Statistica Sinica Preprint No: SS-2017-0303	
Title	Optimal Gaussian Approximation For Multiple Time Series
Manuscript ID	SS-2017-0303
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202017.0303
Complete List of Authors	Sayar Karmakar and Wei Biao Wu
Corresponding Author	Sayar Karmakar
E-mail	sayarkarmakar@uchicago.edu
Notice: Accepted version subject to English editing.	

OPTIMAL GAUSSIAN APPROXIMATION FOR MULTIPLE TIME SERIES

By Sayar Karmakar[†,*] and Wei Biao Wu[*]

[†] *University of Florida and* [*] *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 based on the sample size n and the moment condition. Under the assumption of p-th finite moment, with $p > 2$, this 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

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

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

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

where $B(\cdot)$ is the standard Brownian motion and S_n' is constructed on a richer space such that $(S_i)_{i \le n} \overset{D}{=} (S_i')_{i \le n}$, 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, [32]; 2002a, [33]; 2002b, [34]) 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

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

where $^\top$ denotes matrix transpose, $\mathcal{F}_i = (\epsilon_i, \epsilon_{i-1}, \ldots)$ 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 $(\epsilon_i, \epsilon_{i-1}, \ldots)$ 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 [24, 26, 19] among others. When $d = 1$, Wiener [25] 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, [29]) and Zhou and Wu (2010, [35]) 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 $(S_i^c)_{1\le i\le n} \overset{D}{=} (S_i)_{1\le i\le n}$ 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 p-th 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, [27]) 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 [27] 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, [31]) obtained sub-optimal rates for a multiple non-stationary time series. A critical limitation of the result by [31, 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. Generally speaking, larger values of χ and A means the dependence decays faster. 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. 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. Applications to linear processes and locally stationary non-linear non-Lipschitz processes are given in section 3. The outline of the proof of Theorem 2.1 is sketched in section 4 while a more detailed version appears in online supplementary section 8. The goal of the sketched outline is to give the readers a basic idea of our long and involved derivation. Some useful results used throughout the proofs are presented in online supplementary section 9. These two supplementary sections are collated in the online supplementary material.

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. $x^+ = \max(x, 0)$ and $x^- = -\min(x, 0)$. 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 $a_n \le cb_n$ 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 $A = QDQ^\mathsf{T}$, where Q is orthonormal and $D = (\lambda_1, \ldots, \lambda_d)$ with $\lambda_1 \ge \ldots \ge \lambda_d$, write the Grammian square root $A^{1/2} = QD^{1/2}Q^\mathsf{T}$, $\rho_*(A) = \lambda_d$ 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 r-th 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;

(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 $t \geq 1, l \geq l_*$,

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

The uniform integrability assumption is necessary due to the non-stationarity of the process. The latter is frequently imposed in study of multiple time series.

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 5. 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 8.3, 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

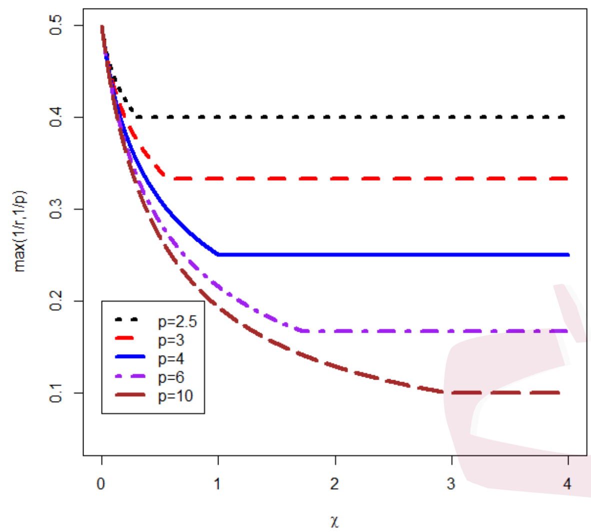


FIG 1. *Optimal bound as a function of* χ

3.1. *Vector linear processes:* Assume 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 $\epsilon_i = (\epsilon_{i1}, \dots, \epsilon_{id})^\mathsf{T}$, ϵ_i are i.i.d. random variables with mean 0 and finite q-th moment for some $q > 2$. Assume

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

where A satisfies (2.6) with p therein replaced by q. The model in (3.1) covers a large class of popularly used multiple time-series models such as vector AR, vector MA or vector ARMA etc. under mild conditions on the coefficient matrices. Specifically for a zero-mean vector ARMA process with lags a and b

$$X_i - \Psi_1 X_{i-1} - \ldots - \Psi_a X_{i-a} = \epsilon_i + \Phi_1 \epsilon_{i-1} + \ldots + \Phi_b \epsilon_{i-b}, \tag{3.3}$$

the stability condition (see [16] for definition) ensures a pure vector MA representation (3.1). The stationarity of the X_i process and finite q th moment ensures condition (2.A) with p therein replaced by q. Write $\Psi_* = I - \Psi_1 - \ldots - \Psi_a$, $\Phi_* = I + \Phi_1 + \ldots + \Phi_b$. Assume Ψ_*, Φ_* and $\Sigma_e = E(e_1 e_1^\top)$ are non-singular. Elementary calculation shows that as $l \to \infty$,

$$Var(S_l/\sqrt{l}) \to \Psi_*^{-1} \Phi_* \Sigma_e \Phi_*^\top \Psi_*^{-\top},$$

which is also non-singular. Thus condition (2.B) holds. Note that, $\|X_i - X_{i,(i-j)}\|_q = O(|B_j|)$ and thus condition (2.3) is satisfied for the X_i process from assumption (3.2). Thus, under suitable moment assumption we can apply Theorems 2.1 and 2.2 to generalize the central limit theory type results to a stronger invariance principle.

Next we discuss the covariance process for X_i that admits a representation as (3.1). Assume $q > 4$. Let the $d(d+1)/2$ dimensional 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 $(X_i)_{i=1}^n$. Write $p = q/2$. Fix two co-ordinates $1 \le r \le s \le d$. Then,

$$
\begin{aligned}
&\|X_{ir}X_{is} - X_{i,(i-j)r}X_{i,(i-j)s}\|_p \\
&\quad \le \|X_{ir}X_{is} - X_{ir}X_{i,(i-j)s}\|_p + \|X_{ir}X_{i,(i-j)s} - X_{i,(i-j)r}X_{i,(i-j)s}\|_p \\
&\quad \le \|X_{ir}\|_q\|X_{is} - X_{i,(i-j)s}\|_q + \|X_{ir} - X_{i,(i-j)r}\|_q\|X_{i,(i-j)s}\|_q \\
&\quad = O(|B_j|),
\end{aligned}
$$

since ϵ_i has finite q-th moment. Thus the condition (3.2) translates to the condition (2.3) for the W process with $p = q/2$. Condition (2.A) is trivially satisfied since the process W_i is stationary and has finite p-th moment. Let $\Sigma_W = \sum_{k=-\infty}^{\infty} Cov(W_0, W_k)$ be the long-run covariance matrix of (W_i). We assume the minimum eigenvalue of Σ_W is positive. This ensures that condition (2.B) holds. By Theorems 2.1 and 2.2, we have

$$
\max_{i \le n} |i\bar{W}_i - iE(W_1) - \Sigma_W^{1/2} IB(i)| = o_P(\tau_n), \tag{3.4}
$$

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 (3.4) is helpful for change point inference for multiple time series based on covariances; see [1, 23] 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

$$\sup_{0\le u\le 1} \sup_{x\ne x'} \frac{\|F(x, \epsilon_i, \theta(u)) - F_i(x', \epsilon_i, \theta(u))\|_p}{|x - x'|} < 1. \tag{3.5}$$

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

$$\delta_{i,p} = O(\beta^i). \tag{3.6}$$

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 analogue of ARCH, GARCH, AR, ARMA type models are prominent examples in this large class of non-stationary models. We discuss an example of threshold AR(1) model (see Tong (1990, [22])) with time-varying coefficients:

$$Y_i = \theta_1(i/n) Y_{i-1}^{+} + \theta_2(i/n) Y_{i-1}^{-} + e_i, \tag{3.7}$$

where e_i are i.i.d. mean-zero innovations. Assuming $\theta(\cdot) = (\theta_1(\cdot), \theta_2(\cdot))^{\mathsf{T}}$ is continuous, we can estimate $\theta(t)$ for $t \in [0,1]$ by

$$(\hat{\theta}_1(t), \hat{\theta}_2(t))^T = \arg\min_{\eta_1,\eta_2} \sum_{i=2}^{n} (Y_i - \eta_1 Y_{i-1}^{+} - \eta_2 Y_{i-1}^{-})^2 K\left(\frac{i/n - t}{b_n}\right), \tag{3.8}$$

where K is a symmetric kernel with bounded variation and compact support and b_n is appropriately chosen bandwidth. For such an estimation choice one has

$$\sqrt{nb_n}M(t)(\hat{\theta}(t)-\theta(t)) = \frac{1}{\sqrt{nb_n}}\sum_{i=2}^{n}\mathbf{v}_i\mathbf{v}_i^{\intercal}\left(\theta\left(\frac{i}{n}\right)-\theta(t)\right)K\left(\frac{i/n-t}{b_n}\right) + \frac{1}{\sqrt{nb_n}}\sum_{i=2}^{n}\mathbf{v}_i e_i K\left(\frac{i/n-t}{b_n}\right), \tag{3.9}$$

where $\mathbf{v}_i = (Y_{i-1}^{+}, Y_{i-1}^{-})^{\intercal}$ and $M(t) = (nb_n)^{-1}\sum_{i=2}^{n}\mathbf{v}_i\mathbf{v}_i^{\intercal}K((i/n-t)/b_n)$. Assuming some mild conditions on the innovation process e_i and the time-varying functions θ_1 and θ_2, one can construct simultaneous confidence interval for θ from (3.9). Assume for some $p > 2$, $\|e_1\|_p < \infty$, e_1 has a density with support $(-\infty, \infty)$ and

$$s = \sup_t(|\theta_1(t)| + |\theta_2(t)|) < 1. \tag{3.10}$$

We verify the conditions of Theorem 2.2 with the bivariate process $X_i = \mathbf{v}_i e_i$. Towards (2.A), it suffices to show uniform integrability for $(|Y_i|^p)_{i\geq 1}$ for the model (3.7). It easily follows since e_i is an i.i.d. innovation process with finite p-th moment and

$$|Y_i| \leq |e_i| + s|Y_{i-1}| \leq \sum_{j=0}^{\infty} s^j |e_{i-j}|.$$

Thus (2.A) holds. Because of independence of e_i and $x^+x^- = 0$,

$$Var(S_{t+l} - S_t) = \sum_{i=t+1}^{t+l} Var(\mathbf{v}_i e_i) = \sum_{i=t+1}^{t+l} \mathrm{diag}(E((Y_{i-1}^{+})^2)E(e_i^2), E((Y_{i-1}^{-})^2)E(e_i^2)).$$

With $D_i = \theta_1(i/n)Y_{i-1}^{+} + \theta_2(i/n)Y_{i-1}^{-}$ and $c_0 = 2\sup_i \|Y_i\|_2$,

$$\begin{aligned} E((Y_{i-1}^{+})^2) = E(((e_{i-1} + D_{i-2})^{+})^2) &\geq E(((e_{i-1} + D_{i-2})^{+})^2 I(|D_{i-2}| \leq c_0)) \\ &\geq E(((e_{i-1} - c_0)^{+})^2) P(|D_{i-2}| \leq c_0) \\ &> c_1(1 - 2\sup_i \|Y_i\|_2^2 / c_0^2), \end{aligned} \tag{3.11}$$

where c_1 is a constant that does not depend on i. We have a similar calculation for $E((Y_{i-1}^{-})^2)$ and thus (2.B) is satisfied. Under the assumption (3.10), since X_i satisfies the geometric moment contraction property (3.5), (2.3) holds for any $\chi > 0$.

For the second term in (3.9), we apply the Gaussian approximation from Theorem 2.2 with rate $\tau_n = n^{1/p}$. Using summation-by-parts, the neglibility criterion for the term with approximation rate requires,

$$n^{1/p}/\sqrt{nb_n} \to 0, \tag{3.12}$$

assuming bounded variation of K (cf. Zhao and Wu (2007,[30])). Now assume $\theta_1(\cdot)$ and $\theta_2(\cdot)$ are Hölder-α continuous for some $\alpha < 1/2$. For the negligibility of the first term in (3.9) we need $\sqrt{nb_n}b_n^{\alpha} \to 0$. This along with (3.12) and $\alpha < 1/2$ requires $p > 4$. This portrays one scenario among many that demands a sharper Gaussian approximation than $n^{1/4}$ and one such is obtained in Theorem 2.2. In the regime of curve estimation thus our result provides a strong tool by relaxing smoothness assumption on the coefficient curves/functions. This example shows how to overcome the unavailability of Taylor series expansion using minimal Hölder-continuity property and a sharper Gaussian approximation.

4. Key ideas of the proof of Theorem 2.1 The proof of Theorem 2.1 is quite involved. Here we discuss a brief outline of the major components of the proof. In particular our discussion will emphasize the difficulties that arise due to non-stationarity and vector-valued process and the techniques we use to circumvent those. Since these techniques allow us to solve this problem in such generality, we believe it might be of independent interest to the reader to at least have an overview of the major steps. The detailed proof is postponed to the online supplementary material.

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. This step differs from the treatment of [2] because of the choice of truncation level; we included the term t_n exploiting the uniform integrability assumption. This is necessary due to the non-stationarity. Secondly, we use the $m-$dependence approximation for a suitably chosen sequence m_n in terms of the decay rate χ. This generalizes the treatment in [2] as it additionally allows for processes where dependence decays slowly. Lastly, the blocking approximation requires some sharp Rosenthal-type inequality that needs γ-th 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. This step needs k-dic decomposition where k is possibly greater or equal to 3 to allow for non-stationarity.

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 (4.3) (Same as (8.9), (8.12) and (8.13) in the detailed version in online supplement A) 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. *Outline of preparation step:* The importance of the preparation steps is two-fold. It creates the platform for the conditional Gaussian approximation and regrouping by creating almost independent blocks. Moreover, these steps allow us to build a system of equation to solve for the approximation rate $\tau_n = n^{1/r}$ as a function of decay rate χ in (2.3). These equations are key in our generic approach to derive the optimal rate for slowly decaying dependence and show how it possibly affects (See Figure 1) the optimal Gaussian approximation rate.

Towards the truncating approximation we exploit the uniform integrability to introduce a sequence $t_n \to 0$ very slowly, such as

$$t_n \log\log n \to \infty \tag{4.1}$$

and use it at the truncation level $t_n n^{1/p}$. The truncation is defined through the operator

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

For the m-dependence approximation step and the blocking approximation, assume

$$m = \lfloor n^L t_n^k \rfloor, \quad 0 < k < (\gamma - p)/(\gamma/2 - 1), \quad 0 < L < 1, \tag{4.2}$$

$$n^{1/2-1/r}\Theta_{m,r} \to 0, \quad n^{1-\gamma/r} m^{\gamma/2-1} \to 0 \quad \text{and} \quad n^{1/p-1/\gamma} \sum_{j=m+1}^{\infty} \delta_{j,p}^{p/\gamma} \to 0, \tag{4.3}$$

where in (4.3) the first one is required for the m-dependence step and the other two are for the blocking approximation. At the end of these approximations we have a partial sum process $S_n^\diamond$ with the following summarized definition

$$\begin{aligned} S_i^\diamond &= \sum_{j=1}^{q_i} A_j \quad \text{with} \quad A_j = \sum_{i=(2jk_0-2k_0)m+1}^{2k_0jm} \tilde{X}_i, \\ \text{where } \tilde{X}_j &= E(T_{t_n n^{1/p}}(X_j)|\epsilon_j, \ldots, \epsilon_{j-m}) - E(T_{t_n n^{1/p}}(X_j)), \end{aligned}$$

and $k_0 = \lfloor \Theta_{0,2}^2/\lambda_* \rfloor + 2, q_i = \lfloor i/(2k_0 m) \rfloor$. For this truncated, m-dependent and blocked process $S_n^\diamond$ we have the approximation

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

See details of this subsection in online supplement section 8.1. Next, in subsection 4.2 and 4.3 we discuss how to obtain a Gaussian approximation for $S_n^\diamond$.

4.2. *Outline of conditional Gaussian approximation:* The blocks created in the preparation 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 8.3) for $d \geq 2$. 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. (8.11)), where λ_* is mentioned in Condition 2.B.

4.3. *Outline of regrouping and unconditional Gaussian approximation:* In the last part of our proof, we obtain the Gaussian approximation for the unconditional process by applying Proposition 8.3 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 (8.20) of Subsection 8.2) 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 8.22 in Subsection 8.2). 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 novel 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 the conditions in (4.3) (Same as equations (8.9), (8.12) and (8.13) in the detailed version of the proof). Elementary calculations show that $r < p$ for $\chi < \chi_0$. Provided $1 - (\chi + 1)p/\gamma < 0$, we have

$$\begin{aligned}\sum_{j=m+1}^{\infty} \delta_{j,p}^{p/\gamma} &\leq \sum_{i=\lfloor \log_2 m \rfloor}^{\infty} \sum_{j=2^i}^{2^{i+1}-1} \delta_{j,p}^{p/\gamma} \leq \sum_{i=\lfloor \log_2 m \rfloor}^{\infty} 2^{i(1-p/\gamma)} \Theta_{2^i,p}^{p/\gamma} \\ &= \sum_{i=\lfloor \log_2 m \rfloor}^{\infty} 2^{i(1-p/\gamma)} O(2^{-\chi i p/\gamma} i^{-Ap/\gamma}) = O(m^{1-p/\gamma-\chi p/\gamma} (\log m)^{-Ap/\gamma}).\end{aligned} \tag{4.4}$$

By (4.1) and (8.15) $\log m \asymp \log n$. Assume that,

$$1/2 - 1/r - \chi L = 0, \quad A > \gamma/p, \tag{4.5}$$

$$1 - \gamma/r + L(\gamma/2 - 1) = 0, \quad 0 < k < (\gamma/2 - 1)^{-1}(\gamma - p) \tag{4.6}$$

$$1/p - 1/\gamma + (1 - (\chi + 1)p/\gamma)L = 0. \tag{4.7}$$

Then conditions in (4.3) hold. Solving equations in (4.5), (4.6) and (4.7), we obtain r given in (2.7),

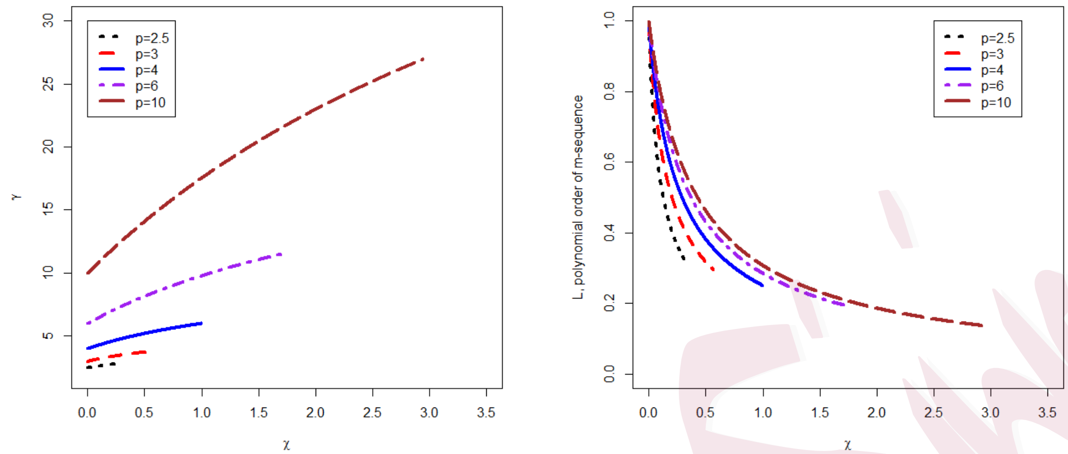
$$\begin{aligned}\gamma &= \frac{(2p+p^2)\chi + p^2 + 3p + 2 + f_5^{1/2}}{2+2p+4\chi},\\ L &= \frac{f_1 - f_2 + \chi\sqrt{(p-2)(f_3-3p)}}{\chi f_4},\end{aligned}$$

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 (8.40).

Remark 4.1. *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.*

5. 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 (8.24) to handle the $(\log n)$ term in choice of l in (8.26). While using the Nagaev inequality to show (8.43), we use a power $\gamma' > \gamma$ while keeping the choice of l (cf 8.26) same as before. We form



(a) (b)

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

a set of new equations

$$\begin{aligned} 1/2 + 1/p - 2/r' + L'(1 - (\chi + 1)p/r') &= 0, \qquad (5.1) \\ 1/p - 1/\gamma' + L' - L'(\chi + 1)p/\gamma' &= 0, \\ 1 - \gamma'/r' + L'(\gamma'/2 - 1) &= 0. \end{aligned}$$

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 9.2 to obtain a new equation corresponding to the m-dependence approximation using a power r' that is little higher than p. The solution of (5.1) has the property

$$\gamma' < 2(1+p+p\chi)/3. \tag{5.2}$$

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 (8.15). We apply the following version of Nagaev-type inequality from Liu, Xiao and Wu (2013, [15]) to obtain

$$\begin{aligned} P(|\tilde{S}_m| \geq \sqrt{lm}) &\lesssim \frac{m}{(lm)^{\gamma'/2}}\nu_R^{\gamma'+1} + \sum_{r=1}^{R} \exp\left(-c_{\gamma'}\frac{\lambda_r^2 l}{\tilde{\theta}_{r,2}^2}\right) + \frac{m^{\gamma'/2}\tilde{\Theta}_{m+1,\gamma'}^{\gamma'}}{(lm)^{\gamma'/2}} \\ &+ \frac{m \sup_i \|T_{t_n n^{1/p}}(X_i)\|_{\gamma'}^{\gamma'}}{(lm)^{\gamma'/2}} + \exp\left(-\frac{c_{\gamma'} l}{\sup_i \|T_{t_n n^{1/p}}(X_i)\|_2^2}\right), \end{aligned} \tag{5.3}$$

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 (5.2), (8.4) 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 (9.4). Combining these in the view of our new set of equations in (5.1), we get $P(|\tilde{S}_m| \geq \sqrt{lm}) = o(m/n)$ which is sufficient to conclude the proof as proposed in (8.43).

The positive-definitization technique introduced in (8.31) is validated in Proposition 8.9. 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 (8.26)) to use the conclusion of Proposition 8.3 by l' without the logarithm term $(\log n)$ in the denominator and without the condition (8.25). 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 8.3 again. The sufficient conditions in that result lead to an unavoidable $(\log n)$ term in choice of l (see 8.26). 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 (8.45), 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 γ-th moment in (5.3). 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 (4.5), (4.6) and (4.7). A careful check reveals that the rest of the proof goes through with this modified m sequence. $\square$

6. Supplementary Material The online supplementary material COMPLETED BY THE TYPESETTER.pdf contains the detailed proofs of Theorem 2.1 and some useful lemmas. The long detailed steps are in section 8 and the lemmas are postponed to section 9.

7. Acknowledgements We are grateful to the Associate Editor and one anonymous referee for their important feedbacks and comments about an earlier version of the paper. These helped us improve the quality and organization of the paper significantly. This research study is partially supported by NSF/DMS 1405410.

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] Tong, H. (1990). *Nonlinear time series. Oxford Statistical Science Series* **6**. The Clarendon Press, Oxford University Press, New York A dynamical system approach, With an appendix

by K. S. Chan, Oxford Science Publications. MR1079320

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

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

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

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

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

[28] 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

[29] Wu, W. B. and Zhao, Z. (2007a). Inference of trends in time series. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **69** 391–410. MR2323759

[30] Wu, W. B. and Zhao, Z. (2007b). Inference of trends in time series. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **69** 391–410. MR2323759

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

[32] 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

[33] 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

[34] 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

[35] 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