GAUSSIAN APPROXIMATION FOR NON-STATIONARY TIME SERIES WITH OPTIMAL RATE AND EXPLICIT CONSTRUCTION

By Soham Bonnerjee[1], Sayar Karmakar[2] and Wei Biao Wu[1]

[1] Department of Statistics, University of Chicago,
[2] Department of Statistics, University of Florida

Abstract: Statistical Inference for time series such as curve estimation for time-varying models or testing for existence of change-point have garnered significant attention. However, these works are restricted to the limiting assumption of independence and/or stationarity at its best. The main obstacle is that the existing optimal Gaussian approximation results for non-stationary processes only provides 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 non-stationary 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 non-stationary time series, obtains optimal rates and yet has good applicability. Building on such approximations, we also show theoretical results for change-point detection and simultaneous inference in presence of non-stationary errors. Finally we substantiate our theoretical results with simulation studies and some real data analyses.

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, non-stationary time series are more ubiquitous, and unfortunately similar Gaussian approximation tools for non-stationary process are either not sharp enough or difficult to apply. Our main goal in this paper is to establish optimal KMT-type Gaussian approximations for non-stationary time series that also provides an explicit construction strategy and thus enables doing asymptotic inference for such series.

We now discuss some motivations for theoretical development for non-stationary 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. Secondly, 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 changes a process that was varying slowly or in a more stationary way. These two approaches are captured in the literature of time-varying models and change-point analysis respectively.

The literature of time-varying model tries to address this issue by allowing model parameters to vary smoothly over time. See [35], [36], [52], [53], [65], [89], [108], [16] among others. The inference questions arise naturally while choosing a time-varying model in contrast of a time-constant one.

Keywords and phrases: Gaussian approximation, non-stationary time series, Simultaneous confidence band, change-point testing

Such hypothesis testing frameworks are discussed in [109], [110], [20], [12], [74], [63], [78], [84], [3] and [64]. Moving from pointwise inference, [114], [101], [57] discussed obtaining more challenging simultaneous confidence bands for linear and non-linear time series. Such simultaneous inference requires Gaussian approximation beyond the central limit theorem and motivates for KMT-type Gaussian approximations as spelt out in (1.1). The second approach- the analysis of change-points was originated in quality control ([79, 80]) but has since become an integral part of a wide variety of fields, among them economics ([83]), finance ([2]), climatology ([90]) and engineering ([96]). 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 [21] among others. Structural break estimation can also be viewed as a model selection problem, see [26], [70], [93]. See [5] and [54] 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 non-stationarity has generally been reflected only in the signal and not in the innovation process. This posits a fundamental challenging problem. The literature on inference for non-stationary 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 [59, 60] obtained an optimal Gaussian approximation: for $S_i := \sum_{j=1}^{i} X_j$,

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

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 this 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 non-stationary 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 non-stationary process (X_t), we view them as outputs from a physical system with the following causal representation:

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

where $(\varepsilon_t)_{t\in\mathbb{Z}}$ are i. i. d. input of this system and $(g_t)_{t\in\mathbb{Z}} : \mathbb{R}^\infty \to \mathbb{R}$ is a set of measurable functions. A Gaussian approximation for such non-stationary processes was obtained by [102], with a suboptimal rate and only for $2 < p \leq 4$. On the other hand, for inferential procedures it is important to establish an approximation for each S_i. They did provide a regularization $G_j = \sum_{i=1}^{j} \Sigma_i^{1/2} Y_i$, where $\Sigma_i = Var(\sum_{k=i}^{\infty}(\mathbb{E}(X_k|\mathcal{F}_i) - \mathbb{E}(X_k|\mathcal{F}_{i-1})))$; however, Σ_i's are not naturally estimable quantities. This result was improved upon by [58], 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, [73] used a local long-run covariance matrix as variance of the limiting Brownian motion. Their proof relies on martingale embedding strategy of [31] to bound Wasserstein distance between $S_j = n^{-1/2} \sum_{i=1}^{j} X_i$ vectors and the Gaussian approximation. Nonetheless their rate is not optimal in n for the Gaussian approximation.

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 $G_i = \sum_{j \leq i} Y_i$ 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 $Cov(X_i, X_j) = Cov(Y_i, Y_j)$ which is sufficient for $\mathbb{E}(S_i^2) = \mathbb{E}(G_i^2)$. Assumption of Gaussianity is frequently used in many different areas of statistics where, as further specification, one puts a structure on the covariance function of X_i and X_j. Our Gaussian approximation provides theoretical validation that for non-stationary 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 approximation, despite being quite natural for non-stationary processes, are rarely discussed in the literature. In particular, for a possible non-stationarity 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 non-stationary processes with unknown covariance structure, the practical implementation with this novel Gaussian approximation remains somewhat challenging. Our second set of Gaussian approximation results first embeds the approximating Gaussian process in a Brownian Motion with evolving variance and then regularizes the latter. As expected the variance 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 [73], 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 the best one can get when variance are estimated. This also means that to achieve such results, assumptions on only slightly higher than 4 moments is enough. 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 show the application of such an invariance principle through the lens of three prominent inference problems, namely the inference problem related to existence of change-point, the simultaneous confidence bands for non-stationary time series and approximating limiting distribution of 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 non-stationary 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 non-stationary 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 [58]. Crucially, we recover the sharp $n^{1/p}$ rate modulo a logarithmic factor without the lower bound assumption of block variance, which is needed in [58].
- 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 non-stationary 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 non-stationary time series and approximating 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 dataset 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 non-stationary 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 non-stationary quadratic forms. We discuss three important inference problems in Section 4. The hypothesis testing related to test existence of change-point is discussed in 4.1. Subsequently, we discuss simultaneous confidence bands for non-stationary time series, which is deferred to Subsection 4.2. Finally, the discussion on wavelet coefficient process is deferred till Section 4.3. Next, we use Section 5 to exhibit through simulations that we achieve better approximations with the regularization spelt out in theoretical results than the prototypical block-sum variance. Then we also show extensive simulation results in for the first two of the above-mentioned application and show advantage of our theory and estimates by analysing a recent archeological dataset.

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_p and C_p for constants that depend only on p and c for universal constants. These might take different values in different lines unless otherwise specified. For two positive sequences a_n and b_n, if $a_n/b_n \to 0$, write $a_n = o(b_n)$. Write $a_n \lesssim b_n$ or $a_n = O(b_n)$ if $a_n \leq Cb_n$ for all sufficiently large n and some constant $C < \infty$. Similarly for a sequence of random variables $(X_n)_{n\geq 1}$ 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 non-stationary time series, we first provide a concise introduction of similar results for the 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 $(X_i^c)_{1\leq i\leq n} =_{\mathbb{D}} (X_i)_{1\leq i\leq n}$ 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 the i.i.d. random variables, the mentioned result (1.1) by [59, 60] represented the culmination of a series of results on *strong invariance principle* starting from [32] and [30]. Subsequently, the seminal paper by Sakhanenko [95] essentially generalized the KMT-type Gaussian approximation for independent but possibly not identically distributed random variables. The following theorem follows easily from [95].

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

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

Then, there exists a Brownian Motion $\mathbb{B}$, such that the following holds

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

The readers can look into [105, 106] and [107] for a review of similar approximations for independent but possibly non-identically distributed random variables. For time series, [9] represents the optimal result for stationary processes in this direction while [58] shows an optimal existential result for non-stationary multivariate processes. However, [58] 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 obvious to conjecture that the approximating Gaussian process have the same second-order structure as that of the original non-stationary process X_t. To deal with such results, we need to characterize the dependency set-up of the wide class of the non-stationary processes we consider in (1.2). This structural premise is laid out in the next section.

2.2. *Functional dependence measure for non-stationary processes.* To deal with the dependency structure of a non-stationary process, we employ the framework of functional dependence measure introduced in Wu [100]. We will work with the representation (1.2), which is quite general and arises naturally from the idea of writing the joint distribution of $(X_1, \ldots, X_n)$ in terms of compositions of conditional quantile functions of i.i.d. uniform random variables. With this system, given $k \in \mathbb{Z}$, 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 $p > 2$, define the uniform functional dependence as follows:

$$\delta_p(k) := \sup_i (\mathbb{E}|X_i - X_{i,\{i-k\}}|^p)^{1/p}, \text{ where } X_{i,\{i-k\}} = g_i(\ldots, \varepsilon_{i-k-1}, \varepsilon'_{i-k}, \varepsilon_{i-k+1}, \ldots, \varepsilon_i) \tag{2.3}$$

is a coupled version of X_i. 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 non-stationary 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

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

Observe that $\|X_i\|_p \leq \Theta_{0,p}$. With this framework, we are able to conveniently propose conditions on temporal dependence for the non-stationary 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 $(X_i)_{i \geq 1}$ is non-stationary, 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:*

$$\text{For any fixed } a > 0, \ \max_{1 \leq i \leq n} \mathbb{E}\left(|X_i|^p \mathbb{I}_{\{|X_i|^p \geq an\}}\right) \to 0 \text{ as } n \to \infty.$$

The classical uniform integrability condition for $(|X_t|^p)_t$ is $\max_i \mathbb{E}(|X_i|^p \mathbb{I}_{\{|X_i|^p \geq k\}}) \to 0$ as $k \to \infty$. Note that Condition 2.2 is weaker. We will also require a mild non-singularity condition on the variance of the original process X_t.

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

This non-singularity condition is a very natural one. A simple counter-example 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 $(\varepsilon_t)_{t \in \mathbb{N}}$ are i.i.d 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

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

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 $(X_t)_{t\geq 1}$, 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

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

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

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

Moreover, it is also possible to get rid of the $\sqrt{\log n}$ factor in (2.9), but it requires us to impose a stronger non-singularity condition as [58].

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 $l \geq l_0$, $\min_{1\leq j\leq n-l+1} \|X_j + \ldots + X_{j+l-1}\|^2/l \geq c$.*

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 [58].

THEOREM 2.3. *Assume the process $(X_t)_{t\geq 1}$ 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

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

2.4. *Gaussian approximation with independent increments.* In addition to having a natural interpretation, the Gaussian approximations in the previous Section 2.3 also enjoys applicability when information about the covariance structure of the original process is available, such as for stationary processes [103] or processes from a defined parametric structure. However, for a general non-stationary processes, the precise correlation structure of X_t process are not estimable, 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), i.e involving Brownian Motion with independent increments, where the involved $\mathbb{E}[S_i^2]$ is estimable. The following two theorems addresses these issues and yields Gaussian approximations which allows practical implementation. Our first result is analogous to Theorem 2.2. However, in this result, we no longer require any non-singularity condition, and yet we almost recover the optimal $n^{1/p}$ rate (up to a log factor). Again, we can exactly recover the optimal rate if our Gaussian approximation involves 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 $\mathbb{B}$, such that*

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

Further, it holds that

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

Note that in Theorem 2.4, again using 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 non-singularity condition of Theorem 2.3, we are able to not only recover the optimal KMT rate of $n^{1/p}$ from using the Y_t themselves, but also to relax the decay rate.

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

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

REMARK 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. Suppose $n > 2$. Let $X_1, \ldots, X_n$ be i.i.d. with $\mathbb{P}(X_i = \pm n^{1/p}) = 1/n$ and $\mathbb{P}(X_i = \pm 1) = 1/2 - 1/n$, $1 \leq i \leq n$. Note that, condition 2.2 is violated since $\mathbb{E}[|X_i|^p \mathbb{I}\{|X_i|^p > n/2\}] = 2$. For the sake of contradiction, suppose the Gaussian approximation (2.14) holds, which implies

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

Since $\mathbb{E}(S_i^2) = i\mathbb{E}(X_1^2)$ and $\mathbb{E}(X_1^2) = 1 - 2/n + 2n^{2/p-1} \to 1$, 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 $\max_{1\leq i\leq n} |X_i| = o_{\mathbb{P}}(n^{1/p})$. However, the latter is false, since as $n \to \infty$

$$\mathbb{P}\left(\max_{1\leq i\leq n} |X_i| > \frac{n^{1/p}}{2}\right) = 1 - (1 - \frac{2}{n})^n \to 1 - e^{-2}.$$

This 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.

2.5. *Examples.* We now show some examples of non-stationary 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

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

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 processes. [1] and [112] considered the special case of multiplicative modulated stationary processes with a linear form. Define the functional dependence measures as

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

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 very general class of non-stationary 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 system in [7] and [38]. Apart from communication and signal detection, cyclostationary processes have enjoyed wide use in econometrics ([81]), atmospheric sciences ([10]) and across many other disciplines- the reader is encouraged to look into [40], [76] 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

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

Dahlhaus [23], [24] introduced locally stationary processes in terms of time-varying spectrum. [91] provided a general asymptotic theory for such processes. For further examples of see [111].

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

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

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 i-th order Volterra kernels. Then elementary calculations show that for a constant c_p depending only on p,

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

2.6. *Outline of the proof of Theorems.* Our proofs are quite involved and is 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 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, which we discuss in the following.

- **Truncation:** In the 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 non-stationarity.
- 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 non-stationary 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 [58] 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 [95]'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 [95] again to obtain the unconditional Gaussian approximation.
- **Regularization of Variance:** From the variance in terms of m-dependent processes in Step 3, 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 bad 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 [77], Priestley [88] and Liu and Wu [68] among others for a comprehensive review of research in this direction. As a special case of kernel-based estimates, blocking techniques has been particularly popular in this area. Carlstein [18] used non-overlapping blocks to consistently estimate $\mathbb{E}(S_i^2)$ for a stationary process. From a bootstrap perspective, Politis and Romano [85] uses non-overlapping blocks of random sizes to define a 'stationary bootstrap'. Using the 'flat-top kernel' methods of [86], [87] 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 [62], which combines ideas from [46], [18], [19] 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 in obtaining optimal Gaussian approximations. See [58], [102] and [66] for relevant references. Naturally, since the statements of our Theorems 2.2-2.5 does 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 Non-overlapping Block Bootstrap (NBB) method proposed in Chapter 2 of [62]. 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 $1 \le a, k, j \le \lceil n/m \rceil$,

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

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

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

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 from 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 non-stationary process is novel to the best of our knowledge. Thus we state it as a standalone result.

3.1. *A maximal quadratic large deviation bound.* Maximum quadratic large deviation bounds, commonly known as Hanson-Wright inequalities for subgaussian random variables have a long history starting from the seminal work by Hanson and Wright [47] and Wright [99], which proved concentration results for quadratic forms of i.i.d. random variables (see [94] for an extensive overview). Subsequent work by [8], [56] and others established *moderate deviation principles* for quadratic forms of stationary Gaussian processes. Moving beyond sub-Gaussianity, Xiao and Wu [103] and Zhang and Wu [111] 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 i.e., derive tail probability bounds for the maximal partial sum, and (ii) relax the stationarity assumption by providing a result for the general non-stationary processes. Our proof is quite similar to the Theorem 6.1 of [111]; 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 [75]. This, in particular, changes the treatment of a few important terms in our proof compared to that by [111]. Moreover, we also tackle the case when $2 < p \le 4$, something that is usually absent from other Hanson-Wright 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*

$$V_k = \sum_{t=(k-1)\mathcal{D}_n+1}^{(k\mathcal{D}_n)\wedge n} \sum_{1\le s\le t} a_{s,t} X_s X_t, \ \textit{for } 1 \le k \le \lceil n/\mathcal{D}_n \rceil, \tag{3.3}$$

and $Q_n = \sum_{j=1}^{\lfloor n/\mathcal{D}_n \rfloor} V_j$. *Then for* $x \geq 0$,

$$\mathbb{P}\left(\max_{1\leq k\leq \lceil n/\mathcal{D}_n\rceil}\left|\sum_{j=1}^{k}(V_j-\mathbb{E}(V_j))\right|\geq x\right)\leq \begin{cases} C_p x^{-p/2} n\mathcal{D}_n^{p/4}\mu_{p,A}^p, & 2<p\leq 4,\\ C_p x^{-p/2} n\mathcal{D}_n^{p/2-1}\mu_{p,A}^p + 4\exp\left(-C_p\frac{x^2}{n\mathcal{D}_n\mu_{4,A}^4}\right), & p>4,\end{cases} \tag{3.4}$$

where C_p *is a constant depending solely on* p.

The proof is given in Appendix Section 10.1.

Remark 2. *In view of* (2.6), $\delta_p^{\oplus}(j) \leq \delta_p(j)$ *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. *The bound in Theorem 3.1 should be contrasted with the bound obtained in Theorem 6 of [111]. In fact, our proof works for* $A > 1/2 - 1/q$ *and matches their non-uniform 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* α. *In view of our maximal inequality holding true for general non-stationary process, Theorem 3.1 is a much stronger result than commonly 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 $s = t$, $|a_{s,t}| \leq 1$ when $1 \leq |s-t| < 2m$ and 0 otherwise. Thus, taking $\mathcal{D}_n = 2m$, Theorem 3.1 implies that,

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

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

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

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 $1 \leq i \leq n$. For the bias part, we need to control the order of the cross-product terms $\mathbb{E}[B_iB_j]$ for $i \neq j$. The following lemma, whose prove 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*

$$\max_{1\leq k\leq \lfloor n/m\rfloor}|\mathbb{E}(B_kB_{k+1})| = O(1), \quad \max_{1\leq k\leq \lfloor n/m\rfloor}\sum_{i:|i-k|\geq 2}|\mathbb{E}(B_iB_k)| = O(m^{1-A}). \tag{3.7}$$

Observe that (3.7) readily yields

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

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:*

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

In particular, with $m \asymp n^{\zeta_1}$, where $\zeta_1 = \min\{1, 2-4/p\}/(1+2A)$, it follows from (3.9) and increment property of Brownian Motions that

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

where $\mathbb{P}^*$ refers to the conditional distribution after observing $X_1, \dots, X_n$.

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 best possible rate for $\max_{1\le i\le n} \left|\mathbb{W}(\mathcal{T}_i) - \mathbb{B}(\mathbb{E}(S_i^2))\right|$. 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 improve 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 sub-optimal rate, which anyway is the best for the empirical version, can we allow decay rate A to be smaller. We answer that question in the following.

Theorem 3.2. *Let $p > 2$. Assume decay condition 2.1 holds with $A > 1$. Further grant the truncated uniform integrability condition 2.2. Then, there exists a Brownian Motion $\mathbb{B}$, such that*

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

Remark 4. *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_iB_{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

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

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

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

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. [61] and [67], Zhou [113] and Mies and Steland [73]), 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 [73] analyzes a different variance for the approximating Gaussian process (defined as a local long-range variance $\sigma^2_{loc_i}$), 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 $\max_{1\le i\le n} |S_i^c - \mathbb{B}(\sigma^2_{loc_i})| = o_{\mathbb{P}}(n^{p/(3p-2)}\sqrt{\log n})$ 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 non-stationary process. 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 [101]. Let

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

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

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

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

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

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 analysis requires providing a rate of Ω_n depending on certain conditions.

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

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

We propose a CUSUM-based testing procedure with test statistic

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

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

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

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 non-stationary noise. Let $0 = t_0 < t_1 < t_2 < \ldots < t_{n-1} < t_n < t_{n+1} = 1$ be an n-length grid on $[0,1]$. Assume that the random variables $\{X_i\}_{i=1}^n$ is observed from the model

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

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 set-up, such as [33], [13] and [101]. 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 obtain the trend function from 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

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

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

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

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

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

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 assumptions of Theorem 2.4 for Z_i, there exists Brownian motion $\mathbb{B}$ such that with $Q_{h_n}(t) = \sum_{i=1}^n w_{h_n}(t,i)\mathbf{Y}_i$, where $\mathbf{Y}_i = \mathbb{B}(\mathbb{E}[S_i^2]) - \mathbb{B}(\mathbb{E}[S_{i-1}^2])$, and it holds that*

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

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 (2.13) of 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 [34], $\mathbb{E}[\hat{\mu}_{h_n}(t)] - \mu(t) = h_n^2\mu''(t)\beta + \mathrm{O}_{\mathbb{P}}\left(h_n^3 + n^{-1}h_n^{-1}\right)$. 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 $C_1 \le f(t) \le C_2$ for all t. □

4.2.1. *Bias Correction.* Using (4.10) in to construct simultaneous confidence band requires estimation of $\mu''(t)$. Following [48], 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 equivalent to using the kernel $K^*(x) = 2K(x) - K(x/\sqrt{2})/\sqrt{2}$; see [114], [101] and [57] 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

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

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, we can calculate the empirical quantile of $\max_{1\leq i\leq n} |\tilde{Q}_{h_n}(i/n)|$, denoted by $q_{1-\alpha}$. Thus, given significance level α, the simultaneous confidence level for $\mu(\cdot)$ can be constructed as

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

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 [45]. Assume the signal length is $n = 2^k$. Then the j-th level Haar Wavelet coefficients with $j \leq k$ are

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

Donoho [29] used wavelet methods to perform non-parametric 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 [55] and von Sachs and MacGibbon [98] extended the results for correlated Gaussian and locally stationary noise processes respectively. Recently, [72] 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 [82]. Here we allow the signal $(X_i)_{1\leq i\leq n}$ to be possibly non-stationary, 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

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

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,

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

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

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

In particular, it holds if $j_* \ge c \log p$ for some constant $c > 2/(p \log 2)$. Similar analysis can be performed with the more general Daubechies Wavelet filters (Daubechies [25]), which have better smoothness properties. The uniform Gaussian approximation (4.16) allows an asymptotic distributional theory for statistics based on wavelet transforms of non-stationary processes.

5. Simulation. In this section, we present the results of some simulation study for our results in Sections 2, 3 and 4. Our aim for our simulation studies are as follows. In Section 5.1, we start off by investigating the accuracy of the two kinds of Gaussian approximation in Sections 2.3 and 2.4. Then in Section 5.2, we inspect the accuracy of our bootstrap Gaussian approximations for finite sample. 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. In Section 5.3, we aim to look at analysing the performance of the CUSUM-based testing procedure for existence of change-point, as discussed in Section 4.1. Finally in Section 5.4, we explore the empirical coverage of our simultaneous confidence band procedure discussed in Section 4.2 under different settings.

5.1. *Empirical accuracy of theoretical Gaussian approximations.* We consider two models:

5.1. $X_t = \theta X_{t-1} + \varepsilon_t,\ \theta \in \{0.9, -0.9\}$.

5.2. $X_t = \theta_t X_{t-1} + \varepsilon_t$, $\theta_t = \theta$ if $t \le n/2$, $\theta_t = -\theta$ if $t > n/2$, $\theta \in \{0.9, -0.9\}$.

We will start off by considering $\varepsilon_t \overset{\text{i.i.d.}}{\sim} t_4/\sqrt{2}$ for both 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 size of $n = 100$. For each of the set-up, we will quantiles of the following three random variables:

$$U_1 := \max_{1\le i\le n} S_i,\ U_2 = \max_{1\le i\le n} \mathbb{B}(\mathbb{E}[S_i^2]),\ U_Y = \max_{1\le i\le n} \sum_{j=1}^{i} Y_i,$$

where $(Y_t)_{t=1}^n$ is a mean-zero Gaussian process with same covariance structure as $(X_t)_{t=1}^n$. The true quantiles are estimated by sample quantiles based on 1000 simulations. Figures 1 and 2 depicts the QQ-plots of U_2 and U_3 against U_1. Clearly, when compared to the Gaussian approximation involving 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 sample size, both the approximations being theoretically valid with optimal rate of convergence, their performances become equally accurate. To show this empirically, we consider two more complicated non-stationary models.

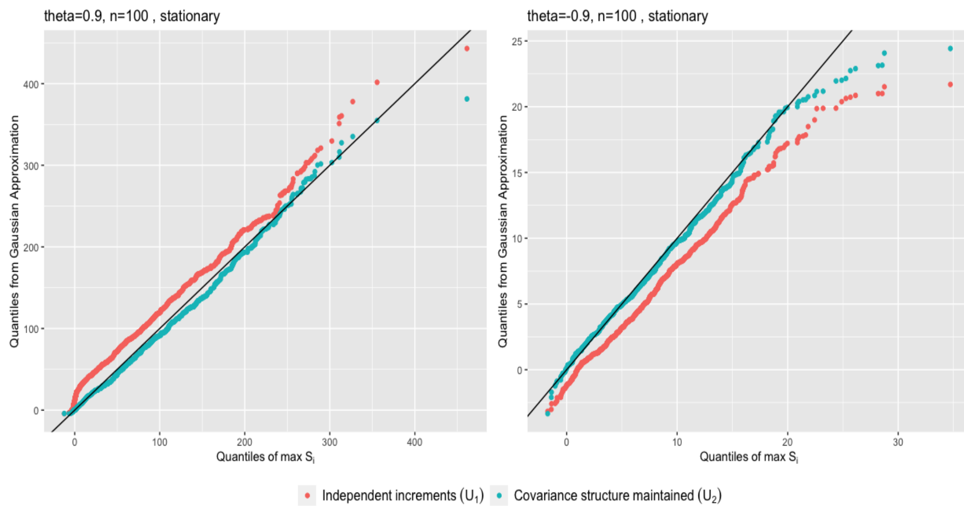


Figure 1: Comparison of theoretical quantiles with the two kinds of Gaussian approximation $X_1, \ldots, X_n \sim$ Model 5.1 with t_4 innovations: with independent increments, and with the approximation maintaining covariance structure.

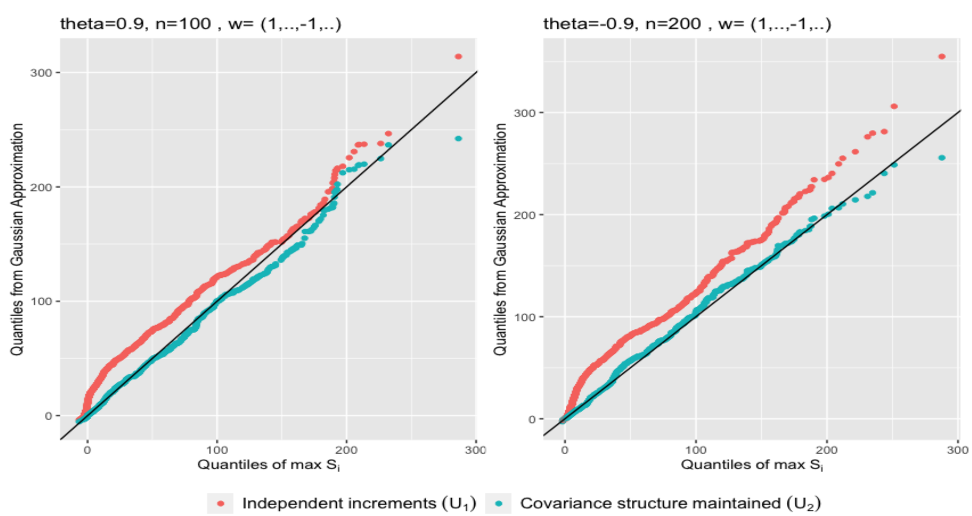


Figure 2: Comparison of theoretical quantiles with the two kinds of Gaussian approximation $X_1, \ldots, X_n \sim$ Model 5.2 with t_4 innovations: with independent increments, and with the approximation maintaining covariance structure.

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

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

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 set-ups, the more natural Gaussian approximation retains an advantage over the Gaussian approximation involving the Brownian Motion.

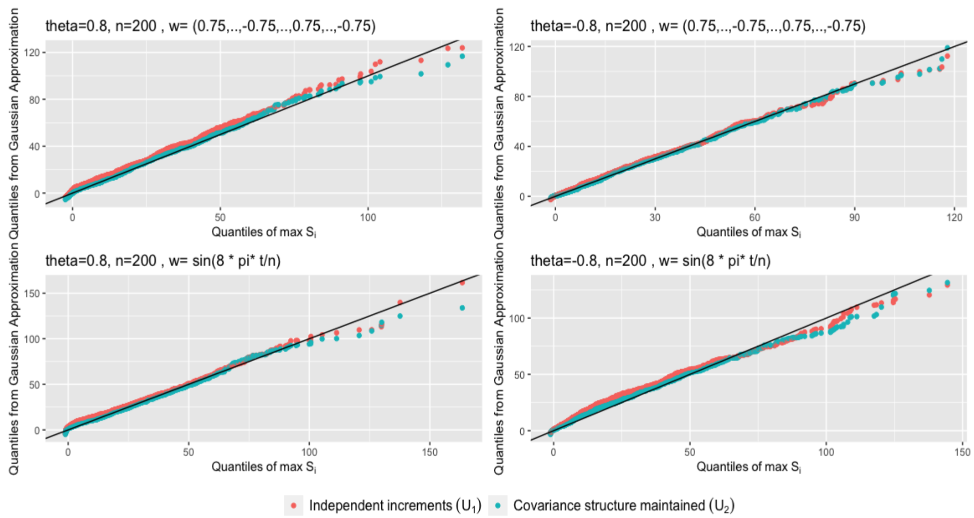


Figure 3: Comparison of theoretical quantiles with the two kinds of Gaussian approximation $X_1, \ldots, X_n \sim$ Model 5.3 with $\chi_1^2 - 1$ innovations: with independent increments, and with the approximation maintaining covariance structure.

5.2. *Empirical accuracy of the Gaussian Approximation with estimated variance.* For the strongly dependent settings with $\theta = -0.8$ and 0.8, we will explore the finite-sample accuracy of our Gaussian approximation when the variance of the Brownian motion is estimated using bootstrap. For a particular model, we take $n = 600$ and $m = \lfloor n^{1/3} \rfloor$, and simulate $B = 1000$ many samples, each of size n to estimate $U_1 := \max_{1 \le i \le n} S_i$. Next we randomly generate a data of size n from that model, and simulate B many bootstrap samples of

$$\hat{U}_2 := \max_{1 \le i \le n} \mathbb{W}(\mathcal{T}_i),\ \hat{U}_3 := \max_{1 \le i \le n} \mathbb{W}(\mathcal{T}_i^-),\ \hat{U}_4 := \max_{1 \le i \le n} \mathbb{W}(\mathcal{T}_i^\diamond),$$

where "$\hat{}$" in $\hat{U}_i$, $i = 2,3,4$ emphasize their dependence on the randomly generated data based on which bootstrap is performed. Figures 5 and 6 depict typical QQ-plots of $\hat{U}_2$, $\hat{U}_3$ and $\hat{U}_4$ against U_1 for $\varepsilon_t \sim N(0,1)$, where the "typical" is used to emphasize that $\hat{U}_i$'s are generated via bootstrap based on one typical draw of $(X_i)_{i=1}^n$ from the corresponding models. We note that as expected from our theoretical discussion, $\mathcal{T}_i$ yields much better approximation to the quantiles of $\max_{1 \le i \le n} S_i$ compared to $\mathcal{T}_i^-$ and $\mathcal{T}_i^\diamond$. More simulation results are added in Appendix Section 12.1.

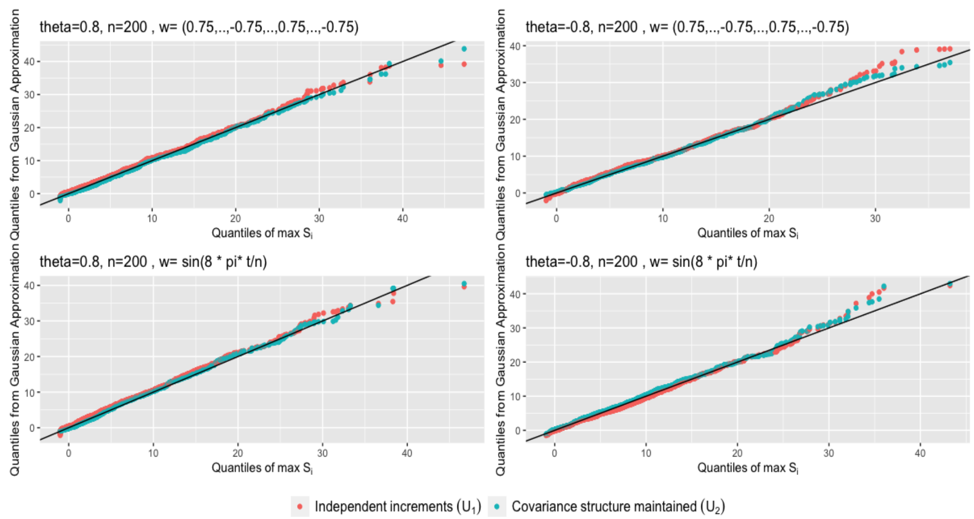


Figure 4: Comparison of theoretical quantiles with the two kinds of Gaussian approximation $X_1, \dots, X_n \sim$ Model 5.4 with $N(0,1)$ innovations: with independent increments, and with the approximation maintaining covariance structure.

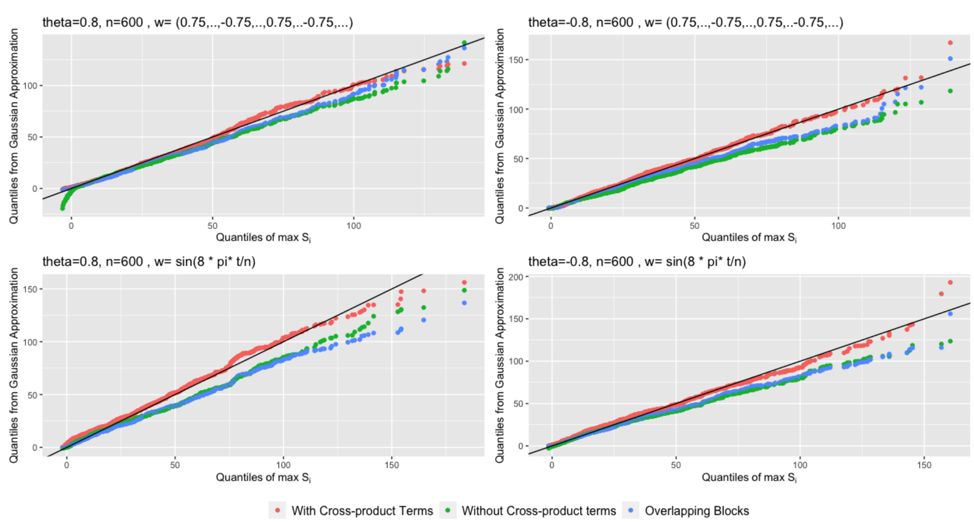


Figure 5: Comparison of theoretical quantiles with the bootstrap Gaussian approximation quantiles based on $X_1, \dots, X_n \sim$ Model 5.3 with $N(0,1)$ innovations, with and without cross-product terms.

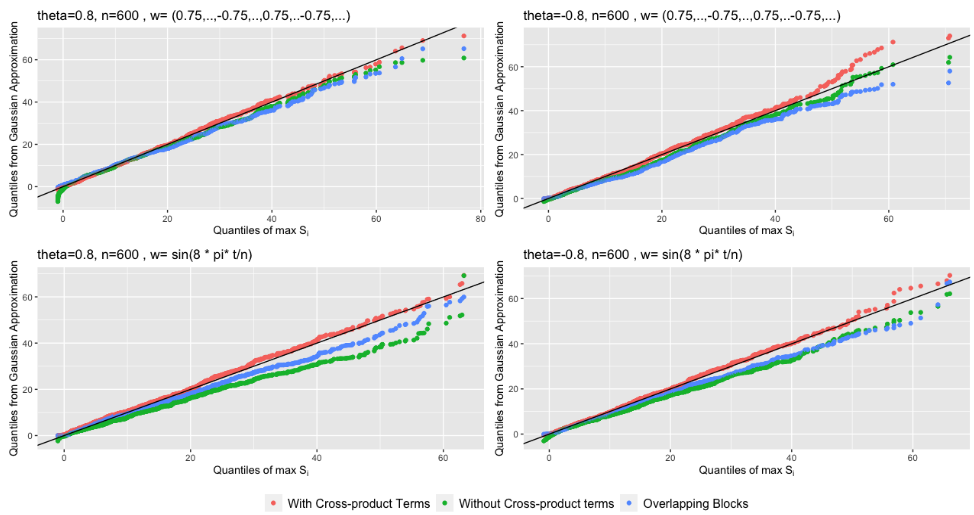


Figure 6: Comparison of theoretical quantiles with the bootstrap Gaussian approximation quantiles based on $X_1, \dots, X_n \sim$ Model 5.4 with $N(0, 1)$ innovations, with and without cross-product terms.

5.3. *Change-Point Detection.* In the simulation below, we consider the following model:

5.5. $X_t = \delta_t + \varepsilon_t \ , \varepsilon_t \sim$ Model 5.3 , $\delta_t = \delta \mathbb{I}\{t > n/2\}$.

For power calculation, we vary $\delta \in \{0.1, 0.2, \dots, 1\}$. Note that $\delta = 0$ leads to the type-1-error.

5.3.1. *Simulation based on Theoretical cut-offs.* The discussion in Section 4.1 enables us to define an approximately valid level-α test ψ_n, which we identify as an oracle test, as follows:

$$\psi_n := \mathbb{I}\{U_n > c_\alpha)\}, \text{ where } c_\alpha := \inf_r \{\mathbb{P}(V > r) \le \alpha\}, \tag{5.1}$$

and $V = \max_{i \le n} n^{-1/2} \left| \mathbb{B}(\mathbb{E}[S_i^2]) - n^{-1} i \mathbb{B}(\mathbb{E}[S_n^2]) \right|$. In this set of simulations, we theoretically calculate $\mathbb{E}[S_i^2]$ for i.i.d. innovations $\varepsilon_t \sim N(0, 1)$, and estimate c_α based on 1000 Monte-Carlo simulations of V. For each δ, power is estimated based on 1000 many samples of each size n, where we vary $n \in \{300, 600\}$. The type-I error and estimated power is shown in the Table 1. As expected, the type-I error of the test is at the nominal level, and even though ψ_n seems slightly conservative, power grows quickly as n and δ increases. Note that for $\theta = -0.8$ and 0.8, the dependence is strong, and this results in slightly lesser power compared to other cases.

	Weights : $w = 0.75, \dots, -0.75, \dots, 0.75, \dots, -0.75, \dots$								Weights: $w = \sin(8\pi t/n)$							
	$n = 300$				$n = 600$				$n = 300$				$n = 600$			
	$\theta = -0.8$	$\theta = -0.4$	$\theta = 0.4$	$\theta = 0.8$	$\theta = -0.8$	$\theta = -0.4$	$\theta = 0.4$	$\theta = 0.8$	$\theta = -0.8$	$\theta = -0.4$	$\theta = 0.4$	$\theta = 0.8$	$\theta = -0.8$	$\theta = -0.4$	$\theta = 0.4$	$\theta = 0.8$
Cutoff	2.409	1.46	1.614	2.414	2.574	1.551	1.588	2.482	2.717	1.508	1.441	2.833	2.86	1.505	1.518	2.84
Type-1 error	0.055	0.066	0.027	0.047	0.035	0.051	0.046	0.042	0.043	0.044	0.059	0.024	0.048	0.036	0.047	0.048
Power: $\delta = 0.1$	0.048	0.120	0.059	0.057	0.065	0.137	0.117	0.065	0.046	0.121	0.117	0.028	0.056	0.177	0.150	0.059
Power: $\delta = 0.2$	0.110	0.283	0.200	0.100	0.129	0.435	0.431	0.195	0.083	0.263	0.305	0.073	0.135	0.486	0.497	0.149
Power: $\delta = 0.3$	0.244	0.592	0.419	0.212	0.337	0.838	0.812	0.418	0.145	0.515	0.587	0.128	0.287	0.851	0.863	0.287
Power: $\delta = 0.4$	0.378	0.810	0.730	0.358	0.601	0.972	0.979	0.668	0.280	0.813	0.819	0.227	0.491	0.976	0.984	0.495
Power: $\delta = 0.5$	0.527	0.944	0.915	0.571	0.852	0.999	0.999	0.844	0.425	0.952	0.953	0.391	0.705	0.998	0.999	0.725
Power: $\delta = 0.6$	0.704	0.994	0.982	0.735	0.946	1	1	0.964	0.617	0.991	0.994	0.549	0.875	1	1	0.863
Power: $\delta = 0.7$	0.847	0.999	0.999	0.866	0.994	1	1	0.994	0.722	0.998	0.999	0.693	0.968	1	1	0.965
Power: $\delta = 0.8$	0.929	1	1	0.954	0.998	1	1	0.999	0.838	0.999	1	0.831	0.991	1	1	0.990
Power: $\delta = 0.9$	0.977	1	1	0.986	1	1	1	1	0.922	1	1	0.908	0.999	1	1	0.998
Power: $\delta = 1$	0.996	1	1	0.996	1	1	1	1	0.967	1	1	0.951	0.999	1	1	0.999

TABLE 1

Type-I error and power of test ψ_n for $X_1, \dots, X_n \sim$ Model 5.5.

5.3.2. *Simulation Based on Bootstrap.* Following our discussion in Section 3 as well as Section 5.2, we can estimate $\mathbb{E}[S_i^2]$ by $\mathcal{T}_i$ as in (3.1), $\mathcal{T}_i^-$ as in (3.12) and $\mathcal{T}_i^\diamond$ as in (3.15) respectively, to yield three bootstrap-based tests. We will numerically compare the efficacy of the these bootstrap procedures in approximating the CUSUM test statistics. In order to estimate the asymptotic distribution under H_0, we estimate $\mathcal{T}_i$, $\mathcal{T}_i^-$ and $\mathcal{T}_i^\diamond$ by plugging in $X_i - \hat{\mu}_i$ instead of Z_i, where $\hat{\mu}_i = \tau^{-1} \sum_{j=1}^{\tau} X_j \mathbb{I}\{i \leq \tau\} + (n-\tau)^{-1} \sum_{j=\tau+1}^{n} X_j \mathbb{I}\{i > \tau\}$, with $\tau = \operatorname{argmax}_t |\sum_{i=1}^{t} (X_t - \bar{X})|/\sqrt{n}$. Similar to Figures 5 and 6, Figures 7 and 8 depict "typical" QQ-plots of the CUSUM test statistic calculated from bootstrap samples based on $\mathcal{T}_i$, $\mathcal{T}_i^-$ and $\mathcal{T}_i^\diamond$ respectively, against the CUSUM statistic calculated from original random sample $\{X_1, \dots, X_n\}$, with $\mathcal{T}_i$ generally providing the best approximation in line with our arguments in Section 3.3.

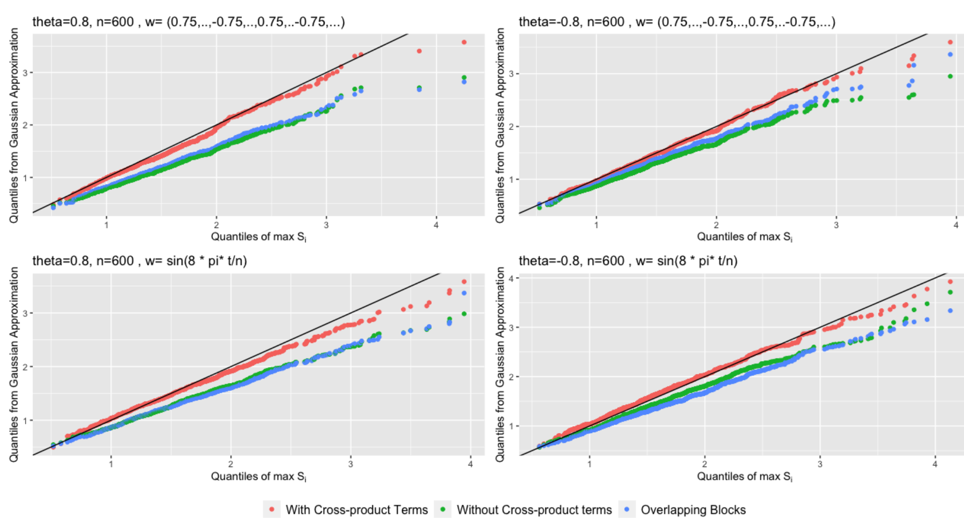


Figure 7: Comparison of theoretical quantiles of CUSUM statistic U_n with quantiles of bootstrap Gaussian approximation of CUSUM quantiles based on $X_1, \dots, X_n \sim$ Model 5.3 with $N(0,1)$ innovations, with and without cross-product terms.

5.4. *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-

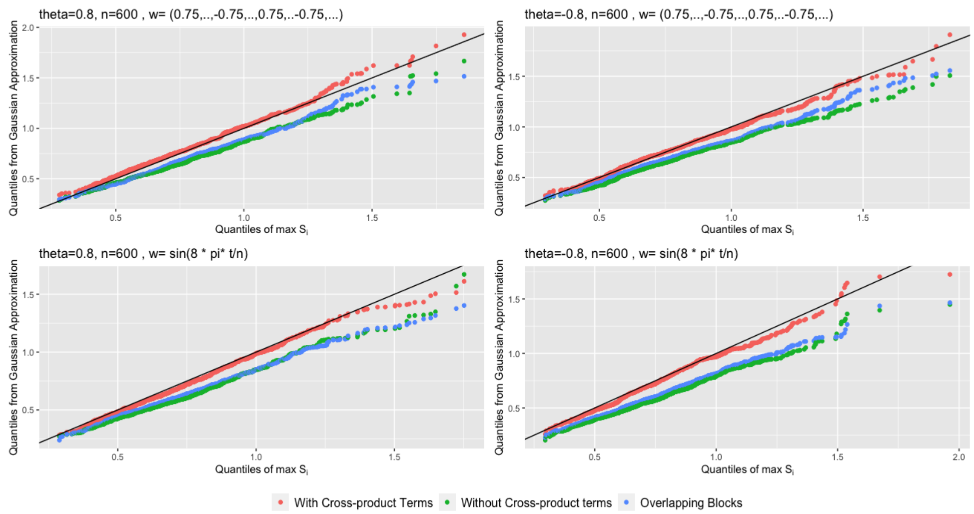


Figure 8: Comparison of theoretical quantiles of CUSUM statistic U_n with quantiles of bootstrap Gaussian approximation of CUSUM quantiles based on $X_1, \ldots, X_n \sim$ Model 5.4 with $N(0,1)$ innovations, with and without cross-product terms.

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 $i = 1, \ldots, n$. 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 $K(x) = \frac{3}{4}(1-x^2)\mathbb{I}\{|x| \leq 1\}$. For each of these model, 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 will 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 [34], the variance of local linear estimator is comparatively high on the boundary points, which affects coverage. Thus, we will report as empirical coverage the percentage of times the estimated SCB contains the true $\mu(t)$ curve in the interval $[0.05, 0.95]$. Generally speaking, the coverage probabilities in Tables 2 and 3 are reasonably close to the nominal level 0.95. Moreover, the bandwidths does not seem to have too large an effect on the coverage probability.

		Weights : $w = (0.75, \ldots, -0.75, \ldots, 0.75, \ldots, -0.75, \ldots)$				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

TABLE 2

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 : $w = (0.75, \ldots, -0.75, \ldots, 0.75, \ldots, -0.75, \ldots)$				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

TABLE 3

Empirical Coverage Probabilities of SCB of X_t from Model (4.6) where $Z_t \sim$ Model 5.4 with t_6 error.

6. Real Data Application.

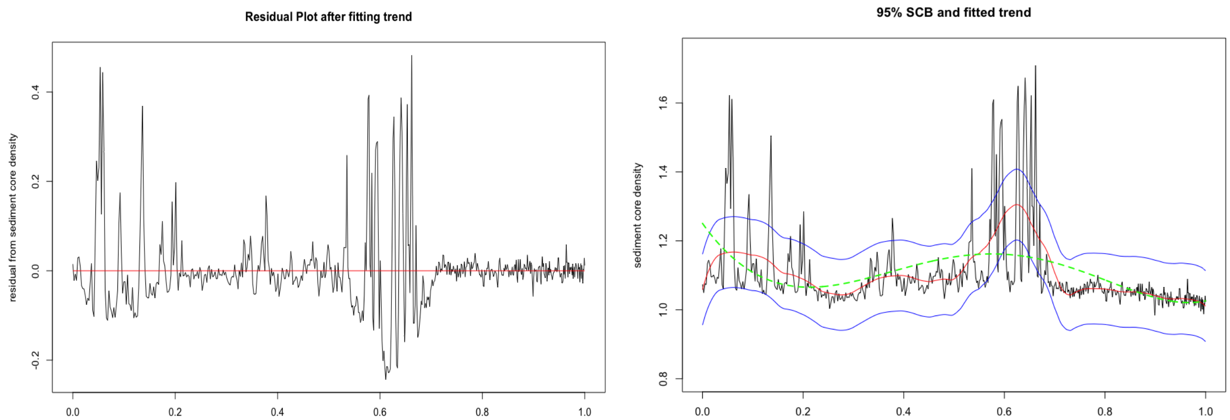
6.1. *Lake Chichancanab Sediment Density data.* The Maya civilization, arguably one of the most important pre-Columbian mesoamerican civilization, underwent a collapse during the last classical period of their history, circa 900-1100 AD ([4], [27], [42], [104]). A severe drought has been hinted at as a primary reason behind this collapse ([37], [44], [97]), despite the Mayans primarily inhabiting a seasonally dry tropical forest ([43]). Drought has also been explored as a possible cause of comparatively less-studied preclassical Maya collapse in 150-200 AD ([41]). [50, 51, 49] analyzed the sediment core density dataset from the Lake Chichancanab in the Yucatan peninsula to analyze the onset pattern of droughts during the Maya civilization. Each sediment is from a particular depth underneath the soil surface of the lake, whose calender age has been taken as the point estimate based on applying an age-depth model to radiocarbon dating. The total number of data points is $n = 564$, and the years corresponding to the observations range from -858 BC to 1994 AD.

We first test the existence of a change-point for this dataset as described in subsection 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 non-existence of a change-point. [41] 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 9b), we select $h = 0.1$. The residual plots 9a of $X_i - \hat{\mu}_L(t_i)$ where $\hat{\mu}_L$ is the locally linear estimate, suggests that the error process is indeed non-stationary. [49] 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 ([71], [97]) but also in dialogues involving climate change ([28]). 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 9b shows that based on our 95% SCB, we cannot accept the trend of (6.1). [17] argued that [51, 49] 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.

7. Discussion. This paper develops an optimal Gaussian approximation for non-stationary univariate time series, that besides being optimal, also provides a clear instructive way as to how



(a) **(b)**

Figure 9: (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.

one can construct such approximations for practical applications. Our results match the best possible rates from other literature on non-stationary time series [59, 60, 9, 58] 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 for constructing approximating Gaussian processes was the non-availability of the notion of a long-run covariance and our paper settles this question while maintaining the sharp rate. This work lays an asymptotic framework, which can be used in many areas of non-stationary time series such as complex non-linear and non-stationary econometric models with smooth or abrupt changes. Moreover, one can further explore beyond the scope of just temporal dependence and wish to obtain similar results for complex spatial, spatio-temporal or tensor processes where non-stationarity is quite intrinsic.

Acknowledgments. The authors would like to thank the Associate Editor and the reviewers for their constructive comments that helped to improve the paper.

References.

[1] Adak, S. (1998). Time-dependent spectral analysis of nonstationary time series. *J. Amer. Statist. Assoc.* **93** 1488–1501. MR1666643

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

[3] Andrews, D. W. K. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica* **61** 821–856. MR1231678

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

[5] Aue, A. and Horváth, L. (2013). Structural breaks in time series. *J. Time Series Anal.* **34** 1–16. MR3008012

[6] Bai, J. and Perron, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica* **66** 47–78. MR1616121

[7] Bennett, W. R. (1958). Statistics of regenerative digital transmission. *Bell System Tech. J.* **37** 1501–1542. MR102138

[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

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

[10] Bloomfield, P., Hurd, H. L. and Lund, R. B. (1994). Periodic correlation in stratospheric ozone data. *J. Time Ser. Anal.* **15** 127–150. MR1263886

[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

[14] Burkholder, D. L. (1973). Distribution function inequalities for martingales. *Ann. Probab.* **1** 19–42. MR365692

[15] Burkholder, D. L. (1988). Sharp inequalities for martingales and stochastic integrals. *Astérisque* **157-158** 75–94. Colloque Paul Lévy sur les Processus Stochastiques (Palaiseau, 1987). MR976214

[16] Cai, Z. (2007). Trending time-varying coefficient time series models with serially correlated errors. *J. Econometrics* **136** 163–188. MR2328589

[17] Carleton, C. (2017). *Archaeological and Palaeoenvironmental Time-series Analysis. Theses (Department of Archaeology)*. Simon Fraser University.

[18] 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. MR856813

[19] 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

[20] Chow, G. C. (1960). Tests of equality between sets of coefficients in two linear regressions. *Econometrica* **28** 591–605. MR0141193

[21] Csörgő, M. and Horváth, L. (1997). *Limit theorems in change-point analysis. Wiley Series in Probability and Statistics*. John Wiley & Sons, Ltd., Chichester With a foreword by David Kendall. MR2743035

[22] Cuny, C., Dedecker, J. and Merlevède, F. (2018). An alternative to the coupling of Berkes-Liu-Wu for strong approximations. *Chaos Solitons Fractals* **106** 233–242. MR3740091

[23] Dahlhaus, R. (1997). Fitting time series models to nonstationary processes. *Ann. Statist.* **25** 1–37. MR1429916

[24] Dahlhaus, R. (2000). A likelihood approximation for locally stationary processes. *Ann. Statist.* **28** 1762–1794. MR1835040

[25] Daubechies, I. (1992). *Ten Lectures on Wavelets*. Society for Industrial and Applied Mathematics, USA.

[26] 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

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

[28] Diaz, H. and Trouet, V. (2014). Some Perspectives on Societal Impacts of Past Climatic Changes. *History Compass* **12** 160-177.

[29] Donoho, D. L. (1995). De-noising by soft-thresholding. *IEEE Trans. Inform. Theory* **41** 613–627. MR1331258

[30] Doob, J. L. (1949). Heuristic approach to the Kolmogorov-Smirnov theorems. *Ann. Math. Statistics* **20** 393–403. MR30732

[31] Eldan, R., Mikulincer, D. and Zhai, A. (2020). The CLT in high dimensions: quantitative bounds via martingale embedding. *Ann. Probab.* **48** 2494–2524. MR4152649

[32] Erdös, P. and Kac, M. (1946). On certain limit theorems of the theory of probability. *Bull. Amer. Math. Soc.* **52** 292–302. MR15705

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

[34] Fan, J. and Gijbels, I. (1996). *Local polynomial modelling and its applications. Monographs on Statistics and Applied Probability* **66**. Chapman & Hall, London. MR1383587

[35] Fan, J. and Zhang, W. (1999). Statistical estimation in varying coefficient models. *Ann. Statist.* **27** 1491–1518. MR1742497

[36] Fan, J. and Zhang, W. (2000). Simultaneous confidence bands and hypothesis testing in varying-coefficient models. *Scand. J. Statist.* **27** 715–731. MR1804172

[37] Faust, B. B. (2001). Maya environmental successes and failures in the Yucatan Peninsula. *Environmental*

Science & Policy **4** 153-169.

[38] Franks, L. E. (1969). *Signal Theory. Information theory series*. Prentice-Hall.

[39] Fuk, D. K. and Nagaev, S. V. (1971). Probability Inequalities for Sums of Independent Random Variables. *Theory of Probability & Its Applications* **16** 643-660.

[40] Gardner, W. A. et al. (1994). *Cyclostationarity in communications and signal processing* **1**. IEEE press New York.

[41] Gill, R. B. (2000). *The Great Maya Droughts: Water, Life, and Death*. University of New Mexico Press.

[42] 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.

[43] Golden, C. W. and Borgstede, G. (2004). *Continuities and Changes in Maya Archaeology: Perspectives at the Millennium*. Routledge.

[44] 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.

[45] Haar, A. (1910). Zur Theorie der orthogonalen Funktionensysteme. *Math. Ann.* **69** 331–371. MR1511592

[46] Hall, P. (1985). Resampling a coverage pattern. *Stochastic Process. Appl.* **20** 231–246. MR808159

[47] Hanson, D. L. and Wright, F. T. (1971). A bound on tail probabilities for quadratic forms in independent random variables. *Ann. Math. Statist.* **42** 1079–1083. MR279864

[48] Härdle, W. (1986). A note on jackknifing kernel regression function estimators. *IEEE Trans. Inform. Theory* **32** 298–300. MR838421

[49] 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). *Quaternary Science Reviews* **24** 1413-1427.

[50] 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.

[51] 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.

[52] 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

[53] 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

[54] 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

[55] 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

[56] Kakizawa, Y. (2007). Moderate deviations for quadratic forms in Gaussian stationary processes. *J. Multivariate Anal.* **98** 992–1017. MR2325456

[57] Karmakar, S., Richter, S. and Wu, W. B. (2022). Simultaneous inference for time-varying models. *J. Econometrics* **227** 408–428. MR4384679

[58] Karmakar, S. and Wu, W. B. (2020). Optimal Gaussian approximation for multiple time series. *Statist. Sinica* **30** 1399–1417. MR4257539

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

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

[61] Kunsch, H. R. (1989). The Jackknife and the Bootstrap for general stationary observations. *Ann. Statist.* **17** 1217 – 1241.

[62] Lahiri, S. N. (2003). *Resampling methods for dependent data. Springer Series in Statistics*. Springer-Verlag, New York. MR2001447

[63] Leybourne, S. J. and McCabe, B. P. M. (1989). On the distribution of some test statistics for coefficient constancy. *Biometrika* **76** 169–177. MR991435

[64] Lin, C.-F. J. and Teräsvirta, T. (1999). Testing parameter constancy in linear models against stochastic stationary parameters. *J. Econometrics* **90** 193–213. MR1703341

[65] Lin, D. Y. and Ying, Z. (2001). Semiparametric and nonparametric regression analysis of longitudinal data. *J. Amer. Statist. Assoc.* **96** 103–126. With comments and a rejoinder by the authors. MR1952726

[66] Liu, W. and Lin, Z. (2009). Strong approximation for a class of stationary processes. *Stochastic Process. Appl.*

119 249–280. MR2485027
[67] 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
[68] Liu, W. and Wu, W. B. (2010). Asymptotics of spectral density estimates. *Econometric Theory* **26** 1218–1245. MR2660298
[69] Liu, W., Xiao, H. and Wu, W. B. (2013). Probability and moment inequalities under dependence. *Statist. Sinica* **23** 1257–1272. MR3114713
[70] 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
[71] Lucero, L. J., Gunn, J. D. and Scarborough, V. L. (2011). Climate Change and Classic Maya Water Management. *Water* **3** 479–494.
[72] 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
[73] Mies, F. and Steland, A. (2023). Sequential Gaussian approximation for nonstationary time series in high dimensions. *Bernoulli* **29** 3114–3140. MR4632133
[74] 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. MR924867
[75] Nagaev, S. V. (1979). Large deviations of sums of independent random variables. *Ann. Probab.* **7** 745–789. MR542129
[76] Napolitano, A. (2016). Cyclostationarity: New trends and applications. *Signal Processing* **120** 385-408.
[77] Newey, W. K. and West, K. D. (1987). A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica* **55** 703–708.
[78] Nyblom, J. (1989). Testing for the constancy of parameters over time. *J. Amer. Statist. Assoc.* **84** 223–230. MR999682
[79] Page, E. S. (1954). Continuous inspection schemes. *Biometrika* **41** 100–115. MR88850
[80] Page, E. S. (1955). A test for a change in a parameter occurring at an unknown point. *Biometrika* **42** 523–527. MR72412
[81] Parzen, E. and Pagano, M. (1979). An approach to modeling seasonally stationary time series. *Journal of Econometrics* **9** 137-153.
[82] 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.
[83] Perron, P. et al. (2006). Dealing with structural breaks. *Palgrave Handbook of Econometrics* **1** 278–352.
[84] Ploberger, W., Krämer, W. and Kontrus, K. (1989). A new test for structural stability in the linear regression model. *J. Econometrics* **40** 307–318. MR994952
[85] Politis, D. N. and Romano, J. P. (1994). The stationary bootstrap. *J. Amer. Statist. Assoc.* **89** 1303–1313. MR1310224
[86] Politis, D. N. and Romano, J. P. (1995). Bias-corrected nonparametric spectral estimation. *J. Time Ser. Anal.* **16** 67–103. MR1323618
[87] Politis, D. N. and White, H. (2004). Automatic block-length selection for the dependent bootstrap. *Econometric Rev.* **23** 53–70. MR2041534
[88] Priestley, M. B. (1981). *Spectral Analysis and Time Series. Probability and mathematical statistics : A series of monographs and textbooks* **v. 1**. Academic Press.
[89] Ramsay, J. O. and Silverman, B. W. (2005). *Functional data analysis*, second ed. *Springer Series in Statistics*. Springer, New York. MR2168993
[90] 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. *Journal of Applied Meteorology and Climatology* **46** 900–915.
[91] Richter, S. and Dahlhaus, R. (2019). Cross validation for locally stationary processes. *Ann. Statist.* **47** 2145–2173. MR3953447
[92] Rio, E. (2009). Moment inequalities for sums of dependent random variables under projective conditions. *J. Theoret. Probab.* **22** 146–163. MR2472010
[93] 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
[94] Rudelson, M. and Vershynin, R. (2013). Hanson-Wright inequality and sub-Gaussian concentration. *Elec-*

tron. Commun. Probab. **18** no. 82, 9. MR3125258

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

[96] 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.

[97] Turner, B. L. and Sabloff, J. A. (2012). Classic Period collapse of the Central Maya Lowlands: Insights about human–environment relationships for sustainability. *Proceedings of the National Academy of Sciences* **109** 13908-13914.

[98] von Sachs, R. and MacGibbon, B. (2000). Non-parametric curve estimation by wavelet thresholding with locally stationary errors. *Scand. J. Statist.* **27** 475–499. MR1795776

[99] Wright, F. T. (1973). A bound on tail probabilities for quadratic forms in independent random variables whose distributions are not necessarily symmetric. *Ann. Probability* **1** 1068–1070. MR353419

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

[101] 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

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

[103] Xiao, H. and Wu, W. B. (2012). Covariance matrix estimation for stationary time series. *Ann. Statist.* **40** 466–493. MR3014314

[104] Yoffee, N. and Cowgill, G. L. (1991). *The Collapse of Ancient States and Civilizations. Book collections on Project MUSE.* University of Arizona Press.

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

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

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

[108] Zhang, W., Lee, S.-Y. and Song, X. (2002). Local polynomial fitting in semivarying coefficient model. *J. Multivariate Anal.* **82** 166–188. MR1918619

[109] Zhang, T. and Wu, W. B. (2012). Inference of time-varying regression models. *Ann. Statist.* **40** 1376–1402. MR3015029

[110] Zhang, T. and Wu, W. B. (2015). Time-varying nonlinear regression models: nonparametric estimation and model selection. *Ann. Statist.* **43** 741–768. MR3319142

[111] 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

[112] Zhao, Z. and Li, X. (2013). Inference for modulated stationary processes. *Bernoulli* **19** 205–227. MR3019492

[113] Zhou, Z. (2013). Heteroscedasticity and autocorrelation robust structural change detection. *J. Amer. Statist. Assoc.* **108** 726–740. MR3174655

[114] 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

8. Appendix A: Proofs of Theorems 2.2 and 2.4. Since both Theorems 2.2 and 2.4 require similar sets of assumptions, we will prove them together. Further, Theorem 2.4 does not require the non-singularity Condition 2.3 for $(X_t)_{t\geq 1}$. Therefore, we begin by proving this result.

8.1. *Proof of Theorem 2.4.* Recall A_0 from (2.7). Define, in the light of the form of $\Theta_{i,p}$ in Condition 2.5 with $A > A_0$,

$$L = \frac{2 - f_1 + f_2 + \sqrt{f_3 + f_2^2}}{2pf_4}, \tag{8.1}$$
$$\alpha = \frac{2 + f_1 + f_2 + \sqrt{f_3 + f_2^2}}{2 + 2p + 2A},$$

with

$$f_1(p, A) = p(3 + A),$$
$$f_2(p, A) = p^2(1 + A),$$
$$f_3(p, A) = 4 - 4p(A - 1) - p^2(7A^2 + 6A + 3) + 2p^3(A^2 - 1),$$
$$f_4(p, A) = p(A + 1)^2 - 2.$$

Specifically, with $A > A_0$, our choice of L and α satisfies the following, which will be used in our proofs:

$$\frac{1}{2} - \frac{1}{p} - \frac{LA}{2} < 0, \tag{8.2}$$
$$L\left(\frac{\alpha}{2} - 1\right) + 1 - \frac{\alpha}{p} < 0, \tag{8.3}$$
$$p < \alpha < 2(1 + p + pA)/3, \tag{8.4}$$
$$1/p - 1/\alpha + L - L(A + 1)p/\alpha = 0. \tag{8.5}$$

These equations feature crucially in our proof, enabling us to read off certain terms as $o(1)$. In particular, they are important in proving the following three results. We will employ the following lemma, which uses the uniform integrability condition to control the p-th moment of the truncated process.

LEMMA 8.1. *Assume Condition 2.2 for the sequence $(X_i)_{i\geq 1}$. Then,*

$$\sup_i \mathbb{E}(|T_{n^{1/p}}(X_i)|^\alpha) = o(n^{\alpha/p-1}).$$

PROOF. Note that, for a fixed $a > 0$, since $n \sup_i \mathbb{P}(|X_i|^p > an) \leq \frac{1}{a} \sup_i \mathbb{E}[|X_i|^p \mathbb{I}\{|X_i|^p > an\}] = o(1)$ using Condition 2.2, therefore,

$$\begin{aligned}\overline{\lim_{n\to\infty}}\, n^{1-\alpha/p} \sup_i \mathbb{E}(|T_{n^{1/p}}(X_i)|^\alpha) &= \overline{\lim_{n\to\infty}}\, n^{1-\alpha/p} \sup_i \mathbb{E}\left[|T_{n^{1/p}}(X_i)|^\alpha \left(\mathbb{I}\{|X_i|^p \leq an\} + \mathbb{I}\{|X_i|^p > an\}\right)\right] \\ &\leq a^{\alpha/p-1} \sup_i \mathbb{E}(|X_i|^p) + \overline{\lim_{n\to\infty}} \sup_i n\mathbb{P}(|X_i|^p > an) \\ &\leq a^{\alpha/p-1} \sup_i \mathbb{E}(|X_i|^p). \qquad (8.6)\end{aligned}$$

Since $\sup_i \mathbb{E}(|X_i|^p) < \infty$ and a can be chosen arbitrarily small, (8.6) completes the proof. $\square$

8.1.1. *Key Lemma.* In this section, first we provide a bound on the p-th moment of maximal partial sums.

LEMMA 8.2. *Consider Condition 2.1 for X_t from (1.2). Let $p \geq 2$. Then for $m \geq 1$ it holds that*

$$\max_a \| \max_{1 \leq k \leq m} |X_{a+1} + \ldots + X_{a+k}| \|_p \leq \frac{p}{\sqrt{p-1}} m^{1/2} \Theta_{0,p}. \tag{8.7}$$

PROOF. Let the projection operator $P_k(X) = \mathbb{E}[X|\mathcal{F}_k] - \mathbb{E}[X|\mathcal{F}_{k-1}]$, $R_k = \sum_{i=1}^{k} X_{a+i}$ and $R_{k,s} = \sum_{i=1}^{k} P_{a+i-s} X_{a+i}$, $s \geq 0$. Note that $R_k = \sum_{s=0}^{\infty} R_{k,s}$. For fixed $s \geq 0$, $(P_{a+i-s} X_{a+i})_{1 \leq i \leq m}$ form martingale differences, and therefore, Burkholder's inequality ([92], Theorem 2.1) entails that

$$\|R_{m,s}\|_p^2 = \| \sum_{i=1}^{m} P_{a+i-s} X_{a+i} \|_p^2 \leq (p-1) \sum_{i=1}^{m} \|P_{a+i-s} X_{a+i}\|_p^2 \leq (p-1) m \delta_{s,p}^2.$$

where the last assertion follows from Theorem 1 of [100] and the uniform definition of our functional dependence measure. Finally, Doob's maximal inequality implies that

$$\| \max_{1 \leq k \leq m} |R_k| \|_p \leq \sum_{s=0}^{\infty} \| \max_{1 \leq k \leq m} |R_{k,s}| \|_p \leq \sum_{s=0}^{\infty} \frac{p}{p-1} \|R_{m,s}\|_p \leq \frac{p}{\sqrt{p-1}} m^{1/2} \Theta_{0,p}, \tag{8.8}$$

which completes the proof of (8.7). $\square$

Next, we present a lemma which is one of the main ingredients of our proof. In this result, we raise the partial sums of the truncated, m-dependent processes to a power $\alpha > p$, and our specific choice of α allows us to provide a sharp upper-bound. We will use this lemma throughout our proof to infer certain quantities are $o(1)$.

LEMMA 8.3. *Assume Conditions 2.1 and 2.2, along with (8.2), (8.3), (8.4) and (8.5) for A, L and α. Let $m = \lfloor n^L \rfloor$ and let*

$$\tilde{R}_{s,t} = \tilde{X}_{s+1} + \ldots + \tilde{X}_{s+t},$$

where $\tilde{X}_i$ is as defined in (8.22). Then

$$\max_s \mathbb{E}\left[\max_{1 \leq t \leq m} |\tilde{R}_{s,t}|^{\alpha} \right] = o(m n^{\alpha/p - 1}). \tag{8.9}$$

REMARK 5. *Lemma 8.3 and its proof should be contrasted with Lemma 7.3 of [58], where one requires a sequence t_n converging slowly to zero in both the definition of m and the truncated process $X_i^{\oplus}$. In contrast, our assumption 2.2 with the help of Lemma 8.1 enables us to circumvent the need of such sequences.*

PROOF. In the following, $\lesssim$ includes constants depending on p and A, emanating from $\mu_{p,A} = O(1)$ from Condition 2.1. Let, $\delta_p^{\oplus}(\cdot)$ and $\tilde{\delta}_p(\cdot)$ denote the functional dependence measure defined for the truncated and the m-dependent processes, respectively. Since the functional dependence measure

(2.3) is defined in an uniform manner, we can ignore the $\max_s$ term and apply the Rosenthal-type bound in [69] to obtain

$$\begin{aligned}\|\max_{1\le t\le m}|\tilde{R}_{s,t}|\|_\alpha &\lesssim m^{1/2}\left[\sum_{j=1}^{m}\tilde{\delta}_2(j)+\sum_{m+1}^{\infty}\tilde{\delta}_\alpha(j)+\sup_i\|T_{n^{1/p}}(X_i)\|\right]\\ &\quad + m^{1/\alpha}\left[\sum_{j=1}^{m}j^{1/2-1/\alpha}\tilde{\delta}_\alpha(j)+\sup_i\|T_{n^{1/p}}(X_i)\|_\alpha\right]\\ &\lesssim (I+II+III+IV),\end{aligned}\tag{8.10}$$

where

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

For I, we note that $\tilde{\delta}_2(j)\le\delta_2^{\oplus}(j)\le\delta_2(j)$, and $\sup_i\|X_i\|\le\Theta_{0,2}$. Thus $I=O(m^{1/2})$, which yields, using (8.3),

$$\frac{n^{1-\alpha/p}}{m}I^\alpha=o(1).\tag{8.11}$$

For both II and III, we start by observing that

$$\begin{aligned}\Big(\tilde{\delta}_\alpha(j)\Big)^\alpha\le\big(\delta_\alpha^{\oplus}(j)\big)^\alpha&=\sup_i\mathbb{E}\left[|T_{n^{1/p}}(X_i)-T_{n^{1/p}}(X_{i,\{i-j\}})|^\alpha\right]\\ &\le n^{\alpha/p}\sup_i\mathbb{E}\left[\left|\min\left(2,\frac{|X_i-X_{i,i-j}|}{n^{1/p}}\right)\right|^\alpha\right]\\ &\le 2^\alpha n^{\alpha/p-1}\delta_p(j)^p,\end{aligned}\tag{8.12}$$

since $\min(2^\alpha,|x|^\alpha)\le 2^\alpha|x|^p$. Hence, for II we use (8.12) to get,

$$\begin{aligned}II\le 2m^{1/2}n^{1/p-1/\alpha}\sum_{j=m+1}^{\infty}\delta_p(j)^{p/\alpha}&\le 2m^{1/2}n^{1/p-1/\alpha}\sum_{l=\lfloor\log_2 m\rfloor}^{\infty}\sum_{j=2^l}^{2^{l+1}-1}\delta_p(j)^{p/\alpha}\\ &\lesssim m^{1/2}n^{1/p-1/\alpha}\sum_{l=\lfloor\log_2 m\rfloor}^{\infty}2^{l(1-p/\alpha)}\Theta_{2^l,p}^{p/\alpha}\\ &\lesssim m^{1/2}n^{1/p-1/\alpha}m^{1-p/\alpha-pA/\alpha}=O(m^{1/2})\end{aligned}$$

using (8.5). Thus (8.3) leads to,

$$\frac{n^{1-\alpha/p}}{m} II^{\alpha} = o(1). \tag{8.13}$$

In light of (8.12), for III, we proceed as following

$$\begin{aligned}\sum_{j=1}^{m} j^{1/2-1/\alpha}\delta_p(j)^{p/\alpha} &\le \sum_{l=1}^{[\log_2(m)]} \sum_{j=2^l}^{2^{l+1}-1} j^{1/2-1/\alpha}\delta_p(j)^{p/\alpha} \\ &\lesssim \sum_{l=1}^{[\log_2(m)]} 2^{l(3/2-1/\alpha-p/\alpha)}\Theta_{2^l,p}^{p/\alpha} = O(1) \quad \text{(Using (8.4))}.\end{aligned} \tag{8.14}$$

Fix J. Using $\tilde{\delta}_\alpha(j)^\alpha \le \delta_\alpha^{\oplus}(j)^\alpha \le c_\alpha \sup_i \mathbb{E}[|T_{n^{1/p}}(X_i)|^\alpha]$ in conjunction with Lemma 8.1 yields

$$n^{1-\alpha/p}(j^{1/2-1/\alpha}\tilde{\delta}_\alpha(j))^\alpha = o(1) \tag{8.15}$$

for each $1 \le j \le J$. Thus, using (8.12) along with (8.14) and (8.15),

$$\begin{aligned}\overline{\lim_{n\to\infty}} \frac{n^{1-\alpha/p}}{m} III^{\alpha} &= \overline{\lim_{n\to 0}}\, n^{1-\alpha/p}\left(\sum_{j=1}^{\infty} j^{1/2-1/\alpha}\tilde{\delta}_\alpha(j)\right)^{\alpha} \\ &\le \overline{\lim_{n\to\infty}}\, c_\alpha n^{1-\alpha/p}\left(\sum_{j=1}^{J} j^{1/2-1/\alpha}\tilde{\delta}_\alpha(j)\right)^{\alpha} + c_\alpha\left(\sum_{j=J+1}^{\infty} j^{1/2-1/\alpha}\delta_p(j)^{p/\alpha}\right)^{\alpha} \\ &= c_\alpha\left(\sum_{j=J+1}^{\infty} j^{1/2-1/\alpha}\delta_p(j)^{p/\alpha}\right)^{\alpha},\end{aligned}$$

which, in view of (8.14), implies

$$\overline{\lim_{n\to\infty}} \frac{n^{1/\alpha-1/p}}{m} III^{\alpha} \le c_\alpha \overline{\lim_{J\to\infty}}\left(\sum_{j=J+1}^{\infty} j^{1/2-1/\alpha}\delta_p(j)^{p/\alpha}\right)^{\alpha} = 0. \tag{8.16}$$

Finally, for IV, using Lemma 8.1, one obtains

$$\frac{n^{1-\alpha/p}}{m} IV^{\alpha} = n^{1-\alpha/p}\sup_i \|T_{n^{1/p}}(X_i)\|_\alpha^\alpha = o(1). \tag{8.17}$$

The proof is now completed combining (8.11), (8.13), (8.16) and (8.17).

□

Now we are ready to prove our main Gaussian approximation result.

Proof of Theorem 2.4. The proof can be divided broadly in seven steps, which we discuss in the following sections.

8.1.2. *Truncation.* Recall $S_i^{\oplus}$ from (2.6). In this section we derive a result showing the effectiveness of the truncated partial sum process in optimally approximating the original partial sum process $(S_i)_{i\geq 1}$.

Proposition 8.1. *Under the conditions of Theorem 2.4,* $\max_{1\leq i\leq n}|S_i - S_i^{\oplus}| = o_{\mathbb{P}}(n^{1/p})$.

Proof. Note that by the truncated uniform integrability Condition 2.2,

$$\max_{1\leq j\leq n}\mathbb{P}(|X_j| > n^{1/p}) \leq \frac{1}{n}\max_{1\leq j\leq n}\mathbb{E}\left(|X_j|^p\mathbb{I}_{|X_j|\geq n^{1/p}}\right) = o(n^{-1}). \tag{8.18}$$

Thus,

$$\mathbb{P}(\max_{1\leq i\leq n}|S_i - \sum_{j=1}^{i}X_i^{\oplus}| > 0) \leq \mathbb{P}(\max_{1\leq j\leq n}|X_j| > n^{1/p}) \leq n\max_{1\leq j\leq n}\mathbb{P}(|X_j| > n^{1/p}) \to 0.$$

Hence,

$$\max_{1\leq i\leq n}|S_i - \sum_{j=1}^{i}X_i^{\oplus}| = o_{\mathbb{P}}(1). \tag{8.19}$$

Next, note that by (8.18),

$$X_i - X_i^{\oplus} = (X_i - n^{1/p})\mathbb{I}\{X_i > n^{1/p}\} + (X_i + n^{1/p})\mathbb{I}\{X_i < -n^{1/p}\}.$$

This immediately implies

$$|X_i - X_i^{\oplus}| \leq |X_i|\mathbb{I}\{|X_i| > n^{1/p}\} \leq n^{1/p-1}|X_i|^p\}\mathbb{I}\{|X_i| > n^{1/p}\},$$

which, upon invoking Condition 2.2, yields

$$\max_{1\leq i\leq n}|\mathbb{E}(X_i - X_i^{\oplus})| = o(n^{\frac{1}{p}-1}). \tag{8.20}$$

Therefore,

$$\max_{1\leq i\leq n}|\mathbb{E}(S_i - \sum_{j=1}^{i}X_j^{\oplus})| \leq \sum_{j=1}^{n}|\mathbb{E}(X_j - X_j^{\oplus})| \leq n\max_{1\leq j\leq n}|\mathbb{E}(X_j - X_j^{\oplus})| = o(n^{1/p}), \tag{8.21}$$

which by (8.19) and (8.21) completes the proof. □

8.1.3. *m-dependence.* m-dependence approximation is a useful tool which allows us to handle the truncated process in terms of the innovations ε_i. This technique has been studied extensively in the literature; see for example [66] and [9]. For a suitably chosen m, one looks at the conditional mean $\mathbb{E}(X_i|\varepsilon_i,\ldots,\varepsilon_{i-m})$. More formally, let $m = \lfloor n^L\rfloor$. Define the m-dependent version of the truncated process $X_i^{\oplus}$ as

$$\tilde{X}_j = \mathbb{E}(X_j^{\oplus}|\varepsilon_j,\ldots,\varepsilon_{j-m}) - \mathbb{E}(X_j^{\oplus}). \tag{8.22}$$

Let $\tilde{S}_i = \sum_{j=1}^{i}\tilde{X}_j$.

PROPOSITION 8.2. *Under the conditions of Theorem 2.4,* $\max_{1\le i\le n}|S_i^{\oplus}-\tilde{S}_i| = o_{\mathbb{P}}(n^{1/p})$.

PROOF. Using Lemma A1 of [66] (although the lemma is for stationary random variables, however the proof can be verified to be readily applicable to the non-stationary case) and (8.2), we have

$$\|\max_{1\le i\le n}|S_i^{\oplus}-\tilde{S}_i|\|_p \le c_p n^{1/2}\Theta_{1+m,p} = o(n^{1/p}), \tag{8.23}$$

which completes the proof. □

8.1.4. *Blocking.* We will form blocks of sum of m-dependent processes obtained above. Such blocking will make it easier to control the dependency structure of our process, resulting in optimal error bounds. Let $m\in\mathbb{N}$, and for $1\le i\le n$, denote

$$l_i = \left\lceil \frac{\lceil i/m\rceil}{3}\right\rceil. \tag{8.24}$$

For $k=1,\ldots,\lceil n/m\rceil$, consider blocks of m-dependent processes

$$\tilde{B}_k = \sum_{j=(k-1)m+1}^{km\wedge n}\tilde{X}_j,$$

with $\tilde{B}_j := 0$ if $j\notin\{1,\ldots,l_n\}$. Similarly for our original process we will define blocks

$$B_k = \sum_{j=(k-1)m+1}^{km\wedge n} X_j. \tag{8.25}$$

For the blocking approximation we approximate the partial sum process $\tilde{S}_i$ by

$$S_i^{\diamond} = \sum_{l=1}^{l_i}\sum_{k=3l-2}^{3l\wedge\lceil n/m\rceil}\tilde{B}_k.$$

The following proposition justifies the blocking approximation.

PROPOSITION 8.3. *Under conditions of Theorem 2.4,* $\max_{1\le i\le n}|\tilde{S}_i - S_i^{\diamond}| = o_{\mathbb{P}}(n^{1/p})$.

PROOF. For $1\le k\le\lceil n/m\rceil$, let $\tilde{S}_{k,l}=\sum_{j=k+1}^{l}\tilde{X}_j$. Note that, for $\varepsilon>0$,

$$\begin{aligned}
\mathbb{P}(\max_{1\le i\le n}|\tilde{S}_i - S_i^{\diamond}| > n^{1/p}\varepsilon) &\le l_n \max_{1\le l\le l_n}\mathbb{P}\left(\max_{3lm\le j\le 3(l+1)m}|\tilde{S}_{3lm,j}| > n^{1/p}\varepsilon\right)\\
&\le C\frac{n}{3m}\max_{1\le l\le k}\frac{\mathbb{E}(\max_{3lm\le j\le 3(l+1)m}|\tilde{S}_{3lm,j}|^{\alpha})}{\varepsilon^{\alpha}n^{\alpha/p}}\\
&= o(1). \text{ (By Lemma 8.3)}
\end{aligned}$$

Therefore, $\max_{1\le i\le n}|\tilde{S}_i - S_i^{\diamond}| = o_{\mathbb{P}}(n^{1/p})$. □

8.1.5. *Conditional Gaussian Approximation.* The blocking step in Section 8.1.4 yields us m-dependent blocks. The dependence between these blocks is induced by the shared innovations ε_i's along the border. In this subsection, we condition on these shared ε_i's and apply Theorem 1 of [95] to obtain a conditional Gaussian approximation.

In order to properly explain the conditioning argument, we require some notations. Let $\boldsymbol{\eta} = (\ldots, \boldsymbol{\eta}_0, \boldsymbol{\eta}_3, \ldots)$, where $\boldsymbol{\eta}_k = (\varepsilon_{(k-1)m+1}, \ldots, \varepsilon_{km})$. We will use an argument conditioning via $\boldsymbol{\eta}$. To facilitate such arguments, denote by $\boldsymbol{a}$ an arbitrary deterministic sequence $(\ldots, \boldsymbol{a}_0, \boldsymbol{a}_3, \ldots)$ with $\boldsymbol{a}_k = (a_{(k-1)m+1}, \ldots, a_{km})$.

Let $(g_i)_{i\in\mathbb{N}}$ be measurable functions such that $\tilde{X}_i = g_i(\varepsilon_{i-m}, \ldots, \varepsilon_i)$. Recall l_n from (8.24). For $1 \le k \le l_n$, define the random functions,

$$\begin{aligned}
\tilde{B}_{3k-2}(\boldsymbol{a}_{3k-3}) &= \sum_{i=(3k-3)m+1}^{(3k-2)m} g_i(a_{i-m}, \ldots, a_{(3k-3)m}, \varepsilon_{(3k-3)m+1}, \ldots, \varepsilon_i), \\
\tilde{B}_{3k-1} &= \sum_{i=(3k-2)m+1}^{(3k-1)m} g_i(\varepsilon_{i-m}, \ldots, \varepsilon_{(3k-2)m}, \ldots, \varepsilon_i), \\
\tilde{B}_{3k}(\boldsymbol{a}_{3k}) &= \sum_{i=(3k-1)m+1}^{3km} g_i(\varepsilon_{i-m}, \ldots, \varepsilon_{(3k-1)m}, a_{(3k-1)m+1}, \ldots, a_i).
\end{aligned}$$

For $1 \le l \le l_n$, let

$$\begin{aligned}
M_{3l}(\boldsymbol{a}_{3l}) &= \mathbb{E}(\tilde{B}_{3l}(\boldsymbol{a}_{3l})), \\
M_{3l-2}(\boldsymbol{a}_{3l-3}) &= \mathbb{E}(\tilde{B}_{3l-2}(\boldsymbol{a}_{3l-3})), \text{ and} \\
Y_l^{\boldsymbol{a}} := Y_l(\boldsymbol{a}_{3l-3}, \boldsymbol{a}_{3l}) &= \tilde{B}_{3l-2}(\boldsymbol{a}_{3l-3}) - M_{3l-2}(\boldsymbol{a}_{3l-3}) + \tilde{B}_{3l-1} + \tilde{B}_{3l}(\boldsymbol{a}_{3l}) - M_{3l}(\boldsymbol{a}_{3l}).
\end{aligned} \tag{8.26}$$

In the rest of this sub-section, unless otherwise specified, we will treat $\boldsymbol{a}$ as fixed. Note that in $Y_l^{\boldsymbol{a}}$'s, we have combined three blocks together to combine an "outer" layer of blocks. Further, due to our conditioning (by $\boldsymbol{\eta} = \boldsymbol{a}$), $Y_l^{\boldsymbol{a}}$'s are independent. The corresponding mean and variance functionals, for $1 \le k \le l_n$, are respectively,

$$M_k(\boldsymbol{a}) = \sum_{l=1}^{k} [M_{3l-2}(\boldsymbol{a}_{3l-3}) + M_{3l}(\boldsymbol{a}_{3l})], \tag{8.27}$$

$$Q_k(\boldsymbol{a}) = \sum_{l=1}^{k} V_l(\boldsymbol{a}_{3l-3}, \boldsymbol{a}_{3l}), \tag{8.28}$$

where $V_l(\boldsymbol{a}_{3l-3}, \boldsymbol{a}_{3l})$ is the variance of $Y_l(\boldsymbol{a}_{3l-3}, \boldsymbol{a}_{3l})$. Let $C_{3l-1}(\boldsymbol{\eta}_{3l-2}) = \mathbb{E}[\tilde{B}_{3l-1} | \boldsymbol{\eta}_{3l-2}]$. We will decompose V_l as follows:

$$\begin{aligned}
&V_l(\boldsymbol{a}_{3l-3}, \boldsymbol{a}_{3l}) \\
&= \mathbb{E}(Y_l(\boldsymbol{a}_{3l-3}, \boldsymbol{a}_{3l})^2) \\
&= \mathbb{E}\left[(\mathbb{E}[Y_l(\boldsymbol{a}_{3l-3}, \boldsymbol{a}_{3l})|\boldsymbol{\eta}_{3l-2}, \boldsymbol{\eta}_{3l-1}] - \mathbb{E}[Y_l(\boldsymbol{a}_{3l-3}, \boldsymbol{a}_{3l})|\boldsymbol{\eta}_{3l-2}])^2\right] + \mathbb{E}\left[(\mathbb{E}[Y_l(\boldsymbol{a}_{3l-3}, \boldsymbol{a}_{3l})|\boldsymbol{\eta}_{3l-2}])^2\right] \\
&= \mathbb{E}\left[\left(\tilde{B}_{3l-1} - C_{3l-1}(\boldsymbol{\eta}_{3l-2}) + \tilde{B}_{3l}(\boldsymbol{a}_{3l}) - M_{3l}(\boldsymbol{a}_{3l})\right)^2\right] \\
&\qquad\qquad + \mathbb{E}\left[\left(\tilde{B}_{3l-2}(\boldsymbol{a}_{3l-3}) - M_{3l-2}(\boldsymbol{a}_{3l-3}) + C_{3l-1}(\boldsymbol{\eta}_{3l-2})\right)^2\right] \\
&:= \tilde{V}_{2l}(\boldsymbol{a}_{3l}) + \tilde{V}_{2l-1}(\boldsymbol{a}_{3l-3}).
\end{aligned}$$

Let $Q_{l_i}(\boldsymbol{a}) = V_k(\boldsymbol{a}_{3l_i}, \boldsymbol{a}_{3l_i+3})$. Further, let

$$V_l^0(\boldsymbol{a}_{3l}) = \tilde{V}_{2l}(\boldsymbol{a}_{3l}) + \tilde{V}_{2l+1}(\boldsymbol{a}_{3l}). \tag{8.29}$$

Then, for all $t \in \mathbb{N}$,

$$\sum_{l=1}^{t} V_l(\boldsymbol{a}_{3l-3}, \boldsymbol{a}_{3l}) = \tilde{V}_1(\boldsymbol{a}_0) + \sum_{l=1}^{t-1} V_l^0(\boldsymbol{a}_{3l}) + \tilde{V}_{2t}(\boldsymbol{a}_{3t}). \tag{8.30}$$

Let, for $\alpha > p$,

$$\begin{aligned}
L_\alpha(\boldsymbol{a}, x) &= \sum_{l=1}^{l_n} \mathbb{E}\min\{|Y_l(\boldsymbol{a}_{3l-3}, \boldsymbol{a}_{3l})/x|^\alpha, |Y_l(\boldsymbol{a}_{3l-3}, \boldsymbol{a}_{3l})/x|^2\} \\
&\le \sum_{l=1}^{l_n} \mathbb{E}\left[|Y_l(\boldsymbol{a}_{3l-3}, \boldsymbol{a}_{3l})/x|^\alpha\right].
\end{aligned}$$

Then by Theorem 1 of [95], there exists a probability space $(\Omega_{\boldsymbol{a}}, \mathcal{A}_{\boldsymbol{a}}, \mathbf{P}_{\boldsymbol{a}})$ where we can define random variables $(\mathcal{R}_l(\boldsymbol{a}))_{1\le l\le l_n} =_{\mathbb{D}} (Y_l(\boldsymbol{a}_{3l-3}, \boldsymbol{a}_{3l}))_{1\le l\le l_n}$, and Brownian Motion $\mathbf{B}_{\boldsymbol{a}}$, such that

$$\mathbb{P}\left(\max_{1\le i\le n} |\Gamma_i(\boldsymbol{a}) - \mathbf{B}_a(Q_{l_i}(\boldsymbol{a}))| \ge c\alpha x\right) \le L_\alpha(\boldsymbol{a}, x), \tag{8.31}$$

where $c > 0$ is an absolute constant and $\Gamma_i(\boldsymbol{a}) = \sum_{j=1}^{l_i} \mathcal{R}_j(\boldsymbol{a})$.

Now we will incorporate the randomness coming from η in our conditional Gaussian approximation (8.31). Using $x = n^{1/p}$ and Lemma 8.3,

$$\mathbb{E}(L_\alpha(\boldsymbol{\eta}, x)) \le \frac{n^{1-\alpha/p}}{m} c_\alpha \max_l \mathbb{E}[|\tilde{S}_{l,m+l}|^\alpha] = o(1).$$

Thus, we have,

$$\max_{1\le i\le n} |\Gamma_i(\boldsymbol{\eta}) - \mathbf{B}_{\boldsymbol{\eta}}(Q_{l_i}(\boldsymbol{\eta}))| = o_{\mathbb{P}}(n^{1/p}). \tag{8.32}$$

Similar to [9], the probability space for the above in-probability convergence is

$$(\Omega_*, \mathcal{A}_*, \mathrm{P}_*) = (\Omega, \mathcal{A}, \mathbb{P}) \times \prod_{\tau\in\Omega} \left(\Omega_{\boldsymbol{\eta}(\tau)}, \mathcal{A}_{\boldsymbol{\eta}(\tau)}, \mathbf{P}_{\boldsymbol{\eta}(\tau)}\right),$$

where $(\Omega, \mathcal{A}, \mathbb{P})$ is the probability space on which the random variables $(\varepsilon_i)_{i\in\mathbb{Z}}$ are defined and, for a set $A \subset \Omega_*$ with $A \in \mathcal{A}_*$, the probability measure P_* is defined as

$$\mathrm{P}_*(A) = \int_\Omega \mathbf{P}_{\boldsymbol{\eta}(\omega)}\left(A_\omega\right) \mathbb{P}(d\omega)$$

where A_ω is the ω-section of A. Here we recall that, for each $\mathbf{a}$, $(\Omega_{\boldsymbol{a}}, \mathcal{A}_{\boldsymbol{a}}, \mathbf{P}_{\boldsymbol{a}})$ is the probability space carrying $\mathbf{B}_{\boldsymbol{a}}$ and $\mathcal{R}_l(\boldsymbol{a})$ given $\boldsymbol{\eta} = \mathbf{a}$. On the probability space $(\Omega_*, \mathcal{A}_*, \mathrm{P}_*)$, the random variable $\mathcal{R}_l(\boldsymbol{\eta})$ is defined as $\mathcal{R}_l(\boldsymbol{\eta})(\omega, \theta(\cdot)) = \mathcal{R}_l(\boldsymbol{\eta}(\omega))(\theta(\omega))$, where $(\omega, \theta(\cdot)) \in \Omega_*, \theta(\cdot)$ is an element in $\prod_{\tau\in\Omega} \Omega_{\boldsymbol{\eta}(\tau)}$ and $\theta(\tau) \in \Omega_{\boldsymbol{\eta}(\tau)}, \tau \in \Omega$. The other random processes $M_{l_i}(\boldsymbol{\eta})$ and $\mathbb{B}_{\boldsymbol{\eta}}\left(Q_{l_i}(\boldsymbol{\eta})\right)$ can be similarly defined.

8.1.6. *Unconditional Gaussian Approximation.* In this subsection, we lift the condition on the shared innovations and work with $\Gamma_i(\boldsymbol{\eta})$ and $Q_{l_i}(\boldsymbol{\eta})$. Again, given $\boldsymbol{a}$, using (8.30), for i.i.d. standard normal random variables $(Z_k^{\boldsymbol{a}})_{k=1}^{l_n}$, let

$$\omega_k(\boldsymbol{a}) := \sum_{l=1}^{k-1} \sqrt{V_l^0(\boldsymbol{a}_{3l})} Z_l^{\boldsymbol{a}},$$

and

$$R_k(\boldsymbol{a}) = \sqrt{\tilde{V}_1(\boldsymbol{a}_0)} Z_0^{\boldsymbol{a}} + \sqrt{\tilde{V}_{2k}(\boldsymbol{a}_{3k})} Z_k^{a},$$

such that we can write

$$\mathbf{B}_a(Q_k(\boldsymbol{a})) = \omega_k(\boldsymbol{a}) + R_k(\boldsymbol{a}).$$

For $1 \le k \le l_n$, denote $Z_k^\eta := Z_k$ to be the i.i.d. standard normal random variables independent of ε's , and define

$$\Phi_i = \sum_{k=1}^{l_i-1} \sqrt{V_k^0(\boldsymbol{\eta}_{3k})} Z_l, \tag{8.33}$$

Note that $\Phi_i | \{\boldsymbol{\eta} = \boldsymbol{a}\} =_{\mathbb{D}} \omega_{l_i}(\boldsymbol{a})$. Hence

$$(\Phi_i)_{\{i\ge1\}} =_{\mathbb{D}} (\omega_{l_i}(\boldsymbol{\eta}))_{\{i\ge1\}}.$$

By Jensen's inequality and Lemma 8.3, we have

$$\max_{1\le k\le l_n} \|\tilde{V}_{2k}(\boldsymbol{\eta}_{3k})\|^{\alpha/2} = o(mn^{\alpha/p-1}), \tag{8.34}$$

which implies

$$\mathbb{P}(\max_{1\le k\le l_n} |\tilde{V}_{2k}(\boldsymbol{\eta}_{3k})| \ge cn^{2/p}) \le \sum_{k=1}^{K} \mathbb{P}(|\tilde{V}_{2k}(\boldsymbol{\eta}_{3k})| \ge cn^{2/p})$$

$$\leq c^{-\alpha/2}\frac{n}{3m}n^{-\alpha/p}\|\tilde{V}_{2k}(\boldsymbol{\eta}_{3k})\|^{\alpha/2} = o(1).$$

Therefore,

$$\max_{1\leq k\leq l_n}|\tilde{V}_{2k}(\boldsymbol{\eta}_{3k})| = o_{\mathbb{P}}(n^{2/p}). \tag{8.35}$$

Similarly, $|\tilde{V}_1(\boldsymbol{\eta}_0)| = o_{\mathbb{P}}(n^{2/p})$. Thus,

$$\max_{1\leq i\leq n}|\mathbf{B}_{\boldsymbol{\eta}}(Q_{l_i}(\boldsymbol{\eta})) - \omega_{l_i}(\eta)| = o_{\mathbb{P}}(n^{1/p}). \tag{8.36}$$

From (8.32) and (8.36), we have

$$\max_{1\leq i\leq n}|\Gamma_i(\boldsymbol{\eta}) - \omega_{l_i}(\eta)| = o_{\mathbb{P}}(n^{1/p}). \tag{8.37}$$

Recall (8.27). We have the following distributional equality

$$(\Gamma_i(\boldsymbol{\eta}) + M_{l_i}(\boldsymbol{\eta}))_{1\leq i\leq n} =_{\mathbb{D}} (S_i^{\diamond})_{1\leq i\leq n}. \tag{8.38}$$

In view of (8.37) and (8.38), we need to prove Gaussian approximation $\Phi_i + M_{l_i}(\boldsymbol{\eta})$. Let for $1\leq l\leq l_n$,

$$\tilde{A}_l = \sqrt{V_l^0(\boldsymbol{\eta}_{3l})}Z_l + M_{3l}(\boldsymbol{\eta}_{3l}) + M_{3l+1}(\boldsymbol{\eta}_{3l}). \tag{8.39}$$

Let $S_i^{\natural} = \sum_{l=1}^{l_i}\tilde{A}_l$. Note that by the same argument as in (8.34) and (8.35), we have

$$\max_{1\leq i\leq n}|\Phi_i + M_{l_i}(\boldsymbol{\eta}) - S_i^{\natural}| = \max_{1\leq i\leq n}|V_{l_i}^0(\eta_{3l}) - M_{3l_i+1}(\boldsymbol{\eta}_{3l}) + M_1(\boldsymbol{\eta}_0)| = o_{\mathbb{P}}(n^{1/p}). \tag{8.40}$$

Hence, by Theorem 1 of [95] (ignoring the technicalities of enriched space), on the same probability space that defines $\tilde{A}_l$'s, we have a standard Brownian Motion $\mathbb{B}$ such that

$$\max_{1\leq i\leq n}|S_i^{\natural} - \mathbb{B}(\sigma_i^2)| = o_{\mathbb{P}}(n^{1/p}), \tag{8.41}$$

where $\sigma_i^2 = \sum_{l=1}^{l_i}\|\tilde{A}_l\|^2$.

8.1.7. *Approximation of the variance.* In this final step of our proof, we aim to provide an approximation to σ_i^2 in terms of the variance of the truncated random process $X_i^{\oplus}$. To that end, we start from the expression of $V_l^0(\boldsymbol{a}_{3l})$ in equation (8.29):

$$\begin{aligned} V_l^0(\boldsymbol{a}_{3l}) = \|\tilde{B}_{3l-1}\|^2 + \|\tilde{B}_{3l}(\boldsymbol{a}_{3l})\|^2 - M_{3l}^2(\boldsymbol{a}_{3l}) + \|\tilde{B}_{3l+1}(\boldsymbol{a}_{3l})\|^2 \\ - M_{3l+1}^2(\boldsymbol{a}_{3l}) - \|C_{3l-1}(\boldsymbol{\eta}_{3l-2})\|^2 + \|C_{3l+2}(\boldsymbol{\eta}_{3l+1})\|^2 \\ + 2\mathbb{E}\left[\tilde{B}_{3l-1}\tilde{B}_{3l}(\boldsymbol{a}_{3l})\right] + 2\mathbb{E}\left[\tilde{B}_{3l+1}(\boldsymbol{a}_{3l})C_{3l+2}(\boldsymbol{\eta}_{3l+1})\right]. \end{aligned} \tag{8.42}$$

Using (8.42) in view of the definition of A_l in (8.39) yields,

$$\tilde{v}_l := \|\tilde{A}_l\|^2 = \|\tilde{B}_{3l-1}\|^2 + \|\tilde{B}_{3l}\|^2 + \|\tilde{B}_{3l+1}\|^2$$

$$+ 2\mathbb{E}(\tilde{B}_{3l-1}\tilde{B}_{3l}) + 2\mathbb{E}(\tilde{B}_{3l}\tilde{B}_{3l+1}) + 2\mathbb{E}(\tilde{B}_{3l+1}\tilde{B}_{3l+2})$$

$$- \|C_{3l-1}(\boldsymbol{\eta}_{3l-2})\|^2 + \|C_{3l+2}(\boldsymbol{\eta}_{3l+1})\|^2. \tag{8.43}$$

Hence,

$$\sigma_i^2 = \sum_{l=1}^{l_i}(\|\tilde{B}_{3l-1}\|^2 + \|\tilde{B}_{3l}\|^2 + \|\tilde{B}_{3l+1}^2\| + 2\mathbb{E}[\tilde{B}_{3l-1}\tilde{B}_{3l} + \tilde{B}_{3l}\tilde{B}_{3l+1} + \tilde{B}_{3l+1}\tilde{B}_{3l+2}])$$
$$- \|C_2(\boldsymbol{\eta}_1)\|^2 + \|C_{3l_i+2}(\boldsymbol{\eta}_{3l_i+1})\|^2.$$

We define the functional dependence measure for the process $\tilde{X}_i$ as

$$\tilde{\delta}_p(k) = \sup_i \|\tilde{X}_i - \tilde{X}_{i,i-k}\|_p,$$

where $p \geq 2$. We can easily have the following simple relation:

$$\tilde{\delta}_p(k) \leq \delta_p(k). \tag{8.44}$$

Lemma 3.1 and (8.44) along with observing that $\max_{1\leq k\leq \lceil n/m\rceil} \mathbb{E}(B_k^2) = O(m)$ yields

$$\max_{1\leq i\leq n} |\|\tilde{S}_i\|^2 - \sigma_i^2| = O(m) = o(n^{(\alpha/p-1)/(\alpha/2-1)}),$$

in view of (8.3). This implies

$$\max_{1\leq i\leq n} |\mathbb{B}(\sigma_i^2) - \mathbb{B}(\|\tilde{S}_i\|^2)| = o_{\mathbb{P}}(n^{(\alpha/p-1)/(\alpha-2)}\log n) = o_{\mathbb{P}}(n^{1/p}). \tag{8.45}$$

Now, using $\|S_n^{\oplus} - \tilde{S}_n\| \leq \sqrt{n}\Theta_{m,2} \leq \sqrt{n}\Theta_{m,p}$ and (8.2), we obtain,

$$|\|S_i^{\oplus}\|^2 - \|\tilde{S}_i\|^2| \leq \|S_i^{\oplus} - \tilde{S}_i\|\|S_i^{\oplus} + \tilde{S}_i\| \leq n\Theta_{m,p}\Theta_{0,p} = O(nm^{-A}) = o(n^{2/p}/\log n).$$

Therefore,

$$\max_{1\leq i\leq n} \left|\mathbb{B}(\|\tilde{S}_i\|^2) - \mathbb{B}(\|S_i^{\oplus}\|^2)\right| = o_{\mathbb{P}}(n^{1/p}), \tag{8.46}$$

which completes the proof of (2.12) in view of Propositions 8.1, 8.2, 8.3, and equations (8.32), (8.37), (8.38), (8.40), (8.41), (8.45) and (8.46).

8.1.8. *Connecting $\|S_i^{\oplus}\|^2$ to $\|S_i\|^2$.* The crux of the proof in this section relies on the following fundamental lemma connecting the variance of the truncated process to that in terms of the original process.

LEMMA 8.4. *Let $T_{n^{1/p}}(X_i)$ and $S_i^{\oplus}$ be defined as in (2.6) and (2.6) respectively. Also assume condition (2.1) for the process $(X_t)_{t\geq 1}$. Then,*

$$\max_{1\leq i\leq n} |\mathbb{E}(S_i^2) - \mathbb{E}(S_i^{\oplus^2})| = o(n^{2/p}). \tag{8.47}$$

Proof. In view of our truncated uniform integrability condition (2.2), one obtains,

$$\max_{1\le i\le n}(\sum_{t=1}^{i}\mathbb{E}[X_t^{\oplus}])^2 = n^2(n^{1/p-1})^2 o(1) = o(n^{2/p}). \tag{8.48}$$

Thus, it is enough to show that

$$\max_{1\le i\le n}|\mathbb{E}[S_i^2-(\sum_{t=1}^{i}X_t^{\oplus})^2]| = o(n^{2/p}).$$

Writing $\mathbb{E}[S_i^2-(\sum_{t=1}^{i}X_t^{\oplus})^2]=\sum_{s=1}^{i}\sum_{t=1}^{i}\mathbb{E}[X_sX_t-X_s^{\oplus}X_t^{\oplus}]$, observe that

$$|\mathbb{E}(X_sX_t)-\mathbb{E}(X_s^{\oplus}X_t^{\oplus})| \le |\mathbb{E}[(X_t-X_t^{\oplus})X_s]| + |\mathbb{E}[X_t^{\oplus}(X_s-X_s^{\oplus})]| := I + II. \tag{8.49}$$

Since (8.49) is symmetric in s and t, we can assume without loss of generality that $s>t$. Recall the causal representation (1.2) for X_s. Denote by $X_{s,\{t\}}=g_i(\varepsilon_s,\varepsilon_{s-1},\ldots,\varepsilon_{t+1},\varepsilon_t',\varepsilon_{t-1}',\ldots)$, where ε_l' is an i.i.d. copy of ε_l. Such coupling has also been used in [22] to obtain weaker conditions for [9]'s result. Since $X_{s,\{t\}}$ is independent of X_t, hence for the term I in (8.49), Hölder's inequality along with condition (2.2) yields,

$$\begin{aligned}\mathbb{E}[(X_t-X_t^{\oplus})X_s] = \mathbb{E}[(X_t-X_t^{\oplus})(X_s-X_{s,\{t\}})] &\le \|X_t-X_t^{\oplus}\|_{\frac{p}{p-1}}\|X_s-X_{s,\{t\}}\|_p\\ &\le C_p n^{2/p-1}\Theta_{s-t,p}o(1),\end{aligned} \tag{8.50}$$

where the last inequality follows from an application of Theorem 1(iii) of [100]. Here C_p is a constant depending on p. Note that here and also in the subsequent equations, the $o(1)$ term is independent of s, t and i, since our condition (2.2) is defined uniformly for all $1\le t\le n$.

Now we will tackle the term II in (8.49). For simplicity, let us denote $X_s-X_s^{\oplus}$ by $r_{n^{1/p}}(X_s)$. Again via Hölder inequality we obtain,

$$|\mathbb{E}[X_t(X_s-X_s^{\oplus})]| = |\mathbb{E}[X_t(r_{n^{1/p}}(X_s)-r_{n^{1/p}}(X_{s,\{t\}}))]| \le \|X_t\|_p\|r_{n^{1/p}}(X_s)-r_{n^{1/p}}(X_{s,\{t\}})\|_{\frac{p}{p-1}}. \tag{8.51}$$

Now, in the light of Hölder's inequality and Condition 2.2, we have,

$$\begin{aligned}\mathbb{E}[|r_{n^{1/p}}(X_s)-r_{n^{1/p}}(X_{s,\{t\}})|^{\frac{p}{p-1}}] &\le \mathbb{E}[|X_s-X_{s,\{t\}}|^{\frac{p}{p-1}}\Big(\mathbb{I}\{|X_s|\ge n^{1/p}\}+\mathbb{I}\{|X_{s,\{t\}}|\ge n^{1/p}\}\Big)]\\ &\le \big(\mathbb{E}[\|X_s-X_{s,\{t\}}\|^p]\big)^{\frac{1}{p-1}}\ \|\mathbb{I}\{|X_s|\ge n^{1/p}\}+\mathbb{I}\{|X_{s,\{t\}}|\ge n^{1/p}\}\|_{\frac{p-1}{p-2}}\\ &\le C_p'\Theta_{s-t,p}^{\frac{p}{p-1}}n^{\frac{2-p}{p-1}}o(1).\end{aligned} \tag{8.52}$$

Therefore, using $\sup_t\|X_t\|_p\le\Theta_{0,p}$, (8.49), (8.50) and (8.51) in conjunction with (8.52) yields, for a fixed $1\le i\le n$,

$$\sum_{s=1}^{i}\max_{1\le t\le i}|\mathbb{E}(X_iX_j)-\mathbb{E}(X_i^{\oplus}X_j^{\oplus})| \le C_p'' o(1)\Theta_{0,p}\sum_{s=1}^{i}\sum_{t=1}^{i}\Theta_{|s-t|,p}n^{2/p-1}.$$

This, in view of $\Theta_{i,p}=O(i^{-A})$ for $A>1$, immediately implies that,

$$\sup_{1\le i\le n}|\mathbb{E}(S_i^2)-\mathbb{E}(S_i^{\oplus^2})| \le C_p''\Theta_{0,p}O(1)n^{2/p}o(1) = o(n^{2/p}),$$

which completes the proof of the lemma. □

The proof of (2.13) is immediate using Theorem 2.4, Lemma 8.4 and increment property of Brownian motion. □

8.2. *Proof of Theorem 2.2.* Before we state the proof of Theorem 2.2, we state and prove a couple of lemmas which are heavily used in the subsequent proofs.

Lemma 8.5 concerns approximating a Gaussian process Y_t by its orthogonal projection on consecutive blocks of size m. On the other hand, given a Brownian Motion $\mathbb{B}_n$ and a process X_t, Lemma 8.6 constructs a Gaussian process Y_t from X_t, such that the Y_t has the same covariance structure as X_t, and partial sums of certain projections of Y_t are $\mathbb{B}_t$.

Lemma 8.5. *Let $(X_t)_{t=1}^n$ be a non-stationary process satisfying Condition 2.1, 2.2 and 2.3. Consider $m \in \mathbb{N}$ with $m/n \to 0$ and $m \to \infty$ as $n \to \infty$. Further let Y_t be a Gaussian process with $\mathrm{Cov}(Y_s, Y_t) = Cov(X_s, X_t)$. Let $\Xi_k = \sum_{j=(k-1)m+1}^{km\wedge n} Y_j$ be the block sums for $1 \le k \le \lceil n/m \rceil$, and let the k-th order projection be defined as*

$$\xi_k = \mathbb{E}[\sum_{i=1}^n Y_i | \Xi_k, \dots, \Xi_1] - \mathbb{E}[\sum_{i=1}^n Y_i | \Xi_{k-1}, \dots, \Xi_1].$$

Then it holds that

$$\max_{1\le i\le n} |\sum_{j=1}^i Y_j - \sum_{k=1}^{\lceil i/m\rceil} \xi_k| = O_{\mathbb{P}}(\sqrt{m}\log n). \tag{8.53}$$

Further, recall B_k from (8.25), defined using m considered in this lemma. Let $\amalg_k = \mathbb{E}[\xi_k^2]$ for $1 \le k \le \lceil n/m \rceil$. Then it holds that

$$\amalg_k = u_{kk} + 2u_{k+1,k} + \boldsymbol{u}_k^T U_{k,k}^{-1} \boldsymbol{u}_k - \boldsymbol{u}_{k-1}^T U_{k-1,k-1}^{-1} \boldsymbol{u}_{k-1} + O(m^{1-A}), \tag{8.54}$$

where $u_{kl} = \mathbb{E}[B_k B_l]$, $U_{k,k} := \mathrm{Cov}\begin{pmatrix} B_1 \\ B_2 \\ \vdots \\ B_k \end{pmatrix}$, and $\boldsymbol{u}_k = (u_{k+1,1}, \dots, u_{k+1,k})^T$ with $\boldsymbol{u}_{\lceil n/m\rceil} = 0$.

Proof. We will require repeated use of results from conditional mean of normal distributions. Observe that Lemma 3.1 implies

$$\mathrm{Var}\begin{pmatrix} B_1 \\ B_2 \\ \vdots \\ B_k \\ B_{k+1} + \dots + B_{\lceil n/m\rceil} \end{pmatrix} = \begin{pmatrix} U_{k,k} & \boldsymbol{u}_k + \boldsymbol{r}_k \\ \boldsymbol{u}_k + \boldsymbol{r}_k & \|B_{k+1} + \dots + B_{\lceil n/m\rceil}\|^2 \end{pmatrix}, \tag{8.55}$$

where $\boldsymbol{r}_k = \sum_{i=k+2}^{\lceil n/m\rceil} (u_{i,1}, \dots, u_{i,k})^T$ with $\boldsymbol{r}_{\lceil n/m\rceil -1} = \boldsymbol{r}_{\lceil n/m\rceil} = 0$. Observe that, by Lemma 3.1,

$$\begin{aligned} \|\boldsymbol{r}_k\| &= O(m^{1-A}), \\ \|\boldsymbol{u}_k\| &= O(1), \\ |\boldsymbol{r}_k^T \boldsymbol{u}_k| &= O(m^{1-A}). \end{aligned} \tag{8.56}$$

Let $\boldsymbol{R}_i = \sum_{j=i+1}^{m\lceil i/m\rceil \wedge n} Y_j$ with $\boldsymbol{R}_{km} = 0$ for $1 \le k \le \lceil n/m \rceil$, be the remainder part of the partial sum. Clearly, (8.55) yields

$$\sum_{k=1}^{\lceil i/m\rceil} (\xi_k - \Xi_k) - \boldsymbol{R}_i = \mathbb{E}\Big[\sum_{k=\lceil i/m\rceil+1}^{\lceil n/m\rceil} \Xi_k | \Xi_{\lceil i/m\rceil}, \dots, \Xi_1\Big] - \boldsymbol{R}_i$$

$$= (\boldsymbol{u}_{\lceil i/m\rceil}^T + \boldsymbol{r}_{\lceil i/m\rceil}^T) U^{-1}_{\lceil i/m\rceil,\lceil i/m\rceil} \begin{pmatrix} \Xi_1 \\ \vdots \\ \Xi_{\lceil i/m\rceil} \end{pmatrix} - \boldsymbol{R}_i.$$

Therefore, for all i,

$$\| \sum_{k=1}^{\lceil i/m\rceil} \xi_k - \sum_{j=1}^{i} Y_j \|^2 \lesssim \mathbb{E}[\boldsymbol{R}_i^2] + (\boldsymbol{u}_{\lceil i/m\rceil}^T + \boldsymbol{r}_{\lceil i/m\rceil}^T) U^{-1}_{\lceil i/m\rceil,\lceil i/m\rceil} (\boldsymbol{u}_{\lceil i/m\rceil} + \boldsymbol{r}_{\lceil i/m\rceil}). \tag{8.57}$$

Observe that, there exists constant $c > 0$ such that

$$\lambda_{\min}(U_{k,k}) \ge \max_{1\le l\le k} [\mathrm{Var}(B_l) - \sum_{j\ne l, 1\le j\le k} |\mathbb{E}(B_l B_j)|] \ge \max_{1\le l\le k} \mathrm{Var}(B_l) - c \to \infty, \tag{8.58}$$

where the first inequality is by Gershgorin circle theorem and the second inequality follows by invoking Lemma 3.1 and noting that $\sum_{j=1,|j-l|=1}^{k} |\mathbb{E}[B_l B_j]| = O(1)$, and $\sum_{j=1,|j-l|\ge 2}^{k} |\mathbb{E}[B_l B_j]| = O(m^{1-A})$. The limit assertion is due to Condition 2.3. Finally, via (8.56), $\|\boldsymbol{u}_k + \boldsymbol{r}_k\|^2 = O(1)$. Thus, (8.58) implies

$$(\boldsymbol{u}_{\lceil i/m\rceil} + \boldsymbol{r}_{\lceil i/m\rceil})^T U^{-1}_{\lceil i/m\rceil,\lceil i/m\rceil} (\boldsymbol{u}_{\lceil i/m\rceil} + \boldsymbol{r}_{\lceil i/m\rceil}) \le \lambda_{\max}(U^{-1}_{\lceil i/m\rceil,\lceil i/m\rceil}) \|\boldsymbol{u}_{\lceil i/m\rceil} + \boldsymbol{r}_{\lceil i/m\rceil}\|^2 \le O(1). \tag{8.59}$$

This directly yields, in view of (8.57), (8.59) and $\max_i \mathbb{E}[\boldsymbol{R}_i^2] \lesssim m\Theta_{0,2}^2 .]$,

$$\max_i \mathbb{E}\Big[\Big| \sum_{k=1}^{\lceil i/m\rceil} \xi_k - \sum_{j=1}^{i} Y_j \Big|^2\Big] = O(m), \tag{8.60}$$

which implies (8.53) via an application of (8.3).

For proving (8.54), (8.55) implies that

$$\xi_k = \Xi_k + (\boldsymbol{u}_k^T + \boldsymbol{r}_k^T) U^{-1}_{k,k} \begin{pmatrix} \Xi_1 \\ \vdots \\ \Xi_k \end{pmatrix} - (\boldsymbol{u}_{k-1}^T + \boldsymbol{r}_{k-1}^T) U^{-1}_{k-1,k-1} \begin{pmatrix} \Xi_1 \\ \vdots \\ \Xi_{k-1} \end{pmatrix}$$

$$= \Xi_k + \Big((\boldsymbol{u}_k^T + \boldsymbol{r}_k^T) U^{-1}_{k,k} + (-(\boldsymbol{u}_{k-1}^T + \boldsymbol{r}_{k-1}^T) U^{-1}_{k-1,k-1} \;\; 1) \Big) \begin{pmatrix} \Xi_1 \\ \vdots \\ \Xi_k \end{pmatrix}. \tag{8.61}$$

In light of (8.56) and (8.58), and observing that $\mathrm{Var}\left(\boldsymbol{u}_k^T U_{k,k}^{-1} \begin{pmatrix} \Xi_1 \\ \vdots \\ \Xi_k \end{pmatrix}\right) = \boldsymbol{u}_k^T U_{k,k}^{-1} \boldsymbol{u}_k$, and $\mathrm{Cov}\left(\Xi_k,\ \boldsymbol{u}_{k-1}^T U_{k-1,k-1}^{-1} \begin{pmatrix} \Xi_1 \\ \vdots \\ \Xi_{k-1} \end{pmatrix}\right) = \boldsymbol{u}_{k-1}^T U_{k-1,k-1}^{-1} \boldsymbol{u}_{k-1}$, (8.61) directly results in

$$\mathrm{II}_k = u_{kk} + \boldsymbol{u}_k^T U_{k,k}^{-1} \boldsymbol{u}_k - \boldsymbol{u}_{k-1}^T U_{k-1,k-1} \boldsymbol{u}_{k-1} + 2\Big(-\boldsymbol{u}_{k-1}^T U_{k-1,k-1}^{-1} \quad 1 \Big) \boldsymbol{u}_k + O(m^{1-A}). \tag{8.62}$$

Write $\boldsymbol{u}_k = (\boldsymbol{s}_k \quad u_{k+1,k})^T$, where $\boldsymbol{s}_k = (u_{k+1,1} \dots u_{k+1,k-1})^T$. Then similar to (8.56), $\|\boldsymbol{s}_k\| = O(m^{1-A})$. Therefore, in light of (8.58), (8.62) can be re-written as (8.54), which completes the proof.

□

Before moving onto the construction of Y^c-processes, we need to introduce a few notation. For $w_1, \dots, w_n \overset{\text{i.i.d.}}{\sim} N(0,1)$, define

$$Y_t^w := Y_t(w_1, \dots, w_t) = \|X_t\| \Big(\sum_{i=1}^t x_i^{(t)} w_i \Big), \tag{8.63}$$

where $x_1^{(1)} = 1$, and for $i \le t$, $x_i^{(t)}$ is obtained by solving the equation system $\sum_{k=1}^i x_k^{(i)} x_k^{(t)} = \mathrm{Corr}(X_i, X_t)$. Observe that by construction, $\mathrm{Cov}(Y_s, Y_t) = \mathrm{Cov}(X_s, X_t)$ for all $1 \le s, t \le n$. For $m \to \infty$ with $m/n \to 0$, define $(\alpha_i^{(k)})_{1 \le i \le km, 1 \le k \le \lceil n/m \rceil}$ such that

$$\mathbb{E}\Big[\sum_{i=1}^n Y_i^w \Big| \Big(\sum_{i=(j-1)m+1}^{jm} Y_i^w \Big)_{j=1}^k \Big] = \sum_{i=1}^{km} \alpha_i^{(k)} w_i. \tag{8.64}$$

Recall the definition of II_k from Lemma 8.5. By (8.63) and (8.64), $\mathrm{II}_k = \sum_{i=1}^{km} \alpha_i^{(k)^2}$. We emphasize that at the present level of construction, Y_t^c are not defined. However, what is known is their distribution, and therefore II_k and $\alpha_i^{(k)}$'s are also known. Now Y_t^c will be defined through this quantities.

Lemma 8.6. *Let $\mathbb{B}$ be a given Brownian Motion, and $(X_i)_{i=1}^n$ be a given stochastic process satisfying Conditions 2.1, 2.2 and 2.3. Let II_k be defined as in Lemma 8.5 with some $m \in \mathbb{N}$ such that $m \to \infty$, $m/n \to 0$ as $n \to \infty$. Consider the following algorithm of constructing a new Gaussian process Y_t^c:*

- *Write $\mathbb{B}(\sum_{j=1}^k \mathrm{II}_j) := \sum_{j=1}^{km} \alpha_i^{(k)} \eta_i$ for i.i.d. standard normal variables $\eta_1, \dots, \eta_n$.*
- *Define $Y_t^c := Y_t(\eta_1, \dots, \eta_t)$ as in (8.63).*

Then it holds that

- *For $1 \le i \le n$, $Y_i^c \sim N(0, \mathrm{Var}(X_t))$, and for $1 \le i \ne j \le n$, $\mathrm{Cov}(Y_i^c, Y_j^c) = \mathrm{Cov}(X_i, X_j)$.*
- $\max_{1 \le i \le n} |\mathbb{B}(\sum_{j=1}^{\lceil i/m \rceil} \mathrm{II}_j) - \sum_{j=1}^i Y_j^c| = O_{\mathbb{P}}(\sqrt{m} \log n)$.

PROOF. The first assertion can be verified directly from our construction. For the second assertion, let Ξ_l^c and ξ_l^c be obtained from Y_t^c as in Lemma 8.5. Then as in (8.64), $\mathbb{B}(\sum_{l=1}^{\lceil i/m \rceil} \text{II}_l)$ can also be represented as $\sum_{l=1}^{\lceil i/m \rceil} \xi_l^c$. Then the result follows from Lemma 8.5. $\square$

PROOF OF THEOREM 2.2. Consider the Brownian Motion $\mathbb{B}$ from the conclusion of Theorem 2.4. Let II_k be defined as in Lemma 8.5 with the original process $(X_t)_{t=1}^n$. From equation (8.54) in Lemma 8.5, it follows that

$$\max_{1\le i\le n} |\sum_{j=1}^{\lceil i/m \rceil} \text{II}_j - \mathbb{E}(S_i^2)| = O(nm^{-A}) = o_{\mathbb{P}}(n^{2/p}/\log n),$$

which yields

$$\max_{1\le i\le n} |\mathbb{B}(\mathbb{E}[S_i^2]) - \mathbb{B}(\sum_{j=1}^{\lceil i/m \rceil} \text{II}_j)| = o_{\mathbb{P}}(n^{1/p})$$

where the last equality follows from (8.3). Lemma 8.6 and Theorem 2.4 completes the proof of (2.8).

For (2.9), define $\text{II}_k^{\oplus}$ based on $X_t^{\oplus}$. Invoking (8.44), Lemma 8.5 and 8.6 holds when the original process is $X_t^{\oplus}$. Therefore, following exactly the above argument, along with Theorem 2.4 completes the proof. $\square$

9. Appendix B: Proofs of Theorems 2.3 and 2.5. Much like Theorems 2.4 and 2.2, we will prove Theorem 2.5 first, and the proof of Theorem 2.3 will follow from it.

9.1. *Proof of Theorem 2.5.* Recall A_0' from (2.10). Define, in the light of the form of $\Theta_{i,p}$ in (2.5) with $A' > A_0'$,

$$L' = \frac{f_1 - f_2 + A'\sqrt{(p-2)(f_3 - 3p)}}{A' f_4}, \tag{9.1}$$
$$\alpha' = \frac{(2p + p^2)A' + p^2 + 3p + 2 + \sqrt{f_5}}{2 + 2p + 4A'},$$

with $f_1(p, A) = Ap^2(A + 1)$, $f_2(p, A) = A(2pA + 3p - 2)$, $f_3(p, A) = p^3(1 + A)^2 + 6f_1 + 4pA - 2$, $f_4(p, A) = 2p(2pA^2+3pA+p-2)$ and $f_5(p, A) = p^2(p^2+4p-12)A^2+2p(p^3+p^2-4p-4)A+(p^2-p-2)^2$. Our choice of L' and α' satisfies the following relations which we use in our proof:

$$\frac{1}{2} - \frac{1}{p} - L'A' < 0, \tag{9.2}$$

$$L'\left(\frac{\alpha'}{2} - 1\right) + 1 - \frac{\alpha'}{p} < 0, \tag{9.3}$$

$$p < \alpha' < 2(1 + p + pA')/3, \tag{9.4}$$

$$1/p - 1/\alpha' + L' - L'(A' + 1)p/\alpha' = 0. \tag{9.5}$$

Note that with this new L' and α', proof of Lemma 8.1 and Lemma 8.3 goes through. We will also use the following lemma in a crucial step of our proof.

LEMMA 9.1. *Under the assumption of Theorem 2.5,*

$$\min_{l\geq 1}\{\Theta_{l,p}+ln^{2/p-1}\}=o\left(\frac{n^{1/p-1/2}}{\sqrt{\log\log n}}\right). \tag{9.6}$$

PROOF. Let $A'>1$. Define $B:=\frac{A'+1}{2}$ and choose $l_0=n^{\frac{1/2-1/p}{B}}$. In light of $1<B<A'$,

$$\begin{aligned}\min_{l\geq 0}\{\Theta_{l,p}+ln^{2/p-1}\}\leq \Theta_{l_0,p}+l_0n^{2/p-1}&=(n^{\frac{1}{2}-\frac{1}{p}})^{-\frac{A'}{B}}+(n^{\frac{1}{2}-\frac{1}{p}})^{\frac{1}{B}-2}\\&=o\left(\frac{n^{1/p-1/2}}{\sqrt{\log\log n}}\right).\end{aligned}$$

□

PROOF OF THEOREM 2.5. For the proof of Corollary 2.5, we proceed exactly as that of Theorem 2.4 with L and α replaced by L' and α', and equations (8.2)-(8.5) replaced by equations (9.2)-(9.5). The arguments up until equation (8.41) goes through verbatim with $m=\lfloor n^{L'}\rfloor$. Thus the only part which requires our attention is the *approximation of variance* step.

9.1.1. *Variance Regularization.* Based on equation (8.43), define, for $1\leq l\leq l_n$,

$$\begin{aligned}v_l:=&\|B_{3l-1}\|^2+\|B_{3l}\|^2+\|B_{3l+1}\|^2+2\mathbb{E}[B_{3l-1}B_{3l}]+2\mathbb{E}[B_{3l}B_{3l+1}]+2\mathbb{E}[B_{3l+1}B_{3l+2}]\\&-\|C_{3l-1}(\boldsymbol{\eta}_{3l-2})\|^2+\|C_{3l+2}(\boldsymbol{\eta}_{3l+1})\|^2+2\mathbb{E}[B_1B_{3l}+\ldots+B_{3l-2}B_{3l}]\\&+2\mathbb{E}[B_1B_{3l+1}+\ldots+B_{3l-1}B_{3l+1}]+2\mathbb{E}[B_1B_{3l+2}+\ldots+B_{3l}B_{3l+2}].\end{aligned} \tag{9.7}$$

Let $B_k^{\oplus}=\sum_{j=(k-1)m+1}^{km}(X_j^{\oplus}-\mathbb{E}(X_j^{\oplus}))$. Note that for all $j\geq 1$,

$$\|\tilde{B}_j-B_j^{\oplus}\|\leq\sqrt{m}\Theta_{m,2}\leq\sqrt{m}\Theta_{m,p}.$$

Let the projection operator P_k be defined as $P_k(\cdot)=\mathbb{E}[\cdot|\mathcal{F}_k]-\mathbb{E}[\cdot|\mathcal{F}_{k-1}]$. Using the Cauchy-Schwarz inequality, and $B_j^{\oplus}=\sum_{k=-\infty}^{j}P_kB_j^{\oplus}$, for $j\geq 1$,

$$|\mathbb{E}(\tilde{B}_j^2)-\mathbb{E}({B_j^{\oplus}}^2)|\leq\|\tilde{B}_j-B_j^{\oplus}\|\|\tilde{B}_j+B_j^{\oplus}\|\leq 4m\Theta_{m,p}\Theta_{0,p}. \tag{9.8}$$

Similarly,

$$|\mathbb{E}(\tilde{B}_j\tilde{B}_{j+1})-\mathbb{E}(B_j^{\oplus}B_{j+1}^{\oplus})|\leq\|B_j^{\oplus}\|\|\tilde{B}_{j+1}-B_{j+1}^{\oplus}\|+\|\tilde{B}_{j+1}\|\|\tilde{B}_j-B_j^{\oplus}\|\leq 4m\Theta_{m,p}\Theta_{0,p}. \tag{9.9}$$

Further, note that uniformly for all $k,l\geq 1$ using uniform integrability condition (2.2), we obtain

$$\begin{aligned}|\mathbb{E}(X_kX_l-X_k^{\oplus}X_l^{\oplus})|=&|\mathbb{E}(X_kX_l\mathbb{I}_{\max\{|X_k|,|X_l|\}\leq n^{1/p}})-\mathbb{E}(X_k^{\oplus}X_l^{\oplus})+\mathbb{E}(X_kX_l\mathbb{I}_{\max\{|X_k|,|X_l|\}>n^{1/p}}|\\=&|-\mathbb{E}(X_k^{\oplus}X_l^{\oplus}\mathbb{I}_{\max\{|X_k|,|X_l|\}>n^{1/p}})+\mathbb{E}(X_kX_l\mathbb{I}_{\max\{|X_k|,|X_l|\}>n^{1/p}}|\\\leq&|\mathbb{E}(X_k^{\oplus}X_l^{\oplus}\mathbb{I}_{\max\{|X_k|,|X_l|\}>n^{1/p}})|+|\mathbb{E}(X_kX_l\mathbb{I}_{\max\{|X_k|,|X_l|\}>n^{1/p}}|\\\leq&\mathbb{E}\left[(|X_k|^2+|X_l|^2)\mathbb{I}_{\max\{|X_k|,|X_l|\}>n^{1/p}}\right]=o(n^{2/p-1}).\end{aligned}$$

In view of (8.44), (10.35) also holds for $X_k^{\oplus} X_l^{\oplus}$. Noting that Condition 2.2 implies $\max_i |\mathbb{E}(X_i^{\oplus})| = o(n^{1/p-1})$, we have, for a fixed $0 \le j \le m-1$ and $l \ge 0$,

$$\begin{aligned}\mathbb{E}(B_{j+1}^2 - B_{j+1}^{\oplus^2}) &= \sum_{k=1}^{m} \mathbb{E}(X_{jm+k}^2 - X_{jm+k}^{\oplus^2}) \\ &\qquad + \sum_{s\neq t}^{m} \mathbb{E}(X_{jm+s}X_{jm+t} - X_{jm+s}^{\oplus}X_{jm+t}^{\oplus}) - \left(\mathbb{E}[\sum_{k=1}^{m} X_{jm+k}^{\oplus}]\right)^2 \\ &\le o(mn^{2/p-1}) + O(lmn^{2/p-1} + m\sum_{s=l+1}^{\infty}\sum_{i=0}^{\infty}\delta_p(i)\delta_p(i+s)),\end{aligned}$$

where the last line follows from using the fact that there are $\le m$ terms of the form $\mathbb{E}(X_k X_{k+s} - X_k^{\oplus} X_{k+s}^{\oplus})$ for a fixed $s \le m$ and applying (10.35) in the proof of Lemma 3.1. Note that

$$\sum_{j=l}^{\infty}\sum_{i=0}^{\infty}\delta_p(i)\delta_p(i+j) \le \Theta_{0,p}\Theta_{l,p}.$$

Hence,

$$\begin{aligned}|\mathbb{E}(B_j^2) - \mathbb{E}(B_j^{\oplus^2})| &= O(mn^{2/p-1} + m\min_{l\ge 0}\{ln^{2/p-1} + \Theta_{l+1,p}\}) \\ &= O(m\min_{l\ge 1}\{ln^{2/p-1} + \Theta_{l,p}\}).\end{aligned} \tag{9.10}$$

Similarly,

$$|\mathbb{E}(B_jB_{j+1}) - \mathbb{E}(B_j^{\oplus}B_{j+1}^{\oplus})| = O(m\min_{l\ge 1}\{ln^{2/p-1} + \Theta_{l,p}\}). \tag{9.11}$$

Therefore, (9.8), (9.9), (9.10) and (9.11) together with Lemma 3.1 yields

$$|\tilde{v}_l - v_l| = O(m\Theta_{m,p} + m\min_{l\ge 1}\{ln^{2/p-1} + \Theta_{l,p}\} + m^{1-A'}) = O(m^{1-A'} + m\min_{l\ge 1}\{ln^{2/p-1} + \Theta_{l,p}\}). \tag{9.12}$$

Applying Lemma 9.1 along with (9.12) leads to the following assertion:

$$\max_l \frac{|\tilde{v}_l - |v_l||}{m} \le \max_l |\tilde{v}_l - v_l|/m \to 0 \text{ as } n \to \infty. \tag{9.13}$$

Recall l_n from (8.24); (3.7) implies that

$$\begin{aligned}\max_{1\le l\le l_n}\Big|&\mathbb{E}[B_{3l-1}B_{3l}] + \mathbb{E}[B_{3l}B_{3l+1}] + \mathbb{E}[B_{3l+1}B_{3l+2}] + \mathbb{E}[B_1B_{3l} + \ldots + B_{3l-2}B_{3l}] \\ &+ \mathbb{E}[B_1B_{3l+1} + \ldots + B_{3l-1}B_{3l+1}] + \mathbb{E}[B_1B_{3l+2} + \ldots + B_{3l}B_{3l+2}]\Big|/m \to 0 \text{ as } n \to \infty.\end{aligned} \tag{9.14}$$

Hence, for large enough n, it follows from the regularity Condition 2.4 along with (9.14) that

$$\inf_l \frac{|v_l|}{3m} \ge \inf_l \Big(\|B_{3l-1}\|^2 + \|B_{3l}\|^2 + \|B_{3l+1}\|^2 + \|C_{3l+2}(\boldsymbol{\eta}_{3l+1})\|^2 - \|C_{3l-1}(\boldsymbol{\eta}_{3l-1})\|^2$$

$$+\, 2|\mathbb{E}[B_{3l-1}B_{3l}] + \mathbb{E}[B_{3l}B_{3l+1}] + \mathbb{E}[B_{3l+1}B_{3l+2}] + \mathbb{E}[B_1B_{3l} + \ldots + B_{3l-2}B_{3l}]$$

$$+\, \mathbb{E}[B_1B_{3l+1} + \ldots + B_{3l-1}B_{3l+1}] + \mathbb{E}[B_1B_{3l+2} + \ldots + B_{3l}B_{3l+2}]|\Big)/(3m) > \frac{c}{3}, \tag{9.15}$$

where we have used $\|C_{3l-1}(\boldsymbol{\eta}_{3l-1})\|^2 \le \|B_{3l-1}\|^2$ by Jensen's inequality. Observe that $\mathbb{B}(\sigma_n^2)$ can be represented as $\sum_{l=1}^{l_n} \sqrt{\tilde{v}_l} Z_l^*$ for i.i.d. standard Gaussian $Z_1^*, \ldots, Z_{l_n}^*$. We define the following Brownian motion:

$$\mathbf{B}_n = \sum_{l=1}^{l_n} \sqrt{|v_l|} Z_l^*. \tag{9.16}$$

Let

$$\Psi_n^2 = \|\mathbb{B}(\sigma_n^2)) - \mathbf{B}_n\|^2 = \sum_{l=1}^{l_n} (\sqrt{\tilde{v}_l} - \sqrt{|v_l|})^2. \tag{9.17}$$

Using (9.15) and (9.13), for large enough n we have,

$$\inf_l \frac{(\sqrt{\tilde{v}_l} + \sqrt{|v_l|})^2}{3m} \ge \inf_l \frac{\tilde{v}_l + |v_l|}{3m} \ge \inf_l \frac{2v_l - |\tilde{v}_l - v_l|}{3m} > \frac{c}{3}. \tag{9.18}$$

Therefore, from (9.12) and (9.18) one obtains for $1 \le l \le l_n$,

$$\sup_l (\sqrt{\tilde{v}_l} - \sqrt{|v_l|})^2 = \frac{(\tilde{v}_l - |v_l|)^2/3m}{(\sqrt{\tilde{v}_l} + \sqrt{|v_l|})^2/3m} = O\left(m^{1-2A'} + m\left(\min_{l\ge1}\{ln^{2/p-1} + \Theta_{l,p}\}\right)^2\right). \tag{9.19}$$

Thus, Lemma 9.1 together with (9.2) implies

$$\Psi_n^2 = O\left(nm^{-2A'} + n\left(\min_{l\ge1}\{ln^{2/p-1} + \Theta_{l,p}\}\right)^2\right) = o(n^{2/p}/\log\log n). \tag{9.20}$$

Note that $\mathbb{B}(\sigma_k^2) - \mathbf{B}_k$ is a Gaussian process with independent increments. Therefore, using Doob's maximal inequality we have

$$\max_{1\le i\le n} |\mathbb{B}(\sigma_i^2) - \mathbf{B}_i| = o_{\mathbb{P}}(n^{1/p}) \tag{9.21}$$

in view of (9.20). Next, definition of v_l in (9.7) along with (3.7) implies

$$|\mathbb{E}(S_i^2) - \sum_{l=1}^{l_i} |v_l|| \le |\mathbb{E}(S_i^2) - \sum_{l=1}^{l_i} v_l| = O(m), \tag{9.22}$$

which readily leads to

$$\max_{1\le i\le n} |\mathbb{B}(\mathbb{E}(S_i^2)) - \mathbf{B}_i| = o_{\mathbb{P}}(n^{1/p}). \tag{9.23}$$

This completes the proof of (2.14) in view of Propositions (8.1), (8.2), (8.3), and equations (8.32), (8.37), (8.38), (8.40), (8.41), (9.21) and (9.23). $\square$

9.2. *Proof of Theorem 2.3.*

PROOF. Now, we will construct a Gaussian process (Y_t^c) with the same covariance structure X_t such that (2.11) holds. Recall II_k as defined in Lemma 8.5, and v_l from (9.7). We wil use the same notations as in the proof of Theorem 2.2. Define

$$\tau_k := u_{kk} + 2u_{k+1,k} + \boldsymbol{u}_k^T U_{k,k}^{-1}\boldsymbol{u}_k - \boldsymbol{u}_{k-1}^T U_{k-1,k-1}^{-1}\boldsymbol{u}_{k-1}.$$

In light of Lemma 8.5, it is easy to observe that

$$\max_{1\le i\le n} \Big|\sum_{l=1}^{l_i} v_l - \sum_{l=1}^{\lceil i/m\rceil} \tau_l\Big| = O(m). \tag{9.24}$$

An argument similar to (9.15) yields that $\min_l |\tau_l|/m > c > 0$. This motivates us to define the following Gaussian process

$$\mathcal{G}_n := \sum_{l=1}^{\lceil n/m\rceil} |\tau_l|^{1/2} Z_l^*,$$

with the same Z_l^*'s as in definition of $\mathbf{B}_n$ in (9.16). In light of (9.24), we obtain

$$\max_{1\le i\le n} |\mathbf{B}_i - \mathcal{G}_i| = o_{\mathbb{P}}(\sqrt{m}\log n) = o_{\mathbb{P}}(n^{(\alpha/p-1)/(\alpha-2)}\log n) = o_{\mathbb{P}}(n^{1/p}). \tag{9.25}$$

Denote by

$$\mathcal{B}_n := \sum_{k=1}^{\lceil n/m\rceil} \text{II}_k^{1/2} Z_k^* = \mathbf{B}\Big(\sum_{k=1}^{\lceil n/m\rceil} \text{II}_k\Big).$$

Now we will show that $\mathcal{B}_n$ is a good enough approximation of $\mathcal{G}_n$. Since by definition of τ_l, $\text{II}_l = \tau_l + O(m^{1-A})$ for $1 \le l \le \lceil n/m\rceil$, thus an application of the argument same as (9.18) implies that

$$\|\mathcal{B}_n - \mathcal{G}_n\|^2 = \sum_{l=1}^{k} (\text{II}_l^{1/2} - \tau_l^{1/2})^2 = O(nm^{-2A'}) = o(n^{2/p}/\log\log n), \tag{9.26}$$

which in turn yields

$$\max_{1\le k\le\lceil n/m\rceil} |\mathcal{B}_k - \mathcal{G}_k| = o_{\mathbb{P}}(n^{1/p}) \tag{9.27}$$

using Doob's maximal inequality.

We will construct our Gaussian approximation $(Y_t^c)_{t=1}^n$ from $\mathbb{B}_k$ exactly as in Lemma 8.6. In light of (9.27), (9.25) and (9.21), Lemma 8.6 completes the proof of (2.11). □

10. Appendix C: Proofs of Section 3. In this section we provide the proofs of the two main results of the estimation step. Firstly, we prove the maximal quadratic deviation inequality in Theorem 3.1.

10.1. *Proof of Theorem 3.1.* In order to prove this theorem, we require the following lemmas. These results are well-known in the literature for $p > 4$ case; however we weaken the condition on moments to allow $p > 2$. For the sake of completeness we state and prove the results as we use it.

Our first lemma is the (one-dimensional) general version of Lemma D.6 of Supplement to [111].

LEMMA 10.1. *Assume that the process (1.2) has $\mathbb{E}(X_t) = 0$ and $\Theta_{0,p} < \infty$ for some $p > 2$. Let $\mathcal{D}_n \le n$ and $\boldsymbol{\eta}_k = (\varepsilon_{(k-1)\mathcal{D}_n+1}, \ldots, \varepsilon_{k\mathcal{D}_n})$. For $1 \le k \le \lfloor n/\mathcal{D}_n \rfloor$, define V_k as in (3.3) and let $V_{k,h} = \mathbb{E}(V_k|\boldsymbol{\eta}_{k-h}, \boldsymbol{\eta}_{k-h+1}, \ldots, \boldsymbol{\eta}_k)$. Then for $h \ge 2$,*

$$\|V_{k,h} - V_{k,h-1}\|_{p/2} \le \begin{cases} C_p \mathcal{D}_n^{1/2+2/p} \Theta_{0,p} \sum_{d=(h-2)\mathcal{D}_n+1}^{(h+1)\mathcal{D}_n} \delta_p(d), & 2 < p \le 4, \\ C_p \mathcal{D}_n \Theta_{0,p} \sum_{d=(h-2)\mathcal{D}_n+1}^{(h+1)\mathcal{D}_n} \delta_p(d), & p > 4. \end{cases} \tag{10.1}$$

PROOF. Let $\mathcal{F}_a^b = (\varepsilon_a, \varepsilon_{a+1}, \ldots, \varepsilon_b)$. Define the backward projection operator as $\mathcal{P}^a X = \mathbb{E}(X|\mathcal{F}_a^b) - \mathbb{E}(X|\mathcal{F}_{a+1}^b)$. Observe that for $h \ge 2$,

$$V_{k,h} - V_{k,h-1} = \mathbb{E}(V_k|\boldsymbol{\eta}_{k-h}, \ldots, \boldsymbol{\eta}_k) - \mathbb{E}(V_k|\boldsymbol{\eta}_{k-h+1}, \ldots, \boldsymbol{\eta}_k) = \sum_{l=1}^{\mathcal{D}_n} \mathcal{P}^{(k-h-1)\mathcal{D}_n+l} V_k,$$

and using Jensen's inequality (see [100]; Theorem 1.(i)) one obtains

$$\|\mathcal{P}^{(k-h-1)\mathcal{D}_n+l} V_k\|_{p/2} \le I + II, \tag{10.2}$$

where

$$I = \Big\| \sum_{t=(k-1)\mathcal{D}_n+1}^{(k\mathcal{D}_n)\wedge n} (X_t - X_{t,\{(k-h-1)\mathcal{D}_n+l\}}) \sum_{s=(t-\mathcal{D}_n)\vee 1}^{t} a_{s,t} X_s \Big\|_{p/2}, \tag{10.3}$$

$$II = \| \sum_{s=[(k-2)\mathcal{D}_n+1]\vee 1}^{(k\mathcal{D}_n)\wedge n} (X_s - X_{s,\{(k-h-1)\mathcal{D}_n+l\}}) \sum_{t=s}^{(s+\mathcal{D}_n)\wedge n} a_{s,t} X_{t,\{(k-h-1)\mathcal{D}_n+l\}} \|_{p/2}. \tag{10.4}$$

In order to tackle I we start off by noting the following assertion. In view of Burkholder's inequality ([92]; Theorem 2.1), it follows that

$$\begin{aligned} \| \sum_{s=1}^{t} c_s X_s \|_p \le \sum_{r=0}^{\infty} \| \sum_{s=1}^{t} c_s \mathcal{P}_{s-r} X_s \|_p &\le C_p \sum_{r=0}^{\infty} \sqrt{\sum_{s=1}^{t} \|c_s \mathcal{P}_{s-r} X_s\|_p^2} \qquad (10.5) \\ &\le C_p \sum_{r=0}^{\infty} \sqrt{\sum_{s=1}^{t} c_s^2 \delta_p(r)} \\ &= C_p \Theta_{0,p} \sqrt{\sum_{s=1}^{t} c_s^2}, \qquad (10.6) \end{aligned}$$

which entails, invoking Hölder's inequality, that,

$$I \leq \sum_{t=(k-1)\mathcal{D}_n+1}^{(k\mathcal{D}_n)\wedge n} \|X_t - X_{t,\{(k-h-1)\mathcal{D}_n+l\}}\|_p \left\| \sum_{s=(t-\mathcal{D}_n)\vee 1}^{t} a_{s,t} X_s \right\|_p$$

$$\leq C_p \Theta_{0,p} \sqrt{\mathcal{D}_n} \sum_{t=(k-1)\mathcal{D}_n+1}^{(k\mathcal{D}_n)\wedge n} \delta_p(t-(k-h-1)\mathcal{D}_n - l). \tag{10.7}$$

Similarly,

$$II \leq C_p \Theta_{0,p} \sqrt{\mathcal{D}_n} \sum_{t=(k-1)\mathcal{D}_n+1}^{(k\mathcal{D}_n)\wedge n} \delta_p(t-(k-h-1)\mathcal{D}_n - l). \tag{10.8}$$

Thus combining (10.7) and (10.8) with (10.2) yields,

$$\|\mathcal{P}^{(k-h-1)\mathcal{D}_n+l} V_k\|_{p/2} \leq C_p \Theta_{0,p} \sqrt{\mathcal{D}_n} \sum_{t=(k-1)\mathcal{D}_n+1}^{(k\mathcal{D}_n)\wedge n} \delta_p(t-(k-h-1)\mathcal{D}_n - l). \tag{10.9}$$

Finally, for $p > 4$, Rio [92]'s version of Burkholder's inequality (Theorem 2.1 of [92]) along with (10.9) implies

$$\|V_{k,h} - V_{k,h-1}\|_{p/2}^2 \leq C_p \sum_{l=1}^{\mathcal{D}_n} \|\mathcal{P}^{(k-h-1)\mathcal{D}_n+l} V_k\|_{p/2}^2 \leq C_p^3 \Theta_{0,p}^2 \mathcal{D}_n \sum_{l=1}^{\mathcal{D}_n} \left(\sum_{d=(h-1)\mathcal{D}_n-l+1}^{(h+1)\mathcal{D}_n-l} \delta_p(d) \right)^2$$

$$\leq C_p^3 \Theta_{0,p}^2 \mathcal{D}_n^2 \left(\sum_{d=(h-2)\mathcal{D}_n+1}^{(h+1)\mathcal{D}_n} \delta_p(d) \right)^2,$$

which completes the proof for $p > 4$. For the case $2 < p \leq 4$, one proceeds using Theorem 3.2 of [14] as follows:

$$\begin{aligned} \|V_{k,h} - V_{k,h-1}\|_{p/2}^{p/2} &\leq C_p \mathbb{E}\left[\left(\sum_{l=1}^{\mathcal{D}_n} |\mathcal{P}^{(k-h-1)\mathcal{D}_n+l} V_k|^2 \right)^{p/4} \right] \\ &\leq C_p \sum_{l=1}^{\mathcal{D}_n} \mathbb{E}[|\mathcal{P}^{(k-h-1)\mathcal{D}_n+l} V_k|^{p/2}] \\ &\leq C_p \Theta_{0,p}^{p/2} \mathcal{D}_n^{p/4} \sum_{l=1}^{\mathcal{D}_n} \left(\sum_{d=(h-2)\mathcal{D}_n+1}^{(h+1)\mathcal{D}_n} \delta_p(d) \right)^{p/2} \\ &\leq C_p \Theta_{0,p}^{p/2} \mathcal{D}_n^{1+p/4} \left(\sum_{d=(h-2)\mathcal{D}_n+1}^{(h+1)\mathcal{D}_n} \delta_p(d) \right)^{p/2}, \end{aligned} \tag{10.10}$$

where we applied $(|a_1| + \ldots + |a_n|)^{p/4} \leq |a_1|^{p/4} + \ldots + |a_n|^{p/4}$ for $2 < p \leq 4$. This completes the proof. □

Next we will use a version of equation 41, Proposition 8 of [103].

LEMMA 10.2. *Assume that the process* (1.2) *has* $\mathbb{E}(X_t) = 0$ *and* $\Theta_{0,p} < \infty$ *and* $p > 2$. *Then,*

$$\| \sum_{s,t=1}^{n} a_{s,t}(X_s X_t - \mathbb{E}(X_s X_t)) \|_{p/2} \le \begin{cases} C_p \mathcal{C} \Theta_{0,p}^2 n^{2/p}, & 2 < p \le 4 \\ C_p \mathcal{C} \Theta_{0,p}^2 \sqrt{n}, & p \ge 4 \end{cases} \tag{10.11}$$

where $\mathcal{C} = \max\{\max_{1 \le t \le n}(\sum_{s=1}^{n} a_{s,t}^2)^{1/2}, \max_{1 \le s \le n}(\sum_{t=1}^{n} a_{s,t}^2)^{1/2}\}$.

PROOF. Let

$$\sum_{s,t=1}^{n} a_{s,t}(X_s X_t - \mathbb{E}(X_s X_t)) = \sum_{r=-\infty}^{n} P_r \left(\sum_{s,t=1}^{n} a_{s,t} X_s X_t \right),$$

where P_r are defined as in the proof of Lemma 3.1. Now, Jensen's inequality yields,

$$\|P_r \left(\sum_{s,t=1}^{n} a_{s,t} X_s X_t \right) \|_{p/2} \le \| \sum_{s,t=1}^{n} a_{s,t} \left(X_s X_t - X_{s,\{r\}} X_{t,\{r\}} \right) \|_{p/2} \le I_r + II_r,$$

where

$$I_r = \| \sum_{s,t=1}^{n} a_{s,t}(X_s - X_{s,\{r\}}) X_t \|_{p/2}, \tag{10.12}$$

$$II_r = \| \sum_{s,t=1}^{n} a_{s,t} X_{s,\{r\}} (X_t - X_{t,\{r\}}) \|_{p/2}. \tag{10.13}$$

To tackle I_r, we employ Hölder's inequality and (10.6), it follows that,

$$I_r \le \sum_{s=1}^{n} \|X_s - X_{s,\{r\}}\|_p \| \sum_{t=1}^{n} a_{s,t} X_t \|_p \le C_p \Theta_{0,p} \mathcal{C} \sum_{s=1}^{n} \delta_p(s-r).$$

The same bound applies to II_r. Now, for $p > 4$, Burkholder's inequality ([92]) implies that

$$\begin{aligned} \| \sum_{s,t=1}^{n} a_{s,t}(X_s X_t - \mathbb{E}(X_s X_t)) \|_{p/2}^2 &\le C_p \sum_{r=-\infty}^{n} \|P_r \left(\sum_{s,t=1}^{n} a_{s,t} X_s X_t \right) \|_{p/2}^2 \\ &\le C_p \Theta_{0,p}^2 \mathcal{C}^2 \sum_{r=-\infty}^{n} \left(\sum_{s=1}^{n} \delta_p(s-r) \right)^2 \le C_p \Theta_{0,p}^4 n \mathcal{C}^2. \end{aligned}$$

As for $2 < p \le 4$, invoking [15] along with elementary inequality $(|a_1| + \ldots + |a_n|)^{p/4} \le |a_1|^{p/4} + \ldots + |a_n|^{p/4}$, yields,

$$\| \sum_{s,t=1}^{n} a_{s,t}(X_s X_t - \mathbb{E}(X_s X_t)) \|_{p/2}^{p/2} \le \left\| \sqrt{\sum_{r=-\infty}^{n} |P_r(\sum_{s,t=1}^{n} a_{s,t}(X_s X_t))|^2} \right\|_{p/2}^{p/2}$$

$$\le \mathbb{E}\left[\sum_{r=-\infty}^{n} |P_r(\sum_{s,t=1}^{n} a_{s,t}(X_sX_t))|^{p/2}\right]$$
$$\le C_p\Theta_{0,p}^{p/2}\mathcal{C}^{p/2}\sum_{r=-\infty}^{n}\left(\sum_{s=1}^{n}\delta_p(s-r)\right)^{p/2}$$
$$\le C_p\mathcal{C}^{p/2}n\Theta_{0,p}^{p}.$$

This completes the proof. □

Finally we will need a Fuk-Nagaev type inequality.

LEMMA 10.3. *Let $Z_1,\dots,Z_n$ be independent zero-mean random variables with $\sup_i \mathbb{E}[|Z_i|^p] < \infty$ for $p > 1$. Let $S_i = \sum_{j=1}^{i} Z_j$. Then, for any $x > 0$,*

$$\mathbb{P}\left(\max_{1\le i\le n}|S_i| \ge x\right) \le \begin{cases} C_p x^{-p}\sum_{i=1}^{n}\mathbb{E}[|Z_i|^p], & 1 < p \le 2, \\ C_p x^{-p}\sum_{i=1}^{n}\mathbb{E}[|Z_i|^p] + \exp\left(-\frac{x^2}{C\sum_{i=1}^{n}\mathbb{E}[Z_i^2]}\right), & p > 2. \end{cases} \tag{10.14}$$

PROOF. First we prove the following:

$$\mathbb{P}\left(|S_n| \ge x\right) \le \begin{cases} C_p x^{-p}\sum_{i=1}^{n}\mathbb{E}[|Z_i|^p], & 1 < p \le 2, \\ C_p x^{-p}\sum_{i=1}^{n}\mathbb{E}[|Z_i|^p] + \exp\left(-\frac{x^2}{C\sum_{i=1}^{n}\mathbb{E}[Z_i^2]}\right), & p > 2. \end{cases} \tag{10.15}$$

The cases $p > 2$ and $1 < p \le 2$ follows directly from Corollary 1.8 and 1.6 of [75] respectively. Now, in view of Borovkov [11]'s argument, the left-hand side $\mathbb{P}(|S_n| \ge x)$ in Theorems 1-4 in [39] can be replaced by $\mathbb{P}\left(\max_{1\le i\le n}|S_i| \ge x\right)$. Thus (10.15) implies (10.14). □

PROOF OF THEOREM 3.1. In the following, any constant C_p depends solely on p. Let $r_p = \max\{1/2 + 2/p, 1\}$ and $U_n = \lceil n/\mathcal{D}_n\rceil$. If $U_n = 1$, the conclusion readily follows from Markov inequality. Therefore let $U_n \ge 2$.

Denote by $\boldsymbol{\eta}_k = (\varepsilon_{(k-1)\mathcal{D}_n+1},\dots,\varepsilon_{k\mathcal{D}_n})$. Recall V_k from (3.3). Let $V_{k,\tau} = \mathbb{E}[V_k|\boldsymbol{\eta}_k,\dots,\boldsymbol{\eta}_{k-\tau}]$. Let $L_n = \lfloor \log U_n/\log 2\rfloor$. We will omit the subscript n from U_n and L_n for presentation purposes, their dependence on n being implicit. Let $\tau_l = 2^l$, $1 \le l \le L-1$, and $\tau_L = U$. Let for $1 \le k \le U$, and $1 \le l \le L$

$$M_{k,l} = \sum_{j=1}^{k}(V_{j,\tau_l} - V_{j,\tau_{l-1}}). \tag{10.16}$$

Define $D_k = \sum_{j=1}^{k} V_j$ for $1 \le k \le U$, and let $D_{k,\tau} = \mathbb{E}[D_k|\boldsymbol{\eta}_k,\dots,\boldsymbol{\eta}_{k-\tau}]$. Note that

$$D_k - \mathbb{E}(D_k) = \sum_{j=1}^{k}(V_j - V_{j,n}) + \sum_{l=2}^{L} M_{k,l} + \sum_{j=1}^{k}(V_{j,2} - \mathbb{E}(V_{j,2})). \tag{10.17}$$

Thus,

$$\max_{1\le k\le U}|D_k - \mathbb{E}(D_k)| \le \max_{1\le k\le U}|\sum_{j=1}^{k}(V_j - V_{j,U})| + \sum_{l=2}^{L}|\max_{1\le k\le U} M_{k,l}| + |\max_{1\le k\le U}\sum_{j=1}^{k}(V_{j,2} - \mathbb{E}(V_{j,2}))|. \tag{10.18}$$

For the first term in the above sum, note that

$$\left\|\max_{1\le k\le U}|D_k-D_{k,U}|\right\|_{p/2}\le\|D_U-D_{U,U}\|_{p/2}+\left\|\max_{1\le i\le U-1}\left|\sum_{k=U-i}^{U}(V_k-V_{k,U})\right|\right\|_{p/2}. \tag{10.19}$$

Now, $V_k-V_{k,U}=\sum_{i=U+1}^{\infty}(V_{k,i}-V_{k,i-1})$. Since $V_{k,i}-V_{k,i-1}$ are martingale difference with respect to $\sigma(\boldsymbol{\eta}_{k-i},\boldsymbol{\eta}_{k-i+1},\ldots)$. Hence, using Doob's Inequality,

$$\left\|\max_{1\le i\le U-1}\left|\sum_{k=U-i}^{U}(V_{k,j}-V_{k,j-1})\right|\right\|_{p/2}\le C_p\|D_{U,j}-D_{U,j-1}\|_{p/2}\,. \tag{10.20}$$

Therefore, Lemma 10.1 along with Burkholder inequality ([14] for $2<p<4$ along with $(|a_1|+\ldots+|a_n|)^{p/4}\le|a_1|^{p/4}+\ldots+|a_n|^{p/4}$, and [92]'s version for $p>4$) implies,

$$\|D_{U,j}-D_{U,j-1}\|_{p/2}\le\begin{cases}C_pU^{2/p}\mathcal{D}_n^{1/2+2/p}\Theta_{0,p}\sum_{d=(j-1)\mathcal{D}_n+1}^{(j+1)\mathcal{D}_n}\delta_p(d), & 2<p\le4\\ C_p\sqrt{U}\mathcal{D}_n\Theta_{0,p}\sum_{d=(j-1)\mathcal{D}_n+1}^{(j+1)\mathcal{D}_n}\delta_p(d), & p>4.\end{cases} \tag{10.21}$$

Therefore, using

$$\left\|\max_{1\le i\le U-1}\left|\sum_{k=U-i}^{U}(V_k-V_{k,U})\right|\right\|_{p/2}\le\sum_{j=U+1}^{\infty}\left\|\max_{1\le i\le U-1}\left|\sum_{k=U-i}^{U}(V_{k,j}-V_{k,j-1})\right|\right\|_{p/2},$$

we have,

$$\left\|\max_{1\le i\le U-1}\left|\sum_{k=U-i}^{U}(V_k-V_{k,n})\right|\right\|_{p/2}\le\begin{cases}C_pU^{2/p}\mathcal{D}_n^{1/2+2/p}\Theta_{0,p}\Theta_{U\mathcal{D}_n+1}, & 2<p\le4\\ C_p\sqrt{U}\mathcal{D}_n\Theta_{0,p}\Theta_{U\mathcal{D}_n+1} & p>4.\end{cases}$$

Now, recall (2.5). Note that from Condition 2.1,

$$\Theta_{U\mathcal{D}_n+1,p}\le C(U\mathcal{D}_n+1)^{-A}\mu_{p,A}\le Cn^{-A}\mu_{p,A},$$

which yields, under condition 2.1,

$$\left\|\max_{1\le i\le U-1}\left|\sum_{k=U-i}^{U}(V_k-V_{k,n})\right|\right\|_{p/2}\le\begin{cases}C_pU^{2/p}\mathcal{D}_n^{1/2+2/p}\mu_{p,A}^2n^{-A}, & 2<p\le4\\ C_p\sqrt{U}\mathcal{D}_n\mu_{p,A}^2n^{-A} & p>4.\end{cases}. \tag{10.22}$$

Proceeding similarly,

$$\||D_U-D_{U,U}|\|_{p/2}\le\begin{cases}C_pU^{2/p}\mathcal{D}_n^{1/2+2/p}\mu_{p,A}^2n^{-A}, & 2<p\le4\\ C_p\sqrt{U}\mathcal{D}_n\mu_{p,A}^2n^{-A} & p>4.\end{cases}.$$

Hence, by Markov Inequality,

$$\mathbb{P}\left(\max_{1\le k\le U}|D_k-D_{k,U}|\ge x\right)\lesssim\begin{cases}x^{-p/2}n^{1-Ap/2}\mathcal{D}_n^{p/4}\mu_{p,A}^p, & 2<p\le4\\ x^{-p/2}n^{p/4-Ap/2}\mathcal{D}_n^{p/4}\mu_{p,A}^p, & p>4.\end{cases} \tag{10.23}$$

For the second term in (10.18), define the following quantities:

$$Y_{h,l} = \sum_{j=(h-1)\tau_l+1}^{(h\tau_l)\wedge U} (V_{j,\tau_l} - V_{j,\tau_{l-1}}), \quad 1 \le h \le \lceil U/\tau_l \rceil := U_0, \tag{10.24}$$

$$R^e_{s,l} = \sum_{h \text{ even}}^{s} Y_{h,l} \ , \ R^o_{k,l} = \sum_{h \text{ odd}}^{s} Y_{h,l}. \tag{10.25}$$

Further let $\{\lambda_j\}_{1\le j\le L}$ be such a sequence that $\sum_{l=1}^{L} \lambda_l \le 1$. For some $s \in \mathbb{N}$, denote by $s_l := s\tau_l \wedge U$. Therefore,

$$\begin{aligned}\mathbb{P}(\max_{1\le k\le U} |M_{k,l}| \ge 3\lambda_l x) \le \mathbb{P}(\max_{1\le s\le U_0} |R^e_{s,l}| \ge \lambda_l x) + \mathbb{P}(\max_{1\le s\le U_0} |R^o_{s,l}| \ge \lambda_l x) + \\ \sum_{s=1}^{U_0} \mathbb{P}(\max_{s_l+1\le j\le (s+1)_l} |M_{j,l} - M_{s_l,l}| \ge \lambda_l x).\end{aligned} \tag{10.26}$$

For the first two terms in (10.26), note that $Y_{h_1,l}$ and $Y_{h_2,l}$ are independent for $|h_1 - h_2| \ge 2$. Therefore, using Lemma 10.3, we obtain

$$\mathbb{P}(\max_{1\le s\le U_0} |R^e_{s,l}| \ge \lambda_l x) \le \begin{cases} C_p \frac{\sum_{h \text{ even}} \mathbb{E}[|Y_{h,l}|^{p/2}]}{(\lambda_l x)^{p/2}}, & 2 < p \le 4, \\ C_p \frac{\sum_{h \text{ even}} \mathbb{E}[|Y_{h,l}|^{p/2}]}{(\lambda_l x)^{p/2}} + 2\exp\left(-C_p \frac{(\lambda_l x)^2}{\sum_{h \text{ even}} \mathbb{E}[|Y_{h,l}|^2]}\right), & p > 4, \end{cases} \tag{10.27}$$

where C_p are constants depending on p. An argument similar to (10.20) and (10.22) yields,

$$\|Y_{h,l}\|_{p/2} \lesssim \begin{cases} \tau_l^{2/p} \mathcal{D}_n^{1/2+2/p} (\tau_l \mathcal{D}_n)^{-A} \mu^2_{p,A}, & 2 < p \le 4, \\ \sqrt{\tau_l} \mathcal{D}_n (\tau_l \mathcal{D}_n)^{-A} \mu^2_{p,A}, & p > 4. \end{cases}$$

Thus,

$$\begin{aligned}&\mathbb{P}(\max_{1\le s\le U_0} |R^e_{s,l}| \ge \lambda_l x) \\ &\le \begin{cases} C_p (\lambda_l x)^{-p/2} \frac{U}{\tau_l} \tau_l^{1-Ap/2} \mathcal{D}_n^{p/4+1-Ap/2} \mu^p_{p,A}, & 2 < p \le 4 \\ C_p (\lambda_l x)^{-p/2} \frac{U}{\tau_l} \tau_l^{p/4-Ap/2} \mathcal{D}_n^{p/2-Ap/2} \mu^p_{p,A} + 2\exp\left(-C_p \frac{(\lambda_l x)^2}{\frac{U}{\tau_l} \tau_l \mathcal{D}_n^2 (\tau_l \mathcal{D}_n)^{-2A} \mu^4_{4,A}}\right), & p > 4 \end{cases} \\ &\lesssim \begin{cases} (\lambda_l x)^{-p/2} \tau_l^{-Ap/2} n \mathcal{D}_n^{p/4-Ap/2} \mu^p_{p,A}, & 2 < p \le 4 \\ (\lambda_l x)^{-p/2} \tau_l^{p/4-Ap/2-1} n \mathcal{D}_n^{p/2-Ap/2-1} \mu^p_{p,A} + 2\exp\left(-C_p \frac{(\lambda_l x)^2 (\tau_l \mathcal{D}_n)^{2A}}{n \mathcal{D}_n \mu^4_{4,A}}\right), & p > 4. \end{cases}\end{aligned} \tag{10.28}$$

A similar inequality holds for $\max_{1\le s\le U_0} |R^o_{s,l}|$. To tackle the third term $\sum_{s=1}^{U_0} \mathbb{P}(\max_{s_l+1\le j\le (s+1)_l} |M_{j,l} - M_{s_l,l}| \ge \lambda_l x)$ in (10.26), we employ an argument similar to (10.19) through (10.23). Write

$$\left\| \max_{s_l+1\le j\le (s+1)_l} |M_{j,l} - M_{s_l,l}| \right\|_{p/2} \le \|M_{(s+1)_l} - M_{s_l,l}\|_{p/2} +$$

$$\left\| \max_{s_l+1 \le j \le (s+1)_l} \left| \sum_{k=j}^{(s+1)_l} (V_{k,\tau_l} - V_{k,\tau_{l-1}} - V_{s_l,\tau_l} + V_{s_l,\tau_{l-1}}) \right| \right\|_{p/2}.$$

Using Doob's Inequality,

$$\left\| \max_{s_l+1 \le j \le (s+1)_l} \left| \sum_{k=j}^{(s+1)_l} (V_{k,\tau_l} - V_{k,\tau_{l-1}} - V_{s_l,\tau_l} + V_{s_l,\tau_{l-1}}) \right| \right\|_{p/2} \le \frac{p/2}{p/2-1} \|M_{(s+1)_l} - M_{s_l,l}\|_{p/2}.$$

An argument similar to (10.21) yields,

$$\|M_{(s+1)_l} - M_{s_l,l}\|_{p/2} \lesssim \begin{cases} \tau_l^{2/p} \mathcal{D}_n^{1/2+2/p} (\tau_l \mathcal{D}_n)^{-A} \mu_{p,A}^2, & 2 < p \le 4, \\ \sqrt{\tau_l} \mathcal{D}_n (\tau_l \mathcal{D}_n)^{-A} \mu_{p,A}^2, & p > 4, \end{cases}$$

Therefore, applying Markov Inequality implies

(10.29)
$$\sum_{s=1}^{U_0} \mathbb{P}\left(\max_{s_l+1 \le j \le (s+1)_l} |M_{j,l} - M_{s_l,l}| \ge \lambda_l x \right) \lesssim \begin{cases} (\lambda_l x)^{-p/2} \tau_l^{-Ap/2} n \mathcal{D}_n^{p/4-Ap/2} \mu_{p,A}^p, & 2 < p \le 4 \\ (\lambda_l x)^{-p/2} \tau_l^{p/4-Ap/2-1} n \mathcal{D}_n^{p/2-Ap/2-1} \mu_{p,A}^p, & p > 4. \end{cases}$$

Thus, combining (10.28) and (10.29) in (10.26), we get,

$$\mathbb{P}(\max_{1 \le k \le U} |M_{k,l}| \ge 3\lambda_l x) \lesssim \begin{cases} (\lambda_l x)^{-p/2} \tau_l^{-Ap/2} n \mathcal{D}_n^{p/4-Ap/2} \mu_{p,A}^p, & 2 < p \le 4 \\ (\lambda_l x)^{-p/2} \tau_l^{p/4-Ap/2-1} n \mathcal{D}_n^{p/2-Ap/2-1} \mu_{p,A}^p & \\ \qquad + 2\exp\left(-C_p \frac{(\lambda_l x)^2 (\tau_l \mathcal{D}_n)^{2A}}{n \mathcal{D}_n \mu_{4,A}^4} \right), & p > 4. \end{cases} \tag{10.30}$$

Using (10.30), we have for the second term in (10.18),

$$\begin{aligned} \mathbb{P}\left(\sum_{l=2}^{L} |\max_{1 \le k \le U} M_{k,l}| \ge 3x \right) &\le \sum_{l=2}^{L} \mathbb{P}(\max_{1 \le k \le U} |M_{k,l}| \ge 3\lambda_l x) \\ &\lesssim \begin{cases} x^{-p/2} n \mathcal{D}_n^{p/4-Ap/2} \mu_{p,A}^p \cdot I_1 & 2 < p \le 4 \\ x^{-p/2} n \mathcal{D}_n^{p/2-Ap/2-1} \mu_{p,A}^p \cdot I_2 + 4 \cdot II, & p > 4, \end{cases} \end{aligned} \tag{10.31}$$

where

$$\begin{aligned} I_1 &= \sum_{l=2}^{L} \lambda_l^{-p/2} \tau_l^{-Ap/2} \\ I_2 &= \sum_{l=2}^{L} \lambda_l^{-p/2} \tau_l^{p/4-Ap/2-1} \\ II &= \sum_{l=2}^{L} \exp\left(-C_p \frac{(\lambda_l x)^2 (\tau_l \mathcal{D}_n)^{2A}}{n \mathcal{D}_n \mu_{4,A}^4} \right). \end{aligned}$$

Let $\lambda_l = (1/l^2)/(\pi^2/3)$ for $1 \le l \le L/2$, and $\lambda_l = (1/(L+1-l)^2)/(\pi^2/3)$ for $L/2 < l \le L$. Clearly $\sum_{l=1}^{L} \lambda_l \le 1$. With our choice of λ_l and τ_l, elementary calculation using $A > 1/2 - 1/p$ and $\min_{l \ge 1} \lambda_l^2 \tau_l^{2A} > 0$ shows that there exists a constant C such that

$$I_1 \le C; I_2 \le C;\ II \le C \exp\left(-C_p \frac{x^2}{\mu_{4,A}^4 n \mathcal{D}_n^{1-2A}}\right). \tag{10.32}$$

Putting (10.32) in (10.31), one obtains

$$\mathbb{P}\left(\sum_{l=2}^{L} |\max_{1\le k\le n} M_{k,l}| \ge 3x\right) \lesssim \begin{cases} x^{-p/2}\mu_{p,A}^p n \mathcal{D}_n^{p/4} & 2 < p \le 4 \\ x^{-p/2}\mu_{p,A}^p n \mathcal{D}_n^{p/2-1} + 4\exp\left(-C_p \frac{x^2}{n\mathcal{D}_n^{1-2A}\mu_{4,A}^4}\right), & p > 4. \end{cases} \tag{10.33}$$

Now finally we tackle the third term in (10.18). Note that as $\boldsymbol{\eta}_k$'s are independent, hence $V_{k,2}$ and $V_{k',2}$ are independent if $|k-k'| > 2$. We again employ Lemma 10.3 and techniques similar to (10.24), (10.25) and (10.27) to obtain,

$$\mathbb{P}\left(|\max_{1\le k\le U} \sum_{j=1}^{k} (V_{j,2} - \mathbb{E}(V_{j,2}))| \ge x\right)$$
$$\lesssim \begin{cases} x^{-p/2}\sum_{j=1}^{U} \mathbb{E}(|V_{j,2} - \mathbb{E}(V_{j,2})|^{p/2}), & 2 < p \le 4, \\ x^{-p/2}\sum_{j=1}^{U} \mathbb{E}(|V_{j,2} - \mathbb{E}(V_{j,2})|^{p/2}) + \\ \qquad 2\exp\left(-c\frac{x^2}{\sum_{j \text{ even}} \mathbb{E}(|V_{j,2} - \mathbb{E}(V_{j,2})|^2)}\right) + 2\exp\left(-c_p\frac{x^2}{\sum_{j \text{ odd}} \mathbb{E}(|V_{j,2} - \mathbb{E}(V_{j,2})|^2)}\right), & p > 4. \end{cases}$$

By conditional Jensen's inequality and Lemma 10.2 (noting that $\mathcal{C} = O(\mathcal{D}_n)$), we get

$$\mathbb{E}(|V_{j,2} - \mathbb{E}(V_{j,2})|^{p/2}) \le \mathbb{E}(|V_j - \mathbb{E}(V_j)|^{p/2}) \lesssim \begin{cases} \left(\mathcal{D}_n^{1/2+2/p}\right)^{p/2} \mu_{p,A}^p, & 2 < p \le 4, \\ \mathcal{D}_n^{p/2}\mu_{p,A}^p, & p > 4, \end{cases}$$

which, yields

$$\mathbb{P}\left(|\max_{1\le k\le U} \sum_{j=1}^{k} (V_{j,2} - \mathbb{E}(V_{j,2}))| \ge x\right) \lesssim \begin{cases} x^{-p/2} n \mathcal{D}_n^{p/4}\mu_{p,A}^p, & 2 < p \le 4 \\ x^{-p/2} n \mathcal{D}_n^{p/2-1}\mu_{p,A}^p + 4\exp\left(-C_p\frac{x^2}{n\mathcal{D}_n\mu_{4,A}^4}\right), & p > 4. \end{cases} \tag{10.34}$$

Combining (10.23), (10.33) and (10.34), we have the result. □

10.2. *Proof of Lemma 3.1.*

PROOF. Define the projection operator P_i as $P_i X = \mathbb{E}[X|\mathcal{F}_i] - \mathbb{E}[X|\mathcal{F}_{i-1}]$ where $\mathcal{F}_i = \sigma(\ldots, \varepsilon_{i-1}, \varepsilon_i)$. Note that for $l > k$,

$$\begin{aligned} |\mathbb{E}(X_k X_l)| &= |\sum_{i\in\mathbb{Z}}\sum_{j\in\mathbb{Z}} \mathbb{E}\left[(P_i X_k)(P_j X_l)\right]| \le \sum_{i\in\mathbb{Z}} \|P_i(X_k)\|\|P_i(X_l)\| \\ &\le \sum_{i=\infty}^{k} \delta_p(k-i)\delta_p(l-i) = \sum_{i=0}^{\infty} \delta_p(i)\delta_p(i+l-k). \end{aligned} \tag{10.35}$$

Using (10.35) repeatedly,

$$\begin{aligned}
&\max_{1\le k\le\lfloor\frac{n}{m}\rfloor}|\mathbb{E}(B_kB_{k+1})| \\
&\lesssim \sum_{j=0}^{2m}(m-|m-j|)\sum_{i=0}^{\infty}\delta_p(i)\delta_p(i+j) \\
&\le \sum_{i=0}^{\infty}\delta_p(i)\left(\Theta_{i+1,p}+\Theta_{i+2,p}+\ldots+\Theta_{i+2m-1,p}\right) \\
&\le \sum_{i=0}^{\infty}\delta_p(i)\sum_{j=1}^{2m-1}\Theta_{j,p}=\Theta_{0,p}\sum_{j=1}^{2m-1}\Theta_{j,p}=\mu_{p,A}O\left(\sum_{j=1}^{2m-1}(j+1)^{-A}\right)=O(1),
\end{aligned}$$

since $A>1$ in Condition 2.1. Moreover, for fixed i,j, via an exact same argument as above, one obtains,

$$|\mathbb{E}(B_iB_j)|\le\Theta_{0,p}\sum_{k=|i-j-1|m+1}^{|i-j+1|m-1}\Theta_{j,p}. \tag{10.36}$$

Then (10.36) in conjunction with (2.5) directly implies that

$$\begin{aligned}
\sup_{1\le k\le\lfloor n/m\rfloor}\sum_{i:|i-k|>2}|\mathbb{E}(B_iB_k)| &\lesssim_d \Theta_{0,p}\sup_{1\le k\le\lfloor n/m\rfloor}\sum_{i=1}^{k-2}\sum_{j=im+1}^{(i+2)m-1}\Theta_{j,p} \\
&\le 2\Theta_{0,p}\sum_{j=m+1}^{\infty}\Theta_{j,p}=O(m^{1-A}).
\end{aligned}\tag{10.37}$$

This completes the proof of (3.7). □

10.3. *Proof of Theorem 3.2.* We will define L and α as in the proof of Theorem 2.4. Let $\nu=\min\{(1+A)/(2+4A),(1+4A/p)/(2+4A)\}$. The theorem follows trivially from Theorem 2.4 if $A>A_0$. Thus let $A\le A_0$. Specifically, with $1<A\le A_0$, our choice of L and α satisfies the following, which will be used in our proofs:

$$\frac{1}{2}-\nu-\frac{LA}{2}<0, \tag{10.38}$$

$$L\left(\frac{\alpha}{2}-1\right)+1-\alpha\nu<0, \tag{10.39}$$

$$\alpha\ge\max\{p,2(1+p+pA)/3\}, \tag{10.40}$$

$$1/p-1/\alpha+L-L(A+1)p/\alpha=0. \tag{10.41}$$

We will need a slightly different version of Lemma 8.3. To avoid confusion, we state and prove it separately.

Lemma 10.4. *Assume Conditions 2.1 and 2.2, along with* (10.38), (10.39), (10.40) *and* (10.41) *for A, L and α. Let $m=\lfloor n^L\rfloor$ and let*

$$\tilde{R}_{s,t}=\tilde{X}_s+\ldots+\tilde{X}_t,$$

where $\tilde{X}_i$ is as defined in (8.22). Then

$$\max_s \mathbb{E}\left[\max_{1\le t\le m} |\tilde{R}_{s,t}|^\alpha\right] = o(mn^{\alpha\nu-1}). \tag{10.42}$$

PROOF. The proof of this lemma is almost same as that of Lemma 8.3. The only point of differences are the use of (10.39) instead of (8.3), as well as a different treatment of the term III in (8.10). The latter difference is necessitated as we no longer have $\alpha < 2(1+p+pA)/3$ as we had in (8.4).

In fact for term III we will proceed as follows. For the case $\alpha > 2(1+p+pA)/3$, using (10.40), and using same argument as (8.14), one obtains

$$\begin{aligned} III = m^{1/\alpha}\sum_{j=1}^m j^{1/2-1/\alpha}\tilde{\delta}_\alpha(j) &\lesssim m^{1/\alpha}n^{1/p-1/\alpha}\sum_{l=1}^{\lfloor \log_2 m\rfloor}\sum_{j=2^l}^{2^{l+1}-1} j^{1/2-1/\alpha}\delta_p(j)^{p/\alpha} \\ &\le m^{1/\alpha}n^{1/p-1/\alpha}\sum_{l=1}^{\lfloor \log_2 m\rfloor} 2^{l(3/2-1/\alpha-p/\alpha)}O(2^{-lAp/\alpha}) \\ &\le m^{3/2-p/\alpha-Ap/\alpha}n^{1/p-1/\alpha} = m^{1/2}, \end{aligned} \tag{10.43}$$

where the last equality follows from (10.41). Therefore, in view of (10.39), we obtain

$$\frac{n^{1-\alpha\nu}}{m} III^\alpha = n^{1-\alpha\nu}m^{-1}O(m^{\alpha/2}) = o(1). \tag{10.44}$$

In case $\alpha = 2(1+p+pA)/3$, same treatment as (10.43) yields,

$$III \lesssim m^{1/\alpha}n^{1/p-1/\alpha}\log_2 m \lesssim Lm^{1/\alpha}n^{1/p-1/\alpha}\log_2 n. \tag{10.45}$$

One immediately obtains,

$$\frac{n^{1-\alpha\nu}}{m} III^\alpha = L(\log_2 n)n^{\alpha/p-\alpha\nu}O(1) = o(1), \tag{10.46}$$

where the last assertion is due to $\nu > 1/p$. This completes the proof of this lemma. □

PROOF OF THEOREM 3.2. The proof follows mostly along the lines of the proof of Theorem 2.4, with $S_i^\oplus$ and $\tilde{S}_i$ defined as in that proof. We list below the point of differences from that proof.

- As above, use (10.38) and Lemma 10.4 instead of whenever (8.2) and Lemma 8.3 is used in the proof of Theorem 2.4.
- Proposition 8.2 now holds with a rate of n^ν.
- Proposition 8.3 also holds with a rate of n^ν. For the proof, we will investigate $\mathbb{P}(\max_{1\le i\le n}|\tilde{S}_i - S_i^\oplus| > n^{1/4}\varepsilon)$.
- Instead of (8.32), we will reach a rate of n^ν using $x = n^\nu$ and Lemma 10.4 in the previous step.
- Investigate $\mathbb{P}(\max_{1\le k\le K}|\tilde{V}_{2k}(\boldsymbol{\eta}_{3k})| \ge cn^{2\nu})$ to obtain a rate of n^ν instead of (8.35).

□

11. Appendix C: Additional Lemma for Theorem 4.1. Here we will prove a technical lemma required to bound the total variation of the weights for the local linear estimate. This lemms also helps control the bias of the estimate $\hat{\mu}_{h_n}(t)$.

LEMMA 11.1 (Consistency). *Let $S_j(t)$ be defined as in (4.7). Then with $h_n \to 0$ and $nh_n \to \infty$, under the assumptions of Theorem 4.1, it holds that*

$$\sup_{t\in[\omega h_n, 1-\omega h_n]} \left| \frac{S_j(t)}{nh_n^{j+1} f(t)} \right| = m_j + o(1), \text{ for } j = 0, 1, 2, \tag{11.1}$$

with $m_0 = 1$, $m_1 = 0$ and $m_2 = 2\beta = \int u^2 K(u)du$. Moreover, for Ω_n in (4.2) with $w_{h_n}(t,i)$ as in (4.8), it holds that $\Omega_n = O(n^{-1}h_n^{-1})$.

PROOF. Observe that

$$S_j(t) = \int_0^n \left(F^{-1}(\frac{\lfloor 1+u \rfloor}{n}) - t\right)^j K\left(\frac{F^{-1}(\lfloor 1+u \rfloor/n) - t}{h_n}\right) \mathrm{d}u.$$

Let $m_j = \int v^j K(v)\mathrm{d}v$. Note that $m_0 = 1, m_1 = 0$ and $m_2 = 2\beta$. Consider the corresponding smoothed version

$$\tilde{S}_j(t) = \int_0^n \left(F^{-1}(\frac{u}{n}) - t\right)^j K\left(\frac{F^{-1}(u/n) - t}{h_n}\right) \mathrm{d}u.$$

Let $v = (F^{-1}(u/n) - t)/h_n$. Also let g be such that $g(v) = (F^{-1}(\lfloor 1+u \rfloor/n) - t)/h_n$. Clearly, since $C_1 \le f(t) \le C_2$ for all t, therefore, $|g(v) - v| = O(n^{-1}h_n^{-1})$ for all t. Now, by change of variables techniques and noting that $t \in [\omega h_n, 1-\omega h_n]$, it holds that

$$S_j(t) - \tilde{S}_j(t) = nh_n^{j+1} \int_{-\omega}^{\omega} \left[(g(v))^j K(g(v)) - v^j K(v)\right] f(vh_n + t)\mathrm{d}v.$$

For $j = 0$, $S_j(t) - \tilde{S}_j(t) = O(h_n^j)$ follows from (4.9) directly. For $j = 1, 2$, note that $(g(v))^j - v^j = O(n^{-1}h_n^{-1})$ since $v \in [-\omega, \omega]$. Therefore, again invoking (4.9) yields that

$$S_j(t) - \tilde{S}_j(t) = O(h_n^j). \tag{11.2}$$

Finally, observing $f(vh_n + t) = f(t) + O(vh_n)$,

$$\begin{aligned} \tilde{S}_j(t) &= nh_n^{j+1} \int_{-\omega}^{\omega} v^j K(v) f(vh_n + t)\mathrm{d}v \\ &= nh_n^{j+1} m_j f(t) + O(nh_n^{j+2}), \end{aligned} \tag{11.3}$$

since $c \le f(t) \le C$. Thus, using $h_n \to 0$ and $nh_n \to \infty$, it holds that

$$\begin{aligned} S_j(t) &= \int_0^n (F^{-1}(\frac{u}{n}) - t)^j K(\frac{F^{-1}(u/n) - t}{h_n}) \mathrm{d}u + O(h_n^j) \\ &= nf(t)h_n^{j+1}(m_j + o(1)). \end{aligned}$$

which completes the proof of (11.1). To observe $\Omega_n = O(n^{-1}h_n^{-1})$, note that for fixed t, $w_n(t,i) = 0$ unless $t_i \in [t - \omega h_n, t + \omega h_n]$, and therefore in (4.8), $|t - t_i| = O(h_n)$. Putting the approximations of $S_j(t)$ in (4.8), and noting that K is bounded, one obtains $|w_n(t,1)| = O(n^{-1}h_n^{-1})$. On the other hand, $\sum_{i=2}^n |w_i(t) - w_{i-1}(t)|$ can be bounded by $O(n^{-1}h_n^{-1}) + O(1/(nh_n)^2)$ by noting that $|\sum_{i=1}^n ((t-t_i)K(t-t_i/h_n) - (t-t_{i-1})K(t-t_{i-1})/h_n)| = O(h_n)$. This completes the proof. $\square$

12. Appendix D: Additional Simulations for Section 5.

12.1. *Empirical Accuracy for Gaussian approximation.* In this section we further explore the performance of the models 5.3 and 5.4. In addition to N(0,1) innovations, we will also consider suitably normalized t_6 innovations (subsequently we will omit "normalized" while describing the errors). Figures 10-11 depict the “typical” Q-Q plots of 1000 data-based bootstrap samples of $\mathbb{B}(\mathcal{T}_i)$, $\mathcal{W}(\mathcal{T}_i^-)$ and $\mathbb{B}(\mathcal{T}_i^\diamond)$ against the theoretical quantiles of $\max_{1\le i\le n} S_i$ (based on 1000 monte carlo samples) for different settings. The conclusions reflect those of Figures 5 and 6, and justify our use of $\mathcal{T}_i$ as a plug-in estimate for $\mathbb{E}[S_i^2]$.

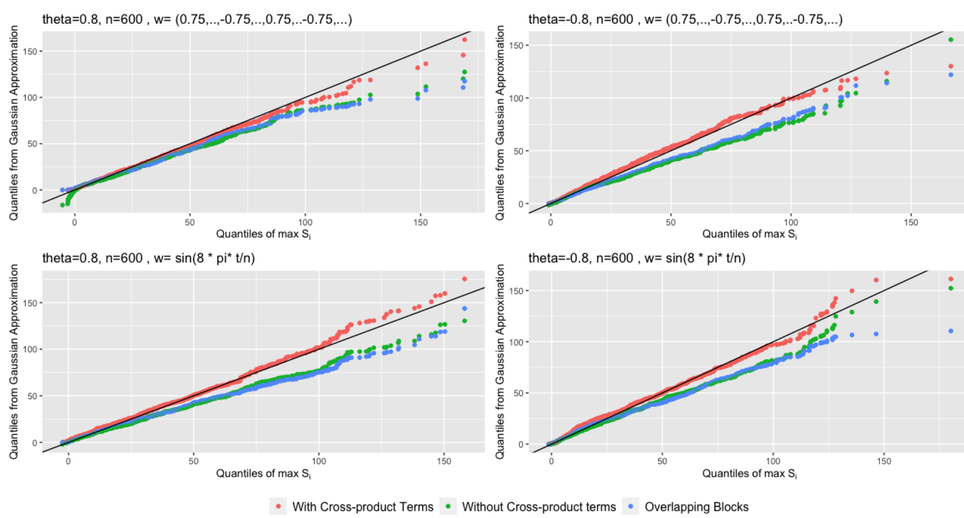


Figure 10: Comparison of theoretical quantiles with the bootstrap Gaussian approximation quantiles based on $X_1, \ldots, X_n \sim$ Model (5.3) with t_6 innovations, with and without cross-product terms.

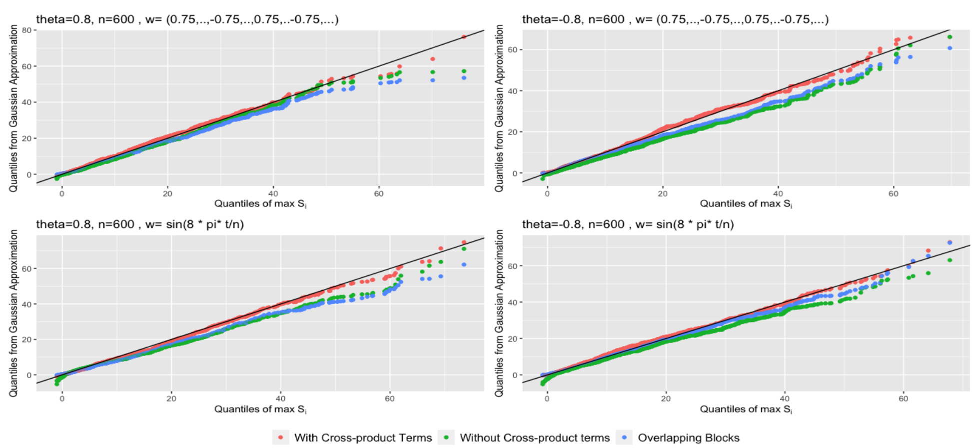


Figure 11: Comparison of theoretical quantiles with the bootstrap Gaussian approximation quantiles based on $X_1, \ldots, X_n \sim$ Model (5.4) with t_6 innovations, with and without cross-product terms.

12.2. *Change-point detection: Simulation based on Bootstrap.* In this section we further explore the performance of the bootstrap-based test as described in Section 5.3.2 for t_6 innovations. We consider $X_t \sim$ Model 5.3 and 5.4 respectively. Figures 12-13 show the plots corresponding to Figures 7 and 8 for each of the models 5.3 and 5.4.

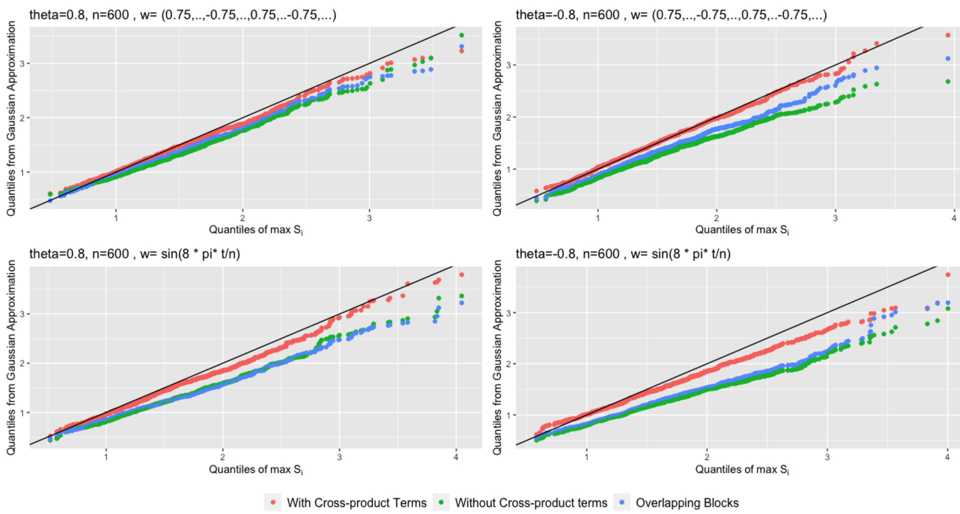


Figure 12: Comparison of theoretical quantiles of CUSUM statistic U_n with quantiles of bootstrap Gaussian approximation of CUSUM quantiles based on $X_1, \ldots, X_n \sim$ Model (5.3) with t_6 innovations, with and without cross-product terms.

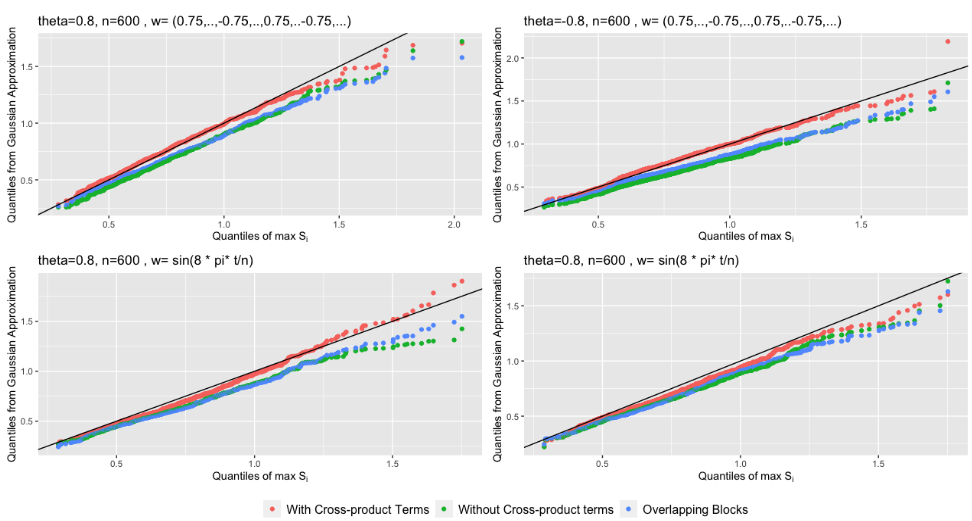


Figure 13: Comparison of theoretical quantiles of CUSUM statistic U_n with quantiles of bootstrap Gaussian approximation of CUSUM quantiles based on $X_1, \ldots, X_n \sim$ Model (5.4) with t_6 innovations, with and without cross-product terms.