

Estimation and inference of time-varying auto-covariance under complex trend: A difference-based approach*

Yan Cui^{1,2}

¹*Department of Statistical Sciences, University of Toronto, Canada*
e-mail: cui@utstat.toronto.edu

²*Institute for Advanced Study in Mathematics at the Harbin Institute of Technology (HIT),
 China*

Michael Levine^{†3}

³*Department of Statistics, Purdue University, USA*
e-mail: mlevins@purdue.edu

Zhou Zhou⁴

⁴*Department of Statistical Sciences, University of Toronto, Canada*
e-mail: zhou@utstat.toronto.edu

Abstract: We propose a difference-based nonparametric methodology for the estimation and inference of the time-varying auto-covariance functions of a locally stationary time series when it is contaminated by a complex trend with both abrupt and smooth changes. Simultaneous confidence bands (SCB) with asymptotically correct coverage probabilities are constructed for the auto-covariance functions under complex trend. A simulation-assisted bootstrapping method is proposed for the practical construction of the SCB. Detailed simulation and a real data example round out our presentation.

Keywords and phrases: Change points, Gaussian approximation, local stationarity, simultaneous confidence bands.

Received July 2020.

1. Introduction

Our discussion begins with a heteroscedastic nonparametric regression model

$$Y_i = \mu(t_i) + \sqrt{V(t_i)}\varepsilon_i, i = 1, \dots, n, \quad (1)$$

where Y_i are the observations, μ is an unknown mean function, $t_i = i/n$ are the design points, $i = 1, \dots, n$, ε_i are the errors with mean zero and variance 1,

*Zhou Zhou's research has been partially supported by NSERC grant 489079.

†Corresponding author.

and V is the variance function. Historically, it has been assumed that the errors ε_i are independent. Variance estimation in regression models with the unknown mean has traditionally been a rather important problem. Accurate variance estimation is required for the purpose of, for example, construction of confidence bands for the mean function, testing the goodness of fit of a model, and also in order to choose the amount of smoothing needed to estimate the mean function; see e.g. [29], [13], [15], and [18]. An extensive survey of the difference sequence approach to estimate the variance in the nonparametric regression setting when the variance function is only a constant can be found in [7].

The situation when the variance is not constant is more complicated. One of the first attempts to estimate the variance function in a regression model was made in [20] who proposed the basic idea of kernel smoothing of squared differences of observations. This idea has been further developed in [22]. [2] introduced a class of difference-based local polynomial regression-based estimators of the variance function $V(x)$ and obtained optimal convergence rates for this class of estimators that are uniform over broad functional classes. [31] obtained the minimax rate of convergence for estimators of the variance function in the model (1) and characterized the effect of not knowing the mean function on the estimation of variance function in detail. Similar approach was used to construct a class of difference-based estimators in [3] when the covariate $X \in \mathbb{R}^d$ for $d > 1$.

All of the above mentioned papers only considered the case where the data are independent. However, difference-based methods have also been used to estimate variance and/or autocovariance in nonparametric regression where the errors are generated by a stationary process. The pioneering approach here was probably that of [21] who proposed estimators based on the first-order differences to estimate (invertible) linear transformations of the variance-covariance matrix of stationary m -dependent errors. Here, by m -dependent errors we mean the errors generated by a stationary process whose autocovariance is equal to zero for any lag greater than some $m > 0$. [19] suggested second order differences to estimate the zero frequency of the spectral density of stationary processes with short-range dependence. In the case of autoregressive errors, [16] proposed \sqrt{n} -root consistent and, under the assumption of normality of errors, efficient estimators of the autocovariance that are also based on differences of observations. Under certain mixing conditions, [26] proposed estimating the autocovariance function by applying difference-based estimators of the first order to the residuals of a kernel-based fit of the signal. [34] provided an optimal difference-based estimator of the variance for smooth nonparametric regression when the errors are correlated. Finally, the closest to us in spirit is, probably, [30] that proposed a class of difference-based estimators for the autocovariance in nonparametric regression when the signal is discontinuous and the errors form a stationary m -dependent sequence. To the best of our knowledge, the problem of autocovariance estimation in a nonparametric regression where the errors form a nonstationary sequence while the signal is discontinuous has not been considered before.

The purpose of this article is to estimate and make inference of the time-varying covariance structure of a locally stationary time series when it is con-

taminated by a complex trend function with both smooth and abrupt changes. Here local stationarity refers to the slowly or smoothly evolving data generating mechanism of a temporal system ([6], [24], [37]). In time series analysis, the estimation and modelling of the autocovariance structure is of fundamental importance in, for example, the optimal forecasting of the series [1], the efficient estimation of time series regression models [17] and the inference of time series regression parameters [1]. When the trend function is discontinuous, removing the trend from the time series and then estimating the autocovariances from the residuals is not a good idea since it is very difficult to estimate the trend function near the points of discontinuity accurately. In this case, the aforementioned difference-based methods offer a good alternative. In this paper, we adopt a difference-based local linear regression method for the aforementioned time-varying autocovariance estimation problem. The method can be viewed as a nonparametric and nonstationary extension to [30]. It is shown that the uniform convergence rate of autocovariance function estimation for the difference-based method under complex trend is the same as that of autocovariance function estimation of a zero-mean time series when the number of points of discontinuity as well as the jump sizes diverge to infinity at a sufficiently slow rate. Therefore, asymptotically, the accuracy of autocovariance function estimation will not be affected by the complex trend when the difference-based nonparametric method is used.

Making inference of the autocovariance functions is an important task in practice as practitioners and researchers frequently test whether certain parametric or semi-parametric models are adequate to characterize the time series covariance structure. For instance, one may be interested in testing whether the autocovariance functions are constant over time so that a weakly stationary time series model is sufficient to forecast the future observations. There is a rich statistical literature on the inference of autocovariance structure of locally stationary time series, particularly on the testing of weak stationarity of such series. See for instance [25], [12], [8], [23], [9] and [11]. To our knowledge, only constant or smoothly time-varying trend were considered in the aforementioned literature of covariance inference. In this paper, simultaneous confidence bands (SCB) with asymptotically correct coverage probabilities are constructed for the time-varying autocovariance functions when estimated by the difference-based local linear method. The SCB serves as an asymptotically correct tool for various hypothesis testing problems of the autocovariance structure under discontinuous mean functions. A general way to perform such hypothesis tests is to estimate the autocovariance functions under the parametric or semi-parametric null hypothesis and then check whether the fitted functions can be fully embedded into the SCB. As the autocovariance functions can be estimated with faster convergence rates under the parametric or semi-parametric null hypothesis, the aforementioned way to perform the test achieves correct Type-I error rate asymptotically. The tests are of asymptotic power 1 for local alternatives whose uniform distances from the null are of the order greater than that of the width of the SCB, see Theorem 2 in [35] for instance. We also propose a simulation-assisted bootstrapping method for the practical construction of the

SCB.

The paper is organized as follows. In Section 2, we introduce the model formulation and some assumptions on x_i and ε_i^k . Section 3 presents the asymptotic theory for local estimate $\beta_k(\cdot)$. Practical implementation including a suitable difference lag and tuning parameters selection procedure, estimation of covariance matrices as well as an assisted bootstrapping method are discussed in Section 4. In Section 5, we conduct some simulation experiments on the performance of our SCBs. A real data application is provided in Section 6. Technical proofs of the main results are deferred to the Section 7.

2. Model formulation

Consider model:

$$y_{i,n} = \mu_{i,n} + x_{i,n} \quad (2)$$

where $\mu_{i,n} := \mu(t_i)$ is a mean function or signal with unknown change points, $y_{i,n} = y(t_i)$ and $x_{i,n} = x(t_i) := G(\frac{i}{n}, \mathcal{F}_i)$ is a zero-mean locally stationary process with $t_i = i/n$, $i = 1, \dots, n$. Eq. (2) covers a wide range of nonstationary linear and nonlinear processes, see [37] for more discussion. We shall omit the subscript n in the sequel if no confusion arises. Let ζ_i , $i \in \mathbb{Z}$, be independent and identically distributed (i.i.d.) random variables, and define $\mathcal{F}_i = (\dots, \zeta_{i-1}, \zeta_i)$. Then, the process $\{x_i\}$ can be written as

$$x_i = G(t_i, \mathcal{F}_i),$$

where $G(\cdot, \cdot)$ is a measurable function such that $G(t, \mathcal{F}_i)$ is well defined for all $t \in [0, 1]$. In this paper, we focus on the case that there exists $0 = a_0 < a_1 < \dots < a_d < a_{d+1} = 1$ such that

$$\mu(t) = \sum_{j=0}^d \mu_j(t) \mathbb{1}_{\{a_j \leq t < a_{j+1}\}},$$

where $\mu_j(t)$ is a Lipschitz continuous function over $[a_j, a_{j+1})$ and d is the total number of change points. Till the end of this paper, we will always assume $d = d_n = \mathcal{O}(n^\alpha)$ and the maximal jump size $\Delta_n = \max_{1 \leq j \leq d} |\mu_j(a_j) - \mu_j(a_j^-)| = \mathcal{O}(n^\beta)$ with $0 \leq \alpha, \beta < 1$.

To estimate the second order structure of the process Eq. (2), we introduce the approach based on the difference sequence of a finite order applied to the observations y_i . Assuming that the number of observations is $n+k$, this difference-based covariance estimation approach would define simple squared differences of the observations, i.e., $\rho_i^k := \rho_k(t_{i-k}) = (y_i - y_{i-k})^2$, $i = k+1, k+2, \dots, k+n$. Notice that for any fixed t , $\{\xi_i(t) := G(t, \mathcal{F}_i)\}_{i \in \mathbb{Z}}$ is a stationary process. For convenience, let us denote $s_i = t_{i-k} = \frac{i-k}{n}$. Then, $\gamma_k(s_i)$ is the k th order autocovariance function of the process $\{x_i\}$ at the fixed time s_i ; in other words, $\gamma_k(s_i) := \text{Cov}(x_i, x_{i-k})$, $k = 0, 1, \dots$. If $k = 0$, then $\gamma_0(s_i)$ turns out to be the variance of x_i at fixed time s_i .

We first introduce some notation that will be used throughout this paper. For any vector $v = (v_1, v_2, \dots, v_p) \in \mathbb{R}^p$, we let $|v| = (\sum_{i=1}^p v_i^2)^{1/2}$. For any random vector V , write $V \in \mathcal{L}^p$ ($p > 0$) if $\|V\|_p := \mathbb{E}(|V|^p)^{1/p} < \infty$. Denote $C^p([0, 1])$ as the function space on $[0, 1]$ of functions that have continuous first p derivatives with integer $p > 0$. Now, we need the following definition and assumptions:

Definition 1 (Physical dependence measure). *For all $t \in [0, 1]$, assume a stochastic system $L(t, \mathcal{F}_j) \in \mathcal{L}^p$, $q > 0$. Let $(\zeta'_j)_{j \in \mathbb{Z}}$ be an i.i.d. copy of $(\zeta_j)_{j \in \mathbb{Z}}$. Then, for any $j \geq 0$, we denote $\mathcal{F}'_j = (\mathcal{F}_{-1}, \zeta'_0, \zeta'_1, \dots, \zeta'_j)$. The physical dependence measure for $L(t, \mathcal{F}_j)$ is defined as*

$$\delta_q(L, j) = \sup_{t \in [0, 1]} \|L(t, \mathcal{F}_j) - L(t, \mathcal{F}'_j)\|_q. \quad (3)$$

If $j < 0$, let $\delta_q(L, j) = 0$. Thus, $\delta_q(L, j)$ measures the dependence of the output $L(t, \mathcal{F}_j)$ on the single input ζ_0 ; see [32] for more details.

Assumption 1. $\delta_8(G, l) = \mathcal{O}(l^{-4})$ for $l \geq 1$.

Assumption 2 (Stochastic Lipschitz continuity). *There exists a constant $C > 0$, such that $\|G(t, \mathcal{F}_i) - G(s, \mathcal{F}_i)\|_4 \leq C|t - s|$ holds for all $t, s \in [0, 1]$ and $\sup_{t \in [0, 1]} \|G(t, \mathcal{F}_i)\|_8 < \infty$.*

Assumption 1 shows that the dependence measure of time series $\{x_i\}$ decays at a polynomial rate, thus indicating short-range dependence. Assumption 2 means that G changes smoothly over time and ensures local stationarity. Here, we show some examples of the locally stationary linear and nonlinear time series that satisfy these assumptions.

Example 1 (Nonstationary linear processes). *Let ζ_i be i.i.d. random variables with $\zeta_i \in \mathcal{L}^q$, $q \geq 1$; let $a_j(\cdot)$, $j = 0, 1, \dots$, be $\mathcal{C}^1([0, 1])$ functions such that*

$$G(t, \mathcal{F}_i) = \sum_{j=0}^{\infty} a_j(t) \zeta_{i-j} \quad (4)$$

is well defined for all $t \in [0, 1]$. Clearly by Proposition 2 of [37], we know that Assumption 1 will be satisfied if $\sup_{t \in [0, 1]} \{|a_l(t)|^2\} = \mathcal{O}(l^{-4})$. Furthermore, if $\sum_{j=0}^{\infty} \{\sup_{t \in [0, 1]} |a'_j(t)|^2\} < \infty$, the stochastic Lipschitz continuity condition in Assumption 2 also holds true.

Example 2 (Nonstationary nonlinear processes). *Let ζ_i be i.i.d. random variables and consider the nonlinear time series framework*

$$\xi_i(t) = R(t, \xi_{i-1}(t), \zeta_i), \quad (5)$$

where R is a measurable function and $t \in [0, 1]$. This form has been introduced by [37] and [36]. Suppose that for some x_0 , we have $\sup_{t \in [0, 1]} \|R(t, x_0, \zeta_i)\|_q < \infty$ for $q > 0$. Denote

$$\chi := \sup_{t \in [0, 1]} L(t), \text{ where } L(t) = \sup_{x \neq y} \frac{\|R(t, x, \zeta_0) - R(t, y, \zeta_0)\|_q}{|x - y|}.$$

It is known from Theorem 6 of [37] that if $\chi < 1$, then Eq. (5) admits a unique locally stationary solution with $\xi_i(t) = G(t, \mathcal{F}_i)$ and the physical dependence measure satisfies that $\delta_q(G, j) \leq C\chi^j$, which shows geometric moment contraction. Hence, the temporal dependence with exponentially decay indicates Assumption 1 holds with $q = 8$. Further by Proposition 4 of [37], we conclude that Assumption 2 holds for $q = 4$ if

$$\sup_{t \in [0, 1]} \|M(G(t, \mathcal{F}_0))\|_q < \infty, \text{ where } M(x) = \sup_{0 \leq t < s \leq 1} \frac{\|R(t, x, \zeta_0) - R(s, x, \zeta_0)\|_q}{|t - s|}.$$

Due to the local stationarity of the process $\{x_i\}$, we have the following lemma which shows that, under mild assumptions, the autocovariance of $\{x_i\}$ also exhibits polynomial decay.

Lemma 1. Suppose Assumptions 1 and 2 hold, then we have $\sup_{t \in [0, 1]} |\gamma_k(t)| = \mathcal{O}(k^{-4})$ for $k \geq 1$.

With the above result, we can choose h large enough such that $\gamma_k(t) \approx 0$ for $k \geq h$. Next we focus on the difference series ρ_i^k for $k = 1, \dots, h$ and we always assume $h = h_n = \mathcal{O}(n^{1/8} \log n)$. By Eq. (2), we know that

$$\begin{aligned} \rho_i^k &= (x_i - x_{i-k})^2 + (\mu_i - \mu_{i-k})^2 + 2(x_i - x_{i-k})(\mu_i - \mu_{i-k}) \\ &:= \alpha_i^k + \lambda_i^k + \theta_i^k. \end{aligned} \quad (6)$$

Recall $s_i = (i - k)/n$ for $i = k + 1, \dots, k + n$ and notice that $\alpha_i^k := \alpha_k(s_i) = (x_i - x_{i-k})^2$ is the squared difference of two locally stationary processes. Therefore, it is also a locally stationary process. As a result, we can define

$$\alpha_i^k = (x_i - x_{i-k})^2 = \beta_k(s_i) + \varepsilon_i^k, \quad k = 1, \dots, h, \quad (7)$$

where $\beta_k(\cdot)$ is the unknown trend function and $\varepsilon_i^k := \varepsilon_k(s_i)$ is a zero-mean process. Then ε_i^k can be written as

$$\varepsilon_i^k = H_k(s_i, \mathcal{F}_i), \quad (8)$$

where H_k is a measurable function similar to G . With Eq. (7), if the trend function is smooth, one can easily obtain the estimator of $\beta_k(\cdot)$. Now, we introduce the following conditions.

Assumption 3. For each $k = 0, \dots, h - 1$, we assume that the k th order autocovariance function $\gamma_k(t) \in \mathcal{C}^3([0, 1])$.

Assumption 4. The smallest eigenvalue of $\sigma_k(t)$ is bounded away from 0 on $[0, 1]$ for $k = 1, \dots, h$, where

$$\sigma_k(t) = \left\{ \sum_{j=-\infty}^{\infty} \text{Cov}(H_k(t, \mathcal{F}_0), H_k(t, \mathcal{F}_j)) \right\}^{1/2}, \quad (9)$$

and $\sigma_k^2(t)$ represents the long-run variance of $\varepsilon_k(t)$ for each fixed $t \in [0, 1]$.

Assumption 5. A kernel $K(\cdot)$ is a symmetric, bounded density function with the compact support $[-1, 1]$.

Assumption 3 guarantees that the trend function $\beta_k(\cdot)$ changes smoothly for each $k = 1, \dots, h$ and is three-times continuously differentiable over $[0, 1]$. Assumption 4 prevents the asymptotic multicollinearity of regressors. Assumption 5 allows popular kernel functions such as Epanechnikov kernel. Now substituting Eq. (7) to Eq. (6), we have

$$\rho_i^k = \beta_k(s_i) + \varepsilon_i^k + \lambda_i^k + \theta_i^k. \quad (10)$$

Since the length of the series $\{\rho_i^k\}_{i=k+1}^{k+n}$ is n , we reset the subscript with respect to i as $\{\rho_i^k\}_{i=1}^n$ and therefore the time point turns out to be $t_i = i/n$ for $i = 1, \dots, n$. Similar notations are used for series $\{\varepsilon_i^k\}$, $\{\lambda_i^k\}$ and $\{\theta_i^k\}$. By Assumption 3 and the Taylor's expansion on $\beta_k(\cdot)$, it is natural to estimate $\beta_k(t)$ using the local linear estimator as follows:

$$(\widehat{\beta}_{k,b}(t), \widehat{\beta}'_{k,b}(t)) = \arg \min_{c_0, c_1 \in \mathbb{R}} \left[\sum_{i=1}^n [\rho_i^k - c_0 - c_1(t_i - t)]^2 K_b(t_i - t) \right], \quad (11)$$

where $t_i = i/n$ with $i = 1, \dots, n$ and $K_b(\cdot) = K(\cdot/b)$ is a kernel function, $b = b_n$ is the bandwidth satisfying $b \rightarrow 0$ and $nb \rightarrow \infty$. Since Eq. (11) is essentially a weighted least squares estimate, we can write the solution of Eq. (11) as

$$\widehat{\beta}_{k,b}(t) = \sum_{i=1}^n \omega_n^b(t, i) \rho_i^k, \quad (12)$$

where $\omega_n^b(t, i) = K_b(t_i - t) \frac{S_2^b(t) - (t_i - t) S_1^b(t)}{S_2^b(t) S_0^b(t) - [S_1^b(t)]^2}$ with $S_j^b(t) = \sum_{i=1}^n (t_i - t)^j K_b(t_i - t)$, $j = 0, 1, 2$. The time domain of t is fixed over $[0, 1]$ and $\omega_n^b(t, i)$, $n(t - b) \leq i \leq n(t + b)$ is the weight given to each observation ρ_i^k .

Next, we will establish the following two lemmas that are useful in establishing asymptotic properties of proposed estimators. Their proofs are given in the Section 7.2.

Lemma 2. Suppose Assumptions 1–2 hold, then we have $\delta_4(H_k, l) = \mathcal{O}(l^{-4})$ for $0 < l < k$ and $\delta_4(H_k, l) = \mathcal{O}(l^{-4}) + \mathcal{O}((l - k + 1)^{-4})$ for $l \geq k$, where $k = 1, \dots, h$.

Lemma 3. Suppose Assumptions 1–3 hold, then we have $\|H_k(t, \mathcal{F}_i) - H_k(s, \mathcal{F}_i)\|_2 \leq C|t - s|$ and $\sup_{t \in [0, 1]} \|H_k(t, \mathcal{F}_i)\|_4 < \infty$.

3. Main results

3.1. Asymptotic theory

By Assumption 3 and for $l = 0, 1, \dots$, define

$$Q_{n,l}^k(t) = \frac{1}{nb} \sum_{i=1}^n \left(\frac{t_i - t}{b} \right)^l K \left(\frac{t_i - t}{b} \right), \quad (13)$$

$$R_{n,l}^k(t) = \frac{1}{nb} \sum_{i=1}^n \rho_i^k \left(\frac{t_i - t}{b} \right)^l K \left(\frac{t_i - t}{b} \right). \quad (14)$$

Then Eq. (11) can be expressed as

$$\begin{pmatrix} \widehat{\beta}_{k,b}(t) \\ b\widehat{\beta}'_{k,b}(t) \end{pmatrix} = \begin{pmatrix} Q_{n,0}^k(t) & Q_{n,1}^k(t) \\ Q_{n,1}^k(t) & Q_{n,2}^k(t) \end{pmatrix}^{-1} \begin{pmatrix} R_{n,0}^k(t) \\ R_{n,1}^k(t) \end{pmatrix} := [Q_n^k(t)]^{-1} R_n^k(t). \quad (15)$$

Let

$$\mu_l = \int_{\mathbb{R}} x^l K(x) dx \text{ and } \phi_l = \int_{\mathbb{R}} x^l K^2(x) dx, \quad l = 0, 1, \dots.$$

Now, we will construct SCBs for $\beta_k(\cdot)$, $k = 1, \dots, h$.

Theorem 1. Suppose that Assumptions 1–5 hold and further assume that

- (1) $\sigma_k(t)$ is Lipschitz continuous on $[0, 1]$.
- (2) $\log(n)/(n^{1/4-\beta}\sqrt{b}) + \log(n)/(n^{3/8-2\beta}\sqrt{b}) + nb^5 \rightarrow 0$.
- (3) $2b < \iota := \min_{1 \leq j \leq d} (a_j - a_{j-1})$.

Then, for each $k = 1, \dots, h$, we have

$$\begin{aligned} \mathbb{P} \left[\sqrt{\frac{nb}{\phi_0}} \sup_{t \in \mathcal{T}} \left| \sigma_k^{-1}(t) \left\{ \widehat{\beta}_{k,b}(t) - \beta_k(t) \right\} \right| - B_K(m^*) \leq \frac{u}{\sqrt{2 \log(m^*)}} \right] \\ = \exp \{-2 \exp(-u)\}, \end{aligned}$$

as $n \rightarrow \infty$, where $\mathcal{T} = [b, 1-b]$, $m^* = 1/b$ and

$$B_K(m^*) = \sqrt{2 \log(m^*)} + \frac{1}{\sqrt{2 \log(m^*)}} \log \left(\frac{1}{\pi} \sqrt{\frac{1}{4\phi_0} \int_{-1}^1 |K'(u)|^2 du} \right).$$

Let us comment on the conditions listed in Theorem 1. Condition (1) shows the smoothness of $\sigma_k(t)$. Condition (2) is an undersmoothing requirement that reduces the bias of the estimators to the second order. Condition (3) guarantees there exists at most one change point in the b -neighborhood of any time t . Furthermore, if the change points $\{a_j\}_{j=1}^d$ are uniformly distributed on the interval $[0, 1]$, it suffices to control $\alpha \leq 1/5$ such that $n^{-\alpha} > 2b$ using $nb^5 \rightarrow 0$ in condition (2).

Notice that $\mathbb{E}(\rho_i^k) = \beta_k(t_i) + \mathbb{E}(\lambda_i^k) = 2\gamma_0(t_i) - 2\gamma_k(t_i) + \tilde{\Delta}_i^2$, where $\tilde{\Delta}_i^2 = \mathcal{O}(1/n^2)$ when there is no change point between observations y_i and y_{i-k} , $\tilde{\Delta}_i^2 \leq \Delta_n^2 = \mathcal{O}(n^{2\beta})$ when there exists at least a change point on $\mu(\cdot)$. However, the estimate of λ_i^k can be viewed as a negligible term (refer to the proof of Theorem 2 in the 7.2). With the previous discussion in mind, we can define

$$\begin{aligned} \widehat{\gamma}_0(t) &= \frac{1}{2} \widehat{\beta}_{h,b_h}(t), \\ \widehat{\gamma}_k(t) &= \frac{1}{2} [\widehat{\beta}_{h,b_h}(t) - \widehat{\beta}_{k,b_k}(t)], \quad k = 1, \dots, h-1, \end{aligned}$$

where b_h and b_k are the bandwidths for estimators $\widehat{\gamma}_0(t)$ and $\widehat{\gamma}_k(t)$, respectively. Making it easy to distinguish, here we use the different notations for the bandwidths which will be selected by some criterion (see Section 4.4). Notice that we require the same bandwidth b_k to compute the estimator of $\gamma_k(t)$. With the above results, the SCB for $\gamma_0(\cdot)$ is straightforward.

Corollary 1. *With the conditions in Theorem 1, we have*

$$\begin{aligned} \mathbb{P} \left[\sqrt{\frac{4nb_h}{\phi_0}} \sup_{t \in \mathcal{T}} |\sigma_h^{-1}(t) \{\widehat{\gamma}_0(t) - \gamma_0(t)\}| - B_K(m^*) \leq \frac{u}{\sqrt{2 \log(m^*)}} \right] \\ = \exp\{-2 \exp(-u)\}. \end{aligned}$$

Furthermore, to facilitate the SCBs for $\gamma_k(\cdot)$, $k = 1, \dots, h-1$, we will consider a linear combination of $\beta_k(\cdot)$. First, define $\tilde{H}_k = (H_h(t, \mathcal{F}_i), H_k(t, \mathcal{F}_i))^\top$ and a 2 by 2 matrix

$$\Sigma_k^2(t) = \sum_{j=-\infty}^{\infty} \text{Cov}(\tilde{H}_k(t, \mathcal{F}_0), \tilde{H}_k(t, \mathcal{F}_j)). \quad (16)$$

We also denote $\tilde{\beta}_k(t) = (\beta_h(t), \beta_k(t))^\top$ as a two-dimensional vector, $C = (1, -1)^\top$ and $\beta_{C,k}(t) = C^\top \tilde{\beta}_k(t)$. The natural estimators for $\tilde{\beta}_k(t)$ and $\beta_{C,k}(t)$ are $\widehat{\beta}_k(t) = (\widehat{\beta}_{h,b_k}, \widehat{\beta}_{k,b_k})^\top$ and $\widehat{\beta}_{C,k}(t) = C^\top \widehat{\beta}_k(t) = \widehat{\beta}_{h,b_k} - \widehat{\beta}_{k,b_k}$, respectively. Furthermore, let $\sigma_{C,k}^2(t) = C^\top \Sigma_k^2(t) C$, similar to Theorem 3 in [38]. At this point, we can obtain the following result.

Corollary 2. *Suppose that the smallest eigenvalue of $\sigma_{C,k}(t)$ is bounded away from 0 on $[0, 1]$ for $k = 1, \dots, h-1$. Moreover, we assume that all of the conditions of Theorem 1 are valid. Then, we have (i)*

$$\begin{aligned} \mathbb{P} \left[\sqrt{\frac{nb_k}{\phi_0}} \sup_{t \in \mathcal{T}} \left| \sigma_{C,k}^{-1}(t) \{\widehat{\beta}_{C,k}(t) - \beta_{C,k}(t)\} \right| - B_K(m^*) \leq \frac{u}{\sqrt{2 \log(m^*)}} \right] \\ = \exp\{-2 \exp(-u)\}, \end{aligned}$$

as $n \rightarrow \infty$. (ii) Furthermore, one can easily deduce the SCB for $\gamma_k(\cdot)$, $k = 1, \dots, h-1$,

$$\begin{aligned} \mathbb{P} \left[\sqrt{\frac{4nb_k}{\phi_0}} \sup_{t \in \mathcal{T}} \left| \sigma_{C,k}^{-1}(t) \{\widehat{\gamma}_k(t) - \gamma_k(t)\} \right| - B_K(m^*) \leq \frac{u}{\sqrt{2 \log(m^*)}} \right] \\ = \exp\{-2 \exp(-u)\}. \end{aligned}$$

Remark 1. *It is noteworthy mentioning that for estimating $\beta_{C,k}(t)$, we use the same bandwidth b_k ; therefore, the entire estimator $\widehat{\beta}_{C,k}(t) = \widehat{\beta}_{h,b_k}(t) - \widehat{\beta}_{k,b_k}(t)$ depends on only a single tuning parameter (bandwidth b_k). This enables us to achieve the conclusion of Corollary 2 (i) based on the result of Theorem 1. As a result, Corollary 2 (ii) also holds true due to this fact.*

After constructing SCBs for the second order structure $\gamma_k(\cdot)$, the following theorem states that $\widehat{\gamma}_k(t)$ are consistent estimators for $\gamma_k(t)$ uniformly in t for all $k = 0, \dots, h - 1$.

Theorem 2. *Under Assumptions 1–5 and suppose condition*

$$\log(n)/(n^{3/5}b) + \log(n)/(n^{7/8-2\beta}b) + \log(n)/(n^{3/4-\beta}b) \rightarrow 0, \quad \iota > 2b$$

holds true. Then, we have

$$\sup_{t \in \mathcal{T}} |\widehat{\beta}_{k,b}(t) - \beta_k(t)| = \mathcal{O}_{\mathbb{P}}(\chi_n), \quad k = 1, \dots, h,$$

$$\text{where } \chi_n = b^2 + \sqrt{\frac{\log(n)}{nb}} + \frac{\log(n)}{n^{3/5}b} + \frac{\log(n)}{n^{7/8-2\beta}b} + \frac{\log(n)}{n^{3/4-\beta}b}.$$

This theorem implies the uniform consistency of $\widehat{\beta}_{k,b}(\cdot)$. Additionally, due to the relationship between β_k and γ_k , we can also easily obtain the following consistency result for $\widehat{\gamma}_k(\cdot)$.

Corollary 3. *With the conditions in Theorem 2, we have*

$$\sup_{t \in \mathcal{T}} |\widehat{\gamma}_k(t) - \gamma_k(t)| = \mathcal{O}_{\mathbb{P}}(\chi_n), \quad k = 0, \dots, h - 1.$$

4. Practical implementation

4.1. Selection of the difference lag

Note that for any fixed time t , $\widehat{\beta}_{k,b_k}(t) = 2\widehat{\gamma}_0(t) - 2\widehat{\gamma}_k(t)$ and recall that $\gamma_k(t) \approx 0$ when $k \geq h$, where h is a large value that has been chosen in advance. Hence, we know that if $k \geq h$, $\widehat{\beta}_{k,b_k}(t) \approx 2\widehat{\gamma}_0(t)$ is practically invariant with respect to t as k increases. This fact suggests the following bandwidth selection procedure.

First, for any fixed t , we choose a large enough value $h_0(t)$ and select $k = h_0(t)$. Next, we calculate $\widehat{\beta}_{k,b_k}(t)$. Then, by successively decreasing the value of k and considering $k = h_0(t) - 1, h_0(t) - 2, \dots$, we calculate the corresponding quantities $\widehat{\beta}_{k,b_k}(t)$ until $\widehat{\beta}_{k,b_k}(t)$ shows an abrupt change. At this point, the optimal difference lag for time t can be selected as the current k plus 1. Intuitively, we can interpret this through the scatterplot of $(k, \widehat{\beta}_{k,b_k}(t))$. When the slope of the function $\widehat{\beta}_{k,b_k}(t)$ shows an obvious change, then we can choose $h^*(t) = k + 1$. Following the above procedure for each time point $t_i = i/n$, $i = 1, \dots, n$, we finally choose the optimal lag as $h = \sum_{i=1}^n h^*(t_i)/n$.

4.2. Covariance matrix estimation

To apply Corollaries 1 and 2 (ii), we need to estimate the long-run variance $\Sigma_k^2(\cdot)$ in Eq. (16) first. This problem is complicated but has been extensively studied by many researchers. Here we adopt the technique considered by [38].

Let $Q_i^k = \sum_{j=-m}^m \tilde{\varepsilon}_{i+j}$, $i = 1, \dots, n$, where $\tilde{\varepsilon}_i = (\varepsilon_i^h, \varepsilon_i^k)^\top$ for $i = 1, \dots, n$. Notice that $\mathbb{E}(\tilde{\varepsilon}_i) = 0$ and denote $N_i^k := Q_i^k Q_i^{k\top} / (2m + 1)$. In the locally stationary case, we can make use of the fact that a block of $\{\tilde{\varepsilon}_i\}$ is approximately stationary when its length is small compared with n . Hence, $\mathbb{E}(N_i^k) \approx \Sigma_k^2(t_i)$ as $m \rightarrow \infty$ and $m/n \rightarrow 0$. Let $\tau = \tau_n$ be the bandwidth and define the covariance matrix estimator as

$$\tilde{\Sigma}_k^2(t) = \sum_{i=1}^n \tilde{\omega}_n^\tau(t, i) N_i^k, \quad \tilde{\omega}_n^\tau(t, i) = \frac{K_\tau(t_i - t)}{\sum_{k=1}^n K_\tau(t_k - t)},$$

with τ being the bandwidth. Therefore, the estimate $\tilde{\Sigma}_k^2(t)$ is guaranteed to be positive semidefinite. The following theorems provide consistency of our covariance matrix estimate.

Theorem 3. Assume that $\Sigma_k^2(t) \in \mathcal{C}^2[0, 1]$, $\delta_4(\tilde{H}_k, j) = \mathcal{O}(\{j \log(j)\}^{-2})$, $m = m_n \rightarrow \infty$, $m = \mathcal{O}(n^{1/3})$, $\tau \rightarrow 0$ and $n\tau \rightarrow \infty$. Then, (i) for each k , $k = 1, \dots, h-1$ and any fixed $t \in (0, 1)$,

$$\left\| \tilde{\Sigma}_k^2(t) - \Sigma_k^2(t) \right\| = \mathcal{O} \left(\sqrt{\frac{m}{n\tau}} + \frac{1}{m} + \tau^2 \right),$$

(ii) for $\mathcal{I} = [\tau, 1 - \tau]$,

$$\left\| \sup_{t \in \mathcal{I}} \left| \tilde{\Sigma}_k^2(t) - \Sigma_k^2(t) \right| \right\| = \mathcal{O} \left(\sqrt{\frac{m}{n\tau^2}} + \frac{1}{m} + \tau^2 \right).$$

In practice, the errors $\tilde{\varepsilon}_j$ cannot be observed, thus we use $\hat{\Sigma}_k^2(t) = \sum_{i=1}^n \tilde{\omega}_n^\tau(t, i) \times \hat{N}_i^k$, where \hat{N}_i^k is defined as N_i^k with $\tilde{\varepsilon}_j$ therein replaced by its estimator $\hat{\varepsilon}_j$.

Theorem 4. Assume that conditions of Theorem 2 and conditions of Theorem 3 hold. Denote $\nu_n = \sqrt{m \log(n)} \chi_n$, where χ_n is defined in Theorem 2 and further assume $\nu_n \rightarrow 0$. Then

$$\sup_{t \in \mathcal{I}} \left| \hat{\Sigma}_k^2(t) - \Sigma_k^2(t) \right| = \mathcal{O}_{\mathbb{P}} \left(\nu_n + \sqrt{\frac{m}{n\tau^2}} + \frac{1}{m} + \tau^2 \right).$$

Note that $\sigma_h^2(t)$ is the first diagonal element of $\Sigma_k^2(t)$ and $\sigma_{C,k}^2(t) = C^\top \Sigma_k^2(t) C$. Thus, the covariance estimates in Corollaries 1 and 2 (ii) can be easily calculated via plugging in the long-run covariance matrix estimate $\hat{\Sigma}_k^2(t)$.

4.3. Simulation assisted bootstrapping method

Now we aim to apply Corollary 1 and Corollary 2 (ii) to construct the SCBs. Let $\hat{\gamma}_0''(t)$ and $\hat{\gamma}_k''(t)$ be uniformly consistent estimators of $\gamma_0''(t)$ and $\gamma_k''(t)$ for $k = 1, \dots, h-1$, respectively. Then the corresponding $(1 - \alpha)$ th SCB with $\alpha \in (0, 1)$ for $\gamma_0(t)$ and $\gamma_k(t)$ are

$$\left[\hat{\gamma}_0(t) \pm \hat{\sigma}_h(t) \sqrt{\frac{\phi_0}{4nb}} \left(B_K(m^*) - \frac{\log[\log(1 - \alpha)^{-1/2}]}{\sqrt{2 \log(m^*)}} \right) \right],$$

$$\left[\hat{\gamma}_k(t) \pm \hat{\sigma}_{C,k}(t) \sqrt{\frac{\phi_0}{4nb}} \left(B_K(m^*) - \frac{\log[\log(1-\alpha)^{-1/2}]}{\sqrt{2\log(m^*)}} \right) \right], \quad k = 1, \dots, h-1.$$

Due to the slow rate of convergence to Gumbel distribution, in practice, the UCB from Corollaries 1 and 2 (ii) may not have good finite-sample performances. To circumvent this problem, we shall adopt a simulation assisted bootstrapping approach.

Proposition 1. *Suppose conditions in Theorem 1 hold and also assume that $\sigma_k(t)$ is Lipschitz continuous for $k = 1, \dots, h$. Then, on a richer probability space, there are i.i.d. standard normal distributed random variables u_i such that*

$$\begin{aligned} \sup_{t \in \mathcal{T}} |\hat{\gamma}_0(t) - \gamma_0(t) - Z_0(t)| &= \mathcal{O}_{\mathbb{P}}(\psi_n), \\ \sup_{t \in \mathcal{T}} |\hat{\gamma}_k(t) - \gamma_k(t) - Z_{C,k}(t)| &= \mathcal{O}_{\mathbb{P}}(\psi_n), \quad k = 1, \dots, h-1, \end{aligned}$$

where $\psi_n = n^{1/2}b^{7/2} + \sqrt{\frac{1}{n^{1-2\alpha-4\beta}b}} + \sqrt{\frac{\log(n)}{n^{1-2\alpha-2\beta}b}} + \sqrt{\frac{b\log(n)}{n}}$, $Z_0(t) = \sigma_h(t)\mu_{b_h}^\dagger(t)$ and $Z_{C,k}(t) = \sigma_{C,k}(t)\mu_{b_k}^\dagger(t)$ with $\mu_b^\dagger(t) = \sum_{i=1}^n \omega_n^b(t, i)u_i/2$.

The proposition implies that the distribution of $\sup_{t \in \mathcal{T}} |\sigma_h^{-1}(t)[\hat{\gamma}_0(t) - \gamma_0(t)]|$ ($\sup_{t \in \mathcal{T}} |\sigma_{C,k}^{-1}(t)[\hat{\gamma}_k(t) - \gamma_k(t)]|$) can be well approximated by that of $\sup_{t \in \mathcal{T}} |\mu_{b_h}^\dagger(t)|$ ($\sup_{t \in \mathcal{T}} |\mu_{b_k}^\dagger(t)|$), which can be obtained by generating a large number of i.i.d. copies via bootstrapping. Therefore, the above proposition provides us with an alternative way to construct the SCB of the autocovariance function without using the asymptotic Gumbel distribution.

For ease of application, we combine procedures mentioned above into a convenient sequence of steps below.

- Choose the difference lag order h by using method that is proposed in Section 4.1.
- Find appropriate bandwidths b_h and b_k for estimating $\beta_h(\cdot)$, $\beta_h(\cdot) - \beta_k(\cdot)$ respectively, and the bandwidth τ for estimating $\Sigma_k^2(\cdot)$.
- Generate i.i.d. random variables $u_1, u_2, \dots \sim N(0, 1)$ and calculate $\sup_{t \in [b_k, 1-b_k]} |\mu_{b_k}^\dagger(t)|$ for $k = 1, \dots, h$.
- Repeat the last step for a large number of times (e.g. 10^4) and obtain the estimated $(1-\alpha)$ th quantile $\hat{q}_{1-\alpha}^k$ of $\sup_{t \in [b_k, 1-b_k]} |\mu_{b_k}^\dagger(t)|$.
- Calculate $\hat{\Sigma}_k^2(t)$ by using the method in Section 4.2. Then, obtaining $\hat{\sigma}_h(t)$ together with $\hat{\sigma}_{C,k}(t)$ is straightforward.
- Construct the $(1-\alpha)$ th SCB of the autocovariance function $\gamma_0(t)$ as $\hat{\gamma}_0(t) \pm \hat{q}_{1-\alpha}^h \hat{\sigma}_h(t)$, and $\hat{\gamma}_k(t) \pm \hat{q}_{1-\alpha}^k \hat{\sigma}_{C,k}(t)$ for $\gamma_k(t)$, $k = 1, \dots, h-1$.

4.4. Selection of tuning parameters

In this subsection, we briefly discuss the practical choices of tuning parameters b , m and τ . Here, we consider the generalized cross-validation (GCV) method by

[5] to choose the bandwidth b . Specifically, we consider two cases of bandwidth selection for $\gamma_0(t)$ and $\gamma_k(t)$, $k = 1, \dots, h - 1$, respectively. For estimating $\gamma_0(t)$, let $P_h = (\rho_1^h, \dots, \rho_n^h)^\top$ and $\hat{P}_h(b) = (\hat{\rho}_1^h(b), \dots, \hat{\rho}_n^h(b))^\top$ be the corresponding fitted values. One can write $\hat{P}_h(b) = H(b)P_h$, where $H(b)$ is an n by n square hat matrix that depends on b . Then, we choose the optimal bandwidth (say b_h) that minimizes

$$\text{GCV}_h(b) = \frac{n^{-1} \sum_{i=1}^n [\rho_i^h - \hat{\rho}_i^h(b)]^2}{[1 - \text{trace}(H(b))/n]^2}.$$

On the other hand, when estimating $\gamma_k(t)$ for $k = 1, \dots, h - 1$, we treat $\hat{\gamma}_k(t)$ as a whole term and choose a joint bandwidth for it. Similarly, denote $P_k = (\rho_1^h - \rho_1^k, \dots, \rho_n^h - \rho_n^k)^\top$ and let $\hat{P}_k(b) = (\hat{\rho}_1^h(b) - \hat{\rho}_1^k(b), \dots, \hat{\rho}_n^h(b) - \hat{\rho}_n^k(b))^\top$ be the corresponding fitted values. As before, one can write $\hat{P}_k(b) = H(b)P_k$. With this in mind, we select as optimal the bandwidth (say b_k) that minimizes the following quantity:

$$\text{GCV}_k(b) = \frac{n^{-1} \sum_{i=1}^n [(\rho_i^h - \rho_i^k) - (\hat{\rho}_i^h(b) - \hat{\rho}_i^k(b))]^2}{[1 - \text{trace}(H(b))/n]^2}.$$

For the choice of m and τ , we now employ the extended minimum volatility method (including two parameters) which was proposed in Chapter 9 of [10]. This method is based on the fact that if a pair of block size and bandwidth is in an appropriate range, then confidence regions for the local mean constructed by $\hat{\Sigma}_k^2(t)$ should be stable. Therefore, we first consider a grid of possible block sizes and bandwidths and then choose the optimal pair that minimizes the volatility of the boundary points of the confidence regions in the neighborhood of this pair. To be more specific, let the grid of possible block sizes and bandwidths be $\{m_1, \dots, m_{M_1}\}$ and $\{\tau_1, \dots, \tau_{M_2}\}$, respectively. Then denote the estimated long-run covariance matrices as $\{\hat{\Sigma}_k^2(m_i, \tau_j, t)\}$ for $i = 1, \dots, M_1$, $j = 1, \dots, M_2$. For each pair (m_i, τ_j) , we need to calculate

$$\text{ISE} \left[\cup_{r=-2}^2 \{\hat{\Sigma}_k^2(m_{i+r}, \tau_j, t)\} \cup \cup_{r=-2}^2 \{\hat{\Sigma}_k^2(\tau_{j+r}, m_i, t)\} \right], \quad k = 1, \dots, h, \quad (17)$$

where ISE denotes the integrated standard error

$$\text{ISE}[\{\hat{\Sigma}_k^2(s, \cdot, t)\}_{s=1}^l] = \int_0^1 \left\{ \frac{1}{l-1} \sum_{s=1}^l \left| \hat{\Sigma}_k^2(s, \cdot, t) - \bar{\hat{\Sigma}}_k^2(\cdot, t) \right|^2 \right\}^{1/2} dt$$

with $\bar{\hat{\Sigma}}_k^2(\cdot, t) = \sum_{s=1}^l \hat{\Sigma}_k^2(s, \cdot, t)/l$ and s being the parameter m or τ . Finally, we choose the pair (m_i^*, τ_j^*) that minimizes Eq. (17).

5. Simulations

To illustrate performance of the proposed estimator of autocovariance, we consider several models. For each model, we obtain the uniform confidence interval

coverage of the true variance function and the autocovariance function at lag 1 for three different sample sizes: $n = 400$, $n = 600$ and $n = 800$. In each case, we use 500 replications. To select bandwidths b_h and b_k we use the grid from 0.15 to 0.45 with the step size 0.01. We also provide a graphical illustration of a confidence interval enclosing the true variance and autocovariance lag 1 functions for each of the models considered.

The first model considered has the errors that are generated by a locally stationary linear process (4) with $a_j(t) = (\frac{t}{2})^j$, $j = 1, 2, \dots$ while the sequence (ζ_i) consists of i.i.d. normal random variables with mean zero and variance 1. In this case the coefficients start with $j = 1$ since otherwise $a_0(t)$ is undefined at $t = 0$. The Assumption 1 is satisfied since $\sup_{t \in [0,1]} a_j^2(t) \leq (\frac{1}{4})^j$ which is, of course, $O(j^{-4})$. Assumption 2 is also satisfied because $\sum_{j=1}^{\infty} \sup_{t \in [0,1]} |a'_j(t)|^2 = \sum_{j=1}^{\infty} \frac{j^2}{4^j} < \infty$. Next, the mean function $\mu(t)$ is taken to be a piecewise constant function with six change-points located at fractions $\frac{1}{6} \pm \frac{1}{36}$, $\frac{3}{36} \pm \frac{2}{36}$, and $\frac{5}{6} \pm \frac{3}{36}$ of the sample size n . In the first segment, $\mu_1 \equiv \mu(t_1) = 0$, in the second it is equal to 1, and in the remaining segments $\mu(t)$ alternates between 0 and 1, starting with 0 in the third segment. This mean function is very similar to the one that has been considered earlier in several other publications; see e.g. [4] and [30].

The second model we consider has exactly the same error structure as Model 1 but the mean function is a slightly different one. In particular, we make the value of the function in the second segment 2 instead of 1 while the remaining segments of $\mu(t)$ alternate between 0 and 1, starting with 0 in the third segment. Since the error process remains the same as before in Model 1, Assumptions 1 and 2 are satisfied.

The third model we consider is where the errors are generated by a locally stationary MA(2) process

$$x_i = \sum_{j=0}^2 a_j(t) \zeta_{i-j}$$

with coefficients being equal to $a_j(t) = \frac{(t+0.05)^j}{2^j}$, $j = 0, 1, 2$. The sequence (ζ_i) consists of i.i.d. $N(0, 0.3)$ random variables. The locally stationary MA process considered is a special case of the general locally stationary linear process. Since the process consists of the finite number of terms, the stochastic Lipschitz continuity condition in the Assumption 2 is satisfied automatically. Because $\sup_{t \in [0,1]} a_j^2(t) \leq (0.525)^{2j}$, the Assumption 1 will also be satisfied. Finally, the mean function stays the same as in the Model 1.

In Table 1, we illustrate coverage probabilities of uniform confidence intervals of the variance function and lag 1 covariance function for all three of the models considered with three sample sizes, $n = 400$, $n = 600$ and $n = 800$. Note that even a relatively small sample size of 400 gives excellent coverage probabilities. It is also worthwhile noting that the coverage probabilities are generally higher for lag 1 autocovariance function than for the variance function.

To investigate a possible breakdown threshold in terms of sample size, we also considered the Model 1 from our simulation setting for two smaller sample

TABLE 1
Empirical coverage probabilities for all three models.

Sample size		Model 1	Model 2	Model 3
$n = 400$	Variance	0.968	0.994	0.962
	Lag 1 autocovariance	0.994	0.998	0.998
$n = 600$	Variance	0.980	0.996	0.962
	Lag 1 autocovariance	0.996	1.000	0.998
$n = 800$	Variance	0.996	0.970	0.972
	Lag 1 autocovariance	0.996	0.996	0.998

sizes $n = 300$ and $n = 100$, with 500 replications in either case. The bandwidth was selected on a grid from 0.15 to 0.45 with the step of 0.01. When the sample size $n = 300$, the resulting empirical coverages were 0.936 for the variance and 0.978 for the lag 1 autocovariance. For the sample size $n = 100$, however, the resulting coverages became completely unsatisfactory; more specifically, they turned out to be 0.312 for the variance γ_0 and 0.51 for the lag 1 autocovariance. These results seem to imply that the sample size threshold necessary for such a breakdown is rather low.

To illustrate the behavior of uniform confidence intervals for each of the three models considered, we also include sample plots of fitted variance/autocovariance curves with corresponding confidence intervals in Figures 1–3. For each model, two plots are given: one with the true variance function, its estimate, and a uniform confidence interval for the estimated variance curve, while the other one contains the true lag 1 autocovariance function, its estimate, and the corresponding uniform confidence interval for the estimated autocovariance curve. In each of the plots, a solid line is used for the true variance/autocovariance curve, a dashed line for the corresponding estimated curve, and red dotted lines for uniform confidence intervals.

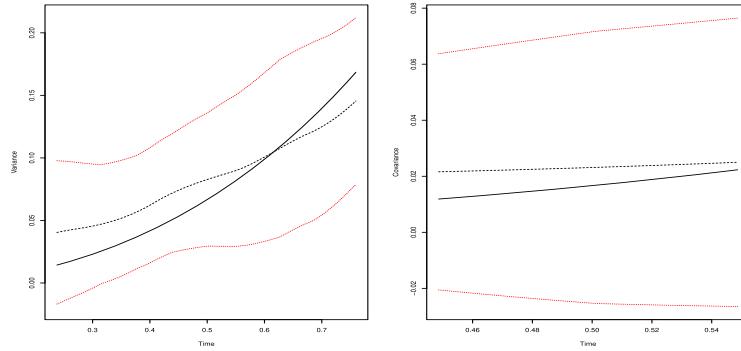


FIG 1. Estimated variance (left) and lag 1 autocovariance functions (right) for Model 1

To illustrate the reasonableness of our method, we also provide a quick comparison of our approach to a very straightforward “naive” method. Such a method would start with a rough estimate of the mean function $\mu(t)$ using

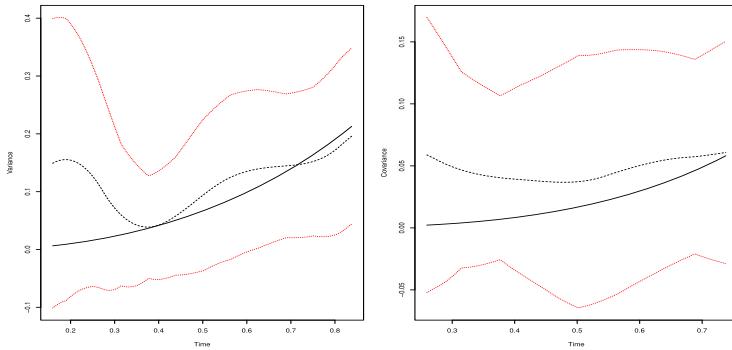


FIG 2. Estimated variance (left) and lag 1 autocovariance functions (right) for Model 2

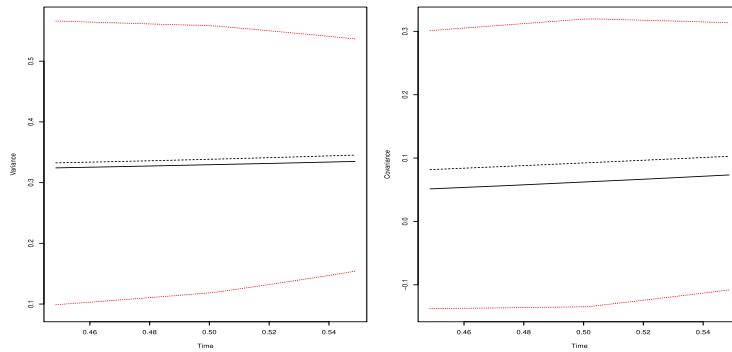


FIG 3. Estimated variance (left) and lag 1 autocovariance functions (right) for Model 3

a local smoother, for example, a local linear regression. The resulting rough estimate of the mean function can then be subtracted from observations to form a series of residuals e_i , $i = 1, \dots, n$. Using this series, a naive approach would estimate the variance function $\gamma_0(t)$ by applying a smoother, e.g. yet again the local linear regression, to squared residuals e_i^2 , $i = 1, \dots, n$. In much the same way, applying the local linear regression to a series $e_i e_{i-1}$, $i = 2, \dots, n$ will result in a “naive” estimate of the lag 1 autocovariance function $\gamma_1(t)$. In both situations, we used a simple generalized cross-validation to obtain the optimal smoothing bandwidth.

It is probably sufficient to say that such a naive approach fails completely in an attempt to estimate the second order structure when the mean is discontinuous and has numerous change points. More specifically, we tried to obtain the coverage of the true variance function $\gamma_0(t)$ by a uniform confidence interval that is based on the “naive” estimate described above. To do so, we used our Model 1 with the sample size $n = 400$ and 500 replications. We found that the coverage is zero, that is, the true variance function $\gamma_0(t)$ is never completely inside the uniform confidence interval. This can be explained properly by noticing

ing that our mean estimate used to obtain residuals is extremely crude. More specifically, in order for the local linear mean function estimator to be consistent at a given point the mean function $\mu(t)$ has to have two continuous derivatives at that point; see e.g. [14] p. 62 for a detailed discussion. This lack of consistency results in a severe bias of the variance function estimator. Thus, such a direct approach seems to be completely inappropriate for determination of the second order structure.

A more sophisticated comparison to existing methods is also desirable but is not easy to conceive of since our general model has not been considered in the literature earlier. Here, we pursue the following approach. First, we generate data from the nonparametric heteroscedastic regression model $y_{i,n} = \mu_{i,n} + x_{i,n}$ with correlated errors. The piecewise-constant mean is taken as in the Model 2 but the error process is now stationary. For easy comparison, we selected the 1-dependent stationary error process described in Section 6 of [30]. Now, it is necessary to estimate the (constant) variance and autocovariance at lag 1 γ_0 and γ_1 in such a model in two different ways.

- First, we apply the difference-based estimators of [30] and construct confidence intervals based on these estimators. To obtain these confidence intervals, we use the results from Theorem 1 of [30] where the bias and variance of these estimators are computed. Let B_i and V_i , $i = 0, 1$ be the bias and variance of $\hat{\gamma}_i$, $i = 0, 1$ obtained in the Theorem 1 [30] directly. Note that, since the depth of dependence $m = 1$, we select as the optimal difference coefficient $d = 1$ according to (2.14) in [30]. Whenever B_i and V_i depend on the true values of γ_i , we plug in the estimated values $\hat{\gamma}_i$ instead. Then, assuming asymptotic normality of their estimators, we can obtain $100(1 - \alpha)\%$ intervals, $0 < \alpha < 1$, for both true variance γ_0 and lag 1 autocovariance γ_1 as $\hat{\gamma}_i + B_i \pm z_{\alpha/2}V(\hat{\gamma}_i)$, $i = 0, 1$.
- Second, we use the difference-based estimators proposed in our manuscript to estimate the unknown γ_0 and γ_1 . This would have been, of course, a natural step if the stationarity of the data was not known in advance. Next, the confidence bands obtained in Corollaries 1 and 2 of our manuscript are used to compute the empirical coverage level of these estimators.

We proceed to compare the two approaches, choosing 500 replications for sample sizes $n = 400, 600$ and 800 . The value of the parameter defining the 1-dependent stationary error process of [30] is set to 0.2. To select bandwidth b_h and b_k , we use the grid from 0.1 to 0.3 with the step size 0.05 when the sample size $n = 400$, from 0.1 to 0.2 with the step size 0.05 when the sample size $n = 600$, and from 0.1 to 0.15 with the step size 0.05 when the sample size $n = 800$. The results are summarized in Tables 2 and 3.

The results suggest that using the method proposed for stationary data offers better empirical coverage probabilities than our method when applied to stationary data. A possible exception is only observed for larger sample sizes (e.g. $n = 600$ and $n = 800$) when estimating the variance function. Note also that, for the variance γ_0 , our method can provide a better performance as the sample size increases, while still remaining below the nominal level for both $n = 400$

TABLE 2
Empirical coverage probabilities for the variance and lag 1 autocovariance in the stationary model using our method

Sample size	400	600	800
Variance coverage	0.826	0.932	0.968
Lag 1 autocovariance	0.952	0.954	0.976

TABLE 3
Empirical coverage probabilities for the variance and lag 1 autocovariance in the stationary model using [30] method

Sample size	400	600	800
Variance coverage	0.906	0.912	0.924
Lag 1 autocovariance	0.972	0.972	0.978

and $n = 600$. This does not happen for lag 1 autocovariance γ_1 where even the sample size $n = 400$ is enough for our method to guarantee the promised coverage level.

Finally, it also seems desirable to see how the method performs if the mean function is more complicated than just a simple piecewise constant pattern. As a first step, we make the mean function more difficult to handle. Now, the mean function $\mu(t)$ is taken to be a piecewise linear function with six change-points located at fractions $\frac{1}{6} \pm \frac{1}{36}$, $\frac{3}{36} \pm \frac{2}{36}$, and $\frac{5}{6} \pm \frac{3}{36}$ of the sample size n . Adding 0 and n to the above sequence, we have a sequence τ_l where $l = 0, 1, \dots, 8$. We denote the mean function over a segment starting at τ_{2k} $k = 0, 1, 2, 3$ μ_{2k} and mean function over a segment starting at τ_{2k+1} , $k = 0, 1, 2$ μ_{2k+1} . In all of the even-numbered segments $\mu_{2k} \equiv 0$ while in all of the odd numbered segments the mean function is a straight line with the equation $\mu_{2k+1}(x) = \frac{x - \tau_{2k+1}}{\tau_{2k+2} - \tau_{2k+1}}$ for any $\tau_{2k+1} \leq x < \tau_{2k+2}$. Second, we set up the error process that would imitate the second order structure of the Northern hemisphere anomalies dataset that is used in our manuscript. Since the autocovariances at lags greater than 1 seem to be effectively zero for this dataset, we decided to make the variance and lag 1 autocovariance functions of the proposed synthetic process similar to those of that dataset. To do so, we use the locally linear MA (1) process $x_{i,n} = \sum_{j=0}^1 a_j(t) \zeta_{i-j}$ where ζ_i are iid $N(0, \sigma^2)$ with some variance σ^2 as our error process. The coefficient functions are selected as $a_j(t) = e^{-(j+1)(t+0.05)}$ for $j = 0, 1$. This implies immediately that the variance of this process is $\gamma_0(t) = \sigma^2[e^{-2(t+0.05)} + e^{-4(t+0.05)}]$ and the lag 1 covariance function is $\gamma_1 = e^{-3t-0.03-2/n}$. We select $\sigma^2 = 0.1$. The resulting mean function and the typical path of the resulting process $x_{i,n}$ are illustrated in the Fig. 4, while its variance and covariance functions are illustrated below in Fig. 5. Note that variance and covariance functions are, in particular, similar to those of the Northern anomalies dataset in Section 6 of our manuscript.

We used our approach to estimate the variance and lag 1 autocovariance functions of the resulting process for two different sample size: $n = 400$ and $n = 600$. First, let $n = 400$ and the number of replications equal to 500 replications. We select the bandwidths b_h and b_k using the grid consisting of only two points,

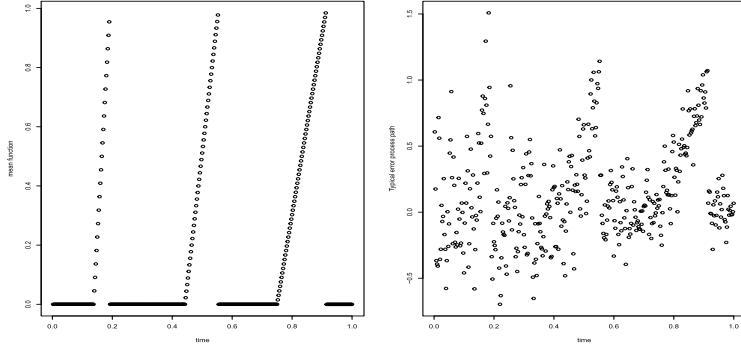


FIG 4. Mean function and the typical MA(1) error process path when $n = 400$

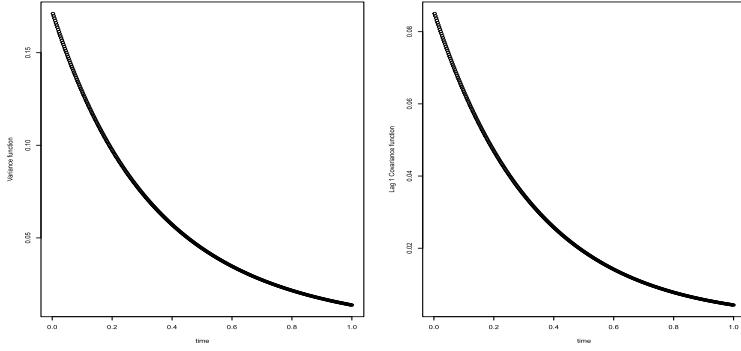


FIG 5. Variance function and lag 1 autocovariance function for the MA(1) error process when $n = 400$

0.10 and 0.15. Then, the resulting coverage is 96.2% for the variance function and 99% for the lag 1 autocovariance function. Using the same setting but with the sample size $n = 600$, we find the coverage 98.2% for the variance function and 99.4% for the lag 1 autocovariance function. Thus, our approach seems capable of handling more complicated (not just piecewise constant) mean functions and various second order structures.

6. Real data application

In this section, we illustrate our approach using a real dataset. There is a rather clear evidence that the global temperatures are nonstationary (see e.g. [28]) and so we use the dataset that consists of monthly temperature anomalies observed during the period from January 1856 to September 2019. A shorter subset of the same series has been used earlier in [27]. The data used are publicly available from the Climate Research Unit of the University of East Anglia, UK at <https://www.cru.uea.ac.uk/>. The anomalies are defined here as the difference

of temperatures from a reference value. The anomaly data are available for both Northern and Southern hemisphere separately. Fig. 6 displays the temperature anomaly data for both hemispheres.

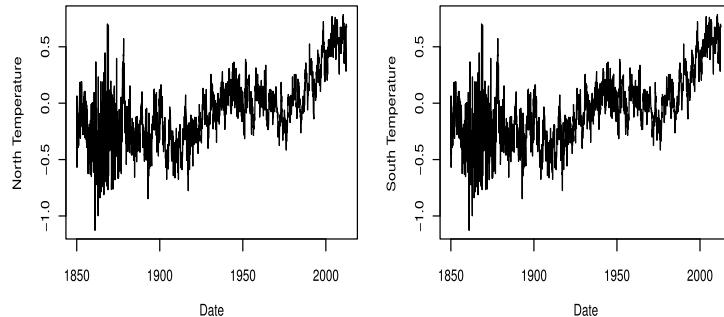


FIG 6. Recorded temperature anomalies for Northern (left) and Southern (right) hemispheres

Our purpose is to estimate the variance and lag 1 autocovariance function of these data as a function of time. For the Northern hemisphere data the approach suggested in our manuscript produces an almost monotonically decaying variance curve that suggests that some nonstationarity is, indeed, present in the data. This monotonic decay is probably due to the increasing number of weather stations recording the data over time. The variance of the Southern hemisphere data is also mostly decreasing although the decay is not as clearly monotonic as for the Northern hemisphere data. Note that, for both sets of data, the lag 1 autocovariance is very small in magnitude; however, the horizontal zero line added to both autocovariance plots is clearly not fully inside the uniform confidence interval, indicating that the temperature series are not white noises.

Now, we are interested in testing whether for these data there exist change points in the mean function, namely,

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_n = \mu \longleftrightarrow H_1 : \mu_i \neq \mu_j \text{ for some } 1 \leq i < j \leq n.$$

To this end, we will use the robust bootstrap test for nonstationary time series proposed in Section 4 of [36]. For Northern Hemisphere data, the robust bootstrap test yields a $< 0.1\%$ p -values with 10000 bootstrap samples, which provides a very strong evidence against the null hypothesis of no structural change in mean. On the other hand, we applied the robust bootstrap to the Southern Hemisphere data. The corresponding p -value of the test with 10000 bootstrap samples is also $< 0.1\%$, which also shows a strong evidence against H_0 . As a result, the test further illustrates the usefulness of our method for constructing SCB with finite change points. Over some time periods, the data with wild fluctuations indicates a change in mean and suggests the nonstationarity, as pointed by [28].

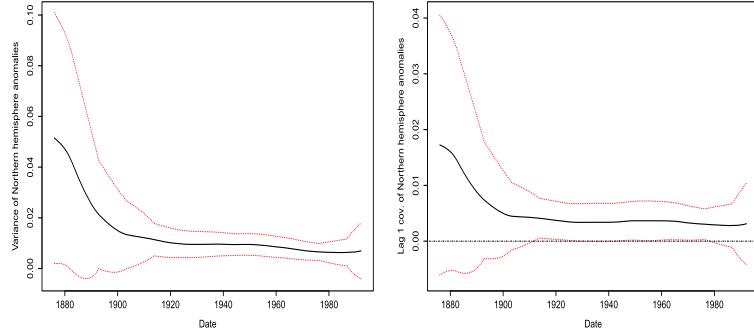


FIG 7. Variance (left) and lag 1 autocovariance function (right) of Northern hemisphere anomalies

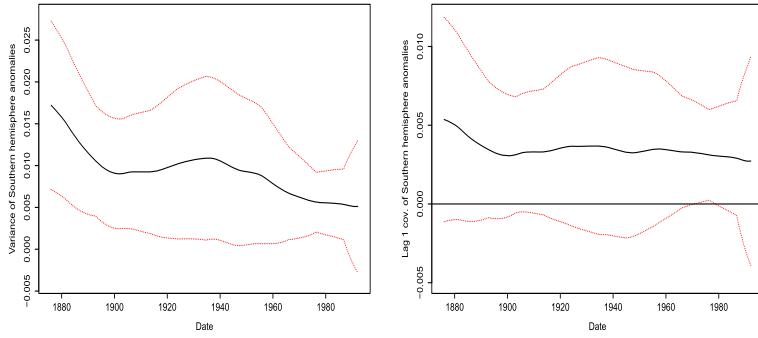


FIG 8. Variance (left) and lag 1 autocovariance function (right) of Southern hemisphere anomalies

7. Appendix

7.1. Additional results

The following theorem provides the Gaussian approximation result for nonstationary multiple time series, which can be found in Theorem 2 of [38].

Theorem 5 (Theorem 2 in [38]). *Suppose that Assumptions 1 and 2 hold. A partial sum process can be defined as $\tilde{S}_i^k = \sum_{j=1}^i \varepsilon_j^k$ for $k = 1, \dots, h$. Then, on a richer probability space, there exist i.i.d. standard normal random variables u_1, u_2, \dots , and a process \hat{S}_i^k such that $\{\tilde{S}_i^k\}_{i=1}^n \stackrel{\mathcal{D}}{=} \{\hat{S}_i^k\}_{i=1}^n$ and*

$$\max_{i \leq n} \left| \hat{S}_i^k - \sum_{j=1}^i \sigma_k(t_j) u_j \right| = \mathcal{O}_{\mathbb{P}} \left(n^{2/5} \log(n) \right),$$

where $\sigma_k(\cdot)$ is defined as $\sigma_k(t) = \left\{ \sum_{j=-\infty}^{\infty} \text{Cov}(H_k(t, \mathcal{F}_0), H_k(t, \mathcal{F}_j)) \right\}^{1/2}$.

Theorem 5 implies the Gaussian approximation for a partial sum of a locally stationary process. Note that, due to the result stated in Lemma 2 of the paper, the physical dependence measure $\delta_4(H_k, i)$ has different types of the polynomial decay under two circumstances, which is more complicated than that of Corollary 2 in [33]. But letting the order of the m -dependence sequence larger than k and making a careful check of the proof of Corollary 2 in [33], we can obtain the same argument. Owing to the non-stationarity, the approximated Gaussian process $\{\sum_{j=1}^i \sigma_k(t_j) u_j\}_{i=1}^n$ has independent but possibly non-identically distributed increments.

7.2. Technical details

Proof of Lemma 1. For $j \in \mathbb{Z}$, define the projection operator $\mathcal{P}_j(\cdot) = \mathbb{E}(\cdot | \mathcal{F}_j) - \mathbb{E}(\cdot | \mathcal{F}_{j-1})$, then we can write $x_i = \sum_{j=-\infty}^i \mathcal{P}_j(x_i)$. Denote $t_i = i/n$, we have

$$\begin{aligned} |\gamma_k(t_i)| &= \left| \mathbb{E} \left(\sum_{j=-\infty}^i \mathcal{P}_j(x_i) \sum_{j=-\infty}^{i-k} \mathcal{P}_j(x_{i-k}) \right) \right| \\ &\leq \left| \mathbb{E} \left(\sum_{j=-\infty}^{\infty} \mathcal{P}_j(x_i) \mathcal{P}_j(x_{i-k}) \right) \right| \\ &\leq \sum_{j=-\infty}^{\infty} \|\mathcal{P}_j(x_i)\|_2 \cdot \|\mathcal{P}_j(x_{i-k})\|_2 \\ &\leq \sum_{j=-\infty}^{i-k-1} \delta_2(G, i-j) \delta_2(G, i-k-j). \end{aligned}$$

The first inequality follows by the orthogonality of $\mathcal{P}_j(\cdot)$ and the second inequality is due to Fubini's theorem and Cauchy-Schwartz inequality. The last inequality follows from the argument in [32, Theorem 1]. Therefore, with Assumption 1, there exists a constant $C > 0$ such that $|\gamma_k(t_i)| \leq Ck^{-4}$. This concludes our proof.

Proof of Lemma 2. Now, we consider the locally stationary process $(x_l - x_{l-k})^2$ and let x'_l be the coupled process of x_l with ζ_0 replaced by an i.i.d. copy ζ'_0 . Then for each $k = 1, \dots, h$,

$$\begin{aligned} &\|H_k(t, \mathcal{F}_l) - H_k(t, \mathcal{F}'_l)\|_4 \\ &= \|(x_l - x_{l-k})^2 - (x'_l - x'_{l-k})^2\|_4 \\ &= \|(x_l - x_{l-k} + x'_l - x'_{l-k})(x_l - x_{l-k} - x'_l + x'_{l-k})\|_4 \\ &\leq \|x_l - x_{l-k} + x'_l - x'_{l-k}\|_8 \cdot \|x_l - x_{l-k} - x'_l + x'_{l-k}\|_8 \\ &\leq 4 \sup_{t \in [0,1]} \|x_l\|_8 \cdot (\sup_{t \in [0,1]} \|x_l - x'_l\|_8 + \sup_{t \in [0,1]} \|x_{l-k} - x'_{l-k}\|_8) \end{aligned}$$

$$\begin{aligned}
&= 4 \sup_{t \in [0,1]} \|x_l\|_8 \begin{cases} \delta_8(G, l), & \text{if } 0 < l < k, \\ \delta_8(G, l) + \delta_8(G, l - k + 1), & \text{if } l \geq k. \end{cases} \\
&= \begin{cases} \mathcal{O}(l^{-4}), & \text{if } 0 < l < k, \\ \mathcal{O}(l^{-4}) + \mathcal{O}((l - k + 1)^{-4}), & \text{if } l \geq k. \end{cases}
\end{aligned}$$

Proof of Lemma 3. By Assumption 3, we know that $\beta_k(\cdot)$ is also Lipschitz continuous, thus with Eq. (7),

$$\begin{aligned}
&\|H_k(t, \mathcal{F}_i) - H_k(s, \mathcal{F}_i)\|_2 \\
&= \|[G(t, \mathcal{F}_i) - G(t - \frac{k}{n}, \mathcal{F}_{i-k})]^2 - \beta_k(t) - [G(s, \mathcal{F}_i) - G(s - \frac{k}{n}, \mathcal{F}_{i-k})]^2 + \beta_k(s)\|_2 \\
&\leq \|\beta_k(t) - \beta_k(s)\|_2 + \|(G(t, \mathcal{F}_i) - G(t - \frac{k}{n}, \mathcal{F}_{i-k}) + G(s, \mathcal{F}_i) - G(s - \frac{k}{n}, \mathcal{F}_{i-k}))\|_4 \\
&\quad \cdot \|(G(t, \mathcal{F}_i) - G(t - \frac{k}{n}, \mathcal{F}_{i-k}) - G(s, \mathcal{F}_i) + G(s - \frac{k}{n}, \mathcal{F}_{i-k}))\|_4 \\
&\leq C|t - s| + 4 \sup_{t \in [0,1]} \|x_i\|_4 \cdot (\|G(t, \mathcal{F}_i) - G(s, \mathcal{F}_i)\|_4 \\
&\quad + \|G(t - \frac{k}{n}, \mathcal{F}_{i-k}) - G(s - \frac{k}{n}, \mathcal{F}_{i-k})\|_4) \\
&\leq C|t - s|.
\end{aligned}$$

The first inequality follows from the triangle inequality and Minkowski's inequality. The second inequality uses elementary calculation and the last line follows by Assumption 2. On the other hand, Assumption 3 implies that $\beta_k(t)$ is bounded on the compact $[0, 1]$. Then,

$$\begin{aligned}
\sup_{t \in [0,1]} \|H_k(t, \mathcal{F}_i)\|_4 &= \sup_{t \in [0,1]} \|(x_i - x_{i-k})(x_i - x_{i-k}) + \beta_k(t)\|_4 \\
&\leq (\sup_{t \in [0,1]} \|x_i - x_{i-k}\|_8)^2 + \sup_{t \in [0,1]} \|\beta_k(t)\|_4 \\
&\leq (\sup_{t \in [0,1]} \|x_i\|_8 + \sup_{t \in [0,1]} \|x_{i-k}\|_8)^2 + \sup_{t \in [0,1]} \|\beta_k(t)\|_4 \\
&< \infty.
\end{aligned}$$

To prove Theorem 1, we need to introduce the following lemmas.

Lemma 4. Let $F_n^k(t) = \sum_{i=1}^n u_i K_b(t_i - t)$, where $\{u_i\}_{i \in \mathbb{Z}}$ be i.i.d. $N(0, 1)$ random variables. Suppose that $b \rightarrow 0$ and $nb/\log(n) \rightarrow \infty$, $m^* = 1/b$. Then,

$$\begin{aligned}
\lim_{n \rightarrow \infty} \left(\mathbb{P} \left[\frac{1}{\sqrt{nb\phi_0}} \sup_{t \in \mathcal{T}} |F_n^k(t)| - B_K(m^*) \leq \frac{u}{\sqrt{2\log(m^*)}} \right] \right) \\
= \exp\{-2 \exp(-u)\}.
\end{aligned}$$

Proof. Similar to Lemma 2 in [38].

Lemma 5. Let $D_k(t) = \frac{1}{nb} \sum_{i=1}^n \varepsilon_i^k K_b(t_i - t)$. Assume that $\sigma_k(t)$ is Lipschitz continuous and bounded away from 0 on $[0, 1]$ and $\log(n)/n^{3/5}b + b\log(n) \rightarrow 0$. Then we have

$$\begin{aligned} \lim_{n \rightarrow \infty} & \left(\mathbb{P} \left[\sqrt{\frac{nb}{\phi_0}} \sup_{t \in \mathcal{T}} |\sigma_k^{-1}(t) D_k(t)| - B_K(m^*) \leq \frac{u}{\sqrt{2 \log(m^*)}} \right] \right) \\ & = \exp\{-2 \exp(-u)\}. \end{aligned}$$

Proof. By summation by part and the result of Theorem 5, we have

$$\sup_{t \in \mathcal{T}} |D_k(t) - \Xi_k(t)| \leq \left| \frac{1}{nb} \sum_{i=1}^n \left[\sum_{j=1}^i [\varepsilon_i^j - \sigma_k(t_j) u_j] K\left(\frac{t_i - t}{b}\right) \right] \right| = \mathcal{O}_{\mathbb{P}}\left(\frac{\log(n)}{n^{3/5}b}\right),$$

where $\Xi_k(t) = \frac{1}{nb} \sum_{i=1}^n \sigma_k(t_i) u_i K_b(t_i - t)$. Since $\sigma_k(\cdot)$ is Lipschitz continuous, then

$$\sup_{t \in \mathcal{T}} \left| \Xi_k(t) - \frac{1}{nb} \sigma_k(t) \sum_{i=1}^n u_i K\left(\frac{t_i - t}{b}\right) \right| = \mathcal{O}_{\mathbb{P}}\left(\sqrt{\frac{b \log(n)}{n}}\right). \quad (18)$$

Hence, Lemma 5 holds following the above equations and the arguments of Lemma 4.

Proof of Theorem 1. Recall Eqs. (13) to (15) and denote $\eta_k(t) = (\beta_k(t), b\beta'_k(t))^\top = [Q_n^k(t)]^{-1} R_n^k(t)$, we have

$$Q_n^k(t)(\hat{\eta}_k(t) - \eta_k(t)) = \begin{pmatrix} b^2 Q_{n,2}^k(t)\{\beta''_k(t)/2 + \mathcal{O}(b)\} \\ b^2 Q_{n,3}^k(t)\{\beta''_k(t)/2 + \mathcal{O}(b)\} \end{pmatrix} + \Lambda_n^k(t) + \Theta_n^k(t) + T_n^k(t), \quad (19)$$

where $T_n^k(t) = (T_{n,0}^k(t), T_{n,1}^k(t))^\top$, $\Lambda_n^k(t) = (\Lambda_{n,0}^k(t), \Lambda_{n,1}^k(t))^\top$, $\Theta_n^k(t) = (\Theta_{n,0}^k(t), \Theta_{n,1}^k(t))^\top$ and

$$\begin{aligned} T_{n,l}^k(t) &= \frac{1}{nb} \sum_{i=1}^n \varepsilon_i^k \left(\frac{t_i - t}{b} \right)^l K\left(\frac{t_i - t}{b}\right), \\ \Lambda_{n,l}^k(t) &= \frac{1}{nb} \sum_{i=1}^n \lambda_i^k \left(\frac{t_i - t}{b} \right)^l K\left(\frac{t_i - t}{b}\right), \\ \Theta_{n,l}^k(t) &= \frac{1}{nb} \sum_{i=1}^n \theta_i^k \left(\frac{t_i - t}{b} \right)^l K\left(\frac{t_i - t}{b}\right), \quad l = 0, 1. \end{aligned}$$

As a consequence of the weak law of large numbers, we know that $Q_n^k(t) \xrightarrow{\mathcal{P}} Q(t)$ in the sense that each element converges in probability, where

$$Q(t) = \begin{pmatrix} 1 & 0 \\ 0 & \mu_2 \end{pmatrix}.$$

Hence, from Eq. (19) and the technical proofs of Theorem 2, we have

$$\sqrt{nb} \sup_{t \in \mathcal{T}} |\hat{\beta}_{k,b}(t) - \beta_k(t) - T_{n,0}^k(t)|$$

$$\begin{aligned}
&= \sqrt{nb} \sup_{t \in \mathcal{T}} |b^2 Q_{n,2}^k(t) \{\beta_k''(t)/2 + \mathcal{O}(b)\} + \Lambda_{n,0}^k(t) + \Theta_{n,0}^k(t)| \\
&= \mathcal{O}_{\mathbb{P}}(\sqrt{nb^5}) + \mathcal{O}_{\mathbb{P}}\left(\frac{\log(n)}{n^{3/8-2\beta}\sqrt{b}}\right) + \mathcal{O}_{\mathbb{P}}\left(\frac{\log(n)}{n^{1/4-\beta}\sqrt{b}}\right) \xrightarrow{\mathcal{P}} 0.
\end{aligned}$$

Furthermore by the Proposition 6 in [37], we have for any $k = 1, \dots, h$,

$$(nb)^{1/2} T_n^k(t) \Rightarrow N(0, \nu_k(t)), \quad \nu_k(t) = \sigma_k^2(t) \operatorname{diag}(\phi_0, \phi_2).$$

Next, we treat $T_{n,0}^k(t)$ as $D_k(t)$ in Lemma 5, then Theorem 1 follows.

Proof of Theorem 2. Recall that our model contains $d = \mathcal{O}(n^\alpha)$ change points with the maximal size $\Delta_n = \mathcal{O}(n^\beta)$ on $\mu(\cdot)$. By condition $\iota > 2b$, we know there exists at most one jump on the b -neighborhood of any time t . For each $k = 1, \dots, h$ and any fixed $t \in [0, 1]$, by Eqs.(6), (7) and (12) in the paper, we have

$$\begin{aligned}
&\sup_{t \in \mathcal{T}} \left| \widehat{\beta}_{k,b}(t) - \beta_k(t) \right| \\
&= \sup_{t \in \mathcal{T}} \left| \sum_{i=1}^n \omega_n^b(t, i)(\beta_k(t_i) + \varepsilon_i^k + \lambda_i^k + \theta_i^k) - \sum_{i=1}^n \omega_n^b(t, i)\beta_k(t) \right| \\
&= \sup_{t \in \mathcal{T}} \left| \sum_{i=1}^n \omega_n^b(t, i)\beta'_k(t_i)(t - t_i) + \sum_{i=1}^n \omega_n^b(t, i) \left[\frac{\beta_k''(t_i)}{2} + \mathcal{O}(b) \right] (t - t_i)^2 \right. \\
&\quad \left. + \sum_{i=1}^n \omega_n^b(t, i)\varepsilon_i^k + \sum_{i=1}^n \omega_n^b(t, i)\lambda_i^k + \sum_{i=1}^n \omega_n^b(t, i)\theta_i^k \right| \\
&\leq \sup_{t \in \mathcal{T}} \left| \sum_{i=1}^n \omega_n^b(t, i) \left[\frac{\beta_k''(t_i)}{2} + \mathcal{O}(b) \right] (t - t_i)^2 \right| + \sup_{t \in \mathcal{T}} \left| \sum_{i=1}^n \omega_n^b(t, i)\varepsilon_i^k \right| \\
&\quad + \sup_{t \in \mathcal{T}} \left| \sum_{i=1}^n \omega_n^b(t, i)\lambda_i^k \right| + \sup_{t \in \mathcal{T}} \left| \sum_{i=1}^n \omega_n^b(t, i)\theta_i^k \right| := I + II + III + IV.
\end{aligned}$$

It is obvious to see $I = \mathcal{O}_{\mathbb{P}}(b^2)$. Then we will apply a chaining argument for calculating II . For $t \in [0, 1]$, define the sampling time points as $t = s/n$, $s = 0, 1, \dots, n$ and let $\Pi_t = \sum_{i=1}^n \omega_n^b(t, i)\varepsilon_i^k$. Then for each time point s/n , we have

$$\begin{aligned}
&\max_{0 \leq s \leq n} |\Pi_{\frac{s}{n}}| = \max_{0 \leq s \leq n} \left| \sum_{i=1}^n \omega_n^b\left(\frac{s}{n}, i\right) \varepsilon_i^k \right| \\
&\leq \max_{0 \leq s \leq n} \left| \sum_{i=1}^n \omega_n^b\left(\frac{s}{n}, i\right) [\varepsilon_i^k - \sigma_k(t_i)u_i] \right| + \max_{0 \leq s \leq n} \left| \sum_{i=1}^n \omega_n^b\left(\frac{s}{n}, i\right) \sigma_k(t_i)u_i \right| \\
&= \max_{0 \leq s \leq n} \left| \sum_{i=1}^n \omega_n^b\left(\frac{s}{n}, i\right) [\varepsilon_i^k - \sigma_k(t_i)u_i] \right|
\end{aligned}$$

$$\begin{aligned}
& \left| -\sum_{j=2}^n [\omega_n^b(\frac{s}{n}, j) - \omega_n^b(\frac{s}{n}, j-1)] \sum_{i=1}^j [\varepsilon_i - \sigma_k(t_i)u_i] \right| \\
& + \max_{0 \leq s \leq n} \left| \sum_{i=1}^n \omega_n^b\left(\frac{s}{n}, i\right) \sigma_k(t_i)u_i \right| \\
& \leq \max_{0 \leq s \leq n} \left\{ \max_{j \leq n} \left| \sum_{i=1}^j [\varepsilon_i - \sigma(t_i)u_i] \right| \left[\left| \omega_n^b(\frac{s}{n}, n) \right| + \sum_{j=2}^n \left| \omega_n^b(\frac{s}{n}, j) - \omega_n^b(\frac{s}{n}, j-1) \right| \right] \right\} \\
& + \max_{0 \leq s \leq n} \left| \sum_{i=1}^n \omega_n^b\left(\frac{s}{n}, i\right) \sigma_k(t_i)u_i \right| \\
& = \mathcal{O}_{\mathbb{P}}\left(\frac{n^{2/5} \log n}{nb}\right) + \mathcal{O}_{\mathbb{P}}\left(\sqrt{\frac{\log n}{nb}}\right) = \mathcal{O}_{\mathbb{P}}\left(\frac{\log n}{n^{3/5}b} + \sqrt{\frac{\log n}{nb}}\right).
\end{aligned}$$

The first inequality uses Theorem 5 and the triangle inequality. Next, we consider the difference between Π_t and $\Pi_{\frac{s}{n}}$.

$$\begin{aligned}
& \sup_{t \in [s/n, (s+1)/n]} |\Pi_t - \Pi_{\frac{s}{n}}| \\
& = \sup_{t \in [s/n, (s+1)/n]} \left| \sum_{i=1}^n [\omega_n^b(t, i) - \omega_n^b(\frac{s}{n}, i)] \varepsilon_i^k \right| \\
& = \sup_{t \in [s/n, (s+1)/n]} \int_{\frac{s}{n}}^t \left| \sum_{i=1}^n \omega_n^{b'}(v, i) \varepsilon_i^k \right| dv \\
& \leq \int_{\frac{s}{n}}^{\frac{s+1}{n}} \left| \sum_{i=1}^n \omega_n^{b'}(v, i) \varepsilon_i^k \right| dv \\
& = \left| \sum_{i=1}^n \left| \omega_n^b\left(\frac{s+1}{n}, i\right) - \omega_n^b\left(\frac{s}{n}, i\right) \right| \varepsilon_i^k \right| \\
& \leq \left| \sum_{i=1}^n \left| \omega_n^b\left(\frac{s+1}{n}, i\right) - \omega_n^b\left(\frac{s}{n}, i\right) \right| [\varepsilon_i^k - \sigma_k(t_i)u_i] \right| \\
& + \left| \sum_{i=1}^n \left| \omega_n^b\left(\frac{s+1}{n}, i\right) - \omega_n^b\left(\frac{s}{n}, i\right) \right| \sigma_k(t_i)u_i \right| \\
& = \mathcal{O}_{\mathbb{P}}\left(\frac{n^{2/5} \log n}{n^2 b^2}\right) + \mathcal{O}_{\mathbb{P}}\left(\frac{\sqrt{\log n}}{nb}\right).
\end{aligned}$$

Thus, we have

$$\begin{aligned}
II & = \sup_{t \in \mathcal{T}} |\Pi_t| = \sup_{t \in \mathcal{T}} |\Pi_t - \Pi_{\frac{s}{n}} + \Pi_{\frac{s}{n}}| \\
& \leq \max_{0 \leq s \leq n} \sup_{t \in [s/n, (s+1)/n]} |\Pi_t - \Pi_{\frac{s}{n}}| + \max_{0 \leq s \leq n} |\Pi_{\frac{s}{n}}|
\end{aligned}$$

$$= \mathcal{O}_{\mathbb{P}} \left(\sqrt{\frac{\log n}{nb}} \right) + \mathcal{O}_{\mathbb{P}} \left(\frac{\log n}{n^{3/5}b} \right).$$

As for III , we consider the following two cases.

- When $\mu(t)$ is continuous on $[0, 1]$. Rewrite $\lambda_k(t_i) = [\mu(t_i) - \mu(t_{i-k})]^2$, then we know function λ_k is also continuous for $t \in [0, 1]$. Moreover, $\lambda'_k = \mathcal{O}(\frac{1}{n^2})$ and $\lambda''_k = \mathcal{O}(\frac{1}{n^2})$. By Taylor's expansion, we have

$$\begin{aligned} & \sup_{t \in \mathcal{T}} \left| \sum_{i=1}^n \omega_n^b(t, i) \lambda_i^k \right| \\ &= \sup_{t \in \mathcal{T}} \left| \omega_n^b(t, i) [\lambda_k(t) + \lambda'_k(t)(t - t_i) + \lambda''_k(t)(t - t_i)^2/2 + \mathcal{O}(b^3/n^2)] \right| \\ &= \sup_{t \in \mathcal{T}} |\lambda_k(t) + \mathcal{O}(b^2/n^2) + \mathcal{O}(b^2/n^3)| \\ &\leq C \left(\frac{k^2}{n^2} + b^2/n^2 + b^2/n^3 \right) = \mathcal{O} \left(\frac{n^{1/4} \log^2(n)}{n^2} \right). \end{aligned}$$

- When $\mu(t)$ is a piecewise continuous function with the maximal jump size $\Delta_n = \mathcal{O}(n^\beta)$. Let \mathcal{T}_0 be the union of the b -neighborhood of all change points, that is $[a_1 - b, a_1 + b] \cup \dots \cup [a_d - b, a_d + b]$.

$$\begin{aligned} & \sup_{t \in \mathcal{T}} \left| \sum_{i=1}^n \omega_n^b(t, i) \lambda_i^k \right| \\ &= \max \left\{ \sup_{t \in \mathcal{T} \setminus \mathcal{T}_0} \left| \sum_{i=1}^n \omega_n^b(t, i) \lambda_i^k \right|, \sup_{t \in \mathcal{T}_0} \left| \sum_{i=1}^n \omega_n^b(t, i) \lambda_i^k \right| \right\} \\ &= \max \left\{ \sup_{t \in \mathcal{T} \setminus \mathcal{T}_0} \left| \sum_{i=1}^n \omega_n^b(t, i) \lambda_i^k \right|, \sup_{t \in \mathcal{T}_0} \left| \sum_{i=n(t-b)}^{n(t+b)} \omega_n^b(t, i) [\mu(t_i) - \mu(t_{i-k})]^2 \right| \right\} \\ &= \max \left\{ \mathcal{O} \left(\frac{n^{1/4} \log^2(n)}{n^2} \right), \mathcal{O} \left(\frac{kn^{2\beta}}{nb} + \frac{1}{nb}(2nb - k) \frac{n^{1/4} \log^2(n)}{n^2} \right) \right\} \\ &= \mathcal{O} \left(\frac{\log(n)}{n^{7/8-2\beta}b} \right) \end{aligned}$$

In summary, $III = \mathcal{O} \left(\frac{\log(n)}{n^{7/8-2\beta}b} \right)$.

As for $IV = \sup_{t \in \mathcal{T}} |\sum_{i=1}^n \omega_n^b(t, i) \theta_i^k|$ where $\theta_i^k = (\mu_i - \mu_{i-k})(x_i - x_{i-k})$. Similar to previous proof, we consider two cases: when $\mu(t)$ is continuous on $[0, 1]$ or piecewise continuous function with maximal jump size $\mathcal{O}(n^\beta)$.

- When $\mu(t)$ is continuous on $[0, 1]$. Denote $\tilde{\lambda}_i^k = \tilde{\lambda}_k(t_i) = \mu(t_i) - \mu(t_{i-k})$, then we know function $\tilde{\lambda}_k$ is also continuous for $t \in [0, 1]$. Moreover, $\tilde{\lambda}'_k = \mathcal{O}(\frac{1}{n})$ and $\tilde{\lambda}''_k = \mathcal{O}(\frac{1}{n})$. Since $\sup_i \|x_i\|_8 < \infty$, we have $\max_{1 \leq i \leq n} |x_i| =$

$$\mathcal{O}_{\mathbb{P}}(n^{1/8}).$$

$$\begin{aligned} IV &= \sup_{t \in \mathcal{T}} \left| \sum_{i=1}^n \omega_n^b(t, i) \theta_i^k \right| \leq \max_{1 \leq i \leq n} |x_i| \sup_{t \in \mathcal{T}} \left| \sum_{i=1}^n \omega_n^b(t, i) \tilde{\lambda}_i^k \right| \\ &\leq Cn^{1/8} \sup_{t \in \mathcal{T}} \left| \sum_{i=1}^n \omega_n^b(t, i) [\tilde{\lambda}_k(t) + \tilde{\lambda}'_k(t)(t - t_i) \right. \\ &\quad \left. + \tilde{\lambda}''_k(t)(t - t_i)^2/2 + \mathcal{O}(b^3/n)] \right| \\ &= Cn^{1/8} \sup_{t \in \mathcal{T}} \left| \tilde{\lambda}_k(t) + \mathcal{O}(b^2/n) + \mathcal{O}(b^2/n^2) \right| = \mathcal{O}(n^{-3/4} \log(n)). \end{aligned}$$

- When $\mu(t)$ is a piecewise continuous function.

$$\begin{aligned} &\sup_{t \in \mathcal{T}} \left| \sum_{i=1}^n \omega_n^b(t, i) \theta_i^k \right| \\ &= \max \left\{ \sup_{t \in \mathcal{T} \setminus \mathcal{T}_0} \left| \sum_{i=1}^n \omega_n^b(t, i) \theta_i^k \right|, \sup_{t \in \mathcal{T}_0} \left| \sum_{i=1}^n \omega_n^b(t, i) \theta_i^k \right| \right\} \\ &\leq Cn^{1/8} \max \left\{ \sup_{t \in \mathcal{T} \setminus \mathcal{T}_0} \left| \sum_{i=1}^n \omega_n^b(t, i) \tilde{\lambda}_i^k \right|, \sup_{t \in \mathcal{T}_0} \left| \sum_{i=n(t-b)}^{n(t+b)} \omega_n^b(t, i) \tilde{\lambda}_i^k \right| \right\} \\ &\leq Cn^{1/8} \max \left\{ \mathcal{O}(n^{-7/8} \log(n)), \mathcal{O}\left(\frac{kn^\beta}{nb} + \frac{1}{nb}(2nb - k)n^{-7/8} \log(n)\right) \right\} \\ &= \mathcal{O}_{\mathbb{P}}\left(\frac{\log(n)}{n^{3/4-\beta} b}\right). \end{aligned}$$

As a result, $IV = \mathcal{O}_{\mathbb{P}}\left(\frac{\log(n)}{n^{3/4-\beta} b}\right)$. With conditions imposed in the theorem, the above four error bounds all converge to 0 as $n \rightarrow \infty$.

Proof of Theorem 3. Similar to the proof of Theorem 4 in [38].

Proof of Theorem 4. Let \mathcal{I}' be a closed interval in $(0, 1)$ such that $\mathcal{I} \subset \mathcal{I}'$ and the two intervals do not share common end points. Recall $\tilde{\beta}_k(t) = (\beta_h(t), \beta_k(t))^\top$ and denote $\hat{\beta}_{k,b_k}(t) = (\hat{\beta}_{h,b_k}(t), \hat{\beta}_{k,b_k}(t))^\top$. According to Theorem 2, it follows that

$$\sup_{i/n \in \mathcal{I}'} |\hat{\varepsilon}_i - \tilde{\varepsilon}_i| = \sup_{i/n \in \mathcal{I}'} \left| \tilde{\beta}_k(t_i) - \hat{\beta}_{k,b_k}(t_i) \right| = \mathcal{O}_{\mathbb{P}}(\chi_n). \quad (20)$$

Note that $Q_i^k/(2m+1)$ is the Nadaraya-Waston smoother of the series $\{\tilde{\varepsilon}_i\}$ at i with the rectangle kernel and bandwidth m/n . Therefore, for each $k = 1, \dots, h$, we have

$$\sup_{i/n \in \mathcal{I}'} |Q_i^k| = \mathcal{O}_{\mathbb{P}}(\sqrt{m \log n}). \quad (21)$$

Let $\widehat{Q}_i^k = \sum_{j=-m}^m \widehat{\varepsilon}_{i+j}$ and $\widehat{N}_i^k = \widehat{Q}_i^k \widehat{Q}_i^{k\top} / (2m + 1)$. Then

$$(2m + 1)(N_i^k - \widehat{N}_i^k) = (Q_i^k - \widehat{Q}_i^k)(Q_i^k)^\top + \widehat{Q}_i^k(Q_i^k - \widehat{Q}_i^k)^\top.$$

Substituting equations (20) and (21) into the above equation, we have $\sup_{i/n \in \mathcal{I}'} |N_i^k - \widehat{N}_i^k| = \mathcal{O}_{\mathbb{P}}(\nu_n)$ with the assumption $\nu_n \rightarrow 0$. By the definitions of $\widetilde{\Sigma}_k(t)$ and $\widehat{\Sigma}_k(t)$, we obtain

$$\sup_{i/n \in \mathcal{I}'} |\widetilde{\Sigma}_k(t) - \widehat{\Sigma}_k(t)| = \mathcal{O}_{\mathbb{P}}(\nu_n).$$

Together with the results of Theorem 3, Theorem 4 holds.

References

- [1] BROCKWELL, P. J. and DAVIS, R. A. (2016). *Introduction to Time Series and Forecasting*. Springer. [MR3526245](#)
- [2] BROWN, L. D. and LEVINE, M. (2007). Variance estimation in nonparametric regression via the difference sequence method. *The Annals of Statistics* **35** 2219–2232. [MR2363969](#)
- [3] CAI, T. T., LEVINE, M. and WANG, L. (2009). Variance function estimation in multivariate nonparametric regression with fixed design. *Journal of Multivariate Analysis* **100** 126–136. [MR2460482](#)
- [4] CHAKAR, S., LEBARBIER, E., LÉVY-LEDUC, C. and ROBIN, S. (2017). A robust approach for estimating change-points in the mean of an AR(1) process. *Bernoulli* **23** 1408–1447. [MR3606770](#)
- [5] CRAVEN, P. and WAHBA, G. (1978). Smoothing noisy data with spline functions. *Numerische mathematik* **31** 377–403. [MR0405795](#)
- [6] DAHLHAUS, R. (1997). Fitting time series models to nonstationary processes. *The Annals of Statistics* **25** 1–37. [MR1429916](#)
- [7] DAI, W., TONG, T. and ZHU, L. (2017). On the choice of difference sequence in a unified framework for variance estimation in nonparametric regression. *Statistical Science* **32** 455–468. [MR3696005](#)
- [8] DETTE, H., PREUSS, P. and VETTER, M. (2010). A measure of stationarity in locally stationary processes with applications to testing. *Journal of the American Statistical Association* **106** 1113–1124. [MR2894768](#)
- [9] DETTE, H., WU, W. and ZHOU, Z. (2019). Change Point Analysis of Correlation in Non-stationary Time Series. *Statistica Sinica* **29** 611–643. [MR3931381](#)
- [10] DIMITRIS N. POLITIS, J. P. R. and WOLF, M. (1999). *Subsampling*. New York: Springer. [MR1707286](#)
- [11] DING, X. and ZHOU, Z. (2019). Estimation and inference for precision matrices of non-stationary time series. *The Annals of Statistics*. arXiv:1803.01188. [MR4134802](#)
- [12] DWIVEDI, Y. and RAO, S. S. (2011). A test for second-order stationarity of a time series based on the discrete Fourier transform. *Journal of Time Series Analysis* **32** 68–91. [MR2790673](#)

- [13] EUBANK, R. L. and SPIEGELMAN, C. H. (1990). Testing the goodness of fit of a linear model via nonparametric regression techniques. *Journal of the American Statistical Association* **85** 387–392. [MR1141739](#)
- [14] FAN, J. and GIJBELS, I. (1994). Local Polynomial Modelling and Its Applications. London: Chapman and Hall. [MR1383587](#)
- [15] GASSER, T., KNEIP, A. and KÖHLER, W. (1991). A flexible and fast method for automatic smoothing. *Journal of the American Statistical Association* **86** 643–652. [MR1147088](#)
- [16] HALL, P. and KEILEGOM, I. V. (2003). Using difference-based methods for inference in nonparametric regression with time series errors. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **65** 443–456. [MR1983757](#)
- [17] HAMILTON, J. D. (1994). *Time Series Analysis*. Princeton University Press. [MR1278033](#)
- [18] HÄRDLE, W. and TSYBAKOV, A. (1997). Local polynomial estimators of the volatility function in nonparametric autoregression. *Journal of Econometrics* **81** 223–242. [MR1484586](#)
- [19] HERRMANN, E., GASSER, T. and KNEIP, A. (1992). Choice of bandwidth for kernel regression when residuals are correlated. *Biometrika* **79** 783–795. [MR1209478](#)
- [20] MÜLLER, H.-G. and STADTMÜLLER, U. (1987). Estimation of heteroscedasticity in regression analysis. *The Annals of Statistics* **15** 610–625. [MR0888429](#)
- [21] MÜLLER, H.-G. and STADTMÜLLER, U. (1988). Detecting dependencies in smooth regression models. *Biometrika* **75** 639–650. [MR0995108](#)
- [22] MÜLLER, H.-G. and STADTMÜLLER, U. (1993). On variance function estimation with quadratic forms. *Journal of Statistical Planning and Inference* **35** 213–231. [MR1220417](#)
- [23] NASON, G. (2013). A test for second-order stationarity and approximate confidence intervals for localized autocovariances for locally stationary time series. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* **75** 879–904. [MR3124795](#)
- [24] NASON, G. P., von SACHS, R. and KROISANDT, G. (2000). Wavelet processes and adaptive estimation of the evolutionary wavelet spectrum. *Journal of the Royal Statistical Society. Series B. Statistical Methodology* **62** 271–292. [MR1749539](#)
- [25] PAPARODITIS, E. (2010). Validating stationarity assumptions in time series analysis by rolling local periodograms. *Journal of the American Statistical Association* **105** 839–851. [MR2724865](#)
- [26] PARK, B. U., LEE, Y. K., KIM, T. Y. and PARK, C. (2006). A Simple Estimator of Error Correlation in Non-parametric Regression Models. *Scandinavian Journal of Statistics* **33** 451–462. [MR2298059](#)
- [27] RAO, S. S. (2004). On multiple regression models with nonstationary correlated errors. *Biometrika* **91** 645–659. [MR2090628](#)
- [28] RAO, T. S. and TSOLAKI, E. (2004). Nonstationary time series analysis of monthly global temperature anomalies. In *Time Series Analysis and*

- Applications to Geophysical Systems* 73–103. Springer. [MR1624100](#)
- [29] RICE, J. et al. (1984). Bandwidth choice for nonparametric regression. *The Annals of Statistics* **12** 1215–1230. [MR0760684](#)
- [30] TECUAPETLA-GÓMEZ, I. and MUNK, A. (2017). Autocovariance Estimation in Regression with a Discontinuous Signal and m -Dependent Errors: A Difference-Based Approach. *Scandinavian Journal of Statistics* **44** 346–368. [MR3658518](#)
- [31] WANG, L., BROWN, L. D., CAI, T. T. and LEVINE, M. (2008). Effect of mean on variance function estimation in nonparametric regression. *The Annals of Statistics* **36** 646–664. [MR2396810](#)
- [32] WU, W. B. (2005). Nonlinear system theory: another look at dependence. *Proceedings of the National Academy of Sciences of the United States of America* **102** 14150–14154. [MR2172215](#)
- [33] WU, W. B. and ZHOU, Z. (2011). Gaussian approximations for nonstationary multiple time series. *Statistica Sinica* **21** 1397–1413. [MR2827528](#)
- [34] ZHOU, Y., CHENG, Y., WANG, L. and TONG, T. (2015). Optimal difference-based variance estimation in heteroscedastic nonparametric regression. *Statistica Sinica* **25** 1377–1397. [MR3409072](#)
- [35] ZHOU, Z. (2010). Nonparametric inference of quantile curves for nonstationary time series. *The Annals of Statistics* **38** 2187–2217. [MR2676887](#)
- [36] ZHOU, Z. (2013). Heteroscedasticity and Autocorrelation Robust Structural Change Detection. *Journal of the American Statistical Association* **108** 726–740. [MR3174655](#)
- [37] ZHOU, Z. and WU, W. B. (2009). Local linear quantile estimation for nonstationary time series. *The Annals of Statistics* **37** 2696–2729. [MR2541444](#)
- [38] ZHOU, Z. and WU, W. B. (2010). Simultaneous inference of linear models with time varying coefficients. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*. **72** 513–531. [MR2758526](#)