



State-domain change point detection for nonlinear time series regression



Yan Cui^{a,b,*}, Jun Yang^c, Zhou Zhou^d

^a Institute for Advanced Study in Mathematics, Harbin Institute of Technology, China

^b School of Mathematics, Jilin University, China

^c Department of Statistics, University of Oxford, United Kingdom

^d Department of Statistical Sciences, University of Toronto, Canada

ARTICLE INFO

Article history:

Received 23 May 2019

Received in revised form 21 November 2021

Accepted 22 November 2021

Available online 17 December 2021

Keywords:

Change-point detection

Nonlinear time series

Nonparametric hypothesis test

State domain

ABSTRACT

Change point detection in time series has attracted substantial interest, but most of the existing results have been focused on detecting change points in the time domain. This paper considers the situation where nonlinear time series have potential change points in the state domain. We apply a density-weighted anti-symmetric kernel function to the state domain and therefore propose a nonparametric procedure to test the existence of change points. When the existence of change points is affirmative, we further introduce an algorithm to estimate the number of change points together with their locations. Theoretical results of the proposed detection and estimation procedures are given and a real dataset is used to illustrate our methods.

© 2021 Elsevier B.V. All rights reserved.

1. Introduction

Consider the following state-domain nonlinear auto-regression

$$X_i = \mu(X_{i-1}) + \epsilon_i, \quad (1)$$

where $\mu(\cdot)$ is an unknown regression function, $\{\epsilon_i\}$ is a martingale difference sequence such that $\mathbb{E}[\epsilon_i | (\epsilon_{i-1}, \epsilon_{i-2}, \dots)] = 0$ almost surely. Special cases of Eq. (1) include threshold AR models (Tong, 1990), exponential AR models (Haggan and Ozaki, 1981) and ARCH models (Engle, 1982), among others. Furthermore, Eq. (1) can be viewed as a discretized version of the diffusion model

$$dX_t = \mu(X_t)dt + dM(t), \quad (2)$$

where $\mu(\cdot)$ is the instantaneous return or drift function, and $\{M(t)\}$ is a continuous-time martingale. In the literature, the special case of Model (2) with $dM(t) = \sigma(X_t)d\mathbb{B}(t)$ has been widely discussed to understand and model nonlinear temporal systems in economics and finance, where $\mathbb{B}(t)$ denotes the standard Brownian motion and $\sigma^2(\cdot)$ is understood as the volatility function. Among others, Stanton (1997), Chapman and Pearson (2000) and Fan and Zhang (2003) considered the nonparametric estimation of $\mu(\cdot)$ and $\sigma^2(\cdot)$. Zhao (2011) addressed the model validation problem for Eq. (2). In particular, Eq. (2) can be used to model the temporal dynamics of financial data with $\{X_t\}$ being interest rates, exchange rates, stock prices or other economic quantities. Among others, Zhao and Wu (2006) considered kernel quantile estimates of Eq. (2)

* Corresponding author at: Institute for Advanced Study in Mathematics, Harbin Institute of Technology, China.

E-mail addresses: cui@utstat.toronto.edu (Y. Cui), jun.yang@stats.ox.ac.uk (J. Yang), zhou@utstat.toronto.edu (Z. Zhou).

for the exchange rates between Pound and USD. Liu and Wu (2010) constructed simultaneous confidence bands for $\mu(\cdot)$ and $\sigma(\cdot)$ with the U.S. Treasury yield curve rates data. See also the latter papers for further references. Observe that we allow the error process to be general martingale differences in (1) which significantly expands the applicability of our theory and methodology in economic applications. As pointed out by one referee, conditional moment restrictions in dynamic economic models routinely arise from Euler/Bellman equations in dynamic programming, which are martingale differences. Furthermore, asset returns, due to the no-arbitrage theory, are (semi)martingales. Hence, their (demeaned) returns are martingale differences.

Throughout this article, following Chapter 6.3 of Fan and Yao (2003), we shall call (1) a state-domain nonlinear regression model. The term “state domain” originated from the celebrated state-space models (e.g. Kalman (1960) and Shumway and Stoffer (2000, Chapter 6)) where the dynamics of a sequence of state variables ($\{X_i\}$ in Eq. (1)) are driven by a group of control variables (ϵ_i in Eq. (1)) through the nonlinear state Equation (1). Therefore in this article the term “state domain” refers to the Euclidean space in which the variables on the axes are the state variables. Observe that the state-domain nonlinear regression (1) aims to characterize the relationship between X_i and past values (states) of the time series through a discretized stochastic differential equation. On the contrary, time-domain nonlinear regression (see e.g. Fan and Yao (2003), Chapter 6.2)

$$X_i = f(i/n) + \varepsilon_i, \quad i = 1, 2, \dots, n \quad (3)$$

with $\mathbb{E}[\varepsilon_i] = 0$ describes the relationship between X_i and time.

To date, most investigations on the nonparametric inference procedure of Eq. (1) are based on the assumption that the underlying regression function $\mu(\cdot)$ is continuous, which may cause serious restrictions in many real applications. In fact, in parametric modeling of nonlinear time series, various choices of $\mu(\cdot)$ with possible discontinuities have drawn much attention in the literature. One of the most prominent examples is the threshold model proposed by Tong and Lim (1980), in which regime switches are triggered by an observed variable crossing an unknown threshold. Also, AR model with regime-switch controlled by a Markov chain mechanism was introduced by Tong (1990). In economics, the expanding phase and contracting phase are not always governed by the same dynamics, see Tiao and Tsay (1994), Durlauf and Johnson (1995), McConnell and Perez-Quiros (2000) and other references therein. As a result, the occurrence of abrupt changes in the state-domain regression function $\mu(\cdot)$ is common and detecting as well as estimating them are of vital importance. Motivated by this, in the current paper we focus on the situation where the regression function $\mu(\cdot)$ is piecewise smooth on an interval of interest $T = [l, u]$ with a finite but unknown number of change points. More precisely, there exist $l = a_0 < a_1 < \dots < a_M < a_{M+1} = u$ such that $\mu(\cdot)$ is smooth on each of the intervals $[a_0, a_1), \dots, [a_M, a_{M+1}]$; that is, on the interval $[l, u]$

$$\mu(x) = \sum_{j=0}^M \mu_j(x) \mathbb{1}(a_j \leq x < a_{j+1}), \quad (4)$$

where M is the total number of change points. Throughout this article, we assume M is fixed.

To our knowledge, there exist no results on change point detection of the state-domain regression function $\mu(\cdot)$ in the literature. The purpose of this paper is twofold. First we want to test whether $\mu(x)$ is smooth or discontinuous on the interval $[l, u]$; that is to test the null hypothesis $H_0 : M = 0$ of Eq. (4). By sliding a density-weighted anti-symmetric kernel through the state domain, we shall suggest a nonparametric test statistic and non-trivially apply the discretized multivariate Gaussian approximation result of Zaitsev (1987) to establish its asymptotic distribution. Additionally, the Gaussian approximation results also directly suggest a finite sample simulation-based bootstrapping method which improves the accuracy of the test in practical implementations. Second, if $M \geq 1$, we reject the null hypothesis and subsequently want to locate all the change points. In this case, we propose an estimation procedure and establish the corresponding asymptotic theory on the accuracy of the estimators. Finally, the above theoretical results are of general interest and could be used for a wider class of state-domain change point detection problems.

There is long-standing literature in statistics discussing jump detection of the time-domain nonlinear regression model (3) where occasional jumps occur in an otherwise smoothly changing time trend $f(\cdot)$. It is impossible to show a complete reference here and we only list some representative works. Müller (1992) and Eubank and Speckman (1994) employed a kernel method to estimate jump points in smooth curves. Wang (1995) suggested using wavelets and provided a review of jump-point estimation. Two-step methods were considered by Müller and Song (1997) and Gijbels et al. (1999) to study the asymptotic convergence properties of the jumps. Later, Gijbels et al. (2007) suggested a compromise estimation method which can preserve possible jumps in the curve. Zhang (2016) considered the situation where the trend function allows a growing number of jump points. In econometrics, there is a significant body of literature discussing time-domain jump detection in jump diffusion models; see for instance Bollerslev et al. (2008), Jiang and Oomen (2008), Lee and Mykland (2012) and the references therein. On the other hand, it is well known that state-domain asymptotic theory is very different from that of the time domain (see, for instance Fan and Yao (2003), Chapter 6). In our specific case, uniform asymptotic behavior of our test statistic on $[l, u]$ is arguably more difficult to establish than the corresponding problem in the time domain. In the current paper, we establish that, unlike time-domain change point detection problems of (3) where the long-run variances of the process are of crucial importance in the asymptotics, state-domain change point detection theory of (1) heavily depends on the conditional variances and densities of the process $\{X_i\}$. We also provide

an estimation procedure using simulated critical values to detect and locate the change points. We show that, when the jump sizes have a fixed and positive lower bound, the method will asymptotically detect all the change points with a preassigned probability and an accuracy c_n which is much smaller than $1/\sqrt{n}$, where n is the length of the time series.

The rest of the paper is organized as follows. In Section 2, we introduce the model framework and some basic assumptions. Section 3 contains our main results, including a nonparametric test to determine the existence of change points and a procedure for estimating the number of change points together with their locations. Practical implementation based on a bootstrap procedure and a suitable method for bandwidth selection are discussed in Section 4. Section 5 reports some simulation studies. A real data application of daily COVID-19 infections in Germany is carried out in Section 6. Section 7 contains all the proofs of the theoretical results in Section 3 and additional proofs are deferred to Section 8.

2. Model formulation and basic assumptions

Throughout this paper, we use the following notations. A random vector $X \in \mathcal{L}^p$ if $\|X\|_p := (\mathbb{E}|X|^p)^{1/p} < \infty$. For two random variables U and V , $F_{U|V}(\cdot)$ denotes the conditional distribution function of U given V and $f_{U|V}(\cdot)$ denotes the conditional density. Furthermore, for function g with $\mathbb{E}|g(U)| < \infty$, we let $\mathbb{E}(g(U) | V) := \int g(x)dF_{U|V}(x)$ be the conditional expectation of $g(U)$ given V . Finally, $\mathbb{1}$ stands for the indicator function.

Assume that the process $\{\epsilon_i\}$ is stationary and causal. Following Wu (2005), we assume that $\{\epsilon_i\}$ is a Bernoulli shift process such that

$$\epsilon_i = G^*(\xi_i), \tag{5}$$

where the function G^* is a measurable function such that the process $\{\epsilon_i\}$ exists and $\xi_i = (\dots, \eta_{i-1}, \eta_i)$ is a shift process, where $\{\eta_i\}$ are independent and identically distributed (i.i.d.) random variables. Furthermore, $\{\epsilon_i\}$ is a martingale difference sequence satisfying $\mathbb{E}[\epsilon_i | (\epsilon_{i-1}, \epsilon_{i-2}, \dots)] = 0$ almost surely. From Eq. (5), one can interpret the transform G^* as the underlying physical mechanism, with ξ_i and $G^*(\xi_i)$ being the input and output of the system, respectively.

Similarly, we assume

$$X_i = G(\xi_i),$$

where G is a measurable function such that X_i exists. To facilitate the main results, we first introduce the time series dependence measures in Wu (2005) associated with X_i and ϵ_i . Assume $X \in \mathcal{L}^p$, and let

$$X'_n = G(\xi'_n), \quad \xi'_n := (\xi_{-1}, \eta'_0, \eta_1, \dots, \eta_n),$$

where X'_n is a coupled process of X_n with η_0 replaced by an i.i.d. copy η'_0 . Then, we define the physical dependence measures of X_i as

$$\theta_{n,p} = \|X_n - X'_n\|_p.$$

Let $\theta_{n,p} = 0$ if $n < 0$. Thus for $n \geq 0$, $\theta_{n,p}$ measures the dependence of the output $G(\xi_n)$ on the single input η_0 . We refer to Wu (2005) for more details on the physical dependence measures.

Similarly, we define the physical dependence measures for the errors as

$$\theta_{n,p}^* = \|\epsilon_n - \epsilon'_n\|_p,$$

where $\epsilon'_n = G^*(\xi'_n)$. Let $\theta_{n,p}^* = 0$ if $n < 0$.

Suppose that $\{X_i\}_{i=1}^n$ is observed. Recall $H_0 : M = 0$ and we aim to test the null hypothesis that the regression function is smooth. To this end, we introduce a density-weighted anti-symmetric kernel function \tilde{K}_n , which is defined by

$$\tilde{K}_n(X, x, b) := \frac{w_n^*(x, b)K\left(\frac{X-x}{b}\right) - w_n(x, b)K^*\left(\frac{X-x}{b}\right)}{w_n(x, b)w_n^*(x, b)},$$

where $K(\cdot)$ is a kernel function supported on $S = [0, 1]$ with $\int_S K(u)du = 1$ and $K^*(u) := K(-u)$. The data-dependent weights $w_n(x, b)$ and $w_n^*(x, b)$ are defined by

$$w_n(x, b) := \frac{1}{nb} \sum_{i=1}^n K\left(\frac{X_i - x}{b}\right), \quad w_n^*(x, b) := \frac{1}{nb} \sum_{i=1}^n K^*\left(\frac{X_i - x}{b}\right),$$

where $b = b_n$ is the bandwidth satisfying $b \rightarrow 0$ and $nb \rightarrow \infty$. Note that $w_n(x, b)$ and $w_n^*(x, b)$ are one-sided kernel density estimators. Hence $\tilde{K}_n(X, x, b)$ can be approximated by $[K(\frac{X-x}{b}) - K^*(\frac{X-x}{b})]/f(x)$, where $f(x)$ is the density function of X_i . Observing that $K(x) - K^*(x)$ is an anti-symmetric function, we then call $\tilde{K}_n(X, x, b)$ a density-weighted anti-symmetric kernel function. By sliding this kernel function \tilde{K}_n through the state domain, we are able to test whether $\mu(x)$ has change points. More specifically, the quantity $\sum_{k=2}^n \tilde{K}_n(X_{k-1}, x, b)X_k/nb$ is a boundary kernel estimation of $\mu(x^+) - \mu(x^-)$, where $\mu(x^+)$ and $\mu(x^-)$ are the right and left limits of $\mu(\cdot)$ at x . Thus, if x is a continuous point of $\mu(\cdot)$, this quantity will be approximately zero at x . However, if $\mu(\cdot)$ is discontinuous at x , the quantity will be approximately equal to the jump size of $\mu(\cdot)$ at x . To establish our first main result, we need the following regularity conditions:

- (a) There exist $0 < \delta_2 \leq \delta_1 < 1$ such that $n^{-\delta_1} = \mathcal{O}(b)$ and $b = \mathcal{O}(n^{-\delta_2})$.
- (b) Assume $\mathbb{E}|\epsilon_i|^p < \infty$ where $p > 2/(1 - \delta_1)$.
- (c) For the same p defined in [Condition \(b\)](#), assume that $X_i \in \mathcal{L}^p$, $\theta_{n,p} = \mathcal{O}(\rho^n)$, and $\theta_{n,p}^* = \mathcal{O}(\rho^n)$ for some $0 < \rho < 1$.
- (d) The density function f of X_i is positive on $[l - \epsilon, u + \epsilon]$ for some $\epsilon > 0$ and there exists a constant $B < \infty$ such that

$$\sup_x \left[|f_{X_n|\xi_{n-1}}(x)| + |f'_{X_n|\xi_{n-1}}(x)| + |f''_{X_n|\xi_{n-1}}(x)| \right] \leq B, \text{ a.s.}$$

- (e) $K(\cdot)$ is differentiable over $(0, 1)$, the right derivative $K'(0+)$ and the left derivative $K'(1-)$ exists and $\sup_{0 \leq u \leq 1} |K'(u)| < \infty$. The Lebesgue measure of the set $\{u \in [0, 1] : K(u) = 0\}$ is zero. Furthermore, $K(0) = K(1) = 0$, $K'(0) > 0$ and $\int_0^1 uK(u)du = 0$.

For the above regularity conditions, [Condition \(a\)](#) specifies the allowable range of the bandwidth. [Condition \(b\)](#) puts a mild moment restriction on ϵ_i . [Condition \(c\)](#) requires that the quantities $\theta_{n,p}$ and $\theta_{n,p}^*$ satisfy the geometric moment contraction (GMC) property. The GMC property is preserved in many linear and nonlinear time series models such as the ARMA models and the ARCH and GARCH models; see [Shao and Wu \(2007\)](#) for more discussions. Furthermore, denote $\Theta_n := \sum_{i=0}^n \theta_{i,2}$, which measures the cumulative dependence of X_0, \dots, X_n on η_0 . Then if [Condition \(c\)](#) holds, it is easy to see that $\Theta_\infty < \infty$ which indicates short-range dependence of $\{X_i\}$. With [Condition \(d\)](#), we require that the density and conditional density of X_i exist and are bounded. Moreover, f has bounded derivatives up to the second order. [Condition \(e\)](#) puts some restrictions on the smoothness and order of the kernel function K . In particular, $\int_0^1 uK(u)du = 0$ indicates that K is a second-order kernel which has both positive and negative parts on $[0, 1]$.

3. State-domain change point detection and estimation

In this section, we propose a test on the existence of change points in $\mu(\cdot)$ and an algorithm to estimate the number and locations of the change points when $\mu(\cdot)$ is discontinuous.

3.1. Test for the existence of change points

With the foregoing discussion, we introduce a nonparametric statistic based on the density-weighted anti-symmetric kernel to test whether model [Eq. \(1\)](#) has change points in the state domain regression function $\mu(\cdot)$ on $[l, u]$. By proper scaling, our test statistic is defined as

$$t_n(x) := \frac{\sqrt{f(x)}}{\sigma(x)} \frac{1}{nb} \sum_{k=1}^n \tilde{K}_n(X_{k-1}, x, b) X_k,$$

where $\sigma^2(x) = \mathbb{E}[\epsilon_i^2 | X_{i-1} = x]$. In practice, since $f(\cdot)$ and $\sigma(\cdot)$ are unknown, we use the kernel density estimator $f_n(x)$ and Nadaraya–Watson (NW) estimator $\sigma_n^2(x)$ to replace $f(x)$ and $\sigma^2(x)$, respectively. The kernel density estimator is given by

$$f_n(x) = \frac{1}{nh} \sum_{k=2}^n W\left(\frac{X_{k-1} - x}{h}\right),$$

where $W(\cdot)$ is a general kernel function with $W(\cdot) \geq 0$ and $\int W(u)du = 1$, $h = h_n$ is the bandwidth sequence satisfying $h \rightarrow 0$ and $nh \rightarrow \infty$. Let $\hat{\epsilon}_k^2 = [X_k - \mu_n(X_{k-1})]^2$ be the square of the estimated residuals, where

$$\mu_n(x) = \frac{1}{nhf_n(x)} \sum_{k=2}^n W\left(\frac{X_{k-1} - x}{h}\right) X_k$$

is the NW estimator of $\mu(\cdot)$, then the NW estimator of $\sigma^2(x)$ is given by

$$\sigma_n^2(x) = \frac{1}{nhf_n(x)} \sum_{k=2}^n W\left(\frac{X_{k-1} - x}{h}\right) \hat{\epsilon}_k^2.$$

The following remark provides the uniform consistency of the estimated density and conditional variance functions.

Remark 3.1. Under [Condition \(a\)](#) for both bandwidths h and b with $0 < \delta_1 < 1/4$, [Condition \(c\)](#), [Condition \(d\)](#), and [Condition \(e\)](#), we have

$$\mathbb{E}f_n(x) - f(x) = f''(x)h^2\psi_W + o(h^2),$$

where $\psi_W := \int u^2W(u)du/2$ and

$$\sup_x |f_n(x) - f(x)| = \mathcal{O}_{\mathbb{P}}\left(\frac{(\log n)^3}{\sqrt{nh}} + h^2 \log n\right).$$

Similarly, for $\sigma_n^2(x)$, under the conditions of [Theorem 3.2](#), we also have

$$\sup_x |\sigma_n^2(x) - \sigma^2(x)| = \mathcal{O}_{\mathbb{P}} \left(\frac{(\log n)^3}{\sqrt{nh}} + h^2 \log n \right).$$

See [Section 8.1](#) for the proof. \triangleleft

Let $f_{\epsilon}(\cdot)$ be the density function of ϵ_i and $\lambda_K = \int K^2(x)dx$. We have the following main result on the asymptotic properties of the proposed test statistic.

Theorem 3.2. *Let $l, u \in \mathbb{R}$ be fixed. Recall the piece-wise formulation of [Eq. \(4\)](#), let T_j^{ϵ} and T^{ϵ} be the ϵ -neighborhoods of the intervals $T_j = [a_j, a_{j+1})$ and $T = [l, u]$, respectively. Let $T_a = \{a_j\}$ be the collection of the change points, T_a^{ϵ} be the ϵ -neighborhood of T_a . Assume that [Condition \(a\)–Condition \(e\)](#) hold with $f_{\epsilon}(\cdot)$, $\sigma(\cdot) \in \mathcal{C}^3(T^{\epsilon})$, $\mu_j(\cdot) \in \mathcal{C}^3(T_j^{\epsilon})$ for some $\epsilon > 0$ and b satisfies*

$$0 < \delta_1 < 1/3, \quad 0 < \delta_2 \leq 1/4, \quad nb^9 \log n = o(1),$$

then

$$\mathbb{P} \left(\sqrt{\frac{nb}{2\lambda_K}} \sup_{x \in T \cap (T_a^b)^c} |t_n(x)| - d_n \leq \frac{z}{(2 \log \bar{b}^{-1})^{1/2}} \right) \rightarrow e^{-2e^{-z}},$$

where $\bar{b} := b/(u - l)$ and

$$d_n := (2 \log \bar{b}^{-1})^{1/2} + \frac{1}{(2 \log \bar{b}^{-1})^{1/2}} \log \frac{\sqrt{K_2}}{\sqrt{2\pi}}$$

with $K_2 := \int_0^1 (K'(u))^2 du / \lambda_K$.

Proof. See [Section 7.1](#). \square

[Theorem 3.2](#) is a general result which establishes the asymptotic theory of the test statistic. In practical implementation, we will use the density estimates $f_n(x)$ and variance estimates $\sigma_n(x)$ instead of $f(x)$ and $\sigma(x)$ to calculate $t_n(x)$ as discussed before. Therefore, we have the following corollary.

Corollary 3.3. *Denote $t_n^*(x) = \frac{\sqrt{f_n(x)}}{\sigma_n(x)} \frac{1}{nb} \sum_{k=1}^n \tilde{K}_n(X_{k-1}, x, b) X_k$. Under the conditions of [Theorem 3.2](#) and further assume the bandwidth $h \leq b$, then the asymptotic result of [Theorem 3.2](#) holds for $t_n^*(x)$; this is*

$$\mathbb{P} \left(\sqrt{\frac{nb}{2\lambda_K}} \sup_{x \in T \cap (T_a^b)^c} |t_n^*(x)| - d_n \leq \frac{z}{(2 \log \bar{b}^{-1})^{1/2}} \right) \rightarrow e^{-2e^{-z}}.$$

Note that in [Corollary 3.3](#), we have added the assumption $h \leq b$ with the purpose of ensuring the consistency of $f_n(x)$ and $\sigma_n(x)$ on $T \cap (T_a^b)^c$. When there is no change point in $\mu(\cdot)$, we have similar results as shown in the following remark, which suggests that under the null hypothesis, after proper scaling and centering, our test statistic converges to a Gumbel distribution asymptotically.

Remark 3.4. *Assume $H_0 : M = 0$ holds. We further assume that $f(\cdot)$, $\sigma(\cdot) \in \mathcal{C}^3(T^{\epsilon})$ and the remaining conditions of [Corollary 3.3](#) hold. Then, $T_a = \emptyset$, $T_a^b = \emptyset$, which implies $T \cap (T_a^b)^c = T$. Therefore, the previous theorem reduces to*

$$\mathbb{P} \left(\sqrt{\frac{nb}{2\lambda_K}} \sup_{x \in T} |t_n^*(x)| - d_n \leq \frac{z}{(2 \log \bar{b}^{-1})^{1/2}} \right) \rightarrow e^{-2e^{-z}}. \quad \triangleleft$$

Denote the jump-size of $\mu(\cdot)$ at a_i as Δ_i ; that is, $\Delta_i := |\mu(a_i+) - \mu(a_i-)|$. Next, we consider the alternative hypothesis $H_a : M \geq 1$ with $\Delta_i \geq \tilde{\Delta} > 0$. When H_a holds true, it is easy to see that the proposed test has an asymptotic power 1 as $n \rightarrow \infty$. In other words, with some preassigned level $\alpha \in (0, 1)$ and as $n \rightarrow \infty$, we have

$$\mathbb{P} \left(\sup_{x \in T} |t_n(x)| \geq \sqrt{\frac{2\lambda_K}{nb}} \left[d_n - \frac{\log\{\log(1 - \alpha)^{-1/2}\}}{(2 \log \bar{b}^{-1})^{1/2}} \right] \right) \rightarrow 1.$$

Once the null hypothesis of no change point is rejected, one would be interested in detecting the number of change points together with their locations, which we discuss in [Section 3.2](#).

3.2. Change-point estimation

Suppose there exist a fixed number M of change points on $[l, u]$, which are denoted by $l < a_1 < \dots < a_M < u$, with the minimum jump size $\min_{1 \leq i \leq M} \Delta_i \geq \tilde{\Delta}_n > 0$. In this paper, we assume $\tilde{\Delta}_n = \mathcal{O}(1)$ which is allowed to decrease with n . The idea for estimating the number and locations of the change points is to search for local maximas of $|t_n(x)|$ which exceed the critical value of the test. To be more specific, we propose in the following a procedure for change point estimation.

- For a fixed level α , perform the bootstrap procedure described in Section 4.1 to determine the critical value, say $C_{n,\alpha}$.
- Set $T_1 := (l, u)$.
- Starting from the interval T_1 , find the largest x of $|t_n(x)|$ that exceeds the critical value and denote its location as $\hat{a}_{(1)}$, then rule out the interval $[\hat{a}_{(1)} - b, \hat{a}_{(1)} + b]$ from T_1 to get $T_2 := T_1 \cap [\hat{a}_{(1)} - b, \hat{a}_{(1)} + b]^c$.
- Repeat the previous step until all significant local maximas are found. In other words, $|t_n(x)|$ on the remaining intervals are all below $C_{n,\alpha}$.
- Denote the number of detected change points by \hat{M} and re-order the estimated change points as $l < \hat{a}_1 < \dots < \hat{a}_{\hat{M}} < u$.

The following theorem provides an asymptotic result for \hat{M} and \hat{a}_i .

Theorem 3.5. *Under the conditions of Theorem 3.2, we further assume that $K'(\cdot)$ is differentiable over $(0, 1)$ with $K'(1) = 0$, the right derivative $K''(0+)$ and the left derivative $K''(1-)$ exist and $\sup_{0 \leq u \leq 1} |K''(u)| < \infty$. The Lebesgue measure of the set $\{u \in [0, 1] : K'(u) = 0\}$ is zero. If $\sqrt{\frac{\log n}{nb}} = o(\tilde{\Delta}_n)$ then for any given level α , we have*

$$\mathbb{P} \left(\left\{ \hat{M} = M \right\} \cap \left\{ \max_{1 \leq i \leq M} |\hat{a}_i - a_i| < c_n \right\} \right) \rightarrow 1 - \alpha,$$

for any c_n such that $1/c_n = \mathcal{O} \left(\tilde{\Delta}_n \sqrt{\frac{n}{b \log n}} \right)$

Proof. See Section 7.2. \square

Theorem 3.5 reveals that for any given small probability α , with asymptotic probability $1 - \alpha$, our proposed procedure will correctly estimate all the change points with an accuracy c_n . It is important to mention that when $\tilde{\Delta}_n = \tilde{\Delta} > 0$, that is, when the jump sizes have a fixed lower bound, the smallest order for c_n is $\sqrt{b \log n/n}$, which is smaller than $n^{-1/2}$. It can also be seen as a product of $\sqrt{\log n}$ and the optimal convergence rate $(\sqrt{b/n})$ of time-domain change-point estimators established in Müller (1992). Hence, we conjecture that our rate c_n is nearly optimal for state-domain change point detection.

4. Practical implementation

4.1. The bootstrap procedure

It is well known that the convergence rate of the Gumbel distribution in Theorem 3.2 is slow. As a result, a very large sample size would be needed for the approximation to be reasonably accurate. To overcome this issue, we propose a simulation-based bootstrap procedure to improve the finite-sample performance of the proposed test. The bootstrap procedure is as follows.

- Generate i.i.d. standard normal random variables U_k , $k = 0, \dots, n$.
- Compute the quantity Π_n^* defined in Eq. (6) for many times and calculate its $(1 - \alpha)$ th quantile as the critical value of our test.

For the proposed bootstrap procedure, we have the following theoretical results which shows that, with proper scaling and centering, Π_n^* has the same asymptotic Gumbel distribution.

Proposition 4.1. *Denote $\Pi_n = \sup_{x \in T} |t_n^*(x)|$ and*

$$\Pi_n^* = \sup_{x \in T} \left| \frac{\sqrt{g(x)}}{nb} \sum_{k=1}^n \tilde{K}_n(U_{k-1}, x, b) U_k \right|, \tag{6}$$

where $\{U_k\}_{k=0}^n$ are i.i.d. standard normal random variables and $g(x)$ is its density. Assume $H_0 : M = 0$, Condition (a), Condition (e) hold and b satisfies

$$0 < \delta_1 < 1/3, \quad 0 < \delta_2 \leq 1/4, \quad nb^9 \log n = o(1).$$

Then we have

$$\mathbb{P}\left(\sqrt{\frac{nb}{2\lambda_K}}\Pi_n^* - d_n \leq \frac{z}{(2 \log \bar{b}^{-1})^{\frac{1}{2}}}\right) \rightarrow e^{-2e^{-z}}, \text{ as } n \rightarrow \infty.$$

Proposition 4.1 shows that Π_n^* and Π_n have the same asymptotic Gumbel distribution with proper scaling and centering under the null hypothesis. Therefore, the $(1-\alpha)$ th quantile of Π_n can be estimated consistently by calculating the empirical $(1-\alpha)$ th quantile $C_{n,\alpha}$ of Π_n^* with a large number of replications by the bootstrap procedure. We reject the null hypothesis at level $\alpha \in (0, 1)$ if $\Pi_n > C_{n,\alpha}$. When implementing the procedure described in Section 3.2 for estimating the change points, we also suggest using $C_{n,\alpha}$ to find the detection region. Our numerical experiments suggest that the bootstrap method yields more accurate results than those based on the asymptotic limiting distribution under small or moderate sample sizes.

4.2. Bandwidth selection

The bandwidth used in $f_n(x)$ can be chosen based on classic bandwidth selectors for nonparametric kernel density estimation. However, the choice of bandwidth b for test statistic $t_n^*(x)$ and h for the estimated variance $\sigma_n^2(x)$ can be quite nontrivial and are usually of practical interest. In this paper, we adopt the standard leave-one-out cross-validation criterion for bandwidth selection suggested by Rice and Silverman (1991):

$$CV(b) = \frac{1}{n} \sum_{k=1}^n [X_{k+1} - \mu_n^{(-k)}(X_k)]^2,$$

$$CV(h) = \frac{1}{n} \sum_{k=1}^n [(X_{k+1} - \mu_n(X_k))^2 - \sigma_n^{2(-k)}(X_k)]^2$$

where $\mu_n^{(-k)}(X_k)$ and $\sigma_n^{2(-k)}(X_k)$ are the kernel estimators of μ and σ^2 computed with all measurements with the k th subject deleted, respectively. For example, a cross-validation bandwidth \hat{b} can be obtained by minimizing $CV(b)$ with respect to b , i.e., $\hat{b} = \arg \min_{b \in \mathcal{B}} CV(b)$, where \mathcal{B} is the allowable range of b . The bandwidth selection for h is similar.

5. Simulation study

In this section, we carry out Monte Carlo simulations to examine the finite-sample performances of our proposed test and estimator. Throughout the numerical experiments, the Epanechnikov kernel $W(x) = 0.75(1 - x^2)\mathbb{1}(|x| \leq 1)$ is used for estimating the density and conditional variances. On the other hand, we adopt the higher-order kernel function in the form $K(x) = b[\tilde{W}(x) - a\tilde{W}(\sqrt{ax})]$ in the expression of \hat{K}_n , where $\tilde{W}(x)$ is the kernel function on $[0,1]$ by shifting and scaling $W(x)$. From Theorem 3.2, one can see that the power of our test increases as λ_K decreases. As a result, we aim to maximize the quantity $Q(a, b) = \frac{\int_0^\infty K(x)dx}{\sqrt{\int_0^\infty K^2(x)dx}}$ with the constraints $\int_0^\infty K(x)dx = 1$ and $\int_0^\infty xK(x)dx = 0$ to choose a and b . It turns out that $Q(a, b)$ is maximized at $a = 0.34$ and $b = \frac{2}{\sqrt{0.34-0.34}}$. Hence, we will use $K(x) = \frac{2}{\sqrt{0.34-0.34}}[\tilde{W}(x) - 0.34\tilde{W}(\sqrt{0.34x})]$ in our simulations and data analysis.

5.1. Accuracy of bootstrap

We run Monte Carlo simulations to study the accuracy of the proposed bootstrap procedure for finite samples $n = 200, 500$ and 800 . Here, we aim to test the null hypothesis H_0 of no change point in the regression function. The number of replications is fixed to be 1000 and the number of bootstrap samples is $B = 2000$ at each replication.

To guarantee the stationarity of the process $\{X_i\}$, $|\mu(x)|$ is required to be less than one (Fan and Yao, 2003, Section 2.1). First, we consider Model A listed below to investigate the robustness of our testing procedure with respect to various levels of persistence in the data generating process. Additional four state-domain nonlinear models (Models B–E listed below) where $\mu(\cdot)$ is of various shapes are further investigated for the accuracy of our test. In our simulations the martingale difference process $\epsilon_i = \sigma(X_{i-1})\epsilon_i^*$ with $\sigma^2(x) = \mathbb{E}(\epsilon_i^2|X_{i-1} = x)$ and $\epsilon_i^* \stackrel{i.i.d.}{\sim} \mathcal{N}(0, 1)$. Note that the error processes $\{\epsilon_i\}$ are specified via different conditional variance functions $\sigma^2(x)$ in Models A–D. On the other hand, in Model E we set $\epsilon_i = 0.5\eta_i(\eta_{i-7} + 1.5)$ where $\eta_i \stackrel{i.i.d.}{\sim} \mathcal{N}(0, 1)$ so that $\{\epsilon_i\}$ has a period of 7 which matches the data generating process observed in the empirical data example in Section 6.

- Model A:

$$\mu(x) = \begin{cases} \kappa_1 x^3, & |x| \leq 1, \\ \kappa_1, & x > 1, \\ -\kappa_1, & x < -1, \end{cases}$$

Table 1
Simulated Type I error rates for Model A with the first order ADCF of the model.

Model A	κ_1	0.2	0.4	0.6	0.8
	$\mathcal{R}(1)$	0.240	0.321	0.412	0.523
$\alpha = 0.05$	$n = 200$	0.046	0.055	0.064	0.067
	$n = 500$	0.048	0.052	0.060	0.065
	$n = 800$	0.053	0.049	0.050	0.065
$\alpha = 0.1$	$n = 200$	0.103	0.109	0.132	0.131
	$n = 500$	0.096	0.092	0.119	0.138
	$n = 800$	0.099	0.092	0.109	0.126

Table 2
Simulated Type I error rates for Model B–E with the first and seventh order ADCF of Model E.

	Model	B	C	D	E	
$\alpha = 0.05$	$n = 200$	0.040	0.044	0.050	0.060	$\mathcal{R}(1)$
	$n = 500$	0.066	0.041	0.054	0.054	0.195
	$n = 800$	0.056	0.051	0.057	0.054	
$\alpha = 0.1$	$n = 200$	0.085	0.106	0.124	0.105	$\mathcal{R}(7)$
	$n = 500$	0.102	0.092	0.114	0.092	0.258
	$n = 800$	0.093	0.101	0.112	0.095	

$$\sigma(x) = 1.5e^{-0.5x^2},$$

where $\kappa_1 = 0.2, 0.4, 0.6, 0.8$ represents various levels of temporal dependence in the series.

- Model B:

$$\mu(x) = 0.2e^{-0.5x^2}, \quad \sigma(x) = \frac{1.5e^x}{1 + e^x}.$$

- Model C:

$$\mu(x) = \frac{0.3e^x}{1 + e^x},$$

$$\sigma(x) = \begin{cases} 0.7(1 + x^2), & |x| \leq 1, \\ 1.4, & \text{otherwise.} \end{cases}$$

- Model D:

$$\mu(x) = 0.8 \sin(x), \quad \sigma(x) = 1.$$

- Model E:

$$\mu(x) = 0.5 \cos(x).$$

Note that the regression functions $\mu(\cdot)$ in Models A–E are all continuous. At nominal significance levels $\alpha = 0.05$ and 0.1 , the simulated Type I error rates for sample sizes $n = 200, 500$ and 800 are reported in Tables 1–2 for Model A and Models B–E, respectively. To measure the strength of the nonlinear temporal dependence, we will employ the auto-distance correlation function (ADCF) investigated in Zhou (2012). In Table 1, we illustrate the first order ADCF (denoted by $\mathcal{R}(1)$) for Model A. Meanwhile, for Model E the first order and the seventh order ADCF are listed in Table 2. One can see that the performance of our testing procedure is reasonably accurate for different sample sizes across the models and the accuracy improves as the sample size increases. On the other hand, from Table 1, we find that as the dependence of the process becomes stronger, the Type I errors tend to be less accurate, but are still in a reasonable range.

5.2. Power of hypothesis testing

In this subsection, we consider the simulated power of our test under various alternatives. Recall the representation $\epsilon_i = \sigma(X_{i-1})\epsilon_i^*$ with $\epsilon_i^* \stackrel{i.i.d.}{\sim} \mathcal{N}(0, 1)$. Here, we consider the following two types of alternatives with a change point of size δ :

- Model F_1 :

$$\mu(x) = \begin{cases} 0.5e^{-x^2}, & x < 0, \\ 0.5e^{-x^2} - \delta, & x \geq 0, \end{cases}$$

$$\sigma(x) = e^{-0.5x^2}.$$

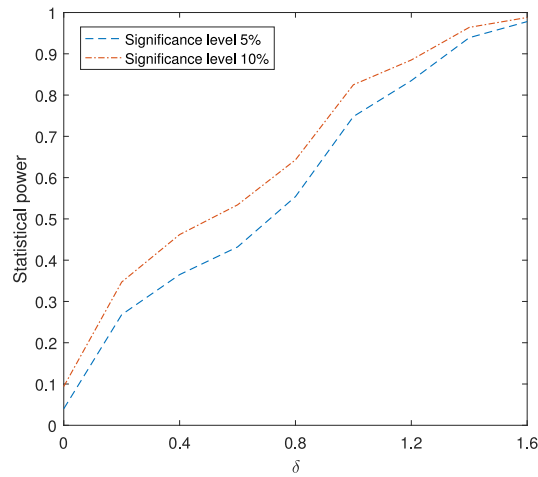


Fig. 1. Simulated rejection rates for testing change point for Model F_1 .

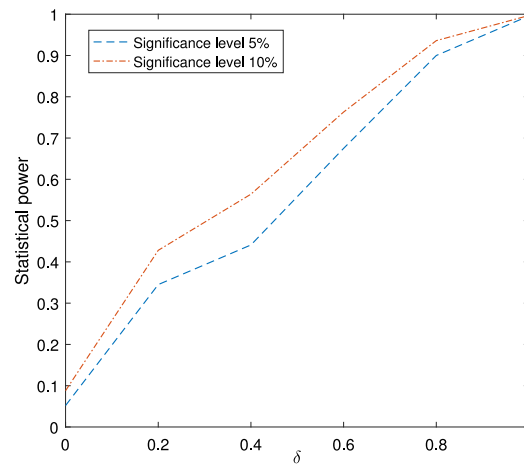


Fig. 2. Simulated rejection rates for testing change point for Model F_2 .

• Model F_2 :

$$\mu(x) = \begin{cases} 0.3 - \delta, & x < 0, \\ 0.3, & x \geq 0, \end{cases}$$

$$\sigma(x) = \frac{e^x}{1 + e^x}.$$

We let the jump size δ range from 0 to 1.6 for Model F_1 and from 0 to 1 for Model F_2 at location $x = 0$. For each model, we investigate the empirical sensitivity of our testing procedure under nominal levels 0.05 and 0.1 with sample size $n = 800$ based on 1000 replications. The simulated power curves for the above models are plotted in Figs. 1 and 2, respectively. According to the plots, the statistical power of the proposed testing procedure increases reasonably fast as δ increases. On the other hand, we also observe that our test shows a slower speed of increase at near alternatives when compared with “classic” power curves of parametric tests. We believe that part of the reason is that our nonparametric test aims at detecting alternatives from a large class of discontinuous functions while tests tailored to some parametric models (such as the threshold model) target a specific class of alternative functions. Therefore our test is expected to be less sensitive to small deviations from the null compared to those parametric tests. See also Section 5.4 for a numerical experiment that compares the sensitivity of our testing procedure with that of a parametric test of the threshold model.

Table 3

Accuracy in estimating the change-point locations and the percentage of correctly estimating the number of change points.

Case 1	n	MADE	MSE	Percentage
$a_1 = 0$	200	0.0451	0.0055	90.51%
	500	0.0195	0.0014	93.77%
	800	0.0134	0.0006	94.51%
Case 2	n	MADE	MSE	Percentage
$a_1 = -0.3$ $a_2 = 0$	200	0.0519	0.0059	81.82%
		0.0757	0.0069	
$a_1 = -0.3$ $a_2 = 0$	500	0.0508	0.0043	86.59%
		0.0496	0.0042	
$a_1 = -0.3$ $a_2 = 0$	800	0.0386	0.0028	89.80%
		0.0362	0.0024	

Note: true change point $a_1 = 0$ for Case 1; true change points $a_1 = -0.3$ and $a_2 = 0$ for Case 2; MADE: mean absolute deviation error; MSE: mean squared error.

5.3. Accuracy in estimating the number of change points and their locations

Utilizing the algorithm listed in Section 3.2, in this subsection we focus on estimating the number of change points and their locations based on 1000 realizations with sample sizes $n = 200, 500$ and 800 . In the simulations, we let the error process $\{\epsilon_i^*\}_{i=1}^n$ be i.i.d. standard normal random variables and consider the following two cases:

- Case 1: A single change point.

$$\mu(x) = \begin{cases} 0.7e^{-x^2}, & x < 0, \\ 0.7e^{-x^2} - 1.6, & x \geq 0, \end{cases}$$

$$\sigma(x) = e^{-0.5x^2}.$$

- Case 2: Two change points.

$$\mu(x) = \begin{cases} 0.8x + 0.8, & x < -0.3, \\ -1, & -0.3 \leq x < 0, \\ -0.2x + 0.5, & x \geq 0, \end{cases}$$

$$\sigma(x) = \frac{e^x}{1 + e^x}.$$

The estimates of the locations of change points are compared in terms of their mean absolute deviation errors (MADE) and mean squared errors (MSE). We also report the simulated percentage of correctly estimating the number of change points. The results are listed in Table 3. One can see from Table 3 that the values of MADE and MSE are all quite small, which suggests the estimated locations by our approach are fairly accurate. Furthermore, as the sample size increases, the percentage of correctly estimating the number of change points increases in both cases.

5.4. Comparison to threshold testing and estimation in threshold model

In this subsection, we compare the accuracy and sensitivity of our nonparametric method with existing threshold testing and estimation methods for the classic threshold AR (TAR) model proposed by Tong and Lim (1980) when the TAR model is indeed the underlying data generating mechanism. We consider the following two-regime TAR(1) model

$$X_i = \begin{cases} 0.5(X_{i-1} + 1) + \epsilon_i, & X_{i-1} < 0.25, \\ \kappa_2(X_{i-1} + 1) + \epsilon_i, & X_{i-1} \geq 0.25, \end{cases}$$

where $\kappa_2 = 0.5, 0.3, 0.1, -0.1, -0.3, -0.5$ and the error process $\epsilon_i \stackrel{i.i.d.}{\sim} \mathcal{N}(0, 0.75^2)$. First, we are interested in comparing the accuracy and power of our nonparametric test with the parametric F -test of threshold nonlinearity proposed in Tsay (1989). Table 4 shows the testing results for nonlinearity of the model based on both the parametric and nonparametric methods, in which the sample size is $n = 800$ and the number of bootstrap samples is $B = 2000$.

We observe that the nonparametric method has slightly higher powers when the scale coefficient κ_2 changes slightly from 0.5. However, as κ_2 becomes 0.1 or smaller, the parametric method has higher powers than the nonparametric method.

Table 4
Simulated rejection rates for testing change point with TAR(1) model.

κ_2		0.5	0.3	0.1	-0.1	-0.3	-0.5
Para.	$\alpha = 0.05$	0.042	0.175	0.831	0.904	1	1
	$\alpha = 0.1$	0.095	0.282	0.897	0.906	1	1
Nonpara.	$\alpha = 0.05$	0.069	0.256	0.406	0.646	0.792	0.910
	$\alpha = 0.1$	0.131	0.378	0.540	0.761	0.861	0.940

Table 5
Estimation accuracy for change-point locations.

	n	MADE	MSE
Nonpara.	200	0.1055	0.0143
	500	0.0519	0.0066
	800	0.0367	0.0041
Para.	200	0.0340	0.0027
	500	0.0178	0.0012
	800	0.0098	0.0004

In addition, we compare the accuracy in change point estimation. We study the following TAR(1) model,

$$X_i = \begin{cases} \frac{2}{3}(X_{i-1} + 1) + \epsilon_i, & X_{i-1} < 0.25, \\ -\frac{2}{3}(X_{i-1} + 1) + \epsilon_i, & X_{i-1} \geq 0.25, \end{cases}$$

where $\epsilon_i \stackrel{i.i.d.}{\sim} \mathcal{N}(0, 0.75^2)$. Note that parametric estimation of the threshold value of the above two-regime TAR(1) process can be done via the R function uTAR in the NTS package (we refer to Liu et al. (2020) for more details). The simulated MADEs and MSEs are listed in Table 5. From Table 5, one can see that both methods provide relatively accurate estimates of the locations of change point (threshold). The parametric method shows more accurate estimation results comparing with those of the nonparametric method. With the above observations, it can be seen that the parametric method is better for testing and detecting change points for the TAR model when the model is well-specified. This result is not surprising since testing sensitivity and estimation accuracy tend to be higher when the model is correctly restricted to a (smaller) parametric class.

6. Illustrative example

In this section, we consider the daily new confirmed cases of Coronavirus disease of 2019 (COVID-19) in Germany. The dataset contains 156 observations from April 28th to September 30th of 2020 which can be downloaded from <https://ourworldindata.org/coronavirus-source-data>. From the COVID-19 timeline, Germany registered the first case on January 28th and later suffered an outbreak of this pandemic from mid March to late April. In the data analysis, we select the aforementioned time span between the first and second waves of COVID-19 so that the time series is approximately stationary. Let X_i be the logarithm of confirmed cases at day $i = 1, \dots, 156$ and $Y_i = X_{i+1} - X_i$. The sample path of $\{X_i\}$ and the ADCF plot of $\{X_i\}$ are shown in Fig. 3. Both plots in Fig. 3 suggest that the time series is approximately stationary and has a moderate seasonal dependence with period $S = 7$. The seasonal behavior probably comes from the reporting lag behind during weekends, which happens in almost every country. We consider the state-domain nonlinear regression model (which is equivalent to Eq. (1)):

$$Y_i = \mu(X_i) + \epsilon_i,$$

where $\{\epsilon_i\}$ is a martingale difference sequence. In this application, $\mu(x)$ represents the expected increase or decrease in percentage of COVID-19 cases in day i when $X_{i-1} = x$.

We apply the proposed method to testing whether $\mu(\cdot)$ contains any change points. We choose $T = [l, u] = [5.7, 7.5]$ which includes 82.69% of X_i so that data are relatively abundant in this region and the test is expected to be accurate. According to the leave-one-out cross-validation criterion, the selected bandwidths b and h are 0.446 and 0.40, respectively. Through the practical implementation in Section 4.1, we calculate the empirical 99% quantile of Π_n^* with 10000 bootstrap samples, which gives $C_{n,\alpha} = 1.596$. Next, we investigate the behavior of the test statistics, which is shown in Fig. 4. Our test rejects the null hypothesis of continuity of $\mu(\cdot)$ at 1% level and flags two change points at $\hat{x}_1 = 6.83$ and $\hat{x}_2 = 7.40$.

Note that Y_i can be viewed as the conditional daily growth rate for COVID-19. For comparison, we also use the nonparametric local polynomial method to fit $\mu(x)$ assuming that there is no change point. The corresponding estimated regression function $\mu_n(x)$ over $[5.7, 7.5]$ is plotted on the left hand side of Fig. 5. On the right hand side of Fig. 5 we plot

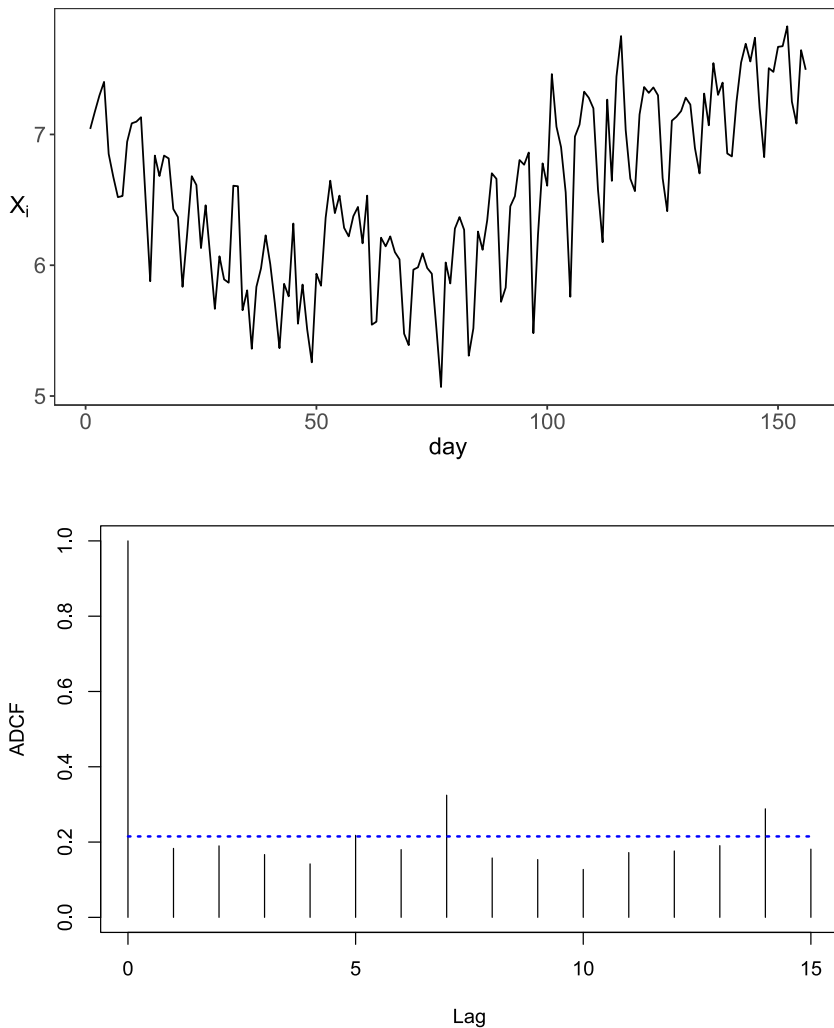


Fig. 3. Top: Logarithm of daily confirmed cases of COVID-19 in Germany from April 28th to September 30th, 2020 (which is denoted as $\{X_i\}_{i=1}^{156}$ in Section 6). Bottom: ADCF plot of $\{X_i\}$.

the fitted drift function $\mu_n(x)$ with the knowledge of the change points. The difference between the two plots in Fig. 5 suggest that, with or without the knowledge of change points, our understanding of the relationship between Y_i and X_i can be quite different. With the knowledge of the change points, we can see that two large jumps exist at $x_1 = 6.83$ and $x_2 = 7.40$, which shows that the growth rate changes abruptly at these two points.

It is obvious to see from the right plot of Fig. 5 that those two change points divide the state domain into three regimes/phases. Furthermore, the latter plot indicates that the nonlinear dynamics can be approximated by a three-regime threshold model with the data generating mechanism switching at the detected change points. Additionally, according to the timeline, we can find out the periods corresponding to each phase. The first phase $x \in [5.7, 6.83]$ contains May 3–5, 10–11, 13, May 15–July 31, August 2–5, 9–10, 16–17, 19, 23–24, 30–31 and September 7 where the trajectory depicts a relatively inactive period of the virus transmission and the conditional infection rate $\mu(x)$ decreases from positive to negative as x increases. The second phase $x \in [6.83, 7.4]$ includes April 28–30, May 2, 6–9, 12, 14, August 1, 7–8, 11–15, 18, 22, 25–29, September 1–6, 8–9, 11, 13–15, 20–21 and 27–28 where the conditional infection rate jumps up when x surpasses 6.83 and then it decreases gradually again. The third phase $x \in [7.4, 7.83]$ corresponds to May 1, August 6, 20–21, September 10, 12, 16–19, 22–26 and 29–30 where a sudden large increase in the conditional infection rate can be found at the left boundary and then it decreases sharply, possibly due to strong governmental interventions.

In summary, the analyzed period from April 28th to September 30th of 2020 of German COVID-19 data shows a complicated nonlinear dynamic balance between disease transmission and government intervention. The proposed method of the paper could help understand this complex nonlinear dynamics by determining the boundaries of phases where the state-domain relationship changes abruptly and subsequently segment the time series into multiple regimes.

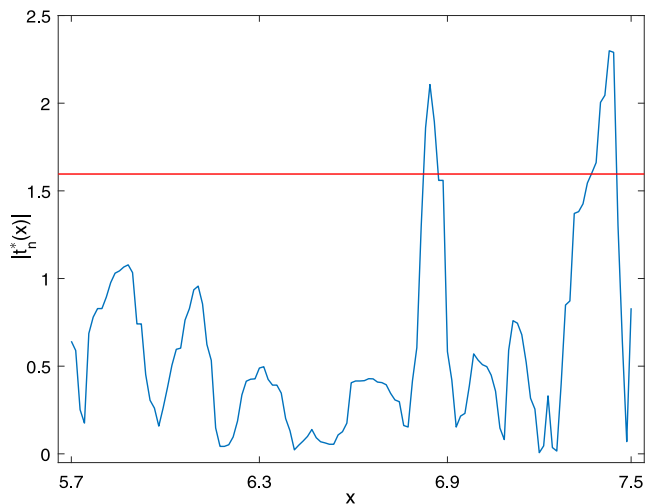


Fig. 4. The absolute value of test statistics $|t_n^*(x)|$ over $[5.7, 7.5]$, red line denotes the 99% sample quantile ($=1.596$) of \mathbb{I}_n^* . (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

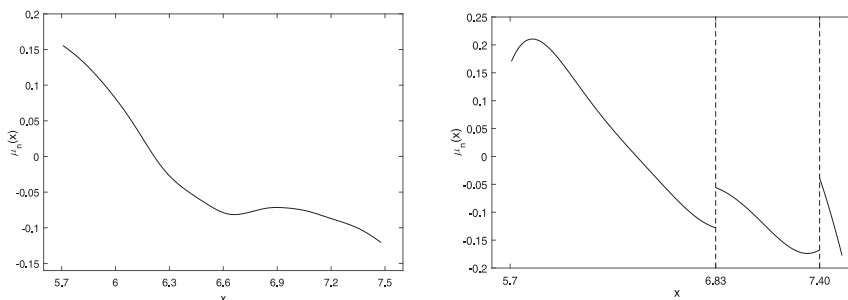


Fig. 5. Left: Smooth fit with no change points; Right: Piece-wise smooth fit with the knowledge of two change points.

7. Proofs of main results

7.1. Proof of Theorem 3.2

The outline of the proof is as follows. Firstly, we use the following decomposition of X_i

$$X_i = \mu(X_{i-1}) + \epsilon_i = [\mu(X_{i-1}) - \mu(x)] + \mu(x) + \epsilon_i,$$

and prove the results involving the first two terms. This is given in Section 7.1.1.

Secondly, we use a technique called m -dependent approximation to approximate the martingale $\{\epsilon_i\}$ using $\{\mathbb{E}[\epsilon_k | \xi_{i,i-m}] - \mathbb{E}[\epsilon_k | \xi_{i-1,i-m}]\}$, where $\xi_{k_1,k_2} := (\eta_{k_1}, \dots, \eta_{k_2})$, for a properly chosen order $m \rightarrow \infty$, which simplifies the sum of a sequence of dependent random variables to a corresponding sum of m -dependent random variables. This is done in Section 7.1.2.

Thirdly, we divide the sequence of n (m -dependent) random variables into alternating big and small blocks, where the length of big blocks has a slightly higher order than that of the small blocks. Furthermore, the length of the small blocks is larger than m . Using this proof technique, we can approximate the sum of n (m -dependent) random variables using the sum of the subsequence which includes the random variables residing in the big blocks. Since the length of small blocks is larger than m , the m -dependent random variables in different big blocks are *independent*. This part of the proof is given in Section 7.1.3.

Fourthly, we only need to deal with a sequence of independent sums of random variables within each big block. In order to get prepared for using the multivariate Gaussian approximation result by Zaitsev (1987), we first compute the asymptotic covariance structure of the sequence of independent sums. This is given in Section 7.1.4.

In the final two steps, we first apply the multivariate Gaussian approximation by Zaitsev (1987), which is given in Section 7.1.5 and then prove the convergence to Gumbel distribution, which is given in Section 7.1.6. The techniques used in these two steps heavily depend on some existing work, particularly, the work by Zhao and Wu (2008) and Liu and Wu (2010), which eventually applied the work by Bickel and Rosenblatt (1973) and Rosenblatt (1976).

7.1.1. Decomposition

First, we substitute $X_i = \mu(X_{i-1}) + \epsilon_i$ into $t_n(x)$ and separate the terms involving K and K^* . We first focus on the term involving K only. That is,

$$\begin{aligned} & \frac{1}{nbw(x, b)} \sum_{k=1}^n K\left(\frac{X_{k-1} - x}{b}\right) [\mu(X_{k-1}) + \epsilon_k], \\ &= \frac{1}{nbw(x, b)} \sum_{k=1}^n K\left(\frac{X_{k-1} - x}{b}\right) [\mu(X_{k-1}) - \mu(x)] \\ & \quad + \frac{1}{nbw(x, b)} \sum_{k=1}^n K\left(\frac{X_{k-1} - x}{b}\right) \mu(x) \\ & \quad + \frac{1}{nbw(x, b)} \sum_{k=1}^n K\left(\frac{X_{k-1} - x}{b}\right) \epsilon_k. \end{aligned} \tag{7}$$

Next it is easy to see that by the definition of $w(x, b)$, the second term of the decomposition on the right hand side of Eq. (7) equals $\mu(x)$. For the first term of the decomposition in Eq. (7), following exactly the proof of Liu and Wu (2010, Lemma 5.2), uniformly over x , we have that

$$\begin{aligned} & \frac{1}{nbw(x, b)} \sum_{k=1}^n K\left(\frac{X_{k-1} - x}{b}\right) [\mu(X_{k-1}) - \mu(x)] \\ &= \frac{b^2 \psi_K [\mu''(x)f(x) + 2\mu'(x)f'(x)]}{\mathbb{E}[w(x, b)] + \mathcal{O}_{\mathbb{P}}(\sqrt{\log n/nb})} + \mathcal{O}_{\mathbb{P}}(b^3) + \mathcal{O}_{\mathbb{P}}(\tau_n) \\ &= \frac{b^2 \psi_K [\mu''(x)f(x) + 2\mu'(x)f'(x)]}{\mathbb{E}[w(x, b)]} + b^2 \mathcal{O}_{\mathbb{P}}(\sqrt{\log n/nb}) \\ & \quad + \mathcal{O}_{\mathbb{P}}\left(\sqrt{\frac{b \log n}{n}} + b^3 + \frac{b}{n} \sqrt{\sum_{k=-n}^{\infty} (\Theta_{n+k} - \Theta_k)^2}\right) \\ &= \frac{b^2 \psi_K [\mu''(x)f(x) + 2\mu'(x)f'(x)]}{\mathbb{E}[w(x, b)]} + \mathcal{O}_{\mathbb{P}}\left(\sqrt{\frac{b \log n}{n}} + b^3\right), \end{aligned}$$

where $\tau_n := \sqrt{\frac{b \log n}{n}} + b^4 + \frac{b}{n} \sqrt{\sum_{k=-n}^{\infty} (\Theta_{n+k} - \Theta_k)^2}$ comes from Zhao and Wu (2008, Lemma 2(ii)), and in the last equality we have applied the assumptions on b and $\sum_{k=-n}^{\infty} (\Theta_{n+k} - \Theta_k)^2$ to get $\frac{b}{n} \sqrt{\sum_{k=-n}^{\infty} (\Theta_{n+k} - \Theta_k)^2} = \mathcal{O}(\sqrt{b \log n/n})$.

7.1.2. m -dependent approximation

For the third term of the decomposition in Eq. (7), recalling that we have defined the notation $\xi_{k_1, k_2} := (\eta_{k_1}, \dots, \eta_{k_2})$, we consider the decomposition of ϵ_k ,

$$\begin{aligned} \epsilon_k &= (\epsilon_k - \mathbb{E}[\epsilon_k \mid \xi_{k, k-m}]) \\ & \quad + (\mathbb{E}[\epsilon_k \mid \xi_{k, k-m}] - \mathbb{E}[\epsilon_k \mid \xi_{k-1, k-m}]) \\ & \quad + \mathbb{E}[\epsilon_k \mid \xi_{k-1, k-m}], \end{aligned}$$

where $m = \lfloor n^\tau \rfloor$ where $\tau < 1 - \delta_1$. The first and last terms in the decomposition can be ignored comparing to the second term. To see this, consider

$$\begin{aligned} \mathbb{E}[\epsilon_k \mid \xi_{k-1, k-m}] &= \mathbb{E}[\epsilon_k \mid \xi_{k-1, k-m}] - \mathbb{E}[\epsilon_k \mid \mathcal{F}_{k-1}] \\ &= \sum_{i=1}^{\infty} \mathbb{E}[\epsilon_k \mid \xi_{k-1, k-i}] - \mathbb{E}[\epsilon_k \mid \xi_{k-1, k-i-1}], \end{aligned}$$

which implies $\|\mathbb{E}[\epsilon_k \mid \xi_{k-1, k-m}]\|_p = \mathcal{O}(\sum_{i=m}^{\infty} \rho^i) = \mathcal{O}(\rho^m)$. Since $m > (\log n)^2$, we have

$$\sqrt{nb} \sup_{x \in T} \left| \frac{1}{nb} \sum_{k=1}^n K\left(\frac{X_{k-1} - x}{b}\right) \mathbb{E}[\epsilon_k \mid \xi_{k-1, k-m}] \right| = \sqrt{\frac{n}{b}} \mathcal{O}_{\mathbb{P}}(\rho^m) = o_{\mathbb{P}}((\log n)^{-2}).$$

Similarly, one can verify in the same way that

$$\sqrt{nb} \sup_{x \in T} \left| \frac{1}{nb} \sum_{k=1}^n K\left(\frac{X_{k-1} - x}{b}\right) (\epsilon_k - \mathbb{E}[\epsilon_k \mid \xi_{k, k-m}]) \right| = o_{\mathbb{P}}((\log n)^{-2}).$$

Furthermore, since the martingale differences are uncorrelated, we have

$$\mathbb{E}[\epsilon_k^2] - \mathbb{E} \left[\left(\mathbb{E}[\epsilon_k | \xi_{k,k-m}] - \mathbb{E}[\epsilon_k | \xi_{k-1,k-m}] \right)^2 \right] = \mathcal{O}(\rho^m).$$

Therefore, defining

$$\zeta_k := \frac{\mathbb{E}[\epsilon_k | \xi_{k,k-m}] - \mathbb{E}[\epsilon_k | \xi_{k-1,k-m}]}{\sqrt{\mathbb{E} \left[\left(\mathbb{E}[\epsilon_k | \xi_{k,k-m}] - \mathbb{E}[\epsilon_k | \xi_{k-1,k-m}] \right)^2 \right]}}$$

we have

$$\sqrt{nb} \sup_{x \in T} \left| \frac{1}{nb} \sum_{k=1}^n K \left(\frac{X_{k-1} - x}{b} \right) \left(\zeta_k - \frac{\epsilon_k}{\sqrt{\mathbb{E}[\epsilon_k^2]}} \right) \right| = o_{\mathbb{P}}((\log n)^{-2}).$$

Next, following exactly the proof of Liu and Wu (2010, Lemma 5.3), we get that uniformly over x ,

$$\begin{aligned} & \frac{1}{nbw(x, b)} \sum_{k=1}^n K \left(\frac{X_{k-1} - x}{b} \right) \epsilon_k \\ &= \frac{1}{nbw(x, b)} \sum_{k=1}^n K \left(\frac{X_{k-1} - x}{b} \right) \sigma(X_k) \zeta_k + \mathcal{O}_{\mathbb{P}} \left(\sqrt{\frac{b \log n}{n}} \right) \\ &= \frac{1}{nb} \frac{1}{\mathbb{E}[w(x, b)] + \mathcal{O}_{\mathbb{P}}(\sqrt{\log n/nb})} \sum_{k=1}^n K \left(\frac{X_{k-1} - x}{b} \right) \sigma(x) \zeta_k + \mathcal{O}_{\mathbb{P}} \left(\sqrt{\frac{b \log n}{n}} \right) \\ &= \frac{1}{nb} \frac{1}{f(x) + \mathcal{O}_{\mathbb{P}}(b^2 + \sqrt{\log n/nb})} \sum_{k=1}^n K \left(\frac{X_{k-1} - x}{b} \right) \sigma(x) \zeta_k + \mathcal{O}_{\mathbb{P}} \left(\sqrt{\frac{b \log n}{n}} \right). \end{aligned}$$

Following the above arguments again we can compute the orders for the decomposition of the term involving K^* and get $t_n(x)$ by the differences. Note that many terms such as $\mu(x)$ in the second term and $\mathcal{O}(b^2)$ term in the first term cancel out. Therefore, overall it can be easily verified that

$$\begin{aligned} t_n(x) &= \frac{\sqrt{f(x)}}{\sigma(x)} \frac{1}{nb f(x)} \sum_{k=1}^n \tilde{K} \left(\frac{X_{k-1} - x}{b} \right) \sigma(x) \zeta_k + \mathcal{O}_{\mathbb{P}} \left(\sqrt{\frac{b \log n}{n}} + b^3 \right) \\ &\quad + \mathcal{O}_{\mathbb{P}}(b^2 + \sqrt{\log n/nb}) \mathcal{O}_{\mathbb{P}}(\sqrt{\log n}), \end{aligned}$$

where $\tilde{K}(\cdot)$ is an anti-symmetric kernel defined by

$$\tilde{K}(u) := K(u) - K^*(u).$$

Now to prove Theorem 3.2, it suffices to show

$$\mathbb{P} \left(\sqrt{\frac{nb}{2\lambda_K}} \sup_{x \in T} \frac{1}{\sqrt{f(x)}} |M_n(x) - M_n^*(x)| - d_n \leq \frac{z}{(2 \log \bar{b}^{-1})^{1/2}} \right) \rightarrow e^{-2e^{-z}},$$

where

$$M_n(x) := \frac{1}{nb} \sum_{k=1}^n K \left(\frac{X_{k-1} - x}{b} \right) \zeta_k, \quad M_n^*(x) := \frac{1}{nb} \sum_{k=1}^n K^* \left(\frac{X_{k-1} - x}{b} \right) \zeta_k.$$

Note that we have $\mathbb{E}[\zeta_i] = 0$ and $\mathbb{E}[\zeta_i^2] = 1$. Next, we define a truncated version of ζ_i by

$$\check{\zeta}_i := \zeta_i \mathbb{1}\{|\zeta_i| \leq (\log n)^{12/(p-2)}\} - \mathbb{E} \left[\zeta_i \mathbb{1}\{|\zeta_i| \leq (\log n)^{12/(p-2)}\} \right].$$

We next define $\tilde{M}_n(x)$ using m -dependent conditional expectations

$$\begin{aligned} \tilde{M}_n(x) &:= \frac{1}{nb} \sum_{k=1}^n \frac{\check{\zeta}_k}{\check{\sigma}^2} \left\{ \mathbb{E} \left[K \left(\frac{X_{k-1} - x}{b} \right) \mid \xi_{k-1,k-m} \right] \right. \\ &\quad \left. - \mathbb{E} \left[K \left(\frac{X_{k-1} - x}{b} \right) \mid \xi_{k-2,k-m} \right] \right\}, \end{aligned}$$

where $\check{\sigma}^2 := \mathbb{E}\check{\zeta}_1^2$.

7.1.3. Alternating big and small blocks

Recall that $m = \lfloor n^\tau \rfloor$. We choose τ_1 such that $\tau < \tau_1 < 1 - \delta_1$ and split $[1, n]$ into alternating big and small blocks $H_1, I_1, \dots, H_{l_n}, I_{l_n}, I_{l_n+1}$ with length $|H_i| = \lfloor n^{\tau_1} \rfloor, |I_i| = \lfloor n^\tau \rfloor, \forall 1 \leq i \leq l_n$, and $|I_{l_n+1}| = n - l_n(\lfloor n^{\tau_1} \rfloor + \lfloor n^\tau \rfloor)$. Note that $l_n = \lfloor n/(\lfloor n^{\tau_1} \rfloor + \lfloor n^\tau \rfloor) \rfloor$. Then we define

$$u_j(x) := \sum_{k \in H_j} \frac{\check{\zeta}_k}{\check{\sigma}^2} \left\{ \mathbb{E} \left[K \left(\frac{X_{k-1} - x}{b} \right) \mid \xi_{k-1, k-m} \right] - \mathbb{E} \left[K \left(\frac{X_{k-1} - x}{b} \right) \mid \xi_{k-2, k-m} \right] \right\}.$$

Then we define

$$\tilde{M}_n(x) := \frac{1}{nb} \sum_{j \in \cup_{i=1}^{l_n} H_i} u_j(x).$$

Next we show in the following that we can approximate $M_n(x)$ by $\tilde{M}_n(x)$ and then approximate $\tilde{M}_n(x)$ by $\check{M}_n(x)$. That is, we show

$$\mathbb{P} \left(\sqrt{nb} \sup_{x \in T} |M_n(x) - \tilde{M}_n(x)| \geq (\log n)^{-2} \right) = o(1). \tag{8}$$

To show Eq. (8), we first follow the proof of Liu and Wu (2010, Lemma 5.1) using Freedman's inequality for martingale differences (Freedman, 1975) to get

$$\mathbb{P} \left(\sqrt{nb} \sup_{x \in T} \left| \frac{1}{nb} \sum_{k=1}^n K \left(\frac{X_{k-1} - x}{b} \right) (\zeta_k - \check{\zeta}_k) \right| \geq 3(\log n)^{-2} \right) = o(1),$$

which implies we can approximate $M_n(x)$ by replacing ζ_k with $\check{\zeta}_k$ in the definition of $M_n(x)$.

Next, we write $K \left(\frac{X_{k-1} - x}{b} \right)$ as a sum of three terms

$$\begin{aligned} & K \left(\frac{X_{k-1} - x}{b} \right) \\ &= \left\{ K \left(\frac{X_{k-1} - x}{b} \right) - \mathbb{E} \left[K \left(\frac{X_{k-1} - x}{b} \right) \mid \xi_{k-1, k-m} \right] \right\} \\ & \quad + \left\{ \mathbb{E} \left[K \left(\frac{X_{k-1} - x}{b} \right) \mid \xi_{k-1, k-m} \right] - \mathbb{E} \left[K \left(\frac{X_{k-1} - x}{b} \right) \mid \xi_{k-2, k-m} \right] \right\} \\ & \quad + \mathbb{E} \left[K \left(\frac{X_{k-1} - x}{b} \right) \mid \xi_{k-2, k-m} \right]. \end{aligned} \tag{9}$$

Note that $\check{\zeta}_k$ is uncorrelated with the second term of the right hand side of Eq. (9). Next, we show that under our assumptions on physical dependence measure, the first term of the right hand side of Eq. (9) becomes very small for large m . In order to rigorously prove this fact, defining

$$Z_k(x) = \check{\zeta}_k \left\{ K \left(\frac{X_{k-1} - x}{b} \right) - \mathbb{E} \left[K \left(\frac{X_{k-1} - x}{b} \right) \mid \xi_{k-1, k-m} \right] \right\},$$

we first approximate $\sum_{k=1}^n Z_k(x)$ by the skeleton process $\sum_{k=1}^{q_n} Z_k(x_j), 1 \leq j \leq q_n$, where $q_n = \lfloor n^2/b \rfloor$ and $x_j = j/(bq_n)$. Following the same arguments as in Liu and Wu (2010, Proof of Lemma 4.2) using Freedman's inequality for martingale differences (Freedman, 1975), we have

$$\sup_{x_{j-1} \leq x \leq x_j} \left| \sum_{k=1}^n (Z_k(x) - Z_k(x_j)) \right| = o_{\mathbb{P}} \left(\sqrt{nb}/(\log b^{-1})^2 \right).$$

Next, we show $\sup_{x \in T} \mathbb{E}|Z_k(x)|$ exponentially decays with m . We consider two cases $|X_{k-1} - \mathbb{E}(X_{k-1} \mid \xi_{k-1, k-m})| \geq \rho_1^m$ and $|X_{k-1} - \mathbb{E}(X_{k-1} \mid \xi_{k-1, k-m})| < \rho_1^m$, where $\rho_1 = \frac{1+\rho}{2}$. Using the assumption $\theta_{n,p} = \mathcal{O}(\rho^n)$, we have

$$\begin{aligned} \sup_{x \in \mathbb{R}} \mathbb{E}|Z_k(x)| &\leq C \mathbb{P}(|X_{k-1} - \mathbb{E}(X_{k-1} \mid \xi_{k-1, k-m})| \geq \rho_1^m) \\ &\quad + C \sup_{x \in \mathbb{R}} \mathbb{P} \left(\left\{ \frac{X_{k-1} - x}{b} \in [-1, 1] \right\} \right) \\ &= \mathcal{O}(\rho/\rho_1)^m + \mathcal{O}(\rho_1^m/b). \end{aligned}$$

Now, we can show the maximum of the skeleton process over $\{x_j\}, j = 1, \dots, q_n$ is small. Recall that m is a polynomial of n , then we have

$$\begin{aligned} & \mathbb{P} \left(\max_{1 \leq j \leq q_n} \left| \sum_{k=1}^n Z_k(x_j) \right| \geq \sqrt{nb}(\log b^{-1})^{-2} \right) \\ & \leq q_n \frac{\max_{1 \leq j \leq q_n} \mathbb{E} \left| \sum_{k=1}^n Z_k(x_j) \right|}{\sqrt{nb}(\log b^{-1})^2} \\ & \leq \frac{nq_n}{\sqrt{nb}(\log b^{-1})^2} \sup_{x \in T} \mathbb{E} |Z_k(x)| = o(1). \end{aligned}$$

Next, we show the third term of the decomposition of $K \left(\frac{X_{k-1} - x}{b} \right)$ in Eq. (9) can also be ignored. In order to show this, we define

$$N_n(x) = \frac{1}{\sqrt{nb}} \sum_{k=1}^n \check{\zeta}_k \mathbb{E} \left[K \left(\frac{X_{k-1} - x}{b} \right) \mid \xi_{k-1, k-m} \right].$$

Using the same argument as in Liu and Wu (2010, Proof of Lemma 4.2), we can approximate $N_n(x)$ by its skeleton process, since $\sup_{x_{j-1} \leq x \leq x_j} |N_n(x) - N_n(x_j)| = o_{\mathbb{P}}(\log n)^{-2}$. We first approximate $\sup_x |N_n(x)|$ by the maximum over the skeleton process. Then we have $\mathbb{P}(\max_{1 \leq j \leq q_n} |N_n(x_j)| \geq (\log n)^{-2}) = o(1)$ using Freedman’s inequality for martingale differences (Freedman, 1975). Therefore, we can approximate $M_n(x)$ by

$$\frac{1}{nb} \sum_{k=1}^n \frac{\check{\zeta}_k}{\mathbb{E}[\check{\zeta}_k^2]} \left\{ \mathbb{E} \left[K \left(\frac{X_{k-1} - x}{b} \right) \mid \xi_{k-1, k-m} \right] - \mathbb{E} \left[K \left(\frac{X_{k-1} - x}{b} \right) \mid \xi_{k-2, k-m} \right] \right\}.$$

Furthermore, since $|1 - \mathbb{E}[\check{\zeta}_k^2]/\mathbb{E}[\zeta_k^2]| = \mathcal{O}((\log n)^{-12/(p-2)})$, we can replace $\check{\zeta}_k/\mathbb{E}[\check{\zeta}_k^2]$ by $\check{\zeta}_k/\check{\sigma}^2$, which leads to the definition of $\tilde{M}_n(x)$. Therefore, we have proved

$$\mathbb{P} \left(\sqrt{nb} \sup_{x \in T} |M_n(x) - \tilde{M}_n(x)| \geq (\log n)^{-2} \right) = o(1).$$

Therefore, in order to finish the proof of Eq. (8), it suffices to show

$$\mathbb{P} \left(\sqrt{nb} \sup_{x \in T} |R_n(x)| \geq (\log n)^{-2} \right) = o(1),$$

where $R_n(x) := \frac{1}{nb} \sum_{j \in \cup_{i=1}^{m+1} I_i} u_j(x)$. Following the same argument as above using skeleton process, we only need to consider the grids $\{x_j, j = 0, \dots, q_n\}$. Using the fact that $\tau < \tau_1$ and $n^{-\delta_1} = \mathcal{O}(b)$, again by Freedman’s inequality for martingale differences, for some constant C that

$$\begin{aligned} & \mathbb{P} \left(\sqrt{nb} \sup_{0 \leq j \leq q_n} |R_n(x_j)| \geq (\log n)^{-3} \right) \\ & \leq 4q_n \exp \left[\frac{(\log n)^{-6} nb}{-2C(\log n)^{-3} \sqrt{nb} - 2C(n^{1-\tau_1+\tau} + n^{\tau_1})b} \right] = o(1), \end{aligned}$$

which finishes the proof of Eq. (8).

Observing that, since $K(\cdot)$ is supported on $[0, 1]$, one of the following two terms must be zero:

$$\begin{aligned} & \mathbb{E} \left[K \left(\frac{X_{k-1} - x}{b} \right) \mid \xi_{k-1, k-m} \right] - \mathbb{E} \left[K \left(\frac{X_{k-1} - x}{b} \right) \mid \xi_{k-2, k-m} \right], \\ & \mathbb{E} \left[K^* \left(\frac{X_{k-1} - x}{b} \right) \mid \xi_{k-1, k-m} \right] - \mathbb{E} \left[K^* \left(\frac{X_{k-1} - x}{b} \right) \mid \xi_{k-2, k-m} \right]. \end{aligned}$$

Hence, defining $\tilde{M}_n^*(x)$ similarly as $\tilde{M}_n(x)$ using $K^*(\cdot)$ instead of $K(\cdot)$, by Eq. (8), we only need to focus on the following term

$$\begin{aligned} \hat{M}_n(x) & := \sqrt{\frac{nb}{2\lambda_{\check{K}} f(x)}} [\tilde{M}_n(x) - \tilde{M}_n^*(x)] \\ & = \frac{1}{\sqrt{nb\lambda_{\check{K}} f(x)}} \sum_{k \in \cup_{i=1}^m H_i} \frac{\check{\zeta}_k}{\check{\sigma}^2} \left\{ \mathbb{E} \left[\tilde{K} \left(\frac{X_{k-1} - x}{b} \right) \mid \xi_{k-1, k-m} \right] \right. \\ & \quad \left. - \mathbb{E} \left[\tilde{K} \left(\frac{X_{k-1} - x}{b} \right) \mid \xi_{k-2, k-m} \right] \right\}. \end{aligned}$$

Clearly, in order to complete the proof of [Theorem 3.2](#), it suffices to show

$$\mathbb{P} \left(\sup_{x \in T} \left| \hat{M}_n(x) \right| - d_n \leq \frac{z}{(2 \log b^{-1})^{1/2}} \right) \rightarrow e^{-2e^{-z}}. \tag{10}$$

7.1.4. Asymptomatic covariance structure

Next, we prove some results on the asymptomatic covariance structure of $\{\hat{M}_n(x)\}$ which will be needed later for Gaussian approximation using the results in [Bickel and Rosenblatt \(1973\)](#). Define the following quantities: $r(s) := \int K(x)K(x+s)dx/\lambda_K$, $\hat{r}(s) := \mathbb{E}\hat{M}_n(x)\hat{M}_n(x+s)$, $\tilde{r}(s) := \int \tilde{K}(x)\tilde{K}(x+s)dx/\lambda_{\tilde{K}}$, and $\tilde{K}_2 := \int_{-1}^1 (\tilde{K}'(x))^2 dx/(2\lambda_{\tilde{K}})$. Note that since $\tilde{K}'(0) > 0$, we have $\int K(u)K^*(u \pm s)du = \mathcal{O}(\int_0^{|s|} x|s-x|dx) = \mathcal{O}(|s|^3) = o(|s|^2)$. Then by the definition of $\tilde{r}(s)$, using $\lambda_{\tilde{K}} = 2\lambda_K$, we have

$$\begin{aligned} \tilde{r}(s) &= \int \tilde{K}(v)\tilde{K}(v+s)dv/\lambda_{\tilde{K}} \\ &= \frac{1}{\lambda_{\tilde{K}}} \int [K(v) - K^*(v)] [K(v+s) - K^*(v+s)] dx \\ &= \frac{1}{2\lambda_K} \left[\int K(v+s)K(v)dv + \int K^*(v+s)K^*(v)dv \right. \\ &\quad \left. - \int K^*(v+s)K(v)dv - \int K(v+s)K^*(v)dv \right] \\ &= r(s) + o(|s|^2). \end{aligned}$$

Next, according to [Bickel and Rosenblatt \(1973, Theorems B1 and B2\)](#), we have $r(s) = 1 - K_2|s|^2 + o(|s|^2)$. Note that

$$\tilde{K}_2 = \int_{-1}^1 (\tilde{K}'(x))^2 dx/(2\lambda_{\tilde{K}}) = \frac{1}{2} \int_{-1}^1 (\tilde{K}'(x))^2 dx/(2\lambda_K) = \frac{1}{2}(2K_2) = K_2.$$

This implies $\tilde{r}(s) = 1 - \tilde{K}_2|s|^2 + o(|s|^2)$, which can also be obtained directly from [Bickel and Rosenblatt \(1973, Theorems B1 and B2\)](#).

Next, we show $\hat{r}(s) = \tilde{r}(s) + \mathcal{O}(b)$. Note that $\{\check{\zeta}_k\}$ are uncorrelated and $\mathbb{E}\check{\zeta}_k = 0$. Then, using $|f(v+s) - \sqrt{f(t)f(s)}| = \mathcal{O}(b)$ uniformly over $|s-t| \leq 2b$ and $|v| \leq 2b$, we have

$$\begin{aligned} &\mathbb{E}\hat{M}_n(t)\hat{M}_n(s) \\ &= \frac{1}{nb\lambda_{\tilde{K}}} \int \frac{1}{\sqrt{f(t)f(s)}} \sum_{k \in \cup_{i=1}^n H_i} \left\{ \mathbb{E} \left[\tilde{K} \left(\frac{X_{k-1}-t}{b} \right) \tilde{K} \left(\frac{X_{k-1}-s}{b} \right) \right] + \mathcal{O}(b^2) \right\} \\ &= \frac{1}{b\lambda_{\tilde{K}}} \int \frac{1}{f(v+s) + \mathcal{O}(b)} \tilde{K} \left(\frac{v-t+s}{b} \right) \tilde{K} \left(\frac{v}{b} \right) f(v+s) dv + \mathcal{O}(b) \\ &= \frac{1}{\lambda_{\tilde{K}}} \int \tilde{K}(v-t+s) \tilde{K}(v) dv + \mathcal{O}(b) = \tilde{r}(t-s) + \mathcal{O}(b). \end{aligned}$$

Therefore, we have proved that, as $s \rightarrow 0$,

$$\tilde{r}(s) = 1 - \tilde{K}_2|s|^2 + o(|s|^2), \quad \tilde{r}(s) = r(s) + o(|s|^2), \quad \hat{r}(s) = \tilde{r}(s) + \mathcal{O}(b). \tag{11}$$

7.1.5. Gaussian approximation

Now, we go back to prove [Eq. \(10\)](#). We use similar techniques as in [Liu and Wu \(2010, Proof of Lemma 4.5\)](#). First, as in [Bickel and Rosenblatt \(1973\)](#), we split the interval T into alternating big and small intervals $W_1, V_1, \dots, W_N, V_N$, where $W_i = [a_i, a_i + w]$, $V_i = [a_i + w, a_{i+1}]$, $a_i = (i-1)(w+v)$, and $N = \lfloor (u-l)/(w+v) \rfloor$. We let w be fixed, and v be small which goes to 0. Since u and l are fixed numbers, without loss of generality, we assume $l = 0$ and $u = 1$ in this proof.

Next, we approximate $\Omega^+ := \sup_{0 \leq t \leq 1} \hat{M}_n(t)$ by big blocks $\{W_k\}$. That is, by $\Psi^+ := \max_{1 \leq k \leq N} \Upsilon_k^+$, where $\Upsilon_k^+ := \sup_{t \in W_k} \hat{M}_n(t)$. Then we further approximate Υ_k^+ via discretization by $\mathcal{E}_k^+ := \max_{1 \leq j \leq \chi} \hat{M}_n(a_k + jax^{-1})$, where $\chi = \lfloor wx/a \rfloor$ with $a > 0$. We define $\Omega^-, \Psi^-, \Upsilon_k^-,$ and \mathcal{E}_k^- similarly by replacing sup or max by inf or min, respectively. Letting $\Omega = \max(\Omega^+, -\Omega^-) = \sup_{0 \leq t \leq 1} |\hat{M}_n(t)|$ and $x_z = d_n + z/(2 \log b^{-1})^{1/2}$, we have

$$\begin{aligned} &|\mathbb{P}(\Omega \geq x_z) - \mathbb{P}(\{\Psi^+ \geq x_z\} \cup \{\Psi^- \leq -x_z\})| \leq R_1 + R_2, \\ &\left| \mathbb{P}(\{\Psi^+ \geq x_z\} \cup \{\Psi^- \leq -x_z\}) - \mathbb{P} \left(\bigcup_{k=1}^N \{\mathcal{E}_k^+ \geq x_z\} \cup \bigcup_{k=1}^N \{\mathcal{E}_k^- \leq -x_z\} \right) \right| \\ &\leq R_3 + R_4. \end{aligned}$$

where

$$\begin{aligned}
 R_1 &:= \mathbb{P} \left(\max_{1 \leq k \leq N} \sup_{t \in V_k} \hat{M}_n(t) \geq x_z \right), \\
 R_2 &:= \mathbb{P} \left(\min_{1 \leq k \leq N} \inf_{t \in V_k} \hat{M}_n(t) \leq -x_z \right), \\
 R_3 &:= \sum_{k=1}^N \left| \mathbb{P}(\mathcal{Y}_k^+ \geq x_z) - \mathbb{P}(\mathcal{E}_k^+ \geq x_z) \right|, \\
 R_4 &:= \sum_{k=1}^N \left| \mathbb{P}(\mathcal{Y}_k^- \leq -x_z) - \mathbb{P}(\mathcal{E}_k^- \leq -x_z) \right|.
 \end{aligned}$$

Next, we are ready to apply Gaussian approximation. We first use discretization for approximating $\hat{M}_n(x)$. Let $s_j = j/(\log n)^6$, $1 \leq j < t_n$, where $t_n = 1 + \lfloor (\log n)^6 t \rfloor$, $s_{t_n} = t$. Write $[s_{j-1}, s_j] = \bigcup_{k=1}^{q_n} [s_{j,k-1}, s_{j,k}]$, where $q_n = \lfloor (s_j - s_{j-1})n^2 \rfloor = \lfloor n^2 / (\log n)^6 \rfloor$ and $s_{j,k} - s_{j,k-1} = (s_j - s_{j-1})/q_n$. Following the same arguments as in Liu and Wu (2010, Proof of Lemma 4.6), we have the following discretization approximation holds for all large enough Q ,

$$\begin{aligned}
 &\mathbb{P} \left(\sup_{0 \leq s \leq t} \hat{M}_n(v + s) \geq x \right) \\
 &\leq \mathbb{P} \left(\max_{1 \leq j \leq t_n} \hat{M}_n(v + s_j) \geq x - (\log n)^{-2} \right) + Cn^{-Q}.
 \end{aligned}$$

Next, we apply the multivariate Gaussian approximation from Zaitsev (1987). To this end, similar to the definition of $u_j(t)$, we first define

$$\begin{aligned}
 \tilde{u}_j(t) &:= \sum_{k \in H_j} \frac{\check{\zeta}_k}{\check{\sigma}^2} \left\{ \mathbb{E} \left[\tilde{K} \left(\frac{X_{k-1} - t}{b} \right) \mid \xi_{k-1, k-m} \right] \right. \\
 &\quad \left. - \mathbb{E} \left[\tilde{K} \left(\frac{X_{k-1} - t}{b} \right) \mid \xi_{k-2, k-m} \right] \right\}, \quad j = 1, \dots, t_n
 \end{aligned}$$

Note that the sequence of random variables $\{\tilde{u}_j(t), j = 1, \dots, t_n\}$ are independent. Then we define

$$\begin{aligned}
 \hat{u}_j(t) &:= \tilde{u}_j(t) \mathbb{1}\{|\tilde{u}_j(t)| \leq \sqrt{nb}(\log n)^{-20p/(p-2)}\} \\
 &\quad - \mathbb{E} \left[\tilde{u}_j(t) \mathbb{1}\{|\tilde{u}_j(t)| \leq \sqrt{nb}(\log n)^{-20p/(p-2)}\} \right].
 \end{aligned}$$

Now we introduce $\hat{M}_n(t) := \frac{1}{\sqrt{nb\lambda_{\check{K}}f(t)}} \sum_{j=1}^{t_n} \hat{u}_j(t)$. Then using (Zaitsev, 1987, Theorem 1.1) as well as $\sup_t \max_{1 \leq j \leq t_n} \|\hat{u}_j(t) - \tilde{u}_j(t)\| \leq Cn^{-Q}$ for large enough Q , we have

$$\begin{aligned}
 &\mathbb{P} \left(\max_{1 \leq j \leq t_n} \hat{M}_n(v + s_j) \geq x - (\log n)^{-2} \right) \\
 &\leq \mathbb{P} \left(\max_{1 \leq j \leq t_n} \hat{M}_n(v + s_j) \geq x - (\log n)^{-2} \right) + Cn^{-Q} \\
 &\leq \mathbb{P} \left(\max_{1 \leq j \leq t_n} Y_n(j) \geq x - 2(\log n)^{-2} \right) + Ct_n^{5/2} \exp \left(-\frac{C(\log n)^{18p/(p-1)}}{t_n^{5/2}} \right) \\
 &\quad + Cn^{-Q},
 \end{aligned}$$

where $(Y_n(1), \dots, Y_n(t_n))$ is a centered Gaussian random vector with covariance matrix $\hat{\Sigma}_n = \text{Cov}(\hat{M}_n(v + s_1), \dots, \hat{M}_n(v + s_{t_n}))$.

Let ψ be the density function of standard Gaussian, and $H_2(a)$ be the Pickands constants (Bickel and Rosenblatt, 1973, Theorem A1, Lemma A1, and Lemma A3). Using Eq. (11), let $t > 0$ be such that $\inf\{s^{-2}(1 - \tilde{r}(s)) : 0 \leq s \leq t\} > 0$. Following exactly the arguments in Liu and Wu (2010, Proof of Lemma 4.6) to apply (Bickel and Rosenblatt, 1973, Lemma A3 and Lemma A4), we can obtain that for $a > 0$,

$$\mathbb{P} \left(\bigcup_{j=1}^{\lfloor tx/a \rfloor} \left\{ \hat{M}_n(v + jax^{-1}) \geq x \right\} \right) = x\psi(x) \frac{H_2(a)}{a} \tilde{K}_2^{1/2} t + o(x\psi(x)),$$

uniformly over $0 \leq v \leq 1$. The limit when $a \rightarrow 0$ also holds, that is

$$\mathbb{P} \left(\bigcup_{0 \leq s \leq t} \{ \hat{M}_n(v+s) \geq x \} \right) = x\psi(x)\tilde{K}_2^{1/2}t/\sqrt{\pi} + o(x\psi(x)), \tag{12}$$

where we have used the Pickands constants $H_2 = \lim_{a \rightarrow 0} H_2(a)/a = 1/\sqrt{\pi}$. The left tail version of the tail bounds also holds with $\geq x$ replaced by $\leq x$. Furthermore, we can show through elementary calculations that

$$\lim_{a \rightarrow 0} \limsup_{v \rightarrow 0} \limsup_{n \rightarrow \infty} R_j = 0, \quad j = 1, \dots, 4.$$

Therefore, it suffices to show the following convergence to Gumbel law

$$\lim_{a \rightarrow 0} \limsup_{v \rightarrow 0} \limsup_{n \rightarrow \infty} \left| \mathbb{P} \left(\bigcup_{k=1}^N \{ \mathcal{E}_k^+ \geq x_z \} \cup \bigcup_{k=1}^N \{ \mathcal{E}_k^- \leq -x_z \} \right) - (1 - e^{-2e^{-z}}) \right| = 0. \tag{13}$$

7.1.6. Convergence to Gumbel distribution

The main steps of the proof of Eq. (13) are as follows. First, we approximate $\hat{M}_n(t)$ by $Y_n(t)$. Then, we approximate $Y_n(t)$ by another quantity $\hat{M}'_n(t)$ which is defined similarly to $\hat{M}_n(x)$ but using a sequence of i.i.d. random variables instead of the dependent time series $\{X_k\}$. Finally, we apply (Rosenblatt, 1976, Theorem) to show convergence to Gumbel distribution.

We define

$$\begin{aligned} \mathbf{B}_{k,j} &:= \{ \hat{M}_n(a_k + jax^{-1}) \geq x \} \cup \{ \hat{M}_n(a_k + jax^{-1}) \leq -x \}, \\ \mathbf{D}_{k,j} &:= \{ Y_n(a_k + jax^{-1}) \geq x \} \cup \{ Y_n(a_k + jax^{-1}) \leq -x \}, \end{aligned}$$

where $Y_n(\cdot)$ is a centered Gaussian process with covariance function

$$\text{Cov}(Y_n(s_1), Y_n(s_2)) = \text{Cov}(\hat{M}_n(s_1), \hat{M}_n(s_2)).$$

First we approximate $\hat{M}_n(t)$ using $Y_n(t)$. Recall that w and v are the lengths of big and small blocks W_i and V_i . Let $N = \lfloor 1/(w+v) \rfloor$. Define a different truncation order for $M_n(t)$ by $\hat{M}'_n(t) := \frac{1}{\sqrt{nb\tilde{\kappa}_f(t)}} \sum_{j=1}^n \hat{u}'_j(t)$ for given d , where

$$\begin{aligned} \hat{u}'_j(t) &:= \tilde{u}_j(t) \mathbb{1} \{ |\tilde{u}_j(t)| \leq \sqrt{nb}(\log n)^{-20dp/(p-2)} \} \\ &\quad - \mathbb{E} \left[\tilde{u}_j(t) \mathbb{1} \{ |\tilde{u}_j(t)| \leq \sqrt{nb}(\log n)^{-20dp/(p-2)} \} \right]. \end{aligned}$$

Then using $\hat{M}'_n(t)$ and following exactly the same proof from Liu and Wu (2010, Proof of Lemma 4.10) to get that, for any fixed integer l that $1 \leq l \leq N/2$,

$$\begin{aligned} &\left| \mathbb{P} \left(\bigcup_{k=1}^N \mathbf{A}_k \right) - \sum_{d=1}^{2l-1} (-1)^{d-1} \left(\sum_{1 \leq i_1 < \dots < i_d \leq N} - \sum_{\mathcal{I}} \right) \mathbb{P} \left(\bigcap_{j=1}^d \mathbf{c}_{i_j} \right) \right| \\ &\leq \frac{C^{2l}}{(2l)!} + \mathcal{O} \left(\frac{1}{\log n} \right), \end{aligned} \tag{14}$$

where $\mathbf{A}_k := \bigcup_{j=1}^{\lfloor wx/a \rfloor} \mathbf{B}_{k,j}$, $\mathbf{C}_k := \bigcup_{j=1}^{\lfloor wx/a \rfloor} \mathbf{D}_{k,j}$, C does not depend on l , and

$$\mathcal{I} := \left\{ 1 \leq i_1 < \dots < i_d \leq N : \min_{1 \leq j \leq d-1} q_j \leq \lfloor 2w^{-1} + 2 \rfloor \right\}.$$

Next, we construct $\hat{M}'_n(t)$ in the following way. Let $\{\eta_i^{(k)}\}$, $i \leq k \leq n$, be i.i.d. copies of $\{\eta_i\}$, and $\xi_j^{(k)} = (\dots, \eta_{j-1}^{(k)}, \eta_j^{(k)})$. Let $X_k^{(k)} = G(\xi_j^{(k)})$. Note that $X_k^{(k)}$, $1 \leq k \leq n$, are i.i.d. Now define \mathbf{A}'_k the same as \mathbf{A}_k except by replacing Y_j and $\{\eta_i\}$ with $X_k^{(k)}$ and $\{\eta_i^{(k)}\}$, respectively. Repeat the previous arguments for getting Eq. (14), we have

$$\begin{aligned} &\left| \mathbb{P} \left(\bigcup_{k=1}^N \mathbf{A}'_k \right) - \sum_{d=1}^{2l-1} (-1)^{d-1} \left(\sum_{1 \leq i_1 < \dots < i_d \leq N} - \sum_{\mathcal{I}} \right) \mathbb{P} \left(\bigcap_{j=1}^d \mathbf{c}_{i_j} \right) \right| \\ &\leq \frac{C^{2l}}{(2l)!} + \mathcal{O} \left(\frac{1}{\log n} \right). \end{aligned}$$

Letting $n \rightarrow \infty$ then $l \rightarrow \infty$, by triangle inequality, we have

$$\limsup_{n \rightarrow \infty} \left| \mathbb{P} \left(\bigcup_{k=1}^N \mathbf{A}_k \right) - \mathbb{P} \left(\bigcup_{k=1}^N \mathbf{A}'_k \right) \right| = 0.$$

Now the key observation is that we can deal with $\{\mathbf{A}'_k\}$ now and \mathbf{A}'_k are defined using $\{X_k^{(k)}\}$ which are i.i.d. Next, we define R'_1 to R'_4 the same as R_1 to R_4 except using $\{X_k^{(k)}\}$ and $\{\eta_i^k\}$ instead of $\{X_k\}$ and $\{\eta_i\}$, then by Eq. (12) and elementary calculations again we have $\lim_{a \rightarrow 0} \limsup_{v \rightarrow 0} \limsup_{n \rightarrow \infty} R'_j = 0$ for $j = 1, \dots, 4$. This implies

$$\lim_{a \rightarrow 0} \lim_{v \rightarrow 0} \limsup_{n \rightarrow \infty} \left| \mathbb{P} \left(\bigcup_{k=1}^N \mathbf{A}'_k \right) - \mathbb{P} \left(\sup_{0 \leq t \leq 1} |\hat{M}'_n(t)| < x \right) \right| = 0,$$

where $\hat{M}'_n(t)$ is defined in the same way as $\hat{M}_n(t)$ by replacing $\{X_k\}$ with $\{X_k^{(k)}\}$, and $\{\eta_i\}$ with $\{\eta_i^k\}$. Finally, since $\{X_k^{(k)}\}$ are i.i.d., we can apply (Rosenblatt, 1976, Theorem), which leads to the convergence of $\mathbb{P} \left(\sup_{0 \leq t \leq 1} |\hat{M}'_n(t)| < x_z \right)$ to $e^{-2e^{-z}}$. This completes the proof of Theorem 3.2.

7.2. Proof of Theorem 3.5

First, let r_n and s_n be positive sequences, then $r_n = \Omega(s_n)$ if $s_n = o(r_n)$. On the other hand, $r_n = \Theta(s_n)$ if both $s_n = \mathcal{O}(r_n)$ and $r_n = \mathcal{O}(s_n)$ hold. Note that

$$\begin{aligned} & \mathbb{P} \left(\left\{ \hat{M} = M \right\} \cap \left\{ \max_{1 \leq i \leq M} |\hat{x}_i - x_i| < c_n \right\} \right) \\ &= \mathbb{P} \left(\max_{1 \leq i \leq M} |\hat{x}_i - x_i| < c_n \mid \hat{M} = M \right) \mathbb{P} \left(\hat{M} = M \right). \end{aligned}$$

We first argue that $\mathbb{P} \left(\hat{M} < M \right) \rightarrow 0$, which implies at least one change point has not been detected, then we can write

$$\mathbb{P} \left(\hat{M} < M \right) \leq \sum_{i=1}^M \mathbb{P} \left(\text{the change point } a_i \text{ is not detected} \right).$$

Then, by the validity of the bootstrap procedure, when $\sqrt{\frac{b \log n}{n}} = o(\tilde{\Delta}_n)$, the power of the test goes to 1 as $n \rightarrow \infty$ which implies that for any i ,

$$\mathbb{P} \left(\text{the change point } a_i \text{ is not detected} \right) \rightarrow 0.$$

This conclude that $\mathbb{P} \left(\hat{M} < M \right) \rightarrow 0$.

Next we argue that $\mathbb{P} \left(\hat{M} > M \right) \rightarrow \alpha$. Note that $\hat{M} > M$ implies there is a set \tilde{T} without any change point in it, however, $\sup_{x \in \tilde{T}} |t_n(x)| \geq C_{n,\alpha}$. Note that by our algorithm, we can consider \tilde{T} to be the largest set constructed by ruling out M intervals from $[l, u]$ such that each interval has length $2b$ and contains one change point. Then since M is a fixed constant and $b \rightarrow 0$, we have $|\tilde{T}| = (|u - l| - 2Mb)^+ \rightarrow |u - l|$. Then we can apply our main result Theorem 3.2 again on \tilde{T} to get that $\mathbb{P} \left(\sup_{x \in \tilde{T}} |t_n(x)| \geq C_{n,\alpha} \right) \rightarrow \alpha$, which implies $\mathbb{P} \left(\hat{M} > M \right) \rightarrow \alpha$.

Therefore, we have $\mathbb{P} \left(\hat{M} = M \right) \rightarrow 1 - \alpha$. Then it suffices to show

$$\mathbb{P} \left(\max_{1 \leq i \leq M} |\hat{x}_i - x_i| < c_n \mid \hat{M} = M \right) \rightarrow 1.$$

Since M is finite, we only need to focus on one change point. Let x_0 be any of the true change point and \hat{x} be its estimate, it suffices to show $\mathbb{P} \left(|\hat{x} - x_0| \geq c_n \mid \hat{M} = M \right) \rightarrow 0$. Without loss of generality, we assume $\hat{x} - x_0 = \hat{c}_n = o_{\mathbb{P}}(b)$ and $t_n(x_0) > 0$. The case $t_n(x_0) < 0$ can be shown using similar arguments. Now we follow similar arguments as in Müller (1992). Define $\zeta(c) := t_n(x_0 + c) - t_n(x_0)$, for $c = o(b)$. Then we can write $\hat{c}_n = \arg \max \zeta(c)$. Therefore, it suffices to show $\hat{c}_n = \mathcal{O}_{\mathbb{P}} \left(\frac{1}{\Delta_n} \sqrt{\frac{b \log n}{n}} \right)$. Suppose b is small enough such that the b -neighborhood of x_0 does not include any other change points, then we apply the previous decomposition in Eq. (7). Note that since x_0 is a change point, without loss of generality, we assume $\mu(x)$ is left continuous at $x = x_0$, then the following term has the order of $\Theta(\tilde{\Delta}_n)$:

$$\frac{1}{nb} \left| \sum_{k=1}^n \tilde{K} \left(\frac{X_{k-1} - x_0}{b} \right) \frac{\mu(x_0)}{f(x_0)} - \sum_{k=1}^n \tilde{K} \left(\frac{X_{k-1} - (x_0 + c)}{b} \right) \frac{\mu(x_0 + c)}{f(x_0 + c)} \right|.$$

Furthermore, using $\int_0^s K(x)dx = \Theta(s^2)$ because of $\tilde{K}'(0) > 0$, considering cases $|X_{k-1} - x_0| \in [0, c]$ and $|X_{k-1} - x_0| \in (c, b]$ separately, we have

$$\begin{aligned} & \left| \frac{1}{nb f(x_0 + c)} \sum_{k=1}^n \tilde{K} \left(\frac{X_{k-1} - (x_0 + c)}{b} \right) [\mu(X_{k-1}) - \mu(x_0 + c)] \right. \\ & \quad \left. - \frac{1}{nb f(x_0)} \sum_{k=1}^n \tilde{K} \left(\frac{X_{k-1} - x_0}{b} \right) [\mu(X_{k-1}) - \mu(x_0)] \right| \\ &= \left[\frac{1}{b} \int_0^c K \left(\frac{x}{b} \right) dx \right] \Theta_{\mathbb{P}}(\tilde{\Delta}_n) \\ & \quad + \left[\frac{1}{b} \int_c^b K \left(\frac{x}{b} \right) dx \right] \mathcal{O}_{\mathbb{P}}(b^3 + \sqrt{b \log n/n}) \\ &= \left[\int_0^{c/b} K(x) dx \right] \Theta_{\mathbb{P}}(\tilde{\Delta}_n) + \mathcal{O}_{\mathbb{P}}(b^3 + \sqrt{b \log n/n}) \\ &= (c/b)^2 \Theta_{\mathbb{P}}(\tilde{\Delta}_n) + \mathcal{O}_{\mathbb{P}}(b^3 + \sqrt{b \log n/n}). \end{aligned}$$

Finally, by the assumptions on K' in [Theorem 3.5](#), we can follow the same arguments in the proof of [Theorem 3.2](#) as the m -dependent approximation [Section 7.1.2](#) and alternating big/small blocks [Section 7.1.3](#) applying to \tilde{K}' instead of \tilde{K} to get

$$\frac{1}{nb f(x)} \sum_{k=1}^n \tilde{K}' \left(\frac{X_{k-1} - x}{b} \right) \epsilon_k = \mathcal{O}_{\mathbb{P}} \left(\sqrt{\frac{\log n}{nb}} \right).$$

Furthermore, using the fact that $|\tilde{K}''(u)|$ is uniformly bounded and mean value theorem, we have

$$\begin{aligned} & \mathbb{E} \left[\frac{1}{f(x)} \left(\tilde{K} \left(\frac{X_{k-1} - x}{b} \right) - \tilde{K} \left(\frac{X_{k-1} - x + c}{b} \right) + \left(\frac{c}{b} \right) \tilde{K}' \left(\frac{X_{k-1} - x}{b} \right) \right)^2 \right] \\ &= \int \frac{1}{f(x)} \left[\tilde{K} \left(\frac{y - x}{b} \right) - \tilde{K} \left(\frac{y - x + c}{b} \right) + \left(\frac{c}{b} \right) \tilde{K}' \left(\frac{y - x}{b} \right) \right]^2 f(y) dy \\ &= \int b \left[\tilde{K}(t) - \tilde{K} \left(t + \frac{c}{b} \right) + \left(\frac{c}{b} \right) \tilde{K}'(t) \right]^2 \frac{f(tb + x)}{f(x)} dt \\ &= b \left(\frac{c}{b} \right)^2 \int \left[\frac{\tilde{K}(t) - \tilde{K}(t + c/b)}{c/b} + \tilde{K}'(t) \right]^2 (1 + \mathcal{O}(b)) dt = \mathcal{O} \left(b \left(\frac{c}{b} \right)^4 \right). \end{aligned}$$

Next, we define a new kernel \check{K} such that

$$\begin{aligned} & \check{K} \left(\frac{X_{k-1} - x}{b} \right) \\ &:= \left(\frac{b}{c} \right)^2 \left[\tilde{K} \left(\frac{X_{k-1} - x}{b} \right) - \tilde{K} \left(\frac{X_{k-1} - x + c}{b} \right) + \left(\frac{c}{b} \right) \tilde{K}' \left(\frac{X_{k-1} - x}{b} \right) \right] \end{aligned}$$

so we have $\mathbb{E} \left[\frac{1}{f(x)} \check{K} \left(\frac{X_{k-1} - x}{b} \right)^2 \right] = \mathcal{O}(b)$. Then we can approximate the following term using the same arguments of m -dependent approximation and alternating big/small blocks as in [Sections 7.1.2](#) and [7.1.3](#) in the proof of [Theorem 3.2](#) applying to this new kernel \check{K} to get

$$\begin{aligned} & \frac{1}{nb f(x)} \sum_{k=1}^n \left[\tilde{K} \left(\frac{X_{k-1} - x}{b} \right) - \tilde{K} \left(\frac{X_{k-1} - x + c}{b} \right) + \left(\frac{c}{b} \right) \tilde{K}' \left(\frac{X_{k-1} - x}{b} \right) \right] \epsilon_k \\ &= \left(\frac{c}{b} \right)^2 \left[\frac{1}{nb f(x)} \sum_{k=1}^n \check{K} \left(\frac{X_{k-1} - x}{b} \right) \epsilon_k \right] = \mathcal{O}_{\mathbb{P}} \left(\left(\frac{c}{b} \right)^2 \sqrt{\frac{\log n}{nb}} \right). \end{aligned}$$

Therefore, we have

$$\begin{aligned} & \frac{1}{nhf(x)} \sum_{k=1}^n \left[\tilde{K} \left(\frac{X_{k-1} - x}{b} \right) - \tilde{K} \left(\frac{X_{k-1} - x + c}{b} \right) \right] \epsilon_k \\ &= \left(\frac{c}{b} \right) \left[\frac{-1}{nhf(x)} \sum_{k=1}^n \tilde{K}' \left(\frac{X_{k-1} - x}{b} \right) \epsilon_k \right] + \mathcal{O}_{\mathbb{P}} \left(\left(\frac{c}{b} \right)^2 \sqrt{\frac{\log n}{nb}} \right) \\ &= \left(\frac{c}{b} \right) \mathcal{O}_{\mathbb{P}} \left(\sqrt{\frac{\log n}{nb}} \right) + \left(\frac{c}{b} \right)^2 \mathcal{O}_{\mathbb{P}} \left(\sqrt{\frac{\log n}{nb}} \right). \end{aligned}$$

Then using $\sqrt{\frac{\log n}{nb}} = o(\tilde{\Delta}_n)$ we can conclude that

$$\zeta(c) = - \left(\frac{c}{b} \right)^2 \mathcal{O}_{\mathbb{P}}(\tilde{\Delta}_n) + \left(\frac{c}{b} \right) \mathcal{O}_{\mathbb{P}} \left(\sqrt{\frac{\log n}{nb}} \right) - \mathcal{O}_{\mathbb{P}}(b^3 + \sqrt{b \log n/n}).$$

Recall that the estimated change point $\hat{x} = x_0 + \hat{c}_n$, where $\hat{c}_n = \arg \max \zeta(c)$, then we have

$$\hat{c}_n = \mathcal{O}_{\mathbb{P}} \left(\frac{b}{\tilde{\Delta}_n} \sqrt{\frac{\log n}{nb}} \right) = \mathcal{O}_{\mathbb{P}} \left(\frac{1}{\tilde{\Delta}_n} \sqrt{\frac{b \log n}{n}} \right),$$

whenever $b^4 = o((\log n)/(n\tilde{\Delta}_n))$ and $b^3 = o((\log n)/(n\tilde{\Delta}_n^2))$. This is always true since we have assumed $\delta_2 \leq 1/4$ which implies $b = \mathcal{O}(n^{-1/4})$ so $b^4 = \mathcal{O}(1/n) = o((\log n)/n)$. Therefore, if we choose $c_n > 0$ such that $\hat{c}_n = o(c_n)$, then we have $\mathbb{P}(|\hat{c}_n| < c_n) \rightarrow 0$, which implies $\mathbb{P}(|\hat{x} - x_0| \geq c_n \mid \hat{M} = M) \rightarrow 0$.

8. Additional proofs

8.1. Proof of Remark 3.1

For $\sigma_n^2(x)$, we first write it as the sum of three terms:

$$\begin{aligned} \sigma_n^2(x) &= \frac{1}{nhf_n(x)} \sum_{k=1}^n W \left(\frac{X_k - x}{h} \right) \epsilon_k^2 \\ &+ \frac{2}{nhf_n(x)} \sum_{k=1}^n W \left(\frac{X_k - x}{h} \right) [\mu(X_k) - \mu_n(X_k)] \epsilon_k \\ &+ \frac{1}{nhf_n(x)} \sum_{k=1}^n W \left(\frac{X_k - x}{h} \right) [\mu(X_k) - \mu_n(X_k)]^2. \end{aligned}$$

For the first term, we first approximate ϵ_k^2 by $\{\mathbb{E}[\epsilon_k^2 \mid \xi_{k,k-m}]\}$ where $m = \lfloor n^\tau \rfloor$ with $\tau > 0$ small enough. Using the same argument as in Section 7.1, we have

$$\sup_x \left| \frac{1}{nh} \sum_{k=1}^n W \left(\frac{X_k - x}{h} \right) \left\{ \epsilon_k^2 - \mathbb{E}[\epsilon_k^2 \mid \xi_{k,k-m}] \right\} \right| = \mathcal{O}_{\mathbb{P}}(\rho^m) = o_{\mathbb{P}}(n^{-1/2}),$$

where we choose $m = c \log n$ with $c > -\frac{1}{2 \log(\rho)}$. We then divide $1, \dots, m$ into $\lfloor n/m \rfloor + 1$ blocks indexed by $1, \dots, \lfloor n/m \rfloor + 1$. Then it is clear that the sum of blocks with odd indices is independent with the sum of blocks with even indices. Following the same argument as the proof of Liu and Wu (2010, Theorem 2.5) for each subsequence of the blocks, and use a union bound, we can get

$$\begin{aligned} & \sup_x \left| \frac{1}{nhf_n(x)} \sum_{k=1}^n W \left(\frac{X_k - x}{h} \right) \mathbb{E}[\epsilon_k^2 \mid \xi_{k,k-m}] \right| \\ &= \mathcal{O}_{\mathbb{P}} \left(2m \left(h^2 + \sqrt{\frac{\log(n/m)}{(n/m)h}} \right) \right) \\ &= \mathcal{O}_{\mathbb{P}} \left(\log n \left(h^2 + \frac{\log n}{\sqrt{nh}} \right) \right) = \mathcal{O}_{\mathbb{P}} \left(h^2 \log n + \frac{(\log n)^2}{\sqrt{nh}} \right) \end{aligned}$$

For the second term, we first approximate $\{\epsilon_k\}$ using $\{\epsilon'_k\}$, where $\epsilon'_k := \mathbb{E}[\epsilon_k | \xi_{k,k-m}] - \mathbb{E}[\epsilon_k | \xi_{k-1,k-m}]$. Then following the same argument as in Section 7.1 we have

$$\sup_x \left| \frac{1}{nh} \sum_{k=1}^n W\left(\frac{X_k - x}{h}\right) (\epsilon_k - \epsilon'_k) \right| = \mathcal{O}(\rho^m).$$

Then, again choosing $m = c \log n$ and divide $1, \dots, n$ into $\lfloor n/m \rfloor + 1$ blocks, by the same argument as in Zhao and Wu (2008, pp. 1875), we can get

$$\begin{aligned} \sup_x \left| \frac{2}{nhf_n(x)} \sum_{k=1}^n W\left(\frac{X_k - x}{h}\right) [\mu(X_k) - \mu_n(X_k)] \epsilon_k \right| \\ = \mathcal{O}_{\mathbb{P}} \left((\log n)^2 \left(\frac{\log n}{nh^{5/2}} + \rho^m \right) \right) = \mathcal{O}_{\mathbb{P}} \left(\frac{(\log n)^3}{nh^{5/2}} + \frac{(\log n)^2}{\sqrt{n}} \right) \end{aligned}$$

Finally, for the last term, we have

$$\begin{aligned} \sup_x \left| \frac{1}{nhf_n(x)} \sum_{k=1}^n W\left(\frac{X_k - x}{h}\right) [\mu(X_k) - \mu_n(X_k)]^2 \right| \\ = \mathcal{O}_{\mathbb{P}} \left(h^4 + \frac{\log n}{nh} \right) \cdot \sup_x \frac{1}{nh} \sum_{k=1}^n \left| W\left(\frac{X_k - x}{h}\right) \right| = \mathcal{O}_{\mathbb{P}} \left(h^4 + \frac{\log n}{nh} \right). \end{aligned}$$

Then, using $0 < \delta_1 < 1/4$ we have that

$$\begin{aligned} \sup_x |\sigma_n^2(x) - \sigma^2(x)| &= \mathcal{O}_{\mathbb{P}} \left(h^2 \log n + \frac{(\log n)^2}{\sqrt{nh}} + \frac{(\log n)^3}{nh^{5/2}} \right) \\ &= \mathcal{O}_{\mathbb{P}} \left(h^2 \log n + \frac{(\log n)^3}{\sqrt{nh}} \right) \end{aligned}$$

For $f_n(x)$, similarly, by the same arguments as the proof for $\sigma_n^2(x)$, following the proof of Liu and Wu (2010, Lemma 4.4), we can obtain $\sup_x |f_n(x) - f(x)| = \mathcal{O}_{\mathbb{P}} \left(\frac{(\log n)^3}{\sqrt{nh}} + h^2 \log n \right)$.

8.2. Proof of Proposition 4.1

Since $\{U_k\}_{k=0}^n$ are i.i.d. standard Gaussian distributed random variables, the proof for this proposition is simpler than Theorem 3.2. We can immediately prove the convergence to Gumbel distribution by using (Rosenblatt, 1976, Theorem 1).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors are grateful to the editor, Prof. Serena Ng, and two anonymous referees for their valuable comments and suggestions which significantly improved the quality of the paper. Zhou Zhou's research has been partially sponsored by NSERC, Canada (No. 489079). Yan Cui is supported by the China Scholarship Council (No. 201806170148).

References

Bickel, P.J., Rosenblatt, M., 1973. On some global measures of the deviations of density function estimates. *Ann. Statist.* 1 (6), 1071–1095.
 Bollerslev, T., Law, T.H., Tauchen, G., 2008. Risk, jumps, and diversification. *J. Econometrics* 144 (1), 234–256.
 Chapman, D.A., Pearson, N.D., 2000. Is the short rate drift actually nonlinear? *J. Finance* 55 (1), 355–388.
 Durlauf, S.N., Johnson, P.A., 1995. Multiple regimes and cross-country growth behaviour. *J. Appl. Econometrics* 10 (4), 365–384.
 Engle, R.F., 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* 50 (4), 987–1007.
 Eubank, R., Speckman, P., 1994. Nonparametric estimation of functions with jump discontinuities. *Lect. Notes-Monogr. Ser.* 130–144.
 Fan, J., Yao, Q., 2003. *Nonlinear Time Series: Nonparametric and Parametric Methods*. Springer.
 Fan, J., Zhang, C., 2003. A re-examination of diffusion estimators with applications to financial model validation. *J. Amer. Statist. Assoc.* 98 (461), 118–134.
 Freedman, D.A., 1975. On tail probabilities for martingales. *Ann. Probab.* 3 (1), 100–118.
 Gijbels, I., Hall, P., Kneip, A., 1999. On the estimation of jump points in smooth curves. *Ann. Inst. Statist. Math.* 51 (2), 231–251.
 Gijbels, I., Lambert, A., Qiu, P., 2007. Jump-preserving regression and smoothing using local linear fitting: a compromise. *Ann. Inst. Statist. Math.* 59 (2), 235–272.

- Haggan, V., Ozaki, T., 1981. Modelling nonlinear random vibrations using an amplitude-dependent autoregressive time series model. *Biometrika* (ISSN: 0006-3444) 68 (1), 189–196.
- Jiang, G.J., Oomen, R.C., 2008. Testing for jumps when asset prices are observed with noise—a “swap variance” approach. *J. Econometrics* 144 (2), 352–370.
- Kalman, R.E., 1960. A new approach to linear filtering and prediction problems. *J. Basic Eng.* 82 (1), 35–45.
- Lee, S.S., Mykland, P.A., 2012. Jumps in equilibrium prices and market microstructure noise. *J. Econometrics* 168 (2), 396–406.
- Liu, X., Chen, R., Tsay, R.S., 2020. NTS: An R package for nonlinear time series analysis. *R J.* 12 (2), 266–292.
- Liu, W., Wu, W.B., 2010. Simultaneous nonparametric inference of time series. *Ann. Statist.* 38 (4), 2388–2421.
- McConnell, M.M., Perez-Quiros, G., 2000. Output fluctuations in the United States: What has changed since the early 1980's? *Amer. Econ. Rev.* 90 (5), 1464–1476.
- Müller, H.-G., 1992. Change-points in nonparametric regression analysis. *Ann. Statist.* 20 (2), 737–761.
- Müller, H.-G., Song, K.-S., 1997. Two-stage change-point estimators in smooth regression models. *Statist. Probab. Lett.* 34 (4), 323–335.
- Rice, J.A., Silverman, B.W., 1991. Estimating the mean and covariance structure nonparametrically when the data are curves. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 53 (1), 233–243.
- Rosenblatt, M., 1976. On the maximal deviation of k -dimensional density estimates. *Ann. Probab.* 4 (6), 1009–1015.
- Shao, X., Wu, W.B., 2007. Asymptotic spectral theory for nonlinear time series. *Ann. Statist.* 35 (4), 1773–1801.
- Shumway, R.H., Stoffer, D.S., 2000. *Time Series Analysis and its Applications*, Vol. 3. Springer.
- Stanton, R., 1997. A nonparametric model of term structure dynamics and the market price of interest rate risk. *J. Finance* 52 (5), 1973–2002.
- Tiao, G.C., Tsay, R.S., 1994. Some advances in non-linear and adaptive modelling in time-series. *J. Forecast.* 13 (2), 109–131.
- Tong, H., 1990. *Nonlinear Time Series Analysis: A Dynamics Approach*. Oxford University Press.
- Tong, H., Lim, K.S., 1980. Threshold autoregression, limit cycles and cyclical data (with discussion). *J. R. Stat. Soc. Ser. B Stat. Methodol.* (ISSN: 1369-7412) 42 (3), 245–292.
- Tsay, R.S., 1989. Testing and modeling threshold autoregressive processes. *J. Amer. Statist. Assoc.* 84 (405), 231–240.
- Wang, Y., 1995. Jump and sharp cusp detection by wavelets. *Biometrika* 82 (2), 385–397.
- Wu, W.B., 2005. Nonlinear system theory: another look at dependence. *Proc. Natl. Acad. Sci. USA* 102 (40), 14150–14154.
- Zaitsev, A.Y., 1987. On the Gaussian approximation of convolutions under multidimensional analogues of SN Bernstein's inequality conditions. *Probab. Theory Related Fields* 74 (4), 535–566.
- Zhang, T., 2016. Testing for jumps in the presence of smooth changes in trends of nonstationary time series. *Electron. J. Stat.* 10 (1), 706–735.
- Zhao, Z., 2011. Nonparametric model validations for hidden Markov models with applications in financial econometrics. *J. Econometrics* 162 (2), 225–239.
- Zhao, Z., Wu, W.B., 2006. Kernel Quantile Regression for Nonlinear Stochastic Models. Technical Report, Department of Statistics, University of Chicago.
- Zhao, Z., Wu, W.B., 2008. Confidence bands in nonparametric time series regression. *Ann. Statist.* 36 (4), 1854–1878.
- Zhou, Z., 2012. Measuring nonlinear dependence in time-series, a distance correlation approach. *J. Time Series Anal.* 33 (3), 438–457.