1-1-2020

A Berry-Esseen bound of order $\frac{1}{\sqrt{n}}$ for martingales

Songqi Wu
Tianjin University

Xiaohui Ma
Tianjin University

Hailin Sang
University of Mississippi

Xiequan Fan
Tianjin University

INSTITUT DE FRANCE
Académie des sciences

Comptes Rendus
Mathématique

Songqi Wu, Xiaohui Ma, Hailin Sang and Xiequan Fan

A Berry–Esseen bound of order $\frac{1}{\sqrt{n}}$ for martingales

Volume 358, issue 6 (2020), p. 701-712.

<https://doi.org/10.5802/crmath.81>

Les Comptes Rendus. Mathématique *sont membres du Centre Mersenne pour l'édition scientifique ouverte*
www.centre-mersenne.org

Comptes Rendus
Mathématique
2020, 358, nº 6, p. 701-712
https://doi.org/10.5802/crmath.81

Probabilities / *Probabilités*

A Berry–Esseen bound of order $\frac{1}{\sqrt{n}}$ for martingales

Une borne de Berry–Esseen d'ordre $\frac{1}{\sqrt{n}}$ pour les martingales

Songqi Wu [a], Xiaohui Ma [a], Hailin Sang [b] and Xiequan Fan [*, a]

[a] Center for Applied Mathematics, Tianjin University, Tianjin, China

[b] Department of Mathematics, The University of Mississippi, University, MS 38677, USA

E-mails: wusongqi7@163.com, maxiaohui@tju.edu.cn, sang@olemiss.edu, fanxiequan@hotmail.com

Abstract. Renz [13] has established a rate of convergence $1/\sqrt{n}$ in the central limit theorem for martingales with some restrictive conditions. In the present paper a modification of the methods, developed by Bolthausen [2] and Grama and Haeusler [6], is applied for obtaining the same convergence rate for a class of more general martingales. An application to linear processes is discussed.

Résumé. Renz [13] a établi un taux de convergence $1/\sqrt{n}$ dans le théorème de la limite centrale pour les martingales avec certaines conditions restrictives. Dans le présent article, une modification des méthodes, développées par Bolthausen [2] et Grama et Haeusler [6], est appliquée pour obtenir le même taux de convergence pour une classe de martingales plus générales. Une application aux processus linéaires est discutée.

Funding. The work has been partially supported by the National Natural Science Foundation of China (Grant no. 11601375). The research of Hailin Sang is partial supported by the Simons Foundation (Grant no. 586789).

Manuscript received 23rd January 2020, accepted 2nd June 2020.

[*] Corresponding author.

1. Introduction and main result

For $n \in \mathbf{N}$, let $(\xi_i, \mathscr{F}_i)_{i=0,\dots,n}$ be a finite sequence of martingale differences defined on some probability space $(\Omega, \mathscr{F}, \mathbf{P})$, where $\xi_0 = 0$ and $\{\emptyset, \Omega\} = \mathscr{F}_0 \subseteq \cdots \subseteq \mathscr{F}_n \subseteq \mathscr{F}$ are increasing σ-fields. Denote

$$X_0 = 0, \quad X_k = \sum_{i=1}^{k} \xi_i, \ k = 1, \dots, n.$$

Then $X = (X_k, \mathscr{F}_k)_{k=0,\dots,n}$ is a martingale. Denote by $\langle X \rangle$ the conditional variance of X:

$$\langle X \rangle_0 = 0, \quad \langle X \rangle_k = \sum_{i=1}^{k} \mathbf{E}\big[\xi_i^2 \big| \mathscr{F}_{i-1}\big], \ k = 1, \dots, n.$$

Define

$$D(X_n) = \sup_{x \in \mathbf{R}} \big| \mathbf{P}(X_n \le x) - \Phi(x) \big|,$$

where $\Phi(x)$ is the distribution function of the standard normal random variable. Denote by $\xrightarrow{\mathbf{P}}$ the convergence in probability as $n \to \infty$. According to the martingale central limit theorem, the "conditional Lindeberg condition"

$$\sum_{i=1}^{n} \mathbf{E}\big[\xi_i^2 \mathbf{1}_{\{|\xi_i| \ge \epsilon\}} \big| \mathscr{F}_{i-1}\big] \xrightarrow{\mathbf{P}} 0, \quad \text{for each } \epsilon > 0,$$

and the "conditional normalizing condition" $\langle X \rangle_n \xrightarrow{\mathbf{P}} 1$ together implies asymptotic normality of X_n, that is, $D(X_n) \to 0$ as $n \to \infty$.

The convergence rate of $D(X_n)$ has attracted a lot of attentions. For instance, Bolthausen [2] proved that if $|\xi_i| \le \epsilon_n$ for a number ϵ_n and $\langle X \rangle_n = 1$ a.s., then $D(X_n) \le c \epsilon_n^3 n \log n$, where, here and after, c is an absolute constant not depending on ϵ_n and n. El Machkouri and Ouchti [3] improved the factor $\epsilon_n^3 n \log n$ in Bolthausen's bound to $\epsilon_n \log n$ under the following more general condition

$$\mathbf{E}\big[|\xi_i|^3 \big| \mathscr{F}_{i-1}\big] \le \epsilon_n \mathbf{E}\big[\xi_i^2 \big| \mathscr{F}_{i-1}\big] \quad a.s. \text{ for all } i = 1, 2, \dots, n.$$

For more related results, we refer to Ouchti [12] and Mourrat [11]. Recently, Fan [4] proved that if there exist a positive constant ρ and a number ϵ_n, such that

$$\mathbf{E}\big[|\xi_i|^{2+\rho} \big| \mathscr{F}_{i-1}\big] \le \epsilon_n^{\rho} \mathbf{E}\big[\xi_i^2 \big| \mathscr{F}_{i-1}\big] \quad a.s. \text{ for all } i = 1, 2, \dots, n,$$

and $\langle X \rangle_n = 1$ a.s., then $D(X_n) \le c_\rho \widehat{\epsilon}_n$, where

$$\widehat{\epsilon}_n = \begin{cases} \epsilon_n^{\rho}, & \text{if } \rho \in (0,1), \\ \epsilon_n |\log \epsilon_n|, & \text{if } \rho \ge 1, \end{cases}$$

and c_ρ is a constant depending only on ρ. Fan [4] also showed that this Berry–Esseen bound is optimal. In particular, if $\epsilon_n \asymp 1/\sqrt{n}$, then we have $\epsilon_n |\log \epsilon_n| \asymp (\log n)/\sqrt{n}$. Thus, we cannot obtain the classical convergence rate $1/\sqrt{n}$ for general martingales.

However, the convergence rate $1/\sqrt{n}$ for martingales is possible to be attained with some additional restrictive conditions. For instance, Renz [13] proved that if there exists a constant $\rho > 0$ such that

$$\mathbf{E}[\xi_i^2 | \mathscr{F}_{i-1}] = 1/n, \quad \mathbf{E}[\xi_i^3 | \mathscr{F}_{i-1}] = 0 \quad \text{and} \quad \mathbf{E}\big[|\xi_i|^{3+\rho} \big| \mathscr{F}_{i-1}\big] \le c n^{-(3+\rho)/2}, \quad \text{a.s.,} \tag{1}$$

then it holds

$$D(X_n) = O\bigg(\frac{1}{\sqrt{n}}\bigg). \tag{2}$$

He also showed that this result is not true for $\rho = 0$. More martingale Berry–Esseen bounds of convergence rate $1/\sqrt{n}$ can be found in Bolthausen [2] and Kir'yanova and Rotar [10].

In this paper we are interested in extending (2) to a class of more general martingales. The following theorem is our main result.

Theorem 1. *Assume that there exist some numbers $\rho \in (0, +\infty)$, $\epsilon_n \in (0, \frac{1}{2}]$ and $\delta_n \in [0, \frac{1}{2}]$ such that for all $1 \le i \le n$,*

$$\left|\langle X\rangle_n - 1\right| \le \delta_n^2, \tag{3}$$

$$\mathbf{E}\left[\xi_i^3 \middle| \mathscr{F}_{i-1}\right] = 0 \tag{4}$$

and

$$\mathbf{E}\left[|\xi_i|^{3+\rho} \middle| \mathscr{F}_{i-1}\right] \le \epsilon_n^{1+\rho} \mathbf{E}\left[\xi_i^2 \middle| \mathscr{F}_{i-1}\right] \quad a.s. \tag{5}$$

Then

$$D(X_n) \le c_\rho(\epsilon_n + \delta_n),$$

where c_ρ depends only on ρ. In addition, it holds $c_\rho = O(\rho^{-1})$, $\rho \to 0$.

Notice that under the conditions of Renz [13], the conditions of Theorem 1 are satisfied with $\delta_n = 0$ and $\epsilon_n \asymp 1/\sqrt{n}$. Thus Theorem 1 extends Renz's result to a class of more general martingales.

Thanks to the additional condition (4), the Berry–Esseen bound (6) improves the bound of Fan [4] by replacing $\epsilon_n |\log \epsilon_n|$ with ϵ_n.

Relaxing the condition (3), we have the following analogue estimation of Fan (cf. [4, (26)]).

Theorem 2. *Assume that there exist some numbers $\rho \in (0, +\infty)$ and $\epsilon_n \in (0, \frac{1}{2}]$ such that for all $1 \le i \le n$,*

$$\mathbf{E}\left[\xi_i^3 \middle| \mathscr{F}_{i-1}\right] = 0$$

and

$$\mathbf{E}\left[|\xi_i|^{3+\rho} \middle| \mathscr{F}_{i-1}\right] \le \epsilon_n^{1+\rho} \mathbf{E}\left[\xi_i^2 \middle| \mathscr{F}_{i-1}\right] \quad a.s.$$

Then, for all $p \ge 1$,

$$D(X_n) \le c_\rho \epsilon_n + c_p \left(\mathbf{E}\left[\left|\langle X\rangle_n - 1\right|^p\right] + \mathbf{E}\left[\max_{1\le i\le n} |\xi_i|^{2p}\right]\right)^{1/(2p+1)}, \tag{6}$$

where c_ρ and c_p depend only on ρ and p, respectively.

It is easy to see that when $p \to \infty$,

$$\left(\mathbf{E}\left[\left|\langle X\rangle_n - 1\right|^p\right]\right)^{1/(2p+1)} \to \|\langle X\rangle_n - 1\|_\infty^{1/2},$$

which coincides with δ_n of Theorem 1.

2. Application

We first extend Theorem 1 to triangular arrays with infinity many terms in each line. For $n \in \mathbf{N}$, let $(\xi_{n,i}, \mathscr{F}_{n,i})_{i=-\infty}^n$ be a sequence of martingale differences defined on some probability space $(\Omega, \mathscr{F}, \mathbf{P})$, where the adapted filtration is $\{\emptyset, \Omega\} = \mathscr{F}_{-\infty} \subset \cdots \subset \mathscr{F}_{n,n-1} \subset \mathscr{F}_{n,n} \subset \mathscr{F}$. Denote $X_{n,k} = \sum_{i=-\infty}^k \xi_{n,i}$, $k \le n$. Then $(X_{n,k}, \mathscr{F}_{n,k})_{k=-\infty}^n$ is a martingale. Let $\langle X\rangle_{n,k} = \sum_{i=-\infty}^k \mathbf{E}[\xi_{n,i}^2 | \mathscr{F}_{n,i-1}]$, $k \le n$. In particular, denote $X_n := X_{n,n}$ and $\langle X\rangle_n := \langle X\rangle_{n,n}$.

With some slight modification on the proof, Theorem 1 still holds in this new setting. Now we apply Theorem 1 with this new setting to the partial sum of linear processes. Let $(\varepsilon_i)_{i\in\mathbf{Z}}$ be a sequence of identically distributed martingale differences adapted to the filtration $(\mathscr{F}_i)_{i\in\mathbf{Z}}$. We consider the causal linear process in the form

$$Y_k = \sum_{j=-\infty}^{k} a_{k-j}\varepsilon_j, \tag{7}$$

where the martingale differences have finite variance and the sequence of real coefficients satisfies $\sum_{i=0}^\infty a_i^2 < \infty$. Without loss of generality, let the variance of the martingale difference to be 1. We say the linear process has long memory if $\sum_{i=0}^{\infty}|a_i| = \infty$. In this case, we assume that $a_0 = 1$ and

$$a_i = \ell(i)i^{-\alpha},\ i > 0, \text{ with } 1/2 < \alpha < 1. \tag{8}$$

Here $\ell(\cdot)$ is a slowly varying function. On the other hand, we say the linear process has short memory if $\sum_{i=0}^{\infty}|a_i| < \infty$ and $\sum_{i=0}^{\infty} a_i \neq 0$. The third case is $\sum_{i=0}^{\infty}|a_i| < \infty$ and $\sum_{i=0}^{\infty} a_i = 0$.

The long memory linear processes covers the well-known fractional ARIMA processes (cf. Granger and Joyeux [7]; Hosking [9]), which play an important role in financial time series modeling and application. As a special case, let $0 < d < 1/2$ and B be the backward shift operator with $B\varepsilon_k = \varepsilon_{k-1}$ and consider

$$Y_k = (1-B)^{-d}\varepsilon_k = \sum_{i=0}^{\infty} a_i\varepsilon_{k-i}, \quad \text{where } a_i = \frac{\Gamma(i+d)}{\Gamma(d)\Gamma(i+1)}.$$

For this example we have $\lim_{n\to\infty} a_n/n^{d-1} = 1/\Gamma(d)$. Note that these processes have long memory because $\sum_{j=0}^{\infty}|a_j| = \infty$.

The partial sum $S_n = \sum_{k=1}^{n} Y_k$ of causal linear process (7) can be written as $S_n = \sum_{i=-\infty}^{n} b_{n,i}\varepsilon_i$, where $b_{n,i} = \sum_{j=0}^{n-i} a_j$ for $0 < i \le n$, and $b_{n,i} = \sum_{j=1-i}^{n-i} a_j$ for $i \le 0$. The variance of S_n is $B_n^2 = \operatorname{var}(S_n) = \sum_{i=-\infty}^{n} b_{n,i}^2$. Now let $X_{n,k} = \sum_{i=-\infty}^{k} b_{n,i}\varepsilon_i/B_n$. Then $X_n = X_{n,n} = S_n/B_n$ and $\langle X\rangle_n = \sum_{i=-\infty}^{n} b_{n,i}^2\mathbf{E}[\varepsilon_i^2|\mathscr{F}_{i-1}]/B_n^2$. If we assume $|\langle X\rangle_n - 1| \le \delta_n^2$ for some $\delta_n \in [0,\frac{1}{2}]$, $\mathbf{E}[\varepsilon_i^3|\mathscr{F}_{i-1}] = 0$ and $\mathbf{E}[|\varepsilon_i|^{3+\rho}|\mathscr{F}_{i-1}] \le d_\rho^{1+\rho}\mathbf{E}[\varepsilon_i^2|\mathscr{F}_{i-1}]$ a.s. for all $i \in \mathbf{Z}$ and some constant d_ρ, then, by Theorem 1,

$$\sup_{x\in\mathbf{R}}|\mathbf{P}(S_n/B_n \le x) - \Phi(x)| \le c_\rho(\epsilon_n + \delta_n),$$

where $\epsilon_n = d_\rho \sup_{i\le n}|b_{n,i}|/B_n$.

In the case that $\sum_{i=0}^{\infty}|a_i| < \infty$, $\sup_{i\le n}|b_{n,i}| \le \sum_{i=0}^{\infty}|a_i| < \infty$ and it is well known that B_n^2 has order n. Hence ϵ_n has order $1/\sqrt{n}$ in this case. In the long memory case $\sum_{i=0}^{\infty}|a_i| = \infty$, if we assume (8), B_n^2 has order $n^{3-2\alpha}\ell^2(n)$ (e.g., Wu and Min [14]) and $\sup_{i\le n}|b_{n,i}|$ has order $n^{1-\alpha}\ell(n)$ (see Beknazaryan et al. [1] for upper bound and Fortune et al. [5] for lower bound in the case $d = 1$). Hence in this case ϵ_n also has order $1/\sqrt{n}$. In either case the Berry–Esseen bound has order $1/\sqrt{n}$ if $\delta_n = O(n^{-1/2})$. In particular, if we in addition assume that the innovations $(\varepsilon_i)_{i\in\mathbf{Z}}$ are independent, then $\delta_n = 0$ and the Berry–Esseen bound $\sup_{x\in\mathbf{R}}|\mathbf{P}(S_n/B_n \le x) - \Phi(x)|$ has order $1/\sqrt{n}$. Here the condition $\mathbf{E}[\varepsilon_i^3|\mathscr{F}_{i-1}] = 0$ is needed to have the Berry–Esseen bound of order $1/\sqrt{n}$. We cannot have this order from the result of Fan [4].

3. Proofs of theorems

3.1. *Preliminary lemmas*

In the proofs of theorems, we need the following technical lemmas. The first two lemmas can be found in Fan [4, Lemmas 3.1 and 3.2].

Lemma 3. *If there exists an $s > 3$ such that*

$$\mathbf{E}[|\xi_i|^s|\mathscr{F}_{i-1}] \le \epsilon_n^{s-2}\mathbf{E}[\xi_i^2|\mathscr{F}_{i-1}],$$

then, for any $t \in [3,s)$,

$$\mathbf{E}[|\xi_i|^t|\mathscr{F}_{i-1}] \le \epsilon_n^{t-2}\mathbf{E}[\xi_i^2|\mathscr{F}_{i-1}].$$

Lemma 4. *If there exists an $s > 3$ such that*

$$\mathbf{E}[|\xi_i|^s|\mathscr{F}_{i-1}] \le \epsilon_n^{s-2}\mathbf{E}[\xi_i^2|\mathscr{F}_{i-1}],$$

then

$$\mathbf{E}[\xi_i^2|\mathscr{F}_{i-1}] \le \epsilon_n^2.$$

The next two technical lemmas are due to Bolthausen (cf. [2, Lemmas 1 and 2]).

Lemma 5. *Let X and Y be random variables. Then*
$$\sup_u \left|\mathbf{P}\big(X \le u\big) - \Phi(u)\right| \le c_1 \sup_u \left|\mathbf{P}\big(X+Y \le u\big) - \Phi(u)\right| + c_2 \left\|\mathbf{E}[Y^2|X]\right\|_\infty^{1/2},$$
where c_1 and c_2 are two positive constants.

Lemma 6. *Let $G(x)$ be an integrable function on $\mathbf{R}$ of bounded variation $\|G\|_V$, X be a random variable and $a, b \neq 0$ are real numbers. Then*
$$\mathbf{E}\left[G\left(\frac{X+a}{b}\right)\right] \le \|G\|_V \sup_u \left|\mathbf{P}\big(X \le u\big) - \Phi(u)\right| + \|G\|_1 |b|,$$
where $\|G\|_1$ is the $L_1(\mathbf{R})$ norm of $G(x)$.

In the proof of Theorem 2, we also need the following lemma of El Machkouri and Ouchti [3].

Lemma 7. *Let X and Y be two random variables. Then, for $p \ge 1$,*
$$D(X+Y) \le 2D(X) + 3\left\|\mathbf{E}\big[Y^{2p}|X\big]\right\|_1^{1/(2p+1)}. \tag{9}$$

3.2. *Proof of Theorem 1*

By Lemma 3, we only need to consider the case of $\rho \in (0,1]$. We follow the method of Grama and Haeusler [6]. Let $T = 1 + \delta_n^2$. We introduce a modification of the conditional variance $\langle X\rangle_n$ as follows:
$$V_k = \langle X\rangle_k \mathbf{1}_{\{k<n\}} + T\mathbf{1}_{\{k=n\}}. \tag{10}$$
It is easy to see that $V_0 = 0, V_n = T$, and that $(V_k, \mathscr{F}_k)_{k=0,\dots,n}$ is a predictable process. Set
$$\gamma = \epsilon_n + \delta_n.$$
Let c_* be some positive and sufficient large constant. Define the following non-increasing discrete time predictable process
$$A_k = c_*^2\gamma^2 + T - V_k, \quad k = 1,\dots,n. \tag{11}$$
Obviously, we have $A_0 = c_*^2\gamma^2 + T$ and $A_n = c_*^2\gamma^2$. In addition, for $u, x \in \mathbf{R}$, and $y > 0$, denote
$$\Phi_u(x,y) = \Phi\left(\frac{u-x}{\sqrt{y}}\right). \tag{12}$$
Let $\mathcal{N} = \mathcal{N}(0,1)$ be a standard normal random variable, which is independent of X_n. Using a smoothing procedure, by Lemma 5, we deduce that
$$\begin{aligned}
\sup_u \left|\mathbf{P}\big(X_n \le u\big) - \Phi(u)\right| &\le c_1 \sup_u \left|\mathbf{P}\big(X_n + c_*\gamma\mathcal{N} \le u\big) - \Phi(u)\right| + c_2\gamma \\
&= c_1 \sup_u \left|\mathbf{E}\big[\Phi_u\big(X_n, A_n\big)\big] - \Phi(u)\right| + c_2\gamma \\
&\le c_1 \sup_u \left|\mathbf{E}\big[\Phi_u\big(X_n, A_n\big)\big] - \mathbf{E}\big[\Phi_u\big(X_0, A_0\big)\big]\right| \\
&\quad + c_1 \sup_u \left|\mathbf{E}\big[\Phi_u\big(X_0, A_0\big)\big] - \Phi(u)\right| + c_2\gamma \\
&= c_1 \sup_u \left|\mathbf{E}\big[\Phi_u\big(X_n, A_n\big)\big] - \mathbf{E}\big[\Phi_u\big(X_0, A_0\big)\big]\right| \\
&\quad + c_1 \sup_u \left|\Phi\left(\frac{u}{\sqrt{c_*^2\gamma^2 + T}}\right) - \Phi(u)\right| + c_2\gamma.
\end{aligned} \tag{13}$$
It is obvious that
$$\left|\Phi\left(\frac{u}{\sqrt{c_*^2\gamma^2 + T}}\right) - \Phi(u)\right| \le c_3\left|\frac{1}{\sqrt{c_*^2\gamma^2 + T}} - 1\right| \le c_4\gamma. \tag{14}$$

Returning to (13), we get

$$\sup_u \left|\mathbf{P}\big(X_n \le u\big) - \Phi(u)\right| \le c_1 \sup_u \left|\mathbf{E}\big[\Phi_u\big(X_n, A_n\big)\big] - \mathbf{E}\big[\Phi_u\big(X_0, A_0\big)\big]\right| + c_5\gamma. \tag{15}$$

By a simple telescoping, we know that

$$\mathbf{E}\big[\Phi_u\big(X_n, A_n\big)\big] - \mathbf{E}\big[\Phi_u\big(X_0, A_0\big)\big] = \mathbf{E}\left[\sum_{k=1}^{n}\big(\Phi_u\big(X_k, A_k\big) - \Phi_u\big(X_{k-1}, A_{k-1}\big)\big)\right]. \tag{16}$$

Taking into account the fact that

$$\frac{\partial^2}{\partial x^2}\Phi_u(x,y) = 2\frac{\partial}{\partial y}\Phi_u(x,y),$$

we get

$$\mathbf{E}\big[\Phi_u\big(X_n, A_n\big)\big] - \mathbf{E}\big[\Phi_u\big(X_0, A_0\big)\big] = J_1 + J_2 - J_3, \tag{17}$$

where

$$\begin{aligned} J_1 = \mathbf{E}\Bigg[\sum_{k=1}^{n}\bigg(&\Phi_u(X_k, A_k) - \Phi_u(X_{k-1}, A_k) - \frac{\partial}{\partial x}\Phi_u(X_{k-1}, A_k)\xi_k \\ &- \frac{1}{2}\frac{\partial^2}{\partial x^2}\Phi_u(X_{k-1}, A_k)\xi_k^2 - \frac{1}{6}\frac{\partial^3}{\partial x^3}\Phi_u(X_{k-1}, A_k)\xi_k^3\bigg)\Bigg], \end{aligned} \tag{18}$$

$$J_2 = \frac{1}{2}\mathbf{E}\left[\sum_{k=1}^{n}\frac{\partial^2}{\partial x^2}\Phi_u(X_{k-1}, A_k)\big(\triangle\langle X\rangle_k - \triangle V_k\big)\right], \tag{19}$$

$$J_3 = \mathbf{E}\left[\sum_{k=1}^{n}\left(\Phi_u(X_{k-1}, A_{k-1}) - \Phi_u(X_{k-1}, A_k) - \frac{\partial}{\partial y}\Phi_u(X_{k-1}, A_k)\triangle V_k\right)\right], \tag{20}$$

where $\triangle\langle X\rangle_k = \langle X\rangle_k - \langle X\rangle_{k-1}$.

Now, we need to give some estimates of J_1, J_2 and J_3. To this end, we introduce some notations. Denote by ϑ_i some random variables satisfying $0 \le \vartheta_i \le 1$, which may represent different values at different places. For the rest of the paper, φ stands for the density function of the standard normal random variable.

Control of J_1. For convenience's sake, let $T_{k-1} = (u - X_{k-1})/\sqrt{A_k}$, $k = 1,2,\ldots,n$. It is easy to see that

$$\begin{aligned} B_k =: &\Phi_u(X_k, A_k) - \Phi_u(X_{k-1}, A_k) - \frac{\partial}{\partial x}\Phi_u(X_{k-1}, A_k)\xi_k \\ &- \frac{1}{2}\frac{\partial^2}{\partial x^2}\Phi_u(X_{k-1}, A_k)\xi_k^2 - \frac{1}{6}\frac{\partial^3}{\partial x^3}\Phi_u(X_{k-1}, A_k)\xi_k^3 \\ = &\Phi\left(T_{k-1} - \frac{\xi_k}{\sqrt{A_k}}\right) - \Phi(T_{k-1}) + \Phi'(T_{k-1})\frac{\xi_k}{\sqrt{A_k}} \\ &- \frac{1}{2}\Phi''(T_{k-1})\left(\frac{\xi_k}{\sqrt{A_k}}\right)^2 + \frac{1}{6}\Phi'''(T_{k-1})\left(\frac{\xi_k}{\sqrt{A_k}}\right)^3. \end{aligned}$$

To estimate the right hand side of the last equality, we distinguish two cases.

Case 1: $|\xi_k/\sqrt{A_k}| \le 2 + |T_{k-1}|/2$. By a four-term Taylor expansion, it is obvious that if $|\xi_k/\sqrt{A_k}| \le 1$, then

$$\begin{aligned} \left|B_k\right| &= \left|\frac{1}{24}\Phi^{(4)}\left(T_{k-1} - \vartheta\frac{\xi_k}{\sqrt{A_k}}\right)\left|\frac{\xi_k}{\sqrt{A_k}}\right|^4\right| \\ &\le \left|\Phi^{(4)}\left(T_{k-1} - \vartheta\frac{\xi_k}{\sqrt{A_k}}\right)\right|\left|\frac{\xi_k}{\sqrt{A_k}}\right|^{3+\rho}. \end{aligned}$$

If $|\xi_k/\sqrt{A_k}| > 1$, by a three-term Taylor expansion, then

$$\begin{aligned}\left|B_k\right| &\le \frac{1}{2}\left(\left|\Phi'''\left(T_{k-1}-\vartheta\frac{\xi_k}{\sqrt{A_k}}\right)\right|+\left|\Phi'''(T_{k-1})\right|\right)\left|\frac{\xi_k}{\sqrt{A_k}}\right|^3\\ &\le \left|\Phi'''\left(T_{k-1}-\vartheta'\frac{\xi_k}{\sqrt{A_k}}\right)\right|\left|\frac{\xi_k}{\sqrt{A_k}}\right|^3\\ &\le \left|\Phi'''\left(T_{k-1}-\vartheta'\frac{\xi_k}{\sqrt{A_k}}\right)\right|\left|\frac{\xi_k}{\sqrt{A_k}}\right|^{3+\rho},\end{aligned}$$

where

$$\vartheta' = \begin{cases}\vartheta, & \text{if } \left|\Phi'''\left(T_{k-1}-\vartheta\frac{\xi_k}{\sqrt{A_k}}\right)\right| \ge |\Phi'''(T_{k-1})|,\\ 0, & \text{if } \left|\Phi'''\left(T_{k-1}-\vartheta\frac{\xi_k}{\sqrt{A_k}}\right)\right| < |\Phi'''(T_{k-1})|.\end{cases}$$

Using the inequality $\max\{|\Phi'''(t)|,|\Phi''''(t)|\} \le \varphi(t)(2+t^4)$, we find that

$$\begin{aligned}\left|B_k\mathbf{1}_{\{|\xi_k/\sqrt{A_k}|\le 2+|T_{k-1}|/2\}}\right| &\le \varphi\left(T_{k-1}-\vartheta_1\frac{\xi_k}{\sqrt{A_k}}\right)\left(2+\left(T_{k-1}-\vartheta_1\frac{\xi_k}{\sqrt{A_k}}\right)^4\right)\left|\frac{\xi_k}{\sqrt{A_k}}\right|^{3+\rho}\\ &\le g_1(T_{k-1})\left|\frac{\xi_k}{\sqrt{A_k}}\right|^{3+\rho},\end{aligned}\tag{21}$$

where

$$g_1(z) = \sup_{|t-z|\le 2+|z|/2} \varphi(t)(2+t^4).$$

Case 2: $|\xi_k/\sqrt{A_k}| > 2+|T_{k-1}|/2$. It is obvious that, for $|\triangle x| > 1+|x|/2$,

$$\begin{aligned}&\left|\Phi(x-\triangle x)-\Phi(x)+\Phi'(x)\triangle x-\frac{1}{2}\Phi''(x)(\triangle x)^2+\frac{1}{6}\Phi'''(x)(\triangle x)^3\right|\\ &\qquad\le \left(\left|\frac{\Phi(x-\triangle x)-\Phi(x)}{|\triangle x|^3}\right|+|\Phi'(x)|+|\Phi''(x)|+|\Phi'''(x)|\right)|\triangle x|^3\\ &\qquad\le \left(8\left|\frac{\Phi(x-\triangle x)-\Phi(x)}{(2+|x|)^3}\right|+|\Phi'(x)|+|\Phi''(x)|+|\Phi'''(x)|\right)|\triangle x|^3\\ &\qquad\le \left(\frac{\widetilde{c}}{(2+|x|)^3}+|\Phi'(x)|+|\Phi''(x)|+|\Phi'''(x)|\right)|\triangle x|^3\\ &\qquad\le \frac{\widehat{c}}{(2+|x|)^3}|\triangle x|^3\\ &\qquad\le \frac{\widehat{c}}{(2+|x|)^3}|\triangle x|^{3+\rho}.\end{aligned}$$

Hence, we have

$$\left|B_k\mathbf{1}_{\{|\xi_k/\sqrt{A_k}|>2+|T_{k-1}|/2\}}\right| \le g_2(T_{k-1})\left|\frac{\xi_k}{\sqrt{A_k}}\right|^{3+\rho},\tag{22}$$

where

$$g_2(z) = \frac{\widehat{c}}{(2+|z|)^3}.$$

Denote

$$G(z) = g_1(z)+g_2(z).$$

Combining (21) and (22) together, we get

$$|B_k| \le G(T_{k-1})\left|\frac{\xi_k}{\sqrt{A_k}}\right|^{3+\rho}.\tag{23}$$

Therefore,

$$\left|J_1\right| = \left|\mathbf{E}\left[\sum_{k=1}^{n} B_k\right]\right| \le \mathbf{E}\left[\sum_{k=1}^{n} G(T_{k-1})\left|\frac{\xi_k}{\sqrt{A_k}}\right|^{3+\rho}\right]. \tag{24}$$

Next, we consider conditional expectation of $|\xi_k|^{3+\rho}$. By condition (5), we get

$$\mathbf{E}\left[|\xi_k|^{3+\rho}\middle|\mathscr{F}_{k-1}\right] \le \epsilon_n^{1+\rho}\,\triangle\langle X\rangle_k, \tag{25}$$

where $\triangle\langle X\rangle_k = \langle X\rangle_k - \langle X\rangle_{k-1}$ and we know that

$$\triangle\langle X\rangle_k = \triangle V_k = V_k - V_{k-1},\ 1 \le k < n,\ \triangle\langle X\rangle_n \le \triangle V_n, \tag{26}$$

then

$$\mathbf{E}\left[|\xi_k|^{3+\rho}\middle|\mathscr{F}_{k-1}\right] \le \epsilon_n^{1+\rho}\,\triangle V_k. \tag{27}$$

By (24) and (27), we obtain

$$|J_1| \le R_1 := \epsilon_n^{1+\rho}\left[\sum_{k=1}^{n} \frac{G(T_{k-1})}{A_k^{(3+\rho)/2}}\triangle V_k\right]. \tag{28}$$

To estimate R_1, we introduce the time change τ_t as follow: for any real $t \in [0,T]$,

$$\tau_t = \min\{k \le n : V_k \ge t\},\ \text{ where } \min\emptyset = n. \tag{29}$$

Obviously, for any $t \in [0,T]$, the stopping time τ_t is predictable. In addition, $(\sigma_k)_{k=1,\dots,n+1}$ (with $\sigma_1 = 0$) stands for the increasing sequence of moments when the increasing and stepwise function $\tau_t, t \in [0,T]$, has jumps. It is easy to see that $\triangle V_k = \int_{[\sigma_k,\sigma_{k+1})}\mathrm{d}t$, and that $k = \tau_t$ for $t \in [\sigma_k,\sigma_{k+1})$. Since $\tau_T = n$, we have

$$\sum_{k=1}^{n} \frac{G(T_{k-1})}{A_k^{(3+\rho)/2}}\triangle V_k = \sum_{k=1}^{n}\int_{[\sigma_k,\sigma_{k+1})} \frac{G(T_{\tau_t-1})}{A_{\tau_t}^{(3+\rho)/2}}\mathrm{d}t = \int_0^T \frac{G(T_{\tau_t-1})}{A_{\tau_t}^{(3+\rho)/2}}\mathrm{d}t. \tag{30}$$

Let $a_t = c_*^2\gamma^2 + T - t$. Because of $\triangle V_{\tau_t} \le 2\epsilon_n^2 + 2\delta_n^2$ (cf. Lemma 4), we know that

$$t \le V_{\tau_t} = V_{\tau_t-1} + \triangle V_{\tau_t} \le t + 2\epsilon_n^2 + 2\delta_n^2, \quad t \in [0,T]. \tag{31}$$

Assume $c_* \ge 2$, then we have

$$\frac{1}{2}a_t \le A_{\tau_t} = c_*^2\gamma^2 + T - V_{\tau_t} \le a_t, \quad t \in [0,T]. \tag{32}$$

Note that $G(z)$ is symmetric and is non-increasing in $z \ge 0$. The last bound implies that

$$R_1 \le 2^{(3+\rho)/2}\epsilon_n^{1+\rho}\int_0^T \frac{1}{a_t^{(3+\rho)/2}}\mathbf{E}\left[G\left(\frac{u - X_{\tau_t-1}}{a_t^{1/2}}\right)\right]\mathrm{d}t. \tag{33}$$

Note also that $G(z)$ is a symmetric integrable function of bounded variation. By Lemma 6, it is obvious that

$$\mathbf{E}\left[G\left(\frac{u - X_{\tau_t-1}}{a_t^{1/2}}\right)\right] \le c_6\sup_z\left|\mathbf{P}\left(X_{\tau_t-1} \le z\right) - \Phi(z)\right| + c_7\sqrt{a_t}. \tag{34}$$

Because of $c_* \ge 2, V_{\tau_t-1} = V_{\tau_t} - \triangle V_{\tau_t},\ V_{\tau_t} \ge t$ and $\triangle V_{\tau_t} \le 2\epsilon_n^2 + 2\delta_n^2$, we obtain

$$V_n - V_{\tau_t-1} = V_n - V_{\tau_t} + \triangle V_{\tau_t} \le 2\epsilon_n^2 + 2\delta_n^2 + T - t \le a_t. \tag{35}$$

Therefore

$$\begin{aligned}
\mathbf{E}\left[\left(X_n - X_{\tau_t-1}\right)^2\middle|\mathscr{F}_{\tau_t-1}\right] &= \mathbf{E}\left[\sum_{k=\tau_t}^{n}\mathbf{E}\left[\xi_k^2\middle|\mathscr{F}_{k-1}\right]\middle|\mathscr{F}_{\tau_t-1}\right] \\
&= \mathbf{E}\left[\langle X\rangle_n - \langle X\rangle_{\tau_t-1}\middle|\mathscr{F}_{\tau_t-1}\right] \\
&\le \mathbf{E}[V_n - V_{\tau_t-1}|\mathscr{F}_{\tau_t-1}] \\
&\le a_t.
\end{aligned}$$

Then, by Lemma 5, we deduce that for any $t \in [0, T]$,

$$\sup_z \left|\mathbf{P}\left(X_{\tau_t - 1} \le z\right) - \Phi(z)\right| \le c_8 \sup_z \left|\mathbf{P}\left(X_n \le z\right) - \Phi(z)\right| + c_9 \sqrt{a_t}. \tag{36}$$

Combining (28), (33), (34) and (36) together, we get

$$|J_1| \le c_{10} \epsilon_n^{1+\rho} \int_0^T \frac{1}{a_t^{(3+\rho)/2}} \mathrm{d}t \sup_z \left|\mathbf{P}\left(X_n \le z\right) - \Phi(z)\right| + c_{11} \epsilon_n^{1+\rho} \int_0^T \frac{1}{a_t^{1+\rho/2}} \mathrm{d}t. \tag{37}$$

Taking some elementary computations, it follows that

$$\int_0^T \frac{1}{a_t^{(3+\rho)/2}} \mathrm{d}t = \int_0^T \frac{1}{(c_*^2 \gamma^2 + T - t)^{(3+\rho)/2}} \mathrm{d}t \le \frac{2}{c_*^{1+\rho} (1+\rho) \gamma^{1+\rho}} \tag{38}$$

and

$$\int_0^T \frac{1}{a_t^{1+\rho/2}} \mathrm{d}t = \int_0^T \frac{1}{(c_*^2 \gamma^2 + T - t)^{1+\rho/2}} \mathrm{d}t \le \frac{2}{c_*^{\rho} \rho \gamma^{\rho}}. \tag{39}$$

This yields

$$\left|J_1\right| \le \frac{c_{12}}{c_*^{1+\rho}} \sup_z \left|\mathbf{P}\left(X_n \le z\right) - \Phi(z)\right| + \frac{c_{\rho,1} \epsilon_n}{\rho}. \tag{40}$$

Control of J_2. Since $0 \le \triangle V_k - \triangle \langle X \rangle_k \le 2\delta^2 \mathbf{1}_{\{k=n\}}$, we have

$$|J_2| \le \mathbf{E}\left[\frac{1}{2A_n} \left|\varphi'(T_{n-1})(\triangle V_n - \triangle \langle X \rangle_n)\right|\right].$$

Denote $\widetilde{G}(z) = \sup_{|z-t| \le 1} |\varphi'(t)|$, and then $|\varphi'(z)| \le \widetilde{G}(z)$ for any real z. Since $A_n = c_*^2 \gamma^2$, then we get the following estimation:

$$|J_2| \le \frac{1}{c_*^2} \mathbf{E}\left[\widetilde{G}(T_{n-1})\right].$$

Note that $\widetilde{G}$ is non-increasing in $z \ge 0$, and thus it has bounded variation on $\mathbf{R}$. By Lemma 6, we get

$$|J_2| \le \frac{c_{13}}{c_*^2} \sup_z \left|\mathbf{P}\left(X_{n-1} \le z\right) - \Phi(z)\right| + c_{*,2}(\epsilon_n + \delta_n). \tag{41}$$

Then, by Lemma 5, we deduce that

$$\sup_z \left|\mathbf{P}\left(X_{n-1} \le z\right) - \Phi(z)\right| \le c_{14} \sup_z \left|\mathbf{P}\left(X_n \le z\right) - \Phi(z)\right| + c_{15} \epsilon_n. \tag{42}$$

This yields

$$|J_2| \le \frac{c_{16}}{c_*^2} \sup_z \left|\mathbf{P}\left(X_n \le z\right) - \Phi(z)\right| + c_{\rho,2}(\epsilon_n + \delta_n). \tag{43}$$

Control of J_3. By a two-term Taylor expansion, it follows that

$$|J_3| = \frac{1}{8} \mathbf{E}\left[\sum_{k=1}^n \frac{1}{(A_k - \vartheta_k \triangle A_k)^2} \varphi'''\left(\frac{u - X_{k-1}}{\sqrt{A_k - \vartheta_k \triangle A_k}}\right) (\triangle A_k)^2\right].$$

Note that $c_* \ge 2, \triangle A_k \le 0$ and, by Lemma 4, $|\triangle A_k| = \triangle V_k \le 2\epsilon_n^2 + 2\delta_n^2$. We obtain

$$A_k \le A_k - \vartheta_k \triangle A_k \le c_*^2 \gamma^2 + T - V_k + 2\epsilon_n^2 + 2\delta_n^2 \le 2A_k. \tag{44}$$

Denote $\widehat{G}(z) = \sup_{|t-z| \le 2} |\varphi'''(t)|$. Then $\widehat{G}(z)$ is symmetric, and is non-increasing in $z \ge 0$. Using (44), we get

$$|J_3| \le (2\epsilon_n^2 + 2\delta_n^2) \mathbf{E}\left[\sum_{k=1}^n \frac{1}{A_k^2} \widehat{G}\left(\frac{T_{k-1}}{\sqrt{2}}\right) \triangle V_k\right]. \tag{45}$$

By an argument similar to that of (40), we get

$$
\begin{aligned}
|J_3| &\le \frac{c_{17}(2\epsilon_n^2+2\delta_n^2)}{c_*^2\gamma^2}\sup_z\big|\mathbf{P}\big(X_n\le z\big)-\Phi(z)\big|+\frac{2c_{18}(2\epsilon_n^2+2\delta_n^2)}{c_*\gamma}\\
&\le \frac{c_{19}}{c_*^2}\sup_z\big|\mathbf{P}\big(X_n\le z\big)-\Phi(z)\big|+\frac{4c_{18}(\epsilon_n+\delta_n)^2}{c_*\gamma}\\
&\le \frac{c_{19}}{c_*^2}\sup_z\big|\mathbf{P}\big(X_n\le z\big)-\Phi(z)\big|+c_{\rho,3}(\epsilon_n+\delta_n).
\end{aligned}
\tag{46}
$$

Combining (17), (40), (43) and (46) together, we get

$$
\big|\mathbf{E}\big[\Phi_u\big(X_n,A_n\big)\big]-\mathbf{E}\big[\Phi_u\big(X_0,A_0\big)\big]\big|\le\frac{c_{20}}{c_*^{1+\rho}}\sup_z\big|\mathbf{P}\big(X_n\le z\big)-\Phi(z)\big|+\frac{\widehat{c}_\rho}{\rho}(\epsilon_n+\delta_n),
$$

By (15), we know that

$$
\sup_z\big|\mathbf{P}\big(X_n\le z\big)-\Phi(z)\big|\le\frac{c_{21}}{c_*^{1+\rho}}\sup_z\big|\mathbf{P}\big(X_n\le z\big)-\Phi(z)\big|+\frac{\widetilde{c}_\rho}{\rho}(\epsilon_n+\delta_n),
$$

from which, choosing $c_*^{1+\rho}=\max\{2c_{21},2^{1+\rho}\}$, we get

$$
\sup_z\big|\mathbf{P}\big(X_n\le z\big)-\Phi(z)\big|\le\frac{2\widetilde{c}_\rho(\epsilon_n+\delta_n)}{\rho}.
\tag{47}
$$

3.3. *Proof of Theorem 2*

Following the method of Bolthausen [2], we enlarge the sequence $(\xi_i,\mathscr{F}_i)_{1\le i\le n}$ to $\big(\widehat{\xi}_i,\widehat{\mathscr{F}}_i\big)_{1\le i\le N}$ such that $\langle\widehat{X}\rangle_N:=\sum_{i=1}^N\mathbf{E}\big[\widehat{\xi}_i^2|\widehat{\mathscr{F}}_{i-1}\big]=1$ a.s., and then apply Theorem 1 to the enlarged sequence. Consider the stopping time

$$
\tau=\sup\{k\le n:\langle X\rangle_k\le 1\}.
\tag{48}
$$

Assume that $0\le\varepsilon\le\epsilon_n$. Let $r=\big\lfloor\frac{1-\langle X\rangle_\tau}{\varepsilon^2}\big\rfloor$, where $\lfloor x\rfloor$ denotes the "integer part" of x. It is easy to see that $r\le\big\lfloor\frac{1}{\varepsilon^2}\big\rfloor$. Set $N=n+r+1$. Let $(\zeta_i)_{i\ge1}$ be a sequence of independent Rademacher random variables, which is independent of the martingale differences $(\xi_i)_{1\le i\le n}$. Consider the random variables $\big(\widehat{\xi}_i,\widehat{\mathscr{F}}_i\big)_{1\le i\le N}$ defined as follows:

$$
\widehat{\xi}_i=\begin{cases}\xi_i\ \text{ a.s.,} & \text{if } i\le\tau,\\ \varepsilon\zeta_i\ \text{ a.s.,} & \text{if } \tau+1\le i\le\tau+r,\\ \big(1-\langle X\rangle_\tau-r\varepsilon^2\big)^{1/2}\zeta_i\ \text{ a.s.,} & \text{if } i=\tau+r+1,\\ 0\ \text{ a.s.,} & \text{if } \tau+r+1\le i\le N,\end{cases}
$$

and $\widehat{\mathscr{F}}_i=\sigma\big(\widehat{\xi}_1,\widehat{\xi}_2,\dots,\widehat{\xi}_i\big)$.

Clearly, $\big(\widehat{\xi}_i,\widehat{\mathscr{F}}_i\big)_{1\le i\le N}$ still forms a martingale difference sequence with respect to the enlarged filtration. Then $\widehat{X}_k=\sum_{i=1}^k\widehat{\xi}_i$, $k=0,\dots,N$, with $\widehat{X}_0=0$, is also a martingale. Moreover, it holds that $\langle\widehat{X}\rangle_N=1$, $\mathbf{E}\big[\widehat{\xi}_i^3|\widehat{\mathscr{F}}_{i-1}\big]=0$ and

$$
\mathbf{E}\big[\big|\widehat{\xi}_i\big|^{3+\rho}\big|\widehat{\mathscr{F}}_{i-1}\big]\le\epsilon_n^{1+\rho}\mathbf{E}\big[\widehat{\xi}_i^2\big|\widehat{\mathscr{F}}_{i-1}\big],\qquad\text{a.s.}
$$

By Theorem 1, we have

$$
D\big(\widehat{X}_N\big)\le\frac{c_\rho\epsilon_n}{\rho}.
\tag{49}
$$

Using Lemma 7, we obtain that

$$
D(X_n)\le 2D\big(\widehat{X}_N\big)+3\big\|\mathbf{E}\big[\big|X_n-\widehat{X}_N\big|^{2p}\big|\widehat{X}_N\big]\big\|_1^{1/(2p+1)}\le\frac{2c_\rho\epsilon_n}{\rho}+3\big(\mathbf{E}\big[\big|\widehat{X}_N-X_n\big|^{2p}\big]\big)^{1/(2p+1)}.
\tag{50}
$$

Since τ is a stopping time and

$$\widehat{X}_N - X_n = \sum_{i=\tau+1}^{N} \big(\widehat{\xi}_i - \xi_i\big), \qquad \text{where put } \xi_i = 0 \text{ for } i > n, \tag{51}$$

$(\widehat{\xi}_i - \xi_i, \widehat{\mathscr{F}}_i)_{i\geq\tau+1}$ still forms a martingale difference sequence. Applying Theorem 2.11 of Hall and Heyde [8], we get

$$\begin{aligned}\mathbf{E}\big[\big|\widehat{X}_N - X_n\big|^{2p}\big] &\leq \mathbf{E}\Big[\max_{\tau+1\leq i\leq N}\big|\widehat{X}_i - X_i\big|^{2p}\Big]\\ &\leq c_p\bigg(\mathbf{E}\bigg[\bigg|\sum_{i=\tau+1}^{N}\mathbf{E}\big[\big(\widehat{\xi}_i-\xi_i\big)^2\big|\widehat{\mathscr{F}}_{i-1}\big]\bigg|^p\bigg] + \mathbf{E}\Big[\max_{\tau+1\leq i\leq N}\big|\widehat{\xi}_i - \xi_i\big|^{2p}\Big]\bigg).\end{aligned} \tag{52}$$

As $\mathbf{E}[\xi_i\widehat{\xi}_i|\widehat{\mathscr{F}}_{i-1}] = 0$ for all $i \geq \tau+1$, we have

$$\sum_{i=\tau+1}^{N}\mathbf{E}\big[\big(\widehat{\xi}_i-\xi_i\big)^2\big|\widehat{\mathscr{F}}_{i-1}\big] = \sum_{i=\tau+1}^{N}\mathbf{E}\big[\widehat{\xi}_i^2\big|\widehat{\mathscr{F}}_{i-1}\big] + \sum_{i=\tau+1}^{n}\mathbf{E}\big[\xi_i^2\big|\widehat{\mathscr{F}}_{i-1}\big] = 1 - 2\langle X\rangle_\tau + \langle X\rangle_n.$$

Noting that $1-\mathbf{E}[\xi_{\tau+1}^2|\mathscr{F}_\tau] \leq \langle X\rangle_\tau$. Consequently, using the inequality $|a+b|^p \leq 2^{p-1}(|a|^p+|b|^p)$, $p\geq 1$, and Jensen's inequality, we derive that

$$\begin{aligned}\bigg|\sum_{i=\tau+1}^{N}\mathbf{E}\big[\big(\widehat{\xi}_i-\xi_i\big)^2\big|\widehat{\mathscr{F}}_{i-1}\big]\bigg|^p &\leq \big|\langle X\rangle_n - 1 + 2\mathbf{E}\big[\xi_{\tau+1}^2\big|\mathscr{F}_\tau\big]\big|^p\\ &\leq 2^{2p-1}\big(\big|\langle X\rangle_n - 1\big|^p + \big|\mathbf{E}\big[\xi_{\tau+1}^2\big|\mathscr{F}_\tau\big]\big|^p\big)\\ &\leq 2^{2p-1}\big(\big|\langle X\rangle_n - 1\big|^p + \mathbf{E}\big[|\xi_{\tau+1}|^{2p}\big|\mathscr{F}_\tau\big]\big).\end{aligned} \tag{53}$$

Taking expectations on both sides of the last inequality, we deduce that

$$\begin{aligned}\mathbf{E}\bigg[\bigg|\sum_{i=\tau+1}^{N}\mathbf{E}\big[\big(\widehat{\xi}_i-\xi_i\big)^2\big|\widehat{\mathscr{F}}_{i-1}\big]\bigg|^p\bigg] &\leq 2^{2p-1}\big(\mathbf{E}\big[\big|\langle X\rangle_n-1\big|^p\big] + \mathbf{E}\big[|\xi_{\tau+1}|^{2p}\big]\big)\\ &\leq 2^{2p-1}\Big(\mathbf{E}\big[\big|\langle X\rangle_n-1\big|^p\big] + \mathbf{E}\Big[\max_{1\leq i\leq n}|\xi_i|^{2p}\Big]\Big).\end{aligned} \tag{54}$$

Similarly, using the inequality $|a+b|^p \leq 2^{p-1}(|a|^p+|b|^p)$, $p\geq 1$,

$$\begin{aligned}\mathbf{E}\Big[\max_{\tau+1\leq i\leq N}\big|\widehat{\xi}_i-\xi_i\big|^{2p}\Big] &\leq 2^{2p-1}\mathbf{E}\Big[\max_{\tau+1\leq i\leq N}\big(|\xi_i|^{2p} + \big|\widehat{\xi}_i\big|^{2p}\big)\Big]\\ &\leq 2^{2p-1}\Big(\mathbf{E}\Big[\max_{1\leq i\leq n}|\xi_i|^{2p}\Big] + \varepsilon^{2p}\Big).\end{aligned} \tag{55}$$

Combining (52), (54) and (55) together, we obtain

$$\mathbf{E}\big[\big|\widehat{X}_N - X_n\big|^{2p}\big] \leq \widehat{c}_p\Big(\mathbf{E}\big[\big|\langle X\rangle_n - 1\big|^p\big] + \mathbf{E}\Big[\max_{1\leq i\leq n}|\xi_i|^{2p}\Big] + \varepsilon^{2p}\Big). \tag{56}$$

Finally, applying the last inequality to (50) and let $\varepsilon\to 0$, then we have

$$D(X_n) \leq \widetilde{c}_\rho\frac{\epsilon_n}{\rho} + \widetilde{c}_p\Big(\mathbf{E}\big[\big|\langle X\rangle_n - 1\big|^p\big] + \mathbf{E}\Big[\max_{1\leq i\leq n}|\xi_i|^{2p}\Big]\Big)^{1/(2p+1)}.$$

This completes the proof of Theorem 2.

References

[1] A. Beknazaryan, H. Sang, Y. Xiao, "Cramer type moderate deviations for random fields", *J. Appl. Probab.* **56** (2019), no. 1, p. 223-245.

[2] E. Bolthausen, "Exact convergence rates in some martingale central limit theorems", *Ann. Probab.* **10** (1982), p. 672-688.

[3] M. El Machkouri, L. Ouchti, "Exact convergence rates in the central limit theorem for a class of martingales", *Bernoulli* **13** (2007), no. 4, p. 981-999.

[4] X. Fan, "Exact rates of convergence in some martingale central limit theorems", *J. Math. Anal. Appl.* **469** (2019), no. 2, p. 1028-1044.

[5] T. Fortune, M. Peligrad, H. Sang, "A local limit theorem for linear random fields", https://arxiv.org/abs/2007.05036, submitted, 2020.

[6] I. Grama, E. Haeusler, "Large deviations for martingales via Cramér's method", *Stochastic Processes Appl.* **85** (2000), no. 2, p. 279-293.

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

[8] P. G. Hall, C. C. Heyde, *Martingale limit theory and its applications*, Probability and Mathematical Statistics, Academic Press Inc., 1980.

[9] J. R. M. Hosking, "Fractional differencing", *Biometrika* **68** (1981), p. 165-176.

[10] L. V. Kir'yanova, V. I. Rotar, "Estimates for the rate of convergence in the central limit theorem for martingales", *Theory Probab. Appl.* **36** (1991), no. 2, p. 289-302.

[11] J.-C. Mourrat, "On the rate of convergence in the martingale central limit theorem", *Bernoulli* **19** (2013), no. 2, p. 633-645.

[12] L. Ouchti, "On the rate of convergence in the central limit theorem for martingale difference sequences", *Ann. Inst. Henri Poincaré, Probab. Stat.* **41** (2005), no. 1, p. 35-43.

[13] J. Renz, "A note on exact convergence rates in some martingale central limit theorems", *Ann. Probab.* **24** (1996), no. 3, p. 1616-1637.

[14] W. B. Wu, W. Min, "On linear processes with dependent innovations", *Stochastic Processes Appl.* **115** (2005), no. 6, p. 939-958.