Bernoulli **30**(2), 2024, 1231–1256
https://doi.org/10.3150/23-BEJ1630

Mean stationarity test in time series: A signal variance-based approach

HON KIU TO[a] and KIN WAI CHAN[b]

Department of Statistics, The Chinese University of Hong Kong, Shatin, Hong Kong, [a]*tojames@link.cuhk.edu.hk,* [b]*kinwaichan@cuhk.edu.hk*

Inference of mean structure is an important problem in time series analysis. Various tests have been developed to test for different mean structures, for example, the presence of structural breaks, and parametric mean structures. However, many of them are designed for handling specific mean structures, and may lose power upon violation of such structural assumptions. In this paper, we propose a new mean stationarity test built around the signal variance. The proposed test is based on a super-efficient estimator which could achieve a convergence rate faster than $\sqrt{n}$. It can detect non-constancy of the mean function under serial dependence. It is shown to have promising power, especially in detecting hardly noticeable oscillating structures. The proposal is further generalized to test for smooth trend structures and relative signal variability.

Keywords: Difference variate; mean stationarity; non-linear time series; relative variability; signal variance; super-efficiency

1. Introduction

Inference of mean stationarity has been widely studied in the literature. It can be divided into various categories according to the structure of alternative hypothesis H_1. The single change point alternative hypothesis has been vastly studied. Popular tests include the Kolmogorov–Smirnov (KS) change point test and its variants (Crainiceanu and Vogelsang, 2007, Csörgő and Horváth, 1997, Górecki, Horváth and Kokoszka, 2018, Horváth, Kokoszka and Steinebach, 1999, Juhl and Xiao, 2009), and the self-normalized (SN) KS test (Shao and Zhang, 2010). A natural extension is the multiple-change-point alternative hypothesis. When the number of change points $\mathscr{J}$ is specified, existing tests include Bai and Perron (1998), and Antoch and Jarušková (2013). When $\mathscr{J}$ is not specified, one can use the unsupervised SN change point test (Chen, Wang and Wu, 2022, Cheng and Chan, 2023, Jiang, Zhao and Shao, 2022, Zhang and Lavitas, 2018) and wild binary segmentation (Fryzlewicz, 2014). On the other hand, some literatures are interested in the stability of the smooth time-varying mean; see, e.g., the mass excess test of relevant change in mean (Dette and Wu, 2019). The inference of mean stationarity with the incorporation of both change point structure and time-varying mean structure has been studied as well; see, e.g., mean constancy test (Dalla, Giraitis and Phillips, 2015, Wu, 2004). Other mean stationarity-related literatures include tests for monotonic trend (Wu, Woodroofe and Mentz, 2001), construction of simultaneous confidence bands for trend (Wu and Zhao, 2007), Kwiatkowski–Phillips–Schmidt–Shin (KPSS) stationarity test and its generalizations (Hobijn, Franses and Ooms, 2004, Kwiatkowski et al., 1992), Priestley–Subba Rao (PSR) test (Priestley and Subba Rao, 1969) and test for parametric assumptions of trends (Chen and Wu, 2019, Zhang and Wu, 2011).

This paper proposes a measure of mean stationarity, namely the signal variance (SV) formally defined in (3), and a difference-based method to conduct statistical inference on the mean stationarity of time series. The proposed method incorporates both change points structure and time-varying mean structure while requiring mild assumptions on the observed time series. The remaining parts of the paper are organized as follows. In Section 2, we shall discuss the mathematical setup of the problem

1350-7265

and the underlying framework for the asymptotic theories. In Section 3, we define aggregated variability and SV, and discuss the proposed estimator and the underlying motivation. In Section 4, a new test for mean stationarity and its extensions are given. The proposed mean-invariance test (MIT), trend stationary test, relative variability test, marginal relative variability test, are defined in (9), (15), (19), and (25), respectively. In Section 5, we discuss the implementation issues. In Section 6, finite sample performances are demonstrated.

2. Problem formulation

2.1. Notation

Throughout the paper, the following notation is adopted. Let $\mathbb{N} = \{1,2,\ldots\}, \mathbb{N}_0 = \mathbb{N} \cup \{0\}$, and $\mathbb{R}^+ = (0,\infty)$. For any statement A, $\mathbb{1}(A)$ denotes the indicator function of A, i.e., $\mathbb{1}(A) = 1$ if A is true, and $\mathbb{1}(A) = 0$ otherwise. For any $\{a_n \in \mathbb{R}^+\}_{n\in\mathbb{N}}$ and $\{b_n \in \mathbb{R}^+\}_{n\in\mathbb{N}}$, $a_n \asymp b_n$ means there exists $C > 0$ such that $1/C \le a_n/b_n \le C$; $a_n \ll b_n$ or $a_n = o(b_n)$ means $a_n/b_n \to 0$; $a_n \lesssim b_n$ or $a_n = O(b_n)$ means there exists $C > 0$ such that $a_n/b_n \le C$ for all large n. Convergence in probability and in distribution are denoted by "$\xrightarrow{\text{pr}}$" and "$\xrightarrow{\text{d}}$", respectively. For a random variable X, denote $\|X\|_p = \mathsf{E}(|X|^p)^{1/p}$, where $p \ge 1$. By convention, $\|X\| = \|X\|_2$. Denote $X \in \mathscr{L}^p$ if $\|X\|_p < \infty$. Independent and identically distributed (i.i.d.) random variables $X_1,\ldots,X_n$ following a distribution F are denoted by $X_i \overset{\text{IID}}{\sim} F$. For any sequence of variables $A_1,\ldots,A_n$, denote $\bar{A}_n = \sum_{i=1}^n A_i/n$. For any estimator $\widehat{\theta}$ of θ, denote $\mathsf{Bias}(\widehat{\theta};\theta) = \mathsf{Bias}(\widehat{\theta}) = \mathsf{E}(\widehat{\theta}) - \theta$ and $\mathsf{MSE}(\widehat{\theta};\theta) = \mathsf{MSE}(\widehat{\theta}) = \mathsf{E}(\widehat{\theta} - \theta)^2$. A function $f : [0,1] \to \mathbb{R}$ is said to be of bounded variation if

$$V(f) := \sup_{P\in\mathscr{P}} \left\{ \sum_{i=1}^{n_P} |f(x_i) - f(x_{i-1})| \right\} < \infty,$$

where the supremum is taken over $\mathscr{P}$ which is the set of all partitions of $[0,1]$.

2.2. Model

Let the observed time series $\{X_i\}_{i=1}^n$ be generated from the signal-plus-noise model:

$$X_i = \mu_i + Z_i, \qquad i = 1,\ldots,n, \tag{1}$$

where the deterministic signals $\{\mu_i\}_{i=1}^n$ and the noises $\{Z_i\}_{i\in\mathbb{Z}}$ are not directly observable. Denote $\mu_i = \mu(i/n)$ for $i = 1,\ldots,n$, where $\mu : [0,1] \to \mathbb{R}$ is a mean function that admits the following form.

Assumption 2.1. The mean function $\mu(\cdot)$ is the sum of a Lipschitz continuous function $c(\cdot)$ and a step discontinuous function $s(\cdot)$, i.e.,

$$\mu(t) = c(t) + s(t), \qquad s(t) = \sum_{j=0}^{\mathcal{J}} \xi_j \mathbb{1}\left(\frac{T_j}{n} \le t < \frac{T_{j+1}}{n}\right), \tag{2}$$

where $\mathcal{J}$ is the number of discontinuities, $1 \equiv T_0 < T_1 < \cdots < T_{\mathcal{J}} < T_{\mathcal{J}+1} \equiv n+1$ are the times of discontinuities, and $\xi_0,\ldots,\xi_{\mathcal{J}}$ are the step sizes such that $\xi_j - \xi_{j-1} \ne 0$ for all $1 \le j \le \mathcal{J}$.

Assumption 2.2. The discontinuities of the mean function $\mu(\cdot)$ are not closely packed in the sense that $\min_{0\le j\le \mathcal{J}} |T_{j+1} - T_j| \gtrsim \ell$, where $\ell \in \mathbb{N}$ is a parameter to be specified.

Assumption 2.1 is a general setting asserting that the mean function $\mu(\cdot)$ is a piecewise Lipschitz continuous function with $\mathcal{J}$ pieces. Similar settings have been used in other related work; see, e.g., Wu and Zhao (2007) and Dalla, Giraitis and Phillips (2015). Assumption 2.2 controls the frequency of jumps such that $\mu(\cdot)$ does not fluctuate too frequently. The parameter ℓ may vary in different cases, thus, ℓ is specified when Assumption 2.2 is needed.

In view of (2), we assess $\mu(\cdot)$ through two measures, $\mathscr{C}$ and $\mathscr{S}$, which denote the smoothness of $c(\cdot)$ and the maximum step magnitude of $s(\cdot)$, respectively. Mathematically,

$$\mathscr{C} = \sup_{0\le t'<t\le 1} \left| \frac{c(t)-c(t')}{t-t'} \right| \qquad \text{and} \qquad \mathscr{S} = \sup_{j=1,\ldots,\mathcal{J}} |\xi_j - \xi_{j-1}|.$$

Note that $\mathscr{C} = 0$ indicates the absence of continuous trend effect, and $\mathscr{S} = 0$ or $\mathcal{J} = 0$ indicates the absence of step discontinuities. The mean function $\mu(\cdot)$ is constant if $\mathscr{C} = \mathscr{S} = 0$.

2.3. Serial dependence structure

We follow the dependence measure framework developed by Wu (2005). It is noted that classical formulations of dependence measures, including strong mixing conditions of various types (Dedecker and Prieur, 2005, Rosenblatt, 1956) and near-epoch dependence conditions (Andrews, 1995, Ibragimov, 1962), are also widely adopted; see Bradley (2005) for a survey. Developing theories based on these frameworks are also interesting, however, it is beyond the scope of this article and is left for future research. Suppose that $Z_i = g(\mathscr{F}_i)$ for some measurable function g where $\mathscr{F}_i = (\ldots, \varepsilon_{i-1}, \varepsilon_i)$ is the shift process of i.i.d. innovations $\{\varepsilon_i\}_{i\in\mathbb{Z}}$. Define the projection operator by $\mathscr{P}_i\cdot := \mathsf{E}(\cdot \mid \mathscr{F}_i) - \mathsf{E}(\cdot \mid \mathscr{F}_{i-1})$. For $p \ge 1$, define the physical dependence measure and its aggregated value by

$$\omega_{p,i} := \left\| Z_i - Z_{i,\{0\}} \right\|_p \qquad \text{and} \qquad \Omega_p := \sum_{i=0}^{\infty} \omega_{p,i},$$

respectively, where $\mathscr{F}_{i,\{0\}} := (\mathscr{F}_{-1}, \varepsilon_0', \varepsilon_1, \ldots, \varepsilon_i)$, ε_0' be an i.i.d. copy of ε_j, and $Z_{i,\{0\}} := g(\mathscr{F}_{i,\{0\}})$. We impose the following weak dependence condition on the noises $\{Z_i\}_{i\in\mathbb{Z}}$.

Assumption 2.3. The noise sequence $\{Z_i\}_{i\in\mathbb{Z}}$ is a mean-zero strictly stationary time series such that $Z_1 \in \mathscr{L}^{4+\iota}$ for some $\iota > 0$, and $\Omega_4 < \infty$.

The finiteness of Ω_4 provides a mild and easily verifiable condition for asymptotic theory; see Wu (2007, 2011). Moreover, Assumption 2.3 implies the absolute summability of the autocovariances of $\{Z_i\}_{i\in\mathbb{Z}}$, i.e., $\sum_{k\in\mathbb{Z}} |\gamma_k| < \infty$, where $\gamma_k = \mathsf{E}(Z_0 Z_k)$, which further implies the existence of the asymptotic variance $v = \lim_{n\to\infty} n\mathsf{Var}(\bar{Z}_n) = \sum_{k\in\mathbb{Z}} \gamma_k$ and assuring some nice properties.

3. Measure of variability of mean function

3.1. Signal variance and aggregated variability

The asymptotic variance v measures the stochastic variability of the noises $\{Z_i\}_{i=1}^n$, but not the intrinsic variability of the signals $\{\mu_i\}_{i=1}^n$ because $n\mathsf{Var}(\bar{X}_n) \to v$ is invariant to the sequence $\{\mu_i\}_{i=1}^n$. As a result, v fails to measure the overall observed variability. Therefore, we need a new variability measure.

A natural measure of variability of $\mu(\cdot)$ is the *signal variance* (SV) defined as

$$\theta = \int_0^1 \{\mu(t) - \bar{\mu}\}^2 \, \mathrm{d}t, \tag{3}$$

where $\bar{\mu} = \int_0^1 \mu(t)\,\mathrm{d}t$. It is clear that $\mu(\cdot)$ is constant if and only if $\theta = 0$. Being a one number summary of the functional characteristic, θ is useful for assessing the variability of $\mu(\cdot)$.

The SV θ can be represented in terms of variance. We introduce randomly permuted signals $M_i = \mu(U_i)$ for $i = 1,\ldots,n$, where $U_1,\ldots,U_n \overset{\text{IID}}{\sim} \text{Unif}(0,1)$ are independent of $\{Z_i\}_{i\in\mathbb{Z}}$, and Unif(0,1) stands for the uniform distribution on $(0,1)$. Then $\lim_{n\to\infty} n\mathsf{Var}(\bar{M}_n) = \theta$. In other words, SV can be interpreted as the asymptotic variance of the randomly permuted signals, i.e., the randomness is generated by permutation. So the mean-permuted time series becomes $X_i^* = \mu(U_i) + Z_i$. It allows us to define the *aggregated variability* as

$$\mathscr{V} = \lim_{n\to\infty} n\mathsf{Var}\left(\bar{X}_n^*\right) = v + \theta, \tag{4}$$

which captures both the stochastic variability from the noises and intrinsic variability from the signals. Besides θ, there are other measures of variability, e.g., $\theta_q = \int_0^1 |\mu(t) - \bar{\mu}|^q \, \mathrm{d}t$, where $q \geq 1$. However, $\theta = \theta_2$ as defined in (3) has a more direct connection to the asymptotic variance than the other measures in the class $\{\theta_q : q \geq 1\}$, and is chosen to measure the variability of mean in the literature; see, e.g., Dalla, Giraitis and Phillips (2015). The aggregated variability $\mathscr{V}$ serves as a unified measure for the variability of the time series, and our goal is to separate the SV from the aggregated variability. It will be seen shortly in Section 3.2 that our parameter of interest, θ, can be readily estimated through the use of differencing.

3.2. Difference-based estimator of θ

3.2.1. Motivation

Our target is to find a good estimator for θ. As θ is a measure of the second-order variability of the mean function, it is natural to use second-order moments of the data to estimate θ. One possible choice is the sample autocovariance at lag $k \in \mathbb{Z}\cap(-n,n)$, i.e., $\widehat{\gamma}_k = \sum_{i=|k|+1}^n (X_i - \bar{X}_n)(X_{i-|k|} - \bar{X}_n)/n$. If $\mu(\cdot)$ is a constant function, then $\widehat{\gamma}_k \overset{\text{pr}}{\to} \gamma_k$ under standard regularity conditions; see, e.g., Brockwell and Davis (1991). However, $\mu(\cdot)$ is a non-constant function. The estimator $\widehat{\gamma}_k$ is biased upward by an amount θ, i.e., $\widehat{\gamma}_k \overset{\text{pr}}{\to} \gamma_k + \theta$; see Proposition 3.1. This typically unwanted bias turns out to be a useful feature in our context because θ is our target estimand. It remains to find a consistent estimator $\widehat{\gamma}_k^{(m)}$ (say) for the nuisance parameter γ_k, so that $\widehat{\gamma}_k - \widehat{\gamma}_k^{(m)}$ is consistent for θ. We propose to construct $\widehat{\gamma}_k^{(m)}$ by using the differencing technique introduced below. Define, for $m \geq 1$, a sequence $\{d_j \in \mathbb{R}\}_{j=0}^m$ is said to be an mth order difference sequence if $\sum_{j=0}^m d_j = 0$. For $m = 0$, denote $d_0 = 1$ as the 0th order difference sequence. Define

$$\Delta_m = \sum_{|k|\leq m} \delta_k^2, \qquad \text{where} \qquad \delta_k = \sum_{j=0}^{m-|k|} d_j d_{j+|k|}.$$

For any lag $h \in \mathbb{N}$, the mth order difference statistics $\{D_i\}_{i=mh+1}^n$ are defined as

$$D_i = \sum_{j=0}^m d_j \left(X_{i-jh} - \bar{X}_n\right) = \begin{cases} X_i - \bar{X}_n, & \text{if } m = 0; \\ \sum_{j=0}^m d_j X_{i-jh}, & \text{if } m > 0. \end{cases} \tag{5}$$

If $\delta_0 = 1$, then $\{d_j\}_{j=0}^m$ and $\{D_i\}_{i=mh+1}^n$ are said to be normalized. We assume $\delta_0 = 1$ throughout the paper. Define a potential estimator of γ_k as

$$\widehat{\gamma}_k^{(m)} = \frac{1}{n} \sum_{i=mh+|k|+1}^{n} D_i D_{i-|k|}, \qquad m \geq 0. \tag{6}$$

When $m = 0$, $\widehat{\gamma}_k^{(0)} = \widehat{\gamma}_k$ reduces to the usual sample autocovariance at lag k. When $m > 0$, the difference statistics $\{D_i\}_{i=mh+1}^n$ are approximately centered, i.e., $\mathsf{E}(D_i) \approx 0$. Thus, $\widehat{\gamma}_k^{(m)}$ is expected to be consistent for γ_k. The asymptotic properties of $\widehat{\gamma}_k^{(m)}, m \geq 0$ are provided in Proposition 3.1.

Proposition 3.1. *Let $m \geq 0$ be fixed, $|k| \lesssim \ell$ and $h \asymp \ell$, where ℓ satisfies $1/\ell + \ell/n = o(1)$. Let Assumptions 2.2 and 2.3 hold. Then, as $n \to \infty$,*

$$\mathsf{E}\left(\widehat{\gamma}_k^{(m)}\right) = \gamma_k + \theta \mathbb{1}(m = 0) + R_{1,1}, \qquad \mathsf{Var}\left(\widehat{\gamma}_k^{(m)}\right) = R_{1,2},$$

where

$$\begin{aligned} R_{1,1} &= -\frac{|k| + mh}{n}\gamma_k + \left(1 - \frac{|k| + mh}{n}\right) \sum_{s=1}^{m} \delta_s \left(\gamma_{k+sh} + \gamma_{k-sh}\right) \\ &\quad + O\left[\frac{\ell}{n}\left\{(\mathscr{C} + \mathcal{S}\mathcal{J})^2 \mathbb{1}(m = 0) + \frac{\mathscr{C}^2 \ell}{n} + \mathcal{S}^2 \mathcal{J}\right\}\right], \\ R_{1,2} &= \begin{cases} O\{(\mathscr{C} + \mathcal{S}\mathcal{J} + 1)^2/n\}, & \text{if } m = 0; \\ O\{(\mathscr{C}\ell/n + \mathcal{S} + 1)^2/n\}, & \text{if } m > 0. \end{cases} \end{aligned}$$

From Proposition 3.1, the mean squared error (MSE) of $\widehat{\gamma}_k^{(m)}$ is given by

$$\mathsf{E}\left[\widehat{\gamma}_k^{(m)} - \{\gamma_k + \theta \mathbb{1}(m = 0)\}\right]^2 = R_{1,1}^2 + R_{1,2},$$

which implies that $\widehat{\gamma}_k^{(m)} \xrightarrow{\text{pr}} \gamma_k + \theta \mathbb{1}(m = 0)$ if $\mathscr{C} + \mathcal{S}\mathcal{J} = o(\sqrt{n/\ell})$. Based on Proposition 3.1, we can construct a preliminary estimator for θ as

$$\widehat{\theta}_k^{(m)} = \widehat{\gamma}_k^{(0)} - \widehat{\gamma}_k^{(m)}, \qquad m \geq 1. \tag{7}$$

Under the conditions in Proposition 3.1, and $\mathscr{C} + \mathcal{S}\mathcal{J} = o(\sqrt{n/\ell})$, we have $\widehat{\theta}_k^{(m)} \xrightarrow{\text{pr}} \theta$.

3.2.2. *Proposed estimator of θ*

In Section 3.2.1, we can construct many preliminary estimators of θ, e.g., $\widehat{\theta}_0^{(m)}, \widehat{\theta}_1^{(m)}, \ldots, \widehat{\theta}_\ell^{(m)}$ for some $m \geq 1$. However, using any single $\widehat{\theta}_k^{(m)}$ is inefficient. It motivates us to aggregate them to obtain a better estimator for θ. We propose a kernel average estimator for θ, which is defined as

$$\widehat{\theta}^{(m)} = \frac{\sum_{|k| \leq \ell} K(k/\ell) \widehat{\theta}_k^{(m)}}{\sum_{|k| \leq \ell} K(k/\ell)}, \tag{8}$$

where $\ell \in \mathbb{N} \cap [1,n)$, and $K : \mathbb{R} \to \mathbb{R}$ is a kernel function such that $K(1) = 0$, $K(t) = K(-t)$ for all $t \in \mathbb{R}$ and $K(t) = 0$ for $|t| > 1$. Denote $\kappa = \int_{-1}^{1} K(t)\,\mathrm{d}t$ and $\kappa_\ell = \sum_{|k|\le \ell} K(k/\ell)/\ell$. If the rectangular kernel $K(t) = \mathbb{1}(|t| \le 1)$ is used, then $\widehat{\theta}^{(m)}$ reduces to the simple average of $\{\widehat{\theta}_k^{(m)} : k = 0, \pm 1, \ldots, \pm \ell\}$.

Now, we derive the asymptotic MSE of $\widehat{\theta}^{(m)}$ for $m \ge 1$. In our asymptotic theories, we require the following assumption.

Assumption 3.1. The kernel $K(\cdot)$ satisfies that (i) $t \mapsto K(t)$ is Lipschitz continuous on $(-1,1)$, and (ii) there exist $q \in \mathbb{N}$ and $B \neq 0$ such that $\lim_{t\downarrow 0}\{K(t) - K(0)\}/|t|^q = B$.

Theorem 3.2. *Let $m \ge 1$ be fixed, $1/\ell + \ell/n = o(1)$ and $h/\ell =: \lambda \in [2,\infty)$. Let Assumptions 2.2, 2.3 and 3.1 hold and $u_q := \sum_{k\in\mathbb{Z}} |k|^q |\gamma_k| < \infty$ for some $q \in \mathbb{N}$. Denote $A = \int_0^1 K^2(u)\,\mathrm{d}u$ and $\mathscr{M} = \mathscr{C}\ell^{1/2} + \mathscr{C}^2\mathscr{J}\ell^2/n + \mathscr{S}\mathscr{J}\ell^{1/2} + \mathscr{S}^2\mathscr{J}^3\ell^2/n$. Then, as $n \to \infty$,*

$$\mathsf{E}\left(\widehat{\theta}^{(m)}\right) - \theta = O\left\{\frac{(\mathscr{C} + \mathscr{S}\mathscr{J})^2}{n} + \frac{\ell}{n}\left(\frac{\mathscr{C}^2\ell}{n} + \mathscr{S}^2\mathscr{J}\right)\right\} + o\left(\frac{1}{\ell^{q+1}}\right),$$

$$\mathsf{Var}\left(\widehat{\theta}^{(m)}\right) = \frac{4v^2 A\,(\Delta_m - \Delta_0)}{n\ell\kappa_\ell^2} + \frac{4v\theta}{n} + R_2,$$

where R_2 satisfies the following properties: (i) $R_2 = 0$ if $\theta = 0$; (ii) $R_2 = o\{1/(n\ell)\}$ if $\mathscr{M} = o(1)$; and (iii) $R_2 = o(1/n)$ if $\mathscr{M}/\ell = o(1)$.

Implications and remarks of Theorem 3.2 are stated as follows:

1. The convergence rate of $\mathsf{Var}(\widehat{\theta}^{(m)})$ depends on whether $\theta = 0$. If $\theta > 0$, then $\mathsf{Var}(\widehat{\theta}^{(m)}) \sim 4v\theta\kappa^2/(n\kappa_\ell^2) \asymp \theta/n$ provided that $\mathscr{C}^2 = o\{\min(n^{1/2}, n^{3/2}/\ell^2)\}$, $\mathscr{S}^2\mathscr{J} = o(n^{1/2}/\ell)$ and $\mathscr{S}^2\mathscr{J}^2 = o(n^{1/2})$. On the other hand, if $\theta = 0$, then $\mathsf{Var}(\widehat{\theta}^{(m)}) \sim 4v^2 A(\Delta_m - \Delta_0)/(n\ell\kappa_\ell^2) \asymp 1/(n\ell) \ll 1/n$. It suggests that $\widehat{\theta}^{(m)}$ achieves super-efficiency at $\theta = 0$. This phenomenon is similar to the well-known Hodge's estimator. We will show that the super-efficiency achieved by $\widehat{\theta}^{(m)}$ is not a pathological property but an important attribute for our proposed mean invariance test in Section 4.
2. To achieve super-efficiency at $\theta = 0$ in terms of MSE, the squared bias of $\widehat{\theta}^{(m)}$ must be negligible relative to the variance, i.e., $\mathsf{Bias}(\widehat{\theta}^{(m)};\theta)^2 \ll \mathsf{Var}(\widehat{\theta}^{(m)})$. In this case, we need to select ℓ such that $n^{1/(1+2q)} \lesssim \ell \ll n$. Since the order of magnitude of $\mathsf{Bias}(\widehat{\theta}^{(m)};\theta)$ increases with ℓ when $\theta \neq 0$, we recommend using the minimal order of ℓ so that the squared bias is negligible for the largest class of mean function. In practice, we suggest choosing $\ell = \lfloor \varphi n^{1/(1+2q)} \rfloor$ for some $\varphi \in \mathbb{R}$ for the construction of $\widehat{T}_{\mathsf{MIT}}^{(m)}$. According to our simulation experiments, choosing $\varphi = 1$ yields good results. By default, we set $\varphi = 1$.
3. By introducing the parametrizations $\mathscr{C} \asymp n^{\mathscr{c}}, \mathscr{S} \asymp n^{\mathscr{s}}$ and $\mathscr{J} \asymp n^{\mathscr{j}}$, where $\mathscr{c}, \mathscr{s} \in \mathbb{R}$ and $\mathscr{j} \in [0,1]$, we illustrate the allowable region for $\mathscr{C}$, $\mathscr{S}$ and $\mathscr{J}$ when our recommended value $\ell = O(n^{1/(1+2q)})$ is used. In particular, consider $q = 2$. To achieve $R_2 = o\{1/(n\ell)\}$, we need to have (i) $2\mathscr{c} + \mathscr{j} < 3/5$, (ii) $\mathscr{c} < -1/10$, (iii) $2\mathscr{s} + 3\mathscr{j} < 3/5$, and (iv) $\mathscr{s} + \mathscr{j} < -1/10$. On the other hand, to achieve $R_2 = o(1/n)$, we need to have (i) $2\mathscr{c} + \mathscr{j} < 4/5$, (ii) $\mathscr{c} < 1/10$, (iii) $2\mathscr{s} + 3\mathscr{j} < 4/5$, and (iv) $\mathscr{s} + \mathscr{j} < 1/10$. In particular, if $\mathscr{C} + \mathscr{S} = O(1)$, then we allow $\mathscr{J} = o(n^{1/10})$ for $R_2 = o(1/n)$.
4. The theorem holds as long as $\lambda \ge 2$. In this case, the bias and variance admit neat and nice forms. Considering the finite sample performance, we recommend using $\lambda = 2$ in practice as it leads to the best differencing effect. If $\lambda < 2$, then $\mathsf{Var}(\widehat{\theta}^{(m)})$ depends on $\{d_j\}_{j=0}^m$ not only through Δ_m but also a more complicated function of $\{d_j\}_{j=0}^m$. Consequently, a case-by-case derivation is needed.

5. One class of kernels that satisfies Assumption 3.1 is $K(t) = (1 - |t|^q)\mathbb{1}(|t| \leq 1)$, where $q \in \mathbb{N}$ is the characteristic exponent; see Parzen (1957) and Chan and Yau (2017). In particular, it reduces to the well-known Bartlett kernel when $q = 1$. See also Vats and Flegal (2022), Chan and Yau (2023), and Liu and Chan (2023) for more kernel choices.
6. To implement the proposed estimator, we suggest choosing $\lambda = 2$, $q = 2$, $K(t) = (1-|t|^2)\mathbb{1}(|t| \leq 1)$, $\varphi = 1$ and $\ell = \lfloor n^{1/5} \rfloor$. The choices of the differencing order m and the optimal difference sequence $\{d_j\}_{j=0}^m$ can be found in Section 5.1. Section 6.4 presents some sensitivity analyses.

Example 3.1. To illustrate, we compare the preliminary estimator with our proposed kernel average estimator using different m. We generate the noises $\{Z_i\}_{i=1}^n$ from an autoregressive moving average (ARMA) model:

$$Z_i = 0.5Z_{i-1} + 0.5\varepsilon_{i-1} + \varepsilon_i,$$

where $\varepsilon_i \overset{\text{IID}}{\sim} \mathrm{N}(0,1)$. We consider the mean function $\mu(t) = \Delta\cos(5\pi t)$ with different Δ so that $\theta = 0, 0.001, 0.005, 0.01$, and generate the time series $\{X_i\}_{i=1}^n$ as in (1). The left panel of Figure 1 shows the estimation performances of the preliminary estimator $\widehat{\theta}_0^{(1)}$ and the proposed estimators $\widehat{\theta}^{(m)}$ for $m = 1,2,3$. We see that the results are in line with Proposition 3.1 and Theorem 3.2. The proposed estimators can achieve a convergence rate faster than $\sqrt{n}$, the convergence rate of preliminary estimator, in estimation of θ. This shows that the proposed estimators are super-efficient relative to the preliminary estimators.

4. A new test for mean invariance

4.1. Test for mean invariance

Suppose we want to test $H_0 : \theta = 0$, which is equivalent to H_0 : "$\mu(\cdot)$ is a constant function", against $H_1 : \theta > 0$. We propose a test for mean invariance based on the estimator $\widehat{\theta}^{(m)}$ because it achieves super-efficiency under H_0. Theorem 4.1 concerns the asymptotic null distribution of $\widehat{\theta}^{(m)}$.

Theorem 4.1. *Let $m \geq 1$ be fixed, $u_q < \infty$ for some $q \in \mathbb{N}$, $n^{1/(1+2q)} \lesssim \ell \ll n$ and $h/\ell = \lambda \in [2,\infty)$. Let Assumptions 2.2, 2.3 and 3.1 hold. If $\theta = 0$, then, as $n \to \infty$,*

$$\sqrt{n\ell\kappa_\ell^2}\left(\widehat{\theta}^{(m)} - 0\right) \xrightarrow{\mathrm{d}} \mathrm{N}\left(0, 4A(\Delta_m - \Delta_0)\nu^2\right).$$

Based on Theorem 4.1, we can construct a mean invariance test (MIT) statistic $\widehat{T}_{\mathsf{MIT}}^{(m)}$ for testing $H_0 : \theta = 0$ as follows:

$$\widehat{T}_{\mathsf{MIT}}^{(m)} = \frac{\widehat{\theta}^{(m)}}{\sqrt{\left\{4\left(\widehat{\nu}^{(m)}\right)^2 A(\Delta_m - \Delta_0)\right\} / n\ell\kappa_\ell^2}}, \tag{9}$$

where $\widehat{\nu}^{(m)} = \sum_{|k|\leq\ell} K(k/\ell)\widehat{\gamma}_k^{(m)}$ is a consistent estimator of ν under both H_0 and H_1; see Chan (2022a). Also see Casini and Perron (2021), Casini (2023) and Chan (2022b) for some alternative estimators of ν. Remark 4.2 discusses using $\widehat{\theta}^{(m)}$ for testing $\theta \leq \theta_0$ for some $\theta_0 > 0$.

One important observation is that the test statistic $\widehat{T}_{\mathsf{MIT}}^{(m)}$ in (9) does not utilize the always correct standard error $\{\mathsf{Var}(\widehat{\theta}^{(m)})\}^{1/2}$ or its estimator for standardization. Instead, it uses a standardizer that

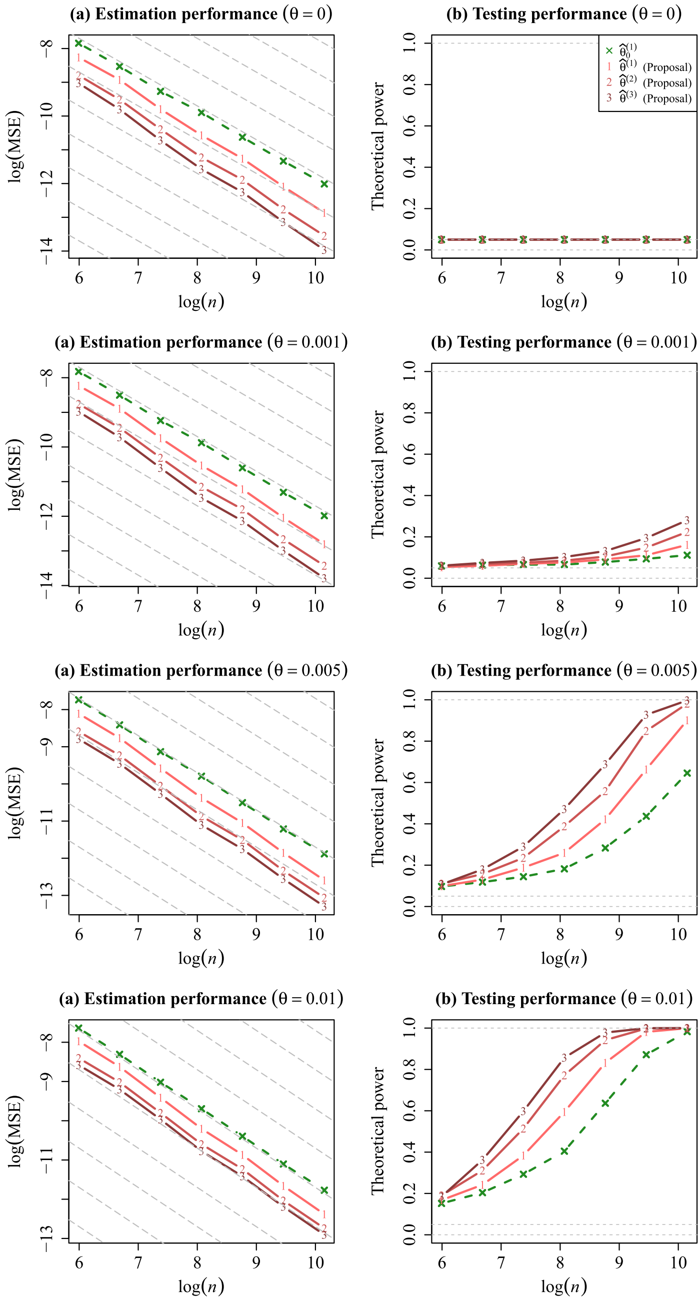

Figure 1. (a) Estimation performance (Example 3.1) and (b) testing performance (Example 4.1) are shown in the first and second columns of plots, respectively, where the preliminary estimator $\widehat{\theta}_0^{(1)}$ and the proposed estimators $\widehat{\theta}^{(m)}$ with $m \in \{1,2,3\}$ are computed. In the first column of plots, the gray dashed lines represent reference lines of slope -1 designating $\sqrt{n}$-consistent estimators. The lines that are steeper than the reference lines correspond to super-efficient estimators.

is only correct under H_0. Interestingly, it leads to a valid and more powerful test; see Remark 4.1 for details. The asymptotic validity and consistency of the proposed test statistic $\widehat{T}_{\mathsf{MIT}}^{(m)}$ are formally stated in Corollary 4.2.

Corollary 4.2. *Let $m \geq 1$ be fixed, $u_q < \infty$ for some $q \in \mathbb{N}$, $n^{1/(1+2q)} \lesssim \ell \ll n$ and $h/\ell = \lambda \in [2, \infty)$. Let Assumptions 2.2, 2.3 and 3.1 hold.*

1. *Under H_0, $\widehat{T}_{\mathsf{MIT}}^{(m)} \xrightarrow{\mathrm{d}} \mathrm{N}(0,1)$ as $n \to \infty$.*
2. *Under H_1, $\widehat{T}_{\mathsf{MIT}}^{(m)} \xrightarrow{\mathrm{pr}} \infty$ as $n \to \infty$ for any fixed $\theta > 0$.*

Corollary 4.2 provides theoretical justification of our proposed mean invariance test. The most appealing feature of our proposed test is that under H_0, it converges to a standard normal distribution at a rate of $\sqrt{n\ell}$ shown in Theorem 4.1. The accelerated rate of convergence makes our proposed test stands out from tests based on $\sqrt{n}$-consistent estimator of θ. Example 4.1 below illustrates this phenomenon.

Example 4.1. We revisit the experiments in Example 3.1. In this example, we compute the power functions of the tests $\widehat{T}_{\mathsf{MIT}}^{(m)}$ based on the estimators $\widehat{\theta}^{(m)}$, for $m = 1, 2, 3$, We also compare it with the test based on the preliminary estimator $\widehat{\theta}_0^{(1)}$, i.e., we reject H_0 if $\widehat{\theta}_0^{(1)} > c_n$, where the critical value c_n satisfies $\mathsf{P}(\widehat{\theta}_0^{(1)} > c_n) = 0.05$ under H_0. The value c_n is obtained through simulation. The right panel of Figure 1 shows the theoretical power functions of the above tests. We see that the proposed tests perform significantly better than the preliminary test, and the powers of the proposed tests increase with m.

Remark 4.1. From Theorem 3.2, a naive test statistic is

$$\breve{T}_{\mathsf{MIT}}^{(m)} = \frac{\widehat{\theta}^{(m)} - 0}{\sqrt{\mathsf{Var}(\widehat{\theta}^{(m)})}} = \frac{\widehat{\theta}^{(m)}}{\sqrt{4(\widehat{\nu}^{(m)})^2 A(\Delta_m - \Delta_0)/(n\ell\kappa_\ell^2) + 4\widehat{\nu}^{(m)}\widehat{\theta}^{(m)}\kappa^2/(n\kappa_\ell^2)}},$$

which always utilizes the true value of $\{\mathsf{Var}(\widehat{\theta}^{(m)})\}^{1/2}$ for standardization no matter under H_0 or H_1. Note that both $\widehat{T}_{\mathsf{MIT}}^{(m)}$ and $\breve{T}_{\mathsf{MIT}}^{(m)}$ are asymptotically $\mathrm{N}(0,1)$ under H_0, and their normalizing constants are approximately the same since $\widehat{\theta}^{(m)}$ is close to 0. However, under H_1, since $\widehat{\theta}^{(m)} \to \theta > 0$, the normalizing constant of $\breve{T}_{\mathsf{MIT}}^{(m)}$ increases with $\widehat{\theta}^{(m)}$, while that of $\widehat{T}_{\mathsf{MIT}}^{(m)}$ remains the minimal value. So, $\widehat{T}_{\mathsf{MIT}}^{(m)}$ has the advantage of achieving higher power under H_1 over $\breve{T}_{\mathsf{MIT}}^{(m)}$. Thus $\widehat{T}_{\mathsf{MIT}}^{(m)}$ is suggested for testing $H_0 : \theta = 0$ against $H_1 : \theta > 0$.

Remark 4.2. Sometimes, we may allow certain amount of fluctuations of the underlying mean function, and test whether the mean is practically non-constant:

$$H_0' : \theta \leq \theta_0 \qquad \text{against} \qquad H_1' : \theta > \theta_0, \tag{10}$$

where $\theta_0 > 0$ is a pre-specified null threshold. We call this type of test relevant mean varying test. However, Theorem 4.1 cannot be applied in this case due to the existence of an additional non-negligible term involved in the asymptotic variance which has a larger order and is independent of m when $\mu(\cdot)$ is non-constant; see Theorem 3.2. A similar limiting result holds under Assumptions 2.2, 2.3 and 3.1,

$m \geq 1$ is fixed, $u_q < \infty$ for some $q \in \mathbb{N}$, $n^{1/(2+2q)} \lesssim \ell \ll \sqrt{n}$ and $h/\ell = \lambda \in [2,\infty)$. Specifically, when $\theta_0 \in \mathbb{R}^+$ is the true SV of $\mu(\cdot)$, we have

$$\sqrt{n\kappa_\ell^2}\left(\widehat{\theta}^{(m)} - \theta_0\right) \xrightarrow{\mathrm{d}} \mathrm{N}\left(0, 4\nu\theta_0\kappa^2\right), \tag{11}$$

as $n \to \infty$. However, since specifying θ_0 can be practically difficult, we recommend the tests in Sections 4.3.2 and 4.3.3 for testing the relevant structural changes.

4.2. Power under local alternatives

Consider a sequence of local alternatives given by $H_1^{(c)} : \theta = t_0/\ell^c$ for some $t_0, c \in \mathbb{R}^+$. Theorem 4.3 shows that the proposed test admits different forms of convergence under different regimes.

Theorem 4.3. *Let Assumptions 2.2, 2.3 and 3.1 hold. Let $m \geq 1$ be fixed and $h/\ell = \lambda \in [2,\infty)$. Consider a sequence of local alternatives in the form of $H_1^{(c)}$.*

1. a) *If $c > 1$, $u_q < \infty$ for some $q \in \{2,3,4,\ldots\}$ and $1/\ell + \ell/n^{1/3} = o(1)$, then as $n \to \infty$,*

$$\sqrt{n\ell\kappa_\ell^2}\left(\widehat{\theta}^{(m)} - \frac{t_0}{\ell^c}\right) \xrightarrow{\mathrm{d}} \mathrm{N}\left(0, 4A\left(\Delta_m - \Delta_0\right)\nu^2\right).$$

   b) *If $0 < c < 1$, $u_q < \infty$ for some $q \in \mathbb{N}$ and $1/\ell + \ell/n^{1/(2+c)} = o(1)$, then as $n \to \infty$,*

$$\sqrt{n\ell^c\kappa_\ell^2}\left(\widehat{\theta}^{(m)} - \frac{t_0}{\ell^c}\right) \xrightarrow{\mathrm{d}} \mathrm{N}\left(0, 4\nu t_0\kappa^2\right).$$

2. *Non-trivial local limiting power is achieved when $\ell^c = \sqrt{n\ell}$, i.e., $\theta = t_0/\sqrt{n\ell}$. In particular, if the nominal size is $\alpha \in (0,1)$, then the local power function is given by*

$$\begin{aligned}\pi_c(t_0) &:= \lim_{n\to\infty} \mathsf{P}_{\theta=t_0/\ell^c}\left(\widehat{T}_{\mathsf{MIT}}^{(m)} > \Phi^{-1}(1-\alpha)\right) \\ &= 1 - \Phi\left(\Phi^{-1}(1-\alpha) - \frac{t_0\kappa}{\sqrt{4A\left(\Delta_m - \Delta_0\right)\nu^2}}\right),\end{aligned}$$

   where $\Phi(\cdot)$ is the distribution function of $\mathrm{N}(0,1)$, and $\mathsf{P}_{\theta=\theta_1}$ denotes the probability measure when $\theta = \theta_1$.

Theorem 4.3 traces the change in convergence across different local alternatives. The difference in the requirement of ℓ arises from controlling the convergence rate of the θ. It suggests that our test achieves $\sqrt{n\ell}$-convergence not only for $H_0 : \theta = 0$, but also for a sequence of local alternatives in the form of $H_1^{(c)}$ when $c > 1$. Moreover, the result is consistent with (11) in the sense that as $c \downarrow 0$, it reduces back to the case of non-local alternative. On the other hand, the case for $c = 1$ is more complicated. We remark that $\sqrt{n\ell\kappa_\ell^2}(\widehat{\theta}^{(m)} - t_0/\ell) = O_p(1)$; see Section A.5 in the Supplementary Material (To and Chan, 2024) for details. Using this property together with Theorem 4.3 (1b), we can show that the local power function is trivial, i.e., $\pi_c(t_0) = 1$, for any $0 < c \leq 1$. Consequently, only the case $c > 1$ leads to a non-trivial local limiting power.

4.2.1. Detectable regions

Recall that the test statistics for the KPSS test (Kwiatkowski et al., 1992) and KS test (Csörgő and Horváth, 1997) are

$$\widetilde{T}_{\mathsf{KPSS}} = \frac{1}{vn^2}\sum_{k=1}^{n}\left\{\sum_{i=1}^{k}\left(X_i - \bar{X}_n\right)\right\}^2 \qquad \text{and} \qquad \widetilde{T}_{\mathsf{KS}} = \sup_{1\le k\le n}\left|\frac{1}{\sqrt{vn}}\sum_{i=1}^{k}\left(X_i - \bar{X}_n\right)\right|,$$

respectively. Here we assume v is known for simplicity. Also define the known-v version of our proposed MIT statistic by

$$\widetilde{T}^{(m)}_{\mathsf{MIT}} = \frac{\widehat{\theta}^{(m)}}{\sqrt{\left\{4v^2 A\left(\Delta_m - \Delta_0\right)\right\}/n\ell\kappa_\ell^2}}.$$

Consider the class of mean function of the form

$$\mu(t) = \sum_{j=1}^{N} g_j\left(\frac{t-\tau_j}{L_j}\right) S_j, \tag{12}$$

where N is the number of segments, S_j measures the amplitude of each segment, $1 \equiv \tau_1 < \tau_2 < \cdots < \tau_N < \tau_{N+1} \equiv n$ are the endpoints of each segment, $L_j = \tau_{j+1} - \tau_j$, and $g_1, \ldots, g_N$, are some template mean functions that are of bounded variation and satisfy $g_j(t) = 0$ for all $t \notin [0,1]$,

$$\int_0^1 g_j(t)\,\mathrm{d}t = 0, \qquad \int_0^s g_j(t)\mathrm{d}t \neq 0, \qquad \text{and} \qquad \int_0^1 g_j^2(t)\,\mathrm{d}t = 1, \tag{13}$$

for all $s \in (0,1)$ and $j = 1, \ldots, N$. Define $S = \max_{j=1,\ldots,N}\left|S_j\right|$ and $L = \max_{j=1,\ldots,N} L_j$. The class of mean functions is defined in a way so each segment has zero mean and unit signal variance for standardization. We allow $N, S_1, \ldots, S_N, L_1, \ldots, L_N$ to be dependent on n, but the functions $g_1, \ldots, g_N$ are independent of n. By construction, this class of mean function "oscillates" for N times, and when N is large, the oscillation effect will be masked by the noises. So, it is challenging. Figure 2 provides an example of such mean function.

Proposition 4.4. *Let the conditions in Theorem 4.1 hold, and $\mu(\cdot)$ admits the form in (12). Then, as $n \longrightarrow \infty$,*

$$\widetilde{T}^{(m)}_{\mathsf{MIT}} = C\sqrt{n\ell}\sum_{j=1}^{N} S_j^2 L_j \int_0^1 g_j^2(s)\,\mathrm{d}s + O\left(\frac{\sqrt{\ell}}{n^{3/2}}N^2S^2\right) + O_p\left(\frac{\sqrt{\ell}}{n}NS + 1\right),$$

$$\widetilde{T}_{\mathsf{KPSS}} = A_{\mathsf{KPSS}} + O_p\left(\sqrt{A_{\mathsf{KPSS}}} + 1\right),$$

$$A_{\mathsf{KS}} \le \widetilde{T}^2_{\mathsf{KS}} \le A_{\mathsf{KS}} + O_p\left(\sqrt{A_{\mathsf{KS}}} + 1\right),$$

where $C = \kappa_\ell/\sqrt{4v^2A\left(\Delta_m - \Delta_0\right)}$, and

$$A_{\mathsf{KPSS}} = n\sum_{j=1}^{N} S_j^2 L_j^3 \int_0^1 \left\{\int_0^t g_j(s)\,\mathrm{d}s\right\}^2 \mathrm{d}t,$$

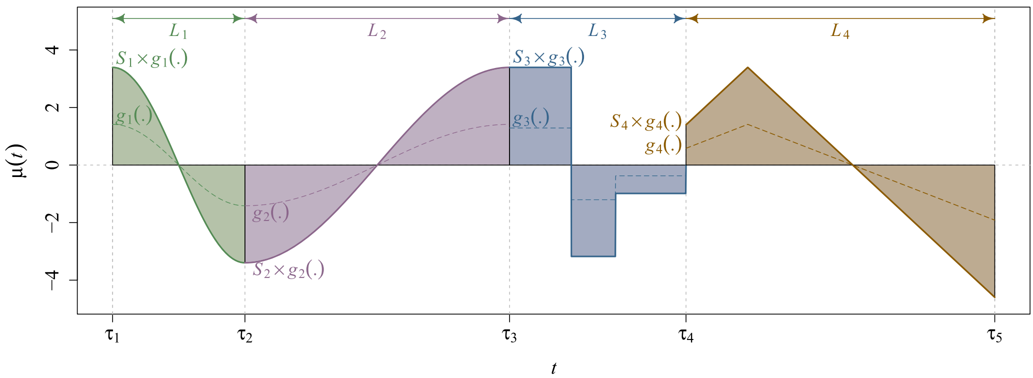

Figure 2. Illustration of $\mu(t)$ satisfying (12) with $N = 4$ segments. The dashed lines denote the template mean functions g_j and the solid lines denote the mean functions after scaling by their corresponding S_j for $j = 1,2,3,4$.

$$A_{\mathsf{KS}} = n \sup_{1 \le j \le N} S_j^2 L_j^2 \sup_{0 \le t \le 1} \left\{ \int_0^t g_j(s)\, \mathrm{d}s \right\}^2 .$$

To illustrate the results in Proposition 4.4, we consider Example 4.2 below.

Example 4.2. Consider $q = 2$ and $\ell = O(n^{1/(1+2q)}) = O(n^{1/5})$. We study the following three cases.

1. Case 1: Sparse and weak signals. Let $|S_j| = S\mathbb{1}(j \le K)$ and $L_j = 1/\sqrt{n}$ for all $j = 1, \dots, N$.
2. Case 2: Sparse and short signals. Let $|S_j| = \mathbb{1}(j \le K)$ and $L_j = L$ for all $j = 1, \dots, N$.
3. Case 3: Weak and short signals. Let $|S_j| = S\mathbb{1}(j = 1)$ and $L_j = L$ for all $j = 1, \dots, N$.

If we introduce the parametrizations $S = n^\psi$, $L = n^\zeta$, $N = n^\nu$ and $K = n^\xi$, where $\psi \in \mathbb{R}$, $\zeta \in (-1, 0)$ and $\nu, \xi \in (0, 1)$, then we can find out the detectable regions, i.e., regions where the power of the test tends to 1 as $n \to \infty$, for non-stationarity by the tests using Proposition 4.4. Figure 3 visualizes the results. We observe that the proposed test performs especially promisingly when the number of oscillations K is large; see cases 1–2. It also demonstrates that the proposed test is able to account for hardly noticeable variation that could possibly tend to 0 when n is large. On the other hand, for case 3, the signal only

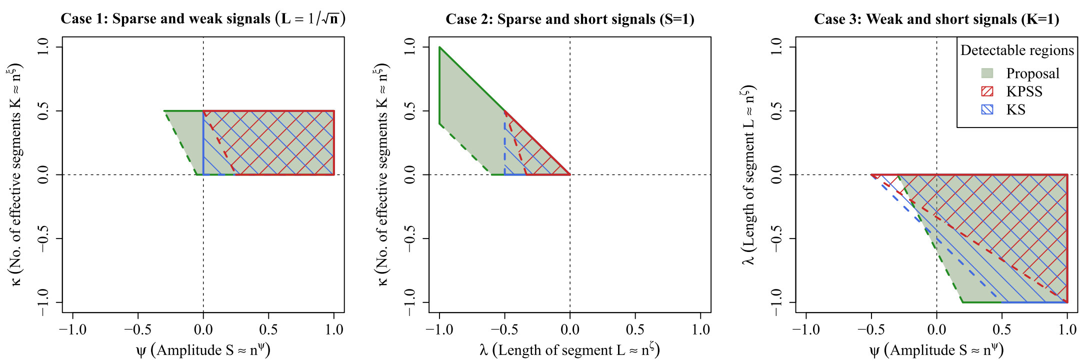

Figure 3. The detectable non-stationarity regions of the proposed test, KPSS test and KS test under each case.

consists of one oscillation. Our proposed test performs worse than the KS and KPSS tests when the signal is long and weak, i.e., large L and small S. This scenario is similar to the one change point situation and favors the KS and KPSS tests as they are specifically designed for handling such mean structure. In a nutshell, the proposed test is more versatile than the KPSS test and KS test in most cases although none of them uniformly dominates the other two tests.

4.3. Generalizations

4.3.1. Test for smooth mean structure

Let $s \in \mathbb{N}_0$ be fixed. Suppose we want to test

$$H_0^{(s)} : \exists a_0, \ldots, a_s \in \mathbb{R} \qquad \text{such that} \qquad \mu(t) = \sum_{j=0}^{s} a_j t^j, \qquad t \in [0,1], \tag{14}$$

against $H_1^{(s)}$: there do not exist $a_0, \ldots, a_s \in \mathbb{R}$ such that (14) holds. Let

$$\theta^{(s)} = \int_0^1 \left\{ \mu(t) - \left(a_0^\star + a_1^\star t + \cdots + a_s^\star t^s \right) \right\}^2 \, \mathrm{d}t,$$

where

$$(a_0^\star, \ldots, a_s^\star) = \underset{(a_0,\ldots,a_s)\in\mathbb{R}^{s+1}}{\arg\min} \int_0^1 \{\mu(t) - (a_0 + a_1 t + \cdots + a_s t^s)\}^2 \, \mathrm{d}t.$$

The quantity $\theta^{(s)}$ measures the signal variance after removing the best fitted sth order polynomial trend. Hence, $H_0^{(s)}$ is true if and only if $\theta^{(s)} = 0$. We can estimate $\mu(t)$ by $\widehat{\mu}_s(t) = \sum_{j=0}^{s} \widehat{a}_j t^j$, where $\widehat{a}_j$ is the least squares estimator of $a_j^\star$ for each $j = 0, \ldots, s$; see (A.24) of the Supplementary Material (To and Chan, 2024) for a detailed formula. To test whether the true mean function is a sth order polynomial, we apply our proposed estimator using the residuals

$$\widehat{Z}_i = X_i - \sum_{j=0}^{s} \widehat{a}_j \left(\frac{i}{n} \right)^j, \qquad i = 1, \ldots, n.$$

Let $D_i^{(m,s)}$ be defined as in (5) by replacing X_i with $\widehat{Z}_i$. Denote

$$\widehat{\gamma}_k^{(m,s)} = \frac{1}{n} \sum_{i=mh+|k|+1}^{n} D_i^{(m,s)} D_{i-|k|}^{(m,s)}, \qquad m \geq 0.$$

For $m \geq 1$, we define an estimator of $\theta^{(s)}$ by

$$\widehat{\theta}^{(m,s)} = \frac{\sum_{|k|\leq\ell} K(k/\ell) \widehat{\theta}_k^{(m,s)}}{\sum_{|k|\leq\ell} K(k/\ell)}, \qquad \text{where} \qquad \widehat{\theta}_k^{(m,s)} = \widehat{\gamma}_k^{(0,s)} - \widehat{\gamma}_k^{(m,s)}.$$

Clearly, when $s = 0$, it reduces back to the original kernel estimator (8), i.e., $\widehat{\theta}^{(m,0)} = \widehat{\theta}^{(m)}$. Theorem 4.5 extends the result in Theorem 4.1 by showing that the proposed test achieves super-efficiency not only for testing mean stationarity, but also for testing trend stationarity.

Theorem 4.5. *Let $m \geq 1$ be fixed, $u_q < \infty$ for some $q \in \mathbb{N}$, $n^{1/(1+2q)} \lesssim \ell \ll n$ and $h/\ell = \lambda \in [2,\infty)$. Let Assumptions 2.2, 2.3 and 3.1 hold. Denote the least squares estimators of $a_0^\star, \ldots, a_s^\star$ by $\widehat{a}_0, \ldots, \widehat{a}_s$, respectively. Under $H_0^{(s)}$ with a fixed $s \in \mathbb{N}_0$, we have, as $n \to \infty$, that*

$$\sqrt{n\ell\kappa_\ell^2}\left(\widehat{\theta}^{(m,s)} - 0\right) \xrightarrow{\mathrm{d}} \mathrm{N}\left(0, 4A\left(\Delta_m - \Delta_0\right)v^2\right).$$

According to Theorem 4.5, our proposed test statistic for testing $H_0^{(s)}$ is given by

$$\widehat{T}_{\mathsf{MIT}}^{(m,s)} = \frac{\widehat{\theta}^{(m,s)}}{\sqrt{4\left(\widehat{v}^{(m,s)}\right)^2 A\left(\Delta_m - \Delta_0\right)/(n\ell\kappa_\ell^2)}}, \tag{15}$$

where $\widehat{v}^{(m,s)} = \sum_{|k|\leq\ell} K(k/\ell)\widehat{\gamma}_k^{(m,s)}$. The test statistic $\widehat{T}_{\mathsf{MIT}}^{(m,s)}$ and $\widehat{T}_{\mathsf{MIT}}^{(m)}$ differ only from the input of data, i.e., we use the observed data $\{X_i\}_{i=1}^n$ for $\widehat{T}_{\mathsf{MIT}}^{(m)}$ while we use the residuals $\{\widehat{Z}_i\}_{i=1}^n$ for $\widehat{T}_{\mathsf{MIT}}^{(m,s)}$. Similarly, when $s = 0$, $\widehat{T}_{\mathsf{MIT}}^{(m,s)}$ reduces back to $\widehat{T}_{\mathsf{MIT}}^{(m)}$, i.e., $\widehat{T}_{\mathsf{MIT}}^{(m,0)} = \widehat{T}_{\mathsf{MIT}}^{(m)}$.

The proposed trend stationarity test can be used as a goodness of fit test for determining whether a polynomial trend model fits the observed time series dataset. We demonstrate it in a real-data example in Appendix D of the Supplementary Material (To and Chan, 2024). Besides, the proposed method can be easily generalized for testing other parametric trend models other than the polynomial null hypothesis in (14).

4.3.2. Test for relative variability

To explore the extent of relevant mean non-invariance, we consider the amount of variability that the underlying mean function contributes to the aggregated variability of the observed time series $\{X_i\}_{i=1}^n$. Based on the aggregated variability $\mathscr{V}$ defined in (4), we define the *relative variability* ϕ_{long} as follows:

$$\phi_{\mathsf{long}} = \frac{\theta}{\mathscr{V}} = \frac{\theta}{v + \theta}, \tag{16}$$

where the subscript “long” is used to emphasize the relationship of ϕ_{long} with the long-run variance v. Clearly, $\phi_{\mathsf{long}} \in [0,1]$. When $\phi_{\mathsf{long}} = 0$, we have $\theta = 0$, which means that $\mu(\cdot)$ is a constant function. When $\phi_{\mathsf{long}} = 1$, $\mu(\cdot)$ completely determines the behaviour of the time series. Recall, from (4), that the aggregated variability $\mathscr{V} = v + \theta$ can be represented as the permuted long-run variance $\mathscr{V} = \lim_{n\to\infty} n\mathsf{Var}(\bar{X}_n^*)$. Hence, one can statistically interpret $1 - \phi_{\mathsf{long}}$ as the proportion of the non-permuted long-run variance $v = \lim_{n\to} n\mathsf{Var}(\bar{X}_n)$ to the permuted counterpart $\mathscr{V}$, i.e.,

$$\frac{\mathsf{Var}(\bar{X}_n)}{\mathsf{Var}(\bar{X}_n^*)} \to 1 - \phi_{\mathsf{long}}.$$

Consequently, $\phi_{\mathsf{long}} = 5\%$ means that permuting the means contributes 5% of the overall variability in terms of the permuted long-run variance $\mathscr{V}$.

In light of this phenomenon, we construct a test for relative variability to test

$$H_0^{\mathsf{long}}: \phi_{\mathsf{long}} \leq \phi_{\mathsf{long},0} \qquad \text{against} \qquad H_1^{\mathsf{long}}: \phi_{\mathsf{long}} > \phi_{\mathsf{long},0}, \tag{17}$$

where $0 < \phi_{\mathsf{long},0} < 1$ is a prescirbed level of relative variability. In practice, it is typically more meaningful to test small values of $\phi_{\mathsf{long},0}$ when testing (17) so that the influences due to the mean function

$\mu(\cdot)$ and the idiosyncratic noise are difficult to distinguish. So, we focus on $0 < \phi_{\mathsf{long},0} \le 0.5$. An estimator for ϕ_{long} can be constructed as follows:

$$\widehat{\phi}^{(m)}_{\mathsf{long}} = \frac{\widehat{\theta}^{(m)}}{\widehat{\nu}^{(m)} + \widehat{\theta}^{(m)}}, \qquad m \ge 1. \tag{18}$$

Theorem 4.6 presents the asymptotic distribution of $\widehat{\phi}^{(m)}_{\mathsf{long}}$.

Theorem 4.6. *Let $m \ge 1$ be fixed. Assume $u_q < \infty$ for some $q \in \mathbb{N}$, $\ell = \lfloor \varphi n^{1/(2q+1)} \rfloor$ for some $\varphi \in \mathbb{R}^+$, and $h/\ell = \lambda \in [2,\infty)$. Let Assumptions 2.2 and 2.3 hold. Assume $\mathscr{C}$, $\mathscr{S}$ and $\mathscr{J}$ are fixed. If the true relative variability of the time series is $\phi_{\mathsf{long},0}$, then, as $n \to \infty$,*

$$n^{q/(1+2q)} \left(\widehat{\phi}^{(m)}_{\mathsf{long}} - \phi_{\mathsf{long},0} \right) \xrightarrow{\mathrm{d}} \mathrm{N}\left(\frac{-B\nu_q \varphi^{-q} \theta}{(\theta+\nu)^2}, 4A\Delta_m \varphi \phi^2_{\mathsf{long},0} \left(\phi_{\mathsf{long},0} - 1\right)^2 \right),$$

where $\nu_q := \sum_{k\in\mathbb{Z}} |k|^q \gamma_k$ and B is as defined in Assumption 3.1.

Based on Theorem 4.6, we construct a test statistic for testing (17) in the following way:

$$\widehat{T}^{(m)}_{\mathsf{long}} = \frac{n^{q/(1+2q)} \max\left(\widehat{\phi}^{(m)}_{\mathsf{long}} - \phi_{\mathsf{long},0}, 0 \right) + \left(B\widehat{\nu}_q \varphi^{-q} \widehat{\theta}^{(m)} \right) / \left(\widehat{\theta}^{(m)} + \widehat{\nu}^{(m)} \right)^2}{\sqrt{4A\Delta_m \varphi \phi^2_{\mathsf{long},0} \left(\phi_{\mathsf{long},0} - 1 \right)^2}}, \tag{19}$$

where $\widehat{\nu}_q = \sum_{|k|\le\ell} |k|^q \widehat{\gamma}^{(m)}_k$ is a consistent estimator of ν_q. Notice that the function $p \mapsto p^2(p-1)^2$ is monotonically increasing on $[0, 0.5]$. Hence in (19), we use $\phi_{\mathsf{long},0}$, the null value of ϕ_{long}, as the normalizing constant so that the test can achieve higher power, similar to that in Remark 4.1. In practice, H^{long}_0 is rejected at size α if $\widehat{T}^{(m)}_{\mathsf{long}} > \Phi^{-1}(1-\alpha)$. This test is asymptotically valid and consistent as $n \to \infty$.

4.3.3. Test for marginal relative variability

The relative variability defined in Section 4.3.2 concerns the long term contribution of the noise variability to the observed data. Besides the long term relative variability, we also consider the short term (marginal) contribution of the noise variability to the observed data. Define the *marginal relative variability* as

$$\phi_{\mathsf{short}} = \frac{\theta}{\theta + \gamma_0}. \tag{20}$$

Clearly, $\phi_{\mathsf{short}} \in [0,1]$ as in ϕ_{long}. Similar to Section 4.3.2, one may statistically interpret $1 - \phi_{\mathsf{short}}$ as the proportion of non-permuted marginal variance $\gamma_0 = \mathsf{Var}(X_i)$ to the permuted marginal variance $\mathsf{Var}(X^*_i) \to \theta + \gamma_0$ for any i, i.e.,

$$\frac{\mathsf{Var}(X_i)}{\mathsf{Var}(X^*_i)} \to 1 - \phi_{\mathsf{short}}.$$

Figure 4 visually compares the quantities ϕ_{long} and ϕ_{short}. It shows that ϕ_{long} and ϕ_{short} may not equal in the time series setting. Intuitively, ϕ_{long} takes all observations $X_1, \ldots, X_n$ into account and tries to describe the relative variability by the average. On the other hand, ϕ_{short} concerns the relative variability at one specific time point only.

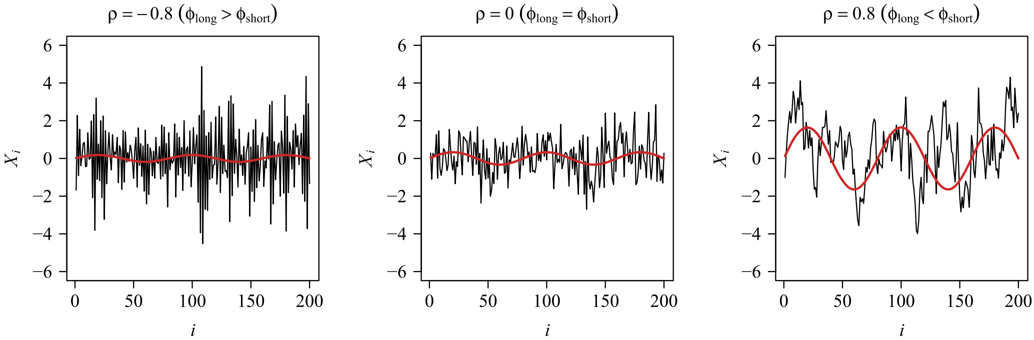

Figure 4. A realization of the time series $X_i = \mu_i + \rho(X_{i-1} - \mu_{i-1}) + \varepsilon_i$ is plotted on the left, middle and right panels for $\rho \in \{-0.8, 0, 0.8\}$, respectively, where $n = 200$, $\varepsilon_i \overset{\text{IID}}{\sim} \mathrm{N}(0,1)$ and $\mu_i = C\sin(5\pi i/n)$ for some scaling constant $C > 0$ to ensure $\phi_{\mathsf{long}} = 0.05$ in each case. When $\rho \in \{-0.8, 0, 0.8\}$, we have $\phi_{\mathsf{long}} > \phi_{\mathsf{short}}$, $\phi_{\mathsf{long}} = \phi_{\mathsf{short}}$, and $\phi_{\mathsf{long}} < \phi_{\mathsf{short}}$, respectively. The red line denotes the mean function in each case.

In view of Proposition 3.1, we know that $\widehat{\gamma}_0^{(0)}$ is a consistent estimator of $\theta + \gamma_0$. So, we propose to estimate ϕ_{short} by

$$\widehat{\phi}^{(m)}_{\mathsf{short}} = \frac{\widehat{\theta}^{(m)}}{\widehat{\gamma}_0^{(0)}}, \qquad m \geq 1. \tag{21}$$

The asymptotic behavior of $\widehat{\phi}^{(m)}_{\mathsf{short}}$ is more complicated than $\widehat{\phi}^{(m)}_{\mathsf{long}}$ because $\widehat{\theta}^{(m)}$ and $\widehat{\gamma}_0^{(0)}$ are structurally different, unlike $\widehat{\phi}^{(m)}_{\mathsf{long}}$, where $\widehat{\theta}^{(m)}$ and $\widehat{\nu}^{(m)}$ are structurally similar. To obtain the asymptotic distribution of $\widehat{\phi}^{(m)}_{\mathsf{short}}$, we need to first study the joint asymptotic distribution of $\widehat{\theta}^{(m)}$ and $\widehat{\gamma}_0^{(0)}$. We begin by considering the random vector $Q_i = (Z_i, Z_i^2 - \gamma_0)^{\mathrm{T}}$ for $i \in \mathbb{Z}$. Assume that $\{Q_i\}_{i\in\mathbb{Z}}$ satisfies the following regularity condition on the covariance structure:

Assumption 4.1. There exist $\lambda_0 > 0$ and $L_0 \in \mathbb{N}$ such that for all $t \geq 1$ and $L \geq L_0$,

$$\lambda_{\mathsf{min}}\left(\mathsf{Var}\left(\sum_{i=t+1}^{t+L} Q_i\right)\right) \geq \lambda_0 L,$$

where $\lambda_{\mathsf{min}}(A)$ denotes the smallest eigenvalue of A.

Assumption 4.1 is indeed the condition (2.B) of Karmakar and Wu (2020), which guarantees a lower bound on eigenvalues of covariance matrices of increment processes. For $i \in \mathbb{N}_0$ and $p \geq 1$, define the uniform functional dependence measure

$$\bar{\omega}_{p,i} = \sup_{j\in\mathbb{Z}} \left\| X_j - X_{j,\{j-i\}} \right\|_p . \tag{22}$$

We impose the following weak dependence conditions on $\{Q_i\}_{i\in\mathbb{Z}}$.

Assumption 4.2. $\bar{\Omega}_p = \sum_{i=0}^{\infty} \bar{\omega}_{p,i} < \infty$, where $p = (4+\iota)/2 > 2$, ι is as defined in Assumption 2.3.

Assumption 4.3. There exists $\chi > \chi_0 = \left\{p^2 - 4 + (p-2)\sqrt{p^2 + 20p + 4}\right\} / 8p > 0$ and $\beta > 0$ such that

$$\bar{\Omega}_{N,p} = \sum_{i=N}^{\infty} \bar{\omega}_{p,i} = O\left(N^{-\chi}(\log N)^{-\beta}\right), \tag{23}$$

as $N \to \infty$, where p is defined as in Assumption 4.2.

Assumption 4.2 implies short range dependence. In Assumption 4.3, $\bar{\Omega}_{N,p}$ is a measure of the tail cumulative dependence. A larger χ or β implies weaker dependence; see Karmakar and Wu (2020). Using the above framework for asymptotic theories, we can establish the joint asymptotic distribution of $\widehat{\theta}^{(m)}$ and $\widehat{\gamma}_0^{(0)}$, as shown in Theorem 4.7.

Theorem 4.7. *Let $m \geq 1$ be fixed, $u_q < \infty$ for some $q \in \mathbb{N}$, $h/\ell = \lambda \in [2, \infty)$, and $\ell/n^{1/2} + 1/\ell = o(1)$. Let Assumptions 2.2 and 2.3 hold. Further let Assumptions 4.1, 4.2 and 4.3 hold for $\{Q_i\}_{i\in\mathbb{Z}}$. Assume $\mathscr{C}$, $\mathscr{S}$ and $\mathscr{J}$ are fixed. Then, as $n \to \infty$,*

$$\sqrt{n}\left(\begin{pmatrix}\widehat{\theta}^{(m)} \\ \widehat{\gamma}_0^{(0)}\end{pmatrix} - \begin{pmatrix}\theta \\ \theta + \gamma_0\end{pmatrix}\right) \xrightarrow{\mathrm{d}} \mathrm{N}\left(\begin{pmatrix}0 \\ 0\end{pmatrix}, \begin{pmatrix}4v\theta & 4v\theta \\ 4v\theta & 4v\theta + \varsigma^2\end{pmatrix}\right),$$

where $\varsigma^2 = \lim_{n\to\infty} n\mathsf{Var}(\sum_{i=1}^n Z_i^2/n)$.

By the Delta method, we obtain the asymptotic distribution of $\widehat{\phi}^{(m)}_{\mathsf{short}}$, as stated in Corollary 4.8, which can be used to test

$$H_0^{\mathsf{short}} : \phi_{\mathsf{short}} \leq \phi_{\mathsf{short},0} \qquad \text{against} \qquad H_1^{\mathsf{short}} : \phi_{\mathsf{short}} > \phi_{\mathsf{short},0}, \tag{24}$$

where $0 < \phi_{\mathsf{short},0} < 1$ is a prescribed level of marginal relative variability.

Corollary 4.8. *Let the conditions of Theorem 4.7 hold. If the true marginal relative variability of the time series is $\phi_{\mathsf{short},0}$, then, as $n \to \infty$,*

$$\sqrt{n}\left(\widehat{\phi}^{(m)}_{\mathsf{short}} - \phi_{\mathsf{short},0}\right) \xrightarrow{\mathrm{d}} \mathrm{N}\left(0, v_{\mathsf{short}}\right),$$

where

$$v_{\mathsf{short}} = \frac{4v\phi_{\mathsf{short},0}\left(1 - \phi_{\mathsf{short},0}\right)^2}{\theta + \gamma_0} + \frac{\varsigma^2\phi^2_{\mathsf{short},0}}{(\theta + \gamma_0)^2}.$$

Similar to (19), we can construct a test statistic for testing (24) as follows:

$$\widehat{T}^{(m)}_{\mathsf{short}} = \frac{\sqrt{n}\max\left(\widehat{\phi}^{(m)}_{\mathsf{short}} - \phi_{\mathsf{short},0}, 0\right)}{\sqrt{\widehat{v}_{\mathsf{short}}}}, \tag{25}$$

where $\widehat{v}_{\mathsf{short}}$ is a consistent estimator of v_{short}. To obtain a consistent estimate for v_{short}, we note that (i) $\widehat{\gamma}_0^{(0)}$ is a consistent estimator of $\theta + \gamma_0$, and (ii) we can estimate ς^2 by

$$\widehat{\varsigma}^2 = \sum_{|k|\leq\ell} K\left(\frac{k}{\ell}\right)\widehat{\eta}_k, \qquad \text{where} \qquad \widehat{\eta}_k = \frac{1}{n}\sum_{i=mh+|k|+1}^{n} D_i^2 D_{i-|k|}^2,$$

where D_i is the same mth order difference statistic define in (5). Note that $\widehat{\varsigma}^2$ is a modified difference-based variance estimator in Chan (2022a) applied to the squared difference statistics $\{D_i^2\}_{i=mh+1}^n$. So, our proposed estimator for v_{short} is

$$\widehat{v}_{\text{short}} = \frac{4\widehat{v}^{(m)}\phi_{\text{short},0}\left(1-\phi_{\text{short},0}\right)^2}{\widehat{\gamma}_0^{(0)}} + \frac{\widehat{\varsigma}^2\phi^2_{\text{short},0}}{\left(\widehat{\gamma}_0^{(0)}\right)^2}.$$

In practice, H_0^{short} is rejected at size α if $\widehat{T}^{(m)}_{\text{short}} > \Phi^{-1}(1-\alpha)$. This test is asymptotically valid and consistent as $n \to \infty$.

We remark that both (19) and (25) can be used as test statistics for testing the existence of a relevant mean stationarity. They can be viewed as alternatives to existing tests, e.g., Dette and Wied (2016) and Dette and Wu (2019). Although these existing tests are also intended to test for relevant changes, they define relevant change as $\lambda(\{t \in [0,1] : |\mu(t) - \mu(0)| > c\}) > \Delta$, where λ is the Lebesgue measure, $c > 0$ is a pre-specified level and $\Delta \in (0,1)$ is a threshold. Hence, their approach is substantially different from our framework. Since they handle different problems, there is no direct way to compare them.

5. Implementation issues

5.1. Choice of difference sequence

Suppose $m \in \mathbb{N}$ is fixed. We suggest choosing a difference sequence that minimizes the MSE of $\widehat{\theta}^{(m)}$. Based on Theorem 3.2, the squared bias is negligible and $\mathsf{Var}(\widehat{\theta}^{(m)})$ is asymptotically proportional to $\Delta_m - \Delta_0 = \sum_{|k|\le m}\delta_k^2 - 1$. Hence, the optimal difference sequence $\{d_j\}_{j=0}^m$ can be chosen as follows:

$$\text{Minimize} \sum_{|k|\le m}\delta_k^2 - 1 \equiv \sum_{\substack{|k|\le m\\ k\neq 0}}\left(\sum_{j=0}^{m-|k|} d_j d_{j+|k|}\right)^2 \text{ subject to } \sum_{j=0}^m d_j = 0 \text{ and } \sum_{j=0}^m d_j^2 = 1, \tag{26}$$

where the constraints are needed due to the definition of $\{d_j\}_{j=0}^m$. The optimal solution $d^\star_{0:m} = (d_0^\star, \ldots, d_m^\star)^{\mathrm{T}}$ can be numerically found and is tabulated in Hall, Kay and Titterington (1990). In particular, $d^\star_{0:1} = (1/\sqrt{2}, -1/\sqrt{2})$, $d^\star_{0:2} = (0.8090, -0.5, -0.3090)$ and $d^\star_{0:3} = (0.1942, 0.2809, 0.3832, -0.8582)$. When $d^\star_{0:m}$ is chosen, we have

$$\sum_{|k|\le m}\delta_k^2 - 1 = \frac{1}{2m}. \tag{27}$$

Theoretically, this suggests that we should choose m as large as possible. However, using a large m leads to poor empirical performance in finite samples. This phenomenon is similar to Hall, Kay and Titterington (1990) and Chan (2022a). In practice, we suggest using $m = 3$ for our proposals; see Section 6 for some simulation evidence.

5.2. Finite-n adjustments

Our proposed tests are asymptotic in nature. So, there is approximation error in finite samples. We mitigate this problem by doing a simple finite-n adjustment. For fixed m, q and φ, we propose the use of finite-n adjusted critical values $c_\alpha(n,\rho,m,q,\varphi)$ by matching the autocorrelation function (ACF) at lag one ρ and the sample size n for a specified level of significance $\alpha \in (0,1)$. We only outline the procedure for $\widehat{T}^{(m)}_{\mathsf{MIT}}$. The procedures for other tests are similar.

Let $Z^\circ_1,\ldots,Z^\circ_n$ be generated from the autoregressive (AR) model: $Z^\circ_i = \rho Z^\circ_{i-1} + \varepsilon^\circ_i$, where $\rho \in \{0,\pm 0.1,\pm 0.2,\ldots,\pm 0.9\}$ and $\varepsilon^\circ_i \overset{\text{IID}}{\sim} \mathrm{N}(0,1)$. Then we compute the value of $\widehat{T}^{(m)}_{\mathsf{MIT}}$ based on $Z^\circ_1,\ldots,Z^\circ_n$. The above procedure is repeated for N times to obtain N simulated values of $\widehat{T}^{(m)}_{\mathsf{MIT}}$, denoted by $\widehat{T}^{\circ(m)}_{\mathsf{MIT},1},\ldots,\widehat{T}^{\circ(m)}_{\mathsf{MIT},N}$, where N is large. Consequently, $c_\alpha(n,\rho,m,q,\varphi)$ is estimated by the $100(1-\alpha)\%$ sample quantile of $\widehat{T}^{\circ(m)}_{\mathsf{MIT},1},\ldots,\widehat{T}^{\circ(m)}_{\mathsf{MIT},N}$. In practice, we estimate ρ by a mean-robust estimator $\widehat{\rho}$ defined as the sample lag-1 ACF of $\{X_{i+\lfloor n^{1/3}\rfloor} - X_i\}_{i=1}^{n-\lfloor n^{1/3}\rfloor}$; see Cheng and Chan (2023). We reject H_0 at size α if

$$\widehat{T}^{(m)}_{\mathsf{MIT}} > c_\alpha(n,\widehat{\rho},m,q,\varphi),$$

where $\widehat{T}^{(m)}_{\mathsf{MIT}}$ is the observed test statistic, and $c_\alpha(n,\widehat{\rho},m,q,\varphi)$ is the finite-n adjusted critical value with linear interpolation if necessary.

6. Monte-Carlo experiment

6.1. Test for mean invariance

In this section, we compare the performances of the proposed test $\widehat{T}^{(m)}_{\mathsf{MIT}}$ with (i) the standard KS change point (CP) test (Csörgő and Horváth, 1997) using the LRV estimator with Bartlett kernel and bandwidth $\lceil 2n^{1/3}\rceil$; (ii) a robust KS CP test using the difference-based LRV estimator proposed by Chan (2022a) with the recommended parameters; (iii) self-normalized KS CP test (Shao and Zhang, 2010); (iv) the constant-mean test proposed by Wu and Zhao (2007); and (v) KPSS stationarity test (Kwiatkowski et al., 1992) using the LRV estimator with Bartlett kernel and bandwidth $\lceil 4(n/100)^{1/4}\rceil$, which is also suggested in Hobijn, Franses and Ooms (2004). We consider a non-linear time series model in this section. We first generate $\{Z'_i\}_{i=1}^n$ from a non-linear autoregressive (NLAR) model as follows:

$$Z'_i = \rho\left|Z'_{i-1}\right| + \sqrt{1-\rho^2}\varepsilon_i, \tag{28}$$

where $\rho \in (-1,1)$ and $\varepsilon_i \overset{\text{IID}}{\sim} \mathrm{N}(0,1)$. Denote $\mu' = \mathsf{E}(Z'_1)$. In the simulation, we set $\rho \in \{0.25,0.5,0.75\}$. Let $\{Z_i\}_{i=1}^n$ be defined as $Z_i = (Z'_i - \mu')/\sqrt{v}$, where v is the LRV of the time series in (28), so that the LRV of $\{Z_i\}_{i=1}^n$ is 1. We consider four types of mean functions:

(a) (Single CP) $\mu(t) = \Delta\mathbb{1}(t>0.2)$;
(b) (Multiple CPs) $\mu(t) = \Delta\{2(\mathbb{1}(t>0.2) - \mathbb{1}(t>0.4) + \mathbb{1}(t>0.6) - \mathbb{1}(t>0.8))\}$;
(c) (Smooth function) $\mu(t) = \Delta\cos(5\pi t)$; and
(d) (Piecewise smooth functions) $\mu(t) = \Delta\{\mathbb{1}(t>0.2) - \cos(5\pi t)\}$,

where $\Delta \in [0,\infty)$. The sample size is $n = 800$. For our proposed test, $q = 2, \varphi = 1$ and $m \in \{1,2,3\}$ are used. In each simulation experiment, 2^{10} replications are used.

From the results shown in Figure 5, we see that the proposed test is less favorable in case (a). It is expected since the tests (i)–(iii) are specialized in the one CP case. In addition, from Section 4.2.1, we see that the test statistics of tests (i) and (v) admit similar forms, so test (v) also inherits the strength in handling one-CP problem. Both test (iv) and our proposed tests are specialized for a general mean

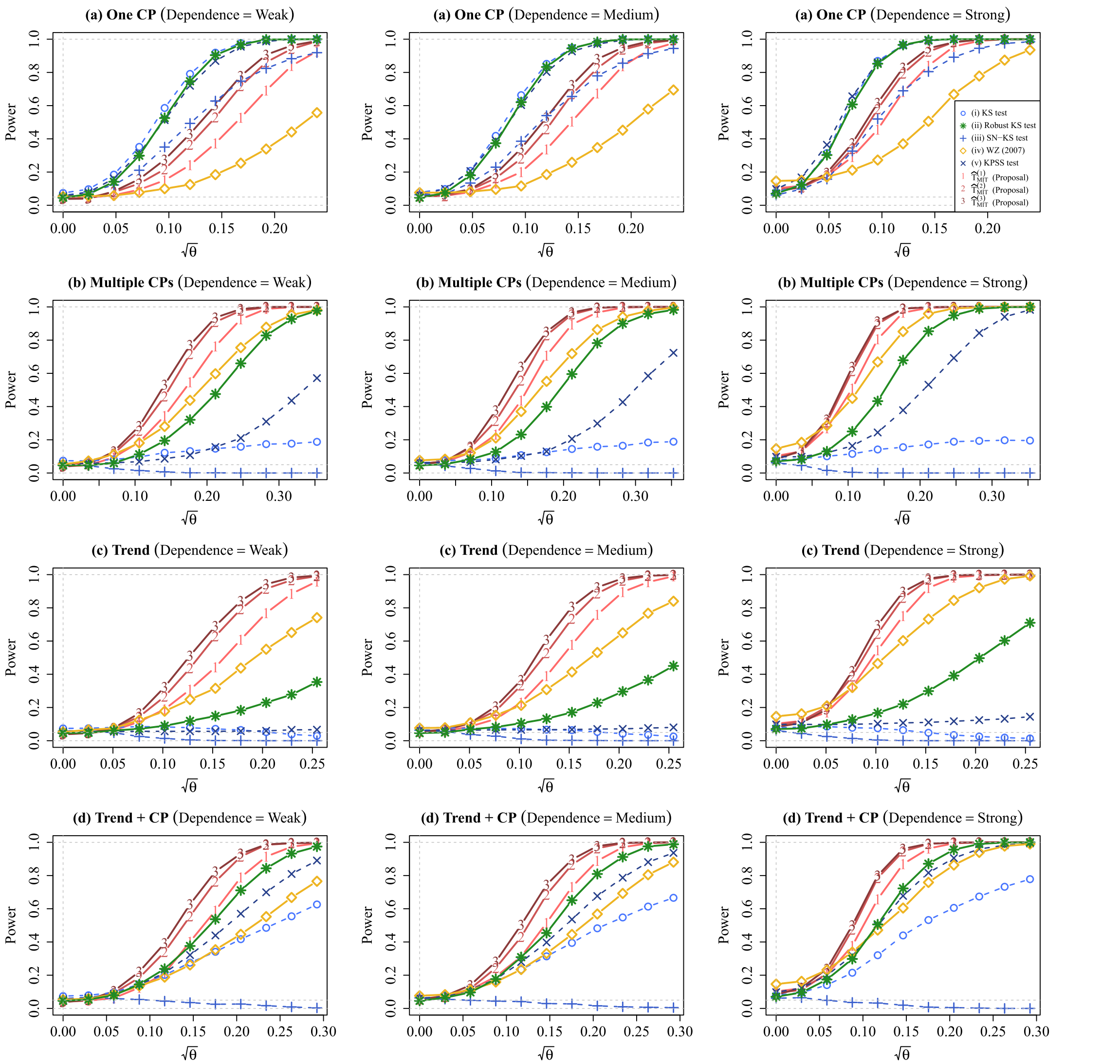

Figure 5. The three columns correspond to $\rho = 0.25$ (weak dependence), $\rho = 0.5$ (medium dependence) and $\rho = 0.75$ (strong dependence), respectively. The four rows correspond to the mean functions (a)–(d), respectively. The nominal size 5% is indicated by the horizontal grey dashed lines. The power curves of each methods are denoted as follows: test (i) standard KS CP test (Csörgő and Horváth, 1997) (- -○- -); test (ii) robust KS CP test (Chan, 2022a) (- -✳- -); test (iii) self-normalized KS CP test (Shao and Zhang, 2010) (–+–); test (iv) constant-mean test (Wu and Zhao, 2007) (–◇–); test (v) KPSS stationarity test (Kwiatkowski et al., 1992) (- -×- -); and the proposed tests $\widehat{T}^{(1)}_{\mathrm{MIT}}$ (–1–), $\widehat{T}^{(2)}_{\mathrm{MIT}}$ (–2–) and $\widehat{T}^{(3)}_{\mathrm{MIT}}$ (–3–).

stationarity problem, but our proposed tests are more powerful than test (iv) in this case. On the other hand, for cases (b)–(d), the proposed test is more powerful than the other tests in general, demonstrating its robustness against a large class of mean functions. Also, the proposed test controls the null rejection rate better, even under strong dependence. We can see that the standard KS test loses power in cases

Test \Δ	Dependence = Weak 0	0.1	0.2	0.3	Dependence = Medium 0	0.1	0.2	0.3	Dependence = Strong 0	0.1	0.2	0.3
$\widehat{T}^{(1)}_{\mathsf{MIT}}$	0.034	0.085	0.343	0.800	0.047	0.103	0.468	0.909	0.077	0.239	0.842	0.997
$\widehat{T}^{(2)}_{\mathsf{MIT}}$	0.028	0.101	0.503	0.935	0.038	0.119	0.604	0.975	0.065	0.274	0.921	0.999
$\widehat{T}^{(3)}_{\mathsf{MIT}}$	0.026	0.098	0.577	0.953	0.035	0.157	0.690	0.985	0.054	0.304	0.956	0.999
KPSS	0.048	0.074	0.228	0.749	0.061	0.098	0.377	0.904	0.122	0.247	0.845	0.999

Table 1. Powers of the proposed test and the KPSS test under $\rho = 0.25$ (weak dependence), $\rho = 0.5$ (medium dependence) and $\rho = 0.75$ (strong dependence) respectively. The nominal size is chosen at 5%.

(b) and (c), and SN based KS test loses power in cases (b)–(d). Tests (iv) and (v) remain monotonically powerful. However, test (iv) is less powerful in cases (c) and (d), and test (v) admits the over-sizing problem under strong dependence. While test (ii) has the advantage of being powerful in all cases with a price of slight size inaccuracy, it is still less powerful than the proposed test in general. Overall, the proposed test is robust against a large class of mean functions with better control of the null rejection rate and higher power in general. On the other hand, we note that the proposed test is increasingly more powerful when m increases. However, the power gain is diminishing. These observations concur with our theoretical findings in previous sections.

6.2. Test for smooth mean structure

In this section, we compare the performances of the proposed test with the KPSS trend stationarity test (Kwiatkowski et al., 1992). We consider the same noise time series model for $Z_1, \ldots, Z_n$ as described in Section 6.1. We are interested in testing $H_0^{(1)}$ defined in (14), i.e., there exist $a_0, a_1 \in \mathbb{R}$ such that $\mu(t) = a_0 + a_1 t$. The mean function under the alternative hypothesis is $\mu(t) = 0.5t + \Delta\cos(5\pi t)$, where $\Delta \geq 0$ is a measure of discrepancy from the null $H_0^{(1)}$. The sample size and the choices of parameters for the proposed test are the same as those in Section 6.1. In Table 1, we see that the proposed test is more powerful with a better control over the null rejection rate. The over-size problem that the KPSS admits is similar to its mean stationarity counterpart. Overall, the proposed test $\widehat{T}^{(3)}_{\mathsf{MIT}}$ gives the most promising performance in terms of size accuracy and power.

6.3. Test for relative variability and marginal relative variability

In this section, we perform a simulation study for testing (17) and (24) with $\phi_{\mathsf{long},0} = 0.05$ and $\phi_{\mathsf{short},0} = 0.05$. We consider the same noise time series model for $Z_1, \ldots, Z_n$ as described in Section 6.1. The mean function $\mu(t) = C\Delta\cos(20\pi t)$, where $C > 0$ is a scaling constant to ensure that $\phi_{\mathsf{long},0} = 0.05$ and $\phi_{\mathsf{short},0} = 0.05$ when $\Delta = 1$ in the respective settings. Note that the null hypotheses in (17) and (24) are true when $\Delta = 1$. When Δ increases, the null hypotheses are more obviously false. We consider $\rho \in \{0.25, 0.5, 0.75\}$ in the simulations. From the results in Figure 6, the power of $\widehat{T}^{(m)}_{\mathsf{long}}$ increases with m, whereas the power of $\widehat{T}^{(m)}_{\mathsf{short}}$ does not depend on m. These observations are consistent with the results in Theorem 4.6 and Corollary 4.8.

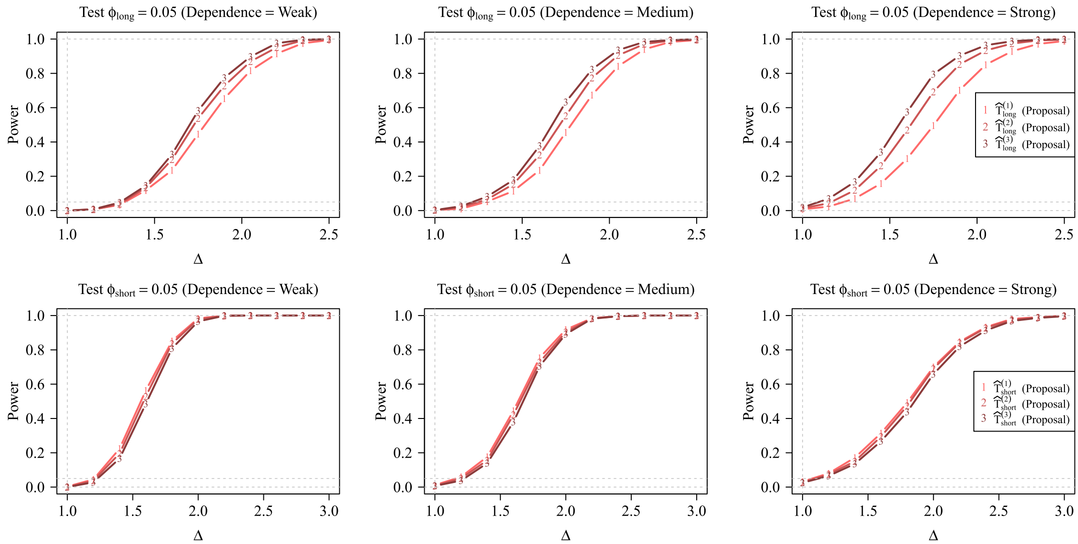

Figure 6. The power curves under $\rho = 0.25$ (weak dependence), $\rho = 0.5$ (medium dependence) and $\rho = 0.75$ (strong dependence). The nominal size 5% is indicated by the horizontal grey dashed lines. The top and bottom rows correspond to the tests for $\phi_{\text{long}} = 0.05$ and $\phi_{\text{short}} = 0.05$, respectively.

6.4. Sensitivity analysis

6.4.1. Sensitivity of bandwidth coefficient

In this section, we examine the effect of different choices of the bandwidth coefficient φ to the performance of our proposed test. The simulation experiments in Section 6.1 are repeated with the same noise time series models and the mean function described in Case (a) when $n = 800$, $q = 2$ and $m = 3$. We consider the bandwidth $\ell = \lfloor \varphi n^{1/5} \rfloor$ for $\varphi \in \{0.5, 1, 1.5, 2, 2.5, 3\}$.

Figure 7 shows the power curves of the tests under different values of φ. We observe that apart from the curve corresponding to a very short bandwidth, i.e., $\varphi = 0.5$, the other curves are close to each other. This observation is true for different strength of dependence. This suggests that $\varphi = 0.5$ is not a good choice, as the asymptotic results (as $\ell \to \infty$) may not kick in under this short bandwidth.

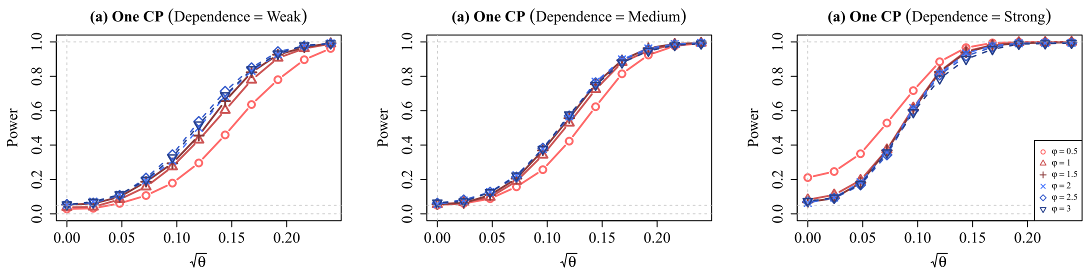

Figure 7. The power curves of the proposed test with $\varphi \in \{0.5, 1, 1.5, 2, 2.5, 3\}$ under $\rho = 0.25$ (weak dependence), $\rho = 0.5$ (medium dependence) and $\rho = 0.75$ (strong dependence). The nominal size 5% is indicated by the horizontal grey dashed lines.

Summary statistics \θ	Dependence = Weak 0	0.048	0.096	Dependence = Medium 0	0.048	0.096	Dependence = Strong 0	0.048	0.096
Mean	5.020	48.809	99.082	5.879	55.977	99.219	7.168	81.250	99.531
Median	5.273	50.000	99.023	5.664	56.738	99.316	7.129	81.543	99.414
Standard deviation	0.686	4.465	0.177	0.439	1.904	0.229	0.686	1.683	0.262
Interquartile range	0.195	6.348	0.293	0.684	1.270	0.391	0.488	1.758	0.391

Table 2. Summary statistics of the proposed test with $\varphi \in \{1, 1.5, 2, 2.5, 3\}$ under $\rho = 0.25$ (weak dependence), $\rho = 0.5$ (medium dependence) and $\rho = 0.75$ (strong dependence) respectively. The nominal size is chosen at 5%. All numbers are rounded to 3 decimal places after multiplying by a factor of 100.

On the other hand, from the summary statistics in Table 2, we see that both the standard deviation and the interquartile range of the power of the test across $\varphi \in \{1, 1.5, 2, 2.5, 3\}$ are very small under different strength of dependence. This observation remains true for both H_0 and H_1. Therefore, these experiments suggest that our proposed test is not too sensitive to the choice of φ.

6.4.2. *Sensitivity of kernel*

In this section, we examine the effect of different choices of kernel functions to the performance of our proposed test. The simulation experiments in Section 6.1 are repeated again with the same noise time series models and the mean function described in Case (a) when $n = 800$, $q = 2$, $m = 3$, and $\ell = \lfloor n^{1/5} \rfloor$. We consider three types of kernel function:

(i) second order polynomial kernel (Parzen, 1957): $K(t) = (1 - |t|^2)\mathbb{1}(|t| \le 1)$;
(ii) Tukey–Hanning kernel (Andrews, 1991): $K(t) = \{1 + \cos(\pi t)\}\mathbb{1}(|t| \le 1)/2$; and
(iii) Parzen kernel (Gallant, 1987): $K(t) = (1 - 6t^2 + 6|t|^3)\mathbb{1}(|t| \le 1/2) + 2(1 - |t|)^3\mathbb{1}(1/2 < |t| \le 1)$.

Figure 8 shows that the three curves are close to each other, demonstrating that our proposed test is not sensitive to the choice of kernel function, with the suggested second order polynomial kernel having a slight edge over the two other kernels.

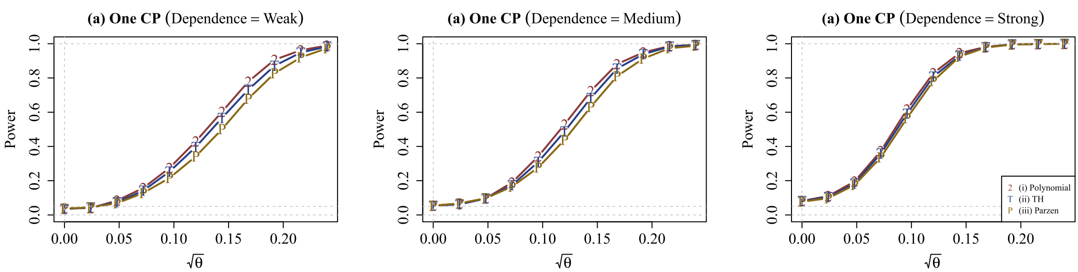

Figure 8. The power curves of the proposed test with (i) second order polynomial kernel (Parzen, 1957), (ii) Tukey–Hanning kernel (Andrews, 1991) and (iii) Parzen kernel (Gallant, 1987) under $\rho = 0.25$ (weak dependence), $\rho = 0.5$ (medium dependence) and $\rho = 0.75$ (strong dependence). The nominal size 5% is indicated by the horizontal grey dashed lines.

Supplementary Material

Supplement to "Mean stationarity test in time series: A signal variance-based approach" (DOI: 10.3150/23-BEJ1630SUPP; .pdf). Appendix A: Proofs of main results. The proofs of Propositions 3.1, 4.4, Theorems 3.2, 4.1, 4.3, 4.5, 4.6, 4.7, Corollaries 4.2, 4.8 and (11) are placed in Sections A.1–A.11, respectively. Appendix B: Auxiliary results. Technical results of independent interest are stated in Sections B.1–B.10. Appendix C: Additional simulation results. It contains additional simulation results for Section 6.1. Appendix D: Real-data application. A real-data application on global land surface temperature data is presented.

References

Andrews, D.W.K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* **59** 817–858. MR1106513 https://doi.org/10.2307/2938229

Andrews, D.W.K. (1995). Nonparametric kernel estimation for semiparametric models. *Econometric Theory* **11** 560–596. MR1349935 https://doi.org/10.1017/S0266466600009427

Antoch, J. and Jarušková, D. (2013). Testing for multiple change points. *Comput. Statist.* **28** 2161–2183. MR3107296 https://doi.org/10.1007/s00180-013-0401-1

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

Bradley, R.C. (2005). Basic properties of strong mixing conditions. A survey and some open questions. *Probab. Surv.* **2** 107–144. Update of, and a supplement to, the 1986 original. MR2178042 https://doi.org/10.1214/154957805100000104

Brockwell, P.J. and Davis, R.A. (1991). *Time Series: Theory and Methods*, 2nd ed. *Springer Series in Statistics*. New York: Springer. MR1093459 https://doi.org/10.1007/978-1-4419-0320-4

Casini, A. (2023). Theory of evolutionary spectra for heteroskedasticity and autocorrelation robust inference in possibly misspecified and nonstationary models. *J. Econometrics* **235** 372–392. MR4602874 https://doi.org/10.1016/j.jeconom.2022.05.001

Casini, A. and Perron, P. (2021). Prewhitened long-run variance estimation robust to nonstationarity. Manuscript.

Chan, K.W. (2022a). Optimal difference-based variance estimators in time series: A general framework. *Ann. Statist.* **50** 1376–1400. MR4441124 https://doi.org/10.1214/21-aos2154

Chan, K.W. (2022b). Mean-structure and autocorrelation consistent covariance matrix estimation. *J. Bus. Econom. Statist.* **40** 201–215. MR4356567 https://doi.org/10.1080/07350015.2020.1796397

Chan, K.W. and Yau, C.Y. (2017). High-order corrected estimator of asymptotic variance with optimal bandwidth. *Scand. J. Stat.* **44** 866–898. MR3730019 https://doi.org/10.1111/sjos.12279

Chan, K.W. and Yau, C.Y. (2023). Asymptotically constant risk estimator of time-average variance constant. Manuscript.

Chen, L., Wang, W. and Wu, W.B. (2022). Inference of breakpoints in high-dimensional time series. *J. Amer. Statist. Assoc.* **117** 1951–1963. MR4528482 https://doi.org/10.1080/01621459.2021.1893178

Chen, L. and Wu, W.B. (2019). Testing for trends in high-dimensional time series. *J. Amer. Statist. Assoc.* **114** 869–881. MR3963187 https://doi.org/10.1080/01621459.2018.1456935

Cheng, C.H. and Chan, K.W. (2023). A general framework for constructing locally self-normalized multiple-change-point tests. Manuscript.

Crainiceanu, C.M. and Vogelsang, T.J. (2007). Nonmonotonic power for tests of a mean shift in a time series. *J. Stat. Comput. Simul.* **77** 457–476. MR2405424 https://doi.org/10.1080/10629360600569394

Csörgő, M. and Horváth, L. (1997). *Limit Theorems in Change-Point Analysis*. *Wiley Series in Probability and Statistics*. Chichester: Wiley. With a foreword by David Kendall. MR2743035

Dalla, V., Giraitis, L. and Phillips, P.C.B. (2015). Testing mean stability of heteroskedastic time series. Cowles Foundation Discussion Papers 2441.

Dedecker, J. and Prieur, C. (2005). New dependence coefficients. Examples and applications to statistics. *Probab. Theory Related Fields* **132** 203–236. MR2199291 https://doi.org/10.1007/s00440-004-0394-3

Dette, H. and Wied, D. (2016). Detecting relevant changes in time series models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **78** 371–394. MR3454201 https://doi.org/10.1111/rssb.12121

Dette, H. and Wu, W. (2019). Detecting relevant changes in the mean of nonstationary processes—a mass excess approach. *Ann. Statist.* **47** 3578–3608. MR4025752 https://doi.org/10.1214/19-AOS1811

Fryzlewicz, P. (2014). Wild binary segmentation for multiple change-point detection. *Ann. Statist.* **42** 2243–2281. MR3269979 https://doi.org/10.1214/14-AOS1245

Gallant, A.R. (1987). *Nonlinear Statistical Models*. *Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics*. New York: Wiley. MR0921029 https://doi.org/10.1002/9780470316719

Górecki, T., Horváth, L. and Kokoszka, P. (2018). Change point detection in heteroscedastic time series. *Econom. Stat.* **7** 63–88. MR3824127 https://doi.org/10.1016/j.ecosta.2017.07.005

Hall, P., Kay, J.W. and Titterington, D.M. (1990). Asymptotically optimal difference-based estimation of variance in nonparametric regression. *Biometrika* **77** 521–528. MR1087842 https://doi.org/10.1093/biomet/77.3.521

Hobijn, B., Franses, P.H. and Ooms, M. (2004). Generalizations of the KPSS-test for stationarity. *Stat. Neerl.* **58** 483–502. MR2106350 https://doi.org/10.1111/j.1467-9574.2004.00272.x

Horváth, L., Kokoszka, P. and Steinebach, J. (1999). Testing for changes in multivariate dependent observations with an application to temperature changes. *J. Multivariate Anal.* **68** 96–119. MR1668911 https://doi.org/10.1006/jmva.1998.1780

Ibragimov, I.A. (1962). Some limit theorems for stationary processes. *Teor. Veroyatn. Primen.* **7** 361–392. MR0148125

Jiang, F., Zhao, Z. and Shao, X. (2022). Modelling the COVID-19 infection trajectory: A piecewise linear quantile trend model. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **84** 1589–1607. MR4515551

Juhl, T. and Xiao, Z. (2009). Tests for changing mean with monotonic power. *J. Econometrics* **148** 14–24. MR2494814 https://doi.org/10.1016/j.jeconom.2008.08.020

Karmakar, S. and Wu, W.B. (2020). Optimal Gaussian approximation for multiple time series. *Statist. Sinica* **30** 1399–1417. MR4257539 https://doi.org/10.5705/ss.202017.0303

Kwiatkowski, D., Phillips, P.C.B., Schmidt, P. and Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *J. Econometrics* **54** 159–178.

Liu, X. and Chan, K.W. (2023). No-lose converging kernel estimation of long-run variance. Manuscript.

Parzen, E. (1957). On consistent estimates of the spectrum of a stationary time series. *Ann. Math. Stat.* **28** 329–348. MR0088833 https://doi.org/10.1214/aoms/1177706962

Priestley, M.B. and Subba Rao, T. (1969). A test for non-stationarity of time-series. *J. Roy. Statist. Soc. Ser. B* **31** 140–149. MR0269062

Rosenblatt, M. (1956). A central limit theorem and a strong mixing condition. *Proc. Natl. Acad. Sci. USA* **42** 43–47. MR0074711 https://doi.org/10.1073/pnas.42.1.43

Shao, X. and Zhang, X. (2010). Testing for change points in time series. *J. Amer. Statist. Assoc.* **105** 1228–1240. MR2752617 https://doi.org/10.1198/jasa.2010.tm10103

To, H.K. and Chan, K.W. (2024). Supplement to "Mean stationarity test in time series: A signal variance-based approach." https://doi.org/10.3150/23-BEJ1630SUPP

Vats, D. and Flegal, J.M. (2022). Lugsail lag windows for estimating time-average covariance matrices. *Biometrika* **109** 735–750. MR4472845 https://doi.org/10.1093/biomet/asab049

Wu, W.B. (2004). A test for detecting changes in mean. In *Time Series Analysis and Applications to Geophysical Systems* (D.R. Brillinger, E.A. Robinson and F. Schoenberg, eds.) **45** 105–121. New York: Springer-Verlag.

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

Wu, W.B. (2007). Strong invariance principles for dependent random variables. *Ann. Probab.* **35** 2294–2320. MR2353389 https://doi.org/10.1214/009117907000000060

Wu, W.B. (2011). Asymptotic theory for stationary processes. *Stat. Interface* **4** 207–226. MR2812816 https://doi.org/10.4310/SII.2011.v4.n2.a15

Wu, W.B., Woodroofe, M. and Mentz, G. (2001). Isotonic regression: Another look at the changepoint problem. *Biometrika* **88** 793–804. MR1859410 https://doi.org/10.1093/biomet/88.3.793

Wu, W.B. and Zhao, Z. (2007). Inference of trends in time series. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **69** 391–410. MR2323759 https://doi.org/10.1111/j.1467-9868.2007.00594.x

Zhang, T. and Lavitas, L. (2018). Unsupervised self-normalized change-point testing for time series. *J. Amer. Statist. Assoc.* **113** 637–648. MR3832215 https://doi.org/10.1080/01621459.2016.1270214

Zhang, T. and Wu, W.B. (2011). Testing parametric assumptions of trends of a nonstationary time series. *Biometrika* **98** 599–614. MR2836409 https://doi.org/10.1093/biomet/asr017

Received September 2022 and revised May 2023