OPTIMAL GAUSSIAN APPROXIMATION FOR MULTIPLE TIME SERIES

By Sayar Karmakar and Wei Biao Wu

University of Chicago

Abstract. We obtain an optimal bound for Gaussian approximation of a large class of vector-valued random processes. Our results substantially generalize earlier ones which assume independence and/or stationarity. Based on the decay rate of functional dependence measure, we quantify the error bound of the Gaussian approximation which can range from the worst $n^{1/2}$ to the optimal $n^{1/p}$ rate.

Key Words and Phrases: Functional central limit theorem, Functional dependence measure, Gaussian Approximation, Weak dependence.

1. Introduction Functional central limit theorem (FCLT) or invariance principle plays an important role in statistics. Let $X_i, i \geq 1$, be independent and identically distributed (i.i.d.) random vectors in $\mathbb{R}^d$ with mean 0 and covariance matrix Σ, and $S_j = \sum_{i=1}^{j} X_i$. The FCLT asserts that

$$\{n^{-1/2} S_{\lfloor nu \rfloor},\, 0 \leq u \leq 1\} \Rightarrow \{\Sigma^{1/2} I\!B(u),\, 0 \leq u \leq 1\}, \tag{1.1}$$

where $\lfloor t \rfloor = \max\{i \in \mathbb{Z} : i \leq t\}$ and $I\!B$ is the standard Brownian motion in $\mathbb{R}^d$, namely it has independent increments and $I\!B(u+v) - I\!B(u) \sim N(0, vI_d)$, $u, v \geq 0$.

In this paper we should substantially generalize (1.1) by developing a convergence rate of (1.1) for multiple time series which can be dependent and non-identically distributed.

The invariance principle was introduced by Erdös and Kac (1946, [9]). Doob (1949, [4]), Donsker (1952, [3]) and Prohorov (1956, [20]) furthered their ideas, which led to the theory of weak convergence of probability measures. There is an extensive literature concerning Gaussian approximation when the dimension $d = 1$. In this case optimal rates for independent random variables were obtained in [11] and [21], among others. When $d = 1$ and X_i are i.i.d. with mean 0, variance σ^2 and have finite p-th moment, $p > 2$, Komlós, Major and Tusnády (1975, 76, [11, 12]) established the very deep result

$$\max_{1 \le i \le n} |S'_i - \sigma B(i)| = o_{\text{a.s.}}(\tau_n), \tag{1.2}$$

where $B(\cdot)$ is the standard Brownian motion and S'_n is constructed on a richer space such that $(S_i)_{i \le n} \stackrel{D}{=} (S'_i)_{i \le n}$, and the approximation rate $\tau_n = n^{1/p}$ is optimal. Results of type (1.2) have many applications in statistics since one can use functionals involving Gaussian processes to approximate statistics of $(X_i)_{i=1}^n$ and thus exploit properties of Gaussian processes. Their result was generalized to independent random vectors by Einmahl (1987a, [6]; 1987b, [7]; 1989, [8]), Zaitsev (2001, [30]; 2002a, [31]; 2002b, [32]) and Götze and Zaitsev (2008, [10]), where optimal and nearly optimal results were obtained.

To generalize (1.2) to multiple time series, we shall consider the possibly non-stationary, d-dimensional, mean 0, vector-valued process

$$X_i = (X_{i1}, \dots, X_{id})^T = H_i(\mathcal{F}_i) = H_i(\epsilon_i, \epsilon_{i-1}, \dots), \quad i \in \mathbb{Z}, \tag{1.3}$$

where ^{T} denotes matrix transpose, $\mathcal{F}_i = (\epsilon_i, \epsilon_{i-1}, \dots)$ and $\epsilon_i, i \in \mathbb{Z}$, are i.i.d. random variables. Here, $H_i(\cdot)$ is a measurable function so that X_i is well-defined. We allow H_i to be possibly non-linear in its argument $(\epsilon_i, \epsilon_{i-1}, \dots)$ to capture a much larger class of processes. If $H_i(\cdot) \equiv H(\cdot)$ does not depend on i, (1.3) defines a stationary causal process. The latter framework is very general; see [23, 25, 19] among others. When $d = 1$, Wiener [24] considered representing stationary processes by functionals of i.i.d. random variables.

Lütkepohl [16] presented many applications of functional central limit theorems for multiple time series analysis. Wu and Zhao (2007, [28]) and Zhou and Wu (2010, [33]) applied Gaussian approximation results with sub-optimal approximation rates to trend estimation and functional regression models. For the class of weakly dependent processes (1.3), we shall show that there exists a probability space (Ω_c, A_c, P_c), on which we can define random vectors X_i^c with the partial sum process $S_i^c = \sum_{t=1}^{i} X_t^c$ and a Gaussian process $G_i^c = \sum_{t=1}^{i} Y_t^c$ with Y_t^c being mean 0 independent Gaussian vectors such that $(S_i^c)_{1 \le i \le n} \stackrel{D}{=} (S_i)_{1 \le i \le n}$ and

$$\max_{i \le n} |S_i^c - G_i^c| = o_P(\tau_n) \quad \text{in } (\Omega_c, A_c, P_c), \tag{1.4}$$

where the approximation bound τ_n is related with the dependence decaying rates. Our result is useful for asymptotic inference for multiple time series. As a primary contribution, we generalize and improve the existing results for Gaussian approximations in several directions. For some $p > 2$, we assume uniform integrability of pth moment and obtain an approximation bound τ_n in terms of p and the decay rate of functional dependence measure. In particular, if the dependence decays fast enough, for τ_n, we are able to achieve the optimal $o_P(n^{1/p})$ bound. In the current literature, optimal results were obtained for some special cases only. We start with a brief overview of them.

For stationary processes with $d = 1$, a sub-optimal rate was derived in Wu (2007, [26]) where the martingale approximation is applied. Berkes, Liu and Wu (2014, [2]) considered causal stationary process (1.3) above and obtained the $n^{1/p}$ bound for $p > 2$. It is considerably more challenging to deal with vector-valued processes. Eberlein (1986, [5]) obtained a Gaussian approximation result for dependent random vectors with approximation error $O(n^{1/2-\kappa})$, for some small $\kappa > 0$. The latter bound can be too crude for many statistical applications. The martingale approximation approach in [26] can not be applied to vector-valued processes since Strassen's embedding generally fails for vector-valued martingales [17]. For a stationary multiple time series with additional constraints, Liu and Lin (2009, [13]) obtained an important result on strong invariance principles for stationary processes with bounds of the order $n^{1/p}$ with $2 < p < 4$. Wu and Zhou (2011, [29]) obtained sub-optimal rates for a multiple non-stationary time series. A critical limitation of the result by [29, 13] was the restriction $2 < p < 4$. It is an open problem on whether the bound $n^{1/p}$ can be achieved when $p \geq 4$.

In this paper, we show that under proper decaying conditions on functional dependence measures for the process (1.3), we can indeed obtain the optimal bound $n^{1/p}$ for $p \geq 4$. Our condition is stated in the form of (2.3), which involves two parameters χ and A to formulate the temporal dependence of the process. With proper conditions on A, we find optimal $\tau_n = \tau_n(\chi)$ for a general $\chi > 0$. In Corollary 2.1 in Berkes, Liu and Wu (2014, [2]) the authors discussed univariate and stationary processes. However, their focus was on larger values of χ that allows them to obtain $\tau_n = n^{1/p}$. In Theorem 2.1 we obtain a rate for any $\chi > 0$ and show that if χ increases from 0 to a certain number χ_0, we obtain the optimal τ_n varying from the worst, $n^{1/2}$, to the optimal, $n^{1/p}$. This work is substantially useful for processes where dependence does not decay fast enough. For the borderline case $\chi = \chi_0$, we can have $o_P(n^{1/p})$ rate for $2 < p < 4$ and for $p \geq 4$ we have $o_P(n^{1/p} \log n)$ rate. However, if $\chi > \chi_0$ we can obtain the optimal $o_P(n^{1/p})$ bound for all $p > 2$.

Our sharp Gaussian approximation result is quite useful for simultaneous inference of curves where the unknown function is not even Lipschitz continuous. There is a huge literature of curve estimation assuming smooth or regular behavior of a function but not so much for functions that are not differentiable or not Lipschitz continuous. Our Gaussian approximation can play a key role in weakening the smoothness assumption and thus enlarging the scope of doing statistical inference. Some applications are mentioned in Karmakar and Wu (2017). Moreover, since the optimal

$o_P(n^{1/p})$ bound for $2 < p < 4$ and stationary processes obtained in [13] has remained a popular choice over the past few years for a multivariate Gaussian approximation, we can apply our sharper invariance principle that generalize ([13])'s one in multiple directions and give optimal rates when $p \geq 4$.

The rest of the article is organized as follows. In Section 2, we introduce the functional dependence measure and present the main result. Theorem 2.1 is proved in Sections 4 and 5. The proof of Theorem 2.2 is given in Section 6. Applications to covariance processes and locally stationary processes are given in Section 3. In Section 4 we discuss the proof strategy briefly to give the readers a basic idea of our long and involved derivation. Some useful results are collected in Section 7.

We now introduce some notation. For a random vector Y, write $Y \in \mathcal{L}_p, p > 0$, if $\|Y\|_p := E(|Y|^p)^{1/p} < \infty$. If $Y \in \mathcal{L}_2$, $Var(Y)$ denotes the covariance matrix. For $\mathcal{L}_2$ norm write $\|\cdot\| = \|\cdot\|_2$. Throughout the text, we use c_p for constants that depend only on p and c for universal constants. These might take different values in different lines unless otherwise specified. For two positive sequences a_n and b_n, if $a_n/b_n \to 0$ (resp. $a_n/b_n \to \infty$), write $a_n \ll b_n$ (resp. $a_n \gg b_n$). Write $a_n \lesssim b_n$ if $a_n \leq cb_n$ for some $c < \infty$. The d-variate normal distribution with mean μ and covariance matrix Σ is denoted by $N(\mu, \Sigma)$. Denote by I_d the $d \times d$ identity matrix. For a matrix $A = (a_{ij})$, we define its Frobenius norm as $|A| = (\sum a_{ij}^2)^{1/2}$. For a positive semi-definite matrix A with spectral decomposition $A = QDQ^T$, where Q is orthonormal and $D = (\lambda_1, \dots, \lambda_d)$ with $\lambda_1 \geq \dots \geq \lambda_d$, write the Grammian square root $A^{1/2} = QD^{1/2}Q^T$, $\rho_*(A) = \lambda_d$ and $\rho^*(A) = \lambda_1$.

2. Main Results We first introduce uniform functional dependence measure on the underlying process using the idea of coupling. Let $\epsilon'_i, \epsilon_j, i, j \in \mathbb{Z}$, be i.i.d. random variables. Assume $X_i \in \mathcal{L}_p, p > 0$. For $j \geq 0$, $0 < r \leq p$, define the functional dependence measure

$$\delta_{j,r} = \sup_i \|X_i - X_{i,(i-j)}\|_r = \sup_i \|H_i(\mathcal{F}_i) - H_i(\mathcal{F}_{i,(i-j)})\|_r, \tag{2.1}$$

where $\mathcal{F}_{i,(k)}$ is the coupled version of $\mathcal{F}_i$ with ϵ_k in $\mathcal{F}_i$ replaced by an i.i.d. copy ϵ'_k,

$$\mathcal{F}_{i,(k)} = (\epsilon_i, \epsilon_{i-1}, \ldots, \epsilon'_k, \epsilon_{k-1}, \ldots) \text{ and } X_{i,(i-j)} = H_i(\mathcal{F}_{i,(i-j)}).$$

Also, $\mathcal{F}_{i,(k)} = \mathcal{F}_i$ if $k > i$. Note that, $\|H_i(\mathcal{F}_i) - H_i(\mathcal{F}_{i,(i-j)})\|_r$ measures the dependence of X_i on ϵ_{i-j}. Since the physical mechanism function H_i can possibly be different for a non-stationary process, we choose to define the functional dependence measure in an uniform manner. The quantity $\delta_{j,r}$ measures the uniform j-lag dependence in terms of the rth moment. Assume throughout the paper that

$$\Theta_{0,p} = \sum_{i=0}^{\infty} \delta_{i,p} < \infty. \tag{2.2}$$

This condition implies short range dependence in the sense that the cumulative dependence of $(X_j)_{j \geq k}$ on ϵ_k is finite. For presentational clarity, in this paper we assume there exists $\chi > 0, A > 0$ such that the tail cumulative dependence measure

$$\Theta_{i,p} = \sum_{j=i}^{\infty} \delta_{j,p} = O\left(i^{-\chi} (\log i)^{-A}\right). \tag{2.3}$$

Larger χ or A implies weaker dependence. Our Gaussian approximation rate τ_n (cf Theorems 2.1 and 2.2) depends on χ and A. Define functions $f_j(\cdot,\cdot)$ by

$$\begin{aligned} f_1 &= f_1(p,\chi) = p^2\chi^2 + p^2\chi, \;\; f_2 = 2p\chi^2 + 3p\chi - 2\chi, \\ f_3 &= p^3(1+\chi)^2 + 6f_1 + 4p\chi - 2, \;\; f_4 = 2p(2p\chi^2 + 3p\chi + p - 2), \\ f_5 &= p^2(p^2+4p-12)\chi^2 + 2p(p^3+p^2-4p-4)\chi + (p^2-p-2)^2. \end{aligned} \tag{2.4}$$

Assume that the process (1.3) satisfies the uniform integrability and the regularity condition on the covariance structure. The latter is frequently imposed in study of multiple time series.

(2.A) The series $(|X_i|^p)_{i\geq 1}$ is uniformly integrable: $\sup_{i\geq 1} E(|X_i|^p \mathbf{1}_{|X_i|\geq u}) \to 0$ as $u \to \infty$;

(2.B) (Lower bound on eigenvalues of covariance matrices of increment processes) There exists $\lambda_* > 0$ and $l_* \in \mathbb{N}$, such that for all $t \geq 1, l \geq l_*$,

$$\rho_*(Var(S_{t+l} - S_t)) \geq \lambda_* l.$$

Theorem 2.1. *Assume $E(X_i) = 0$, (2.A)-(2.B) and (2.3) holds with*

$$0 < \chi < \chi_0 = \frac{p^2 - 4 + (p-2)\sqrt{p^2+20p+4}}{8p}, \tag{2.5}$$

$$A > \frac{(2p+p^2)\chi + p^2 + 3p + 2 + f_5^{1/2}}{p(1+p+2\chi)}. \tag{2.6}$$

Then (1.4) holds with the approximation bound $\tau_n = n^{1/r}$, where

$$\frac{1}{r} = \frac{f_1 + p^2\chi + p^2 - 2p + f_2 - \chi\sqrt{(p-2)(f_3-3p)}}{f_4}. \tag{2.7}$$

THEOREM 2.2. *Assume $E(X_i) = 0$, (2.A)-(2.B), (2.3). Recall (2.5) for χ_0. (i) If $\chi > \chi_0$, and $A > 0$, we can achieve (1.4) with $\tau_n = n^{1/p}$ for all $p > 2$. For $\chi = \chi_0$, assume that A satisfies (2.6). (ii) If $2 < p < 4$, we have $\tau_n = n^{1/p}$; (iii) if $p \geq 4$, we have $\tau_n = n^{1/p} \log n$.*

Theorems 2.1 and 2.2 concern the two cases $\chi < \chi_0$ and $\chi \geq \chi_0$, respectively, and they are proved in Sections 4 and 6. The proof of Theorem 2.2 requires a more refined treatment so that the optimal rate can be derived. For Theorem 2.1 and Theorem 2.2(i) and (iii), we apply Götze and Zaitsev (2008, [10]); see Proposition 7.1, while for Theorem 2.2(ii), Proposition 1 from Einmahl (1987, [6]) is applied. The expression of r is complicated. Figure 1 plots the power $\max(1/r, 1/p)$. As $\chi \to 0$, $r \to 2$, and $r = p$ if $\chi = \chi_0$.

REMARK 2.3. *The lower bound of A for the case $\chi = \chi_0$ can be further simplified to*

$$A > \frac{p^2 + 8p + 4 + (p-2)\sqrt{p^2 + 20p + 4}}{6p}.$$

3. Applications

3.1. *Covariance Processes:* Assuming that X_i is a vector linear process

$$X_i = \sum_{j=0}^{\infty} B_j \epsilon_{i-j}, \tag{3.1}$$

where B_j are $d \times d$ coefficient matrix, and $\epsilon_i = (\epsilon_{i1}, \ldots, \epsilon_{id})^T$, ϵ_{ir} are i.i.d. random variables with mean 0 and finite qth moment, $q > 4$. Let the $d(d+1)/2$ dimensional

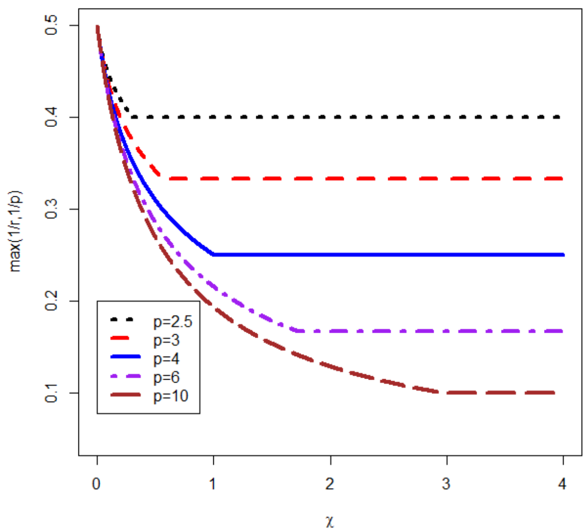


FIG 1. *Optimal bound as a function of* χ

vector $W_i = (X_{ir}X_{is})_{1\le r\le s\le d}$. Then $\bar{W}_n := \sum_{i=1}^n W_i/n$ gives sample covariances of $(X_i)_{i=1}^n$. Assume

$$\sum_{j=t}^{\infty} |B_j| = O(t^{-\chi}(\log t)^{-A}), \tag{3.2}$$

where A satisfies (2.6). Write $p = q/2$. Let $\Sigma = \sum_{k=-\infty}^{\infty} Cov(W_0, W_k)$ be the long-run covariance matrix of (W_i). By Theorems 2.1 and 2.2, we have

$$\max_{i\le n} |i\bar{W}_i - iE(W_1) - \Sigma^{1/2} IB(i)| = o_P(\tau_n), \tag{3.3}$$

where τ_n takes the value $n^{1/r}$ (see (2.7)), and $n^{1/p}$ based on $\chi < \chi_0$ and $\chi > \chi_0$ respectively, IB is a centered standard Brownian motion. Result (3.3) is helpful for change point inference for multiple time series based on covariances; see [1, 22] among others.

3.2. *Nonlinear Non-stationary Time Series:* Consider the process

$$X_i = F(X_{i-1}, \epsilon_i, \theta(i/n)),\ 1 \le i \le n,$$

where ϵ_i are i.i.d. random variables, F is a measurable function, $\theta : [0,1] \to \mathbb{R}$ is a parametric function such that $\max_{0 \le u \le 1} \|F(x_0, \epsilon_i, \theta(u))\|_p < \infty$, and

$$\sup_{0 \le u \le 1} \sup_{x \ne x'} \frac{\|F(x, \epsilon_i, \theta(u)) - F_i(x', \epsilon_i, \theta(u))\|_p}{|x - x'|} < 1. \tag{3.4}$$

Then the process (X_i) satisfies the Geometric moment contraction: for some $0 < \beta < 1$,

$$\delta_{i,p} = O(\beta^i). \tag{3.5}$$

Thus (2.3) holds for any $\chi > 0$ and Theorem 2.2 is applicable with rate $\tau_n = n^{1/p}$. This facilitates inference for the unknown parametric function θ. Time-varying ARCH and GARCH are prominent examples in this large class of models.

4. Proof of Theorem 2.1 The proof of Theorem 2.1 is quite involved. Here we discuss the major components of the proof. Some technical details are given in Section 5.

4.1. *Preparation:-Truncation, m-dependence and blocking approximations:* The first part of our proof consists of series of approximations to create almost independent blocks. The first of them, the truncation approximation will ensure the optimal $n^{1/p}$ bound. Secondly, we use the $m-$dependence approximation for a suitably chosen sequence m_n in terms of the decay rate χ. Lastly, the blocking approximation requires some sharp Rosenthal-type inequality that needs γth moment of the block-sums in the numerator with $\gamma > p$. It is essential to use a power higher than p to obtain a better rate.

To maintain clarity, we defer the exact choice of γ and m_n in terms of χ and A to Subsection 4.4. Instead, in this subsection we come up with conditions (4.9), (4.13) and (4.14) to ensure $n^{1/r}$ rate and solve γ, m_n and r later to obtain the best possible choices for this sequences. Henceforth, we drop the suffix of m_n for our convenience.

4.1.1. *Truncation approximation:* Truncation approximation is necessary to allow higher moments manipulations. Additionally, we need a very slowly converging sequence $t_n \to 0$ based on the uniform integrability condition (2.A). For every $t > 0$, we have

$$\sup_i \frac{1}{t^p} E(|X_i|^p \mathbf{1}_{|X_i| > tn^{1/p}}) = 0 \text{ and } n \sup_i E \min(\frac{|X_i|^\gamma}{t^\gamma n^{\gamma/p}}, 1) \to 0 \text{ as } n \to \infty, \tag{4.1}$$

where $\gamma > p$. The second relation follows from Lemma 7.2. Clearly (4.1) implies that

$$\sup_i \frac{1}{t_n^p} E(|X_i|^p \mathbf{1}_{|X_i| > t_n n^{1/p}}) + n \sup_i E \min(\frac{|X_i|^\gamma}{t_n^\gamma n^{\gamma/p}}, 1) \to 0 \text{ as } n \to \infty, \tag{4.2}$$

holds for a sequence $t_n \to 0$ very slowly. Without loss of generality we can let

$$t_n \log\log n \to \infty \tag{4.3}$$

since otherwise we can replace t_n by $\max(t_n, (\log\log n)^{-1/2})$ (say). For $b > 0$ and $v = (v_1, \ldots, v_d)^T \in \mathbb{R}^d$, define

$$T_b(v) = (T_b(v_1), \ldots, T_b(v_d))^T, \text{ where } T_b(w) = \min(\max(w, -b), b). \tag{4.4}$$

The truncation operator T_b is Lipschitz continuous with Lipschitz constant 1. Let

$$R_{c,l} = \sum_{i=1+c}^{l+c} X_i^{\oplus} = \sum_{i=1+c}^{l+c} [T_{t_n n^{1/p}}(X_i) - ET_{t_n n^{1/p}}(X_i)]. \tag{4.5}$$

PROPOSITION 4.1. *Assume Condition (2.A). For t_n satisfying (4.2), we have*

$$\max_{1\le i\le n} |S_i - S_i^{\oplus}| = o_P(n^{1/p}), \text{ where } S_l^{\oplus} = R_{0,l} = \sum_{i=1}^{l} X_i^{\oplus}. \tag{4.6}$$

PROOF. By (4.2), we have $P(\max_{i\le n} |S_i - \sum_{j=1}^{i} T_{t_n n^{1/p}}(X_j)| = 0) \to 1$ in view of

$$\sup_j P\left(|X_j| > t_n n^{1/p}\right) \le \sup_j \frac{1}{nt_n^p} E\left(|X_j|^p I\left(|X_j| > t_n n^{1/p}\right)\right) = o(1/n).$$

Also by (4.2), $\max_{j\le n} |E(X_j - T_{t_n n^{1/p}}(X_j))| = o(n^{1/p-1})$. Hence (4.6) follows. □

4.1.2. *m-dependence approximation:* The m-dependence approximation is a very important tool that is extensively used in literature; see for example the Gaussian approximation in Liu and Lin (2009, [13]) and Berkes, Liu and Wu (2014, [2]). For a suitably chosen sequence m, we look at the conditional mean $E(X_i|\epsilon_i, \ldots \epsilon_{i-m})$. This gives a very simple yet effective way to handle the original process in terms of a collection of ϵ_i's. As the dependence of X_i and X_{i+k} slowly decrease as k grows, if we can divide the partial sum process in blocks of sufficiently long, their behavior is close to that of a block-independent process. This strategy allows us to apply the existing Gaussian approximation results in the literature suitable for independent process. Define the partial sum process

$$\tilde{R}_{c,l} = \sum_{i=1+c}^{l+c} \tilde{X}_j, \text{ where } \tilde{X}_j = E(T_{t_n n^{1/p}}(X_j)|\epsilon_j, \ldots, \epsilon_{j-m}) - E(T_{t_n n^{1/p}}(X_j)). \tag{4.7}$$

Write $\tilde{R}_{0,i} = \tilde{S}_i$. From Lemma A1 in Liu and Lin (2009, [13]), we have

$$\| \max_{1 \le l \le n} |S_l^{\oplus} - \tilde{S}_l| \|_r \le c_r n^{1/2} \Theta_{1+m,r}. \tag{4.8}$$

The proofs in [13] are for stationary processes. Since our $\delta_{j,r}$ in (2.1) is defined in an uniform manner, the proof goes through for the non-stationary case as well. Assume

$$n^{1/2-1/r} \Theta_{m,r} \to 0. \tag{4.9}$$

By (4.8) and (4.9), we have $n^{1/r}$ convergence in the m-dependence approximation step

$$\max_{1 \le i \le n} |S_i^{\oplus} - \tilde{S}_i| = o_P(n^{1/r}). \tag{4.10}$$

4.1.3. *Blocking approximation:* We now define functional dependence measure for the truncated process $(T_{t_n n^{1/p}}(X_i))_{i \le n}$ as

$$\delta_{j,l}^{\oplus} = \sup_i \| T_{t_n n^{1/p}}(X_i) - T_{t_n n^{1/p}}(X_{i,(i-j)}) \|_l, \text{ where } l \ge 2.$$

Similarly, define the functional dependence measure for the m-dependent process $(\tilde{X}_i)$ as

$$\tilde{\delta}_{j,l} = \sup_i \| \tilde{X}_i - \tilde{X}_{i,(i-j)} \|_l.$$

For these dependence measures, the following inequality holds for all $l \ge 2$:

$$\tilde{\delta}_{j,l} \le \delta_{j,l}^{\oplus} \le \delta_{j,l}. \tag{4.11}$$

Proposition 4.2 gives the blocking approximation result. For $j \geq 0$, define

$$A_{j+1} = \sum_{i=2jk_0m+1}^{(2k_0j+2k_0)m} \tilde{X}_i, \text{ where } k_0 = \lfloor \Theta_{0,2}^2/\lambda_* \rfloor + 2. \tag{4.12}$$

In the blocking approximation step, we shall approximate the partial sum process $\tilde{S}_i$ by sums of A_j. To this end, we will need the following two conditions, for some $\gamma > p$,

$$n^{1-\gamma/r} m^{\gamma/2-1} \to 0, \tag{4.13}$$

$$n^{1/p-1/\gamma} \sum_{j=m+1}^{\infty} \delta_{j,p}^{p/\gamma} \to 0. \tag{4.14}$$

We need another condition for the blocking approximation (see (7.4) in the proof of Lemma 7.4). However, we skip it here and choose m and γ such that conditions (4.9), (4.13) and (4.14) are met. These will automatically imply this fourth one in view of (2.3). We assume an almost polynomial rate for m: for some $0 < L < 1$,

$$m = \lfloor n^L t_n^k \rfloor, \quad 0 < k < (\gamma - p)/(\gamma/2 - 1). \tag{4.15}$$

Proposition 4.2. *Assume (4.13) and (4.14) for some $\gamma > p$. Moreover, assume (4.15) for the m-sequence and (2.3) for the decay rate of $\Theta_{i,p}$ with some $A > \gamma/p$. Then*

$$\max_{1 \leq i \leq n} |\tilde{S}_i - S_i^\diamond| = o_P(n^{1/r}), \text{ where } S_i^\diamond = \sum_{j=1}^{q_i} A_j, \; q_i = \lfloor i/(2k_0 m) \rfloor. \tag{4.16}$$

PROOF. Let $\mathcal{S} = \{2ik_0m, 0 \le i \le q_n\}$ and $\phi_n = (n^{1-\gamma/r}m^{\gamma/2-1})^{1/(2\gamma)}$. Then

$$P\left(\max_{1\le l\le n} |\tilde{R}_{0,l} - \sum_{j=1}^{\lfloor l/(2k_0m)\rfloor} A_j| \ge \phi_n n^{1/r}\right) \le \frac{n}{2k_0m}\max_{c\in\mathcal{S}} P(\max_{1\le l\le 2k_0m} |\tilde{R}_{c,l}| \ge \phi_n n^{1/r})$$
$$\le n\max_{c\in\mathcal{S}} \frac{E(\max_{1\le l\le 2k_0m} |\tilde{R}_{c,l}|^\gamma)}{2k_0m\phi_n^\gamma n^{\gamma/r}} = O(\phi_n^\gamma),$$

from the assumption (4.13) and Lemma 7.4. Since $\phi_n \to 0$, (4.16) follows. □

Summarizing (4.6), (4.10) and (4.17), we can work on $S_i^\diamond$ in view of

$$\max_{1\le i\le n} |S_i - S_i^\diamond| = o_P(n^{1/r}). \tag{4.17}$$

In the next steps, we obtain a Gaussian approximation for $S_n^\diamond$; see backgrounds in Sections 4.2 and 4.3 and detailed argument in Section 5.

4.2. *Conditioning and Gaussian Approximation:* The blocks created in the first steps are not independent because two successive blocks share some ϵ_i's in their shared border. In this second stage, we look at the partial sum process conditioned on these borderline ϵ_i's, which implies conditional independence. Berkes, Liu and Wu (2014, [2]) did a similar treatment with triadic decomposition for stationary scalar processes and applied Sakhanenko's (2006, [21]) Gaussian Approximation result on the conditioned process.

Since the result from Sakhanenko (2006, [21]) is only valid for $d = 1$, we need to use the Gaussian approximation result from Götze and Zaitsev (2008, [10]) (see Proposition 7.1) for $d \ge 2$. This comes with the cost of verifying a very technical sufficient condition on the covariance matrices of the independent vectors. Verification of such a condition is quite involved in our case since we are dealing with the conditional process. We opt for a k-dic decomposition instead of the triadic decomposition in [2]. This is necessary to accommodate the non-stationarity of the process. We need $k_0 > \Theta_{0,2}^2/\lambda_*$ (cf. (4.12)), where λ_* is mentioned in Condition 2.B.

4.3. *Removing the conditioning and regrouping:* In the last part of our proof, we obtain the Gaussian approximation for the unconditional process by applying Proposition 7.1 one more time. In the second part of our proof, we look at the conditional variance (cf $V_j(\bar{a}_{2k_0j}, \bar{a}_{2k_0j+2k_0}) = Var(Y_j(\bar{a}_{2k_0j}, \bar{a}_{2k_0j+2k_0}))$ in (5.3) of Subsection 5.1) of the blocks. These conditional variances are 1-dependent. In order to apply Götze and Zaitsev (2008, [10])'s result, we rearrange the sums of these variances into sums of independent blocks (cf 5.5 in Subsection 5.1). Due to the non-stationarity, this regrouping is completely different and much more involved than Berkes, Liu and Wu (2014, [2]). In particular, the regrouping procedure leads to matrices that may not be positive definite hence cannot be used directly as possible covariance matrices of Gaussian processes. We overcome this obstacle by introducing a positive-definitization that does not affect the optimal rate.

4.4. *Conclusion of the Proof:* This subsection discusses the specific choice of the sequence m, γ and the rate $\tau_n = n^{1/r}$ starting from (4.9), (4.13) and (4.14). Elementary calculations show that $r < p$ for $\chi < \chi_0$. Provided $1 - (\chi + 1)p/\gamma < 0$, we have

$$\sum_{j=m+1}^{\infty} \delta_{j,p}^{p/\gamma} \leq \sum_{i=\lfloor \log_2 m \rfloor}^{\infty} \sum_{j=2^i}^{2^{i+1}-1} \delta_{j,p}^{p/\gamma} \leq \sum_{i=\lfloor \log_2 m \rfloor}^{\infty} 2^{i(1-p/\gamma)} \Theta_{2^i,p}^{p/\gamma} \tag{4.18}$$
$$= \sum_{i=\lfloor \log_2 m \rfloor}^{\infty} 2^{i(1-p/\gamma)} O(2^{-\chi i p/\gamma} i^{-Ap/\gamma}) = O(m^{1-p/\gamma-\chi p/\gamma} (\log m)^{-Ap/\gamma}).$$

By (4.3) and (4.15) $\log m \asymp \log n$. Assume that,

$$1/2 - 1/r - \chi L = 0, \quad A > \gamma/p, \tag{4.19}$$

$$1 - \gamma/r + L(\gamma/2 - 1) = 0, \quad 0 < k < (\gamma/2-1)^{-1}(\gamma - p) \tag{4.20}$$

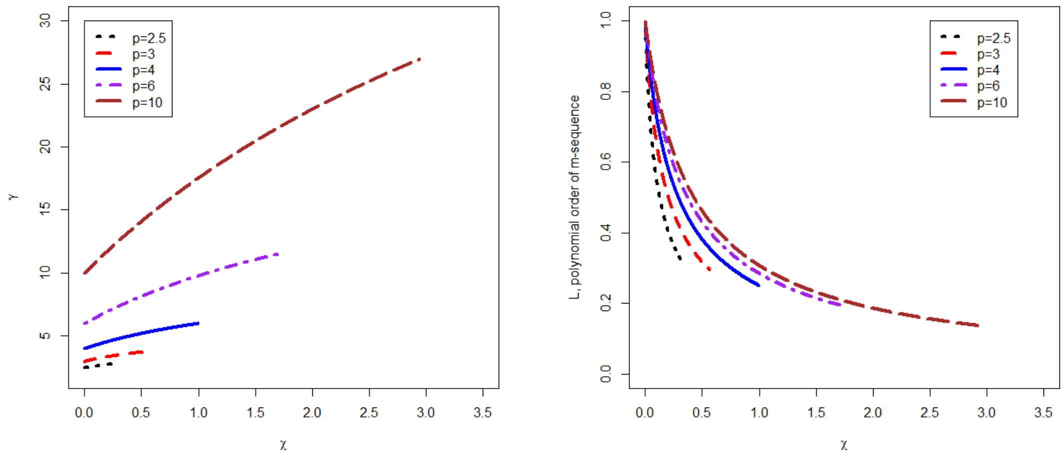
$$1/p - 1/\gamma + (1 - (\chi+1)p/\gamma)L = 0. \tag{4.21}$$

Then conditions (4.9), (4.13) and (4.14) hold. Solving equations in (4.19), (4.20) and (4.21), we obtain r given in (2.7),

$$\gamma = \frac{(2p+p^2)\chi + p^2 + 3p + 2 + f_5^{1/2}}{2 + 2p + 4\chi},$$
$$L = \frac{f_1 - f_2 + \chi\sqrt{(p-2)(f_3 - 3p)}}{\chi f_4},$$

with $f_1, \ldots, f_5$ given in (2.4). Moreover, we specifically choose $A > 2\gamma/p$ for a crucial step in the proof of our Gaussian approximation; see (5.21).

Remark 4.3. *Figure 2 depicts how γ and L change with p and χ for $\chi < \chi_0$. Note that in Figure 2, L, the power of n in the expression of m is close to 1 if χ is small. This is intuitive since if dependence decays very slowly, to make blocks of size m or a multiple of m behave almost independently, one needs a larger L.*



(a) (b)

Fig 2. *(a) γ as a function of χ, (b) L as a function of χ*

5. Remaining Steps of the Proof of Theorem 2.1 In this section we shall provide details of the arguments for steps mentioned in Sections 4.2 and 4.3. Section 5.1 presents the conditional Gaussian approximation, where we shall apply Proposition 7.1 stated in Section 7. Section 5.2 deals with unconditional Gaussian approximation and regrouping.

5.1. *Conditional Gaussian Approximation:* The blocks A_j created in (4.12) after the blocking approximation are weakly independent; except they share some dependence on the border. In this subsection, we look at the conditional process given the ϵ_i the blocks share in their borders. Demeaning the conditional process, we apply the Proposition 7.1 for the Gaussian approximation. For $1 \leq i \leq n$, let $\tilde{H}_i$ be a measurable function such that

$$\tilde{X}_i = \tilde{H}_i(\epsilon_i, \ldots, \epsilon_{i-m}). \tag{5.1}$$

Recall Proposition 4.2 for the definition of q_i. Let $q = q_n$. For $j = 1, \ldots, q$, define

$$\bar{a}_{2k_0 j} = \{a_{(2k_0 j-1)m+1}, \ldots, a_{2k_0 jm}\} \text{ and } a = \{\ldots, \bar{a}_0, \bar{a}_{2k_0}, \bar{a}_{4k_0}, \ldots\}.$$

Given a, define, for $2k_0 jm + 1 \le i \le (2k_0 j + 1)m$,

$$\tilde{X}_i(\bar{a}_{2k_0 j}) = \tilde{H}_i(\epsilon_i, \ldots, \epsilon_{2k_0 jm+1}, a_{2k_0 jm}, \ldots, a_{i-m})$$

and for $(2k_0 j + 2k_0 - 1)m + 1 \le i \le (2k_0 j + 2k_0)m$,

$$\tilde{X}_i(\bar{a}_{2k_0 j+2k_0}) = \tilde{H}_i(a_i, \ldots, a_{(2k_0 j+2k_0-1)m+1}, \epsilon_{(2k_0 j+2k_0-1)m}, \ldots, \epsilon_{i-m}).$$

Further, define the blocks as following,

$$\begin{aligned} F_{4j+1}(\bar{a}_{2k_0 j}) &= \sum_{i=2k_0 jm+1}^{(2k_0 j+1)m} \tilde{X}_i(\bar{a}_{2k_0 j}), \\ F_{4j+2} &= \sum_{i=(2k_0 j+1)m+1}^{(2k_0 j+k_0)m} \tilde{X}_i, \quad F_{4j+3} = \sum_{i=(2k_0 j+k_0)m+1}^{(2k_0 j+2k_0-1)m} \tilde{X}_i, \\ F_{4j+4}(\bar{a}_{2k_0 j+2k_0}) &= \sum_{i=(2k_0 j+2k_0-1)m+1}^{(2k_0 j+2k_0)m} \tilde{X}_i(\bar{a}_{2k_0 j+2k_0}). \end{aligned} \tag{5.2}$$

Similarly, for $j = 1, \ldots, q$, define

$$\bar{\vartheta}_{2k_0 j} = \{\epsilon_{(2k_0 j-1)m+1}, \ldots, \epsilon_{2k_0 jm}\} \text{ and } \vartheta = \{\ldots, \bar{\vartheta}_0, \bar{\vartheta}_{2k_0}, \bar{\vartheta}_{4k_0}, \ldots\}.$$

Recall A_j from (4.12). We have

$$A_{j+1} = F_{4j+1}(\bar{\vartheta}_{2k_0j}) + F_{4j+2} + F_{4j+3} + F_{4j+4}(\bar{\vartheta}_{2k_0j+2k_0}).$$

Define the mean functions

$$\Lambda_{4j+1}(\bar{a}_{2k_0j}) = E^*(F_{4j+1}(\bar{a}_{2k_0j})) \text{ and } \Lambda_{4j+4}(\bar{a}_{2k_0j+2k_0}) = E^*(F_{4j+4}(\bar{a}_{2k_0j+2k_0})),$$

where E^* refers to the conditional moment given a. In the sequel, with slight abuse of notation, we will simply use the usual E to denote moments of random variables conditioned on a. Introduce the centered process

$$\begin{aligned} Y_j(\bar{a}_{2k_0j}, \bar{a}_{2k_0j+2k_0}) &= F_{4j+1}(\bar{a}_{2k_0j}) - \Lambda_{4j+1}(\bar{a}_{2k_0j}) + F_{4j+2} \\ &\quad + F_{4j+3} + F_{4j+4}(\bar{a}_{2k_0j+2k_0}) - \Lambda_{4j+4}(\bar{a}_{2k_0j+2k_0}). \end{aligned} \tag{5.3}$$

Following the definition of $S_n^\diamond$, we let

$$S_i(a) = \sum_{j=0}^{q_i-1} Y_j(\bar{a}_{2k_0j}, \bar{a}_{2k_0j+2k_0}).$$

The mean and variance function of $S_i(a)$ are respectively denoted by

$$\begin{aligned} M_i(a) &= \sum_{j=0}^{q_i-1}[\Lambda_{4j+1}(\bar{a}_{2k_0j}) + \Lambda_{4j+4}(\bar{a}_{2k_0j+2k_0})], \\ Q_i(a) &= \sum_{j=0}^{q_i-1} V_j(\bar{a}_{2k_0j}, \bar{a}_{2k_0j+2k_0}), \end{aligned}$$

where $V_j(\bar{a}_{2k_0j}, \bar{a}_{2k_0j+2k_0})$ is the dispersion matrix of $Y_j(\bar{a}_{2k_0j}, \bar{a}_{2k_0j+2k_0})$. Define

$$V_{j0}(\bar{a}_{2k_0j}) = E(F_{4j-2}F_{4j-1}^T + F_{4j-1}F_{4j-2}^T) + Var(F_{4j-1} + F_{4j}(\bar{a}_{2k_0j}) - \Lambda_{4j}(\bar{a}_{2k_0j}))$$

$$+Var(F_{4j+1}(\bar{a}_{2k_0j}) - \Lambda_{4j+1}(\bar{a}_{2k_0j}) + F_{4j+2}). \quad (5.4)$$

Note that, the following identity holds for all t:

$$\sum_{j=0}^{t} V_j(\bar{a}_{2k_0j}, \bar{a}_{2k_0j+2k_0}) = L(\bar{a}_0) + \sum_{j=1}^{t-1} V_{j0}(\bar{a}_{2k_0j}) + U_t(\bar{a}_{2k_0t+2k_0}), \quad (5.5)$$

where $L(\bar{a}_0) = Var(F_1(\bar{a}_0) + F_2)$ and

$$U_{t-1}(\bar{a}_{2k_0t}) = E(F_{4t-2}F_{4t-1}^T + F_{4t-1}F_{4t-2}^T) + Var(F_{4t-1} + F_{4t}(\bar{a}_{2k_0t}) - \Lambda_{4t}(\bar{a}_{2k_0t})). \quad (5.6)$$

Define

$$L_\gamma^a = \sum_{j=0}^{q-1} E(|Y_j(\bar{a}_{2k_0j}, \bar{a}_{2k_0j+2k_0})|^\gamma).$$

In the sequel, we suppress $Y_j(\bar{a}_{2k_0j}, \bar{a}_{2k_0j+2k_0})$, $Y_j(\bar{\vartheta}_{2k_0j}, \bar{\vartheta}_{2k_0j+2k_0})$, $V_j(\bar{a}_{2k_0j}, \bar{a}_{2k_0j+2k_0})$, $V_{j0}(\bar{a}_{2k_0j})$, $V_j(\bar{\vartheta}_{2k_0j}, \bar{\vartheta}_{2k_0j+2k_0})$ and $V_{j0}(\bar{\vartheta}_{2k_0j})$ as just Y_j^a,Y_j^ϑ, V_j^a, V_{j0}^a, V_j^ϑ and V_{j0}^ϑ respectively. We apply Proposition 7.1 to the independent mean 0 random vectors Y_j^a. We need to find a suitable sequence η_k that allows us to get constants C_1, C_2 in (7.1) and C_3 in (7.2). There are roughly $q = n/(2k_0m)$ many Y_j^a random variables. Define

$$l = \lfloor q^{2/\gamma}/\log^2 q \rfloor. \quad (5.7)$$

To apply Proposition 7.1, we choose the sequence $\eta_k = kl$ and $s \asymp q/l$. This choice is justified by proving the following series of propositions.

Proposition 5.1. *Recall λ_* and A_j from (2.B) and (4.12) respectively. There exists a constant $\delta > 0$ such that*

$$2(\lambda_* + \delta)k_0m \le \rho_*(Var(A_j)) \le \rho^*(Var(A_j)) \le \|A_j\|^2 \le 2k_0m\Theta_{0,2}^2.$$

Proposition 5.2. *We can get positive constants c_1 and c_2 such that for all j,*

$$c_1 m \le \rho_*(Var(Y_j^\vartheta)) \le \rho^*(Var(Y_j^\vartheta)) \le E(|Y_j^\vartheta|^2) \le c_2 m. \tag{5.8}$$

Proposition 5.3. *For l in (5.7), there exists constant c_3 such that,*

$$P\left(\max_{1\le t\le q/l} |Var\left(\sum_{j=(t-1)l}^{tl-1} Y_j^a\right) - E\left(Var\left(\sum_{j=(t-1)l}^{tl-1} Y_j^a\right)\right)| \ge c_3 lm\right) \to 0.$$

Proposition 5.4. *We can get constants c_4 and c_5 such that*

$$P(c_4 q^{2/\gamma} m \le (L_\gamma^a)^{2/\gamma} \le c_5 q^{2/\gamma} m) \to 1.$$

Proposition 5.5. *Choose $\eta_k = kl$ with l being defined in (5.7). Then we can get C_1 and C_2 such that (7.1) is satisfied. Moreover, with l in (5.7), we can get C_3 such that (7.2) holds.*

Thus, we use Proposition 7.1 to construct d-variate mean 0 normal random vectors N_j^a and random vectors E_j^a such that

$$E_j^a \stackrel{D}{=} Y_j^a \text{ and } Var(N_j^a) = Var(Y_j^a), \quad 0 \le j \le q-1,$$

$$P_a\left(\max_{1\le i\le n} |\Pi_i^a - D_i^a| \ge c_0 z\right) \le C\frac{L_\gamma^a}{z^\gamma}, \text{ where } \Pi_i^a = \sum_{j=0}^{q_i-1} E_j^a,\ D_i^a = \sum_{j=0}^{q_i-1} N_j^a \tag{5.9}$$

and C is a constant depending on $\gamma, c_1, \ldots, c_5$ and C_3. These constants are free of a. We can create a set $\mathcal{A}$ with $P(\mathcal{A}) \to 1$ so that $a \in \mathcal{A}$ implies the statements in Proposition 5.4 and Proposition 5.3 hold. Putting $z = n^{1/r}$ above in (5.9), by Lemma 7.4 and the restriction (4.20), we have, as $n \to \infty$,

$$E(L_\gamma^a n^{-\gamma/r}) \le \frac{q}{n^{\gamma/r}} c_\gamma \max_c E(|\tilde{R}_{c,2k_0m}|^\gamma) = O(n^{1-\gamma/r} m^{\gamma/2-1}) \to 0, \tag{5.10}$$

using

$$E(|Y_j(\bar{\vartheta}_{2k_0j}, \bar{\vartheta}_{2k_0j+2k_0})|^\gamma) \leq c_\gamma \max_c E(|\tilde{R}_{c,2k_0m}|^\gamma) = O(m^{\gamma/2}).$$

Hence, conditioning on whether a lies in $\mathcal{A}$ or not, from (5.10) we obtain,

$$\max_{i \leq n} |\Pi_i^\vartheta - D_i^\vartheta| = o_P(n^{1/r}). \tag{5.11}$$

5.2. *Unconditional Gaussian Approximation and Regrouping:* Here we shall work with the processes $\Pi_i^\vartheta, \mu_i^\vartheta$ and D_i^ϑ. Note that, $V_{j0}(\bar{a}_{2k_0j})$ defined in (5.4) is a function of ϑ and might not be positive definite in an uniform fashion. For a constant $0 < \delta_* < \lambda_*$, let

$$V_{j1}(\bar{a}_{2k_0j}) = \begin{cases} V_{j0}(\bar{a}_{2k_0j}) & \text{if } \rho_*(V_{j0}^a) \geq \delta_* m, \\ (\delta_* m) I_d & \text{otherwise}, \end{cases} \tag{5.12}$$

which is a positive-definitized version of $V_{j0}(\bar{a}_{2k_0j})$. The following proposition shows that partial sums of $V_{j0}(\bar{a}_{2k_0j})$ and $V_{j1}(\bar{a}_{2k_0j})$ are close to each other.

Proposition 5.6. *For some $\iota > 0$, we have*

$$\max_{i \leq n} E\left(\left|\sum_{j=1}^{max(1,q_i-1)} (V_{j0}(\bar{a}_{2k_0j}) - V_{j1}(\bar{a}_{2k_0j}))\right|\right) = o_P(n^{2/r-\iota}).$$

Henceforth in the sequel we will slightly abuse $max(1, q_i - 1) = max(1, \lfloor i/(2k_0m) \rfloor - 1)$ and simply use $q_i - 1 = \lfloor i/(2k_0m) \rfloor - 1$ for presentational clarity.

Proof. of Proposition 5.6. Recall (5.2) for the definition of $F_{4j+1}(.), F_{4j+2}$ etc. Define

$$F_{21} = \sum_{i=m+1}^{2m} \tilde{X}_i.$$

Define the projection operator P_i by

$$P_i Y = E(Y|\mathcal{F}_i) - E(Y|\mathcal{F}_{i-1}), \quad Y \in \mathcal{L}_1.$$

For $1 \le j \le m$, $\|P_j F_{21}\| \le \sum_{i=m+1-j}^{m} \delta_{i,2}$. Since $\|E(F_{21}^T|\mathcal{F}_m)\|^2 = \sum_{j=1}^{m} \|P_j F_{21}\|^2$, we have

$$\begin{aligned} |E(F_1(\bar{a}_0)F_2^T)| &= |E(F_1(\bar{a}_0)F_{21}^T)| = |E(F_1(\bar{a}_0)E(F_{21}^T|\mathcal{F}_m))| \\ &\le \|F_1(\bar{a}_0)\|(\sum_{j=1}^{m}(\sum_{i=m+1-j}^{m} \delta_{i,2})^2)^{1/2}. \end{aligned} \tag{5.13}$$

Under the decay condition on $\Theta_{i,p}$ in (2.3), we have

$$E(|E(F_1(\bar{a}_0)F_{21}^T)|^\gamma) = O(m^{\max(\gamma/2, \gamma - \chi\gamma)}).$$

We expand the last term of $V_{j0}(\bar{a}_{2k_0 j})$ (see (5.4)). Also note that,

$$|E(F_{4j-2}F_{4j-1}^T) + E(F_{4j-1}F_{4j-2}^T)| \ll m \text{ and } \rho_*(Var(F_{4j+2})) \ge (k_0 - 1)\lambda_* m.$$

Then Proposition 5.6 follows from the fact that our solution of γ from (4.19), (4.20), and (4.21) satisfy $\gamma > \max(2, 4\chi)$ for $\chi \le \chi_0$ and

$$\begin{aligned} n \max_j P\left(\rho_*(V_{j0}^a) < \delta_* m\right) &\le 2n \max_j P(|E(F_{4j+1}(\bar{a}_{2k_0 j})F_{4j+2}^T)| \ge -\theta m/2) \\ &= O(n)\frac{m^{\max(\gamma/2, \gamma-\chi\gamma)}}{m^\gamma} = o(n^{2/r-\iota}), \end{aligned}$$

for some $\iota > 0$ since we can choose δ_* such that $\theta = (k_0 - 1)\lambda_* - \delta_* > 0$. $\square$

Recall (5.6) for the definition of U_j. By Lemma 7.4 and Jensen's inequality, we obtain

$\max_j \|U_j(\bar\vartheta_{2k_0j+2k_0})\|_{\gamma/2} = O(m^{1/2})$. By (4.20), $\phi_n := q^{1/\gamma}m^{1/2}n^{-1/r} \to 0$. Then

$$\begin{aligned} P\left(\max_{0\le j\le q-1} |U_j(\bar\vartheta_{2k_0j+2k_0})| \ge \phi_n n^{2/r}\right) &\le \sum_{j=0}^{q-1} P\left(|U_j(\bar\vartheta_{2k_0j+2k_0})| \ge \phi_n n^{2/r}\right) \\ &= O(\phi_n^{-\gamma/2} n^{1-\gamma/r} m^{\gamma/2-1}) = O(\phi_n^{\gamma/2}) \to 0. \end{aligned}$$

Similarly, $|L(\bar\vartheta_0)| = o_P(n^{2/r})$. Thus, by (5.5) and Proposition 5.6, since $Var(Y_j^a) = Var(N_j^a)$, one can construct i.i.d. $N(0, I_d)$ normal vectors $Z_l^a, l \in \mathbb{Z}$, such that

$$\max_{i\le n} |D_i^\vartheta - \varsigma_i(\vartheta)| = o_P(n^{1/r}), \text{ where } \varsigma_i(a) = \sum_{j=1}^{q_i-1} V_{j1}^0(\bar a_{2k_0j})^{1/2} Z_j^a.$$

By (5.11), we have

$$\max_{i\le n} |\Pi_i^\vartheta - \varsigma_i(\vartheta)| = o_P(n^{1/r}).$$

Let $Z_l^*, l \in \mathbb{Z}$, independent of $(\epsilon_j)_{j\in\mathbb{Z}}$, be i.i.d. $N(0, I_d)$ and define

$$\Psi_i = \sum_{j=1}^{q_i-1} V_{j1}(\bar\vartheta_{2k_0j})^{1/2} Z_j^*.$$

From the distributional equality,

$$(\Pi_i^\vartheta + M_i(\vartheta))_{1\le i\le n} \stackrel{D}{=} (S_i^\diamond)_{1\le i\le n}, \tag{5.14}$$

we need to prove Gaussian approximation for the process $\Psi_i + M_i(\vartheta)$. Define

$$B_j = V_{j1}(\bar\vartheta_{2k_0j})^{1/2} Z_j^* + \Lambda_{4j}(\bar\vartheta_{2k_0j}) + \Lambda_{4j+1}(\bar\vartheta_{2k_0j}),$$

which are independent random vectors for $j = 1, \ldots, q$ and let

$$S_i^\sharp = \sum_{j=1}^{q_i-1} B_j \text{ and } W_i^\sharp = \Psi_i + M_i(\vartheta) - S_i^\sharp.$$

Note that,

$$\max_{i\le n}|W_i^\sharp| = \max_{i\le n}|\Lambda_{4q_i}(\vartheta_{2k_0q_i}) + \Lambda_1(\vartheta_0)| = o_P(n^{1/r}). \tag{5.15}$$

Conditions (7.1) and (7.2) can be verified easily with this unconditional process $(S)_i^\sharp$ to use the Proposition 7.1. Thus, there exists B_j^{new} and Gaussian random variable B_j^{gau}, such that $(B_j^{new})_{j\le q-1} \stackrel{D}{=} (B_j)_{j\le q-1}$ and corresponding $B_j^{gau} \sim N(0, Var(B_j))$, such that

$$\max_{i\le n}|\sum_{j=1}^{\lfloor i/2k_0m\rfloor-1} B_j^{new} - \sum_{j=1}^{\lfloor i/2k_0m\rfloor-1} B_j^{gau}| \quad = \quad o_P(n^{1/r}). \tag{5.16}$$

By (4.17), (5.14), (5.15) and (5.16), we can construct a process S_i^c and B_j^c such that $(S_i^c)_{i\le n} \stackrel{D}{=} (S_i)_{i\le n}$ and $(B_j^c)_{j\le q-1} \stackrel{D}{=} (B_j^{gau})_{j\le q-1}$ and

$$\max_{i\le n}|S_i^c - \sum_{j=1}^{\lfloor i/(2k_0m)\rfloor-1} B_j^c| \quad = \quad o_P(n^{1/r}). \tag{5.17}$$

Relabel this final Gaussian process as

$$G_i^c = \sum_{j=1}^{\lfloor i/2k_0m\rfloor-1} (Var(B_j))^{1/2} Y_j^c,$$

where Y_j^c are i.i.d. $N(0, I_d)$. This concludes the proof of Theorem 2.1. $\square$

PROOF. of Proposition 5.1. Without loss of generality, we prove it for $j = 1$. Note that

$$2k_0m\lambda_* \le \rho_*(Var(S_{2k_0m})) \le \rho^*(Var(S_{2k_0m})) \le \|\sum_{i=1}^{2k_0m} X_i\|^2 \le 2k_0m\Theta_{0,2}^2. \tag{5.18}$$

Recall $X_i^{\oplus}$ and $\tilde{X}_i$ from (4.5) and (4.7). The same upper bound works for $S_i^{\oplus}$ and $\tilde{S}_i$. Note that, $\|S_{2k_0m}^{\oplus} - S_{2k_0m}\| = o(m)$ and from [14], we have

$$\|A_1 - S_{2k_0m}^{\oplus}\| = O(\sqrt{2k_0m}\Theta_{m,2}) = o(\sqrt{2k_0m}).$$

This concludes the proof using the Cauchy-Schwartz inequality. □

Proof. of Proposition 5.2. As A_j is the block sum of the m-dependent processes with length $2k_0m$, we have, using (5.18), for all j,

$$2k_0m(\lambda_* + \delta) \leq E(|A_j|^2) \leq 2k_0m\Theta_{0,2}^2,$$

for some small $\delta > 0$. We conclude the proof by using

$$|E(|Y_j^{\vartheta}|^2) - E(|A_{j+1}|^2)| = |\Lambda_{4j+1}(\bar{\vartheta}_{2k_0j})|^2 + |\Lambda_{4j+4}(\bar{\vartheta}_{2k_0j+2k_0})|^2 \leq 2m\Theta_{0,2}^2$$

and $k_0 > \Theta_{0,2}^2/\lambda_* + 1$. Using similar arguments, (5.8) follows. □

Proof. of Proposition 5.3. Note that, without loss of generality, we can assume V_j^a to be independent for different j since otherwise we can always break the probability statement in even and odd blocks and prove the statement separately. We use Corollary 1.6 and Corollary 1.7 from Nagaev (1979, [18]) respectively for the case $\gamma < 4$ and $\gamma \geq 4$ on $|V_j^a - E(V_j^a)|$ to deduce that it suffices to show the following

$$q \max_{1\leq t\leq q/l} \max_{t(l-1)+1\leq j\leq tl} P(|V_j^a - E(V_j^a)| \geq lm) \to 0. \tag{5.19}$$

We expand and write V_j^a as follows:

$$\begin{aligned} V_j^a &= Var(F_{4j+1}(\bar{a}_{2k_0j}) - \Lambda_{4j+1}(\bar{a}_{2k_0j})) + Var(F_{4j+2} + F_{4j+3}) \\ &+ E((F_{4j+1}(\bar{a}_{2k_0j}) - \Lambda_{4j+1}(\bar{a}_{2k_0j}))F_{4j+2}^T) + E(F_{4j+2}(F_{4j+1}(\bar{a}_{2k_0j}) - \Lambda_{4j+1}(\bar{a}_{2k_0j}))^T) \\ &+ E(F_{4j+3}(F_{4j+4}(\bar{a}_{2k_0j+2k_0}) - \Lambda_{4j+4}(\bar{a}_{2k_0j+2k_0}))^T) \\ &+ E((F_{4j+4}(\bar{a}_{2k_0j+2k_0}) - \Lambda_{4j+4}(\bar{a}_{2k_0j+2k_0}))F_{4j+3}^T) \\ &+ Var(F_{4j+4}(\bar{a}_{2k_0j+2k_0}) - \Lambda_{4j+4}(\bar{a}_{2k_0j+2k_0})). \end{aligned} \tag{5.20}$$

Using derivation similar to (5.13), it suffices to show (5.19) for only the first and last term in (5.20). Moreover, we assume $d = 1$ and $j = 1$ to simplify notations. The proofs and the theorems used can be easily extended to vector-valued processes. Denote by $\tilde{S}_{m,\{j\}}$ for the sum $\tilde{S}_m$ with ϵ_j replaced by an i.i.d. copy ϵ_j'. For the first term, by Burkholder's inequality,

$$\begin{aligned} E(|Var(F_1(\bar{a}_0)) - E(Var(F_1(\bar{a}_0)))|^{\gamma/2}) &= E(|E(\tilde{S}_m^2|a_{1-m}, \dots, a_0) - E(\tilde{S}_m^2)|^{\gamma/2}) \\ &= \|\sum_{j=-m}^{0} P_j\tilde{S}_m^2\|_{\gamma/2}^{\gamma/2} \le c_\gamma(\sum_{j=-m}^{0}\|P_j\tilde{S}_m^2\|_{\gamma/2}^2)^{\gamma/4} \end{aligned}$$

For $-m \le j \le 0$, $\|P_j\tilde{S}_m^2\|_{\gamma/2} \le \|\tilde{S}_m^2 - \tilde{S}_{m,\{j\}}^2\|_{\gamma/2} \le \|\tilde{S}_m - \tilde{S}_{m,\{j\}}\|_\gamma\|\tilde{S}_m + \tilde{S}_{m,\{j\}}\|_\gamma$. Note that $\|\tilde{S}_m\|_\gamma = O(m^{1/2})$ and $\|\tilde{S}_m - \tilde{S}_{m,\{j\}}\|_\gamma \le \sum_{r=1}^m \tilde{\delta}_{r-j,\gamma}$. By Lemma 7.3, $\tilde{\delta}_{k,\gamma} \le 2n^{1/p-1/\gamma}t_n^{1-p/\gamma}\delta_{k,p}^{p/\gamma}$. Then since $3 > 2(\chi+1)p/\gamma$ for $\chi \le \chi_0$, we have

$$\begin{aligned} \sum_{j=-m}^{0}\|P_j\tilde{S}_m^2\|_{\gamma/2}^2 &= O(m)\sum_{j=-m}^{0}\sum_{r=1}^{m}(\tilde{\delta}_{r-j,\gamma})^2 \\ &= O(m)n^{2/p-2/\gamma}t_n^{2-2p/\gamma}\sum_{j=0}^{m}(\sum_{r=1}^{m}\delta_{r+j,p}^{p/\gamma})^2 \\ &= O(m)n^{2/p-2/\gamma}t_n^{2-2p/\gamma}m^{3-2(\chi+1)p/\gamma}(\log m)^{-2Ap/\gamma}, \end{aligned} \tag{5.21}$$

by (2.3) and the Hölder inequality. Then, since $A > 2\gamma/p$ and $\log m \asymp \log q \asymp \log n$,

$$\begin{aligned} & qE(|Var(F_1(\bar{a}_0)) - E(Var(F_1(\bar{a}_0)))|^{\gamma/2}) \\ \lesssim\ & qm^{\gamma-(\chi+1)p/2} n^{\gamma/2p-1/2} t_n^{\gamma/2-p/2} (\log n)^{-Ap/2} = o((lm)^{\gamma/2}), \end{aligned} \tag{5.22}$$

using (4.3), (4.21) and the choice of l in (5.7). For the last term in (5.20), we view $E(F_4(\bar{a}_{2k_0})^2)$ as

$$E(F_4(\bar{a}_{2k_0})^2) = E((\tilde{S}_{2k_0 m} - \tilde{S}_{(2k_0-1)m})^2 | a_{(2k_0-1)m+1}, \ldots a_{2k_0 m})$$

and show that it is close to $(\tilde{S}_{2k_0 m} - \tilde{S}_{(2k_0-1)m})^2$. Let $\mathcal{F}_j^m = (\epsilon_j, \ldots, \epsilon_m)$. Note that,

$$\begin{aligned} \|\tilde{S}_m^2 - E(\tilde{S}_m^2 | a_m, \ldots, a_1)\|_{\gamma/2}^{\gamma/2} &\lesssim \Big(\sum_{j=-m-1}^{0} \|E(\tilde{S}_m^2 | \mathcal{F}_j^m) - E(\tilde{S}_m^2 | \mathcal{F}_{j+1}^m)\|_{\gamma/2}^2 \Big)^{\gamma/4} \\ &\le cm^{\gamma-(\chi+1)p/2} n^{\gamma/2p-1/2} t_n^{\gamma/2-p/2} (\log m)^{-Ap/2} \\ &= o(q^{-1}(lm)^{\gamma/2}), \end{aligned} \tag{5.23}$$

similar to the derivation in (5.21). By (5.22) and (5.23), it suffices to show that

$$\frac{n}{m} P(|\tilde{S}_m| \ge \sqrt{lm}) \to 0. \tag{5.24}$$

Using the Nagaev-type inequality from Wu and Wu (2016, [27]) we obtain

$$P(|\tilde{S}_m| \ge \sqrt{lm}) \ \le \ C_1 \frac{m^{\max\{1, p(1/2-\chi)\}}}{(lm)^{p/2}} + C_2 \exp(-C_3 l), \tag{5.25}$$

where C_1, C_2 and C_3 depend on χ and p. The second term in (5.25) is $o(m/n)$ since $e^{-l} \to 0$ very fast. For the first term in (5.25), if $\chi < 1/2 - 1/p$, then

$$\frac{n}{m} \frac{m^{p(1/2-\chi)}}{(lm)^{p/2}} = (\log n)^p n^{1-p/\gamma+L(p/\gamma-p\chi-1)} t_n^{k(p/\gamma-p\chi-1)} = o(1),$$

as from (4.21) we have $1 - p/\gamma + L(p/\gamma - p\chi - 1) = L(p/\gamma - 1)(\chi p + p + 1) < 0$. If $1/2 - 1/p \le \chi < \chi_0$ and consequently $r < p$, then we have, for the first term in (5.25),

$$\frac{n}{m}\frac{m}{(lm)^{p/2}} = (\log n)^p n^{p(1/p-1/\gamma+L(1/\gamma-1/2))} t_n^{k(p/\gamma-p/2)} = o(1), \tag{5.26}$$

using (4.3), $r < p$ and the fact that r satisfy $1/r - 1/\gamma + L(1/\gamma - 1/2) = 0$. $\square$

Proof. of Proposition 5.4. By Lemma 7.4, $E(L_\gamma^a) \asymp qm^{\gamma/2}$. Then it suffices to prove

$$P(|L_\gamma^a - E(L_\gamma^a)| \ge cqm^{\gamma/2}/\log q) \to 0, \tag{5.27}$$

holds for some constant $c > 0$. Note that $E(|Y_j^a|^\gamma)$ are even indices j (also for odd indices j). Thus we can prove the statement separately by breaking L_γ^a in sum of even and odd $E(|Y_j^a|^\gamma)$. Without loss of generality, we assume all $E(|Y_j^a|^\gamma)$ are independent and proceed. Define $J_j = (2k_0m)^{-\gamma/2}E(|\tilde{S}_{2k_0mj} - \tilde{S}_{2k_0m(j-1)}|^\gamma|\bar{a}_{2k_0(j-1)}, \bar{a}_{2k_0j})$ and $\theta = l^{\gamma/2} = q/(\log q)^\gamma$. Recall the truncation operator T from (4.4). Noting $E(J_j) = O(1)$ from Lemma 7.4, we have

$$P(|\sum_{j=1}^q T_\theta(J_j) - E(T_\theta(J_j))| \ge \phi) \le \frac{q}{\phi^2}\max_j E(T_\theta(J_j)^2) = O(\theta q/\phi^2) = o(1),$$

where $\phi = q/\log q$, and

$$\max_j P(J_j \ge \theta) \le \max_j P(E(|\tilde{S}_{2k_0mj} - \tilde{S}_{2k_0m(j-1)}|^2|\bar{a}_{2k_0(j-1)}, \bar{a}_{2k_0j}) \ge 2k_0lm) = o(q^{-1}),$$

from (5.22), (5.23) and (5.24). Thus $P(|\sum_{j=1}^q J_j - \sum_{j=1}^q E(J_j)| \ge \phi) \to 0$ which is a restatement of (5.27). $\square$

Proof. of Proposition 5.5. We showed in Proposition 5.4 that

$$P(cqm^{\gamma/2} \leq L_\gamma \leq Cqm^{\gamma/2}) \to 1,$$

for some constants c and C. Let l be as given in (5.7). Let $S = \{0, l, 2l, \cdots\}$. Proposition 5.2 and Proposition 5.3 show that, for some constants c and C,

$$P(clk_0 m \leq \min_{i \in S} \rho_* \left(Var \left(\sum_{j=i}^{i+l-1} Y_j^a \right) \right) \leq \max_{i \in S} \rho^* \left(Var \left(\sum_{j=i}^{i+l-1} Y_j^a \right) \right) \leq Clk_0 m) \to 1.$$

We choose $\eta_k = kl$ and $s \asymp q/l$. Starting with the conditional block sum process Y_j^a for $0 \leq j \leq q-1$, this choice of η_k satisfies (7.1) for a given a with probability going to 1. The other condition, (7.2) can be easily verified for such a choice of η-sequence using ideas similar to the proof of Proposition 5.4. We skip the details of that derivation. □

6. Proof of Theorem 2.2

Proof. *Case 1 ($\chi > \chi_0$):-* Note that the optimal power γ and the optimal bound $1/r$ increases and decreases with χ respectively (see also Figures 1 and 2). This is a motivation behind tweaking our proof for the verification of (7.1) to handle the $(\log n)$ term in choice of l in (5.7). While using the Nagaev inequality to show (5.24), we use a power $\gamma' > \gamma$ while keeping the choice of l (cf 5.7) same as before. We form a set of new equations

$$\begin{aligned} 1/2 + 1/p - 2/r' + L'(1 - (\chi+1)p/r') &= 0, \\ 1/p - 1/\gamma' + L' - L'(\chi+1)p/\gamma' &= 0, \\ 1 - \gamma'/r' + L'(\gamma'/2 - 1) &= 0. \end{aligned} \tag{6.1}$$

The intuition behind the first of these equations is to use a higher power than p in the m-dependence approximation. However, we only defined moments up to p. So we use Lemma 7.3 to obtain a new equation corresponding to the m-dependence approximation using a power r' that is little higher than p. The solution of (6.1) has the property

$$\gamma' < 2(1 + p + p\chi)/3. \tag{6.2}$$

for $\chi > \chi_0$. Also we observe $L' < L(\chi_0)$ (cf Figure 2) and hence $m^{1-\gamma'/2} \ll m'^{1-\gamma'/2}$ where m' is taken as $n^{L'} t_n^k$ following (4.15). We apply the following version of Nagaev-type inequality from Liu, Xiao and Wu (2013, [15]) to obtain

$$\begin{aligned} P(|\tilde{S}_m| \geq \sqrt{lm}) &\lesssim \frac{m}{(lm)^{\gamma'/2}} \nu_R^{\gamma'+1} + \sum_{r=1}^{R} \exp\left(-c_{\gamma'} \frac{\lambda_r^2 l}{\tilde{\theta}_{r,2}^2}\right) + \frac{m^{\gamma'/2} \tilde{\Theta}_{m+1,\gamma'}^{\gamma'}}{(lm)^{\gamma'/2}} \\ &+ \frac{m \sup_i \|T_{t_n n^{1/p}}(X_i)\|_{\gamma'}^{\gamma'}}{(lm)^{\gamma'/2}} + \exp\left(-\frac{c_{\gamma'} l}{\sup_i \|T_{t_n n^{1/p}}(X_i)\|_2^2}\right), \end{aligned} \tag{6.3}$$

where $\nu_R = \sum_{r=1}^{R} \mu_r$, $\mu_r = (\tau_r^{\gamma'/2-1} \tilde{\theta}_{r,\gamma'}^{\gamma'})^{1/(\gamma'+1)}$, $\lambda_r = \mu_r/\nu_R$, $\tilde{\theta}_{r,t} = \sum_{i=1+\tau_{r-1}}^{\tau_r} \tilde{\delta}_{i,t}$ for some sequence $0 = \tau_0 < \tau_1 < \ldots < \tau_R = m$. For the choice $\tau_r = 2^{r-1}$ for $1 \leq r \leq R-1 = \lfloor \log_2 m \rfloor$, we obtain $\nu_R^{\gamma'+1} = O(n^{\gamma'/p-1} t_n^{\gamma'-p})$ using (6.2), (4.2) under the decay condition on $\Theta_{i,p}$ in (2.3). The third term and the exponential terms are straightforward to deal with. The fourth term is dealt similar to (7.6). Combining these in the view of our new set of equations in (6.1), we get $P(|\tilde{S}_m| \geq \sqrt{lm}) = o(m/n)$ which is sufficient to conclude the proof as proposed in (5.24).

The positive-definitization technique introduced in (5.12) is validated in Proposition 5.6. This step requires $\gamma > 4\chi$ for the case $\chi > \max(1/2, \chi_0)$. We observe that $\gamma' - 4\chi = 0$ has a root $\chi_1 > \chi_0$. This allows us to replace χ in the decay condition of $\Theta_{i,p}$ by $\min(\chi, \chi_1)$ and the proof goes through. The arguments for the rest of the proof of Theorem 2.1 remains valid.

Case 2 ($\chi = \chi_0, 2 < p < 4$):- We shall apply Proposition 1 from Einmahl (1987, [6]). He proved a Gaussian approximation result for independent but not necessarily identical vectors with diagonal covariance matrix. The two remarks following the proposition mention that the diagonal nature of every covariance matrix can be relaxed if these matrices have bounded eigenvalues. A careful check of his proof reveals that it can be further relaxed to the assumption of bounded eigenvalues of the covariance matrix of a normalized block sum only. This allows us to replace the l (see (5.7)) to use the conclusion of Proposition 7.1 by l' without the logarithm term ($\log n$) in the denominator and without the condition (7.2). Thus we obtain $o_P(n^{1/p})$ rate for all $2 < p < 4$.

Case 3 ($\chi = \chi_0, p \geq 4$):- In this case we do not have a similar optimal Gaussian approximation result for independent but not identically distributed random vectors. Instead we shall apply Proposition 7.1 again. The sufficient conditions in that result lead to an unavoidable $(\log n)$ term in choice of l (see 5.7). This, in turn leads to $o_P(n^{1/p}\log n)$ rate. Note that, $\chi_0 > 1/2 - 1/p$ for all $p > 2$. From the proof for the case $0 < \chi < \chi_0$, consider (5.26), observe that if $\chi = \chi_0$, then

$$\frac{n}{m}P(|\tilde{S}_m| \geq \sqrt{lm}) = O((\log n)^p t_n^{k(p/\gamma - p/2)}),$$

which may diverge to ∞. To deal with this difficulty in this special case, we choose a different m sequence. Our new set of conditions with $\tau_n = n^{1/p}(\log n)^\delta$ are

$$\begin{aligned} n^{1/2-1/p}m^{-\chi}(\log n)^{-A-\delta} &\to 0, \\ n^{1/p-1/\gamma}m^{1-(\chi+1)p/\gamma}(\log n)^{-Ap/\gamma} &\to 0, \\ n^{1-\gamma/p}(\log n)^{-\gamma\delta}m^{\gamma/2-1} &\to 0, \\ (\log n)^\gamma m^{1-\gamma/2}n^{\gamma/p-1}t_n^{\gamma-p} &\to 0. \end{aligned}$$

where the last one is obtained using γth moment in (6.3). Let $m = \lfloor n^L(\log n)^{2\gamma/(\gamma-2)}t_n^k\rfloor$ with $0 < k < (\gamma/2-1)^{-1}(\gamma-p)$, we can achieve $\delta = 1$. We still have the same set of equations for L, γ and r as (4.19), (4.20) and (4.21). A careful check reveals that the rest of the proof goes through with this modified m sequence. $\square$

7. Some Useful Results Proposition 7.1 concerns Gaussian approximation for independent vectors. There are several types of Gaussian approximations in literature for independent vectors. We find the following result by Götze and Zaitsev (2008, [10]) particularly useful since it provides an explicit and good approximation bound for the partial sums. This has been used several times in our proof.

Proposition 7.1. *Let $\xi_1, \dots, \xi_n$ be independent $\mathbb{R}^d$-valued mean 0 random vectors. Assume that there exist $s \in \mathbb{N}$ and a strictly increasing sequence of non-negative integers $\eta_0 = 0 < \eta_1 < \dots < \eta_s = n$ satisfying the following conditions. Let*

$$\zeta_k = \xi_{\eta_{k-1}+1} + \dots + \xi_{\eta_k}, \quad Var(\zeta_k) = B_k, \quad k = 1, \dots, s$$

and $L_\gamma = \sum_{j=1}^n E(|\xi_j|^\gamma)$, $\gamma \geq 2$, and assume that, for all $k = 1, \dots, s$,

$$C_1 w^2 \leq \rho_*(B_k) \leq \rho^*(B_k) \leq C_2 w^2, \tag{7.1}$$

where $w = (L_\gamma)^{1/\gamma} / \log^* s$, with some positive constants C_1 and C_2. Suppose the quantities

$$\lambda_{k,\gamma} = \sum_{j=\eta_{k-1}+1}^{\eta_k} E\|\xi_j\|^\gamma, \quad k = 1, \dots s,$$

satisfy, for some $0 < \epsilon < 1$ and constant C_3,

$$C_3 d^{\gamma/2} s^\epsilon (\log^* s)^{\gamma+3} \max_{1 \leq k \leq s} \lambda_{k,\gamma} \leq L_\gamma. \tag{7.2}$$

Then one can construct on a probability space independent random vectors $X_1, \dots, X_n$ and a corresponding set of independent Gaussian vectors $Y_1, \dots, Y_n$ so that $(X_j)_{j=1}^n \stackrel{D}{=} (\xi_j)_{j=1}^n$, $E(Y_j) = 0$, $Var(Y_j) = Var(X_j), 1 \leq j \leq n$, and for any $z > 0$,

$$P\left(\max_{t \leq n} |\sum_{i=1}^t X_i - \sum_{i=1}^t Y_i| \geq z\right) \leq C_* L_\gamma z^{-\gamma}.$$

where C_* is a constant that depends on d, γ, C_1, C_2 and C_3.

Lemma 7.2. *Let $p < \gamma$. Assume (2.A). Then $\sup_i E \min\{|X_i|^\gamma n^{-\gamma/p}, 1\} = o(n^{-1})$.*

PROOF. Choose $k_n = \lfloor 2(\log n)/((p+\gamma)\log 2)\rfloor$. Then $n = o(2^{\gamma k_n})$ and $2^{pk_n} = o(n)$. Let $Z = |X_i|n^{-1/p}$. The lemma follows from

$$\begin{aligned} E(\min\{Z^\gamma, 1\}) &\leq P(Z \geq 1) + \sum_{k=0}^{k_n} 2^{-k\gamma} P(2^{-1-k} \leq Z < 2^{-k}) + 2^{-\gamma(k_n+1)} \\ &\leq E(Z^p \mathbf{1}_{Z\geq 1}) + \sum_{k=0}^{k_n} 2^{p(k+1)-k\gamma} E(Z^p \mathbf{1}_{Z\geq 2^{-1-k}}) + 2^{-\gamma(k_n+1)} = o(n^{-1}), \end{aligned}$$

in view of the uniform integrability condition (2.A) and $n^{1/2}/2^{k_n} \to \infty$. □

LEMMA 7.3. *The functional dependence measures defined on the truncated process* $(X_i^\oplus)$ *and the* m*-dependent process* $(\tilde{X}_i)$*, satisfy* $\tilde{\delta}_{j,\gamma} \leq \delta_{j,\gamma}^\oplus \leq 2n^{1/p-1/\gamma} t_n^{1-p/\gamma} \delta_{j,p}^{p/\gamma}$.

PROOF. Since the truncation operator T is Lipschitz continuous,

$$\begin{aligned} (\delta_{j,\gamma}^\oplus)^\gamma &= \sup_i E(|T_{t_n n^{1/p}}(X_i) - T_{t_n n^{1/p}}(X_{i,(i-j)})|^\gamma) \\ &= n^{\gamma/p} t_n^\gamma \sup_i E\left(\left|\min\left(2, \left|\frac{X_i - X_{i,(i-j)}}{t_n n^{1/p}}\right|\right)\right|^\gamma\right) \leq 2^\gamma n^{\gamma/p-1} t_n^{\gamma-p} \delta_{j,p}^p. \end{aligned}$$

The first inequality $\tilde{\delta}_{j,\gamma} \leq \delta_{j,\gamma}^\oplus$ follows from (4.11). □

LEMMA 7.4. *Rosenthal Type Moment Bound Recall (4.2) and (4.3) for* t_n*. Assume (4.9), (4.13), (4.14) along with (2.6) on* A *related to the restriction on* $\Theta_{i,p}$ *as mentioned in (2.3). Moreover, assume* $m = \lfloor n^L t_n^k \rfloor$ *with* k *satisfying* $k < (\gamma/2 - 1)^{-1}(\gamma - p)$*. Then, we have*

$$\max_t E(\max_{1\leq l\leq m} |\tilde{R}_{t,l}|^\gamma) = O(m^{\gamma/2}). \tag{7.3}$$

PROOF. Since the functional dependence measure is defined in an uniform manner, we can ignore the $\max_t$ in (7.3) and use the Rosenthal-type inequality for stationary processes in Liu, Xiao and Wu (2013, [15]). By [15], there is a constant c, depending only on γ, such that

$$\begin{aligned}\|\max_{1\le l\le m}|\tilde{R}_{t,l}|\|_\gamma &\le cm^{1/2}[\sum_{j=1}^m \tilde{\delta}_{j,2}+\sum_{j=1+m}^\infty \tilde{\delta}_{j,\gamma}+\sup_i \|T_{t_n n^{1/p}}(X_i)\|] \\ &\quad +cm^{1/\gamma}[\sum_{j=1}^m j^{1/2-1/\gamma}\tilde{\delta}_{j,\gamma}+\sup_i\|T_{t_n n^{1/p}}(X_i)\|_\gamma] \\ &\le c(I+II+III+IV),\end{aligned}$$

where

$$\begin{aligned} I &= m^{1/2}\sum_{j=1}^m \tilde{\delta}_{j,2}+m^{1/2}\|X_1\|_2, \\ II &= m^{1/2}\sum_{j=m+1}^\infty \tilde{\delta}_{j,\gamma}, \quad III = m^{1/\gamma}\sum_{j=1}^\infty j^{1/2-1/\gamma}\tilde{\delta}_{j,\gamma}, \\ IV &= m^{1/\gamma}\sup_i \|T_{t_n n^{1/p}}(X_i)\|_\gamma.\end{aligned}$$

For the first term I, since $\sum_{j=1}^\infty \delta_{j,2}+\sup_i\|X_i\|_2 \le 2\Theta_{0,2}$ and $\tilde{\delta}_{j,2}\le\delta_{j,2}$, we have $I=O(m^{1/2})$. Starting with II, we apply Lemma 7.3 to obtain

$$II = m^{1/2}\sum_{j=m+1}^\infty \tilde{\delta}_{j,\gamma} \lesssim m^{1/2}n^{1/p-1/\gamma}t_n^{1-p/\gamma}\sum_{j=m+1}^\infty \delta_{j,p}^{p/\gamma}.$$

The rest follows from the derivation in (4.18) and (4.21). For the third term, we have

$$\begin{aligned} III &\lesssim m^{1/\gamma}n^{1/p-1/\gamma}t_n^{1-p/\gamma}\sum_{j=1}^m j^{1/2-1/\gamma}\delta_{j,p}^{p/\gamma} \qquad (7.4) \\ &\le m^{1/\gamma}n^{1/p-1/\gamma}t_n^{1-p/\gamma}\sum_{l=1}^{\lfloor\log_2 m\rfloor+1}\sum_{j=2^{l-1}}^{2^l-1} j^{1/2-1/\gamma}\delta_{j,p}^{p/\gamma} \\ &\le m^{1/\gamma}n^{1/p-1/\gamma}t_n^{1-p/\gamma}\sum_{l=1}^{\lfloor\log_2 m\rfloor+1} 2^{l(3/2-1/\gamma-p/\gamma)}O(2^{-l\chi p/\gamma}l^{-Ap/\gamma}).\end{aligned}$$

Recall the definition of χ_0 from (2.5). If $\chi \le \chi_0$, then our solution for γ satisfies

$$3/2 - 1/\gamma - (\chi+1)p/\gamma \ge 0,$$

with equality holding only for $\chi = \chi_0$. Hence, if $\chi < \chi_0$, we have

$$m^{-1/2} III = m^{1-(\chi+1)p/\gamma} n^{1/p-1/\gamma} t_n^{1-p/\gamma} (\log n)^{-Ap/\gamma} O(1) = o(1),$$

from (4.21), (4.15) and (4.3). If $\chi = \chi_0$, since $A > \gamma/p$ from (2.6) [The lower bound for A there is just $2\gamma/p$ as mentioned in (4.19)], we have

$$m^{-1/2} III = m^{1/\gamma-1/2} n^{1/p-1/\gamma} t_n^{1-p/\gamma} O(1) = o(1), \tag{7.5}$$

since (4.20) is true. Also for the case of $\chi > \chi_0$ in the proof of Theorem 2.2, the way we define our three conditions in (6.1) the new solution also satisfy $\gamma' = 2(1+p+p\chi)/3$ and thus (7.5) holds. For the fourth term IV, we use (4.2) to derive

$$\begin{aligned} m^{-\gamma/2} IV^\gamma &= m^{1-\gamma/2} \sup_i \|T_{t_n n^{1/p}}(X_i)\|^\gamma \qquad (7.6) \\ &\le m^{1-\gamma/2} t_n^\gamma n^{\gamma/p} \sup_i E\left(\min\{\frac{|X_i|^\gamma}{t_n^\gamma n^{\gamma/p}}, 1\}\right) \\ &= m^{1-\gamma/2} t_n^\gamma n^{\gamma/p-1} o(1) = o(1), \end{aligned}$$

in the light of (4.20). □

References

[1] Aue, A., Hörmann, S., Horváth, L. and Reimherr, M. (2009). Break detection in the covariance structure of multivariate time series models. *The Annals of Statistics* **37** 4046–4087.

[2] Berkes, I., Liu, W. and Wu, W. B. (2014). Komlós-Major-Tusnády approximation under dependence. *Ann. Probab.* **42** 794–817. MR3178474

[3] Donsker, M. D. (1952). Justification and extension of Doob's heuristic approach to the Komogorov-Smirnov theorems. *Ann. Math. Statistics* **23** 277–281. MR0047288

[4] Doob, J. L. (1949). Heuristic approach to the Kolmogorov-Smirnov theorems. *Ann. Math. Statistics* **20** 393–403. MR0030732

[5] Eberlein, E. (1986). On strong invariance principles under dependence assumptions. *Ann. Probab.* **14** 260–270. MR815969

[6] Einmahl, U. (1987a). A useful estimate in the multidimensional invariance principle. *Probab. Theory Related Fields* **76** 81–101. MR899446

[7] Einmahl, U. (1987b). Strong invariance principles for partial sums of independent random vectors. *Ann. Probab.* **15** 1419–1440. MR905340

[8] Einmahl, U. (1989). Extensions of results of Komlós, Major, and Tusnády to the multivariate case. *J. Multivariate Anal.* **28** 20–68. MR996984

[9] Erdös, P. and Kac, M. (1946). On certain limit theorems of the theory of probability. *Bull. Amer. Math. Soc.* **52** 292–302. MR0015705

[10] Götze, F. and Zaitsev, A. Y. (2008). Bounds for the rate of strong approximation in the multidimensional invariance principle. *Teor. Veroyatn. Primen.* **53** 100–123. MR2760567

[11] Komlós, J., Major, P. and Tusnády, G. (1975). An approximation of partial sums of independent RV's and the sample DF. I. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **32** 111–131. MR0375412

[12] Komlós, J., Major, P. and Tusnády, G. (1976). An approximation of partial sums of independent RV's, and the sample DF. II. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **34** 33–58. MR0402883

[13] Liu, W. and Lin, Z. (2009). Strong approximation for a class of stationary processes. *Stochastic Process. Appl.* **119** 249–280. MR2485027

[14] Liu, W. and Wu, W. B. (2010). Asymptotics of spectral density estimates. *Econometric Theory* **26** 1218–1245. MR2660298

[15] Liu, W., Xiao, H. and Wu, W. B. (2013). Probability and moment inequalities under dependence. *Statist. Sinica* **23** 1257–1272. MR3114713

[16] Lütkepohl, H. (2005). *New introduction to multiple time series analysis*. Springer-Verlag, Berlin. MR2172368

[17] Monrad, D. and Philipp, W. (1991). The problem of embedding vector-valued martingales in a Gaussian process. *Theory of Probability & Its Applications* **35** 374–377.

[18] Nagaev, S. V. (1979). Large deviations of sums of independent random variables. *Ann. Probab.* **7** 745–789. MR542129

[19] Priestley, M. B. (1988). *Nonlinear and nonstationary time series analysis*. Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], London. MR991969

[20] Prohorov, Y. V. (1956). Convergence of random processes and limit theorems in probability theory. *Teor. Veroyatnost. i Primenen.* **1** 177–238. MR0084896

[21] Sakhanenko, A. I. (2006). Estimates in the invariance principle in terms of truncated power moments. *Sibirsk. Mat. Zh.* **47** 1355–1371. MR2302850

[22] Trapani, L., Urga, G. and Kao, C. (2017). Testing for instability in covariance structures. *Bernoulli* To Appear.

[23] Tsay, R. S. (2010). *Analysis of financial time series*, third ed. *Wiley Series in Probability and Statistics*. John Wiley & Sons, Inc., Hoboken, NJ. MR2778591

[24] Wiener, N. (1958). *Nonlinear Problems in Random Theory*. Wiley, New York.

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

[26] Wu, W. B. (2007). Strong invariance principles for dependent random variables. *Ann. Probab.* **35** 2294–2320. MR2353389

[27] Wu, W.-B. and Wu, Y. N. (2016). Performance bounds for parameter estimates of high-dimensional linear models with correlated errors. *Electron. J. Stat.* **10** 352–379. MR3466186

[28] Wu, W. B. and Zhao, Z. (2007). Inference of trends in time series. *J. R. Stat. Soc. Ser. B*

Stat. Methodol. **69** 391–410. MR2323759

[29] Wu, W. B. and Zhou, Z. (2011). Gaussian approximations for non-stationary multiple time series. *Statist. Sinica* **21** 1397–1413. MR2827528

[30] Zaitsev, A. Y. (2000). Multidimensional version of a result of Sakhanenko in the invariance principle for vectors with finite exponential moments. I. *Teor. Veroyatnost. i Primenen.* **45** 718–738. MR1968723

[31] Zaitsev, A. Y. (2001a). Multidimensional version of a result of Sakhanenko in the invariance principle for vectors with finite exponential moments. III. *Teor. Veroyatnost. i Primenen.* **46** 744–769. MR1971831

[32] Zaitsev, A. Y. (2001b). Multidimensional version of a result of Sakhanenko in the invariance principle for vectors with finite exponential moments. II. *Teor. Veroyatnost. i Primenen.* **46** 535–561. MR1978667

[33] Zhou, Z. and Wu, W. B. (2010). Simultaneous inference of linear models with time varying coefficients. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **72** 513–531. MR2758526