( \min_{1 \leq j \leq p} \Sigma_{jj} \geq C_2 \).

**Theorem 4.5.** Assume the dimensionality \( p \), the significance level \( \alpha \) and the thresholding parameter \( \tau \) satisfying \( \log(p/\alpha) = o(n^{1/3}) \) and \( \tau \gg n^{1/3} \). Then for \( \alpha \in (0, 1) \), and \( t_0 \) satisfying the equation

\[
\frac{t_0}{1 + t_0^2/n} = \Phi^{-1}(1 - \frac{\alpha}{2p}),
\]

we have

\[
\sum_{i=1}^n f(Z_{ij}) = t_0 \frac{t_0}{n} + \frac{1}{n} \sum_{i=1}^n (f(Z_{ij}))^2 - \frac{1}{n} \left[ \sum_{i=1}^n f(Z_{ij}) \right]^2 \triangleq (L_j, U_j), \quad 1 \leq j \leq p
\]

are the \( 1 - \alpha - o(1) \) simultaneous confidence intervals for \( (\mu_j)_{j=1}^p \), where \( f(\cdot) \) is the function defined in (4.2).

The proof of Theorem 4.5 will be provided in Section A.15 in the Supplementary Material.

**5. Proofs.** In this section, we present proofs of Theorem 2.1, Theorems 3.1–3.3 and Theorems 4.1. Throughout the rest of this section, \( A \) and \( C \) denote positive absolute constants that may take different values at each appearance.

**5.1. Proof of Theorem 2.1.** We prove the theorem for the two scenarios \( 0 < x \leq 3 \) and \( x > 3 \), respectively. First, we prove it for \( 0 < x \leq 3 \). For this range, it is sufficient to prove a Berry-Esseen bound as the following proposition will show. The proof of Proposition 5.1 is postponed to Section A.1 the Supplementary Material.

**Proposition 5.1.** For \( 0 < x \leq 3 \), there exists an absolute constant \( A > 0 \) such that

\[
|\mathbb{P}(S_n > xV_n + c) - [1 - \Phi(x + c)]| \leq AL_{3,n}.
\]

Note that \( 1 - \Phi(3.6) \leq 1 - \Phi(x + c) \leq 1 \) for \( 0 < x \leq 3 \) and \( |c| \leq x/5 \). Thus, it follows from Proposition 5.1 that for \( 0 < x \leq 3 \),

\[
\mathbb{P}(S_n > xV_n + c) = [1 - \Phi(x + c)](1 + O(1 + x)L_{3,n}).
\]

Moreover, it holds for \( 0 < x \leq 3 \) satisfying (2.6) that

\[
|(\Psi_x^*)^{-1} - 1| \leq Ax^3L_{3,n} \leq AL_{3,n},
\]

which combining with (5.2) entails that

\[
\mathbb{P}(S_n > xV_n + c) = [1 - \Phi(x + c)]\Psi_x^*(1 + O(1 + x)L_{3,n}).
\]

Consequently, we have

\[
\mathbb{P}(S_n > xV_n + c) = [1 - \Phi(x + c)]\Psi_x^*e^{O_1R_x}(1 + O(1 + x)L_{3,n}),
\]

where the quantity \( |O_1| \leq A \) for some absolute constant. This completes the proof for \( 0 < x \leq 3 \).

Next we deal with the case \( x > 3 \). By applying the elementary inequality

\[
1 + s/2 - s^2 \leq (1 + s)^{1/2} \leq 1 + s/2,
\]
for \( s = V_n^2 - 1 \), we obtain
\[
\frac{1}{2}(V_n^2 + 1) - (V_n^2 - 1)^2 \leq V_n \leq \frac{1}{2}(V_n^2 + 1).
\]

Therefore, plugging in the above upper and lower bounds yields
\[
\mathbb{P}(S_n > xV_n + c) \geq \mathbb{P}(2xS_n - x^2V_n^2 \geq x^2 + 2xc)
\]
and
\[
\mathbb{P}(S_n > xV_n + c) \leq \mathbb{P}(2xS_n - x^2V_n^2 \geq x^2 + 2xc - x\Delta_n)
+ \mathbb{P}\left(S_n > xV_n + c, |V_n^2 - 1| > x^{-1}(1 \vee 6R_x^{1/2})\right),
\]
where \( \Delta_n = \min\{2x(V_n^2 - 1)^2, x^{-1}(2\vee 72R_x)\} \) and the notation \( a \vee b \) means the maximum of \( a \) and \( b \). The upper bound holds because
\[
\mathbb{P}\left(S_n > xV_n + c, |V_n^2 - 1| \leq x^{-1}(1 \vee 6R_x^{1/2})\right)
\leq \mathbb{P}\left(2xS_n - x^2V_n^2 \geq x^2 + 2xc - 2x^2(V_n^2 - 1)^2, |V_n^2 - 1| \leq x^{-1}(1 \vee 6R_x^{1/2})\right)
\leq \mathbb{P}(2xS_n - x^2V_n^2 \geq x^2 + 2xc - x\Delta_n).
\]

The following Propositions 5.2–5.4 draw an outline of the proof for the case \( x > 3 \). Their proofs are relegated to Sections A.2–A.4 in the Supplementary Material.

**Proposition 5.2.** There exists an absolute constant \( A \) such that
\[
\mathbb{P}(2xS_n - x^2V_n^2 \geq x^2 + 2xc) = [1 - \Phi(x+c)] \Psi_x e^{O_1 R_x} \{1 + O_2(1 + x)L_{3,n}\},
\]
for \( x > 3 \) satisfying (2.6) and (2.7) and \( |c| < x/5 \), where \( |O_1| \leq A \) and \( |O_2| \leq A \).

**Proposition 5.3.** There exist absolute constants \( A_1 \) and \( A_2 \) such that
\[
\mathbb{P}(2xS_n - x^2V_n^2 \geq x^2 + 2xc - x\Delta_n) \leq [1 - \Phi(x+c)] \Psi_x e^{A_1 R_x} \{1 + A_2(1 + x)L_{3,n}\},
\]
for \( x > 3 \) satisfying (2.6) and (2.7), and \( |c| < x/5 \).

**Proposition 5.4.** There exist absolute constants \( A_1 \) and \( A_2 \) such that
\[
\mathbb{P}\left(S_n \geq xV_n + c, |V_n^2 - 1| > x^{-1}(1 \vee 6R_x^{1/2})\right)
\leq A_1 R_x[1 - \Phi(x+c)] \Psi_x e^{A_2 R_x},
\]
for \( x > 3 \) satisfying (2.6) and (2.7), and \( |c| < x/5 \).

We obtain by substituting the results in Propositions 5.3–5.4 into (5.5) that
\[
\mathbb{P}(S_n > xV_n + c)
\leq [1 - \Phi(x+c)] \Psi_x e^{A_1 R_x} \{1 + A_1 R_x + A_2(1 + x)L_{3,n}\}
\leq [1 - \Phi(x+c)] \Psi_x e^{A R_x} \{1 + A(1 + x)L_{3,n}\}
\]
which together with the result in Proposition 5.2 yields the desired result (2.5) for \( x > 3 \). The proof is completed.
5.2. Proof of Theorem 3.1. The main idea is to apply the big-block-small-block technique to construct a general self-normalized sum based on an independent sequence to which our main result Theorem 2.1 can be applied. Denote $B_n = \sum_{i=1}^{n} \xi_i$. We first apply Berry-Esseen bound for sum of one-dependent random variables to cope with the case $0 \leq x \leq O(\sqrt{\log n})$. Note that

$$P(S_n \geq x V_n) - \left[1 - \Phi\left(\frac{x}{\sqrt{1+2\rho_n}}\right)\right]$$

$$\leq \left|P\left(S_n \geq x B_n (1 - n^{-1/3})^{1/2}\right) - \left[1 - \Phi\left(\frac{x}{\sqrt{1+2\rho_n}}\right)\right]\right|$$

$$+ \left[P\left(S_n \geq x B_n (1 + n^{-1/3})^{1/2}\right) - \left[1 - \Phi\left(\frac{x}{\sqrt{1+2\rho_n}}\right)\right]\right]$$

$$+ P\left(|V_n^2 - B_n^2| > n^{-1/3} B_n^2\right)$$

$$:= E_1 + E_2 + E_3.$$}

Recalling the definition of $\rho_n$ in (3.1), we have $\text{Var}(S_n) = (1 + 2\rho_n)B_n^2$. By noticing the assumptions

$$E\xi^4_i \leq a_1^4, \quad E\xi^2_i \geq a_2^2, \quad a = a_1/a_2, \quad -1/2 < \rho \leq \rho_n \leq 1/2$$

and applying the Berry-Esseen bound for sums of one-dependent random variables (see Shergin (1980)), we obtain

$$E_1 \leq A(\rho) a^3 \sqrt{n} + \Phi\left(\frac{x}{\sqrt{1+2\rho_n}}\right) - \Phi\left(\frac{x(1 - n^{-1/3})^{1/2}}{\sqrt{1+2\rho_n}}\right)$$

$$\leq A(\rho) a^3 (n^{-1/2} + x^2 n^{-1/3}) \leq A(\rho) a^3 (1 + x)^2 n^{-1/3},$$

where $A(\rho)$ is a positive constant depending on $\rho$ and may take different values at each appearance. In the same manner, this above bound applies to $E_2$ as well. As for $E_3$, it follows by Chebyshev’s inequality that

$$E_3 \leq (n^{1/3} B_n^{-2})^2 E[(V_n^2 - B_n^2)^2] \leq A a^4 n^{-1/3}.$$}

Therefore,

$$P(S_n \geq x V_n) - \left[1 - \Phi\left(\frac{x}{\sqrt{1+2\rho_n}}\right)\right] \leq A(\rho) a^4 (1 + x)^2 n^{-1/3},$$

Moreover, observe that $1 - \Phi\left(\frac{x}{\sqrt{1+2\rho_n}}\right) \geq 1 - \Phi\left(\frac{2\sqrt{3}}{\sqrt{1+2\rho}}\right)$ for $0 \leq x \leq 2\sqrt{3}$ and

$$1 - \Phi\left(\frac{x}{\sqrt{1+2\rho_n}}\right) \geq A(\rho) x^{-1} \exp\left\{-\frac{x^2}{2(1+2\rho)}\right\}$$

for $x > 2\sqrt{3}$. As a consequence, there exists a constant $c_{\rho}$ depending on $\rho$ such that (3.2) holds for $0 \leq x \leq c_{\rho} \sqrt{\log n}.

Next we turn to the proof for $x > c_{\rho} \sqrt{\log n}$. Let the length of big blocks be $l = \lfloor n^{\alpha}\rfloor$, and each small block contains only one random variable. Denote $k = \lfloor n/(l+1)\rfloor$. Without loss of generality, let us assume $n/(l+1)$ to be an integer, then the sequence of $\{\xi_i\}_{1 \leq i \leq n}$ can be divided into $k$ big blocks and $k$ small blocks. Observe that if $n/(l+1)$ is not an integer, the sequence of $\{\xi_i\}_{1 \leq i \leq n}$ will be divided into $k + 1$ big blocks and $k$ small blocks, in which the first $k$ big blocks are of size $l$ and the last one is of size $n - (l+1)\lfloor n/(l+1)\rfloor$. Although the
size of the last big block might be different, our analysis also applies under this scenario. For \(1 \leq j \leq k\), the \(j\)-th big block and the corresponding block sums are given by

\[
H_j = \{i: (j-1)(l+1) + 1 \leq i \leq j(l+1) - 1\} \quad \text{and} \quad X_j = \sum_{i \in H_j} \xi_i, \quad Y_j^2 = \sum_{i \in H_j} \xi_i^2,
\]

Moreover, we denote

\[
S_{n1} = \sum_{j=1}^{k} X_j, \quad V_{n1}^2 = \sum_{j=1}^{k} Y_j^2, \quad B_{n1}^2 = \mathbb{E}V_{n1}^2, \tag{5.11}
\]

\[
S_{n2} = \sum_{j=1}^{k} \xi_j(l+1), \quad V_{n2}^2 = \sum_{j=1}^{k} \xi_j^2(l+1), \quad B_{n2}^2 = \mathbb{E}V_{n2}^2. \tag{5.12}
\]

Observe that \(\{X_j\}_{1 \leq j \leq k}\) and \(\{\xi_j(l+1)\}_{1 \leq j \leq k}\) both consist of independent random variables. As \(S_{n1}\) is the sum of \(\xi_i\)'s in big blocks and it is the main part of \(\sum_{i=1}^{n} \xi_i\), while \(S_{n2}\) corresponds to the small blocks. The big-block-small block technique splits the sum \(\sum_{i=1}^{n} \xi_i\) into two parts \(S_{n1}\) and \(S_{n2}\), each of which is a sum of independent random variables. Let \(\tau = B_{n2}/x\) and we do truncations \(\hat{\xi}_i = \xi_i 1(|\xi_i| \leq \tau)\) only for the \(\xi_i\)'s in small blocks, that is, \(i = j(l+1)\) for \(1 \leq j \leq k\), so

\[
\mathbb{P}(S_n \geq xV_n) \leq \mathbb{P}(\hat{S}_n \geq x\hat{V}_n) + \mathbb{P}(S_n \geq xV_n, \max_{1 \leq j \leq k} |\xi_j(l+1)| > \tau), \tag{5.13}
\]

\[
\mathbb{P}(S_n \geq xV_n) \geq \mathbb{P}(\hat{S}_n \geq x\hat{V}_n) - \mathbb{P}(S_n \geq x\hat{V}_n, \max_{1 \leq j \leq k} |\xi_j(l+1)| > \tau), \tag{5.14}
\]

where \(\hat{S}_n = S_{n1} + \hat{S}_{n2}\), \(\hat{V}_n = V_{n1} + \hat{V}_{n2}\) with \(\hat{S}_{n2} = \sum_{j=1}^{k} \hat{\xi}_j(l+1)\) and \(\hat{V}_{n2} = \sum_{j=1}^{k} \hat{\xi}_j^2(l+1)\). For a positive number \(d_1 > 0\), we have the upper bound

\[
\mathbb{P}(\hat{S}_n \geq x\hat{V}_n) \leq \mathbb{P}(S_{n1} \geq xV_{n1} - d_1n^{-\frac{\alpha}{2}}xB_n) + \mathbb{P}(\hat{S}_{n2} \geq d_1n^{-\frac{\alpha}{2}}xB_n) \tag{5.15}
\]

and the lower bound

\[
\mathbb{P}(\hat{S}_n \geq x\hat{V}_n) \geq \mathbb{P}(S_{n1} \geq xV_{n1} + d_1n^{-\frac{\alpha}{2}}xB_n) - \mathbb{P}(S_{n1} \geq xV_{n1}, V_{n1}^2 < B_{n1}^2/4) - \mathbb{P}(\hat{S}_{n2} < -d_1n^{-\frac{\alpha}{2}}xB_n + x(\hat{V}_n - V_{n1}), V_{n2}^2 > B_{n2}^2/4). \tag{5.16}
\]

We can obtain the following bounds for the terms involved in (5.15) and (5.16). The proofs of Propositions 5.5 and 5.6 are given in Sections A.5 and A.6 in the Supplementary Material.

**Proposition 5.5.** There exists an absolute positive constant \(d_0\) and a constant \(A(\rho, d_0)\) depending on \(\rho\) and \(d_0\) such that, for \(d_1 = \kappa_\rho d_0^2\) with some sufficiently large constant \(\kappa_\rho\),

\[
\mathbb{P}\left(S_{n1} \geq xV_{n1} + d_1n^{-\alpha/2}xB_n\right) = \left[1 - \Phi\left(x\sqrt{\frac{1}{1+2\rho_{n1}}}\right)\right]\left(1 + O_1\left(\frac{a^4x^4}{n^{1-\alpha}} + \frac{a^2x^2}{n^{\alpha/2}} + \frac{a^3x}{n^{\frac{\alpha}{2}}}\right)\right), \tag{5.17}
\]

uniformly for \(x \in (2, d_0a^{-1}\min\{n^{\alpha/4}, n^{(1-\alpha)/4}\})\), where \(|O_1| \leq A(\rho, d_0)\). A similar result holds for \(\mathbb{P}(S_{n1} \geq xV_{n1} - d_1n^{-\alpha/2}xB_n)\). Moreover,

\[
\mathbb{P}\left(S_{n1} \geq xV_{n1}, V_{n1}^2 < \frac{B_{n1}^2}{4}\right) \tag{5.18}
\]
\[
\leq \frac{A a^4 x^4}{n^{1-\alpha}} \left[ 1 - \Phi \left( \frac{x}{\sqrt{1 + 2\rho n}} \right) \right] \exp \left\{ \frac{A a^4 x^4}{n^{1-\alpha}} + A a^3 \frac{x^3}{\sqrt{n}} \right\},
\]
uniformly for \( x \in (2, d_0 a^{-1} n^{\frac{1}{2} (1-\alpha)}) \).

**Proposition 5.6.** For \( d_1 = \kappa_\rho a^2 \) with \( \kappa_\rho \geq 10 \), we have

\[
\mathbb{P}(\hat{S}_{n_2} > d_1 n^{-\alpha/2} x B_n) \leq \exp\{-\kappa_\rho x^2 / 6\},
\]
(5.19) \[
\mathbb{P}(\hat{S}_{n_2} < -d_1 n^{-\alpha/2} x B_n + x(\hat{V}_n - V_{n_1}), V_{n_1}^2 > B_n^2 / 4) \leq \exp\{-\kappa_\rho x^2 / 14\}.
\]
(5.20)

Note that for \( x > c_\rho \sqrt{\log n} \) and sufficiently large \( \kappa_\rho \),

\[
\exp\{-\kappa_\rho x^2 / 6\} \leq \exp\{-\kappa_\rho x^2 / 12\} \leq n^{-1/4}.
\]
(5.21)

By choosing \( \alpha = 1/2 \) and combining (5.17)-(5.21), we obtain

\[
\mathbb{P}\left(\hat{S}_n \geq x \hat{V}_n, \max_{1 \leq j \leq k} |\xi_{j(l+1)}| > \tau\right) \leq A(\rho, d_0) a^4 \frac{(1 + x)^4}{n^{1/2}} \left[ 1 - \Phi \left( \frac{x}{\sqrt{1 + 2\rho n}} \right) \right]
\]
(5.22)

uniformly for \( x \in (c_\rho \sqrt{\log n}, d_0 a^{-1} n^{1/8}) \).

In addition, we can bound the error terms in (5.13) and (5.14) as follows. The proof of Proposition 5.7 is shown in Section A.7 in the Supplementary Material.

**Proposition 5.7.** Under the conditions in Theorem 3.1, we have for \( x \in (c_\rho \sqrt{\log n}, d_0 a^{-1} n^{1/8}) \) that

\[
\mathbb{P}\left(\hat{S}_n \geq x \hat{V}_n, \max_{1 \leq j \leq k} |\xi_{j(l+1)}| > \tau\right) \leq A(\rho, d_0) a^4 \frac{(1 + x)^4}{n^{1/2}} \left[ 1 - \Phi \left( \frac{x}{\sqrt{1 + 2\rho n}} \right) \right]
\]
(5.23)

and

\[
\mathbb{P}\left(S_n \geq x V_n, \max_{1 \leq j \leq k} |\xi_{j(l+1)}| > \tau\right) \leq A(\rho, d_0) a^4 \frac{(1 + x)^4}{n^{1/2}} \left[ 1 - \Phi \left( \frac{x}{\sqrt{1 + 2\rho n}} \right) \right].
\]
(5.24)

Consequently, substituting (5.22)–(5.24) into (5.13) and (5.14) yields the desired result (3.2). This completes the proof of Theorem 3.1.

**5.3. Proof of Theorem 3.2.** The proof for Theorem 3.2 again builds on the big-block-small-block technique, and also exploits a lemma in Shao and Yu (1996) to replace the weakly dependent big blocks and small blocks by independent random variables, respectively. We begin the proof by introducing three essential lemmas from the literature. Lemma 5.1 (Theorem 4.1 of Shao and Yu (1996)) and Lemma 5.2 (Theorem 10.1.b of Lin and Bai (2010)) concern the bound of moments under weak dependence while Lemma 5.3 (Lemma 2.1 of Berbee (1987)) shows that a \( \beta \)-mixing sequence of random variables can be replaced by an independent sequence of random variables in a domain whose measure is at least \( 1 - \sum_{i=1}^{n} \beta^{(i)} \).
LEMMA 5.1. (Theorem 4.1 in Shao and Yu (1996).) Let \( \{X_i, i \geq 1\} \) be a sequence of zero-mean random variables with \( \mathbb{E}|X_i|^r \leq \mu^r \) for \( r > 2 \) and \( \mu > 0 \). Assume that mixing condition (3.6) holds, then
\[
\mathbb{E}
\left[
\left|
\sum_{i=k}^{i=k+m} X_i
\right|^{r'}
\right]
\leq C m^{r'/2} \mu^{r'},
\]
for any \( 2 \leq r' < r, m \geq 1 \) and \( k \geq 0 \), where \( C \) is a constant that depends on \( r', r, a_1, a_2 \) and \( \tau \).

LEMMA 5.2. (Theorem 10.1.b of Lin and Bai (2010).) Assume \( \{X_i\}_{i \geq 1} \) is a sequence of random variables and \( \beta(n) \) is the \( \beta \)-mixing coefficient defined in (3.5). Denote by \( \sigma_1^k \) and \( \sigma_{k+n}^\infty \) the \( \sigma \)-fields generated by \( \{X_i\}_{1 \leq i \leq k} \) and \( \{X_i\}_{i \geq k+n} \), respectively. For \( X \in L_p(\sigma_1^k) \) and \( Y \in L_q(\sigma_{k+n}^\infty) \) with \( p, q, r \geq 1 \) and \( \frac{1}{p} + \frac{1}{q} + \frac{1}{r} = 1 \), we have
\[
\|\mathbb{E}XY - \mathbb{E}X\mathbb{E}Y\| \leq 8\beta(n)^{1/r}\|X\|_p\|Y\|_q.
\]

For two random variables (or vectors) \( X \) and \( Y \), define
\[
\beta(X, Y) = \frac{1}{2} \sup_A \left( (\mathbb{P}_{X,Y} - \mathbb{P}_X \times \mathbb{P}_Y)(A) - (\mathbb{P}_{X,Y} - \mathbb{P}_X \times \mathbb{P}_Y)(A^c) \right),
\]

LEMMA 5.3. (Lemma 2.1 of Berbee (1987).) Let \( \{\xi_i, 1 \leq i \leq n\} \) 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 \( \xi_i \) distributed as \( \xi_i \) such that \( \{\xi_i\}_{1 \leq i \leq n} \) are independent and
\[
\mathbb{P}(\xi_i \neq \xi_i, \text{for some } 1 \leq i \leq n) \leq \beta^{(1)} + \cdots + \beta^{(n-1)}.
\]

Recall the definition of block sums \( \{Y_j\}_{1 \leq j \leq k} \) in (3.7). We set the size of big blocks as \( m_1 = [n^\alpha] \) for some \( 0 < \alpha < 1 - \alpha \) and the size of small blocks as \( 1 \). Denote \( k_1 = k/(m_1 + 1) \), where \( k = [n/l] \). For simplicity of presentation, we assume \( k/(m_1 + 1) \) to be an integer, as explained in the proof of Theorem 3.1. The \( u \)-th big block is given by
\[
I_u = \left\{ j : (m_1 + 1)(u - 1) + 1 \leq j \leq (m_1 + 1)u - 1 \right\}, \quad \text{for } 1 \leq u \leq k_1.
\]

Define
\[
S_k = \sum_{j=1}^{k} Y_j, \quad V_k^2 = \sum_{j=1}^{k} Y_j^2, \quad \xi_u = \sum_{j \in I_u} Y_j, \quad \eta_u^2 = \sum_{j \in I_u} Y_j^2,
\]
\[
S_{k,1} = \sum_{u=1}^{k_1} \xi_u, \quad V_{k,1}^2 = \sum_{u=1}^{k_1} \eta_u^2, \quad S_{k,2} = \sum_{u=1}^{k_1} Y_u(m_1 + 1), \quad V_{k,2}^2 = \sum_{u=1}^{k_1} Y_u(m_1 + 1),
\]
\[
B_n^2 = \sum_{j=1}^{k} \mathbb{E}Y_j^2, \quad B_{n,1}^2 = \sum_{u=1}^{k_1} \sum_{j \in I_u} \mathbb{E}Y_j^2, \quad B_{n,2}^2 = \sum_{u=1}^{k_1} \mathbb{E}Y_u(m_1 + 1)^2.
\]

By Lemma 5.2, it is easy to see that under the mixing condition (3.6),
\[
\frac{B_{n,1}^2}{B_n^2} = 1 + O\left(\frac{k_1}{k}\right) = 1 + O(n^{-\alpha}), \quad \frac{\mathbb{E}S_{k,1}^2}{B_{n,1}^2} = 1 + O(n^{-\alpha}) + O(n^{-\alpha}).
\]
Let $\hat{Y}_j = Y_j \mathbb{1} (|Y_j| \leq b)$, where $b = B_{n,2} / (1 + x)$. Parallel to one-dependent case, we separate the big blocks and small blocks right after truncating the terms inside the small blocks. Denote

$$\hat{S}_{k,2} = \sum_{u=1}^{k_1} \hat{Y}_{u(m_1+1)}, \quad \hat{S}_k = S_{k,1} + \hat{S}_{k,2}$$

and

$$\hat{V}_{k,2} = \sum_{u=1}^{k_1} \hat{Y}_{u(m_1+1)}, \quad \hat{V}_k = V_{k,1} + \hat{V}_{k,2}.$$  

It is straightforward that

$$P(S_k \geq xV_k) \leq P(\hat{S}_k \geq x\hat{V}_k) + P\left(S_k \geq xV_k, \max_{1 \leq u \leq k_1} |Y_{u(m_1+1)}| > b\right),$$

$$P(S_k \geq xV_k) \geq P(\hat{S}_k \geq x\hat{V}_k) - P\left(S_k \geq xV_k, \max_{1 \leq u \leq k_1} |Y_{u(m_1+1)}| > b\right).$$

Further, for the main term $P(\hat{S}_k \geq x\hat{V}_k)$, we choose $\varepsilon = d_1 n^{-\alpha_1/2} \log n$ with a positive number $d_1 > 0$ and obtain

$$P\left(\hat{S}_k \geq x\hat{V}_k\right) \leq P(S_{k,1} \geq xV_{k,1} - \varepsilon xB_n) + P\left(\hat{S}_{k,2} \geq \varepsilon xB_n\right),$$

$$P\left(\hat{S}_k \geq x\hat{V}_k\right) \geq P(S_{k,1} \geq xV_{k,1} + \varepsilon xB_n) - P\left(S_{k,1} \geq xV_{k,1} + \varepsilon xB_n, V_{k,1}^2 \leq \frac{1}{4} B_n^2\right)$$

$$- P\left(S_{k,1} \geq xV_{k,1}, \frac{\hat{V}_{k,2}}{V_k + \hat{V}_k} < -\varepsilon xB_n, V_{k,1}^2 > \frac{1}{4} B_n^2\right).$$

The estimate for dominated terms $P(S_{k,1} \geq xV_{k,1} - \varepsilon xB_n)$ and $P(S_{k,1} \geq xV_{k,1} + \varepsilon xB_n)$ is presented in the following lemma and the proof will be shown in Section A.8.

**Proposition 5.8.** Assume $\varepsilon = d_1 n^{-\alpha_1/2} \log n$ for a positive number $d_1 > 0$ and $\alpha_1 \leq \alpha \tau$. Under the conditions of Theorem 3.2, there exist a positive constant $c_0$ depending on $d_1, \mu_1 / \mu_2, a_1, a_2, \alpha$ and $\tau$ such that

$$P(S_{k,1} \geq xV_{k,1} \pm \varepsilon xB_n) / \left[1 - \Phi(x)\right] = 1 + O\left(\frac{(1 + x)^4}{n^{1-\alpha-\alpha_1}} + \frac{1 + x}{n^{(1-\alpha-\alpha_1)/2}} + \frac{(1 + x)^2}{n^{\alpha}} + \frac{(1 + x)^2 \log n}{n^{\alpha_1/2}}\right)$$

uniformly for $x \in (0, c_0 \min\{n^{(1-\alpha-\alpha_1)/4}, n^{\alpha/2}, \max\{C_1, C_2\}\})$.

For small-block-related error terms, $S_{k,2}$ and $V_{k,2}^2$ can be replaced with the sum of independent random variables by Lemma 5.6. In addition, following a similar proof to Proposition 5.6, we obtain that under $\varepsilon = d_1 n^{-\alpha_1/2} \log n$ for $d_1$ being some positive number depending on $\mu_1 / \mu_2$, there exist positive numbers $C_1$ and $C_2$ depending on $a_1, a_2, \mu_1, \mu_2, \alpha$ and $\tau$ such that

$$P\left(\hat{S}_{k,2} \geq \varepsilon xB_n\right) \leq \exp\{-C_1 (1 + x)^2 d_1 \log n\} + C_2 \exp\{-a_2 n^{\alpha_1/2}\},$$

$$\exp\{-C_1 (1 + x)^2 d_1 \log n\} + C_2 \exp\{-a_2 n^{\alpha_1/2}\},$$
uniformly for $x \in (3, c_0 \min\{n^{1-\alpha_1}/4, n^{\alpha/2}\})$. When $0 < x \leq 3$, it follows from Lemmas 5.1 and 5.3 and Chebyshev inequality that under condition (3.6),
\[
\mathbb{P}(S_{k,1} \geq x V_{k,1}, V_{k,1}^2 \leq \frac{1}{4} B_n^2) \leq \mathbb{P}(V_{k,1}^2 \leq \frac{1}{4} B_n^2) \\
\leq a_1 n^{\alpha_1} e^{-a_2 \alpha \tau} + \exp\left\{-\frac{B_{n,1}^2 - B_n^2/4}{2 \sum_{u=1}^k \mathbb{E} n_u^4}\right\} \\
\leq a_1 n^{\alpha_1} e^{-a_2 \alpha \tau} + \exp\{-A_1 n^{1-\alpha_1} \mu_2/\mu_1\} \\
\leq A_2(1 + x)^4 \frac{1}{n^{1-\alpha_1}} \left[1 - \Phi(x)\right].
\]
Substituting (5.39), (5.33), (5.34), and (5.36) into (5.30) and (5.31) yields
\[
\mathbb{P}(\hat{S}_k \geq x \hat{V}_k)/[1 - \Phi(x)] \\
=1 + O\left(\frac{(1 + x)^4}{n^{1-\alpha_1} + \frac{1}{n^{(1-\alpha_1)/2}} + \frac{(1 + x)^2}{n^{\alpha_1}} + \frac{(1 + x)^2 \log n}{n^{\alpha_1/2}}\right)
\]
uniformly for $x \in (0, c_0 \min\{n^{(1-\alpha_1)/4}, n^{\alpha/2}, n^{\alpha/2}, (\log n)^{-1/2} n^{\alpha_1/4}\})$.

To avoid redundancy, we omit the analysis of the truncation errors in (5.28) and (5.29) as their proofs share the same fashion with (A.59) and (A.60). Consequently, (5.37) also holds for $\mathbb{P}(S_{k} \geq x V_{k})$. Finally, we need to balance the error terms by choosing $\alpha_1$ and seeking the best convergence rate or largest range for convergence. As a result, we choose $\alpha_1 = (1 - \alpha)/2$ when $(1 - \alpha)/2 \leq \alpha \tau$, and choose $\alpha_1 = \alpha \tau$ when $(1 - \alpha)/2 > \alpha \tau$, and then the desired result follows. This completes the proof for Theorem 3.2.

5.4. Proof of Theorem 3.3. The main idea is to use one-dependent random variables to approximate $\{Y_j\}_{1 \leq j \leq k}$ and then apply Theorem 3.1. Recall that $m = [n^{\alpha}]$ and $k = [n/m]$. Let
\[
\hat{Y}_j = \mathbb{E}(Y_j|\xi_l, m_j - 2m + 1 \leq l \leq m_j)
\]
and

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

As \( \{\varepsilon_t\}_{t \in \mathbb{Z}} \) are i.i.d. random variables, \( \{\tilde{Y}_j\}_{j \geq 1} \) are one-dependent. Note that by conditional Jensen’s inequality, for \( 2 \leq r \leq 4 \),

$$
\left\| X_i - \mathbb{E}\left( X_i | \varepsilon_{\ell} : mj - 2m + 1 \leq \ell \leq i \right) \right\|^r_r \\
= \mathbb{E} \left\{ \left| X_i - G_k \left( \mathcal{F}_{mj-2m}, \varepsilon_{mj-2m+1}, \ldots, \varepsilon_i \right) \right| \right\}^r \\
\leq \mathbb{E} \left[ \left| X_i - G_k \left( \mathcal{F}_{mj-2m}, \varepsilon_{mj-2m+1}, \ldots, \varepsilon_i \right) \right|^r \right] \\
\leq [\Delta_r(i - mj + 2m)]^r,
$$

which together with the assumption (3.16) and the fact that \( mj - 1 + 1 \leq i \leq mj \) for \( i \in H_j \) yields

$$
\| Y_j - \tilde{Y}_j \|_r \leq \sum_{i \in H_j} \left\| X_i - \mathbb{E}\left( X_i | \varepsilon_{\ell} : i - 2m + 1 \leq \ell \leq i \right) \right\|_r \\
\leq m a_1 e^{-a_2 m r}.
$$

The above bound shows that \( \{Y_j\}_{1 \leq j \leq k} \) can be well approximated by the one-dependent sequence \( \{\tilde{Y}_j\}_{1 \leq j \leq k} \). We can derive Theorem 3.3 by aggregating the following two propositions. The proofs of Propositions 5.9 and 5.10 will be provided in Sections A.9 and A.10 in the Supplementary Material, respectively.

**PROPOSITION 5.9.** Under conditions of Theorem 3.3, we have there exists a positive constant \( d_0 \) depending on \( \tau, \alpha, w_1, a_1 \) and \( a_2 \) such that

$$
P(\tilde{T}_k \geq x) = [1 - \Phi(x)] \left( 1 + O\left( \frac{1 + x^2}{n(1-\alpha)/4} + \frac{1 + x^2}{n^\alpha} \right) \right),
$$

Uniformly for \( x \in (0, d_0 \min\{n(1-\alpha)/8, n^{\alpha/2}\}) \).

**PROPOSITION 5.10.** Under conditions of Theorem 3.3, we have for \( x > 0 \),

$$
P(T_k \geq x) \leq P(\tilde{T}_k \geq x - C_1 n^{-1}) + C_2(e^{-a_2 n^\alpha} + e^{-O(n^{1-\alpha})})
$$

and

$$
P(T_k \geq x) \geq P(\tilde{T}_k \geq x + C_1 n^{-1}) - C_2(e^{-a_2 n^\alpha} + e^{-O(n^{1-\alpha})}).
$$

Applying Proposition 5.9 to \( P(\tilde{T}_k \geq x + O(n^{-1})) \) yields

$$
P(\tilde{T}_k \geq x + O(n^{-1})) \\
= [1 - \Phi(x + O(n^{-1}))] \left( 1 + O\left( \frac{1 + x^2}{n(1-\alpha)/4} + \frac{1 + x^2}{n^\alpha} \right) \right) \\
= [1 - \Phi(x)] \left( 1 + O\left( \frac{1 + x^2}{n(1-\alpha)/4} + \frac{1 + x^2}{n^\alpha} \right) \right)
$$

uniformly for \( x \in (0, d_0 \min\{n(1-\alpha)/8, n^{\alpha/2}\}) \). Note that the error term \( e^{-a_2 n^\alpha} + e^{-O(n^{1-\alpha})} \) decays at an exponential rate, which is always faster than the polynomial rate. By substituting the above result into (5.40) and (5.41) leads to the desired result (3.21). The proof of Theorem 3.3 is completed.
Acknowledgements. The authors would like to thank the Co-Editors, Associate Editor, and referees for their constructive comments that have helped improve the paper significantly.

SUPPLEMENTARY MATERIAL

Supplement to “Refined Cramér Type Moderate Deviation Theorems for General Self-normalized Sums with Applications to Dependent Random Variables and Winsorized Mean”. The supplement Gao, Shao and Shi (2021) contains all the technical details of the proofs.

REFERENCES

BENTKUS, V., BLOZNELIS, M. and GÖTZE, F. (1996). A Berry-Esseen bound for Student’s statistic in the non-i.i.d. case. J. Theoret. Probab. 9 765–796. MR1400598
BENTKUS, V. and GÖTZE, F. (1996). The Berry-Esseen bound for Student’s statistic. Ann. Probab. 24 491–503. MR1387647
BERBEE, H. (1987). Convergence rates in the strong law for bounded mixing sequences. Probab. Theory Related Fields 74 255–270. MR871254
CHANG, J., TANG, C. Y. and WU, Y. (2016). Local independence feature screening for nonparametric and semiparametric models by marginal empirical likelihood. Ann. Statist. 44 515–539. MR3476608
CHEN, X., SHAO, Q.-M., WU, W. B. and XU, L. (2016). Self-normalized Cramér-type moderate deviations under dependence. Ann. Statist. 44 1593–1617. MR3519934
CRAMÉR, H. (1938). Sur un nouveau théorème-limite de la théorie des probabilités. Actual. Sci. Ind. 736 5–23.
DE LA PEÑA, V. H., LAI, T. L. and SHAO, Q.-M. (2009). Self-normalized processes. Probability and its Applications (New York). Springer-Verlag, Berlin. Limit theory and statistical applications. MR2488094
DELAIGLE, A. and HALL, P. (2009). Higher criticism in the context of unknown distribution, non-independence and classification. In Perspectives in mathematical sciences. I. Stat. Sci. Interdiscip. Res. 7 109–138. World Sci. Publ., Hackensack, NJ. MR2581742
DIXON, W. J. (1960). Simplified estimation from censored normal samples. Ann. Math. Statist. 31 385–391. MR19358
FAN, J. and FAN, Y. (2008). High-dimensional classification using features annealed independence rules. Ann. Statist. 36 2605–2637. MR2485009
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
GAO, L., SHAO, Q.-M. and SHI, J. (2021). Supplement to “Refined Cramér Type Moderate Deviation Theorems for General Self-normalized Sums with Applications to Dependent Random Variables and Winsorized Mean”. GINE, E., GÖTZE, F. and MASON, D. M. (1997). Higher criticism in the context of unknown distribution, non-independence and classification. In Perspectives in mathematical sciences. I. Stat. Sci. Interdiscip. Res. 7 109–138. World Sci. Publ., Hackensack, NJ. MR2581742
HALL, P. (1987). Edgeworth expansion for Student’s t statistic under minimal moment conditions. Ann. Probab. 15 920–931. MR893906
HSING, T. and WU, W. B. (2004). On weighted U-statistics for stationary processes. Ann. Probab. 32 1600–1631. MR2060311
HUBER, P. J. (1964). Robust estimation of a location parameter. Ann. Math. Statist. 35 73–101. MR161415
HUBER, P. J. (1973). Robust regression: asymptotics, conjectures and Monte Carlo. Ann. Statist. 1 799–821. MR356373
JING, B.-Y., SHAO, Q.-M. and WANG, Q. (2003). Self-normalized Cramér-type large deviations for independent random variables. Ann. Probab. 31 2167–2215. MR2016616
LIN, Z. and BAI, Z. (2010). Probability inequalities. Science Press Beijing, Beijing; Springer, Heidelberg. MR2789096
LIU, W. and SHAO, Q.-M. (2010). Cramér-type moderate deviation for the maximum of the periodogram with application to simultaneous tests in gene expression time series. Ann. Statist. 38 1913–1935. MR2662363
LIU, W. and SHAO, Q.-M. (2013). A Cramér moderate deviation theorem for Hotelling’s T2-statistic with applications to global tests. Ann. Statist. 41 296–322. MR3059419
ROTHENBERG, T. J., FISHER, F. M. and TILANSUS, C. B. (1964). A note on estimation from a Cauchy sample. J. Amer. Statist. Assoc. 59 460–463. MR166872
SHAO, Q.-M. (1999). A Cramér type large deviation result for Student’s t-statistic. J. Theoret. Probab. 12 385–398. MR1684750
SHAO, Q. (2018). On necessary and sufficient conditions for the self-normalized central limit theorem. Sci. China Math. 61 1741–1748. MR3856964

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

SHAO, Q.-M. and ZHOU, W.-X. (2016). Cramér type moderate deviation theorems for self-normalized processes. Bernoulli 22 2029–2079. MR3498022

SHERGIN, V. (1980). On the convergence rate in the central limit theorem for m-dependent random variables. Theory of Probability & Its Applications 24 782–796.

WANG, Q. (2011). Refined self-normalized large deviations for independent random variables. J. Theoret. Probab. 24 307–329. MR2795041

WANG, Q. and JING, B.-Y. (1999). An exponential nonuniform Berry-Esseen bound for self-normalized sums. Ann. Probab. 27 2068–2088. MR1742902

WU, W. B. (2005). Nonlinear system theory: another look at dependence. Proc. Natl. Acad. Sci. USA 102 14150–14154. MR2172215

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

ZHOU, W.-X., BOSE, K., FAN, J. and LIU, H. (2018). A new perspective on robust M-estimation: finite sample theory and applications to dependence-adjusted multiple testing. Ann. Statist. 46 1904–1931. MR3845005