Statistica Sinica **21** (2011), 1397-1413

GAUSSIAN APPROXIMATIONS FOR NON-STATIONARY MULTIPLE TIME SERIES

Wei Biao Wu and Zhou Zhou

University of Chicago and University of Toronto

Abstract: We obtain an invariance principle for non-stationary vector-valued stochastic processes. It is shown that, under mild conditions, the partial sums of non-stationary processes can be approximated on a richer probability space by sums of independent Gaussian random vectors with nearly optimal bounds. The latter Gaussian approximation result has a wide range of applications in the study of multiple non-stationary time series.

Key words and phrases: Central limit theorem, functional linear models, Gaussian approximation, local stationarity, non-stationary nonlinear multiple time series.

1. Introduction

Let ε_i, $i \in \mathbb{Z}$, be independent and identically distributed (i.i.d.) random elements. Consider the d-dimensional random vector process

$$X_i = H_i(\mathcal{F}_i) = (X_{i1}, \ldots, X_{id})^T, \tag{1.1}$$

where $\mathcal{F}_i = (\ldots, \varepsilon_{i-1}, \varepsilon_i)$, H_i is a measurable function such that X_i is a well-defined random vector, and ^{T} denotes matrix transpose. The primary goal of the paper is to study approximations of partial sums S_i of the X_i by Gaussian processes. Generically speaking, such a Gaussian approximation scheme means that, on a richer probability space, there exists a Gaussian process $\check{G}_i$ and a process $\check{S}_i$ such that $(\check{S}_i)_{i\in\mathbb{N}}$ and $(S_i)_{i\in\mathbb{N}}$ are identically distributed, denoted by $(\check{S}_i)_{i\in\mathbb{N}} \overset{\mathcal{D}}{=} (S_i)_{i\in\mathbb{N}}$, and

$$\max_{1\le i\le n} |\check{S}_i - \check{G}_i| = O(r_n). \tag{1.2}$$

Here r_n is the rate of the approximation and $O(r_n)$ in (1.2) can be $O_{\mathbb{P}}(r_n)$ or the almost sure rate $O_{\text{a.s.}}(r_n)$. Results of this sort are traditionally called Hungarian embedding and they have many applications in statistics. Roughly speaking, with the approximation (1.2), if r_n is sufficiently small, then statistics involving the partial sum process $(S_i)_{i=1}^n$ can be approximated by functionals of the

Gaussian process $(\check{G}_i)_{i=1}^n$; these are generally easier to deal with since Gaussian processes have many nice properties. For a recent application, Wu and Zhao (2007) considered nonparametric inference of trends in time series by using Gaussian applications of type (1.2).

The problem of Gaussian approximation has a substantial history. For i.i.d. random variables with $d = 1$, see Komlós, Major, and Tusnády (1975, 1976) and Csörgő and Révész (1981). For independent but not identically distributed random variables, see Shao (1995) and Sakhanenko (1984), among others. There is an extensive literature on strong approximations under dependence; see Philipp and Stout (1975), Berkes and Philipp (1979), Bradley (1983), Shao (1993), Rio (1995), Lin and Lu (1996), Volný (1999), Dedecker and Prieur (2004), and Wu (2007). A challenging problem is to generalize the Gaussian approximation results to vector-valued processes. Such results are very useful in statistical inference of multiple time series. Eberlein (1986) obtained a Gaussian approximation result for dependent random vectors with approximation error $O(n^{1/2-\kappa})$, for some $\kappa > 0$. The latter approximation rate can be substantially improved if one assumes that the random vectors are independent. Einmahl (1987a,b, 1989) and Zaitsev (2001, 2002a,b) obtained deep results on Gaussian approximations for independent random vectors with optimal and nearly optimal rates. For stationary multiple time series, Liu and Lin (2009) obtained an important result on strong invariance principles with optimal bounds. Here we focus on the Gaussian approximation problem for non-stationary multiple time series.

We introduce some notation. Denote by Id_d the d-dimensional identity matrix, and by $N(\mu, \Sigma)$ the multivariate Gaussian distribution with mean vector μ and covariance matrix Σ. For a matrix $A = (a_{ij})_{i \le I, j \le J}$, let $|A| = (\sum_{i,j} a_{ij}^2)^{1/2}$, so $|A|^2 = \mathrm{trace}(AA^T)$. If A is a $d \times d$ symmetric nonnegative definite matrix with eigen-decomposition $A = Q\Lambda Q^T$, where Q is an orthonormal matrix satisfying $QQ^T = \mathrm{Id}_d$ and $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_d)$ is a diagonal matrix with $\lambda_1 \ge \cdots \ge \lambda_d \ge 0$, we define its root $A^{1/2} = Q\Lambda^{1/2}Q^T$, where $\Lambda^{1/2} = \mathrm{diag}(\lambda_1^{1/2}, \ldots, \lambda_d^{1/2})$. For a random vector Y, write $Y \in \mathcal{L}^p$, $p > 0$, if $\|Y\|_p := [\mathbb{E}(|Y|^p)]^{1/p} < \infty$. For $\mathcal{L}^2$ norm write $\|\cdot\| = \|\cdot\|_2$. Define the projection operator $\mathcal{P}_i$ by

$$\mathcal{P}_i Y = \mathbb{E}(Y|\mathcal{F}_i) - \mathbb{E}(Y|\mathcal{F}_{i-1}), \quad Y \in \mathcal{L}^1. \tag{1.3}$$

Define the floor function $\lfloor u \rfloor = \max\{k \in \mathbb{Z} : k \le u\}$, $u \in \mathbb{R}$. Throughout the paper C_p denotes a constant whose value depends only on p.

The rest of the paper is structured as follows. Main results are presented in Section 2 and proved in Section 3. Section 2 also presents examples of linear and nonlinear non-stationary multiple time series for which our results are applicable.

2. Main Results

Our main results assert that the partial sum process $S_i = \sum_{j=1}^{i} X_j$ can be "regularized" by a Gaussian process. Namely, under suitable weak dependence conditions on the process (X_i), expressed in terms of $\delta_{i,p}$ (cf (2.1)), the partial sum process $S_i = \sum_{j=1}^{i} X_j$ can be approximated by a Gaussian process G_i with independent but not necessarily identically distributed increments, and the bound of the approximation error can be explicitly given. If the dependence is sufficiently weak, then the approximation error is optimal within a multiplicative logarithmic factor. Such Gaussian approximation results substantially generalize the classical Central Limit Theorem which states that sums of independent random variables that are not necessarily Gaussian, under proper normalization converge to Gaussian distributions.

To develop Gaussian approximations, we need to introduce dependence measures on the underlying process (X_i). To this end, we use the idea of coupling. Let $(\varepsilon_i')_{i\in\mathbb{Z}}$ be an independent copy of $(\varepsilon_i)_{i\in\mathbb{Z}}$. Assume that X_j has mean 0 and $X_j \in \mathcal{L}^p$, $p > 0$. For $j \geq 0$, define the physical dependence measure

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

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

$$\mathcal{F}_{i,\{k\}} = (\ldots, \varepsilon_{k-1}, \varepsilon_k', \varepsilon_{k+1}, \ldots, \varepsilon_{i-1}, \varepsilon_i). \tag{2.2}$$

Note that $\mathcal{F}_{i,\{k\}} = \mathcal{F}_i$ if $k > i$. Wu (2005) introduced a physical dependence measure for stationary processes in which the function $H_i(\cdot)$ does not depend on i. Following Wu (2005), $\|H_i(\mathcal{F}_i) - H_i(\mathcal{F}_{i,\{i-j\}})\|_p$ measures the dependence of X_i on ε_{i-j}. Hence $\delta_{j,p}$ can be interpreted as the uniform lag j dependence measure in the setting of non-stationary processes in which the underlying physical mechanism H_i is time-varying. Assume throughout the paper that

$$\Theta_{m,p} := \sum_{i=m}^{\infty} \delta_{i,p} < \infty. \tag{2.3}$$

The preceding condition implies short-range dependence in the sense that the cumulative dependence of $(X_j)_{j\geq k}$ on ε_k is finite.

Theorem 1 deals with multiple non-stationary nonlinear time series. Earlier, Wu (2007) and Liu and Lin (2009) considered stationary one-dimensional and higher-dimensional processes, respectively.

Theorem 1. *Let $2 < p \leq 4$. Assume that for all i, $\mathbb{E}(X_i) = 0$ and that, for some $c_0 < \infty$, $\sup_i \|X_i\|_p \leq c_0$. Further assume $\Theta_{m,p} = O(m^{-\gamma})$, $\gamma > 0$. Then*

$D_i := \sum_{j=i}^{\infty} \mathcal{P}_i X_j \in \mathcal{L}^p$ and $\sup_i \|D_i\|_p < \infty$. Let $\Sigma_i = \mathbb{E}(D_i D_i^T)$ and assume there exists $\nu_0 > 0$ such that $\Sigma_i - \nu_0 \mathrm{Id}_d$ is positive definite. Then on a richer probability space, there exists a centered Gaussian process $\check{G}_i$ with independent increments and a process $\check{S}_i$ such that $[\check{S}_1, \ldots, \check{S}_n] \overset{\mathcal{D}}{=} [S_1, \ldots, S_n]$ and

$$\max_{1\le i\le n} |\check{S}_i - \check{G}_i| = O_{\mathbb{P}}(\tau_n),$$
$$\text{where } \tau_n = n^{(1/2-1/p+\gamma/p)/(1/2-1/p+\gamma)}(\log n)^{(\gamma+\gamma/p)/(1/2-1/p+\gamma)}. \tag{2.4}$$

Additionally, on a richer probability space, there exists another Gaussian process $\hat{G}_i$ and i.i.d. d-dimensional standard Gaussian random vectors $Y_1, \ldots, Y_n \sim N(0, \mathrm{Id}_d)$ such that $[\hat{G}_1, \ldots, \hat{G}_n] \overset{\mathcal{D}}{=} [\check{G}_1, \ldots, \check{G}_n]$ and

$$\max_{1\le i\le n} |\hat{G}_i - G_i| = O_{\mathbb{P}}(\tau_n), \text{ where } G_i = \sum_{j=1}^{i} \Sigma_j^{1/2} Y_j. \tag{2.5}$$

Larger γ implies weaker dependence. If $\gamma \to \infty$, then the exponent $(1/2 - 1/p + \gamma/p)/(1/2 - 1/p + \gamma)$ in (2.4) converges to $1/p$. In the context of stationary multiple time series, Liu and Lin (2009) obtained the almost sure bound $O_{\mathrm{a.s.}}(n^{1/p})$ when $\gamma \ge (p-2)/(8-2p)+\theta$ for some $\theta > 0$ and some other conditions hold.

Corollary 1. *Let $2 < p \le 4$. Assume that for all i, $\mathbb{E}(X_i) = 0$ and that, for some $c_0 < \infty$, $\sup_i \|X_i\|_p \le c_0$. Further assume $\delta_{m,p} = O(\rho^m)$ for some $\rho \in (0,1)$. Let $D_i = \sum_{j=i}^{\infty} \mathcal{P}_i X_j \in \mathcal{L}^p$, and assume there exists $\nu_0 > 0$ such that $\Sigma_i - \nu_0 \mathrm{Id}_d$ is positive definite. Then on a richer probability space, there exists a centered Gaussian process $\check{G}_i$ with independent increments and a process $\check{S}_i$ such that $[\check{S}_1, \ldots, \check{S}_n] \overset{\mathcal{D}}{=} [S_1, \ldots, S_n]$ and*

$$\max_{1\le i\le n} |\check{S}_i - \check{G}_i| = O_{\mathbb{P}}[n^{1/p}(\log n)^{3/2}]. \tag{2.6}$$

The condition $\delta_{m,p} = O(\rho^m)$ in Corollary 1 is called the *geometric moment contraction* (*GMC*) condition. It is satisfied for many nonlinear stationary time series (Shao and Wu (2007)). In the examples below, we show that GMC holds for a wide class of non-stationary nonlinear time series.

Example 1. Let ε_i be i.id. random elements; let X_i be recursively defined by

$$X_i = F_i(X_{i-1}, \varepsilon_i), \tag{2.7}$$

where F_i satisfies (i) for some x_0, $\sup_i \|F_i(x_0, \varepsilon_1)\|_p < \infty$ and (ii),

$$\sup_i \mathbb{E}(|L_i|^p) < 1, \text{ where } L_i = \sup_{x\ne x'} \frac{|F_i(x,\varepsilon_i) - F_i(x',\varepsilon_i)|}{|x-x'|}. \tag{2.8}$$

Under condition (2.8), iterations of (2.7) ensure that X_i has representation (1.1) and, additionally, we have the GMC $\delta_{m,p} = O(\rho^m)$ with $\rho = \sup_i \|L_i\|_p$. The latter claim can be proved by using the method in Wu and Shao (2004). Details are omitted.

Example 2. (Time-Varying GARCH) Consider the time-varying generalized autoregressive conditional heteroskedasticity (GARCH(1,1)) process

$$Y_t = \varepsilon_t V_t^{1/2}, \quad V_t = c_t + \alpha_t Y_{t-1}^2 + \beta_t V_{t-1},$$

where ε_t are i.i.d. and c_t, α_t, β_t are nonnegative parameters. If c_t, α_t, β_t do not depend on t, then the model becomes Bollerslev's (1986) GARCH(1, 1) process with constant parameters. Time-varying ARCH models have received some attention recently; see Fryzlewicz, Sapatinas, and Subba Rao (2008) and Dahlhaus and Subba Rao (2006). In the latter two papers the parametrization $\alpha_t = \alpha(t/n)$ is used, where $\alpha(\cdot)$ is a continuous function.

Let $X_t = (Y_t, V_t)^T - \mathbb{E}(Y_t, V_t)^T$. Here we give conditions for which Corollary 1 is applicable. Let $W_t = (Y_t^2, V_t)^T$. As in Bougerol and Picard (1992), we can write

$$W_t = M_t W_{t-1} + c_t \begin{pmatrix} \varepsilon_t^2 \\ 1 \end{pmatrix}, \text{ where } M_t = \begin{pmatrix} \alpha_t \varepsilon_t^2 & \beta_t \varepsilon_t^2 \\ \alpha_t & \beta_t \end{pmatrix}.$$

Then M_t has two eigenvalues: 0 and $\alpha_t \varepsilon_t^2 + \beta_t$. Assume $\mathbb{E}(\varepsilon_t) = 0$, $\mathbb{E}(\varepsilon_t^2) = 1$, $\varepsilon_t \in \mathcal{L}^{2p}$, $2 < p \le 4$, $\sup_t c_t < \infty$, and $\sup_t \|\alpha_t \varepsilon_t^2 + \beta_t\|_p < 1$. Then (2.8) holds for (W_t) and (W_t) has the GMC property. By Lemma 1 in Wu and Shao (2004), the process $(V_t^{1/2})$ also satisfies GMC. So (X_t) has the GMC property as well. We now compute $D_i = \sum_{t=i}^{\infty} \mathcal{P}_i X_t$. Without loss of generality let $i = 0$. Let $d_t = \alpha_t + \beta_t$ and $g_t = \alpha_{t+1} \sum_{j=0}^{\infty} \prod_{l=1}^{j} d_{t+l}$. Note that V_i is $\mathcal{F}_{i-1}$-measurable. If $t \ge 2$, then $\mathcal{P}_0 V_t = d_t \mathcal{P}_0 V_{t-1}$, since $\mathcal{P}_0 Y_{t-1}^2 = \mathcal{P}_0(\varepsilon_{t-1}^2 V_{t-1}) = \mathcal{P}_0 V_{t-1}$ and $\mathbb{E}(\varepsilon_{t-1}^2) = 1$. Since $\mathcal{P}_0 V_1 = \alpha_1 \mathcal{P}_0(\varepsilon_0^2 V_0) = \alpha_1(\varepsilon_0^2 - 1)V_0$, $D_0 = (Y_0, g_0(\varepsilon_0^2 - 1)V_0)^T$. Let $\Sigma_0 = \mathbb{E}(D_0 D_0^T)$ have entries $\sigma_{11} = \mathbb{E}(Y_0^2) = \mathbb{E}V_0$, $\sigma_{22} = \mathbb{E}(g_0^2(\varepsilon_0^2 - 1)^2 V_0^2)$, and $\sigma_{12} = \mathbb{E}(Y_0 g_0(\varepsilon_0^2 - 1)V_0) = g_0 \mathbb{E}(V_0^{3/2})\mathbb{E}(\varepsilon_0^3 - \varepsilon_0)$. By Schwarz's inequality,

$$\frac{\sigma_{12}^2}{\sigma_{11}\sigma_{22}} = \frac{(\mathbb{E}(V_0^{3/2}))^2(\mathbb{E}(\varepsilon_0^3 - \varepsilon_0))^2}{(\mathbb{E}V_0)(\mathbb{E}(\varepsilon_0^2 - 1)^2)(\mathbb{E}(V_0^2))} \le \tau, \text{ where } \tau = \frac{(\mathbb{E}(\varepsilon_0^3 - \varepsilon_0))^2}{\mathbb{E}(\varepsilon_0^2)\mathbb{E}(\varepsilon_0^2 - 1)^2} \le 1.$$

Assume $\tau < 1$. Then the smaller eigenvalue of Σ_0 is

$$\frac{\sigma_{11} + \sigma_{22} - \sqrt{(\sigma_{11} - \sigma_{22})^2 + 4\sigma_{12}^2}}{2} \ge \frac{\sigma_{11}\sigma_{22} - \sigma_{12}^2}{\sigma_{11} + \sigma_{22}} \ge \frac{1 - \tau}{\sigma_{11}^{-1} + \sigma_{22}^{-1}}.$$

Note that $v_t = \mathbb{E}V_t$ satisfies the recursion $v_t = c_t + d_t v_{t-1}$. Then $v_t = c_t + d_t c_{t-1} + d_t d_{t-1} c_{t-2} + \ldots$. Assume $v_* := \inf_t v_t > 0$ and $g_* := \inf_t g_t > 0$. So $\Sigma_0 - \nu_0 \mathrm{Id}_2$ is positive definite for any $\nu_0 > 0$ for which $\nu_0 < (1-\tau)/(1/v_* + 1/(g_*^2 v_*^2 \mathbb{E}(\varepsilon_0^2 - 1)^2)))$.

Theorem 1 and Corollary 1 concern nonlinear multiple time series. It turns out that, for the special case of linear multiple time series, we can obtain a better bound.

Theorem 2. *Let ε_i, $i \in \mathbb{Z}$, be i.i.d. d-dimensional random vectors with mean* 0*, covariance matrix* Id_d*, and $\varepsilon_i \in \mathcal{L}^p$, $p > 2$. Let $A_{i,j}$ be $d \times d$ matrices satisfying $\sum_{j=m}^{\infty} \sup_{l \geq j} |A_{l,j}| = O(m^{-\gamma})$, $\gamma > 0$. Let $B_i = \sum_{l=i}^{\infty} A_{l,l-i}$ and*

$$X_i = \sum_{j=0}^{\infty} A_{i,j} \varepsilon_{i-j}. \tag{2.9}$$

Assume that there exists $\nu_0 > 0$ such that $B_i B_i^T - \nu_0 \mathrm{Id}_d$, is positive definite for all i. Then on a richer probability space, there exists i.id. d-dimensional standard Gaussian random vectors $Y_1, \ldots, Y_n \sim N(0, \mathrm{Id}_d)$ and a process $\check{S}_i$ such that $[\check{S}_1, \ldots, \check{S}_n] \overset{\mathcal{D}}{=} [S_1, \ldots, S_n]$ and

$$\max_{1 \leq i \leq n} |\check{S}_i - G_i| = O_{\mathbb{P}}[n^{1/2-\gamma} + n^{1/p} (\log n)^{1-2/p}], \ \textit{where } G_i = \sum_{j=1}^{i} B_j Y_j. \tag{2.10}$$

An important class of non-stationary processes is the so-called locally stationary processes (Dahlhaus (1997); Draghicescu, Guillas, and Wu (2009)). Consider the process

$$X_i = H(\frac{i}{n}; \mathcal{F}_i), \quad 1 \leq i \leq n, \tag{2.11}$$

where $H(\cdot; \mathcal{F}_i)$ is stochastic continuous in the sense that, for any $t_0 \in [0, 1]$,

$$\lim_{t \to t_0} \|H(t; \mathcal{F}_i) - H(t_0; \mathcal{F}_i)\| = 0. \tag{2.12}$$

Additionally, we say that $H(\cdot; \mathcal{F}_i)$ is *stochastic Lipschitz continuous* if there exists a constant $L_H < \infty$ such that

$$\sup_{0 \leq t < t' \leq 1} \frac{\|H(t; \mathcal{F}_i) - H(t'; \mathcal{F}_i)\|}{|t - t'|} \leq L_H. \tag{2.13}$$

For the process (2.11), the physical mechanism $H(\cdot; \cdot)$ generating X_i is time-varying and the stochastic continuity condition (2.12) or (2.13) suggests locally stationarity, namely the underlying physical mechanism is changing smoothly. Such processes appear frequently in practice (Mallat, Papanicolaou, and Zhang (1998)).

Corollary 2. *Let $2 < p \leq 4$. Assume* (2.13)*, that for all i, $\mathbb{E}(X_i) = 0$, and that for some $c_0 < \infty$, $\sup_i \|X_i\|_p \leq c_0$. Suppsoe $\Theta_{m,p} = O(m^{-\gamma})$, $\gamma > 0$, and let*

$$D_i(t) = \sum_{j=i}^{\infty} \mathcal{P}_i H(t; \mathcal{F}_j), \quad 0 \leq t \leq 1. \tag{2.14}$$

If $\Sigma(t) = \mathbb{E}[D_i(t) D_i^T(t)]$ is positive definite for all $t \in [0,1]$, then (2.4) *and* (2.5) *of Theorem 1 hold, and*

$$\max_{1 \leq i \leq n} |\bar{G}_i - G_i| = O_{\mathbb{P}}(n^{1/2-\gamma/(2+\gamma)}), \ \textit{where}\ \bar{G}_i = \sum_{j=1}^{i} \Sigma^{1/2}(\frac{j}{n}) Y_j. \tag{2.15}$$

Remark 1. In our Gaussian approximations (2.4) and (2.10), we obtain in probability bounds $O_{\mathbb{P}}(\cdot)$, while the classical strong invariance principle usually asserts almost sure bounds. Using the argument in Liu and Lin (2009), it is possible to derive almost sure bounds. We decide not to pursue this direction of research since the derivation is very tedious and since the bounds $O_{\mathbb{P}}(\cdot)$ are typically powerful enough for one to obtain asymptotic distributions of statistics involving partial sum processes. Additionally, the strong approximation scheme does not seem to be suitable for processes of type (2.11) which have a triangular array form.

Example 3. Functional linear models. Consider the functional linear regression model

$$y_i = \mathbf{x}_i^T \boldsymbol{\beta}_i + \varepsilon_i, \quad i = 1, 2, \ldots, n, \tag{2.16}$$

where $\mathbf{x}_i \in \mathbb{R}^p$ are regressors, $\mathbb{E}(\varepsilon_i | \mathbf{x}_i) = 0$, and the regression parameter $\boldsymbol{\beta}_i = \boldsymbol{\beta}(i/n)$ is a smooth function. The functional linear model (2.16) has many applications in climatology, finance, econometrics, and other areas. With the regression parameter $\boldsymbol{\beta}(\cdot)$ being time-varying, one is able to explore the dynamic or time-varying associations between the response series $\{y_i\}$ and the explanatory series $\{\mathbf{x}_i\}$; see Robinson (1989, 1991), Orbe, Ferreira, and Rodriguez-Poo (2005, 2006), and Cai (2007).

A fundamental problem in the inference of functional linear models is to test whether $\boldsymbol{\beta}(\cdot)$ is of a certain parametric form. To this end, one needs to construct simultaneous confidence regions (SCR) (Fan and Zhang (2008)). The construction of SCR with asymptotically correct coverage probabilities requires (i) an extreme value theory of Gaussian processes, and (ii) a Gaussian approximation result for the partial sum process. Wu, Chiang, and Hoover (1998) mention that the fundamental difficulty is (ii), a lack of development of Gaussian approximations in the presence of dependence. With Corollary 2, (ii) is solved. With the latter result, Zhou and Wu (2010) applied Lindgren's (1980) extreme value theory for Gaussian processes and constructed SCR for $\boldsymbol{\beta}(\cdot)$.

3. Proofs

The proof of Theorem 1 is quite complicated. For stationary processes with $d = 1$, Wu (2007) established strong invariance principles by first approximating S_n by martingales, and then used Strassen's embedding results and approximated martingales by Brownian motions. For multiple time series with $d \geq 2$, the martingale approximation technique alone does not work well since, generally speaking, vector-valued martingales cannot be embedded into a Gaussian process (Monrad and Philipp (1991)). Here we apply the martingale approximation and approximate X_i by m-dependent processes. The latter technique has been applied in Liu and Lin (2009). Section 3.1 presents a proof of Theorem 1. It uses results in Section 3.4 concerning tail probabilities of martingales and bounds for martingale approximations.

The proof of Theorem 2 is relatively easy, due to the special linearity structure of X_i in (2.9). In this case the approximating martingale is a sum of independent random vectors that can be further approximated by Gaussian processes by applying results in Zaitsev (2001). A detailed derivation is given in Section 3.2.

3.1. Proof of Theorem 1

By Theorem 1 in Wu (2005), $\|\mathcal{P}_i X_j\|_p \leq \delta_{j-i,p}$. So $\|D_i\|_p \leq \Theta_{0,p}$.
Let $\ell = \lfloor (n \log^{-3} n)^{1/(1+4\gamma)} \rfloor$ and

$$\tilde{S}_i = \sum_{j=1}^{i} \tilde{X}_j, \text{ where } \tilde{X}_j = \mathbb{E}(X_j | \varepsilon_j, \ldots, \varepsilon_{j-\ell}). \tag{3.1}$$

By Lemma A1 in Liu and Lin (2009) (see also Theorem 1(ii) in Wu (2007)), we have

$$\left\| \max_{1 \leq i \leq n} |\tilde{S}_i - S_i| \right\|_p \leq C_p \sqrt{n} \Theta_{\ell,p} = O(\sqrt{n} \ell^{-\gamma}) \tag{3.2}$$

for a constant C_p depending only on p. (The proofs in Lemma A1 in Liu and Lin (2009) and Theorem 1(ii) in Wu (2007) are for stationary processes. It is easily seen that their arguments are also valid for nonstationary processes of form (1.1).)

Next we approximate $\tilde{S}_i$ by the martingale

$$\tilde{M}_i = \sum_{j=1}^{i} \tilde{D}_j, \text{ where } \tilde{D}_j = \sum_{k=j}^{\infty} \mathcal{P}_j \tilde{X}_k = \sum_{k=j}^{j+\ell} \mathcal{P}_j \tilde{X}_k. \tag{3.3}$$

Note that $\mathcal{P}_j \tilde{X}_k = 0$ if $k > j + \ell$. Let $\tilde{\delta}_{j,p}$ be the physical dependence measure for $(\tilde{X}_k)_{k\in\mathbb{Z}}$. Observe that $\tilde{\delta}_{j,p} = 0$ if $j > \ell$. Let $\tilde{R}_n = \tilde{S}_n - \tilde{M}_n$ and $\tilde{\Theta}_{i,p} = \sum_{j=i}^{\infty} \tilde{\delta}_{j,p} = \sum_{j=i}^{\ell} \tilde{\delta}_{j,p}$. By Lemma 2, we have

$$\|\tilde{R}_h\|_p^2 \le C_p \sum_{i=1}^{h} \tilde{\Theta}_{i,p}^2 = C_p \sum_{i=1}^{\ell\wedge h} \tilde{\Theta}_{i,p}^2 = \begin{cases} O[\log(\ell\wedge h)] & \text{if } \gamma = \frac{1}{2}, \\ O[1 + (\ell\wedge h)^{1-2\gamma}] & \text{if } \gamma \ne \frac{1}{2}. \end{cases} \tag{3.4}$$

Let $g \in \mathbb{N}$ satisfy $2^{g-1} < n \le 2^g$. If $\gamma < 1/2$, by Proposition 1(i) in Wu (2007), (3.4) implies

$$\begin{aligned} \left\| \max_{i\le 2^g} |\tilde{R}_i| \right\|_p &\le \sum_{r=0}^{g} \left[\sum_{k=1}^{2^{g-r}} \|\tilde{R}_{2^r k} - \tilde{R}_{2^r(k-1)}\|_p^p \right]^{1/p} \\ &\le \sum_{r=0}^{g} [2^{g-r} (\ell \wedge 2^r)^{(1-2\gamma)p/2}]^{1/p}. \end{aligned} \tag{3.5}$$

Observe that the above bound is $O(2^{g/p})$, $O(2^{g/p} g)$, and $O(2^{g/p}\ell^{1/2-\gamma-1/p})$, for $1/p > 1/2 - \gamma$, $1/p = 1/2 - \gamma$, and $1/p < 1/2 - \gamma$, respectively. Since $\ell = o(n)$, we have

$$\left\| \max_{i\le 2^g} |\tilde{R}_i| \right\|_p = O(n^{1/p} + n^{1/2}\ell^{-\gamma}). \tag{3.6}$$

Elementary manipulations show that (3.6) also holds for $\gamma \ge 1/2$ by using (3.4) and (3.5).

Let $m = \lfloor \ell^{1+2\gamma} \rfloor$, $B_1 = \{1, \ldots, m\}$, $A_1 = \{m+1, \ldots, m+3l\}$ and, for $j \ge 2$,

$$B_j = \{i + (j-1)(m+3\ell),\ i \in B_1\}, \quad A_j = \{i + (j-1)(m+3\ell),\ i \in A_1\}.$$

Further let $\underline{b}_j = \min\{k : k \in B_j\}$, $\bar{b}_j = \max\{k : k \in B_j\}$, $A = \cup_{j=1}^{\infty} A_j$, $B = \cup_{j=1}^{\infty} B_j$, $\tilde{U}_j = \sum_{i\in A_j} \tilde{D}_i$, and $\tilde{V}_j = \sum_{i\in B_j} \tilde{D}_i$. Then $\tilde{U}_1, \tilde{U}_2, \ldots$ are independent and $\tilde{V}_1, \tilde{V}_2, \ldots$ are also independent. Let $\tilde{M}_{i,A} = \sum_{j=1}^{i} \tilde{D}_j \mathbf{1}_{j\in A}$ and $\tilde{M}_{i,B} = \sum_{j=1}^{i} \tilde{D}_j \mathbf{1}_{j\in B}$. Observe that $A \cap B = \emptyset$, $A \cup B = \mathbb{N}$, and $\tilde{M}_i = \tilde{M}_{i,A} + \tilde{M}_{i,B}$.

Since $\tilde{D}_j \mathbf{1}_{j\in A}$, $j = 1, 2, \ldots$, are martingale differences and $p/2 > 1$, by Doob's martingale inequality (cf Chow and Teicher (1988)) and Theorem 2.1 in Rio (2009),

$$\begin{aligned} \left\| \max_{j\le n} |\tilde{M}_{j,A}| \right\|_p^2 &\le C_p \|\tilde{M}_{n,A}\|_p^2 \\ &\le C_p (p-1) \sum_{j=1}^{n} \|\tilde{D}_j \mathbf{1}_{j\in A}\|_p^2 = O(\ell \frac{n}{m}). \end{aligned} \tag{3.7}$$

Let $f_i = \max\{\bar{b}_k : \bar{b}_k \le i\}$, $k_i = \max\{k : \bar{b}_k \le i\}$, and $H_i = \tilde{M}_{i,B} - \tilde{M}_{f_i,B}$, $i \in \mathbb{N}$; let $q_n > 0$ satisfy $q_n \to \infty$ and $\lambda_n = n^{1/p}\ell^{1/2-1/p}(\log n)^{1/p} + (m \log n)^{1/2}$. We now apply Lemma 1 with $u = \lambda_n q_n$. Choose $d \in \mathbb{N}$ such that $2^{d-1} < m \le 2^d$. Let $L_i = \tilde{D}_1 + \cdots + \tilde{D}_i$ and $\Delta_r = \max_{1\le j\le 2^{d-r}} |L_{2^r j} - L_{2^r(j-1)}|$, $0 \le r \le d$. By the argument in the proof of Proposition 1(i) in Wu (2007), we have $\max_{i\le m}|L_i| \le \sum_{r=0}^{d} \Delta_r$. Hence, by Lemma 1,

$$\mathbb{P}\left(\max_{j\le m}|L_j| > 2du\right) \le \sum_{r=0}^{d} \mathbb{P}(\Delta_r > u) \le \sum_{r=0}^{d} 2^{d-r}\frac{O[\min(2^{rp/2}, 2^r\ell^{p/2-1})]}{u^p}. \quad (3.8)$$

Since $k_n \sim n/(m+3\ell) \sim n/m$ and $m = \lfloor \ell^{1+2\gamma} \rfloor$, by elementary calculations, (3.8) implies

$$\begin{aligned}\mathbb{P}\left(\max_{i\le n}|H_i| > 2du\right) &\le \sum_{k=1}^{k_n} \mathbb{P}\left(\max_{l\in B_k}|\tilde{D}_{\underline{b}_k} + \cdots + \tilde{D}_l| > 2du\right)\\ &\le O(k_n)\sum_{r=0}^{d} 2^{d-r}\frac{O[\min(2^{rp/2}, 2^r\ell^{p/2-1})]}{u^p} = o(1). \quad (3.9)\end{aligned}$$

Since in $u = \lambda_n q_n$ the speed of $q_n \to \infty$ can be arbitrarily slow, by (3.9), we have

$$\max_{i\le n}|H_i| = \max_{i\le n}|\tilde{M}_{i,B} - \tilde{M}_{f_i,B}| = O_{\mathbb{P}}(\lambda_n d) = O_{\mathbb{P}}(\lambda_n \log n). \quad (3.10)$$

Observe that $\tilde{M}_{f_i,B} = \sum_{j=1}^{k_i} \tilde{V}_j = I_{k_i} + J_{k_i}$, where

$$I_k = \sum_{j=1}^{k} \frac{\tilde{V}_j \mathbf{1}_{|\tilde{V}_j|<Q\lambda_n} - \mathbb{E}(\tilde{V}_j \mathbf{1}_{|\tilde{V}_j|<Q\lambda_n})}{\sqrt{m}},$$

$$J_k = \sum_{j=1}^{k} \frac{\tilde{V}_j \mathbf{1}_{|\tilde{V}_j|\ge Q\lambda_n} - \mathbb{E}(\tilde{V}_j \mathbf{1}_{|\tilde{V}_j|\ge Q\lambda_n})}{\sqrt{m}}.$$

Since summands of I_k are independent and bounded by $Q\lambda_n/\sqrt{m}$, by Thereom 1.2 in Zaitsev (2001) there exists i.i.d. standard d-dimensional Gaussian random vectors $Z_1, \ldots, \sim N(0, \mathrm{Id}_d)$ such that

$$\max_{i\le n}\Big|I_{k_i} - \sum_{j=1}^{k_i} F_j Z_j\Big| = \max_{k\le k_n}\Big|I_k - \sum_{j=1}^{k} F_j Z_j\Big| = O_{\mathbb{P}}\Big[\Big(\frac{Q\lambda_n}{\sqrt{m}}\Big)\log n\Big], \quad (3.11)$$

where $F_i = m^{-1/2}[\mathbb{E}(\tilde{W}_i \tilde{W}_i^T)]^{1/2}$, and $\tilde{W}_i = \tilde{V}_i \mathbf{1}_{|\tilde{V}_i|<Q\lambda_n} - \mathbb{E}(\tilde{V}_i \mathbf{1}_{|\tilde{V}_i|<Q\lambda_n})$.

By Lemma 1, for sufficiently large $Q > 0$,

$$\mathbb{E}(|V_k|\mathbf{1}_{|V_k|\ge Q\lambda_n}) = \int_{Q\lambda_n}^{\infty} \mathbb{P}(|V_k| \ge u)du = \int_{Q\lambda_n}^{\infty} \frac{O(m\ell^{p/2-1})}{u^p}du. \quad (3.12)$$

Hence $\mathbb{E}(|V_k|\mathbf{1}_{|V_k|\geq Q\lambda_n}) = O(m\ell^{p/2-1}\lambda_n^{1-p})$, and by the definition of λ_n,

$$\mathbb{E}\left[\max_{i\leq n}|J_{k_i}|\right] = \mathbb{E}\left[\max_{k\leq k_n}|J_k|\right] \leq \sum_{k=1}^{k_n}\frac{2\mathbb{E}(|V_k|\mathbf{1}_{|V_k|\geq Q\lambda_n})}{\sqrt{m}} = \frac{O(\lambda_n)}{\sqrt{m}}. \quad (3.13)$$

To conclude the proof, we write

$$S_i = (S_i - \tilde{S}_i) + \tilde{R}_i + \tilde{M}_{i,A} + H_i + \sqrt{m}J_{k_i} + \sqrt{m}I_{k_i}. \quad (3.14)$$

By (3.2), (3.6), (3.7), (3.10), (3.11), and (3.13), after some elementary manipulations, we have (2.4) by letting $\breve{G}_i = \sum_{j=1}^{k_i} F_j Z_j \sqrt{m}$, and the approximation error is

$$O_{\mathbb{P}}(\sqrt{n}\ell^{-\gamma}) + O_{\mathbb{P}}(n^{1/p} + n^{1/2}\ell^{-\gamma}) + O_{\mathbb{P}}[(\frac{\ell n}{m})^{1/2}] + O_{\mathbb{P}}(\lambda_n \log n) = O_{\mathbb{P}}(\tau_n). \quad (3.15)$$

Clearly $\breve{G}_i$ has independent increments.

We now prove (2.5). By (3.2), (3.6), and Lemma 2, we have

$$\begin{aligned}\|M_n - \tilde{M}_n\| &\leq \|\tilde{S}_n - S_n\| + \|\tilde{S}_n - \tilde{M}_n\| + \|M_n - S_n\| \\ &= O(\sqrt{n}\Theta_{\ell,p}) + O(1)\Big[\sum_{i=1}^{\ell}\tilde{\Theta}_{i,p}^2\Big]^{1/2} + O(1)\Big[\sum_{i=1}^{n}\Theta_{i,p}^2\Big]^{1/2} \\ &= O(\sqrt{n}\ell^{-\gamma}) + \Big[\sum_{i=1}^{n}O(i^{-2\gamma})\Big]^{1/2} = O(\phi_n), \end{aligned} \quad (3.16)$$

where $\phi_n = \sqrt{n}\ell^{-\gamma} + (\sum_{i=1}^n i^{-2\gamma})^{1/2}$. Recall $\tilde{V}_j = \sum_{i\in B_j}\tilde{D}_i$. Let $V_j = \sum_{i\in B_j} D_i$. By (3.16) and the orthogonality of martingale differences,

$$\sum_{j=1}^{k_n}\|\tilde{V}_j - V_j\|^2 = O(\phi_n^2). \quad (3.17)$$

Since $\Sigma_i - \nu_0\mathrm{Id}_d$ is positive definite, $m^{-1}\mathbb{E}(V_iV_i^T) - \nu_0\mathrm{Id}_d = m^{-1}\sum_{i\in B_j}\Sigma_i - \nu_0\mathrm{Id}_d$ is also positive definite. Let $\Psi_i = (\tilde{W}_i - V_i)/\sqrt{m}$. Elementary matrix manipulations show that

$$|m^{-1/2}[\mathbb{E}(V_iV_i^T)]^{1/2} - \{\mathbb{E}[(\frac{V_i}{\sqrt{m}} + \Psi_i)(\frac{V_i^T}{\sqrt{m}} + \Psi_i^T)]\}^{1/2}| = O(\|\Psi_i\|). \quad (3.18)$$

Similarly as (3.12), by Lemma 1, for sufficiently large Q,

$$\begin{aligned}\|\tilde{W}_i - \tilde{V}_i\|^2 &= \|\tilde{V}_i\mathbf{1}_{|\tilde{V}_i|\geq Q\lambda_n} - \mathbb{E}(\tilde{V}_i\mathbf{1}_{|\tilde{V}_i|\geq Q\lambda_n})\|^2 \leq \|\tilde{V}_i\mathbf{1}_{|\tilde{V}_i|\geq Q\lambda_n}\|^2 \\ &= 2\int_{Q\lambda_n}^{\infty} u\mathbb{P}(|\tilde{V}_i| \geq u)du = O(m\ell^{p/2-1}\lambda_n^{2-p}).\end{aligned}$$

Let $\theta_n = k_n m \ell^{p/2-1} \lambda_n^{2-p} + \phi_n^2$. Hence by (3.17),

$$\sum_{j=1}^{k_n} |m^{-1/2}[\mathbb{E}(V_i V_i^T)]^{1/2} - F_i|^2 = \sum_{j=1}^{k_n} O(\|\Psi_i\|^2) = O(m^{-1})\theta_n. \tag{3.19}$$

Let $Y_1, Y_2, \ldots, \sim N(0, \mathrm{Id}_d)$ be i.i.d. d-dimensional standard Gaussian random vectors and $G_i^\circ = \sum_{j=1}^{k_i} \sum_{l \in B_j} \Sigma_l^{1/2} Y_l$ and $G_i = \sum_{j=1}^{i} \Sigma_j^{1/2} Y_j$; let

$$\hat{G}_i = \sum_{j=1}^{k_i} \sqrt{m} F_j \Big(\sum_{l \in B_j} \Sigma_l \Big)^{-1/2} \sum_{l \in B_j} \Sigma_l^{1/2} Y_l.$$

Since Y_i are i.i.d. standard Gaussian random vectors, $(\sum_{l \in B_j} \Sigma_l)^{-1/2} \sum_{l \in B_j} \Sigma_l^{1/2} Y_l$, $j = 1, 2, \ldots$, are also i.i.d. standard Gaussian random vectors. Hence by (3.19),

$$\max_{i \le n} |G_i^\circ - \hat{G}_i| = O_{\mathbb{P}}(\theta_n^{1/2}).$$

Observe that $[\hat{G}_1, \ldots, \hat{G}_n] \stackrel{\mathcal{D}}{=} [\check{G}_1, \ldots, \check{G}_n]$ and, as in (3.7),

$$\max_{i \le n} |G_i^\circ - G_i| = \max_{i \le n} \Big| \sum_{j=1}^{i} \Sigma_j^{1/2} Y_j \mathbf{1}_{j \in A} \Big| = O_{\mathbb{P}}\Big(\sqrt{\frac{\ell n}{m}} \Big).$$

By elementary manipulations, $\theta_n = O(\tau_n^2)$, and also $\ell n / m = O(\tau_n^2)$. So (2.5) follows.

Proof of Corollary 1. In the proof of Theorem 1, we let $m = \lfloor \ell^{1/2} n^{1/2} (\log n)^{-3/2} \rfloor$ and $\ell = \lfloor c \log n \rfloor$, where $c > 0$ is a sufficiently large constant. Then by the same argument of Theorem 1, we have (2.6) in view of (3.15).

Remark 2. In Theorem 1, we call the condition that $\Sigma_i - \nu_0 \mathrm{Id}_d$ is positive definite for all i *the uniform positive definiteness condition*. A careful check of the proof of Theorem 1 reveals that the uniform positive definiteness condition can be replaced by the following slightly weaker version: there exists $m_0 \in \mathbb{N}$ such that $\sum_{j=i}^{i+m_0} \Sigma_j - \nu_0 \mathrm{Id}_d$ is positive definite for all i.

3.2. Proof of Theorem 2

Observe that $\delta_{j,p} = \sup_{l \ge j} |A_{l,j}| \|\varepsilon_0 - \varepsilon_0'\|_p$. Let $R_h = S_h - M_h$, where $M_h = \sum_{i=1}^{h} D_i$ and $D_i = B_i \varepsilon_i$ are independent random vectors. By Lemma 2, since $\Theta_{i,p} = O(i^{-\gamma})$,

$$\|R_h\|_p^2 = \sum_{i=1}^{h} O(i^{-2\gamma}) = \begin{cases} O(\log h) & \text{if } \gamma = \frac{1}{2}, \\ O(1 + h^{1-2\gamma}) & \text{if } \gamma \ne \frac{1}{2}. \end{cases} \tag{3.20}$$

As in (3.4) and (3.5), for $g \in \mathbb{N}$ satisfying $2^{g-1} < n \le 2^g$, we have

$$\begin{aligned}\Big\|\max_{i\le 2^g}|R_i|\Big\|_p &\le \sum_{r=0}^{g}\Big[\sum_{k=1}^{2^{g-r}}\|R_{2^r k}-R_{2^r(k-1)}\|_p^p\Big]^{1/p} \\ &= O\{\max[2^{g/p}, 2^{g(1/2-\gamma)}]\} + O(g2^{g/p}\mathbf{1}_{1/p=1/2-\gamma}) \qquad (3.21)\end{aligned}$$

by considering the cases $1/p > 1/2-\gamma$, $1/p = 1/2-\gamma$, and $1/p < 1/2-\gamma$, respectively. To establish (2.10), as in the proof of Theorem 1, we use a truncation argument. Let $\lambda_n' = n^{1/p}(\log n)^{-2/p}$, $q_n \to \infty$, and

$$M_h^\diamond = \sum_{i=1}^h B_i\eta_i, \text{ where } \eta_i = \varepsilon_i\mathbf{1}_{|\varepsilon_i|\le q_n\lambda_n'} - \mathbb{E}(\varepsilon_i\mathbf{1}_{|\varepsilon_i|\le q_n\lambda_n'}). \qquad (3.22)$$

Let $\Upsilon_n = [\mathbb{E}(\eta_i\eta_i^T)]^{1/2}$. As in (3.11), since summands of $M_h^\diamond$ are independent and bounded by $2|B_i|q_n\lambda_n'$, by Theorem 1.2 in Zaitsev (2001) there exists i.i.d. standard d-dimensional Gaussian random vectors $Z_1,\ldots,\sim N(0,\mathrm{Id}_d)$ such that

$$\max_{h\le n}|M_h^\diamond - G_h^\diamond| = O_{\mathbb{P}}(q_n\lambda_n'\log n), \text{ where } G_h^\diamond = \sum_{i=1}^h B_i\Upsilon_n Z_i. \qquad (3.23)$$

On the other hand, by Doob's inequality,

$$\begin{aligned}\Big\|\max_{h\le n}|M_h - M_h^\diamond|\Big\|^2 &= \sum_{i=1}^n O(1)\|B_i[\varepsilon_i\mathbf{1}_{|\varepsilon_i|>q_n\lambda_n'} - \mathbb{E}(\varepsilon_i\mathbf{1}_{|\varepsilon_i|>q_n\lambda_n'})]\|^2 \\ &= \sum_{i=1}^n O(1)(q_n\lambda_n')^{2-p}\mathbb{E}(|\varepsilon_i|^p) = o[n(\lambda_n')^{2-p}]. \qquad (3.24)\end{aligned}$$

Since $\mathbb{E}(\varepsilon_i\varepsilon_i^T) = \mathrm{Id}_d$ and $\varepsilon_i \in \mathcal{L}^p$, as in (3.18), we have by Markov's inequality that $|\Upsilon_n - \mathrm{Id}_d| = O(\|\varepsilon_i\mathbf{1}_{|\varepsilon_i|>q_n\lambda_n'}\|) = O[(q_n\lambda_n')^{1-p/2}]$. Hence

$$\mathbb{E}\Big[\max_{h\le n}|G_h - G_h^\diamond|^2\Big] = \sum_{i=1}^n O\{[(q_n\lambda_n')^{1-p/2}]^2\} = O[q_n^{2-p}n^{2/p}(\log n)^{2-4/p}]. \quad (3.25)$$

By (3.21), (3.23), (3.24), and (3.25), we conclude that (2.10) holds since $q_n \to \infty$ can be arbitrarily slow.

3.3. Proof of Corollary 2

Under the condition $\Theta_{m,p} = O(m^{-\gamma})$, $\gamma > 0$, we have $D_i = \sum_{j=i}^\infty \mathcal{P}_i X_j \in \mathcal{L}^p$, and also $D_i(t) \in \mathcal{L}^p$. Let $l_n = \lfloor n^{1/(2+\gamma)}\rfloor$. By (2.13), we have

$$\|D_i(\frac{i}{n}) - D_i\| \le \sum_{j=i}^\infty \|\mathcal{P}_i[H(\frac{i}{n};\mathcal{F}_j) - H(\frac{j}{n};\mathcal{F}_j)]\|$$

$$\leq \sum_{j=i}^{\infty} \min(\frac{L_H(j-i)}{n}, 2\delta_{j-i,p})$$
$$\leq \sum_{j=i}^{i+l_n} \frac{L_H(j-i)}{n} + \sum_{j=i+l_n+1}^{\infty} 2\delta_{j-i,p} = O(n^{-\gamma/(2+\gamma)}). \quad (3.26)$$

Similarly, $\|D_i(t) - D_i(t_0)\| \to 0$ as $t \to t_0$. So $\Sigma(t)$ is continuous in t. Since $\Sigma(t) = \mathbb{E}[D_i(t)D_i^T(t)]$ is positive definite for all $t \in [0,1]$, there exists $\nu_0 > 0$ such that $\Sigma(t) - 2\nu_0 \mathrm{Id}_d$ is positive definite for all $t \in [0,1]$. By (3.26), the condition that $\Sigma_j - \nu_0 \mathrm{Id}_d$ is positive definite in Theorem 1 is satisfied, and hence (2.4) and (2.5) hold.

To prove (2.15), using the argument in (3.18) and (3.19), by (3.26), we have

$$\sum_{i=1}^{n} |\Sigma_i^{1/2} - \Sigma^{1/2}(\frac{j}{n})|^2 = nO(n^{-\gamma/(2+\gamma)})^2.$$

Hence we have (2.15) since $Y_1, Y_2, \ldots,$ are i.i.d. standard Gaussian random vectors.

3.4. Some useful results

Lemma 1. *Let D_i, $i \geq 1$, be ℓ-dependent martingale differences with respect to the filter $\mathcal{G}_i$; let $T_m = \sum_{i=1}^m D_i$. Assume $\sup_i \|D_i\|_p \leq f_0 < \infty$, $p > 2$. Then there exist constants $C_{1,p}$ and $C_{2,p}$, depending only on p, such that*

$$\mathbb{P}(|T_m| \geq u) \leq C_{1,p} \frac{\min(m^{p/2}, m\ell^{p/2-1})}{u^p} f_0^p \quad \text{if} \quad u \geq C_{2,p}(m \log m)^{1/2} f_0. \quad (3.27)$$

Proof. By Burkholder's and Minkowski's inequalities, $\|T_m\|_p \leq C_{3,p} f_0 m^{1/2}$. Here and below $C_{3,p}, C_{4,p}, \ldots$ are constants only depending on p. So $\mathbb{P}(|T_m| \geq u) \leq C_{3,p}^p f_0^p m^{p/2}/u^p$, which implies (3.27) if $m \leq 4\ell$. If $m > 4\ell$, let $A_k = \sum_{i=1+2\ell(k-1)}^{2\ell k-\ell} D_i$ and $B_k = \sum_{i=1-\ell+2\ell k}^{2\ell k} D_i$, $1 \leq k \leq K$, where $K = \lfloor m/(2\ell) \rfloor$. For technical convenience we assume that $m/(2\ell)$ is an integer. Let $S_A = \sum_{k=1}^K A_k$ and $S_B = \sum_{k=1}^K B_k$. Since $\|A_k\|_p \leq C_{3,p} f_0 \ell^{1/2}$, $\sum_{k=1}^K \|A_k\|_p^p \leq C_{3,p}^p f_0^p m \ell^{p/2-1}$, and $\mathbb{E} S_A^2 \leq m f_0^2$, by Corollary 1.8 in Nagaev (1979),

$$\mathbb{P}(|S_A| \geq \frac{u}{2}) \leq C_{4,p} \frac{m\ell^{p/2-1}}{u^p} f_0^p + 2\exp\Big(-\frac{C_{5,p} u^2}{m f_0^2}\Big). \quad (3.28)$$

In (3.27) if we choose sufficient large $C_{2,p}$, then by elementary calculations, the second term in the right hand side of (3.28) will be smaller than $m f_0^p/u^p$. Therefore (3.27) follows from (3.28) in view of $|T_m| \leq |S_A| + |S_B|$.

Lemma 2. *Assume that* (X_i) *defined by* (1.1) *satisfies* $\mathbb{E}(X_i) = 0$, $X_i \in \mathcal{L}^p$, $p > 1$, *and* $\Theta_{m,p} < \infty$. *Then* $D_i := \sum_{j=i}^{\infty} \mathcal{P}_i X_j \in \mathcal{L}^p$, $\sup_i \|D_i\|_p < \infty$, *and for* $M_n = \sum_{j=1}^{n} D_j$ *we have*

$$\|S_n - M_n\|_p^{p'} \le C_p \sum_{i=1}^{n} \Theta_{i,p}^{p'}, \ \textit{where } p' = \min(2, p). \tag{3.29}$$

Lemma 2 can be proved by using the argument in the proof of Theorem 1(ii) in Wu (2007), where the latter paper considers rates of martingale approximations for stationary processes. Details of the derivation of (3.29) are omitted since there are no essential additional difficulties involved.

Acknowledgements

The authors are grateful to three referees for many helpful comments. The work was supported in part by NSF grants DMS-0478704 and 0906073.

References

Berkes, I. and Philipp, W. (1979). Approximation theorems for independent and weakly dependent random vectors. *Ann. Probab.* **7**, 29-54.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroscedasticity. *J. Econometrics* **31**, 307-327.

Bougerol, P. and Picard, N. (1992). Stationarity of GARCH processes and of some nonnegative time series. *J. Econometrics* **52**, 115-127.

Bradley, R. C. (1983). Approximation theorems for strongly mixing random variables. *Michigan Math. J.* **30**, 69-81.

Cai, Z. (2007). Trending time-varying coefficient time series models with serially correlated errors. *J. Econometrics* **136**, 163-188.

Chow, Y. S. and Teicher, H. (1988). *Probability Theory.* 2nd edition. Springer, New York.

Csörgő, M. and Révész, P. (1981). *Strong Approximations in Probability and Statistics.* Academic Press, New York.

Dahlhaus, R. (1997). Fitting time series models to nonstationary processes. *Ann. Statist.* **25**, 1-37.

Dahlhaus, R. and Subba Rao, S. (2006). Statistical inference for time-varying ARCH processes. *Ann. Statist.* **34**, 1075-1114.

Dedecker, J. and Prieur, C. (2004). Coupling for τ-dependent sequences and applications. *J. Theoret. Probab.* **17**, 861-885.

Draghicescu, D., Guillas, S. and Wu, W. B. (2009). Quantile curve estimation and visualization for non-stationary time series. *J. Comput. Graph. Statist.* **18**, 1-20.

Eberlein, E. (1986). On strong invariance principles under dependence assumptions. *Ann. Probab.* **14**, 260-270.

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

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

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

Fan, J. and Zhang, W. (2008). Statistical methods with varying coefficient models. *Statistics and its Interface* **1**, 179-195.

Fryzlewicz, P., Sapatinas, T. and Subba Rao, S. (2008). Normalized least-squares estimation in time-varying ARCH models. *Ann. Statist.* **36**, 742-786.

Komlós, J., Major, P. and Tusnády, G. (1975). An approximation of partial sums of independent RV's and the sample DF. I. *Z. Wahrsch. Verw. Gebiete* **32**, 111-131.

Komlós, J., Major, P. and Tusnády, G. (1976). An approximation of partial sums of independent RV's and the sample DF. II. *Z. Wahrsch. Verw. Gebiete* **34**, 33-58.

Lin, Z. and Lu, C. (1996). *Limit Theory for Mixing Dependent Random Variables.* Science Press, New York.

Lindgren, G. (1980). Extreme values and crossings for the χ^2-process and other functions of multidimensional Gaussian processes, with reliability applications. *Adv. Appl. Probab.* **12**, 746-774.

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

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

Mallat, S., Papanicolaou, G. and Zhang, Z. (1998). Adaptive covariance estimation of locally stationary processes. *Ann. Statist.* **26**, 1-47.

Monrad, D. and Philipp, W. (1991). The problem of embedding vector-valued martingales in a Gaussian process. *Theory Probab. Appl.* **35**, 374-377.

Orbe, S., Ferreira, E. and Rodriguez-Poo, J. (2005). Nonparametric estimation of time varying parameters under shape restrictions. *J. Econometrics* **126**, 53-77.

Orbe, S., Ferreira, E. and Rodriguez-Poo, J. (2006). On the estimation and testing of time varying constraints in econometric models. *Statist. Sinica* **16**, 1313-1333.

Philipp, W. and Stout, W. (1975). Almost sure invariance principles for partial sums of weakly dependent random variables. *Mem. Amer. Math. Soc.* **2**.

Rio, E. (1995). The functional law of the iterated logarithm for stationary strongly mixing sequences. *Ann. Probab.* **23**, 1188-1203.

Rio, E. (2009). Moment inequalities for sums of dependent random variables under projective conditions. *J. Theoret. Probab.* **22**, 146-163.

Robinson, P. M. (1989). Nonparametric estimation of time-varying parameters. In *Statistical Analysis and Forecasting of Economic Structural Change.* (Edited by P. Hackl), 164-253. Springer, Berlin.

Robinson, P. M. (1991). Time-varying nonlinear regression. In *Economic Structure Change Analysis and Forecasting.* (Edited by Hackl, P., Westland, A. H.), 179-190. Springer, Berlin.

Sakhanenko, A. I. (1984). Rate of convergence in the invariance principle for variables with exponential moments that are not identically distributed. In *Limit Theorems for Sums of Random Variables*, Trudy Inst. Mat. **3**, 4-49.

Shao, Q. M. (1993). Almost sure invariance principles for mixing sequences of random variables. *Stochastic Process. Appl.* **48**, 319-334.

Shao, Q. M. (1995). Strong approximation theorems for independent random variables and their applications, *J. Multivariate Anal.* **52**, 107-130.

Shao, X. and Wu, W. B. (2007). Asymptotic spectral theory for nonlinear time series. *Ann. Statist.* **35**, 1773-1801.

Volný, D. (1999). Invariance principles and Gaussian approximation for strictly stationary processes. *Trans. Amer. Math. Soc.* **351**, 3351-3371.

Wu, C. O., Chiang, C. T. and Hoover, D. R. (1998). Asymptotic confidence regions for kernel smoothing of a varying-coefficient model with longitudinal data. *J. Amer. Statist. Assoc.* **93**, 1388-1402.

Wu, W. B. (2005). Nonlinear system theory: Another look at dependence. *Proc. Natl. Acad. Sci. USA* **102**, 14150-14154.

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

Wu, W. B. and Shao, X. (2004). Limit theorems for iterated random functions. *J. Appl. Probab.* **41**, 425-436.

Wu, W. B. and Zhao, Z. (2007). Inference of trends in time series. *J. Roy. Statist. Soc. Ser. B* **69**, 391-410.

Zaitsev, A. (2001). Multidimensional version of a result of Sakhanenko in the invariance principle for vectors with finite exponential moments. I *Theory Probab. Appl.* **45**, 624-641.

Zaitsev, A. (2002a). Multidimensional version of a result of Sakhanenko in the invariance principle for vectors with finite exponential moments. II *Theory Probab. Appl.* **46**, 490-514.

Zaitsev, A. (2002b). Multidimensional version of a result of Sakhanenko in the invariance principle for vectors with finite exponential moments. III *Theory Probab. Appl.* **46**, 676-698.

Zhou, Z. and Wu, W. B. (2010). Simultaneous inference of linear models with time-varying coefficients. *J. Roy. Statist. Soc. Ser. B*, to appear.

Department of Statistics, The University of Chicago. 5734 S. University Ave., Chicago, IL 60637, USA.

E-mail: wbwu@galton.uchicago.edu

Department of Statistics, University of Toronto. 100 St. George St., Toronto, ON M5S3G3, Canada.

E-mail: zhou@utstat.toronto.edu

(Received August 2008; accepted November 2009)