The Annals of Statistics
2024, Vol. 52, No. 5, 2293–2317
https://doi.org/10.1214/24-AOS2436

GAUSSIAN APPROXIMATION FOR NONSTATIONARY TIME SERIES WITH OPTIMAL RATE AND EXPLICIT CONSTRUCTION

BY SOHAM BONNERJEE[1,a], SAYAR KARMAKAR[2,c] AND WEI BIAO WU[1,b]

[1]*Department of Statistics, University of Chicago,* [a]*sohambonnerjee@uchicago.edu,* [b]*wbwu@galton.uchicago.edu*
[2]*Department of Statistics, University of Florida,* [c]*sayarkarmakar@ufl.edu*

Statistical inference for time series such as curve estimation for time-varying models or testing for existence of a change point have garnered significant attention. However, these works are generally restricted to the assumption of independence and/or stationarity at its best. The main obstacle is that the existing Gaussian approximation results for nonstationary processes only provide an existential proof, and thus they are difficult to apply. In this paper, we provide two clear paths to construct such a Gaussian approximation for nonstationary series. While the first one is theoretically more natural, the second one is practically implementable. Our Gaussian approximation results are applicable for a very large class of nonstationary time series, obtain optimal rates and yet have good applicability. Building on such approximations, we also show theoretical results for change-point detection and simultaneous inference in presence of nonstationary errors. Finally, we substantiate our theoretical results with simulation studies and real data analysis.

1. Introduction. Statistical inference for time series is an important topic that has garnered significant attention over the past several decades. There is a well-developed asymptotic theory of Gaussian approximation for stationary processes that in turn yields a solid foundation for doing asymptotic inference. However, in practice, nonstationary time-series processes are more ubiquitous, and unfortunately, similar Gaussian approximation tools for nonstationary processes are either not sharp enough or difficult to apply. Our main goal in this paper is to establish optimal KMT-type Gaussian approximations for nonstationary time series that also provide an explicit construction strategy, and thus enable asymptotic inference for such series.

We now discuss some motivations for theoretical development for nonstationary time series. Stationarity is an idealized assumption for any real-life series observed over a long period of time. In the parlance of analyzing such long series, when parametric models are used, typically this translates to systematic deviation of the parameters. Even without such a parametric guide, one can observe intrinsic changes in how the dependence evolves over time. Apart from these, different external factors such as recession, war, politics, pandemic, etc. affect time series and can introduce abrupt paradigm shifts. Such shifts could be of different types—either a shift in mean, or shock events that change a process that was varying slowly or in a more stationary way. The two approaches are captured in the literature of time-varying models and change-point analyses, respectively.

The literature of time-varying models tries to address this issue by allowing model parameters to vary smoothly over time; see [14, 32, 33, 48, 49, 61, 84, 105] among others. The inference questions arise naturally while choosing a time-varying model in contrast of a time-constant one. Such hypothesis testing frameworks are discussed in [4, 12, 18, 59, 69, 73, 79, 103, 104] and [60]. Moving from pointwise inference, [53, 95, 108]

Received February 2024; revised May 2024.
MSC2020 subject classifications. 62M10, 62M15, 62E17.
Key words and phrases. Gaussian approximation, nonstationary time series, simultaneous confidence band, change-point testing.

discussed obtaining more challenging simultaneous confidence bands. Such simultaneous inference requires Gaussian approximation beyond the central limit theorem, and motivates for KMT-type Gaussian approximations as spelled out in (1.1). The second approach—the analysis of change points, originated in quality control literature [74, 75], but has since become an integral part of a wide variety of fields, among them economics [78], finance [2], climatology [85] and engineering [90]. Building on estimation techniques, these problems discuss different types of inference problems such as the existence of change point or creating confidence bands for means of different pieces, etc. The test statistic for testing existence of change points may be viewed as two-sample tests adjusted for the unknown break location, thus leading to max-type procedures. Such tests also need a Gaussian approximation as mentioned in (1.1) to provide correct cut-off. For some useful references on these, see [6] and [19] among others. Structural break estimation can also be viewed as a model selection problem; see [23, 65] and [87]. See also [5] and [50] for excellent reviews on change-point inference literature.

However, in both of these paradigms, typically the error process is assumed to be stationary, and thus the techniques involved do not go beyond what we already know for stationary series. In other words, the nonstationarity has generally been reflected only in the signal and not in the noise process. This posits a challenging but a fundamental problem. The literature on inference for nonstationary time series is sparse due to difficulty of obtaining a sharp, explicit Gaussian approximation. The existing results are either not as sharp as those for stationary processes, or are difficult to construct.

We now proceed to mathematically introduce the problem. For independent and identically distributed X_i with $\mathbb{E}(X_i) = 0, \mathbb{E}(|X_i|^p) < \infty$, $p > 2$, Komlós, Major and Tusnády [55, 56] obtained an optimal Gaussian approximation: for $S_i := \sum_{j=1}^{i} X_j$,

$$\max_{1\le j\le n} \left|S'_j - \mathbb{B}\left(\mathbb{E}\left(S_j^2\right)\right)\right| = o_{\text{a.s.}}(\tau_n), \tag{1.1}$$

where $\mathbb{E}(S_j^2) = j\mathbb{E}(X_1^2)$, $\mathbb{B}(\cdot)$ is the standard Brownian motion and S'_n is constructed on a richer space such that $(S_i)_{i\ge1} =_{\mathbb{D}} (S'_i)_{i\ge1}$, and the approximation rate $\tau_n = n^{1/p}$ is optimal when only finite pth moment is assumed. Henceforth, throughout this paper, we will assume $p > 2$ unless specified explicitly. The Gaussian approximation (1.1) substantially generalizes the central limit theorem $S_n/\sqrt{n} \Rightarrow N(0, \mathbb{E}(X_1^2))$, and it allows for a systematic study of statistical properties of estimators based on independent data. The optimal rate of $n^{1/p}$ was matched for a large class of stationary time series in the seminal work by Berkes, Liu and Wu [9]. In the latter work, they assume the stationary causal representation for X_i, and are able to replace $\mathbb{E}(S_j^2) = j\mathbb{E}(X_1^2)$ in (1.1) by $j\sigma_\infty^2$ where $\sigma_\infty^2 = \sum_{i\in\mathbb{Z}} \mathbb{E}(X_0X_i)$ is the long-run variance of the time series. One can see that $\sigma_\infty^2 = \lim_{n\to\infty} \mathbb{E}(S_n^2)/n$, and thus S_i being approximated by a Gaussian process with variance $i\sigma_\infty^2$ makes intuitive sense from the idea of preserving a second-order property. Unfortunately, for a nonstationary process, one does not have the notion of such a long-run variance, and thus the existing Gaussian approximation results are somewhat abstract and unclear.

To characterize the nonstationary process (X_t), we view X_t as outputs from a physical system with the following causal representation:

$$X_t = g_t(\mathcal{F}_t), \quad \text{with } \mathcal{F}_t = (\ldots, \varepsilon_{t-1}, \varepsilon_t), \tag{1.2}$$

where $(\varepsilon_i)_{i\in\mathbb{Z}}$ are i.i.d. inputs of this system and $g_t : \mathbb{R}^\infty \to \mathbb{R}$ are measurable functions. A Gaussian approximation for such nonstationary processes was obtained by [96], with a suboptimal rate and only for $2 < p \le 4$. On the other hand, for inferential procedures it is important to establish an approximation for the process $(S_i)_{i=1}^n$. They did provide a regularization $G_j = \sum_{i=1}^{j} \Sigma_i^{1/2} Z_i$, where $\Sigma_i = \mathrm{Var}(\sum_{k=i}^{\infty}(\mathbb{E}(X_k|\mathcal{F}_i) - \mathbb{E}(X_k|\mathcal{F}_{i-1})))$ and Z_i are i.i.d. Gaussian; however, Σ_i's are not naturally estimable quantities. This result was improved upon by [54], who obtained optimal rate $n^{1/p}$ rate for all $p > 2$. However, even their approximating Gaussian process is not regularized as it only provides approximation for blocks of partial sums, and not all S_j as (1.1) does. Moreover, the variance of the approximating Gaussian process was difficult to interpret and connect with that of the original process. Recently, [68] used a local long-run covariance matrix as a proxy to the variance of the approximating Brownian motion. Their proof relies on martingale embedding strategy of [28] to bound Wasserstein distance of the partial sums and their Gaussian analogues. Nonetheless, their rate is suboptimal.

Keeping the main goal of regularizing the approximating Gaussian process, we note that it is possible to preserve the second-order property without the notion of long-run variance if the approximating (of S_j) Gaussian process can be written as $G_i = \sum_{j \le i} Y_i$ with $\mathbb{E}(S_i^2) = \mathbb{E}(G_i^2)$. We start with one such approximation, which ensures this; in fact, we are able to establish a Gaussian approximation that ensures $\mathrm{Cov}(X_i, X_j) = \mathrm{Cov}(Y_i, Y_j)$, which entails $\mathbb{E}(S_i^2) = \mathbb{E}(G_i^2)$. Assumption of Gaussianity is frequently used in many areas of statistics where, as further specification, one puts a covariance structure on (X_i). Our Gaussian approximation provides theoretical validation that for the nonstationary process, one can still obtain an approximating Gaussian process that matches the covariance at a modular level. To the best of our knowledge, such covariance-matching Gaussian approximations, despite being quite natural for nonstationary processes, are rarely discussed in the literature. In particular, for a possible nonstationarity in covariance, such second-order preserving approximation seems to be a first such result that additionally maintains optimal rate.

Our first result is applicable in situations where the practitioner knows the covariance structure of the observed processes. However, for general nonstationary processes with unknown covariance structure, the practical implementation with this novel Gaussian approximation remains somewhat challenging. Our second set of Gaussian approximation results first embed the approximating Gaussian process in a Brownian motion with evolving variance and then regularize the latter. As expected, the variance generally does not increase linearly as it does in [9] for the stationary case. However, in our approximation S_j is approximated by a Brownian motion valued at $\mathbb{E}(S_j^2)$, which is same as (1.1). Unlike [68], the variance of our approximating Gaussian process is simply $\mathbb{E}(S_i^2)$, which immediately suggests intuitive estimators of that variance.

Next, we address the issue of estimating the variance of the approximating Gaussian processes. We first derive a block version of our theoretical Gaussian approximation, which in turn, yields a conditional Gaussian approximation where estimated block variances are used to construct the variances of the approximating theoretical Gaussian process. We are able to achieve $n^{1/4+\varepsilon}$ rate here, which is nearly optimal when variances are to be estimated. This also means that to achieve such results, assumptions on only slightly higher than 4th moments suffice. Here, we also reflect on an alternative estimation procedure, and show that our "Block-based Running Variance (BRV)" estimate gives better rates for all $p > 2$. Finally, we apply our results to three prominent inference problems, namely the inference problem related to existence of change point, the simultaneous confidence bands for nonstationary time series and asymptotic distribution of the wavelet coefficient process. As mentioned above already, stationarity and/or Gaussianity were standard assumptions in all these literature throughout and this paper erases this barrier and establishes theoretical guarantees for a much larger class of time series.

Our main contributions are summarized below.

- We obtain the sharp KMT-type Gaussian approximations of the order $n^{1/p}$ for nonstationary time series with minimal conditions. In particular,

- in our first result, we observe a novel Gaussian approximation, which matches the covariance structure. Despite being intuitively very natural for nonstationary processes, ours is probably the first such approximation result in the literature.
- We also explore a second type of Gaussian approximation, which involves embedding a Brownian motion much like [9] or [54]. Crucially, we recover the sharp $n^{1/p}$ rate modulo a logarithmic factor without the lower bound assumption of block variance needed in [54].

- We discuss estimation of the running variance of the approximating Brownian motion and show consistency of such estimators using uniform deviation inequalities. Such maximal deviation bounds for quadratic forms based on nonstationary processes may be of independent interest.
- Finally, we show applications of such Gaussian approximation through the lens of three prominent inference problems, namely the inference problem related to change point, the simultaneous confidence bands for nonstationary time series and asymptotic distributions of wavelet coefficient processes. As mentioned above already, stationarity and/or Gaussianity were standard assumptions in all these literature throughout and this paper overcomes these limitations to arrive at much more general results.
- We also provide some simulations to corroborate our Gaussian approximations and an analysis of an interesting data set that highlights our applications.

1.1. *Organization of the paper.* The rest of the paper is organized as follows. In Section 2.2, we discuss a functional dependence measure that allows us to encode dependence in a mathematically tractable way for a large class of nonstationary time series. We also discuss other general assumptions there. Sections 2.3 and 2.4 discuss the two Gaussian approximations, which are the main theoretical contributions of our paper. Next, Section 3 is used to describe the block-bootstrap Gaussian approximation and related results, featuring a result on a novel deviation inequality for nonstationary quadratic forms. We discuss three important inference problems in Section 4. The hypothesis testing related to test existence of change point is discussed in Section 4.1. Subsequently, we discuss simultaneous confidence bands for nonstationary time series, which is deferred to Section 4.2. Finally, the discussion on wavelet coefficient process is deferred until Section 4.3. Next, we use Section 5 to demonstrate through simulations that we achieve better approximations with the regularization spelled out in theoretical results than the prototypical block-sum variance. We also show extensive simulation results for the first two of the above mentioned applications. For space constraint, some of these simulations are deferred to Appendix Section 12 [109]. Finally, we show advantage of our theory and estimates by analyzing a recent archaeological data set in Section 6. All the proofs are postponed to Appendix Sections 8, 9, 10 and 11.

1.2. *Notation.* For a random variable 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 $\|\cdot\| = \|\cdot\|_2$. Throughout the text, we use C for constants that might take different values in different lines unless otherwise specified. For two positive sequences a_n and b_n, if $a_n/b_n \to 0$, write $a_n = o(b_n)$. Write $a_n \lesssim b_n$ or $a_n = O(b_n)$ if $a_n \le Cb_n$ for all sufficiently large n and some constant $C < \infty$. Similarly, for a sequence of random variables $(X_n)_{n\ge1}$ and a positive sequence y_n, if $X_n/y_n \to 0$ in probability, we write $X_n = o_{\mathbb{P}}(y_n)$, and if X_n/y_n is stochastically bounded, we write $X_n = O_{\mathbb{P}}(y_n)$.

2. Gaussian approximation results. Before we proceed to discuss the Gaussian approximation results for a general class of nonstationary time series, we first provide a concise introduction of similar results for independent random variables. Note that in principle such Gaussian approximations for random variables $(X_i)_{i=1}^n$ require a common, possibly enriched probability space $(\Phi_c, \mathcal{A}_c, \mathbb{P}_c)$ on which the approximating Gaussian processes and random variables $(X_i^c)_{1\le i\le n} =_{\mathbb{D}} (X_i)_{1\le i\le n}$ can be defined. In order for better readability, we omit this technicality and simply state our results in terms of the original random variables X_i's.

2.1. *Gaussian approximation for independent random variables.* For i.i.d. random variables, the mentioned result (1.1) by [55, 56] represented the culmination of a series of results on *strong invariance principle* starting from [29] and [27]. Subsequently, the seminal paper by Sakhanenko [89] essentially generalized the KMT-type Gaussian approximation for independent but possibly not identically distributed random variables. The following theorem follows easily from [89].

THEOREM 2.1. *Let $(X_i)_{1\le i\le n}$ be independent but possibly not identically distributed random variables with $\mathbb{E}(X_i) = 0$ and for a $p > 2$, $\max_{1\le i\le n} \|X_i\|_p = O(1)$, and there exists $\gamma \ge 2$ such that*

$$\sum_{i=1}^{n} \mathbb{E}\bigl[\min\bigl\{|X_i|^\gamma/n^{\gamma/p}, |X_i|^2/n^{2/p}\bigr\}\bigr] = o(1). \tag{2.1}$$

Then there exists a Brownian motion $\mathbb{B}(\cdot)$ such that the following holds:

$$\max_{1\le i\le n} \bigl|S_i - \mathbb{B}\bigl(\mathbb{E}(S_i^2)\bigr)\bigr| = o_{\mathbb{P}}\bigl(n^{1/p}\bigr). \tag{2.2}$$

The readers can look into [99, 100] and [101] for a review of similar approximations for independent but possibly nonidentically distributed random variables. For time series, [9] represents the optimal result for stationary processes in this direction, while [54] shows an optimal existential result for nonstationary multivariate processes. However, [54] does not provide any result about the covariance structure of the approximating Gaussian processes, apart from them having independent increments. However, in the search for an explicit covariance regularization of the Gaussian approximations, it is natural to conjecture that the approximating Gaussian processes have the same second-order structure as that of the original nonstationary process X_t. To deal with such results, we need to characterize the dependency setup of the wide class of the nonstationary processes that we consider in (1.2). This structural premise is laid out in the next section.

2.2. *Functional dependence measure for nonstationary processes.* To deal with the dependency structure of a nonstationary process, we employ the framework of functional dependence measure [94]. We will work with (1.2), which is quite general and arises naturally from writing the joint distribution of $(X_1, \dots, X_n)$ in terms of compositions of conditional quantile functions of i.i.d. uniform random variables. With this system, given $k \ge 0$, a time lag, we measure the dependence from how much the outputs X_i of this system will change if we replace the input information at time $i - k$ with an i.i.d. copy ε'_{i-k}. For $p \ge 1$, define the uniform functional dependence

$$\begin{aligned} &\delta_p(k) := \sup_i \bigl(\mathbb{E}|X_i - X_{i,\{i-k\}}|^p\bigr)^{1/p} \\ &\qquad \text{where } X_{i,\{i-k\}} = g_i\bigl(\dots, \varepsilon_{i-k-1}, \varepsilon'_{i-k}, \varepsilon_{i-k+1}, \dots, \varepsilon_i\bigr) \end{aligned} \tag{2.3}$$

is a coupled version of X_i. We will assume $\mathbb{E}(X_i) = 0$. Note that $(\mathbb{E}|X_i - X_{i,\{i-k\}}|^p)^{1/p}$ encapsulates the dependence of X_i in ε_{i-k}. Since X_i is a nonstationary process, the physical mechanism process g_i is allowed to be different for every i. Thus, we have defined the functional dependence measure in a uniform manner, by taking supremum over all i. This measure (2.3) is directly related to the data-generating mechanism, and we will express our dependence condition in terms of

$$\Theta_{i,p} = \sum_{k=i}^{\infty} \delta_p(k), \quad i \geq 0. \tag{2.4}$$

Observe that $\sup_i \|X_i\|_p \leq \Theta_{0,p}$. With this framework, we are able to conveniently propose conditions on temporal dependence for the nonstationary time series models we will use.

2.3. *Gaussian approximation maintaining covariance structure.* As discussed in Section 2.2, to state our Gaussian approximation result, we need to properly control the temporal decay by putting mild assumptions on $\Theta_{i,p}$. In particular, we will need that $\Theta_{i,p}$ decays with a polynomial rate.

CONDITION 2.1. *Consider* (1.2). *Suppose that $\Theta_{0,p} < \infty$ for some $p > 2$. Assume there exists $A > 1$ and constant $C > 0$ such that the uniform dependency-adjusted norm*

$$\mu_{p,A} := \sup_{i \geq 0} (i+1)^A \Theta_{i,p} \leq C < \infty. \tag{2.5}$$

Condition 2.1 is satisfied by a large class of processes. Some examples are mentioned in Section 2.5. The assumption $\Theta_{0,p} < \infty$ can be interpreted as the cumulative dependence of $(X_i)_{i \geq k}$ on ε_k being finite. If it fails, the process can be long-range dependent, and in such cases the Brownian motion approximations of the partial sum processes may fail. Since the process $(X_i)_i$ is nonstationary, in order to better control its distributional behavior, we need a uniform integrability condition.

CONDITION 2.2. *For the same p as in Condition 2.1, the series $(|X_i|^p)$ satisfies the truncated uniform integrability condition*:

$$\text{For any fixed } a > 0,\ \sup_i \mathbb{E}\big(|X_i|^p \mathbb{I}_{\{|X_i|^p \geq an\}}\big) \to 0 \text{ as } n \to \infty.$$

The classical uniform integrability condition for $(|X_t|^p)_t$ is $\sup_i \mathbb{E}(|X_i|^p \mathbb{I}_{\{|X_i|^p \geq k\}}) \to 0$ as $k \to \infty$. Note that Condition 2.2 is weaker. To avoid degeneracy, we will also require a mild nonsingularity condition on the block variance of the original process (X_t).

CONDITION 2.3. *For all sequences $(m_n) \in \mathbb{N}$ with $m_n \to \infty$ and $m_n < n$, the process (X_i) satisfies that* $\lim_{n\to\infty} \min_{1 \leq i \leq n-m_n} \|X_i + \cdots + X_{i+m_n}\|^2 = \infty$.

This nonsingularity condition is a very natural one. A simple counterexample may be given for the case where absence of such assumption entails failure of even the central limit theorem. For $t \in \mathbb{N}$, consider the process $X_t = \varepsilon_t - \varepsilon_{t-1}$, and ε_i are i.i.d. non-Gaussian with mean 0 and variance $\sigma^2 > 0$. Then for $n \in \mathbb{N}$, clearly $S_i = \varepsilon_i - \varepsilon_0$ for $1 \leq i \leq n$, and thus both Condition 2.3 and central limit theorem $S_n/\|S_n\| \Rightarrow N(0,1)$ fails to hold. With this condition, we begin by presenting a Gaussian approximation for the truncated partial sum process

$$S_i^{\oplus} := \sum_{j=1}^{i} \big(X_j^{\oplus} - \mathbb{E}(X_j^{\oplus})\big) \quad \text{where } X_i^{\oplus} = T_{n^{1/p}}(X_i),\ i = 1, \dots, n, \tag{2.6}$$

with $T_b(w) = \max\{\min\{w, b\}, -b\}$. The following is the first main result of this paper.

THEOREM 2.2. *Let $p > 2$. For the process $(X_t)_t$, assume Conditions 2.2, 2.3 and 2.1 with*

$$A > A_0 := \max\left\{\frac{p^2 - p - 2 + (p-2)\sqrt{p^2 + 10p + 1}}{4p}, 1\right\}. \tag{2.7}$$

Then there exists a Gaussian process Y_t with $\mathrm{Cov}(X_s, X_t) = \mathrm{Cov}(Y_s, Y_t)$ such that

$$\max_{1 \le i \le n}\left|S_i - \sum_{j=1}^{i} Y_j\right| = o_{\mathbb{P}}\big(n^{1/p}\sqrt{\log n}\big). \tag{2.8}$$

In fact, there also exists a Gaussian process $Y_t^{\oplus}$, with $\mathrm{Cov}(Y_s^{\oplus}, Y_t^{\oplus}) = \mathrm{Cov}(X_s^{\oplus}, X_t^{\oplus})$ such that

$$\max_{1 \le i \le n}\left|S_i - \sum_{j=1}^{i} Y_j^{\oplus}\right| = o_{\mathbb{P}}\big(n^{1/p}\big). \tag{2.9}$$

Here, it is important to note that, although (2.9) has a better rate than (2.8), the approximating process has covariance structure matched with the truncated value of the original process X_i. However, we still present this result since it shows that theoretically it is possible to achieve the optimal $n^{1/p}$ rate without the stronger nonsingularity condition as [54]. Proving such a result also necessitates novel techniques, which are different compared to both [54] and [9].

Finally, if one were to assume the nonsingularity condition as written below, we show that it is possible to achieve $n^{1/p}$ rate even with the approximating process matching covariances exactly with the original (X_t) process.

CONDITION 2.4. *The series (X_i) satisfies the following condition: There exists a constant $c > 0$ and $l_0 \in \mathbb{N}$ such that for all $l \ge l_0$, $\min_{1 \le j \le n-l+1} \|X_j + \cdots + X_{j+l-1}\|^2 / l \ge c$.*

At the cost of making this extra assumption, we are also able to improve the decay rate condition on $\Theta_{i,p}$ from that in Theorem 2.2, matching exactly the optimal cut-off given in [54].

THEOREM 2.3. *Assume the process $(X_t)_{t \ge 1}$ satisfies Conditions 2.2, 2.4 and 2.1 with*

$$A > A_0' := \max\left\{\frac{p^2 - 4 + (p-2)\sqrt{p^2 + 20p + 4}}{8p}, 1\right\}. \tag{2.10}$$

Then there exists a Gaussian process (Y_t) with $\mathrm{Cov}(Y_s, Y_t) := \mathrm{Cov}(X_s, X_t)$ such that

$$\max_{1 \le i \le n}\left|S_i - \sum_{j=1}^{i} Y_j\right| = o_{\mathbb{P}}\big(n^{1/p}\big). \tag{2.11}$$

2.4. *Gaussian approximation with independent increments.* In addition to having a natural interpretation, the Gaussian approximations in the previous Section 2.3 also enjoy applicability when information about the covariance structure of the original process is available, such as for stationary processes [97] or processes from a defined parametric structure. However, for a general nonstationary processes, the precise correlation structure of X_t process may not be available and, therefore, simulating the Y_t process becomes a challenge. Therefore, it is important to investigate if we can further obtain a Gaussian approximation of the form (2.2), that is, involving Brownian motion with independent increments, where the involved $\mathbb{E}(S_i^2)$ is estimable. The following two theorems address these issues and yield Gaussian approximations with this desired structure. Our first result is analogous to Theorem 2.2. However, in this result, we no longer require any nonsingularity condition, and yet we almost recover the optimal $n^{1/p}$ rate (up to a log factor). Again, we recover the exact optimal rate if our Gaussian approximation involves the moments of the truncated process.

THEOREM 2.4. *For the process $(X_t)_{t\geq 1}$, assume Conditions 2.2 and 2.1 with $A > A_0$; see (2.7). Then there exists a Brownian motion $\mathbb{B}(\cdot)$ such that*

$$\max_{1\leq j\leq n} \left|S_j - \mathbb{B}\left(\mathbb{E}\left(S_j^{\oplus^2}\right)\right)\right| = \mathrm{o}_{\mathbb{P}}\left(n^{1/p}\right). \tag{2.12}$$

Further, it holds that

$$\max_{1\leq j\leq n} \left|S_j - \mathbb{B}\left(\mathbb{E}\left(S_j^2\right)\right)\right| = \mathrm{o}_{\mathbb{P}}\left(n^{1/p}\sqrt{\log n}\right). \tag{2.13}$$

A similar remark to the one following Theorem 2.2 is in order. Note that, in Theorem 2.4, again using the moments of the original process in the Gaussian approximation entails a penalty of $\sqrt{\log n}$ in our rate. However, it turns out that under the more stringent nonsingularity condition of Theorem 2.3, we are not only able to recover the optimal rate of $n^{1/p}$ from using the X_t process itself, but also able to relax the decay rate.

THEOREM 2.5. *Under conditions of Theorem 2.3, there exists a Brownian motion $\mathbb{B}(\cdot)$ such that*

$$\max_{1\leq j\leq n} \left|S_j - \mathbb{B}\left(\mathbb{E}\left(S_j^2\right)\right)\right| = \mathrm{o}_{\mathbb{P}}\left(n^{1/p}\right). \tag{2.14}$$

REMARK 2.1. Necessity of the truncated uniform integrability Condition 2.2: We show that the uniform integrability condition is necessary as otherwise the Gaussian approximation might fail. Let $n > 2$. Let $X_1, X_2, \ldots$ be independent with $\mathbb{P}(X_i = \pm(i+1)^{1/p}) = 1/(i+1)$ and $\mathbb{P}(X_i = \pm 1) = 1/2 - 1/(i+1)$. Note that Condition 2.2 is violated since $\max_{1\leq i\leq n} \mathbb{E}[|X_i|^p \mathbb{I}\{|X_i|^p > n/2\}] = 2$. For the sake of contradiction, suppose the Gaussian approximation (2.14) holds, which implies

$$\max_{1\leq i\leq n} \left|X_i - \left(\mathbb{B}\left(\mathbb{E}\left(S_i^2\right)\right) - \mathbb{B}\left(\mathbb{E}\left(S_{i-1}^2\right)\right)\right)\right| = o_{\mathbb{P}}\left(n^{1/p}\right). \tag{2.15}$$

Since X_i's are independent, and $\max_{1\leq i\leq n} \mathbb{E}(X_i^2) \leq 2^{2/p} + 1$ therefore by property of increments of Brownian motion, $\max_{1\leq i\leq n} |\mathbb{B}(\mathbb{E}(S_i^2)) - \mathbb{B}(\mathbb{E}(S_{i-1}^2))| = O_{\mathbb{P}}((\log n)^{1/2})$. Thus, if one assumes that (2.15) is true, then we will have $\max_{1\leq i\leq n} |X_i| = o_{\mathbb{P}}(n^{1/p})$. Now we show that the latter is false. Clearly, $|X_i| \leq n^{1/p}/2$ w.p. 1 if $i \leq n/2^p - 1$ and, therefore,

$$\begin{aligned}\mathbb{P}\left(\max_{1\leq i\leq n} |X_i| > \frac{n^{1/p}}{2}\right) = 1 - \prod_{i=\lceil n/2^p\rceil\vee 1}^{n} \mathbb{P}\left(|X_i| \leq \frac{n^{1/p}}{2}\right) \qquad &\geq 1 - \left(1 - \frac{2}{n+1}\right)^{n(1-\frac{1}{2^p-1})} \\ \to 1 - e^{2^{2-p}-2},&\end{aligned}$$

as $n \to \infty$. This contradiction shows that Theorem 2.5 fails to hold. This vouches for the necessity of our uniform integrability condition; clearly, the reason the Gaussian approximation fails to hold in this example is due to Condition 2.2 not being satisfied. It can be noted that, in this example, Theorem 2.1 does not apply; (2.1) can be verified to be violated in this case.

2.5. *Examples*. We now show some examples of nonstationary time series, which satisfy Condition 2.1. For $t \in \mathbb{Z}$, let $\mathcal{F}_t = (\dots, \varepsilon_{t-1}, \varepsilon_t)$, where ε_t are i.i.d. random variables. Consider the model

$$X_t = g(\theta_t, \mathcal{F}_t), \quad 1 \le t \le n, \tag{2.16}$$

where $\theta_t \in \Gamma$, a parameter space and $g(\cdot, \mathcal{F}_t) : \Gamma \to \mathbb{R}$ is a progressively measurable function such that the process $X_t(\theta) = g(\theta, \mathcal{F}_t)$ is well-defined. We can view (2.16) as a general modulated stationary process. Adak [1] and [106] considered the special case of multiplicative modulated stationary processes with a linear form. Define the functional dependence measures as

$$\begin{aligned} \delta_p^{\Gamma}(k) &:= \sup_{\theta \in \Gamma} \| g(\theta, \mathcal{F}_t) - g(\theta, \mathcal{F}_{t,\{t-k\}}) \|_p \\ &\ge \sup_t \| g(\theta_t, \mathcal{F}_t) - g(\theta_t, \mathcal{F}_{t,\{t-k\}}) \|_p =: \delta_p^X(k). \end{aligned} \tag{2.17}$$

Thus, we only need to assume that $\Theta_{i,p}^{\Gamma} := \sum_{k=i}^{\infty} \delta_p^{\Gamma}(k)$ satisfies Condition (2.1). We mention a couple of examples from the general class of nonstationary processes satisfied by (2.16).

2.5.1. *Cyclostationary process*. Taking $\theta_t = \phi_{t \bmod T}$ in (2.16) for some period T, and $\{\phi_t\}_{t=1}^T \in \Gamma$ yields cyclostationary process. These can be thought of as generalizations of stationary processes, incorporating periodicity in its properties, and were introduced as a model of communications systems in [7] and [35]. Apart from communication and signal detection, cyclostationary processes have enjoyed wide use in econometrics [76], atmospheric sciences [10] and across many other disciplines—the reader is encouraged to look into [36, 71] and the references therein for an introduction and a comprehensive list of all its applications. Despite this huge literature, there is no unified asymptotic distributional theory for the cyclostationary processes. Our Gaussian approximation result allows a systematic study of asymptotic distributions of statistics of such processes.

2.5.2. *Locally stationary process*. In (2.16), let $\Gamma = [0, 1]$. Assume that g is stochastic Lipschitz continuous for some constant $L > 0$ such that for all θ, θ',

$$\| g(\theta, \mathcal{F}_t) - g(\theta', \mathcal{F}_t) \|_p \le L |\theta - \theta'|. \tag{2.18}$$

Then the processes $X_{t,n} := g(t/n, \mathcal{F}_t)$ are locally stationary in view of the approximation

$$\| X_{t,n} - X_t(\theta) \|_p \le L |t/n - \theta| \quad \text{if } t/n \in (\theta - \Delta, \theta + \Delta) \text{ for some } \Delta > 0.$$

Dahlhaus [20, 21] introduced locally stationary processes in terms of a time-varying spectrum. Richter and Dahlhaus [86] provided a general asymptotic theory for such processes. For further examples, see [102].

Consider the special case of locally stationary version of Volterra processes, defined as follows:

$$X_t = \sum_{0 \le j_1 < \cdots < j_i} a\left(j_1, \dots, j_i, \frac{t}{n}\right) \varepsilon_{t-j_1} \dots \varepsilon_{t-j_i}, \tag{2.19}$$

where ε_i's are i.i.d. with mean 0, $\|\varepsilon_0\|_p < \infty$, $p > 2$ and $a : \mathbb{R}^i \times [0, 1] \to \mathbb{R}$ are called ith-order Volterra kernels. Then elementary calculations show that for a constant c_p depending only on p,

$$\begin{aligned} &\delta_p(l)^2 \le c_p \|\varepsilon_0\|_p^{2i} \sup_k A_{k,l,i}, \\ &\text{where } A_{k,l,i} = \sum_{0 \le j_1 < \cdots < j_i, l \in \{j_1, \dots, j_i\}} a^2\left(j_1, \dots, j_i, \frac{k}{n}\right) < \infty. \end{aligned} \tag{2.20}$$

2.6. *Outline of the proof of theorems.* Our proofs are quite involved and are given in Sections 8 and 9. In particular, Theorems 2.2 and 2.4 are based on similar assumptions (in fact Theorem 2.4 works with a weaker set of conditions), and in the same vein, Theorems 2.3 and 2.5 require exactly the same conditions. Therefore, these two pairs of theorems are proven with each other. In particular, all the four theorems follow a general recipe of the proof outlined below.

- *Truncation:* In Proposition 8.1, we truncate our process at level $n^{1/p}$ in order to exploit the uniform integrability condition, which is necessary due to nonstationarity.
- *m-dependence:* In the second step, we use the m-dependence approximation in Proposition 8.2 where m increases with n. This limits the arbitrary nonstationary dependency structure to those only up to m lags, and enables us to treat our series much like a stationary time series. We provide an optimal choice of m so that the error rate of $n^{1/p}$ is achieved.
- *Blocking:* Our blocking step in Proposition 8.3 is quite different from that in [54] as well as [9]; we consider a two-step blocking, with an inner layer of blocks of size m being then combined into an outer layer of blocks of size 3. This enables us to do the required mathematical manipulation to obtain an explicit form of the variance in terms of m-dependent processes.
- *Conditional and unconditional Gaussian approximation:* With the blocking step as mentioned above, we condition on the shared ε's between the outer blocks (that occur at both the boundaries of each block). This results in conditional independence, and thus we can use [89]'s Theorem 1. Then we lift the conditioning random variables (the boundary ε's) by taking another expectation over them, and apply the Theorem 1 from [89] again to obtain the unconditional Gaussian approximation.
- *Regularization of variance:* From the variance in terms of m-dependent blocked processes as mentioned above, in order to obtain the variance approximation in a practically usable form as mentioned in the theorem, in this step we approximate it by $\mathbb{E}((S_i^{\oplus})^2)$ or by variances of sum of blocks in terms of original random process.
- *Final Gaussian approximation:* In this final step, we connect the approximated variance $\mathbb{E}((S_i^{\oplus})^2)$ to the new Gaussian process $(Y_i)_{i=1}^n$ (for Theorems 2.2 and 2.3), via Propositions 8.5 and 8.6, or to the final variance $\mathbb{E}(S_i^2)$ (for Theorems 2.4 and 2.5).

3. Estimating the variance of the approximating Gaussian process. In this section, we address estimating the variance of the approximating process. It is well known in the time series literature that S_i^2 is a poor estimate for $\mathbb{E}(S_i^2)$. The usual practice is to use a kernel function or a particular weighing mechanism. Such methods have been used throughout the literature to estimate spectral density matrices for one-dimensional or low-dimensional cases. For stationary processes, we recommend works by Newey and West [72], Priestley [83] and Liu and Wu [64] among others for a comprehensive review of research in this direction. As a special case of kernel-based estimates, blocking techniques have been particularly popular in this area. Carlstein [16] used nonoverlapping blocks to consistently estimate $\mathbb{E}(S_i^2)$ for a stationary process. From a bootstrap perspective, Politis and Romano [80] use nonoverlapping blocks of random sizes to define a "stationary bootstrap." Using the "flat-top kernel" methods of [81, 82] obtains $O(n^{1/3})$ for the expected optimal block size for the stationary bootstrap. For detailed discussion, readers are encouraged to look into Lahiri [58], which combines ideas from [16, 17, 42] and many others to deduce various resampling schemes for estimating the variance of a stationary process.

The blocking method has been quite popular in the literature as a proof technique for obtaining optimal Gaussian approximations. See [54, 96] and [63] for relevant references. Naturally, since the statements of our Theorems 2.2–2.5 do not involve any blocks, one may question if we can reach the optimal rate by expressing the variance directly in terms of some blocking mechanism. In the next section, we will provide a result that answers the above question in affirmative. The blocking mechanism we use is somewhat related to the nonoverlapping block bootstrap (NBB) method proposed in Chapter 2 of [58]. We describe the scheme in the following. Usually, the block length m is taken so as $m \to \infty$ with $m/n \to 0$. Define for $1 \le a, k, j \le \lceil n/m \rceil$,

$$\begin{aligned} B_a &:= \sum_{i=(a-1)m+1}^{am \wedge n} X_i; \qquad T_k = \sum_{a=1}^{k} B_a^2 + 2\sum_{a=1}^{k-1} B_a B_{a+1}; \\ R_j &:= \mathbb{I}\{j/m \notin \mathbb{N}\} \sum_{i=\lfloor j/m \rfloor m+1}^{j} X_i. \end{aligned} \tag{3.1}$$

Note that $S_j = \sum_{a=1}^{k} B_a + R_j$, where $k = \lfloor j/m \rfloor$. We shall estimate $\mathbb{E}(S_j^2)$ by the following "Block-based Running Variance" (BRV) estimator $\mathcal{T}_j$ where

$$\mathcal{T}_j := T_{\lfloor j/m \rfloor} + R_j^2 + 2B_{\lfloor j/m \rfloor} R_j \quad \text{for all } 1 \le j \le n \tag{3.2}$$

simultaneously. Since $\mathcal{T}_j$'s may be negative, so instead of Brownian motion we use two-sided Brownian motion. A two-sided Brownian motion is defined as $\mathbb{W}(t) = \mathbb{B}_1(t)\mathbf{1}_{t \ge 0} + \mathbb{B}_2(-t)\mathbf{1}_{t<0}$, where $\mathbb{B}_1$ and $\mathbb{B}_2$ are two independent standard Brownian motions starting at 0.

Next, we provide some theoretical properties of the BRV estimator $\mathcal{T}_j$. In particular, we bound the uniform deviation probability of $\mathcal{T}_j$. Such a deviation inequality for nonstationary processes is novel to the best of our knowledge. Thus, we state it as a standalone result.

3.1. *A maximal quadratic large deviation bound.* Quadratic large deviation bounds have a long history that started with the seminal work by Hanson and Wright [43] and Hanson and Wright [93]. See [88] for an extensive overview. These are popularly referred as Hanson–Wright type inequalities in the literature. Subsequent work by [8, 52] and others established *moderate deviation principles* for quadratic forms of stationary Gaussian processes. Moving beyond sub-Gaussianity, Xiao and Wu [97] and Zhang and Wu [102] generalized the Hanson–Wright inequality for stationary process with finite polynomial moments and locally stationary processes, respectively. In this section, we aim to (i) develop a maximal inequality, that is, derive tail probability bounds for the maximal partial sum, and (ii) relax the stationarity assumption by providing a result for the general nonstationary processes. Our proof is similar to the Theorem 6.1 of [102]; however, it differs in a crucial step. Since we aim to provide a maximal inequality, we use Borovkov's version of Nagaev inequality [11], instead of the usual bound of [70]. This, in particular, changes the treatment of a few important terms in our proof compared to that in [102]. Moreover, we also tackle the case when $2 < p \le 4$, something that is usually absent from other Hanson–Wright type inequalities in the literature.

THEOREM 3.1. *Let $p > 2$. Assume Condition 2.1 holds for $\Theta_{i,p}$. Let $Q_n = \sum_{1 \le s \le t \le n} a_{s,t} X_s X_t$, with $a_{s,t} = 0$ if $|s-t| > \mathcal{D}_n$ for some $\mathcal{D}_n \le n$, and $\sup |a_{s,t}| \le 1$. Denote*

$$\begin{aligned} R_k &= \sum_{j=1}^{k} \big(V_j - \mathbb{E}(V_j)\big), \\ &\text{where } V_k = \sum_{t=(k-1)\mathcal{D}_n+1}^{(k\mathcal{D}_n) \wedge n} \sum_{1 \le s \le t} a_{s,t} X_s X_t, \text{ for } 1 \le k \le \lceil n/\mathcal{D}_n \rceil. \end{aligned} \tag{3.3}$$

Then there exists constants C_p, depending only on p such that for all $x > 0$,

$$\mathbb{P}\Big(\max_{1\le k\le \lceil n/\mathcal{D}_n\rceil} |R_k| \ge x\Big) \le \begin{cases} C_p x^{-p/2} n \mathcal{D}_n^{p/4} \mu_{p,A}^p, & 2 < p \le 4, \\ C_p x^{-p/2} n \mathcal{D}_n^{p/2-1} \mu_{p,A}^p + C_p \exp\Big(-\dfrac{C_p x^2}{n\mathcal{D}_n \mu_{4,A}^4}\Big), & p > 4. \end{cases} \tag{3.4}$$

The proof is given in Appendix Section 10.1. We emphasize that to avoid notational cumbersomeness, in (3.4) we have used same notation C_p to denote multiple constants, each depending solely on p.

REMARK 3.1. In view of (2.6), $\delta_p^{\oplus}(j) \le \delta_p(j)$ is satisfied by the functional dependence measure of the truncated process. Therefore, Theorem 3.1 also holds for X_s replaced by $X_s^{\oplus} - \mathbb{E}(X_s^{\oplus})$.

REMARK 3.2. The bound in Theorem 3.1 should be contrasted with the bound obtained in Theorem 6 of [102]. In fact, our proof works for $A > 1/2 - 1/q$ and matches their nonuniform bound for the corresponding case. A similar argument can be followed to yield a bound for a process satisfying $\mu_{p,A} < \infty$ for some general A. In view of our maximal inequality holding true for a general nonstationary process, Theorem 3.1 is a more general result than those found in the literature.

3.2. *Gaussian approximation rate with estimated variance.* Theorem 3.1 is useful in arriving at the estimation error of $\mathcal{T}_i$ as an estimate of $\mathbb{E}(S_i^2)$. To begin with, note that $\mathcal{T}_i/2$ can be written in the form (3.3) with $a_{s,t} = 1/2$ when $s = t$, and in general $|a_{s,t}| = 0$ when $|s - t| \ge 2m$ and $\sup |a_{s,t}| \le 1$. Thus, taking $\mathcal{D}_n = 2m$, Theorem 3.1 implies that

$$\max_{1\le k\le \lfloor n/m\rfloor} \Big| \sum_{j=1}^{k} \big(B_j^2 + 2B_jB_{j+1} - \mathbb{E}[B_j^2 + 2B_jB_{j+1}]\big)\Big| = O_{\mathbb{P}}\big(n^{\max\{2/p,1/2\}} m^{1/2}\big). \tag{3.5}$$

Moreover, by Lemma 8.2, $\max_{1\le j\le \lfloor n/m\rfloor} \mathbb{E}[\max_{1\le k\le m} |X_{mj+1} + \cdots + X_{mj+k}|^p] = O(m^{p/2})$. Hence,

$$\max_{1\le i\le n} \Big| \mathcal{T}_i - \sum_{j=1}^{\lfloor i/m\rfloor} \big(B_j^2 + 2B_jB_{j+1}\big)\Big| = O_{\mathbb{P}}\big(n^{\max\{2/p,1/2\}} m^{1/2}\big) \tag{3.6}$$

by Markov's inequality. Note that (3.6) takes care of the stochastic error of $\mathcal{T}_i$ as an estimate of $\mathbb{E}(S_i^2)$ for $1 \le i \le n$. For the bias part, we need to control the order of the cross-product terms $\mathbb{E}(B_iB_j)$ for $i \ne j$. The following lemma, whose proof we give in Section 10.2, is thus necessitated.

LEMMA 3.1. *Let Condition 2.1 hold with $A > 1$. Then for B_j as defined in (3.1), it holds that*

$$\max_{1\le k\le \lfloor n/m\rfloor} \big|\mathbb{E}(B_kB_{k+1})\big| = O(1), \qquad \max_{1\le k\le \lceil n/m\rceil} \sum_{i:|i-k|\ge 2} \big|\mathbb{E}(B_iB_k)\big| = O\big(m^{1-A}\big). \tag{3.7}$$

Observe that (3.7) readily yields

$$\max_{1\le i\le n} \Big| \mathbb{E}(S_i^2) - \sum_{j=1}^{\lfloor i/m\rfloor} \mathbb{E}\big(B_j^2 + 2B_jB_{j+1}\big)\Big| = O\big(nm^{-A}\big). \tag{3.8}$$

Now (3.5), (3.6) and (3.8) can be summarized into the following proposition.

PROPOSITION 3.1. *Assume $p > 2$ and let Condition 2.1 hold for $\Theta_{i,p}$ with $A > 1$. Recall B_j from (3.1), for a general $m \in \mathbb{N}$. Then the following holds*:

$$\max_{1\le i\le n}\left|\mathcal{T}_i - \mathbb{E}(S_i^2)\right| = O_{\mathbb{P}}\left(n^{\max\{2/p,1/2\}}m^{1/2} + nm^{-A}\right). \tag{3.9}$$

In particular, with $m \asymp n^{\zeta_1}$, where $\zeta_1 = \min\{1, 2-4/p\}/(1+2A)$, (3.9) implies

$$\max_{1\le i\le n}\left|\mathbb{W}(\mathcal{T}_i) - \mathbb{B}_1\left(\mathbb{E}(S_i^2)\right)\right| = O_{\mathbb{P}^*}\left(n^{(1-A\zeta_1)/2}\sqrt{\log n}\right), \tag{3.10}$$

where $\mathbb{P}^*$ refers to the conditional distribution after observing $X_1, \ldots, X_n$, and $\mathbb{B}_1(\cdot)$ is the same Brownian motion defining the positive half-line of $\mathbb{W}(\cdot)$.

Our choice of m balances the bias (nm^{-A}) and the stochastic error ($n^{\max\{2/p,1/2\}}m^{1/2}$) together, and yields the rate in (3.10) by increment property of Brownian motions. However, the approximation rate in (3.10) is worse than what we obtain in Section 2. But this also means that one can only assume moments slightly higher than 4 and still achieve this rate. More importantly, a natural question is if we can relax our decay condition in Theorem 2.4 when we are allowed to assume p finite moments but want to achieve this comparatively large approximation rate. In other words, at the cost of the suboptimal rate, which anyway is the best for the empirical version, can we allow decay rate A to be smaller? In what follows, we answer this question in affirmative.

THEOREM 3.2. *Let $p > 2$. Assume that the decay Condition 2.1 holds with $A > 1$. Further grant the truncated uniform integrability Condition 2.2. Then there exists a Brownian motion $\mathbb{B}(\cdot)$ such that*

$$\max_{1\le j\le n}\left|S_j - \mathbb{B}\left(\mathbb{E}(S_j^2)\right)\right| = o_{\mathbb{P}}\left(n^{(1-A\zeta_1)/2}\sqrt{\log n}\right). \tag{3.11}$$

REMARK 3.3. Note that in (3.11) we no longer need the lower bound (2.10).

3.3. *Gaussian approximation without cross product blocks*. Having explored the asymptotic properties of BRV estimator $\mathcal{T}_j$ as an estimate of $\mathbb{E}(S_j^2)$ for $1 \le j \le n$, let us discuss a natural variant of $\mathcal{T}_j$. Interestingly, in $\mathcal{T}_j$ we have included the cross-product terms $B_i B_{i+1}$, as opposed to another possible estimate $\mathcal{T}_i^-$, which can be defined without them:

$$\mathcal{T}_i^- = \sum_{j=1}^{\lfloor i/m\rfloor} B_j^2 + R_i^2. \tag{3.12}$$

An application of Theorem 3.1 and (3.7) similar to that in Proposition 3.1 show $\mathcal{T}_i^-$ satisfies

$$\max_{1\le i\le n}\left|\mathcal{T}_i^- - \mathbb{E}(S_i^2)\right| = O_{\mathbb{P}}\left(n^{\max\{2/p,1/2\}}m^{1/2} + nm^{-1}\right) \tag{3.13}$$

under Condition 2.1. The above bound is worse than (3.9) and it is minimized at $m \asymp n^{\zeta_2}$, $\zeta_2 = \min\{1, 2-4/p\}/3$. Since $A > 1$, $\zeta_2 < \zeta_1$ and, therefore,

$$\max_{1\le i\le n}\left|\mathbb{W}(\mathcal{T}_i^-) - \mathbb{B}\left(\mathbb{E}(S_i^2)\right)\right| = O_{\mathbb{P}^*}\left(n^{(1-\zeta_2)/2}\sqrt{\log n}\right). \tag{3.14}$$

Thus, the conditional version (3.10) using $\mathcal{T}_i^-$ is also worse.

Following the idea of the moving or overlapping block bootstrap method (cf. [57] and [62], Zhou [107] and Mies and Steland [68]) consider the following estimate of $\mathbb{E}(S_i^2)$ by

$$\mathcal{T}_i^{\diamond} = \sum_{t=m}^{i}\frac{1}{m}\left(\sum_{s=t-m+1}^{t} X_s\right)^2. \tag{3.15}$$

A treatment similar to Proposition 3.1 shows that $\mathcal{T}_i^\diamond$ satisfies (3.13) as well. Thus, $\mathcal{T}_i$ has the best rate for estimating the variance of the Brownian motion among the three estimators discussed here. It should be noted that [68] analyzes a different variance for the approximating Gaussian process (defined as a local long-range variance $\sigma^2_{\mathrm{loc}_i}$), and $\mathcal{T}_i^\diamond$ has been proposed in that context. However, we point out that for fast enough decay, their rate of Gaussian approximation $\max_{1\le i\le n}|S_i^c - \mathbb{B}(\sigma^2_{\mathrm{loc}_i})| = o_{\mathbb{P}}(n^{p/(3p-2)}\sqrt{\log n})$ is suboptimal in n.

4. Applications of Gaussian approximations. In this section, we are interested in obtaining Gaussian approximations of functionals of the form

$$W(t) := \sum_{i=1}^n e_i w_i(t),$$

where $w_i(\cdot) : [0,1] \to \mathbb{R}$ are weight functions and $(e_i)_{1\le i\le n}$ are real-valued, mean-zero, possibly nonstationary processes. Such quantities are ubiquitous in various statistics of change point estimation, wavelet transform and forming a simultaneous confidence band, among others. One can employ (2.13) of Theorem 2.4 to deal with such quantities. A similar treatment is included in [95]. Let

$$W^\diamond(t) = \sum_{i=1}^n w_i(t)\big(\mathbb{B}\big(\mathbb{E}(S_i^2)\big) - \mathbb{B}\big(\mathbb{E}(S_{i-1}^2)\big)\big) \tag{4.1}$$

be the Gaussian process that we want to use to approximate $W(t)$, where $S_i = \sum_{j=1}^i e_j$. Let

$$\Omega_n = \sup_{t\in(0,1)}\left\{\big|w_1(t)\big| + \sum_{i=2}^n\big|w_i(t) - w_{i-1}(t)\big|\right\} \tag{4.2}$$

be the maximum variation of the weights $w_i(t)$. Then

$$\sup_{t\in(0,1)}\big|W(t) - W^\diamond(t)\big| \le \Omega_n \max_{1\le i\le n}\big|S_i - \mathbb{B}\big(\mathbb{E}(S_i^2)\big)\big| = \Omega_n o_{\mathbb{P}}\big(n^{1/p}\sqrt{\log n}\big). \tag{4.3}$$

In the following, we detail three applications—testing for change point, simultaneous confidence band building and wavelet transform—using the above analysis. Each of these analyses requires providing a rate of Ω_n depending on certain conditions.

4.1. *Change-point detection.* Assume $X_i = \mu_i + Z_i$, $i = 1, \dots, n$, where (Z_i) is a mean zero nonstationary process. We want to test for the existence of change point in means, that is, we want to test for $H_0 : \mu_i = \mu_0$ for all i versus the alternative hypothesis,

$$H_1 : \mu_i = \mu_0 + \delta\mathbb{I}\{i > \tau\} \quad \text{holds for some } 1 < \tau < n \text{ and } \delta \ne 0. \tag{4.4}$$

We propose a CUSUM-based testing procedure with test statistic

$$U_n := \max_{t\in(0,1)}\left|\sum_{i\le nt}(X_i - \bar{X})\right|/\sqrt{n}, \tag{4.5}$$

where we reject our null hypothesis if U_n is larger than some suitable cut-off value. Under the null hypothesis, we can write $U_n = \max_{t\in(0,1)}|U_{n,t}|$, where $U_{n,t} := \sum_{i=1}^n w_i(t)Z_i$ and the weights $w_i(t) = ((1-1/n)\mathbb{I}\{i \le nt\} - (1/n)\mathbb{I}\{i > nt\})/\sqrt{n}$. Let

$$V_n = \max_{t\in(0,1)} V_{n,t} \quad \text{where } V_{n,t} := \sum_{i=1}^n w_i(t)\big(\mathbb{B}\big(\mathbb{E}(S_i^2)\big) - \mathbb{B}\big(\mathbb{E}(S_{i-1}^2)\big)\big).$$

By (4.3), we have $|U_n - V_n| = o_{\mathbb{P}}(1)$ since $\Omega_n = (2-1/n)/\sqrt{n}$ and $\Omega_n n^{1/p}\sqrt{\log n} \to 0$.

4.2. *Simultaneous confidence band.* In this section, we discuss construction of simultaneous confidence band for a time-varying signal-plus-noise model with possibly irregularly spaced observed data and possibly nonstationary noise. Let $0 = t_0 < t_1 < t_2 < \cdots < t_{n-1} < t_n < t_{n+1} = 1$ be an n-length grid on [0, 1]. Consider

$$X_i = \mu(t_i) + Z_i, \quad i = 1, \dots, n, \tag{4.6}$$

where $\mu(\cdot) \in \mathcal{C}^3[0, 1]$. The case $t_i = i/n$ has been thoroughly analyzed in the literature for stationary and i.i.d. setup, such as [13, 30] and [95]. Here, we let $t_i = F^{-1}(i/n)$, where $F(t) = \int_0^t f(u)\,\mathrm{d}u$ for some density $f \in C^3[0, 1]$. We will estimate the trend function from observed data (X_i) using the local linear estimate, and denote the result by $\hat{\mu}_{h_n}(\cdot)$, where h_n is the bandwidth parameter. Define

$$S_j(t) = \sum_{i=1}^{n} (t - t_i)^j K\big((t - t_i)/h_n\big). \tag{4.7}$$

Theorem 4.1 below provides a Gaussian approximation for the local linear estimate

$$\hat{\mu}_{h_n}(t) := \sum_{i=1}^{n} w_{h_n}(t, i) X_i \quad \text{where } w_{h_n}(t, i) = K\left(\frac{t - t_i}{h_n}\right) \frac{S_2(t) - (t - t_i) S_1(t)}{S_2(t) S_0(t) - S_1^2(t)}. \tag{4.8}$$

Assume that K is a smooth symmetric kernel with bounded support $[-\omega, \omega]$, satisfying

$$\begin{aligned} &\int_{\mathbb{R}} \Psi_K(u; \delta)\,\mathrm{d}u = O(\delta) \\ &\quad \text{as } \delta \to 0, \text{ where } \Psi_K(u; \delta) = \sup\big\{\big|K(y) - K(u)\big| : |y - u| \le \delta\big\}. \end{aligned} \tag{4.9}$$

THEOREM 4.1. *Assume $\mu, f \in \mathcal{C}^3[0, 1]$ and, for some constants $C_1, C_2 > 0$, $C_1 \le f(t) \le C_2$ for all $t \in [0, 1]$. Then under the assumptions of Theorem 2.4 for Z_i, there exists Brownian motion $\mathbb{B}(\cdot)$ such that with $Q_{h_n}(t) = \sum_{i=1}^n w_{h_n}(t, i) \mathbf{Y}_i$, where $\mathbf{Y}_i = \mathbb{B}(\mathbb{E}(S_i^2)) - \mathbb{B}(\mathbb{E}(S_{i-1}^2))$, the following is true*:

$$\sup_{t \in [\omega h_n, 1 - \omega h_n]} \big|\hat{\mu}_{h_n}(t) - \mu(t) - h_n^2 \beta \mu''(t) - Q_{h_n}(t)\big| = o_{\mathbb{P}}\big(h_n^{-1} n^{1/p - 1} \sqrt{\log n}\big), \tag{4.10}$$

for any $h_n \to 0$ satisfying $h_n^4 = O(n^{1/p-1})$ and $nh_n \to \infty$ with $\beta = \int u^2 K(u)\,du/2$.

PROOF. We apply Theorem 2.4 to $(Z_i)_{i=1}^n$. Note that $Q_{h_n}(t)$ is obtained by fitting the same local linear regression with bandwidth h_n to $(\mathbf{Y}_i)_{i=1}^n$. By the argument in Theorem 3.1 in [31], $\mathbb{E}[\hat{\mu}_{h_n}(t)] - \mu(t) = h_n^2 \mu''(t) \beta + O(h_n^3 + n^{-1} h_n^{-1})$. Then (4.10) follows by applying (4.3) to $\hat{\mu}_{h_n}(t) - \mathbb{E}[\hat{\mu}_{h_n}(t)] - Q_{h_n}(t)$ and noting that $\Omega_n = O(1/(nh_n))$ using Lemma 11.1 and $C_1 \le f(\cdot) \le C_2$. $\square$

4.2.1. *Bias correction.* Using (4.10) to construct simultaneous confidence band requires estimation of $\mu''(t)$. Following [44], we use the jackknife-based bias corrected estimator

$$\tilde{\mu}_{h_n}(t) = 2\hat{\mu}_{h_n}(t) - \hat{\mu}_{h_n\sqrt{2}}(t). \tag{4.11}$$

Using (4.11) is asymptotically equivalent to using the kernel $K^*(x) = 2K(x) - K(x/\sqrt{2})/\sqrt{2}$; see [95, 108] and [53] among others. Based on (4.11), one can observe $\mathbb{E}[\tilde{\mu}_{h_n}(t)] - \mu(t) = O(h_n^3 + n^{-1} h_n^{-1})$. Thus, one can get rid of the $h_n^2 \mu''(t)$ term from the left-hand side of the (4.11) to obtain

$$\sup_{t \in [\omega h_n, 1 - \omega h_n]} \big|\tilde{\mu}_{h_n}(t_i) - \mu(t_i) - \tilde{Q}_{h_n}(t_i)\big| = o_{\mathbb{P}}\big(h_n^{-1} n^{1/p-1} \sqrt{\log n}\big). \tag{4.12}$$

4.2.2. *Choice of bandwidth* h_n. Since our Gaussian approximation Theorem 2.4 holds with $n^{1/4}$ rate for $p \geq 4$, $A > A_0$, for this subsection, assume $p = 4$. Ignoring the log factors, we obtain a rate of $O_{\mathbb{P}}(n^{-3/4}/h_n)$ from (4.10), which readily allows a large range of h_n:

$$n^{-3/4} \leq h_n \leq n^{-3/16}. \tag{4.13}$$

In particular, (4.13) allows for $h_n \asymp n^{-1/5}$, which is the mean-square error optimal bandwidth. As equation (4.12) suggests, $\tilde{Q}_{h_n}$ is a good simultaneous approximation for $\tilde{\mu}_{h_n} - \mu$ in distribution. Therefore, for our bootstrap algorithm, $\tilde{Q}_{h_n}$ is generated based on $(\mathbf{Y}_i)$, which is simulated from our Gaussian approximation where we estimate $\mathbb{E}(S_i^2)$ by $\mathcal{T}_i$'s formed by Z_i as in (3.1). Using this, for $0 < \alpha < 1$, we can calculate $q_{1-\alpha}$, the empirical $(1-\alpha)$-th quantile of $\max_{1 \leq i \leq n} |\tilde{Q}_{h_n}(i/n)|$. Thus, given significance level α, the simultaneous confidence level for $\mu(\cdot)$ can be constructed as

$$\left[\tilde{\mu}_{h_n}(t) - q_{1-\alpha}, \tilde{\mu}_{h_n}(t) + q_{1-\alpha}\right], \quad t \in [0, 1]. \tag{4.14}$$

4.3. *Wavelet coefficient process*. Wavelet transform is a way of representing a time series locally both in time and frequency windows. Mathematically speaking, wavelength coefficients are simply the coefficients when the signal $(X_i)_{1 \leq i \leq n}$ is decomposed in terms of some orthonormal basis of $L^2(\mathbb{R})$. The simplest discrete wavelet transform used is called the Haar transform [41]. Assume the signal length is $n = 2^k$. Then the jth level Haar wavelet coefficients with $j \leq k$ are

$$\begin{aligned} &W_{j,t} = \sum_{l=1}^{2^j} h_{j,l} X_{2^j t - l + 1}, \\ &t = 1, \dots, 2^{k-j}, \text{ where } h_{j,l} = \begin{cases} -2^{-j/2} & \text{if } 1 \leq l \leq 2^{j-1}, \\ 2^{-j/2} & \text{if } 2^{j-1} < l \leq 2^j. \end{cases} \end{aligned} \tag{4.15}$$

Donoho [26] used wavelet methods to perform nonparametric signal estimation via soft thresholding; however, their threshold value crucially depends on the assumptions of the noise process being i.i.d. Gaussian. Johnstone and Silverman [51] and von Sachs and MacGibbon [92] extended the results for correlated Gaussian and locally stationary noise processes, respectively. Recently, [67] considered locally stationary wavelet processes as the noise processes for estimation of signal. Stationarity assumption also features crucially in the wavelet variance estimation mechanism of Percival and Mondal [77]. Here, we allow the signal $(X_i)_{1 \leq i \leq n}$ to be possibly nonstationary, and focus on applying our Theorem 2.4 to provide a Gaussian approximation result for the wavelet coefficient process $W_{j,t}$. Note that $W_{j,t}$ can be written as $\sum_{i=1}^n w_i(j,t) X_i$, where $w_i(j,t) = h_{j, 2^j t - i + 1}$. Let

$$W_{j,t}^{\diamond} = \sum_{i=1}^n w_i(j,t) \left(\mathbb{B}(\mathbb{E}(S_i^2)) - \mathbb{B}(\mathbb{E}(S_{i-1}^2))\right).$$

With Ω_n as defined as in (4.2), it can be easily seen that $\Omega_n = O(2^{-j/2})$. Thus, using (4.3), we get

$$\max_{j_* \leq j \leq k} \max_{1 \leq t \leq n/2^j} \left|W_{j,t} - W_{j,t}^{\diamond}\right| = o_{\mathbb{P}}\left(2^{-j_*/2} n^{1/p} \sqrt{\log n}\right). \tag{4.16}$$

To ensure a uniform Gaussian approximation, we require the highest resolution level j_* to satisfy

$$j_* - \frac{2}{\log 2}\left(\frac{1}{p} \log n + \frac{1}{2} \log \log n\right) \to \infty. \tag{4.17}$$

In particular, it holds if $j_* \geq c \log n$ for some constant $c > 2/(p \log 2)$. Similar analysis can be performed for the more general Daubechies wavelet filters (Daubechies [22]), with better smoothness properties. The uniform Gaussian approximation (4.16) allows an asymptotic distributional theory for statistics based on wavelet transforms of nonstationary processes.

5. Simulation. This section presents a simulation study for some of our results in Sections 2, 3 and 4 while some more are postponed to the Appendix Section 12. Our aims are as follows. In Section 5.1, we start off by investigating the accuracy of the two kinds of theoretical Gaussian approximations in Sections 2.3 and 2.4. We postpone inspecting the accuracy of our bootstrap Gaussian approximations for finite sample to Appendix Section 12.1. In particular, in Section 3.3, having argued that excluding the cross-product terms results in a worse rate and a less accurate approximation compared to (3.10), we compare their finite sample accuracy for some simple cases. Moving on to showing simulation-based evidences for our applications, in Section 5.2, we explore the empirical coverage of our simultaneous confidence band procedure discussed in Section 4.2 under different settings. We again defer analysing the performance of the CUSUM-based testing procedure for existence of change-point, as discussed in Section 4.1 to Appendix Section 12.3.

5.1. *Empirical accuracy of theoretical Gaussian approximations.* Consider two models:
5.1. Model 5.1: $X_t = \theta X_{t-1} + \varepsilon_t, \theta \in \{0.9, -0.9\}$.
5.2. Model 5.2: $X_t = \theta_t X_{t-1} + \varepsilon_t$, $\theta_t = \theta$ if $t \leq n/2$, $\theta_t = -\theta$ if $t > n/2$, $\theta \in \{0.9, -0.9\}$.
We will start off by letting $\varepsilon_t \overset{\text{i.i.d.}}{\sim} t_4/\sqrt{2}$ for both of the models. Observe that, with $N(0,1)$ innovations, $(X_t)_{t=1}^n$ is already a Gaussian process for both Models 5.1 and 5.2 and, therefore, the approximation error is trivially zero. This motivates the use of some other mean-zero error for this model. We will initially consider a small sample of size $n = 100$. For each of the setups, we will compare the quantiles of the following three random variables:

$$U_X := \max_{1 \leq i \leq n} S_i, \qquad U_1 = \max_{1 \leq i \leq n} \mathbb{B}(\mathbb{E}(S_i^2)), \qquad U_2 = \max_{1 \leq i \leq n} \sum_{j=1}^{i} Y_i,$$

where $(Y_t)_{t=1}^n$ is a centered Gaussian process with same covariance structure as $(X_t)_{t=1}^n$. The true quantiles are estimated by sample quantiles based on 10^3 repetitions. Figures 1 and 2 depict the QQ-plots of U_1 and U_2 against U_X. Clearly, when compared with U_1, which involves Brownian motion, our Gaussian approximation of Section 2.3 maintaining covariance structure, performs much better for such a small sample size $n = 100$. However, as we increase n, both the approximations being theoretically valid with optimal rate of convergence, their performances become comparable. To show this empirically, we consider two more complicated nonstationary models.
5.3. Let $w_1 = \underbrace{0.75, \ldots}_{n/4}, \underbrace{-0.75, \ldots}_{n/4}, \underbrace{0.75, \ldots}_{n/4}, \underbrace{-0.75, \ldots}_{n/4}$, $w_2 = (\sin(8\pi t/n))_{t=1}^n$, and

$$X_t = \theta_t X_{t-1} + \varepsilon_t, \qquad \theta_t = \theta w_{it}, \qquad X_0 = 0, \quad i \in \{1, 2\}, \theta \in \{-0.8, 0.8\}.$$

5.4. $X_t = \sin(Y_t)$, where $Y_t \sim$ Model 5.3.
To further show the efficacy of our approximation, we consider a skewed error for Model 5.3 with i.i.d. $\chi_1^2 - 1$ errors. We consider i.i.d. $N(0,1)$ innovations for Model 5.4. Note that due to the sin transformation, Model 5.4 is no longer Gaussian. The corresponding QQ-plots are shown in Figures 3 and 4. It can be seen that both Gaussian approximations show excellent accuracy for a somewhat increased sample size $n = 200$. In fact, in some of the setups, the more natural Gaussian approximation retains an advantage over the Gaussian approximation involving the Brownian motion.

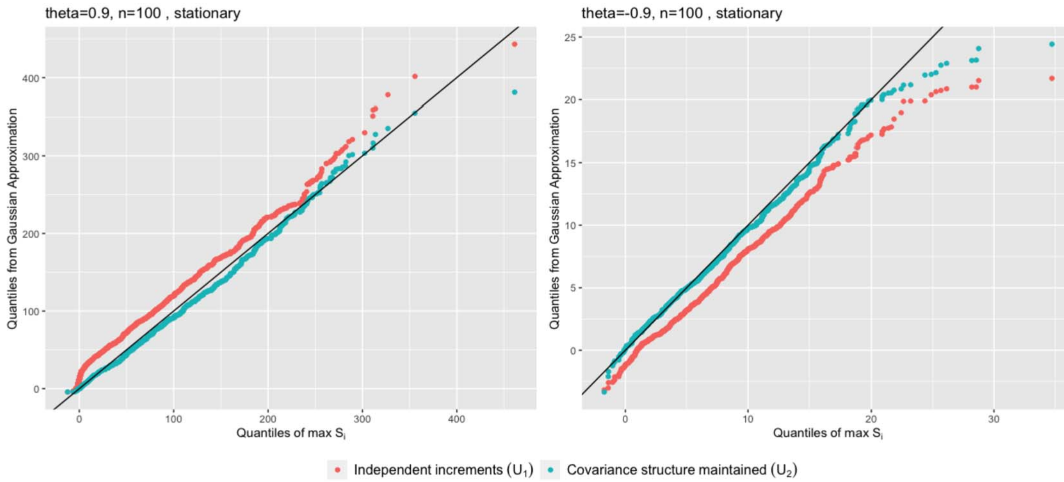


FIG. 1. *Comparison of theoretical quantiles with the two kinds of Gaussian approximation* $X_1, \dots, X_n \sim$ *Model* 5.1 *with* t_4 *innovations*: *with independent increments*, *and with the approximation maintaining covariance structure*.

5.2. *Simulation for simultaneous confidence bands.* In this subsection, we will explore the empirical coverage probabilities for our 95% SCBs constructed as in (4.14). We will use the jackknife-based bias corrected version of the local linear estimate, as in (4.11). We generate data from the model (4.6) with $\mu(t) = 0.5\cos(2\pi t - 0.7) + 0.3\exp(-t)$, with $t_i = i/n$ for $i = 1, \dots, n$. We consider the two models (5.3) and (5.4) with innovations $\varepsilon_t \sim t_6\sqrt{2/3}$ for our error generating process Z_t, and consider the two weighing schemes for each model with $\theta \in \{-0.8, -0.4, 0.4, 0.8\}$ in (5.3). We will estimate the mean curve using the Epanechnikov kernel $K(x) = \frac{3}{4}(1 - x^2)\mathbb{I}\{|x| \le 1\}$. For each of these models, we consider data of sizes $n = 600$ and 800, and bandwidths $h_n = 0.11, 0.13$ and 0.15. For each such setting, we perform 1000 replications each with 500 bootstrap samples of size n each. Following our theoretical result in Theorem 4.1 as well as the discussion at Section 3.2.5 of [31], the variance of local linear estimator is comparatively high on the boundary points, which affects coverage. Thus, we report as empirical coverage the percentage of times the estimated SCB

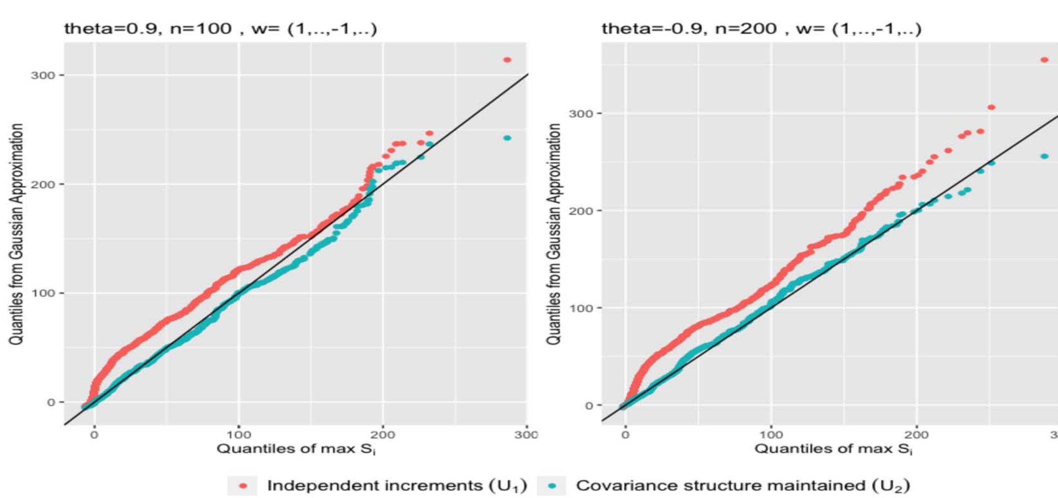


FIG. 2. *Comparison of theoretical quantiles with the two kinds of Gaussian approximation* $X_1, \dots, X_n \sim$ *Model* 5.2 *with* t_4 *innovations*: *with independent increments*, *and with the approximation maintaining covariance structure*.

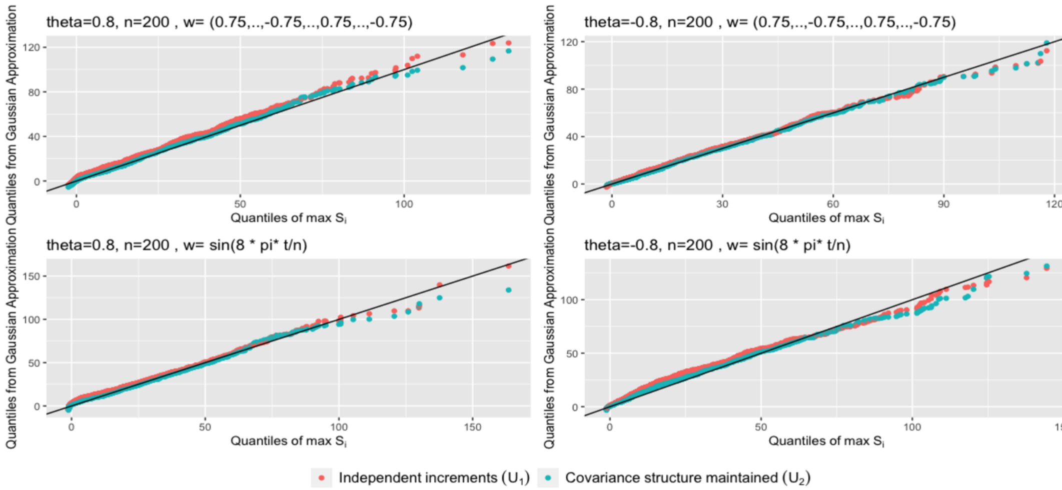


FIG. 3. *Comparison of theoretical quantiles with the two kinds of Gaussian approximation* $X_1, \dots, X_n \sim$ *Model* 5.3 *with* $\chi_1^2 - 1$ *innovations*: *with independent increments*, *and with the approximation maintaining covariance structure*.

contains the true $\mu(t)$ curve in the interval $[0.05, 0.95]$. Generally speaking, the coverage probabilities in Tables 1 and 2 are reasonably close to the nominal level 0.95. Moreover, the bandwidths do not seem to have too large an effect on the coverage probability.

6. Real data application: Analysis of Lake Chichancanab sediment density data. The Maya civilization, arguably one of the most important pre-Columbian mesoamerican civilizations, underwent a collapse during the last classical period of their history, circa 900–1100 AD [3, 24, 38, 98]. A severe drought has been hinted at as a primary reason behind this collapse [34, 40, 91], despite the Mayans primarily inhabiting a seasonally dry tropical forest [39]. Drought has also been explored as a possible cause of a comparatively less studied, preclassical Maya collapse in 150–200 AD [37]. [45–47] analyzed the sediment core density data set from the Lake Chichancanab in the Yucatan peninsula to analyze the onset

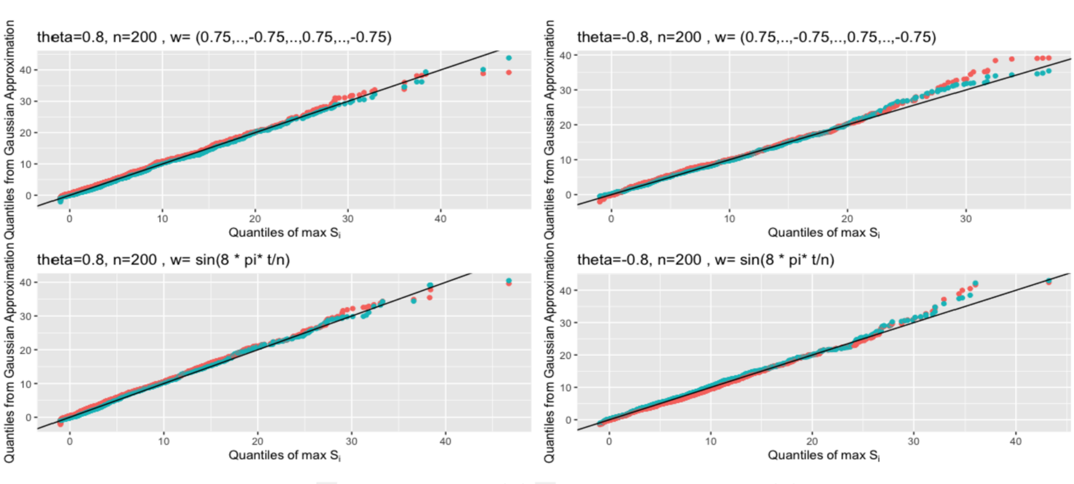


FIG. 4. *Comparison of theoretical quantiles with the two kinds of Gaussian approximation* $X_1, \dots, X_n \sim$ *Model* 5.4 *with* $N(0, 1)$ *innovations*: *with independent increments*, *and with the approximation maintaining covariance structure*.

TABLE 1
Empirical coverage probabilities of SCB of X_t from Model (4.6) where $Z_t \sim$ Model 5.3 with normalized t_6 error

		Weights: $w =$ $(0.75, \dots, -0.75, \dots, 0.75, \dots, -0.75, \dots)$				Weights: $w = \sin(8\pi t/n)$			
n	h_n	$\theta = -0.8$	$\theta = -0.4$	$\theta = 0.4$	$\theta = 0.8$	$\theta = -0.8$	$\theta = -0.4$	$\theta = 0.4$	$\theta = 0.8$
600	0.11	0.922	0.949	0.929	0.913	0.930	0.951	0.959	0.916
	0.13	0.946	0.952	0.951	0.938	0.951	0.956	0.963	0.950
	0.15	0.950	0.963	0.951	0.950	0.956	0.964	0.964	0.959
800	0.11	0.948	0.963	0.954	0.932	0.952	0.962	0.951	0.952
	0.13	0.954	0.963	0.960	0.956	0.958	0.966	0.958	0.962
	0.15	0.955	0.965	0.965	0.953	0.959	0.966	0.971	0.970

pattern of droughts during the Maya civilization. An age-depth model of radiocarbon dating is used to estimate the calendar age of depth of each sediment. The total number of data points is $n = 564$, and the corresponding years range from 858 BC to 1994 AD.

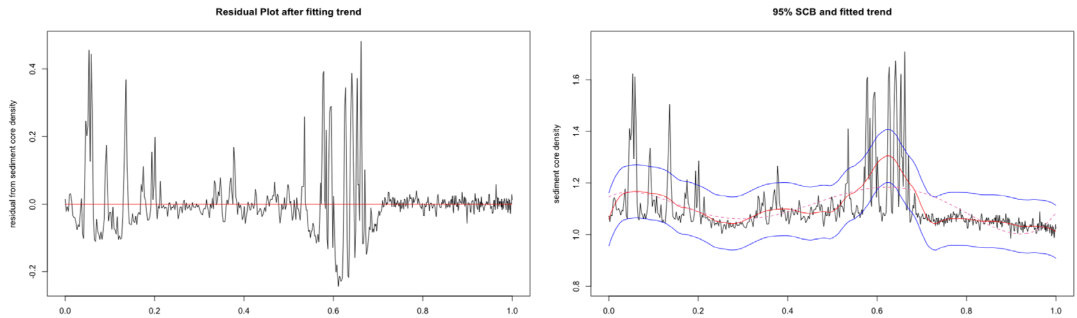
We first test the existence of a change point for this data set as described in Section 4.1. For this, we choose $m = 20$. The p-value of our test ψ_{n1} comes out to be 0.09, and thus we fail to reject nonexistence of a change point. Gill [37] posited that between 800 and 1000 AD, the Yucatan peninsula was hit by a massive drought, triggering the Mayan collapse. However, in light of our findings, such a hypothesis seems unlikely. Next, we move on to building a simultaneous confidence band as in (4.14), which we will subsequently use to test the existence of certain trend. For the local linear estimates (Figures 5b), we select $h = 0.1$. The residual plots 5a of $X_i - \hat{\mu}_L(t_i)$ where $\hat{\mu}_L$ is the locally linear estimate, suggest that the error process is indeed nonstationary. Hodell, Brenner and Curtis [45] concluded that the Yucatan peninsula experienced two drought cycles of period 208 and 50 years. This hypothesis has been very influential in shaping academic discussion not only around classical Mayan collapse [66, 91] but also in dialogues involving climate change [25]. In order to test this hypothesis, we fit the following trend function to our data:

$$\mu(t) = \alpha_0 t + \boldsymbol{\alpha_1}^T f_S(2\pi t\theta_1) + \boldsymbol{\alpha_2}^T f_S(2\pi t\theta_2), \tag{6.1}$$

where $\theta_1 = 208/N$ and $\theta_2 = 50/N$ with N=range of the years in observation, and $f_S(x) = (\sin(x), \cos(x))^T$. Figure 5b shows that based on our 95% SCB, we cannot accept the trend of (6.1). Carleton [15] argued that [45, 46] used interpolation to turn the irregularly spaced data-points into a regularly spaced one before applying their methods, and the obtained periodicity might have been the superficial result of such method.

TABLE 2
Empirical coverage probabilities of SCB of X_t from Model (4.6) where $Z_t \sim$ Model 5.4 with t_6 error

		Weights: $w =$ $(0.75, \dots, -0.75, \dots, 0.75, \dots, -0.75, \dots)$				Weights: $w = \sin(8\pi t/n)$			
n	h_n	$\theta = -0.8$	$\theta = -0.4$	$\theta = 0.4$	$\theta = 0.8$	$\theta = -0.8$	$\theta = -0.4$	$\theta = 0.4$	$\theta = 0.8$
600	0.11	0.940	0.951	0.943	0.946	0.941	0.954	0.958	0.938
	0.13	0.957	0.951	0.947	0.951	0.953	0.951	0.962	0.950
	0.15	0.950	0.962	0.954	0.942	0.959	0.959	0.958	0.957
800	0.11	0.943	0.967	0.956	0.941	0.953	0.959	0.971	0.938
	0.13	0.953	0.961	0.967	0.953	0.956	0.958	0.961	0.952
	0.15	0.946	0.965	0.968	0.949	0.966	0.958	0.959	0.963



(a) (b)

FIG. 5. *(a) Plot of the residual* $X_i - \hat{\mu}_i$. *(b) 95% SCB in blue and the fitted local linear estimate in red. The fitted line (6.1) is in dashed green.*

7. Discussion. This paper develops an optimal Gaussian approximation for nonstationary univariate time series, that besides being optimal, also provides a clear instructive way as to how one can construct such approximations for practical applications. Our results match the best possible rates from other literature on nonstationary time series [9, 54–56], etc. with relaxed assumptions.

Our first result is an approximation result that preserves the population second-order properties in the approximating Gaussian analogue. Our second, and probably more practically usable result states that the approximating Gaussian process can be embedded in a Brownian motion with evolving variances. A major difficulty in constructing approximating Gaussian processes was the nonavailability of the notion of a long-run covariance, and our paper settles this question while maintaining the sharp rate. This work lays out an asymptotic framework, which can be used in many areas of nonstationary time series, such as complex nonlinear and nonstationary econometric models with smooth or abrupt changes. Moreover, one can further explore beyond just temporal dependence and wish to obtain similar results for complex spatial, spatiotemporal or tensor processes where nonstationarity is quite intrinsic.

Acknowledgments. The authors would like to thank the Associate Editor and the reviewers for their constructive feedbacks that helped improving the paper significantly.

Funding. The second author thanks NSF Grant DMS-2124222 for supporting their research.

The third author's research is partially supported by NSF Grants DMS-1916351, DMS-2027723 and DMS-2311249.

SUPPLEMENTARY MATERIAL

Supplement to "Gaussian approximation for nonstationary time series with optimal rate and explicit construction" (DOI: 10.1214/24-AOS2436SUPP; .pdf). Contains all proofs in Sections 8, 9, 10 and 11, and some additional simulation results in Section 12.

REFERENCES

[1] ADAK, S. (1998). Time-dependent spectral analysis of nonstationary time series. *J. Amer. Statist. Assoc.* **93** 1488–1501. MR1666643 https://doi.org/10.2307/2670062

[2] ANDREOU, E. and GHYSELS, E. (2009). Structural breaks in financial time series. *Handbook of Financial Time Series* 839–870.

[3] ANDREWS, A. P., ANDREWS, E. W. and CASTELLANOS, F. R. (2003). The Northern Maya collapse and its aftermath. *Ancient Mesoamerica* **14** 151–156.

[4] ANDREWS, D. W. K. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica* **61** 821–856. MR1231678 https://doi.org/10.2307/2951764

[5] AUE, A. and HORVÁTH, L. (2013). Structural breaks in time series. *J. Time Series Anal.* **34** 1–16. MR3008012 https://doi.org/10.1111/j.1467-9892.2012.00819.x

[6] BAI, J. and PERRON, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica* **66** 47–78. MR1616121 https://doi.org/10.2307/2998540

[7] BENNETT, W. R. (1958). Statistics of regenerative digital transmission. *Bell Syst. Tech. J.* **37** 1501–1542. MR0102138 https://doi.org/10.1002/j.1538-7305.1958.tb01560.x

[8] BERCU, B., GAMBOA, F. and ROUAULT, A. (1997). Large deviations for quadratic forms of stationary Gaussian processes. *Stochastic Process. Appl.* **71** 75–90. MR1480640 https://doi.org/10.1016/S0304-4149(97)00071-9

[9] BERKES, I., LIU, W. and WU, W. B. (2014). Komlós–Major–Tusnády approximation under dependence. *Ann. Probab.* **42** 794–817. MR3178474 https://doi.org/10.1214/13-AOP850

[10] BLOOMFIELD, P., HURD, H. L. and LUND, R. B. (1994). Periodic correlation in stratospheric ozone data. *J. Time Series Anal.* **15** 127–150. MR1263886 https://doi.org/10.1111/j.1467-9892.1994.tb00181.x

[11] BOROVKOV, A. (1973). Notes on inequalities for sums of independent variables. *Theory Probab. Appl.* **17** 556.

[12] BROWN, R. L., DURBIN, J. and EVANS, J. M. (1975). Techniques for testing the constancy of regression relationships over time. *J. Roy. Statist. Soc. Ser. B* **37** 149–192. MR0378310

[13] BÜHLMANN, P. (1998). Sieve bootstrap for smoothing in nonstationary time series. *Ann. Statist.* **26** 48–83. MR1611804 https://doi.org/10.1214/aos/1030563978

[14] CAI, Z. (2007). Trending time-varying coefficient time series models with serially correlated errors. *J. Econometrics* **136** 163–188. MR2328589 https://doi.org/10.1016/j.jeconom.2005.08.004

[15] CARLETON, C. (2017). Archaeological and Palaeoenvironmental Time-series Analysis. Theses, Department of Archaeology, Simon Fraser Univ.

[16] CARLSTEIN, E. (1986). The use of subseries values for estimating the variance of a general statistic from a stationary sequence. *Ann. Statist.* **14** 1171–1179. MR0856813 https://doi.org/10.1214/aos/1176350057

[17] CARLSTEIN, E., DO, K.-A., HALL, P., HESTERBERG, T. and KÜNSCH, H. R. (1998). Matched-block bootstrap for dependent data. *Bernoulli* **4** 305–328. MR1653268 https://doi.org/10.2307/3318719

[18] CHOW, G. C. (1960). Tests of equality between sets of coefficients in two linear regressions. *Econometrica* **28** 591–605. MR0141193 https://doi.org/10.2307/1910133

[19] CSÖRGŐ, M. and HORVÁTH, L. (1997). *Limit Theorems in Change-Point Analysis*. *Wiley Series in Probability and Statistics*. Wiley, Chichester. MR2743035

[20] DAHLHAUS, R. (1997). Fitting time series models to nonstationary processes. *Ann. Statist.* **25** 1–37. MR1429916 https://doi.org/10.1214/aos/1034276620

[21] DAHLHAUS, R. (2000). A likelihood approximation for locally stationary processes. *Ann. Statist.* **28** 1762–1794. MR1835040 https://doi.org/10.1214/aos/1015957480

[22] DAUBECHIES, I. (1992). *Ten Lectures on Wavelets*. *CBMS-NSF Regional Conference Series in Applied Mathematics* **61**. SIAM, Philadelphia, PA. MR1162107 https://doi.org/10.1137/1.9781611970104

[23] DAVIS, R. A., LEE, T. C. M. and RODRIGUEZ-YAM, G. A. (2006). Structural break estimation for nonstationary time series models. *J. Amer. Statist. Assoc.* **101** 223–239. MR2268041 https://doi.org/10.1198/016214505000000745

[24] DEMAREST, A. (2004). *Ancient Maya*: *The Rise and Fall of a Rainforest Civilization*. *Case Studies in Early Societies*. Cambridge Univ. Press, Cambridge.

[25] DIAZ, H. and TROUET, V. (2014). Some perspectives on societal impacts of past climatic changes. *History Compass* **12** 160–177.

[26] DONOHO, D. L. (1995). De-noising by soft-thresholding. *IEEE Trans. Inf. Theory* **41** 613–627. MR1331258 https://doi.org/10.1109/18.382009

[27] DOOB, J. L. (1949). Heuristic approach to the Kolmogorov–Smirnov theorems. *Ann. Math. Stat.* **20** 393–403. MR0030732 https://doi.org/10.1214/aoms/1177729991

[28] ELDAN, R., MIKULINCER, D. and ZHAI, A. (2020). The CLT in high dimensions: Quantitative bounds via martingale embedding. *Ann. Probab.* **48** 2494–2524. MR4152649 https://doi.org/10.1214/20-AOP1429

[29] ERDÖS, P. and KAC, M. (1946). On certain limit theorems of the theory of probability. *Bull. Amer. Math. Soc.* **52** 292–302. MR0015705 https://doi.org/10.1090/S0002-9904-1946-08560-2

[30] EUBANK, R. L. and SPECKMAN, P. L. (1993). Confidence bands in nonparametric regression. *J. Amer. Statist. Assoc.* **88** 1287–1301. MR1245362

[31] FAN, J. and GIJBELS, I. (1996). *Local Polynomial Modelling and Its Applications*. *Monographs on Statistics and Applied Probability* **66**. CRC Press, London. MR1383587

[32] FAN, J. and ZHANG, W. (1999). Statistical estimation in varying coefficient models. *Ann. Statist.* **27** 1491–1518. MR1742497 https://doi.org/10.1214/aos/1017939139
[33] FAN, J. and ZHANG, W. (2000). Simultaneous confidence bands and hypothesis testing in varying-coefficient models. *Scand. J. Stat.* **27** 715–731. MR1804172 https://doi.org/10.1111/1467-9469.00218
[34] FAUST, B. B. (2001). Maya environmental successes and failures in the Yucatan Peninsula. *Environ. Sci. Policy* **4** 153–169.
[35] FRANKS, L. E. (1969). *Signal Theory. Information Theory Series*. Prentice-Hall, New York.
[36] GARDNER, W. A. et al. (1994). *Cyclostationarity in Communications and Signal Processing* **1**. IEEE Press, New York.
[37] GILL, R. B. (2000). *The Great Maya Droughts: Water, Life, and Death*. Univ. New Mexico Press.
[38] GILL, R. B., MAYEWSKI, P. A., NYBERG, J., HAUG, G. H. and PETERSON, L. C. (2007). Drought and the Maya Collapse. *Ancient Mesoamerica* **18** 283–302.
[39] GOLDEN, C. W. and BORGSTEDE, G. (2004). *Continuities and Changes in Maya Archaeology: Perspectives at the Millennium*. Routledge, London.
[40] GUNN, J. D., MATHENY, R. T. and FOLAN, W. J. (2002). Climate-change studies in the Maya area: A diachronic analysis. *Ancient Mesoamerica* **13** 79–84.
[41] HAAR, A. (1910). Zur Theorie der orthogonalen Funktionensysteme. *Math. Ann.* **69** 331–371. MR1511592 https://doi.org/10.1007/BF01456326
[42] HALL, P. (1985). Resampling a coverage pattern. *Stochastic Process. Appl.* **20** 231–246. MR0808159 https://doi.org/10.1016/0304-4149(85)90212-1
[43] HANSON, D. L. and WRIGHT, F. T. (1971). A bound on tail probabilities for quadratic forms in independent random variables. *Ann. Math. Stat.* **42** 1079–1083. MR0279864 https://doi.org/10.1214/aoms/1177693335
[44] HÄRDLE, W. (1986). A note on jackknifing kernel regression function estimators. *IEEE Trans. Inf. Theory* **32** 298–300. MR0838421 https://doi.org/10.1109/TIT.1986.1057142
[45] HODELL, D. A., BRENNER, M. and CURTIS, J. H. (2005). Terminal classic drought in the northern Maya lowlands inferred from multiple sediment cores in Lake Chichancanab (Mexico). *Quat. Sci. Rev.* **24** 1413–1427.
[46] HODELL, D. A., BRENNER, M., CURTIS, J. H. and GUILDERSON, T. (2001). Solar forcing of drought frequency in the Maya lowlands. *Science* **292** 1367–1370. https://doi.org/10.1126/science.1057759
[47] HODELL, D. A., CURTIS, J. H. and BRENNER, M. (1995). Possible role of climate in the collapse of Classic Maya civilization. *Nature* **375** 391–394.
[48] HOOVER, D. R., RICE, J. A., WU, C. O. and YANG, L.-P. (1998). Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. *Biometrika* **85** 809–822. MR1666699 https://doi.org/10.1093/biomet/85.4.809
[49] HUANG, J. Z., WU, C. O. and ZHOU, L. (2004). Polynomial spline estimation and inference for varying coefficient models with longitudinal data. *Statist. Sinica* **14** 763–788. MR2087972
[50] JANDHYALA, V., FOTOPOULOS, S., MACNEILL, I. and LIU, P. (2013). Inference for single and multiple change-points in time series. *J. Time Series Anal.* **34** 423–446. MR3070866 https://doi.org/10.1111/jtsa.12035
[51] JOHNSTONE, I. M. and SILVERMAN, B. W. (1997). Wavelet threshold estimators for data with correlated noise. *J. Roy. Statist. Soc. Ser. B* **59** 319–351. MR1440585 https://doi.org/10.1111/1467-9868.00071
[52] KAKIZAWA, Y. (2007). Moderate deviations for quadratic forms in Gaussian stationary processes. *J. Multivariate Anal.* **98** 992–1017. MR2325456 https://doi.org/10.1016/j.jmva.2006.07.004
[53] KARMAKAR, S., RICHTER, S. and WU, W. B. (2022). Simultaneous inference for time-varying models. *J. Econometrics* **227** 408–428. MR4384679 https://doi.org/10.1016/j.jeconom.2021.03.002
[54] KARMAKAR, S. and WU, W. B. (2020). Optimal Gaussian approximation for multiple time series. *Statist. Sinica* **30** 1399–1417. MR4257539 https://doi.org/10.5705/ss.202017.0303
[55] 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. MR0375412 https://doi.org/10.1007/BF00533093
[56] 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. MR0402883 https://doi.org/10.1007/BF00532688
[57] KÜNSCH, H. R. (1989). The jackknife and the bootstrap for general stationary observations. *Ann. Statist.* **17** 1217–1241. MR1015147 https://doi.org/10.1214/aos/1176347265
[58] LAHIRI, S. N. (2003). *Resampling Methods for Dependent Data. Springer Series in Statistics*. Springer, New York. MR2001447 https://doi.org/10.1007/978-1-4757-3803-2
[59] LEYBOURNE, S. J. and MCCABE, B. P. M. (1989). On the distribution of some test statistics for coefficient constancy. *Biometrika* **76** 169–177. MR0991435 https://doi.org/10.1093/biomet/76.1.169

[60] LIN, C.-F. J. and TERÄSVIRTA, T. (1999). Testing parameter constancy in linear models against stochastic stationary parameters. *J. Econometrics* **90** 193–213. MR1703341 https://doi.org/10.1016/S0304-4076(98)00041-4

[61] LIN, D. Y. and YING, Z. (2001). Semiparametric and nonparametric regression analysis of longitudinal data. *J. Amer. Statist. Assoc.* **96** 103–126. MR1952726 https://doi.org/10.1198/016214501750333018

[62] LIU, R. Y. and SINGH, K. (1992). Moving blocks jackknife and bootstrap capture weak dependence. In *Exploring the Limits of Bootstrap* (*East Lansing*, *MI*, 1990). *Wiley Ser. Probab. Math. Statist. Probab. Math. Statist.* 225–248. Wiley, New York. MR1197787

[63] LIU, W. and LIN, Z. (2009). Strong approximation for a class of stationary processes. *Stochastic Process. Appl.* **119** 249–280. MR2485027 https://doi.org/10.1016/j.spa.2008.01.012

[64] LIU, W. and WU, W. B. (2010). Asymptotics of spectral density estimates. *Econometric Theory* **26** 1218–1245. MR2660298 https://doi.org/10.1017/S026646660999051X

[65] LU, Q., LUND, R. and LEE, T. C. M. (2010). An MDL approach to the climate segmentation problem. *Ann. Appl. Stat.* **4** 299–319. MR2758173 https://doi.org/10.1214/09-AOAS289

[66] LUCERO, L. J., GUNN, J. D. and SCARBOROUGH, V. L. (2011). Climate change and classic Maya water management. *Water* **3** 479–494.

[67] MCGONIGLE, E. T., KILLICK, R. and NUNES, M. A. (2022). Modelling time-varying first and second-order structure of time series via wavelets and differencing. *Electron. J. Stat.* **16** 4398–4448. MR4474578 https://doi.org/10.1214/22-ejs2044

[68] MIES, F. and STELAND, A. (2023). Sequential Gaussian approximation for nonstationary time series in high dimensions. *Bernoulli* **29** 3114–3140. MR4632133 https://doi.org/10.3150/22-bej1577

[69] NABEYA, S. and TANAKA, K. (1988). Asymptotic theory of a test for the constancy of regression coefficients against the random walk alternative. *Ann. Statist.* **16** 218–235. MR0924867 https://doi.org/10.1214/aos/1176350701

[70] NAGAEV, S. V. (1979). Large deviations of sums of independent random variables. *Ann. Probab.* **7** 745–789. MR0542129

[71] NAPOLITANO, A. (2016). Cyclostationarity: New trends and applications. *Signal Process.* **120** 385–408.

[72] NEWEY, W. K. and WEST, K. D. (1987). A simple, positive semidefinite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* **55** 703–708. MR0890864 https://doi.org/10.2307/1913610

[73] NYBLOM, J. (1989). Testing for the constancy of parameters over time. *J. Amer. Statist. Assoc.* **84** 223–230. MR0999682

[74] PAGE, E. S. (1954). Continuous inspection schemes. *Biometrika* **41** 100–115. MR0088850 https://doi.org/10.1093/biomet/41.1-2.100

[75] PAGE, E. S. (1955). A test for a change in a parameter occurring at an unknown point. *Biometrika* **42** 523–527. MR0072412 https://doi.org/10.1093/biomet/42.3-4.523

[76] PARZEN, E. and PAGANO, M. (1979). An approach to modeling seasonally stationary time series. *J. Econometrics* **9** 137–153.

[77] PERCIVAL, D. B. and MONDAL, D. (2012). 22—a wavelet variance primer. In *Time Series Analysis*: *Methods and Applications* (T. Subba Rao, S. Subba Rao and C. R. Rao, eds.). *Handbook of Statistics* **30** 623–657. Elsevier.

[78] PERRON, P. et al. (2006). Dealing with structural breaks. *Palgrave Handbook of Econometrics* **1** 278–352.

[79] PLOBERGER, W., KRÄMER, W. and KONTRUS, K. (1989). A new test for structural stability in the linear regression model. *J. Econometrics* **40** 307–318. MR0994952 https://doi.org/10.1016/0304-4076(89)90087-0

[80] POLITIS, D. N. and ROMANO, J. P. (1994). The stationary bootstrap. *J. Amer. Statist. Assoc.* **89** 1303–1313. MR1310224

[81] POLITIS, D. N. and ROMANO, J. P. (1995). Bias-corrected nonparametric spectral estimation. *J. Time Series Anal.* **16** 67–103. MR1323618 https://doi.org/10.1111/j.1467-9892.1995.tb00223.x

[82] POLITIS, D. N. and WHITE, H. (2004). Automatic block-length selection for the dependent bootstrap. *Econometric Rev.* **23** 53–70. MR2041534 https://doi.org/10.1081/ETC-120028836

[83] PRIESTLEY, M. B. (1981). *Spectral Analysis and Time Series*. *Probability and Mathematical Statistics*: *A Series of Monographs and Textbooks v.* 1. Academic Press.

[84] RAMSAY, J. O. and SILVERMAN, B. W. (2005). *Functional Data Analysis*, 2nd ed. *Springer Series in Statistics*. Springer, New York. MR2168993

[85] REEVES, J., CHEN, J., WANG, X. L., LUND, R. and LU, Q. Q. (2007). A review and comparison of changepoint detection techniques for climate data. *J. Appl. Meteorol. Climatol.* **46** 900–915.

[86] RICHTER, S. and DAHLHAUS, R. (2019). Cross validation for locally stationary processes. *Ann. Statist.* **47** 2145–2173. MR3953447 https://doi.org/10.1214/18-AOS1743

[87] ROBBINS, M. W., LUND, R. B., GALLAGHER, C. M. and LU, Q. (2011). Changepoints in the North Atlantic tropical cyclone record. *J. Amer. Statist. Assoc.* **106** 89–99. MR2816704 https://doi.org/10.1198/jasa.2011.ap10023

[88] RUDELSON, M. and VERSHYNIN, R. (2013). Hanson–Wright inequality and sub-Gaussian concentration. *Electron. Commun. Probab.* **18** no. 82. MR3125258 https://doi.org/10.1214/ECP.v18-2865

[89] SAKHANENKO, A. I. (2006). Estimates in the invariance principle in terms of truncated power moments. *Sibirsk. Mat. Zh.* **47** 1355–1371. MR2302850 https://doi.org/10.1007/s11202-006-0119-1

[90] STOUMBOS, Z. G., REYNOLDS JR., M. R., RYAN, T. P. and WOODALL, W. H. (2000). The state of statistical process control as we proceed into the 21st century. *J. Amer. Statist. Assoc.* **95** 992–998.

[91] TURNER, B. L. and SABLOFF, J. A. (2012). Classic period collapse of the Central Maya Lowlands: Insights about human–environment relationships for sustainability. *Proc. Natl. Acad. Sci. USA* **109** 13908–13914.

[92] VON SACHS, R. and MACGIBBON, B. (2000). Non-parametric curve estimation by wavelet thresholding with locally stationary errors. *Scand. J. Stat.* **27** 475–499. MR1795776 https://doi.org/10.1111/1467-9469.00202

[93] WRIGHT, F. T. (1973). A bound on tail probabilities for quadratic forms in independent random variables whose distributions are not necessarily symmetric. *Ann. Probab.* **1** 1068–1070. MR0353419 https://doi.org/10.1214/aop/1176996815

[94] WU, W. B. (2005). Nonlinear system theory: Another look at dependence. *Proc. Natl. Acad. Sci. USA* **102** 14150–14154. MR2172215 https://doi.org/10.1073/pnas.0506715102

[95] 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 https://doi.org/10.1111/j.1467-9868.2007.00594.x

[96] WU, W. B. and ZHOU, Z. (2011). Gaussian approximations for non-stationary multiple time series. *Statist. Sinica* **21** 1397–1413. MR2827528 https://doi.org/10.5705/ss.2008.223

[97] XIAO, H. and WU, W. B. (2012). Covariance matrix estimation for stationary time series. *Ann. Statist.* **40** 466–493. MR3014314 https://doi.org/10.1214/11-AOS967

[98] YOFFEE, N. and COWGILL, G. L. (1991). *The Collapse of Ancient States and Civilizations*. *Book Collections on Project MUSE*. Univ. Arizona Press.

[99] ZAITSEV, A. Y. (2000). Multidimensional version of a result of Sakhanenko in the invariance principle for vectors with finite exponential moments. I. *Teor. Veroyatn. Primen.* **45** 718–738. MR1968723 https://doi.org/10.1137/S0040585X97978555

[100] ZAITSEV, A. Y. (2001). Multidimensional version of a result of Sakhanenko in the invariance principle for vectors with finite exponential moments. II. *Teor. Veroyatn. Primen.* **46** 535–561. MR1978667 https://doi.org/10.1137/S0040585X97979123

[101] ZAITSEV, A. Y. (2001). Multidimensional version of a result of Sakhanenko in the invariance principle for vectors with finite exponential moments. III. *Teor. Veroyatn. Primen.* **46** 744–769. MR1971831 https://doi.org/10.1137/S0040585X97979305

[102] ZHANG, D. and WU, W. B. (2021). Convergence of covariance and spectral density estimates for high-dimensional locally stationary processes. *Ann. Statist.* **49** 233–254. MR4206676 https://doi.org/10.1214/20-AOS1954

[103] ZHANG, T. and WU, W. B. (2012). Inference of time-varying regression models. *Ann. Statist.* **40** 1376–1402. MR3015029 https://doi.org/10.1214/12-AOS1010

[104] ZHANG, T. and WU, W. B. (2015). Time-varying nonlinear regression models: Nonparametric estimation and model selection. *Ann. Statist.* **43** 741–768. MR3319142 https://doi.org/10.1214/14-AOS1299

[105] ZHANG, W., LEE, S.-Y. and SONG, X. (2002). Local polynomial fitting in semivarying coefficient model. *J. Multivariate Anal.* **82** 166–188. MR1918619 https://doi.org/10.1006/jmva.2001.2012

[106] ZHAO, Z. and LI, X. (2013). Inference for modulated stationary processes. *Bernoulli* **19** 205–227. MR3019492 https://doi.org/10.3150/11-BEJ399

[107] ZHOU, Z. (2013). Heteroscedasticity and autocorrelation robust structural change detection. *J. Amer. Statist. Assoc.* **108** 726–740. MR3174655 https://doi.org/10.1080/01621459.2013.787184

[108] 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 https://doi.org/10.1111/j.1467-9868.2010.00743.x

[109] BONNERJEE, S., KARMAKAR, S. and WU, W. B. (2024). Supplement to "Gaussian approximation for nonstationary time series with optimal rate and explicit construction." https://doi.org/10.1214/24-AOS2436SUPP