2017
arXiv: arXiv:0000.0000

ASYMPTOTIC THEORY FOR TIME-SERIES WITH CYCLIC AND TREND COMPONENTS

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

University of Chicago, November 30, 2017

Analyzing periodic pattern is a topic of interest in various fields ranging from signal processing to environment science. Processes with intrinsic periodicity are often referred in literature as cyclostationary processes. Apart from the cyclic component, a smooth trend is often added to the model description and whether it is of any particular parametric form is sought from an a posteriori data. In this paper, we study inference on mean zero cyclostationary processes in presence of a Hölder-α continuous trend and a possibly non-stationary dependent error. We view the dependence within the error/noise process through the framework of functional dependence measure proposed by Wu (2005, [8]). In particular, we obtain consistent estimator of the period, periodic component and a simultaneous confidence band of the unknown trend function. Our theory substantially generalizes the earlier ones by relaxing the strong mixing conditions of the noise process and putting some mild and easily verifiable conditions instead. It also relaxes the smoothness assumption on the trend function. To substantiate our theoretical results, we perform some simulation studies and conclude by applying the methods to two types of climate data.

Keywords: Trend estimation, cyclostationary process, Hölder-continuous function, Gaussian approximation, Weak dependence, Nonlinear time series, Simultaneous confidence band, Bandwidth, Bias, Variance

MSC 2000:

1. Introduction. Natural processes have a tendency to show periodic pattern over time. Our failure to observe or measure an exact periodic sequence is due to different types of noises that perturb the original periodic process. Such processes commonly arise in climatology, tele-communications and signal processing fields. Various scientific phenomena where we measure random data showing rhythmic or seasonal variation comes under this wide class of processes. These processes, hinting at their underlying periodical functionality, are widely termed as cyclostationary processes. These processes form a special class of non-stationary processes and can be considered as a bridge between the stationary and non-stationary processes.

1.1. *Brief historical background for cyclostationary process.* Gudzenko (1959, [**?**]), Gladyshev (1963, [**?**]), Kayatskas (1968, [**?**]) etc. were notable among the first to contribute to the literature of cyclostationary stochastic processes. Hurd (1969, [?]) and Gardner (1972, [?]) are also considered as pioneers in the research of second-order cyclostationary processes. Boyles et al. (1983, [1]) and Gardner (1985, [?], 1990, [?]) discussed second-order cyclostationary processes in reference to discrete-time and continuous-time stochastic processes respectively. Theory for periodically correlated processes was discussed by Hurd and Miamee (2007, [4]). Spectral analysis was discussed in Dragan (1971, [?], Honda (1982, [?]) and Hurd and Koski (2004, [?]) among others. Wavelet analysis was discussed in Cambanis and Houdré (1995, [2]) and Touati and Pesquet (2001, [?]). Random shifts are discussed in Hurd (1974, [3]), Miamee (1990, [?]) and Gardner (1994, [?]) .

The notion of periodicity has been discussed in literature in many different forms. Out of these, there are two widely used ones. See Gardner (1985 ,[?]). Strict cyclostationary process refers to the scenario where every finite dimensional joint distribution is periodic whereas the weak one has milder assumption of periodicity of only the covariance structure. In this paper, we will take a little different approach of decomposing the stochastic process in different types of components and explore their estimation and asymptotic properties. We explore mean cyclostationary stochastic processes, i.e, there exists integer $\theta_0 > 0$, so that for all i,

$$E(X_i) = E(X_{i+\theta_0}),$$

where X_i is the observed data. We allow dependence for the random noise process $X_i - E(X_i)$. In the literature, dependence has been widely formulated by strong mixing conditions. In this paper, we use the framework by Wu (2005, [8]) and the functional dependence measure introduced therein. These conditions based on moment of the error process is mild and easily verifiable compared to the strong mixing dependence. Moreover, this allows us to form an unified theory for the error process arising from a large class of time series models.

Often the observed data cannot be simplified into just a periodic component and a mean-zero error process. For example, in the field of climatology, it is usual to add a trend function in addition to the seasonal variation. Vogt and Linton (2014, [7]) discussed non-parametric estimation of the period in the presence of a smooth trend function with two derivatives. In this work, however, we add a Hölder-1/2 continuous function as the trend component to the mean cyclostationary process. With the help of the sharp Gaussian approximation and the multiplier bootstrap method by Karmakar and Wu (2017, []), we are able to do inference for a function that is only Hölder-α continuous with $\alpha \geq 1/2$. This can be seen as a two-fold generalization of the usual treatments in the literature of simultaneous confidence band for models of the type

$$y_i = \mu(i/n) + e_i.$$

We decompose the μ in trend and periodic components and construct the simultaneous band for the trend component. Moreover, usually μ is considered to have two smooth derivatives whereas we are relaxing that assumption to Hölder-1/2 continuity only.

We decompose the cyclostationary process into three important components. The periodic component μ_i has an unknown period θ_0 i.e. $\mu_i = \mu_{i+\theta_0}$. Suppose $(\epsilon_i)_{i \in \mathcal{Z}}$ are independent random variables and let $\mathcal{F}_i$ be the σ-field generated by $(\epsilon_i, \epsilon_{i-1}, ..)$. Our model is

$$X_{i,n} = g(i/n) + \mu_i + e_i = g(i/n) + \mu_i + +H_i(\mathcal{F}_i), \tag{1.1}$$

where H_i are functions such that e_i is well-defined random variable. We assume the trend function g has $[0,1]$ as its support. This is done for meaningful asymptotics in inference of g. For example, if our n is large, we have information available in a more dense set in [0,1] to allow us explore local properties of g. Moreover, to make the periodic component and the trend component identifiable, we impose the condition $\int_0^1 g(u)du = 0$.

One can see our observed data considered in (1.1) comes from a triangular array $X_{i,n}$ suggesting that the dynamics of the process change based on the sample size. Later, in Section 5, we allow this to grow as an increasing function of n. Little has been known in the literature of a cyclostationary process where the period can also grow with the size of the data. In this paper, we focus on process with fixed periods and thus drop the second suffix of $X_{i,n}$ henceforth. As a possible variant, we also provide similar theoretical results for the case of unbounded period in Section 5.

1.2. *Organization.* The rest of the paper is organized as follows. We first introduce the functional dependence measure to describe the structure of dependence within the error process. In Section 3, we describe our methods to estimate the period, the periodic components and the trend function. The next Section 4 is devoted to the asymptotic results related to these estimators. Section 5 discusses two variants, first the multivariate version of our model inspired by different climate data and secondly the scenario where the period can grow with sample size. We discuss some simulation results in Section 6 and applications to climate data of Central England in Section 7. All the proofs are postponed to Section 8.

2. Our framework.

We introduce some notation. For a matrix $A = (a_{ij})$ we define it Frobenius norm norm as $|A| = (\sum a_{ij}^2)^{1/2}$. For a random vector Y , write $Y \in \mathcal{L}_p, p > 0$, if $\|Y\|_p := [E(|Y|^p)]^{1/p} < \infty$. For L_2 norm write $\|.\| = \|.\|_2$. Define the projection operator P_i by

$$P_i Y = E(Y|\mathcal{F}_i) - E(Y|\mathcal{F}_{i-1}), \qquad Y \in \mathcal{L}_1. \tag{2.1}$$

Throughout the text, $\lfloor x \rfloor$ refers to the greatest integer less than or equal to x. C_p would refer to a constant that depends only on p but could take different values on different occurrences. $N_p(\mu, \Sigma)$ means p-variate normal distribution with mean μ and covariance matrix Σ. For a random vector Y, $Var(Y)$ stands for the variance-covariance matrix of Y. For a positive semi-definite matrix A, $A^{1/2}$ refers to the usual Grammian square root of

A. If $A = QDQ^T$ is the spectral decomposition of the matrix A then $A^{1/2} = QD^{1/2}A^T$. If two quantities A and B satisfy $A <= cB$ for some $c < \infty$ then we write $A \ll B$ or $B \gg A$. If both $A \ll B$ and $B \ll A$ then we write $A \asymp B$. We use the same symbols if such relationships holds for large n.

2.1. *Dependence structure.* We allow the error/noise process e_i to be a very general non-stationary and possibly dependent process. The dependence can happen in many different ways and hence it is important to put a structure to it.

$$e_i = H_i(\mathcal{F}_i) = H_i(\epsilon_i, \epsilon_{i-1}, \ldots), \tag{2.2}$$

where H_i are functions such that e_i are well-defined random variables. This formulation allows us to capture a huge class of well-known time-series processes. Strong mixing conditions has been extensively used in literature to formulate dependence but it suffers from easy verifiability. Wu(2005, [8]) suggested the idea of coupling to explain the dependence between the noise process e_i. We extend it to the non-stationary case in the following way

$$\delta_{j,r} = \sup_i \|e_i - e_{i,(i-j)}\|_r = \sup_i \|H_i(\mathcal{F}_i) - H_i(\mathcal{F}_{i,(i-j)}\|_r, \tag{2.3}$$

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

$$\mathcal{F}_{i,k} = (\epsilon_i, \epsilon_{i-1}, \cdots, \epsilon'_k, \epsilon_{k-1}, \cdots) \tag{2.4}$$

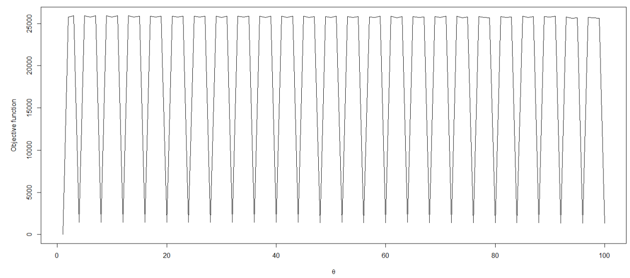
and $e_{i,\{i-j\}} = H_i(\mathcal{F}_{i,\{i-j\}})$. Clearly, $\mathcal{F}_{i,k} = \mathcal{F}_i$ is $k > i$. As Wu(2005, [8]) suggests, $\|H_i(\mathcal{F}_i) - H_i(\mathcal{F}_{i,(i-j)})\|_r$ measures the dependence of e_i on e_{i-j}. $\delta_{j,r}$ measures for the uniform j lag dependence in terms of rth moment. Define the cumulative sum $\Theta_{i,p}$ as $\sum_{j=i}^{\infty} \delta_{j,p}$.

3. Methods. The usual periodogram technique to estimate the period fails here due to the presence of dependent noise and the trend function. Our estimation method is sequential. The first and key step of our estimation procedure relies on a penalized least-square based method to estimate the unknown period. Using this estimate, one can get estimated periodic components. Plugging in the periodic component, we obtain the estimate of the unknown trend function from the residuals. Lastly, a simultaneous confidence band is obtained using invariance principle and extreme value theory for Gaussian processes.

3.1. *Known period-length, Oracle estimator.* Consider a model without the g component and independent Gaussian error e. Suppose we know that the true period is θ. The maximum likelihood estimator of μ_i for $1 \leq i \leq \theta$ is is

$$\hat{\mu}_i = \bar{X}_{i,\theta} = \frac{1}{|A_{\theta,i_\theta}|} \sum_{j \in i_\theta} X_j, \tag{3.1}$$

FIG 1. *Objective function*



where $i_\theta = i \mod \theta$ and

$$A_{\theta,t} = \{i : i_\theta = t\}, \tag{3.2}$$

for $1 \le t \le \theta$. Intuitively speaking, if the dependence decays fast enough then this estimator should work well even for the dependent errors. Also, since we have some well-established Gaussian approximation results in the literature if the error process is non-Gaussian we can still use something that works for a Gaussian process. This motivates us to estimate the ith periodic component using the same estimator as in 3.1 In fact, we can use the same estimator in the presence of the trend g. As we have assumed $\int_0^1 g(u)du = 0$. so the effect of g balances out.

3.2. *Regularization parameter.* We compare some small candidate θ. Say we have n observations $X_1, X_2, \cdots X_n$. For this given sample size n, we fix an upper bound U_n to our search region for true period θ_0.

For each integers $1 \le \theta \le U_n$ compute

$$Q(\theta) = \sum_i (X_i - \bar{X}_{i,\theta})^2. \tag{3.3}$$

We can minimize $Q(\theta)$ over the choices of θ but this clearly chooses some minimizer that is a multiple of the true period.

We first show some results based on the famous Central England data , where we have seasonal average of temperature from 1659 to 2015. Even if we work without detrending the data one can clearly see, that whenever our candidate θ assumes a multiple of 4, the objective function takes a very small value compared to when it is not.

To circumvent this difficulty, we penalize the likelihood in the following manner

$$Q(\theta, \lambda_n) = Q(\theta) + \lambda_n \theta. \tag{3.4}$$

Choosing the proper rate of the penalty parameter λ_n is a tricky question. We cannot choose it to be too big to end up estimating true θ_0 as 1. We cannot choose it too small as then we end up choosing some multiples of the true period as our estimator. We discuss the appropriate rate of λ_n that enables us to consistently estimate the period length.

3.3. *Estimating the periodic components.* We are using the estimate $\hat{\theta}$ to obtain the estimates for the mean of $Y_i = g(i/n) + e_i$.
Our estimates for the periodic components are

$$\hat{m(i)} = \frac{1}{|A_{\hat{\theta},i_{\hat{\theta}}}|} \sum_{j \in A_{\hat{\theta},i_{\hat{\theta}}}} X_j, \tag{3.5}$$

which is in other words the average of all the X_j such that $\hat{\theta}$ divides $i - j$.

3.4. *Estimating the trend function.* Recall the model in (1.1). This model is similar to the popular sequence model in literature.

$$y_i = \mu(i/n) + e_i. \tag{3.6}$$

There is a huge amount of literature where the error process in the above model is assumed to have i.i.d. Gaussian distribution. For a dependent error process there have been previous works by Wu and Zhou(2007, [9]) but that considers only functions μ with two smooth derivatives. In order to accommodate functions that are only Hölder-α continuous, we need much sharper Gaussian approximation than the one used in ([9]). With the help of the main theorem in Karmakar, Wu(2016, []) we can do inference about the above model. This was shown in ([]) using a kernel based approach. Suppose K is a symmetric kernel with bounded variation. We assume the support of the kernel to be [-1,1], $\int_{-1}^{1} K(x)dx = 1, \int_{-1}^{1} K(x)dx = \phi$

$$\hat{\mu}(t) = \frac{1}{nb_n} \sum_{i=1}^{n} K\left(\frac{i/n - t}{nb_n}\right) y_i \tag{3.7}$$

However, a similar estimate for $g(t)$ does not work for this paper as our model

$$y_i = g(i/n) + m_i + e_i, \tag{3.8}$$

involves the periodic sequence m. We instead use

$$\hat{g}(t) = \frac{1}{nb_n} \sum_{i=1}^{n} K\left(\frac{i/n - t}{nb_n}\right) (y_i - \hat{m}_i), \tag{3.9}$$

where $\hat{m_i} = \bar{y}_{i,\theta_0}$ is the estimator of the periodic sequence m. Under very mild conditions on K, we obtain the simultaneous confidence band for the function $g(t)$ using the Gaussian approximation result in Karmakar and Wu (2016, []). We are using the Pristley-Chao estimate instead of the standard Nadaraya Watson estimate although both the estimates behave similarly.

A significant contribution of our work is to provide a simultaneous confidence band (SCB) of the unknown trend function g. This improves upon the pointwise confidence band of the trend function shown in [7]. We also improve upon the usual smoothness assumption of the trend function g. There is a huge amount of literature that deals with function estimation assuming the related function possess two smooth derivatives. One of the biggest advantage of such assumption is that it facilitates taylor series expansion and hence makes the inference easier.

In our work, we only assume that the trend function g is Hölder-α continuous for $\alpha \leq 1/2$. Intuitively speaking, we are allowing our trend function to have some abrupt changes in a continuous fashion. This type of functions can often be very useful in econometrics, finances and related fields. using the sharp Gaussian approximation obtained in Karmakar, Wu (2017, []) and the extreme value theory in Lindgren (1980, [5]), we obtain the 95 % SCB of the trend function. We discuss our result in Theorem 4.4.

Even though the asymptotic coverage of the SCB reported in Theorem 4.4 is 95 %, because of the logarithmic convergence one needs a huge n to achieve a coverage around 95 %. To circumvent such difficulty we propose a bootstrap based method to easily obtain such SCB. Theorem 4.5 discusses that result.

4. Asymptotic Results.

4.1. *Assumptions.* General Assumptions

- The error process e_i is in $\mathcal{L}^p$ where $p \geq 4$. [7] needed a finite moment condition for $p > 4$
- Short range dependency of the error process:

$$\Theta_{0,p} = \sum_{i=0}^{\infty} \delta_{i,p} < \infty. \tag{4.1}$$

- In order to be able to identify the trend function g, we need to put the condition $\int_0^1 g(u)du = 0$. We also assume that g is Hölder-α continuous where $\alpha \geq 1/2$.

Assumption for period estimation For our proofs, we introduce new sequence ν_n diverging to ∞. We assume polynomial rate for all sequences and study the relationship between the exponent to facilitate asymptotic inference. We emphasize that such restrictive assumption of polynomially varying sequence can be relaxed. Assume $U_n \asymp n^{\delta_1}$, $\lambda_n \asymp n^{\delta_2}$ and $\nu_n \asymp n^{\delta_3}$, we need the following conditions on δ_1, δ_2 and δ_3. Recall that, $p > 4$, g is Hölder-α continuous with $1/2 \leq \alpha < 1$.

We summarize the assumptions on these parameters in the following conditions

B.1)
$$U_n \asymp n^{\delta_1}, \lambda_n \asymp n^{\delta_2} \text{ and } \nu_n \asymp n^{\delta_3},$$

B.2)
$$0 < \delta_1 < 2\alpha/(3+2\alpha),$$

B.3)
$$(2/p + 1/2)\delta_1 < \delta_2 < 1 - \delta_1 \text{ and}$$

B.4)
$$\delta_1(1+\alpha) + 1/2 - \alpha < \delta_3 < (1-\delta_1)/2.$$

If $\alpha = 1/2$

4.2. *Asymptotics for period estimation.*

THEOREM 4.1. *Under assumptions (B.1-B.4), we have,*

$$\hat{\theta} - \theta_0 \xrightarrow{P} 0. \tag{4.2}$$

4.3. *Asymptotics for periodic component estimation.*

THEOREM 4.2. *Uniform Consistency*

$$\max_{1 \le i \le n} |\hat{m}(i) - m(i)| = O_p(n^{-\alpha}). \tag{4.3}$$

4.4. *Asymptotics for trend estimation.* Before we state the main result of this section, we first state a Gaussian approximation result by Karmakar and Wu (2017). This is an involved result with potentially many applications. We use this to obtain the simultaneous confidence band. Apart from the general assumptions in (), for this part we assume the following for the error process (e_i),

(4.A) Upper bound on the functional dependence measure:

$$\Theta_{i,p} = O(i^{-\chi}), \chi > 0, \tag{4.4}$$

where larger χ denotes weaker dependence.

(4.B) Uniform integrability: e_i is uniformly integrable, namely

$$\sup_{i \ge 1} E(|e_i|^p |\mathbf{1}_{|e_i| \ge u}) \to 0 \text{ as } u \to \infty. \tag{4.5}$$

(4.C) Lower bound on variance of incremental process: There exists a constant $\lambda_* > 0$ such that, for all $t, l \ge 1$ and any unit vector v

$$\inf_t \rho_*(Var(\sum_{i=t+1}^{t+l} e_i)) \ge \lambda_* l. \tag{4.6}$$

Result 4.3. *Under the assumptions (4.A)-(4.C), if,*

$$\chi > \chi_0 = \frac{p^2 - 4 + (p-2)\sqrt{p^2 + 20p + 4}}{8p} \tag{4.7}$$

then there exists a probability space (Ω_c, A_c, P_c) on which we can define random vectors X_i^c with the partial sum process $S_n^c = \sum_{i=1}^n X_i^c$ and a Gaussian process G_i^c with independent increments such that $S_i^c \overset{D}{=} (S_i)_{i \leq n}$ and

$$\max_{i \leq n} |S_i^c - G_i^c| = o_P(n^{1/p}), \qquad \text{in } (\Omega_c, A_c, P_c), \tag{4.8}$$

where $G_i^c = \sum_{t=1}^i Y_t^c$ with Y_t^c being independent Gaussian random variables with mean 0.

First we state a result for constructing simultaneous confidence band for a special class of non-stationary error process, namely locally stationary process using a Gaussian approximation result obtained by Wu and Zhou (2011, [10]) for the case $2 < p < 4$. Assume $e_i = H(t_i, \mathbb{F}_i)$, where H_i satisfies the following

$$\sup_{0 \leq t < s \leq 1} \frac{\|H(t, \mathcal{F}_i) - H(s, \mathcal{F}_i)\|}{|t-s|} \leq C. \tag{4.9}$$

for some $C < 1$. Let

$$\Sigma_e(t) = (Var(D_i(t)))^{1/2} = (Var(\sum_{j=i}^{\infty} P_i H(t, \mathcal{F}_j)))^{1/2}.$$

Using $\Sigma_e(t)$, we state the following theorem for constructing the simultaneous confidence band of $\mu(t)$.

Theorem 4.4. *Assume the kernel function K is of bounded variation and b_n satisfy*

$$\sqrt{\log(1/b_n)} \left(\frac{n^{1/p}}{\sqrt{nb_n}} + b_n \log n + n^{1/2} b_n^{\alpha + \frac{1}{2}} + b_n^{-1/2} n^{-1/2} \right) \to 0. \tag{4.10}$$

Then we have

$$P(\frac{\sqrt{nb_n}}{\phi_0} sup_{t \in \tau} |\Sigma_e(t)^{-1} \{\hat{g}(t) - g(t)\}| - B(r) \leq \frac{u}{\sqrt{2 \log r}}) \to exp\{-2exp(-u)\}, \tag{4.11}$$

where $r = 1/b_n$, $\tau = [b_n, 1 - b_n]$, and

$$B(r) = \sqrt{2 \log(r)} + \frac{\log(C_K) + (d/2 - 1/2)\log(\log r) - \log(2)}{\sqrt{2 \log(r)}} \tag{4.12}$$

with

$$C_K = \frac{(\int_{-1}^{1} |K'(u)|^2 du)^{1/2}}{\gamma s/2}.$$

Remark:

- By Theorem 4.4, if $\hat{\Sigma}_e(t)$ is the uniformly consistent estimate of $\Sigma_e(t)$ then the $100(1-\alpha)\%$ confidence interval for $g(t)$is

$$\hat{g}(t) + \sqrt{\frac{\phi_0}{nb_n}}[B_K(r) - \frac{\log(\log((1-\alpha)^{-1/2}))}{\sqrt{2\log(1/b_n)}}]\hat{\Sigma}_e(t)\mathcal{B}_d \tag{4.13}$$

  where $\mathcal{B}_d$ is he d-dimensional unit ball.
- If we assume $b_n \asymp n^{-\delta}$ then by (4.10), we need

$$\frac{1}{2} < (\alpha + \frac{1}{2})\delta, \frac{2}{p} < (1-\delta), 0 < \delta < 1. \tag{4.14}$$

One can see if $\alpha = 1/2$, we need $p > 4$ but unfortunately, such a result is not available. As a remedy, the improved Gaussian approximation Result 4.3 valid for all $p > 4$ can be used. However, we do not have a similar corollary attached to the main Gaussian approximation result for a locally stationary process. This motivates us to solve this problem in its most generality by constructing a result similar to Theorem 4.4 for a general non-stationary error process. Our original Gaussian approximation in Theorem 4.1 says

$$\max_{i \le n} |S_i - \sum_{j=1}^{[i/2k_0m]} \Sigma_j^{1/2} Z_j| = o_p(n^{1/p}), \tag{4.15}$$

where $\Sigma_j = Var(B_j)$ is an estimate of the variance of the block $X_{2k_0m(j-1)+1} + \cdots + X_{2k_0mj}$. We introduce a smooth version to facilitate using properties of $\sum Z_j$.

$$\Sigma(t) = \begin{cases} \sum_{2ik_0m/n \le t} \Sigma_i & \text{if } t = 2jk_0m/n \\ \alpha\Sigma_{j-1} + (1-\alpha)\Sigma_j & \text{if } 2(j-1)k)m/n < t < 2jk_0m/n. \end{cases} \tag{4.16}$$

It can be easily shown that,

$$\sup_{0<t\le 1} |\Sigma(t) - \sum_{2ik_0m/n \le t} \Sigma_i| = o_p(n^{2/p}), \tag{4.17}$$

Thus we can use results like Theorem 4.4 to exploit the theory for extreme value Gaussian processes.

4.5. *Bootstrap SCB for known covariance of the error process.* Although we have a nice theoretical result Theorem 4.4 and its analogue for a general non-stationary error, the convergence result is still logarithmic, and thus the simultaneous band does not have the correct coverage in a practical sense unless n is huge. We propose a bootstrap technique to circumvent this difficulty. Note that,

$$\mu_{b_n}(t) - \mu(t) = \frac{1}{nb_n}\sum_{i=1}^{n} K\left(\frac{i/n - t}{b_n}\right) e_i + Bias_n, \tag{4.18}$$

where

$$Bias_n = \frac{1}{nb_n}\sum_{i=1}^{n} K\left(\frac{i/n - t}{b_n}\right)(\mu(i/n) - \mu(t)) + \mu(t)\{\frac{1}{nb_n}\sum_{i=1}^{n} K\left(\frac{i/n - t}{b_n}\right) - 1\}.$$

Also,

$$\sqrt{nb_n} Bias_n = O(\sqrt{n} b_n^{\alpha+1/2}). \tag{4.19}$$

We will use our Gaussian block multiplier to create the quantiles of

$$(nb_n)^{-1/2}\sum_{i=1}^{n} K\left((i/n - t)/nb_n\right) e_i.$$

Assuming K is of bounded variation, by summation-by-parts formula we have,

$$\max_{1\le i\le n} |\frac{1}{nb_n}\sum_{i=1}^{n} K\left(\frac{i/n - t}{b_n}\right) e_i - \frac{1}{nb_n}\sum_{i=1}^{n} K\left(\frac{i/n - t}{b_n}\right) g_i| = o_P(n^{1/p}) O\left(\frac{1}{nb_n}\right), \tag{4.20}$$

where partial sums of the Gaussian random variables g_i form the process G_i^c as described in (4.8). Define the Gaussian analogue of $\hat{m}u(t)$ process as following

$$\mu^Z(t) = \frac{1}{nb_n}\sum_{i=1}^{n} K\left(\frac{i/n - t}{b_n}\right) g_i.$$

From (4.19) and (4.20), we create the following condition on b_n,

$$\frac{n^{1/p}}{\sqrt{nb_n}} + \sqrt{n} b_n^{\alpha+1/2} = o(1). \tag{4.21}$$

Under the above condition on the bandwidth b_n, we have the following theorem for validating our algorithm.

Theorem 4.5. *Recall* $\mu^Z(t)$ *from* (**??**) *where* $g_1, \dots, g_n$ *satisfy* (4.8)*. Then*

$$\sup_t |\sqrt{nb_n}(\mu_{b_n}(t) - \mu(t) - \mu^Z(t))| = o_P(1). \tag{4.22}$$

Proof. Trivially follows from (4.21) and (4.20). □

Assuming we can create Gaussian process (G_i^c) as described in (4.8), we use the following bootstrap algorithm.

Algorithm for simulation

- For $i = 1, 2, 10000$ obtain

$$\sup_{t \in (b_n, 1-b_n)} |\mu^{Z_1}(t)|, \cdots, \sup_{t \in (b_n, 1-b_n)} |\mu^{Z_{10000}}(t)|.$$

- Get u, the $100(1-\alpha)\%$ quantiles of the empirical process $\sup_{t \in (b_n, 1-b_n)} |\mu^{Z^i}(t)|$.
- The SCB for $\mu(t)$ using (4.20) is $\mu_{b_n}(t) - u, \mu_{b_n}(t) + u$.

4.6. *Estimation of covariance:.* In reality, we need to propose some method to simulate $g_1, \dots g_n$ satisfying (4.8). Define, for a suitably chosen sequence $m \to \infty$ (See Karmakar and Wu (2017, []) for more details)

$$W_l = \sum_{|i-j| \le m, 1 \le i,j \le l} e_i e_j^T. \tag{4.23}$$

Choose another sequence M such that $m/M \to 0$. Define $\Delta_F = W_{Mf} - W_{Mf-M}$ for a suitably

Proposition 4.6. *Under the conditions of Theorem 4.1,* Δ_f *is positive definite with probability going to* 1*.*

Let

$$W_l = \sum_{1 \le i < j \le l} I(|i-j| < m) e_i e_j^T.$$

Define,

$$G_i^\Delta = \sum_{f=1}^{[i/M]} \Delta_f^{1/2} Z_f,$$

where $\Delta_f = W_{Mf} - W_{Mf-M}$ and Z_f are i.i.d. d-dimensional standard normal variables independent of (X_i). Note that, here Δ_f is also a random variable.

4.7. *Validity for Bootstrap.* First, we prove a deterministic version of the main theorem of this section.

THEOREM 4.7. *Define the process* (G_i^E) *as follows*

$$G_i^E = \sum_{f=1}^{[i/M]} (E(\Delta_f))^{1/2} Z_f. \tag{4.24}$$

Then we have,

$$\max_{i \le n} |G_i^{raw} - G_i^E| = o_P(n^{1/2-(1-2/p)\chi/(1+\chi)}). \tag{4.25}$$

COMMENT 4.8. *From (4.25), we achieve the optimal* $n^{1/p}$ *bound if* $\chi > 1$*. Also note that for* $p \ge 4$*,* $\chi_0 \ge 1$*.*

Next we discuss the approximating bound for the process G_i^*, constructed from the covariance matrix estimate from the data. The following well-known result for Gaussian process play a key-role in proving efficiency of our bootstrap procedure.
We have the following theorem for the validity of the bootstrap procedure. Let $Z_i, i \in \mathbb{Z}$, be i.i.d. standard normal random variables that are independent of (X_i); let

$$G_i^* = \sum_{f=1}^{i/M} \Delta_f^{1/2} Z_f. \tag{4.26}$$

THEOREM 4.9.

$$\max_{i \le n} |G_i^{\Delta} - G_i^{raw}| = o_{P^*}((nm)^{1/4} (\log n)^{1/2}), \tag{4.27}$$

where P^* *refers to the conditional probability given the data* $\{X_i\}$*.*

The e_i's are unobserved and their covariance structure is unknown. We propose an estimate of G_i^* with e_i therein replaced by $\hat{e}_i = y_i - \hat{\mu}(i/n) - \hat{m}_i$. To show the validity of such a replacement, we need to bound the covariance estimate arising from $E(e_i e_j^T)$ and that from $\hat{e}_i \hat{e}_j^T$.

5. Variants.

5.1. *Unbounded Period.* In this section, we try to further generalize to the case where the true period can be an increasing function of sample size n. The observed data comes from the triangular array

$$X_{i,n} = g(i/n) + m_{i,n} + e_i,$$

where $m_{i,n}$ is a periodic sequence with period $\theta_{0,n} \to \infty$ as n grows. Recall in the case, $\theta_{0,n} = \theta_0$ for all n we required the upper bound of the search region to follow $U_n \ll \sqrt{n}$. Due to the fact that $\theta_{0,n}$ can grow too, we need the search region to be shorter, i.e $U_n \theta_{0,n}^2 \ll n$. Along with this, we need the following condition on ζ_s (cf 8.4)/

Either $\zeta_s = 0$ or $|\zeta_s| > c$ where c is a positive number that does not depend on n.

A careful check of our proofs reveal that these small modification in the rates of the search region and the property of the periodic sequence allows us to extend our derivations to the scenario where the period can grow with the sample size.

5.2. *Multiple Time series.* In this subsection, we discuss a possible extension of our results to a multivariate set-up. As will be described in Section 7, we want to see if the temporal effect are similar for two different co-ordinates of the datasets. Thus it is natural to explore the possibility of extending our results to a multivariate set-up. Consider the following model

$$\mathbf{y_i} = \mathbf{g(i/n)} + \mathbf{m_i} + \mathbf{e_i} = \mathbf{g(i/n)} + \mathbf{m_i} + \mathbf{H_i}(\epsilon_\mathbf{i}, \epsilon_\mathbf{i-1}, \ldots),$$

where $\mathbf{y_i}$, $\mathbf{m_i}$ and $\mathbf{e_i}$ are in $\mathbb{R}^d$ and g is a function from $[0, 1]$ to $\mathbb{R}^d$. Since the Gaussian approximation result by Karmakar and Wu (2017, []) accommodates multiple time-series, our proofs will go through with some minor and obvious modifications.

6. Simulation Study.

6.1. *Choosing penalization parameter for finite sample.* In finite sample, choosing the penalization parameter λ_n is very crucial. In Theorem 4.1 we discuss how λ_n can vary with the sample size but in real life scenario as n is finite, we have to be really careful in choosing this parameter as it affects the estimation of the period length crucially.
Let us consider a simplified model without the trend function. Towards a possible suggestion for the suitable penalty parameter we have the following result.

Proposition 6.1. *Assume e_i are i.i.d. with mean 0 and variance σ^2. Consider the model*

$$X_i = m_i + e_i.$$

Then,

$$\frac{1}{k-1} E(X^T (I - \Pi_{k\theta_0})X - X^T (I - \Pi_{\theta_0})X) = \sigma^2 \theta_0. \tag{6.1}$$

The proof follows from straight-forward algebra. See [7] for details. Next, we first prove that this result is true asymptotically even after we incorporate the trend function.

PROPOSITION 6.2. *Assume e_i are i.i.d. with mean 0 and variance σ^2. Consider the model*

$$X_i = g(i/n) + m_i + e_i.$$

Then,

$$\frac{1}{k-1} E(X^T(I - \Pi_{k\theta_0})X - X^T(I - \Pi_{\theta_0})X) = \sigma^2\theta_0. \tag{6.2}$$

Next we incorporate the serial dependence of stationary errors. We assume the coupling framework as discussed in Section 3 and using the properties of the functional dependence measure we arrive at our next theorem.

PROPOSITION 6.3. *Let us assume*

$$X_i = g(i/n) + m_i + e_i,$$

where e_i admits the causal representation as (2.2) with $H_i = H$ for all i. Then,

$$\frac{1}{k-1} E(X^T(I - \Pi_{k\theta_0})X - X^T(I - \Pi_{\theta_0})X) - \theta_0\sigma^2 \to 0 \tag{6.3}$$

where σ^2 is the long run covariance of the non-stationary error process.

For non-stationary error process e_i however, there is no such notion of a long run covariance. However, $E(S_n^2)/n$ can be used in place of σ^2 as a reasonable estimate. Inspired from results of the type (6.1), (6.2) and (6.3), [7] suggests plugging in a sequence $\hat{\sigma}^2\kappa_n$ as λ_n where κ_n is a slowly diverging sequence and $\hat{\sigma^2}$ is obtained from the following algorithm.

- First estimate $\hat{\theta}$. With probability going to 1 it is a multiple of the true θ_0.
- Using this obtain $\hat{g}$ and $\hat{m}(i)$.
- Obtain $\hat{\epsilon}_i = X_i - \hat{g}(i/n) - \hat{m}(i)$.
- Obtain $\hat{\sigma}^2 = \text{Sample Var}(\hat{\epsilon})$.

It is possible to choose λ_n in a different way that also ensures consistent estimation of θ_0. However, to maintain simplicity we stick to this above algorithm.

6.2. *Estimation of periodic and trend components.* To see the efficacy of our period estimation method, we choose three values for n, three corresponding values for θ_0, the true period and three values for ρ determining the amount of dependence in the error process. Our results based on 1000 iterations are shown in the following three figures.

Fig 2. *Objective function*

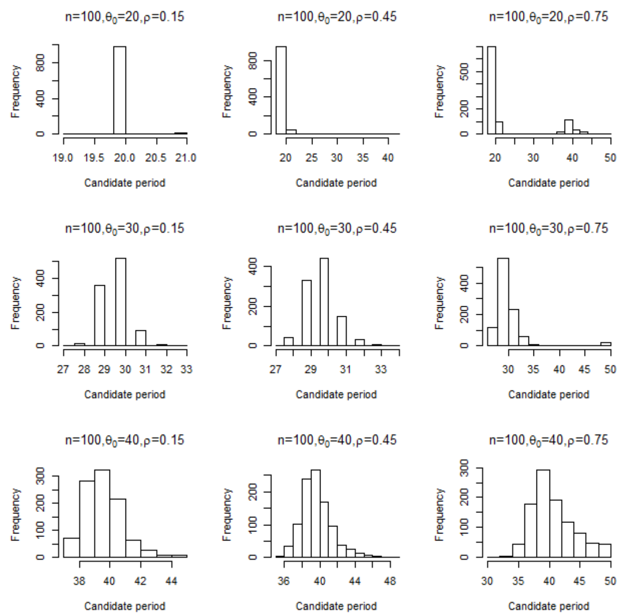


Fig 3. *Objective function*

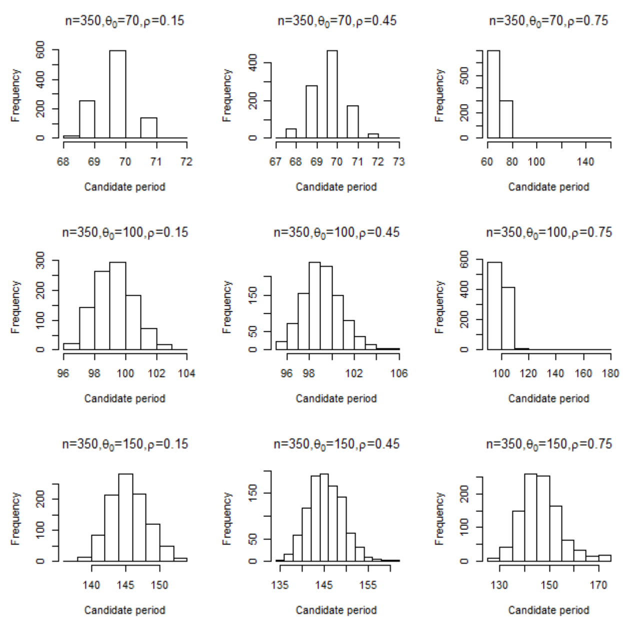
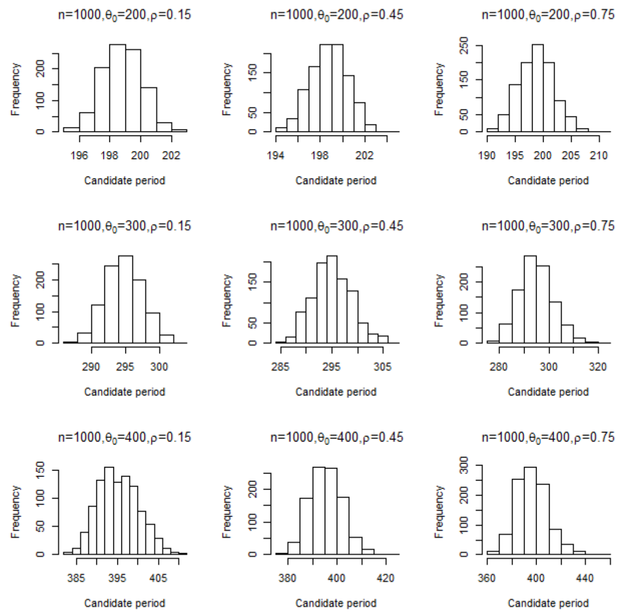


FIG 4. *Objective function*



Remark

Remark

We report the empirical asymptotic coverage probability based on 10000 iterations.

TABLE 1
Coverage probability; 5 cycles

	$n = 100$	$n = 350$	$n = 1000$
$\rho = 0.15$			
$\rho = 0.45$			
$\rho = 0.75$			

Remark

7. Applications.

FIG 5. *Objective function*

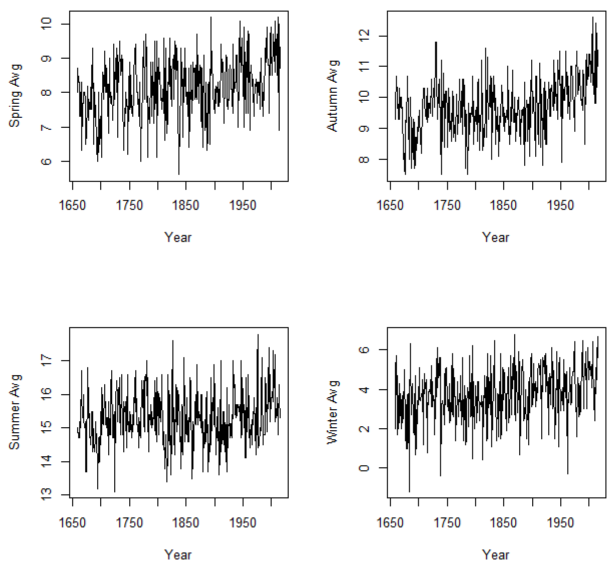


TABLE 2
Coverage probability; 3 cycles

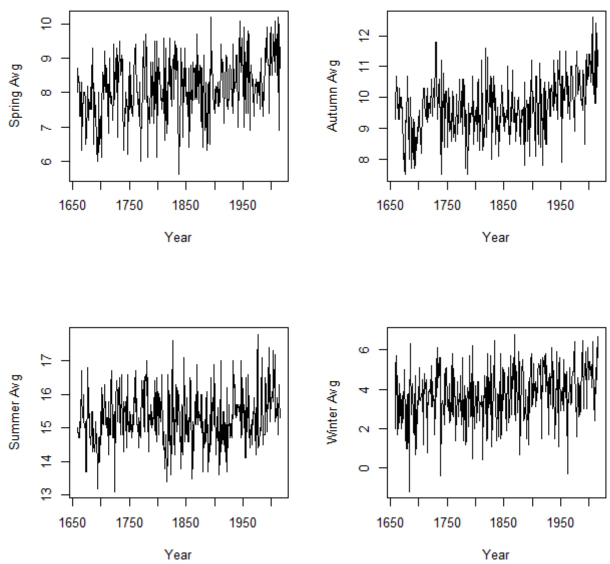
	$n = 100$	$n = 350$	$n = 1000$
$\rho = 0.15$			
$\rho = 0.45$			
$\rho = 0.75$			

TABLE 3
Coverage probability; 2 cycles

	$n = 100$	$n = 350$	$n = 1000$
$\rho = 0.15$			
$\rho = 0.45$			
$\rho = 0.75$			

7.1. *Temperature data.* We use temperature data from central England. The data comprises of 357 years record of seasonal averages of the four season of the years. An obvious period for this data is 4. We tested our methods to verify its consistency. One can see in Fig 1, we have spikes for every multiple of 4. Once we choose proper penalty, we could estimate the true period to be 4. However, in this section, we focus on the interesting study of periodic nature across years. It has been known in the literature of climatology, that the global temperature shows a periodic pattern every 60 years. This is popularly referred as 60-year circle.

FIG 6. *Objective function*



Since, we have data for four seasons, we choose to fit five different models with the fifth one using the average temperature data for each year. Our estimated periods for the central England are around 65-70 for the seasons and for the averaged out temperature of the years. One can see from Figure 4, our simultaneous confidence band for the average trend is smaller than the individual trends. Also, from figure 4 one can see the winter temperature is more fluctuating than the other three seasons.

Our sample size is $n = 357$. We use technique discussed in Section 7 to choose λ_n. U_n, the upper bound for search region is taken to be 100. We chose a bandwidth of 0.1. Slight deviations from this bandwidth also gave similar results.

Figure 7 shows the penalized loss function in (3.4) for the average temperature across years and one can see there is a sharp drop around 70.

In Figure 4, we show that the original data and the data after removing the periodic component look similar however their range is different. This is intuitive since the periodic component is average of the observations at certain distance and are of similar magnitude of the original data.

Fig 7. *Objective function*

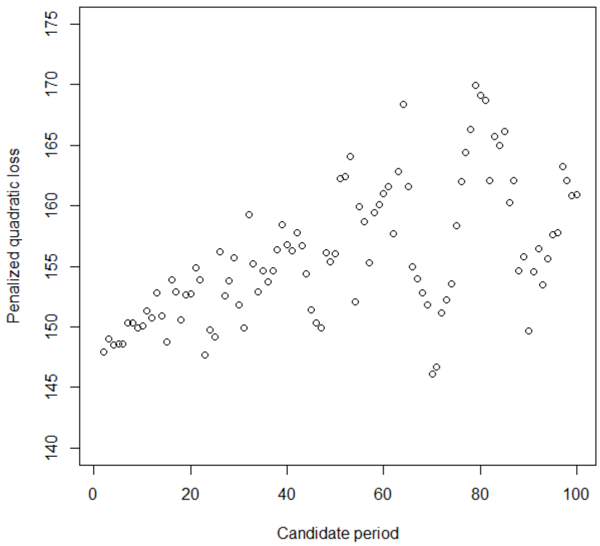


Fig 8. *Objective function*

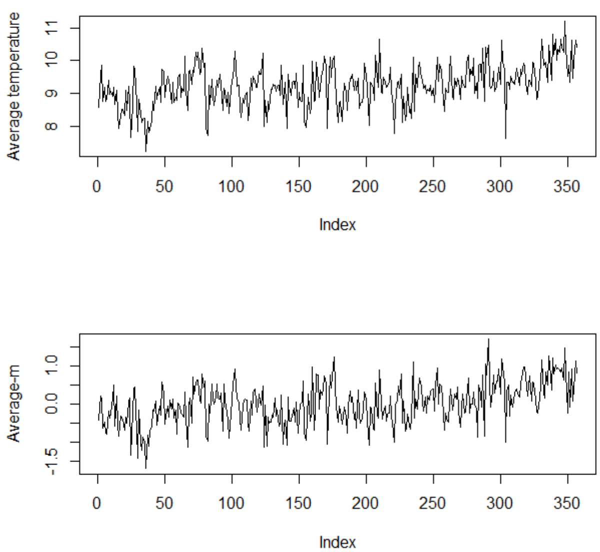
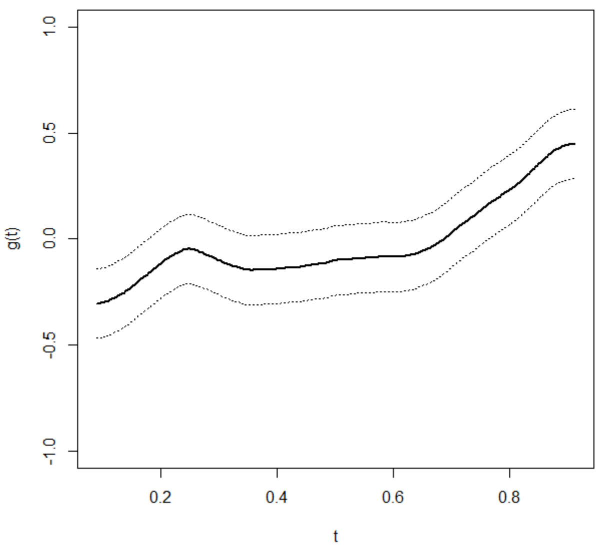


Fig 9. *Objective function*



The estimated trend along with the simultaneous confidence band is shown in the following Figure.

One can see the bands are wider for individual seasons. They all show a steep increase from the beginning of the last century probably owing to the world war.

Sample autocorrelation plot of the series after removing periodic and the trend component is shown below. It does not indicate strong dependence over time.

7.2. *Precipitation data.*

7.3. *Temperature and Precipitation data:-Joint distribution.* In this subsection, we provide an analysis for the joint distribution of the temperature and the precipitation data in conjunction the variant discussed in

7.4. *ssc:multiple.* .

Interesting remarks.

FIG 10. *Objective function*

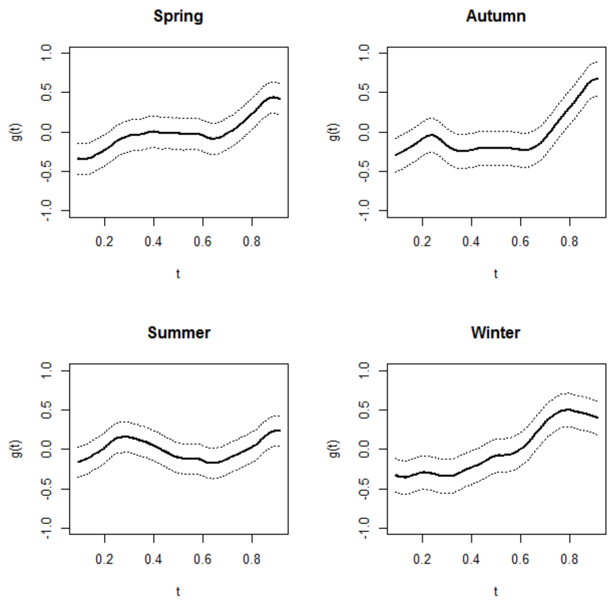
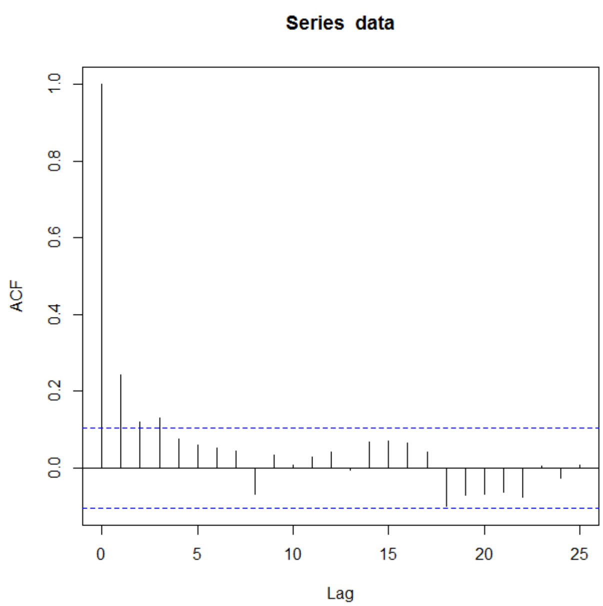


FIG 11. *Objective function*



8. Proofs. The basic steps of the proof of Theorem 1 is closely related to a similar theorem in [7]. However, we do not use the strong mixing properties for our derivations and depend on the dependence framework discussed in Section 2. We provide a detailed proof for the sake of completeness. The proof of Theorem 2 is closely related to Karmakar and Wu (2017, []) however the authors therein did not consider the presence of a trend function and hence our proof is considerably more involved. Before we present the proof of the theorem, for the convenience of the readers we mention some small results we will use throughout. Recall the definition of $|A_{\theta,i_\theta}|$ from equation 3.2. It is equal to $\lfloor \frac{n}{\theta} \rfloor$ or $\lfloor \frac{n}{\theta} \rfloor + 1$ depending on whether $i_\theta < n_\theta$ or not. Note that, we can write

$$Q(\theta, \lambda_n) = X^T (I - \Pi_\theta) X + \lambda_n \theta, \tag{8.1}$$

where $(\Pi_\theta)_{i,j} = 0$ if θ does not divide $i - j$. For pairs (i, j) such that θ divides $i - j$, $(\Pi_\theta)_{i,j} = |A_{\theta,i_\theta}|^{-1} \asymp \theta/n = O(U_n/n)$. The term $Q(\theta_0, \lambda_n) - Q(\theta, \lambda_n)$ involves the matrix $A = \Pi_\theta - \Pi_{\theta_0}$. Using the periodic nature of the vector m, it is possible to simplify some of the related terms. Consider the expression

$$Am = (\Pi_\theta - \Pi_{\theta_0}) m, \tag{8.2}$$

since m has period θ_0, $(I - \Pi_{\theta_0}) m = 0$. Hence

$$Am = (I - \Pi_\theta) m = (\gamma_1, \gamma_2, ..\gamma_{\theta*}, \gamma_1, \gamma_2, ...\gamma_{\theta*}, ..)^T, \tag{8.3}$$

where θ^* is the lcm of θ and θ_0 and

$$\gamma_s = m(s) - \frac{1}{|A_{s_\theta}|} \sum_{k=1}^{|A_{s_\theta}|} m((k - -1)\theta + s_\theta).$$

Write $\gamma_s = \zeta_s + R_s$ where

$$\zeta_s = m(s) - \frac{1}{\theta_0} \sum_{k=1}^{\theta_0} m((k - 1)\theta + s_\theta). \tag{8.4}$$

Observe that, (also see [7])

$$|R_s| \leq \frac{C\theta_0}{|A_{s_\theta}|}. \tag{8.5}$$

We first prove all the results for the case with constant period and then in the next subsection we discuss small tweaks needed to incorporate the case for the unbounded period.

Proof. of Theorem 4.1
Suppose the true period is θ_0. We show that

$$P(\hat{\theta} \neq \theta_0) = \text{small}, \tag{8.6}$$

which is equivalent to show

$$\sum_{\theta} P(X^T(\Pi_{\theta_0} - \Pi_\theta)X \leq \lambda'(\theta_0 - \theta)) \to 0. \tag{8.7}$$

Recall that

$$X_i = g(i/n) + m_i + e_i.$$

For convenience, henceforth, we call $g(i/n)$ as g_i and $\Pi_{\theta_0} - \Pi_\theta$ as a matrix A. Note that we can write

$$X^TAX = g^TAg + m^TAm + e^TAe + 2g^TAe + 2g^TAm + 2m^TAe, \tag{8.8}$$

where $X = (X_1, X_2, \cdots X_n)^T$ and g, m, e bear similar meaning. Since m has period θ_0 i.e $m(s + \theta_0) = m(s)$ for all s. Define the following quantities,

$$S_e^e = e^TAe, \quad S_g^g = g^TAg, \quad S_m^m = m^TAm, \quad S_g^m = g^TAm, \quad S_g^e = g^TAe, \quad S_m^e = m^TAe.$$

Proposition 8.1. *We have the following upper bound on $E(S_e^e)$ for every $1 \leq \theta \leq U_n$,*

$$|E(S_e^e)| < 4\Theta_{0,2}^2 U_n.$$

Proof. We will show $E(X^T\Pi_\theta X) \leq 2\Theta_{0,2}U_n$ for all $1 \leq \theta \leq U_n$. Moreover, we assume θ is a factor of n.

$$E(X^T\Pi_\theta X) \;=\; \sum\sum(\Pi_\theta)_{ij}E(e_ie_j) = \sum_i \sum_{j \in A_{\theta,i_\theta}} \frac{\theta}{n}\zeta_2(|i-j|) \leq 2U_n\sum_{k=0}^{\infty}\zeta_2(k) \leq 2U_n\Theta_{0,2}^2,$$

where $\zeta_2(k) = \sum_{i=0}^{\infty} \delta_{i,2}\delta_{i+k,2}$.

□

Proposition 8.2. *For any $1 \leq \theta \leq U_n$, we have,*

$$S_g^g = O(U_n^{2\alpha}n^{1-2\alpha}). \tag{8.9}$$

PROOF. We prove that $g^T \Pi_\theta g = O(U_n^{2\alpha} n^{1-2\alpha})$. Using Lemma 8.6 we have,

$$
\begin{aligned}
(8.10) g^T \Pi_\theta g &= \sum_i \sum_{j:\theta \text{ divides } i-j} \frac{1}{A_{\theta,i_\theta}} g_i g_j = \sum_{s=1}^{\theta} \frac{1}{|A_{\theta,s_\theta}|} \left(\sum_{k=1}^{A_{\theta,s_\theta}} g\left(\frac{(k-1)\theta+s}{n}\right)\right)^2 \\
&\leq \frac{C\theta}{|A_{\theta,s_\theta}|} (|A_{\theta,s_\theta}|)^{2-2\alpha} \asymp C\theta(n/\theta)^{1-2\alpha} = O(U_n^{2\alpha} n^{1-2\alpha}).
\end{aligned}
$$

□

PROPOSITION 8.3.

$$
\sum_{\theta\neq\theta_0} P(|S_g^e| > \nu_n) = O\left(\frac{U_n^{2+2\alpha} n^{1-2\alpha}}{\nu_n^2}\right). \tag{8.11}
$$

PROOF. It suffices to prove $E(|S_g^e|^2) = O(U_n^{1+2\alpha} n^{1-2\alpha})$. Since the matrix Π_θ is idempotent for any θ,

$$
\begin{aligned}
E(|S_g^e|^2) &\leq 2(E((g^T \Pi_\theta e)^2 + E((g^T \Pi_{\theta_0} e)^2) \\
&\leq 2g^T \Pi_\theta g E(e^T \Pi_\theta e) + 2g^T \Pi_{\theta_0} g E(e^T \Pi_{\theta_0} e).
\end{aligned} \tag{8.12}
$$

The result then follows from Proposition 8.1 and Proposition 8.2.

□

We are now ready to prove Theorem 4.1.

PROOF. of Theorem 4.1 We separate out the following two cases.

- case A: θ is a multiple of θ_0
- case B: θ is not a multiple of θ_0

□

In each of the above cases, we show that,

$$
\sum_{\theta\neq\theta_0, \theta\leq U_n} P(X^T A X \leq \lambda_n(\theta_0 - \theta)) = o(1).
$$

Recall the definition of γ_s, ζ_s and $R_{s,t}$. We have the following Proposition about ζ_s.

PROPOSITION 8.4. *Let θ^* be the lcm of θ and θ_0. For case B, there exists an index $s \in \{1, 2, ...\theta^*\}$ with $\zeta_s \neq 0$. Moreover, we can get a small constant $c_* > 0$ such that $|\zeta_s| \geq c_*$ whenever $\zeta_s > 0$. For case A, $\zeta_s = 0$ for all s*

Proof. The proof for claim corresponding to case A is trivial as in this case $\gamma_s = 0$ for all s. case B. Let θ^* be the lcm of θ and θ_0. We proceed through contradiction. Suppose $\zeta_s = 0$ for all $s \in \{1, 2, \cdots \theta^*\}$. Then

$$\frac{1}{\theta_0}\sum_{k=1}^{\theta_0} m((k-1)\theta + s_\theta) = \frac{1}{\theta_0}\sum_{k=1}^{\theta_0} m((k-1)\theta + (s + r\theta)_\theta). \tag{8.13}$$

Hence, as we have assumed $\zeta_s = \zeta_{s+r\theta} = 0$ we get that $m(s) = m(s + r\theta)$. for all s and r. As θ is not a multiple of θ_0 this would imply θ_{θ_0} is a period of the sequence $\{m(s)\}_{s\in Z}$ which clearly contradicts the fact that θ_0 is the smallest true period of the same.
Now, as we are assuming θ_0 to remain constant over different choices of sample size T, note that $\frac{1}{\theta_0}\sum_{k=1}^{\theta_0} m((k-1)\theta+s_\theta)$ is average of θ_0 many different elements of the sequence $\{m(s)\}_{s\in Z}$. Hence it can only take finitely many possible values. Hence ζ_s can also take only finite number of possible values. So the claim is proved.

□

Define d to be the number of $\zeta_s \neq 0$. Also denote $\mathbb{S}$ as the set of $s, 1 \le s \le \theta^*$ for which $\zeta_s \neq 0$.

Case A For case A, Since θ is a multiple of θ_0,$\lambda_n(\theta_0 - \theta)$ is always negative. Also, $S_g^m = S_Y^m = S_m^m = 0$ since

$$Am = (\Pi_{\theta_0} - \Pi_\theta)m = (I - \Pi_\theta)m - (I - \Pi_{\theta_0})m = 0.$$

Using the assumptions in (4.1)

$$\begin{aligned}
&P(Q(\theta,\lambda_n) \le Q(\theta_0,\lambda_n)) \qquad (8.14)\\
&= P(S_e^e \le -2S_g^e - S_g^g + \lambda_n(\theta_0 - \theta)\\
&\le P(S_e^e \le \nu_n + MU_n^{2\alpha}n^{1-2\alpha} + \lambda_n(\theta_0 - \theta)) + \lambda_n(\theta_0 - \theta), |S_g^e| \le \nu_n/2) + P(|S_g^e| > \nu_n/2)\\
&\le P(S_e^e \le \nu_n + MU_n^{2\alpha}n^{1-2\alpha} + \lambda_n(\theta_0 - \theta)) + O\left(\frac{U_n^{1+2\alpha}n^{1-2\alpha}}{\nu_n^2}\right)\\
&\le P(S_e^e \le -C\lambda_n) + O\left(\frac{U_n^{1+2\alpha}n^{1-2\alpha}}{\nu_n^2}\right),
\end{aligned}$$

for some $C > 0$. Now, since by Proposition 8.1 $|E(S_e^e)| = O(U_n) = o(\lambda_n)$ by our assumption $\delta_2 > \delta_1$ in (4.1). We apply the Lemma 8 from Xiao and Wu (2011, [11]) again to deduce that the above probability is small.

$$\begin{aligned}
U_nP(S_e^e \le -C\lambda_n) &\le U_nP(|S_e^e - E(S_e^e)| \le -C_1\lambda_n) + O\left(\frac{U_n^{2+2\alpha}n^{1-2\alpha}}{\nu_n^2}\right)\\
&= O(U_n^{1+p/4}\lambda_n^{-p/2}n^{-p/4}) + O(\frac{U_n^{2+2\alpha}n^{1-2\alpha}}{\nu_n^2}) = o(1),
\end{aligned} \tag{8.15}$$

by our assumptions on λ_n, μ_n and ν_n.

Case B We have the following proposition for case B.

PROPOSITION 8.5. *We can choose n sufficiently large such that we have the following results for case B.*

$$S_m^m \gtrsim n/U_n, \quad |S_g^m| \lesssim U_n^\alpha n^{1-\alpha}, \quad P(|S_m^e| > \nu_n\sqrt{\frac{dn}{\theta}}) \le C\nu_n^{-2}. \tag{8.16}$$

PROOF. Note that Since we assume θ_0 to be constant over n,

$$\theta_0 \le \theta^* = \frac{\theta\theta_0}{\gcd(\theta,\theta_0))} \le \theta\theta_0. \tag{8.17}$$

Also,

$$|\gamma_s - \zeta_s| = |R_s| \le CU_n/n,$$

for all s. Thus

$$S_m^m = m^T(I-\Pi_\theta)m = \quad \textstyle\sum_{s=1}^n \gamma_s^2 > \frac{n}{\theta*}\sum_{s\in\mathbb{S}}\gamma_s^2 \gtrsim \frac{dn}{\theta^*}. \tag{8.18}$$

from (8.17). The second inequality is obtained using Lemma 8.6, (8.17) and

$$|g^T(I-\Pi_\theta)m| = |\sum_{s=1}^{\theta^*}\gamma_s|A_{\theta^*,s_{\theta^*}}|\left(\frac{1}{|A_{\theta^*,s_{\theta^*}}|}\sum_k g\left(\frac{s+(k-1)\theta}{n}\right)\right)| \lesssim d(n/\theta^*)^{1-\alpha}.$$

For the third one, note that,$S_m^e = e^T(I-\Pi_\theta)m = \sum_s \gamma_s e_s$. We use the same result from Liu and Wu (2010, [6]) again to deduce that,

$$P(|S_m^e| > C\nu_n\sqrt{\frac{dn}{\theta^*}}) \quad \le \quad \frac{\|S_m^e\|^2}{C\nu_n^2 dn}\theta^* = \frac{\sum\gamma_s^2\theta^*}{C\nu_n^2 dn} \le \frac{M}{\nu_n^2}, \tag{8.19}$$

for a large enough M.

□

Using the assumptions in (4.1), Proposition 8.5, Proposition 8.1 and derivation similar to (8.14) imply, for case B,

$$P(Q(\theta,\lambda_n) \le Q(\theta_0,\lambda_n)) \le P(S_e^e \le -Cn/U_n) + \frac{C'}{\nu_n^2} \le P(S_e^e - E(S_e^e) \le -Cn/U_n) + \frac{C'}{\nu_n^2},$$

for some constants C, C' which does not depend on n.. We apply Lemma 8 from Xiao and Wu (2011, [11]) to deduce that

$$\sum_{\theta\ne\theta_0} P(S_e^e - E(S_e^e) \le -C\frac{n}{U_n}) \quad \le \quad \frac{U_n^p}{n^p}\|S_e^e - E(S_e^e)\|_p^p = U_n^{1+3p/4}n^{-p/2} = o(1), \tag{8.20}$$

by our assumptions in (4.1). This concludes the proof of Theorem 4.1.

□

8.1. *Proof of Theorem 4.2.* Suppose $\tilde{m}$ denotes the estimate of the sequence $\{m(s)\}_{s\in\mathbb{Z}}$ when the true period θ_0 is known.

$$\tilde{m}(i) = \frac{1}{|A_{\theta_0}, i_{\theta_0}|} \sum_{j\in A_{\theta_0}, i_{\theta_0}} X_j \tag{8.21}$$

and $\tilde{m}(s+k\theta_0) = \tilde{m}(s)$ for all $s = 1, 2, ..., \theta_0$ and for all $k \in \mathbb{N}$. Write

$$\sqrt{n}(\hat{m}(t) - m(t)) = \sqrt{n}(\hat{m}(t) - \tilde{m}(t)) + \sqrt{n}(\tilde{m}(t) - m(t)). \tag{8.22}$$

For any $\delta > 0$,

$$P(|\sqrt{n}(\hat{m}(t) - \tilde{m}(t))| > \delta) \le P(|\sqrt{n}(\hat{m}(t) - \tilde{m}(t))| > \delta, \hat{\theta} = \theta_0) + P(\hat{\theta} \ne \theta_0). \tag{8.23}$$

The second term is $o_p(1)$ from Theorem 4.1. The first term is identically equal to 0. Hence,

$$\begin{aligned} \sqrt{n}(\hat{m}(t) - m(t)) &= \sqrt{n}(\tilde{m}(t) - m(t)) + o_p(1) \\ &= Q_1(t) + Q_2(t), \end{aligned} \tag{8.24}$$

where

$$\sup_t Q_1(t) = \sup_t \frac{\sqrt{n}}{|A_{\theta_0,t_{\theta_0}}|} \sum_{k=1}^{|A_{\theta_0,t_{\theta_0}}|} g(\frac{t_{\theta_0} + (k-1)\theta_0}{T}) = O(n^{1/2-\alpha}),$$

form Lemma 8.6 and

$$Q_2(t) = \frac{\sqrt{n}}{|A_{\theta_0,n_{\theta_0}}|} \sum_{k=1}^{|A_{\theta_0,t_{\theta_0}}|} e_{t_{\theta_0}+(k-1)\theta_0}.$$

8.2. *Proof of Theorem 4.4.* Let K_t denotes a $n \times 1$ vector v with co-ordinates

$$v_i = \frac{1}{nb_n} K\left(\frac{t_i - t}{b_n}\right). \tag{8.25}$$

Note that,

$$\hat{g}(t) = \frac{1}{nb_n} \sum_{i=1}^{n} K(\frac{t_i - t}{b_n})(y_i - \hat{m}_i) = \frac{1}{nb_n} \sum_{i=1}^{n} K\left(\frac{t_i - t}{b_n}\right)(y_i - \bar{y}_{i,\hat{\theta}}) = K_t^T (I - \Pi_{\hat{\theta}})y. \tag{8.26}$$

We first approximate $\hat{g}(t)$ by the oracle estimator

$$\tilde{g}(t) = \frac{1}{nb_n} \sum_{i=1}^{n} K\left(\frac{t_i - t}{b_n}\right)(y_i - \bar{y}_{i,\theta_0}) = K_t'(I - \Pi_{\theta_0})y.$$

For any c_n,

$$\begin{aligned} P(c_n(|\hat{g}(t)-g(t)|>\delta) &\leq P(c_n(|\hat{g}(t)-g(t)|>\delta,\hat{\theta}=\theta_0)+P(\hat{\theta}\neq\theta_0) \\ &= P(c_n(|\hat{g}(t)-g(t)|>\delta,\hat{\theta}=\theta_0)+o(1) \\ &= P(c_n(|\tilde{g}(t)-g(t)|>\delta,\hat{\theta}=\theta_0)+o(1). \end{aligned} \tag{8.27}$$

Hence,

$$c_n|\hat{g}(t)-g(t)| = c_n|\tilde{g}(t)-g(t)|+o_p(1).$$

We write $(\tilde{g}(t)-g(t))$ as sum of stochastic part $S(t)$ and bias part $B(t)$ where $S(t) = (\tilde{g}(t)-E(\tilde{g}(t)))$ and $B(t) = (E(\tilde{g}(t))-g(t))$. We start with the bias part $B(t)$. We use the following two facts. For $j=0$ and $j=\alpha$,

$$\sup_{t\in\tau}\{\int_0^n |K_j(\frac{\lfloor 1+v\rfloor-nt}{nb_n})-K_j(\frac{v-nt}{nb_n})dv\} = O(1) \text{ and } \frac{1}{nb_n}\int_0^n K_j(\frac{v-nt}{nb_n})dv = \int_{\mathbb{R}} K_j(u)du. \tag{8.28}$$

where $K_j(u)$ denotes $K(u)u^j$.

$$\begin{aligned} E(\tilde{g}(t))-g(t) &= \frac{1}{nb_n}\sum_{i=1}^n K\left(\frac{t_i-t}{b_n}\right)(g(i/n)+m_i-\bar{g}_{i/n,\theta_0}-\bar{m}_{i,\theta_0})-g(t) \\ &= \frac{1}{nb_n}\sum_{i=1}^n K\left(\frac{t_i-t}{b_n}\right)g(i/n)-g(t)+\frac{1}{nb_n}\sum_{i=1}^n K\left(\frac{t_i-t}{b_n}\right)\bar{g}_{i/n,\theta_0} \\ &= B_1(t)+B_2(t). \end{aligned} \tag{8.29}$$

Now,

$$|B_1(t)| \leq \frac{C}{nb_n}\sum_{i=1}^n K(\frac{t_i-t}{b_n})|i/n-t|^\alpha = \frac{Cb_n^\alpha}{nb_n}\sum_{i=1}^n K(\frac{i/n-t}{b_n})|(\frac{i-nt}{nb_n})|^\alpha = O(b_n^\alpha). \tag{8.30}$$

The last line follows from (8.28) and (8.28). From Lemma 8.6, we have $|\bar{g}_{i/n,\theta_0}| \leq C(\theta_0/n)^\alpha$. Plugging that in expression of $B_2(t)$ and using (8.28) for $j=0$

$$\sup_{tin\tau}|B_2(t)| = O(n^{-\alpha}), \tag{8.31}$$

assuming $\theta_0<\infty$ and does not change with n.
In the context of Theorem 4.4, we therefore need

$$(\sqrt{\log 1/b_n})n^{1/2}b_n^{1/2+\alpha}+n^{1/2-\alpha}b_n^{1/2}\to 0. \tag{8.32}$$

Next, we handle the stochastic part $S(t)$. Clearly, as m has period θ_0 we can write

$$S(t) = \tilde{g}(t) - E(\tilde{g}(t))) = \frac{1}{nb_n}\sum_{i=1}^{n} K\left(\frac{t_i - t}{b_n}\right) e_i - \frac{1}{nb_n}\sum_{i=1}^{n} K\left(\frac{t_i - t}{b_n}\right) \bar{e}_{i,\theta_0} = .\tilde{K}_t^T e.$$

where K_t is as described in (8.25) and $\tilde{K}_t = (I - \Pi_{\theta_0})K_t$. From Karmakar and Wu (2017, []), we know,

$$\max_{j\le n} |\sum_{i=1}^{j} e_i - \sum_{i=1}^{j} \Sigma_e(t)V_i| = o(n^{1/p}). \tag{8.33}$$

From (8.33) we use summation by parts formula to deduce that

$$\sup_{t\in\tau} |S(t) - \frac{1}{nb_n}\sum_{i=1}^{n} \tilde{K}\left(\frac{t_i - t}{b_n}\right)| = O_p(\frac{n^{1/p}}{nb_n}), \tag{8.34}$$

provided $\tilde{K}_t = \tilde{v}(t)$ satisfies

$$\max_t |\tilde{v}(t)_1| + \sum_{i=2}^{n} |\tilde{v}(t)_i - \tilde{v}(t)_{i-1}| = O(\frac{1}{nb_n}). \tag{8.35}$$

Next we show that, under the assumption of K having bounded variation, with the special Π_{θ_0} we achieve eq. (8.35) .

PROOF. Let $\bar{K}(\frac{t_i - t}{b_n}, \theta_0)$ denote

$$\bar{K}(\frac{t_i - t}{b_n}, \theta_0) = \frac{1}{|A_{i,\theta_0}|}\sum_{j\in A_{i,\theta_0}} K(\frac{t_j - t}{b_n}).$$

As described before while discussing structure of Π_θ we know that

$$|A_{i,\theta_0}| = O(\theta_0/n). \tag{8.36}$$

As K have bounded variation,

$$\sum_{i=1}^{n} |K\left(\frac{t_i - t}{b_n}\right) - K\left(\frac{t_{i-1} - t}{b_n}\right)| = O(1).$$

We start with the second term in (8.35).

$$
\begin{aligned}
nb_n \sum_{i=2}^{n} |\tilde{v}(t)_i - \tilde{v}(t)_{i-1}| &\leq \sum_{i=1}^{n} |K\left(\frac{t_i - t}{b_n}\right) - K(\frac{t_{i-1}-t}{b_n})| + \sum_{i=1}^{n} |\bar{K}(\frac{t_i - t}{b_n}, \theta_0) - \bar{K}(\frac{t_{i-1}-t}{b_n}, \theta_0)| \\
&\leq O(1) + \sum_{i=1}^{n} |\frac{1}{|A_{i,\theta_0}|} \sum_{j \in A_{i,\theta_0}} K(\frac{t_j - t}{b_n}) - \frac{1}{|A_{i-1,\theta_0}|} \sum_{j \in A_{i-1,\theta_0}} K(\frac{t_j - t}{b_n})| \\
&\leq O(1) + \sum_{i=2}^{n} O(\frac{\theta_0}{n}) \sum_{j \in A_{i,\theta_0}} |K(\frac{t_j - t}{b_n}) - K(\frac{t_{j-1}-t}{b_n})| \\
&\leq O(1) + O(\frac{\theta_0}{n}) \sum_{j \in A_{i,\theta_0}} \sum_{i=2}^{n} |K(\frac{t_j - t}{b_n}) - K(\frac{t_{j-1}-t}{b_n})| \\
&= O(1) + O(\frac{\theta_0}{n}) |A_{i,\theta_0}| O(1) = O(1). \qquad (8.37)
\end{aligned}
$$

□

Hence we have (4.20). We can then use Lindgren(1980, [5]) to conclude the proof of Theorem 4.4. See Lemma 1 and Lemma 2 in Zhou, Wu (2011,[10]) for more detail.

8.3. *Proofs for the case where the true period is unbounded.*

8.4. *Lemmas.*

LEMMA 8.6. *If g is a Hölder-α continuous function with support [0,1] and for each k, $0 = x_0, x_1, x_2, \cdots x_k = 1$ be a partition of [0,1] such that $x_i - x_{i-1} = 1/k$ for all i, then for large k,*

$$|\frac{1}{k} \sum_{i=1}^{k} g(x_i) - \int_0^1 g(u)du| \leq C.k^{-\alpha}. \qquad (8.38)$$

PROOF. This follows from the fact that $\int_0^1 g(u)du = \limsup U(g, \Pi) = \liminf L(g, \Pi)$ where Π is a partition of [0,1] and U and L denotes upper and lower rectangular sums. If we take the partition as $[x_0, x_1, \cdots x_k]$ then the errors in each small sub interval is $O(k^{-1-\alpha})$. □

LEMMA 8.7. *Suppose g and $\tilde{g}$ are two Hölder-α continuous function satisfying $\int_0^1 g(u)du = \int_0^1 \tilde{g}(u)du = 0$. m and $\tilde{m}$ are two periodic sequences with smallest fundamental period θ_0 and $\tilde{\theta}_0$. If*

$$\tilde{g}(i/n) + \tilde{m}_i = g(i/n) + m_i,$$

for all i and n, then $\tilde{g} = g, \tilde{m} = m$ with $\tilde{\theta}_0 = \theta_0$.

PROOF. This proof follows from similar argument as [7]. As g and $\tilde{g}$ is Hölder-α continuous their arguments go through in our extended case as well. $\square$

References.

[1] BOYLES, R. A. and GARDNER, W. A. A. (1983). Cycloergodic properties of discrete-parameter nonstationary stochastic processes. *IEEE Trans. Inform. Theory* **29** 105–114. MR711279

[2] CAMBANIS, S. and HOUDRÉ, C. (1995). On the continuous wavelet transform of second-order random processes. *IEEE Trans. Inform. Theory* **41** 628–642. MR1331259

[3] HURD, H. L. (1974). STATIONARIZING PROPERTIES OF random shifts. *SIAM J. Appl. Math.* **26** 203–212. MR0334318

[4] HURD, H. L. and MIAMEE, A. (2007). *Periodically Correlated Random Sequences: Spectral Theory and Practice (Wiley Series in Probability and Statistics)*. Wiley-Interscience.

[5] LINDGREN, G. (1980). Extreme values and crossings for the χ^2-process and other functions of multidimensional Gaussian processes, with reliability applications. *Adv. in Appl. Probab.* **12** 746–774. MR578847

[6] LIU, W. and WU, W. B. (2010). Asymptotics of spectral density estimates. *Econometric Theory* **26** 1218–1245. MR2660298

[7] VOGT, M. and LINTON, O. (2014). Nonparametric estimation of a periodic sequence in the presence of a smooth trend. *Biometrika* **101** 121–140. MR3180661

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

[9] 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

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

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