Journal of the Royal Statistical Society
Statistical Methodology
Series B

J. R. Statist. Soc. B (2015)

A new non-parametric stationarity test of time series in the time domain

Lei Jin,

Texas A&M University, Corpus Christi, USA

Suojin Wang

Texas A&M University, College Station, USA

and Haiyan Wang

Kansas State University, Manhattan, USA

[Received August 2012. Final revision August 2014]

Summary. We propose a new double-order selection test for checking second-order stationarity of a time series. To develop the test, a sequence of systematic samples is defined via Walsh functions. Then the deviations of the autocovariances based on these systematic samples from the corresponding autocovariances of the whole time series are calculated and the uniform asymptotic joint normality of these deviations over different systematic samples is obtained. With a double-order selection scheme, our test statistic is constructed by combining the deviations at different lags in the systematic samples. The null asymptotic distribution of the statistic proposed is derived and the consistency of the test is shown under fixed and local alternatives. Simulation studies demonstrate well-behaved finite sample properties of the method proposed. Comparisons with some existing tests in terms of power are given both analytically and empirically. In addition, the method proposed is applied to check the stationarity assumption of a chemical process viscosity readings data set.

Keywords: Autocovariance; Order selection; Stationarity test; Systematic samples; Time series; Walsh functions

1. Introduction

In time series analysis, many models are based on the assumption that the data are from a second-order stationary process. For example, second-order stationarity is a required assumption for the widely used auto-regressive moving average (ARMA) models which form an important class of linear time series models. This assumption guarantees a sound asymptotic theory which is essential for statistical inference. However, many series in reality show non-stationary behaviour and various non-stationary models have been proposed to address this issue. One important class of these models is the class of locally stationary processes (Dahlhaus, 1997) that can be approximated by stationary processes locally. Many statistical methods designated for locally stationary processes have been developed, such as the wavelet method in Nason *et al.* (2000) and the smooth localized complex exponential method in Ombao *et al.* (2005). Although models and methods for stationary series are well developed, powerful and usually convenient,

Address for correspondence: Suojin Wang, Department of Statistics, Texas A&M University, College Station, TX 77843, USA.
E-mail: sjwang@stat.tamu.edu

 1369–7412/15/77000

to avoid the risk of obtaining misleading conclusions one should not use these methods on non-stationary series. Therefore, it is critical to check whether a time series is second-order stationary.

Many statistical tests for checking the second-order stationarity property have been proposed in the literature. One early work to detect non-stationarity was by Priestley and Subba Rao (1969). von Sachs and Neumann (2000) proposed a test based on empirical wavelet coefficients. Lee *et al.* (2003) proposed a cumulative sum test for checking whether a few autocovariances change over time. Employing the uncorrelated property of the discrete Fourier transform of a stationary time series at the canonical frequencies, Dwivedi and Subba Rao (2011) proposed a portmanteau-type test for checking stationarity. A few L_2-distance-based tests have also been proposed to test the stationarity in recent years, including Paparoditis (2009, 2010) and Dette *et al.* (2011). The test statistic in Paparoditis (2010) is related to the maximum L_2-distance between local periodograms defined in rolling windows and the spectral density estimate assuming stationarity. The test statistic in Dette *et al.* (2011) is related to the minimal L_2-distance between the spectral density and its stationary approximation. It is consistent under local alternatives which converge to the null hypothesis at any rate slower than $T^{-1/4}$, where T is the length of a time series. As is shown in this work, such a convergence rate of local alternatives can be significantly improved to increase the power of a test which would be especially appealing when analysing short to moderate time series. More recently, Paparoditis and Preuß (2013) studied the asymptotic properties of some recent stationarity tests in the frequency domain under a sequence of local alternatives. Nason (2013a) introduced a test of stationarity by assessing the significance of Haar wavelet coefficients of the wavelet periodogram of a time series. This test works for non-Gaussian data and is capable of identifying the location of non-stationarity. The theory of confidence intervals for localized auocovariance was also provided but the asymptotic properties of the test under local alternatives were not studied.

In this paper, we propose a double-order selection approach that is based on autocovariances and systematic samples in the time domain. The test proposed is simple and applicable with non-Gaussian data but comparable with or more powerful than most existing methods in various situations. Similarly to order selection tests (see Eubank and Hart (1992) and Kim (2000)) designed for other problems, our order selection test is asymptotically consistent under any fixed alternatives of non-stationarity. Furthermore, our test statistic is shown to be consistent under a sequence of local alternatives which converges to the null hypothesis at any rate slower than $T^{-1/2}$. In addition, our test is computationally easy as the construction of the test mainly involves simple quantities such as the sample autocovariances of time series, sample autocovariances of systematic samples and the corresponding estimated covariance matrix. Another feature of our test is that it can work well even when the length of time series is relatively short.

The main idea behind our method is that the autocovariances of any systematic (non-random) sample of the time series equals the corresponding autocovariance of the whole time series if a time series is second-order stationary. Therefore, if there is some significant deviation from a sample autocovariance of a systematic sample from that of the whole time series, it implies that the time series is not stationary. To detect the existence of such deviations, a sequence of systematic samples is defined via the class of Walsh functions (see Ahmed and Rao (1975) and Stoffer (1987)), which consist of a complete orthogonal basis for the space of square integrable functions on $[0, 1)$. Such a device defines global contrasts which can lead to more power in detecting global changes while a similar device via wavelet functions defines local contrasts. The deviations of the sample autocovariances of a systematic sample based on a Walsh function from those of the whole time series have similar asymptotic properties and forms to those of the sample autocovariances. Moreover, the completeness of the Walsh functions enables the sequence of systematic samples to capture any deviation in an autocovariance function from a constant. The test statistic proposed employs a double-order selection strategy, which makes good use of the deviations of the sample autocovariances of different systematic samples at different lags from those of the whole time series, leading to good power properties. Our simulation studies also show that the test proposed has good empirical type I error rates for various models even when the length of time series is quite short and the empirical power appears to perform quite well.

The rest of the paper is organized as follows. In Section 2 we introduce the class of Walsh functions and describe our testing procedure. We then investigate the null asymptotic distribution of the test statistic and its consistency under both fixed and local alternatives. Section 3 presents our investigation of the finite sample properties of the proposed test via simulation studies and an application to check the stationarity assumption of a time series on chemical process viscosity readings. The data set was originally published by Box and Jenkins (1970). It is currently available from `http://robjhyndman.com/tsdldata/data/boxjenk5.dat`. The R code to perform the test proposed can be downloaded from `http://faculty.tamucc.edu/ljin1/TestStationarity/tsRcode.txt`. The main function for the test is `Ts_StationarityTest(ts)`, where `ts` is the vector of the time series data to be analysed. Some concluding remarks are given in Section 4 and the technical proofs are provided in Appendix A.

2. Proposed methodology

2.1. Walsh functions and systematic samples

In this subsection, a sequence of systematic samples is defined on the basis of Walsh functions (Ahmed and Rao, 1975). Some applications of Walsh functions in statistics have also been discussed in Stoffer (1987, 1991). The class of Walsh functions $\{W_l(t),\ l=0,1,2,\ldots\}$ defined below forms a complete orthogonal basis for the square integrable functions on [0,1):

$$W_0(x)=1, \qquad x\in[0,1),$$
$$W_1(x)=\begin{cases}1, & x\in[0,0.5),\\ -1, & x\in[0.5,1),\end{cases}$$

and recursively, for any $l=1,2,\ldots,$

$$W_{2l}(x)=\begin{cases}W_l(2x), & x\in[0,0.5),\\ (-1)^l\ W_l(2x-1), & x\in[0.5,1),\end{cases}$$
$$W_{2l+1}(x)=\begin{cases}W_l(2x), & x\in[0,0.5),\\ (-1)^{l+1}\ W_l(2x-1), & x\in[0.5,1).\end{cases}$$

By rearranging all the Walsh functions to $\{W_k(t), k=0,1,2,\ldots\}$ so that the number of sign changes in a function is in increasing order, the subscript k of $W_k(t)$ is the sequency, which is the number of zero-crossings of the function in $[0,1)$. Fig. 1 illustrates the first nine Walsh functions. For each $k<T$, the kth discrete Walsh function on $\mathcal{T}=\{1,2,\ldots,T\}$ is a T-dimensional vector

$$\boldsymbol{\mathcal{W}}_k=(\mathcal{W}_k(1),\mathcal{W}_k(2),\ldots,\mathcal{W}_k(T))^{\mathrm{T}},$$

where $\mathcal{W}_k(t)=W_k\{(t-1)/T\}$ for $t=1,2,\ldots,T$. Some properties of discrete Walsh functions are given in Appendix A.

Given a time series $\{X_t, t\in\mathcal{T}\}$, a sequence of systematic samples $\mathcal{T}_k,\ k=1,2,\ldots,M,$ in the time domain is defined on the basis of Walsh functions as follows:

$$\mathcal{T}_k=\{t\in\mathcal{T},\ \text{where } (-1)^{k-1}\,\mathcal{W}_k(t)=1\}. \qquad (1)$$

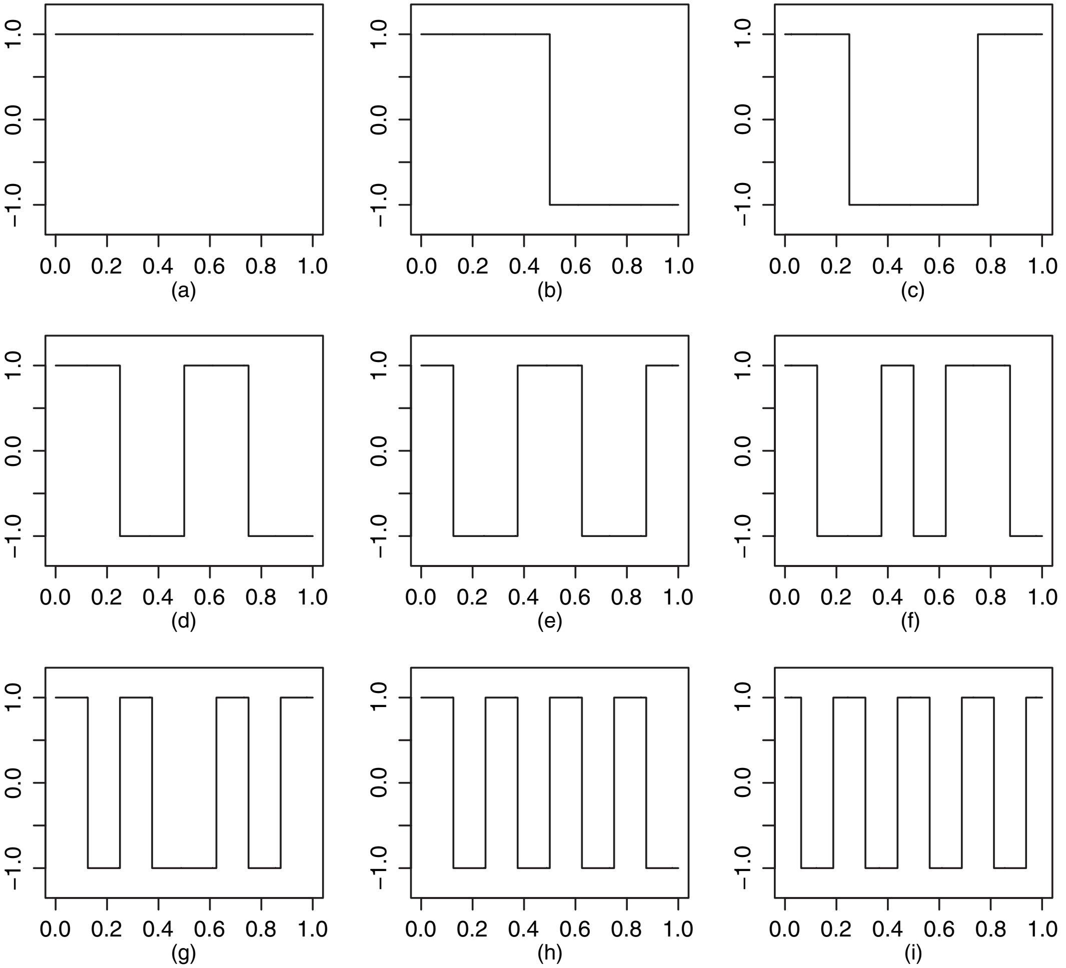

Fig. 1. First nine Walsh functions in sequency order k (pieces over neighbouring intervals are connected by vertical lines to show the pattern: (a) $k=0$; (b) $k=1$; (c) $k=2$; (d) $k=3$; (e) $k=4$; (f) $k=5$; (g) $k=6$; (h) $k=7$; (i) $k=8$

Let $\lfloor a \rfloor$ be the greatest integer that is no more than a. According to the definition in equation (1), the first systematic sample $\mathcal{T}_1 = \{1, 2, \ldots, \lfloor T/2 \rfloor\}$ is the first half of the whole time domain; the second sample $\mathcal{T}_2 = \{\lfloor T/4 \rfloor + 1, \lfloor T/4 \rfloor + 2, \ldots, \lfloor 3T/4 \rfloor\}$ is the middle half of the time domain; the third systematic sample $\mathcal{T}_3 = \{1, \ldots, \lfloor T/4 \rfloor\} \cup \{\lfloor T/2 \rfloor + 1, \lfloor T/2 \rfloor + 2, \ldots, \lfloor 3T/4 \rfloor\}$ consists of two blocks. Let N_k be the smallest power of 2 that is at least k. In general, a systematic sample $\mathcal{T}_k$ consists of $\lfloor (k+1)/2 \rfloor$ blocks, with the length of each block being at least $\lfloor T/N_k \rfloor$. Since k is the number of zero crossings of $\boldsymbol{\mathcal{W}}_k$ and $\mathcal{W}_k(1)$ is always 1 for all k, the last block of $\boldsymbol{\mathcal{W}}_k$ is all -1 when k is odd and all 1 when k is even. Therefore, by using the multiplier $(-1)^{k-1}$ in equation (1), the kth systematic sample $\mathcal{T}_k$ always excludes the last $\lfloor T/N_k \rfloor$ elements of $\mathcal{T}$.

As is seen in equation (1), a Walsh function works as an indicator function naturally for a systematic sample. It is a piecewise constant function taking on only values 1 and -1. In addition, any two Walsh functions are orthogonal and thus the information about stationarity on two different systematic samples is not overlapped or related (this point is reflected in theorem 1 in Section 2.2). The Haar wavelet basis has some similar properties. Roughly speaking, Walsh functions are global whereas wavelets are localized small waves. The systematic samples defined via Walsh functions always consist of one half of the observations, though these via the Haar wavelet functions without appropriate manipulation or restrictions may contain a much smaller number of observations and thus the corresponding variation of statistic could be much larger. With the smaller variations in the statistic based on the systematic deviations from a constant autocovariance, we can develop a simple and powerful method to detect non-stationarity. In principle, it is also possible to develop this method by using the Haar wavelet basis in parallel by thresholding, adding some regularization restrictions or combining several systematic samples. The test in von Sachs and Neumann (2000) is based on the wavelet coefficients, which can be interpreted as local contrasts on the spectral density in the time direction. In contrast, our test is based on Walsh coefficients, which can be considered as global contrasts of autocovariances comparing one half of the observations with the other half in the whole time domain. As will be seen later, the asymptotic variance of the estimated Walsh coefficients is shown to be of order $O(T^{-1})$.

2.2. Difference between two estimates of autocovariances: one based on systematic samples and the other based on the whole time series

For a stationary time series, $\gamma_h = \mathrm{cov}(X_t, X_{t+h})$ is the autocovariance function at lag h. If the time series is a zero-mean process, a natural estimate of γ_h is the sample autocovariance based on the entire time series:

$$\hat{\gamma}_h = \frac{1}{T} \sum_{t=1}^{T-h} X_t X_{t+h}.$$

Similarly, when $h < T/N_k$, which is the case in our asymptotic development, an estimate of γ_h based on the kth systematic sample $\mathcal{T}_k$ is

$$\hat{\gamma}_h^{(k)} = \frac{2}{T} \sum_{t \in \mathcal{T}_k} X_t X_{t+h} = \frac{1}{T} \sum_{t=1}^{T-h} X_t X_{t+h} \{1 + (-1)^{k-1} \mathcal{W}_k(t)\}. \tag{2}$$

Thus, the difference between the sample autocovariance of the kth systematic sample and that of the whole time series is

$$\hat{\gamma}_h^{(k)} - \hat{\gamma}_h = \frac{(-1)^{k-1}}{T} \sum_{t=1}^{T-h} X_t X_{t+h} \mathcal{W}_k(t). \tag{3}$$

The difference in equation (3) has a form of the Walsh transformation which can be easily generalized by using other orthogonal transformations. Not surprisingly, some general tests can be developed in the same way with different orthogonal basis functions instead of Walsh functions. However, the special features of Walsh functions lead to some desirable asymptotic properties of $\hat{\gamma}_h^{(k)} - \hat{\gamma}_h$ at different lags and systematic samples, which are mainly described in theorems 1 and 2 and lemma 1.

To have the desired asymptotic properties of $\hat{\gamma}_h^{(k)} - \hat{\gamma}_h$ under stationarity, the following assumption is needed for the stationary process $\{X_t\}$. Let $v(l) = 1$ if $|l| \leqslant 1$ or $v(l) = \{|l| \log^{1+\kappa}(|l|)\}^{-1}$ if $|l| > 1$ for some $\kappa > 0$. Clearly, $\Sigma_{l=0}^{\infty} |v(l)| < \infty$. The log-function without specifying its base refers to the natural logarithm in the paper.

Assumption 1. Assume that $\{X_t\}$ is a stationary process satisfying

$$X_t = \sum_{l=0}^{\infty} \phi_l Z_{t-l}, \tag{4}$$

where $\{Z_t\}$ are independent and identically distributed (IID) random variables with mean 0 and finite standard deviation σ, $|\phi_l| \leqslant K_0\, v(l)$, $E(Z_t^{4\nu}) < \infty$ for some $\nu > 1$ and $K_0 > 0$.

The condition $|\phi_l| \leqslant K_0\, v(l)$ is from Dahlhaus (2009) and is adapted to the stationary case here. It guarantees that $\Sigma_{l=0}^{\infty}|\phi_l| < \infty$ and $\Sigma_{h=-\infty}^{\infty}|\gamma_h| < \infty$. Moreover, it implies that $|\gamma_h| = |\Sigma_{l=0}^{\infty}\phi_l\phi_{l+h}| \leqslant K_0\, v(h)\,\Sigma_{l=0}^{\infty}|\phi_l| = o(h^{-1})$.

Let R be the maximum number of autocovariances and M be the maximum number of systematic samples to be considered for a given T. By the definition of a systematic sample in equation (1), M is also the maximum number of Walsh functions to be used when T is given. In the paper, we require that $R \to \infty$ and $M \to \infty$ as $T \to \infty$. We impose the following assumption on the rates of R and M.

Assumption 2. Assume that

$$R = O[\log(T)/\log\{\log(T)\}] \text{ and } M = O(T^{1/3}).$$

The following two theorems are on the asymptotic covariance of $\hat{\gamma}_h^{(k)} - \hat{\gamma}_h$.

Theorem 1. Under assumptions 1 and 2, we have

$$\max_{k_1 \neq k_2 \leqslant M,\, 0 \leqslant i \leqslant j \leqslant R} |T \operatorname{cov}(\hat{\gamma}_i^{(k_1)} - \hat{\gamma}_i, \hat{\gamma}_j^{(k_2)} - \hat{\gamma}_j)| \to 0 \qquad \text{as } T \to \infty.$$

Theorem 2. Under assumptions 1 and 2, for any integers $i, j \geqslant 0$ and $k > 0$, we have

$$T \operatorname{cov}(\hat{\gamma}_i^{(k)} - \hat{\gamma}_i, \hat{\gamma}_j^{(k)} - \hat{\gamma}_j) \to \Gamma_{i+1,j+1}$$

as $T \to \infty$, where

$$\Gamma_{i+1,j+1} = \kappa_4\gamma_i\gamma_j + \sum_{v=-\infty}^{\infty}(\gamma_v\gamma_{v-i+j} + \gamma_{v+i}\gamma_{v-j})$$

and κ_4 is the kurtosis of Z_t.

The proofs of theorems 1 and 2 are given in Appendix A.

Let $\hat{\boldsymbol{\gamma}}_r^{(k)} = (\hat{\gamma}_0^{(k)}, \hat{\gamma}_1^{(k)}, \ldots, \hat{\gamma}_{r-1}^{(k)})^{\mathrm{T}}$ and $\hat{\boldsymbol{\gamma}}_r = (\hat{\gamma}_0, \hat{\gamma}_1, \ldots, \hat{\gamma}_{r-1})^{\mathrm{T}}$. Denote $\mathcal{G}_a^b = \sigma\{X_t | a \leqslant t \leqslant b\}$ the σ-field that is generated by random variables $\{X_a, \ldots, X_b\}$. Under the strong mixing condition

$$\sup_v\{|P(E \cap F) - P(E)\,P(F)|, E \in \mathcal{G}_{-\infty}^n, F \in \mathcal{G}_{n+v}^{\infty}\} = \alpha(v) \leqslant A_0 v^{-\varsigma}, \tag{5}$$

where $A_0 > 0$ is a constant and $\varsigma > 1$, Keenan (1997) obtained the uniform weak convergence of the joint distribution of the first R sample autocovariances to multivariate normal. We also have the uniform weak convergence of $\hat{\boldsymbol{\gamma}}_R^{(k)} - \hat{\boldsymbol{\gamma}}_R$ as stated below.

Lemma 1. Under assumptions 1 and 2, and the strong mixing condition given in equation (5), we have

$$\lim_{T\to\infty} \sup_{1 \leqslant k \leqslant M, \mathbf{x}} |P\{T^{-1/2}(\hat{\boldsymbol{\gamma}}_R^{(k)} - \hat{\boldsymbol{\gamma}}_R) \leqslant \mathbf{x}\} - P\{(g_1, \ldots, g_R)^{\mathrm{T}} \leqslant \mathbf{x}\}| \to 0, \tag{6}$$

where $\mathbf{x} = (x_1, x_2, \ldots, x_R)^{\mathrm{T}}$ is an R-dimensional vector of real numbers and $\{g_i\}_{i \geqslant 1}$ is a zero-mean Gaussian process with $\operatorname{cov}(g_i, g_j) = \Gamma_{ij}$.

Keenan (1997) used a splitting procedure to prove his result. Our proof of lemma 1 takes advantage of the same splitting procedure. A detailed proof is given in Appendix A.

Some sufficient conditions for strong mixing are given in chapter 14 of Davidson (1994). Briefly, if the innovations Z_t of a linear process are from a continuous distribution with some moment conditions, a certain summability condition on the coefficients in equation (4) implies a strong mixing condition with a specified rate. In our case, the condition $\phi_l \leqslant K_0 l^{-1.8}\, v(l)$

is sufficient to establish the strong mixing condition in equation (5), assuming that Z_t has a continuous distribution with $E(|Z_t|^{4\nu}) < \infty$ for $\nu = 1.5$. Note that the sufficient condition on ϕ_l can be slightly weakened with a larger ν. In addition, the strong mixing condition in equation (5) can possibly be relaxed by using the ideas in Jirak (2011), in which Keenan's result was obtained under weaker conditions without the assumption of strong mixing.

Clearly, for any fixed number r, lemma 1 implies that

$$T^{-1/2}(\hat{\gamma}_r^{(k)} - \hat{\gamma}_r) \rightarrow N(0, \mathbf{\Gamma}_r) \tag{7}$$

in distribution uniformly for all $k \leqslant M$, where $\mathbf{\Gamma}_r$ is an $r \times r$ matrix with (i, j)th entry $\Gamma_{i,j}$ defined in theorem 2.

2.3. *Test statistic and its asymptotic distribution under stationarity*

To estimate the matrix $\mathbf{\Gamma}_r$, we define $\hat{\mathbf{\Gamma}}_r^{(k)}$ with $(i+1, j+1)$th element

$$\hat{\Gamma}^{(k)}_{i+1,j+1} = \hat{\kappa}_4 \hat{\gamma}_i \hat{\gamma}_j + \sum_{v=-Q_T}^{Q_T} A_{k,v} (\hat{\gamma}_v \hat{\gamma}_{v-i+j} + \hat{\gamma}_{v+i} \hat{\gamma}_{v-j})$$

for $0 \leqslant i, j \leqslant r-1$, $A_{k,v} = 1 - (2k+1)|v|/T$, $Q_T = \lfloor T^{\lambda} \rfloor$ and $\lambda \in (0, 0.5)$, and $\hat{\kappa}_4$ is an estimate of κ_4. A specific version of $\hat{\kappa}_4$ for our test will be given in equation (15) in Section 2.4 which will be seen to have preferable large sample consistency properties under both the null and the alternative hypotheses. As in Lee *et al.* (2003) define $\hat{\mathbf{\Gamma}}_r$ with the $(i+1, j+1)$th element

$$\hat{\Gamma}_{i+1,j+1} = \hat{\kappa}_4 \hat{\gamma}_i \hat{\gamma}_j + \sum_{v=-Q_T}^{Q_T} (\hat{\gamma}_v \hat{\gamma}_{v-i+j} + \hat{\gamma}_{v+i} \hat{\gamma}_{v-j})$$

for $0 \leqslant i, j \leqslant r-1$. Lee *et al.* (2003) pointed out that under stationarity $\hat{\mathbf{\Gamma}}_r$ is consistent for $\mathbf{\Gamma}_r$ if $\hat{\kappa}_4$ is a consistent estimate of κ_4 when $\lambda \in (0, (\nu-1)/(2\nu))$. Although $\hat{\mathbf{\Gamma}}_r^{(k)}$ and $\hat{\mathbf{\Gamma}}_r$ are asymptotically equivalent, our empirical studies show that the test statistic using $\hat{\mathbf{\Gamma}}_r^{(k)}$ generally has better performance under the null than the test statistic using $\hat{\mathbf{\Gamma}}_r$, especially when T is small to moderate (such as $T \leqslant 128$).

Lemma 2. Under the assumptions in theorem 1, given that $\hat{\kappa}_4$ is a consistent estimate of κ_4,

$$\max_{1 \leqslant k \leqslant M, 1 \leqslant i,j \leqslant R} |\hat{\Gamma}_{ij}^{(k)} - \Gamma_{ij}| \rightarrow 0 \tag{8}$$

in probability as $T \rightarrow \infty$.

The proof of lemma 2 is given in Appendix A. Lemma 2 indicates that $\hat{\mathbf{\Gamma}}_r^{(k)}$ is a consistent estimate for $\mathbf{\Gamma}_r$ uniformly for $r = 1, 2, \ldots, R$. It is easy to see that $\mathbf{\Gamma}_r$, $r = 1, 2, \ldots$, are positive semidefinite. However, matrix $\mathbf{\Gamma}_r$ being singular implies that $\hat{\gamma}_i, i = 0, \ldots, r-1$, are asymptotically linearly dependent random variables. That is not so in general especially when r is not very large. For technical convenience, $\hat{\mathbf{\Gamma}}_r^{(k)}, r = 1, \ldots, R, k = 1, \ldots, M$, are assumed to be non-singular. In our empirical studies, we have never encountered numerical problems in obtaining $(\hat{\mathbf{\Gamma}}_r^{(k)})^{-1}$.

We are now ready to define our test statistic. First, consider an order selection statistic

$$\hat{D}_R^k = \max_{1 \leqslant r \leqslant R} \{T(\hat{\gamma}_r^{(k)} - \hat{\gamma}_r)^{\mathrm{T}} (\hat{\mathbf{\Gamma}}_r^{(k)})^{-1} (\hat{\gamma}_r^{(k)} - \hat{\gamma}) - 2r\}$$

to check whether the first R autocovariances of the kth systematic sample are different from those of the whole time series. This step is analogous to the 'maximum value of estimated risk' considered in Hart (1997), section 7.6.3. The penalty of $2r$ in $\hat{D}_R^k$ is related to the well-known Akaike information criterion. We have the following result.

Theorem 3. Assume that $\mathbf{\Gamma}_r$ is positive definite for $r=1,2,\ldots,$ and $\{X_t\}$ is a stationary process satisfying the conditions in lemma 1. Then we have

$$\max_{1\leqslant k\leqslant M}\left|P(\hat{D}_R^k<x)-P\left\{\sup_{r\geqslant 1}\left(\sum_{i=1}^r e_{k,i}^2-2r\right)<x\right\}\right|\rightarrow 0$$

as $T\rightarrow\infty$, where $\{e_{k,i}, k=1,2,\ldots,i=1,2,\ldots\}$ are all mutually independent standard normal random variables.

Clearly, $\Sigma_{i=1}^r e_{k,i}^2$ above has a χ_r^2-distribution. We use this summation notation to show the dependence structure among $e_{k,1}^2,\ldots,\ \Sigma_{i=1}^{r-1} e_{k,i}^2, \Sigma_{i=1}^r e_{k,i}^2$ and their dependence on k. The dependence is critical to the asymptotic properties of the test statistic under the null hypothesis (such as those in theorems 3 and 4). With another order selection over all M systematic samples, the double-order selection test statistic is defined as

$$\hat{D}_{M,R}=\max_{1\leqslant k\leqslant M}[\max_{1\leqslant r\leqslant R}\{T(\hat{\gamma}_r^{(k)}-\hat{\gamma}_r)^{\mathrm{T}}(\hat{\mathbf{\Gamma}}_r^{(k)})^{-1}(\hat{\gamma}_r^{(k)}-\hat{\gamma})-2r\}-(k-1)^{1/2}].$$

Theorem 4 Under the same conditions as those in theorem 3,

$$\hat{D}_{M,R}\rightarrow\sup_{k\geqslant 1}\left\{\sup_{r\geqslant 1}\left(\sum_{i=1}^r e_{k,i}^2-2r\right)-(k-1)^{1/2}\right\}\tag{9}$$

in distribution.

The proofs of theorems 3 and 4 are given in Appendix A.

The penalty term $(k-1)^{1/2}$ in $\hat{D}_{M,R}$ is selected mainly for two considerations. The first consideration is the asymptotic convergence of $\hat{D}_{M,R}$. Suggested by equation (34) in Appendix A, to guarantee convergence, the penalty term on the kth systematic sample is required to be at least $p_0\log(k)$ asymptotically as $k\rightarrow\infty$, where $p_0>4$ is a constant. The second consideration is the empirical power of the test against alternatives when T is not large. For this consideration $p_0\log(k)$ is not a good choice for the penalty. This is because it would overpenalize the components of the test statistic in the first few systematic samples and thus the empirical power to detect the deviations from the autocovariances on these systematic samples to the autocovariances of the whole time series would be weakened too much. A small and slowly increasing penalty function of k is useful for the first few systematic samples. Since $(k-1)^{1/2}$ satisfies both considerations and provides good empirical power, it is employed in our test.

2.4. Consistency of the test statistic under fixed and local alternatives

We consider the class of zero-mean, locally stationary processes of the form

$$X_{t,T}=\sum_{l=-\infty}^{\infty}\phi_{l,t,T}Z_{t-l}\tag{10}$$

which satisfies assumption 2.1 of Dahlhaus (2009). In particular, one condition in that assumption is

$$\sup_{t,T}|\phi_{l,\,t,\,T}|\leqslant K_0\,v(l),\tag{11}$$

where $v(l)$ was defined in Section 2.2 and K_0 is a constant independent of T. For brevity we refer to Dahlhaus (2009) for other regularity assumptions. When $\phi_{l,t,T}$ are not constants with respect to t, the process is not second-order stationary. The time varying autocovariance function at lag h by the Fourier transformation is

$$\gamma_h(u) = \int_{-\pi}^{\pi} f(\omega, u) \exp(\mathrm{i}\omega h) \mathrm{d}\omega,$$

where $f(\omega, u)$ is the time-dependent spectral density of the underlying process at $u \in (0, 1)$.

An alternative hypothesis of non-stationarity in the framework of locally stationary processes states that

$$H_{\mathrm{a}} : \int_0^1 \left\{ \gamma_h(\mu) - \int_0^1 \gamma_h(s)\,\mathrm{d}s \right\}^2 \mathrm{d}\mu > 0, \qquad \text{for at least one lag } h = 0, 1, 2, \ldots. \tag{12}$$

It essentially says that $\gamma_h(\mu)$ at some lag h is not a constant in the almost everywhere sense under H_{a}. By the completeness of the class of Walsh functions, for any time varying autocovariance function, we have the Walsh representation in the almost everywhere sense

$$\gamma_h(\mu) = \sum_{k=0}^{\infty} w_{h,k}\, W_k(\mu) = \int_0^1 \gamma_h(s)\,\mathrm{d}s + \sum_{k=1}^{\infty} w_{h,k}\, W_k(\mu),$$

where $w_{h,k} = \int_0^1 \gamma_h(u)\, W_k(u)\,\mathrm{d}u$ is the kth Walsh coefficient at lag h. It is not difficult to see that the alternative hypothesis in equation (12) is equivalent to

$$\sum_{h=0}^{\infty} \sum_{k=1}^{\infty} w_{h,k}^2 > 0.$$

Clearly, under any alternative in equation (12), there must be some $h = h_0$ and $k = k_0$ such that $w_{h_0,k_0} \neq 0$. Note that the values h_0 and k_0 depend on the particular alternative hypothesis.

Equation (17) in Dahlhaus (2009) states that

$$\sum_{t=1}^{T} |\mathrm{cov}(X_{t,T}, X_{t+h_0,T}) - \gamma_{h_0}(t/T)| \leqslant K_0.$$

Therefore, it is easy to see that, as $T \to \infty$,

$$E(\hat{\gamma}_{h_0}) = \frac{1}{T} \sum_{t=1}^{T-h_0} \mathrm{cov}(X_{t,T}, X_{t+h_0,T}) = \frac{1}{T} \sum_{t=1}^{T-h_0} \gamma_{h_0}\left(\frac{t}{T}\right) + O\left(\frac{1}{T}\right) \to \int_0^1 \gamma_{h_0}(\mu)\,\mathrm{d}\mu$$

and

$$\begin{aligned} E(\hat{\gamma}_{h_0}^{(k_0)} - \hat{\gamma}_{h_0}) &= \frac{(-1)^{k_0-1}}{T} \sum_{t=1}^{T-h_0} E(X_t X_{t+h_0})\, \mathcal{W}_{k_0}(t) \\ &= \frac{(-1)^{k_0-1}}{T} \sum_{t=1}^{T} \gamma_{h_0}\left(\frac{t}{T}\right) \mathcal{W}_{k_0}(t) + O\left(\frac{h_0}{T}\right) \\ &\to (-1)^{k_0-1} w_{h_0,k_0} \neq 0. \end{aligned} \tag{13}$$

The convergence of $E(\hat{\gamma}_{h_0}^{(k_0)} - \hat{\gamma}_{h_0})$ to a non-zero constant provides power to detect the non-stationarity. In addition, a non-zero Walsh coefficient $w_{h,k} \neq 0$ may indicate a certain pattern of the deviations from a constant at the lag h time varying autocovariance function. For example, if $w_{h,1} \neq 0$, the lag h time varying autocovariance function at $\mu \in [0, 0.5)$ is different from that at $\mu \in [0.5, 1)$.

A sequence of local alternatives indexed by T in the time domain can be expressed as

$$H_{\mathrm{a},T} : \gamma_h(u) = c_h + l_T\, c_{f,h}(u), \qquad h = 0, 1, \ldots, \tag{14}$$

where c_h are constants, $c_{f,h}(u)$ are time varying autocovariance functions satisfying equation (12) and $l_T \to 0$. The sequence of local alternatives converges to the null hypothesis at the rate of

l_T when $T \to \infty$. The following theorem establishes the consistency of the proposed test under the local alternatives. As before, in theorem 5 and the following corollary we assume that R and M satisfy assumption 2 and that $\hat{\boldsymbol{\Gamma}}_r^{(k)}$, $r = 1, \ldots, R$, $k = 1, \ldots, M$, are non-singular.

Theorem 5. Under a sequence of local alternatives in equation (14) and $l_T T^{1/2} \to \infty$, with an estimate $\hat{\kappa}_4$ being $O_p(1)$, we have $\hat{D}_{M,R} \to \infty$ in probability as $T \to \infty$.

The proof of theorem 5 is given in Appendix A. Theorem 5 says that the test proposed is consistent under any sequence of local alternatives if the convergence rate of l_T is slower than $T^{-1/2}$, as was pointed out in Section 1.

From the proof of theorem 5 we can easily obtain the following result for the consistency of the proposed test under any fixed alternative hypothesis.

Corollary 1. Under any fixed H_a in equation (12), with an estimate $\hat{\kappa}_4$ being $O_p(1)$, we have

$$\hat{D}_{M,R} \to \infty$$

in probability as $T \to \infty$.

To estimate κ_4, Paparoditis (2009) proposed

$$\tilde{\kappa}_4 = \frac{2\pi \hat{f}_{2,b_1}(0) - 4\pi \int_{-\pi}^{\pi} \hat{f}_{b_1}^2(\omega)\, \mathrm{d}\omega}{\left\{ \int_{-\pi}^{\pi} \hat{f}_{b_1}(\omega)\, \mathrm{d}\omega \right\}^2},$$

where

$$\hat{f}_{b_1}(\omega) = T^{-1} \sum_{j=1}^{T} K_{b_1}(\omega - \omega_j)\, I_T(\omega_j),$$

$K(\cdot)$ is a kernel function, $K_{b_1}(\omega) = b_1^{-1} K(\cdot/b_1)$ is the scaled kernel, $b_1 > 0$ is a proper bandwidth, $\omega_j = 2\pi j/T$,

$$I_T(\omega) = (2\pi T)^{-1} \left| \sum_{t=1}^{T} X_t \exp(-\mathrm{i}t\omega) \right|^2, \qquad \omega \in (-\pi, \pi),$$

and $\hat{f}_{2,b_1}(\cdot)$ is the smoothed periodogram of the squared process $\{X_t^2\}$. According to Paparoditis (2009), the estimate is consistent under the null hypothesis. Under the alternative and some regularity conditions, according to equation (2.9) in Paparoditis (2009), $\hat{f}_{b_1}$ is a uniformly consistent estimate of the time-averaged local spectral density $\int_0^1 f(\omega, u)\, \mathrm{d}u$. We can see that both $\int_{-\pi}^{\pi} \hat{f}_{b_1}^2(\omega)\, \mathrm{d}\omega$ and $\int_{-\pi}^{\pi} \hat{f}_{b_1}(\omega)\, \mathrm{d}\omega$ are $O_p(1)$. However, $\hat{f}_{2,b_1}(0)$ may not be $O_p(1)$ under the alternative. Therefore, we suggest modifying $\tilde{\kappa}_4$ by replacing $\hat{f}_{2,b_1}(0)$ by its asymptotic equivalence $\hat{f}_2(0)$ under the null, which is based on local periodograms of the squared process and is $O_p(1)$ under the alternative.

One such estimate $\hat{f}_2(0)$ can be found in Kreiss and Paparoditis (2015). Define

$$I_{L,2}(\omega, \mu) = (2\pi L)^{-1} \left| \sum_{t=1}^{L} (X^2_{\lfloor \mu T - L/2 \rfloor + t} - \hat{\mu}_{\lfloor \mu T - L/2 \rfloor + t}) \exp(-\mathrm{i}t\omega) \right|^2,$$

where L is the size of a local window, $\mu \in [L/(2T), 1 - L/(2T)]$ and $\hat{\mu}_v$ is the local average of L observations of the squared process around the vth observation X_v^2. Intuitively, $I_{L,2}(\omega, \mu)$ is the local periodogram around time u for the local centred squared process. By dividing the whole time series into N equal size non-overlap pieces such that $N \to \infty$ and $NT^{-1/2} \to 0$, let $L = \lfloor T/N \rfloor$ and

$$\hat{f}_2(0) = N^{-1} \sum_{l=1}^{N} \sum_{j=1}^{\lfloor L/2 \rfloor} I_{L,2}(w_j, \mu_l)\, K_{b_2}(w_j),$$

where $w_j = 2\pi j/L$, $\mu_l = (l - 0.5)/N$ and $b_2 > 0$ is a proper bandwidth. Lemma 2.1 in Kreiss and Paparoditis (2015) indicates that $\hat{f}_2(0)$ is a consistent estimate for the average spectral density over time at frequency 0 of the squared process under both the null and the alternative hypothesis. Therefore, the resulting estimate

$$\hat{\kappa}_4 = \left\{ 2\pi\, \hat{f}_2(0) - 4\pi \int_{-\pi}^{\pi} \hat{f}_b^2(\omega)\, \mathrm{d}\omega \right\} \bigg/ \left\{ \int_{-\pi}^{\pi} \hat{f}_b(\omega)\, \mathrm{d}\omega \right\}^2 \tag{15}$$

is consistent under the null and $O_p(1)$ under the alternative.

2.5. Algorithm implementation

In the implementation of the procedure, we need to select two appropriate parameters R and M for a given T. The proposed test with a selection of large R and M can detect relatively complex (high frequency) departures from a constant autocovariance. A potential issue is that the empirical distribution of $T^{1/2}(\hat{\gamma}_R^{(k)} - \hat{\gamma}_R)$ may not be sufficiently close to the multivariate normal distribution when R and M are too large with a given T. By the assumptions and our empirical experience, we recommend the use of $R = \lfloor \log_2(T)^{0.99} - 3 \rfloor$ and $M = \lfloor T^{1/3} \rfloor$ for given T, both of which tend to ∞ as $T \to \infty$ and satisfy the rate constraints given in assumption 2. Adjacent values of R and M are also possible. The numerical values of the suggested R and M and some other options are listed in Table 6 in Section 3.2 for a few T-values as an illustration. Our numerical experience indicates that using other values close to the suggested R and M yields similar type I error rates and power; see the last paragraph of Section 3.2 for more discussion. To estimate the nuisance parameter κ_4, we set the length of a local piece $L = T$ if $T \leqslant 255$, and $L = 256$ if $256 \leqslant T \leqslant 1024$, and used the Daniell kernel with bandwidths $b_1 = c_1 T^{-1/3}$ and $b_2 = c_2 L^{-1/3}$ with c_j being 1.2 multiplied by a corresponding crude scale estimate. Values in [1, 1.5] other than 1.2 could also be used for the coefficients with similar results.

Although the asymptotic null distribution of $\hat{D}_{M,R}$ can be obtained with equation (9), it is not easy to calculate the critical values analytically. In addition, a very large T may be required for the asymptotic distribution to be an accurate approximation for the distribution of $\hat{D}_{M,R}$, mainly because the penalty $(k-1)^{1/2}$ on the kth systematic sample is relatively small for the first 10–20 systematic samples. Therefore, we conducted Monte Carlo simulations to obtain the critical values for the test for selected finite samples. In particular, 200000 independent Gaussian white noise time series of length T were generated to determine the empirical distribution of $\hat{D}_{M,R}$ for $T = 64$, 128, 256, 512, 1024. In addition, the known value of $\kappa_4 = 0$ was used. The resulting critical values are reported in on-line supplementary material available from `http://faculty.tamucc.edu/ljin1/supp/suppteststationarity.pdf`. Different from bootstrap methods, the critical values that are obtained here do not depend on the time series data. In the simulation studies, we observe that these critical values are good approximations for our testing purposes.

2.6. Discussion of a test statistic based on auto-correlations

As pointed out by a referee, the test statistic for testing stationarity can be constructed almost in the same way if we use the auto-correlations instead of the autocovariances. Similarly, the difference between the sample auto-correlation of the kth systematic sample and that of the whole time series is

$$\hat{\rho}_h^{(k)} - \hat{\rho}_h = \frac{(-1)^{k-1}}{T\hat{\gamma}_0} \sum_{t=1}^{T-h} X_t X_{t+h} \mathcal{W}_k(t),$$

where $\hat{\rho}_h^{(k)} = \hat{\gamma}_h^{(k)}/\hat{\gamma}_0$ and $\hat{\rho}_h = \hat{\gamma}_h/\hat{\gamma}_0$. Similarly to the proof of theorem 7.2.1 in Brockwell and Davis (1991), page 231, we see that the asymptotic covariance matrix of $\hat{\boldsymbol{\rho}}_R^{(k)} - \hat{\boldsymbol{\rho}}_R = (\hat{\rho}_1^{(k)} - \hat{\rho}_1, \ldots, \hat{\rho}_R^{(k)} - \hat{\rho}_R)^{\mathrm{T}}$ is independent of κ_4. Therefore, one advantage of a test statistic based on auto-correlations is that the calculation does not involve estimating κ_4. However, such a test may not be consistent without considering the joint distribution of $\hat{\gamma}_0$ and $\hat{\boldsymbol{\rho}}_R^{(k)} - \hat{\boldsymbol{\rho}}_R$, which has a covariance matrix involving κ_4 again. Without including γ_0, the test using only auto-correlations cannot detect the non-stationarity when $E(X_t X_{t+h})$, $h = 1, 2, \ldots,$ are all constants over time whereas $\gamma_0(t)$ is time varying.

3. Simulation studies and a real data example

In this section, we consider empirical studies to examine the finite sample performance of the proposed procedure under the null and some alternative hypotheses. We also make extensive empirical comparisons with several state of the art methods such as those of Dette *et al.* (2011), Dwivedi and Subba Rao (2011), Nason (2013a), Paparoditis (2010) and Preuß *et al.* (2013). We used 1000 replications for each case considered below. For easy reading, the significance levels α and the rejection rates are given in percentages in all the tables in this section.

3.1. Empirical type I error rates

In this subsection, we investigate the empirical type I error rates of the test proposed. First we consider time series of length $T = 64$, 128, 256, 512 which were generated from several ARMA models with independent and identical Gaussian innovations. Let ϕ_i be the ith AR parameter and θ_j be the jth MA parameter in the ARMA models. Model I is the white noise model; models II and III are AR(1) models with AR parameter $\phi_1 = 0.9$ and $\phi_1 = -0.9$ respectively; models IV and V are MA(1) models with MA parameter $\theta_1 = 0.8$ and $\theta_1 = -0.8$ respectively; model VI is an AR(2) model with AR parameters $\phi_1 = 0.75$ and $\phi_2 = -0.4$ respectively. The results are shown in Table 1. We observe that the empirical type I error rates for all the cases above are generally close to the corresponding nominal levels.

To see the empirical type I error rates of the proposed test for some heavy-tailed stationary time series, we also considered several ARMA(1,1) time series with various parameters driven by t innovations with degrees of freedom 4, 5 and 9. Dette *et al.* (2011) considered these ARMA(1,1) models driven by Gaussian innovations. Our results are shown in Table 2. Theoretically, the method proposed is not guaranteed to work asymptotically for t_4 since its fourth moment does not exist. We include this case to check the robustness of the test when the regularity conditions are violated. At $T = 64$, the empirical type I error rates for time series driven by these heavy-tailed innovations deviate considerably (not shown in Table 2) from the corresponding nominal levels, especially with t_4- and t_5-innovations. For time series driven by t_4-innovations, the empirical type I error rates for some cases are clearly below the nominal levels when $T = 128$ and they improve when $T = 256$ and $T = 512$. For all the time series driven by t_9 that we considered, the empirical type I error rates are reasonably close to the corresponding nominal levels. This empirical study suggests that the test proposed can have reasonably good type I error rates even when the innovations are relatively heavy tailed when T is reasonably large.

Table 1. Empirical type I error rates

Model	α *(%)*	*Error rates (%) for the following values of T:*			
		$T=64$	$T=128$	$T=256$	$T=512$
I	10	10.1	9.7	9.3	9.4
	5	5.4	4.8	4.5	4.6
	1	0.6	1.1	0.8	0.6
II	10	6.4	6.0	6.1	9.0
	5	3.5	3.7	3.0	5.3
	1	0.6	1.5	1.4	1.6
III	10	10.5	9.3	9.6	9.4
	5	4.8	5.1	5.3	4.7
	1	1.1	1.8	1.5	1.4
IV	10	13.3	11.3	11.1	11.4
	5	7.2	6.4	5.2	5.6
	1	2.5	2.3	1.0	1.3
V	10	13.2	10.3	10.2	11.4
	5	6.6	5.8	5.7	5.1
	1	1.7	1.5	1.3	1.3
VI	10	12.7	11.5	12.3	11.3
	5	6.5	6.5	6.3	6.2
	1	1.8	2.3	1.5	1.2

3.2. Empirical power of the test

To investigate the empirical power of the test proposed, we use the following non-stationary time series models:

$$X_t = 1.1\cos\{1.5 - \cos(4\pi t/T)\} Z_{t-1} + Z_t$$

(model NI);

$$X_t = 0.6\sin(4\pi t/T) X_{t-1} + Z_t$$

(model NII);

$$X_t = \begin{cases} 0.5X_{t-1} + Z_t, & \text{for } \{1 \leqslant t \leqslant T/4\} \cup \{3T/4 < t \leqslant T\}, \\ -0.5X_{t-1} + Z_t, & \text{for } T/4 < t \leqslant 3T/4 \end{cases}$$

(model NIII);

$$X_t = \begin{cases} -0.5X_{t-1} + Z_t, & \text{for } \{1 \leqslant t \leqslant T/2\} \cup \{T/2 + T/64 < t \leqslant T\}, \\ 4Z_t, & \text{for } T/2 < t \leqslant T/2 + T/64 \end{cases}$$

(model NIV).

Here Z_t are independent standard normal random variables. All these models were considered in Dette *et al.* (2011) and models NI–NIII were studied in Paparoditis (2010). First, we make comparisons with the tests in Dette *et al.* (2011) and Nason (2013a).

1000 independent time series of lengths $T = 64$, 128, 256, 512 were generated for each model. The results of models NI–NIV are presented in Table 3. Dette *et al.* (2011) reported the empirical power for some of our settings when $T = 256$ and $T = 512$. We quote their results directly in Table 3 for ease of comparisons. The results for Nason's test were obtained with the function `hwtos2` in R package `locits` version 1.4; see Nason (2013b). The test proposed has much

Table 2. Empirical type I error rates of the proposed test at significance level α for ARMA(1,1) time series with the AR parameter ϕ_1 and the MA parameter θ_1 driven by t-innovations with degrees of freedom df

T	α (%)	Error rates (%) for the following values of (ϕ_1, θ_1):				
		$(-0.5, -0.5)$	$(-0.25, -0.25)$	$(0.5, 0.5)$	$(0.25, -0.25)$	$(0.25, 0.25)$
$df=4$						
128	10	13.1	8.4	12.6	4.0	3.8
	5	8.7	3.5	7.1	1.6	2.3
256	10	12.0	7.5	12.2	7.1	6.8
	5	7.3	3.1	6.8	2.6	2.7
512	10	13.2	7.0	12.5	6.7	7.8
	5	7.4	3.2	6.7	3.0	4.4
$df=5$						
128	10	12.8	8.7	12.5	8.3	9.1
	5	6.6	4.1	7.1	3.7	4.9
256	10	13.5	9.1	10.7	8.5	6.5
	5	7.3	5.0	5.7	3.6	3.4
512	10	9.8	8.6	10.2	8.1	9.8
	5	5.5	3.6	6.2	4.0	4.9
$df=9$						
128	10	11.8	10.0	10.9	6.9	10.2
	5	5.9	4.2	5.0	3.3	5.7
256	10	11.3	9.1	9.6	9.3	10.1
	5	6.4	4.9	5.1	4.0	5.0
512	10	10.2	10.1	10.5	8.6	10.3
	5	5.5	5.8	5.8	4.7	4.9

better power than the tests of Dette *et al.* (2011) and Nason (2013a) in all models except for model NIV. In particular, for models NI–NIII, the empirical power of the proposed test for $T=256$ is comparable with or better than the test of Dette *et al.* (2011) for $T=512$. For model NIII, the power of our test for $T=128$ is also comparable with the test of Dette *et al.* (2011) for $T=256$. Similarly, the empirical power of our test for $T=128$ is better than Nason's test for $T=512$ for models NI–NIII. These three models were studied by Paparoditis (2010) with $T=256$. The empirical power of our test is also higher than the best reported results at different selections of the bandwidth parameter and the rolling window size in Paparoditis (2010) for all these three models (see Table 2 of Paparoditis (2010)). Model NIV has autocovariance structure varying only in a very short time period. For this model, the period of varying autocovariance structure is too short to be significant when $T \leqslant 128$ and thus the empirical power of our test is only slightly higher than the nominal levels. When $T=256$, our results and the results in Dette *et al.* (2011) are close; when $T=512$, our empirical power is noticeably higher than that in Dette *et al.* (2011). In contrast, for model NIV Nason's test had the highest power at significance levels of 0.05 and 0.1 for both $T=256$ and $T=512$ whereas our test outperformed at significance level 0.01. It appears that the test proposed has the advantage in detecting global changes whereas the wavelet method has the advantage in detecting local changes.

Following the Associate Editor's comment, we also compared the proposed test with the test of Dette *et al.* (2011) by using a time varying AR model with parameter b:

$$X_t = 2Z_t - \{1 + b\cos(2\pi t/T)\}\, Z_{t-1}. \tag{16}$$

Table 3. Empirical power at significance level α

Model	α *(%)*	*Powers (%) for the following tests and values of T:*							
		Proposed test				*Dette et al. (2011)*†		*Nason (2013a)*	
		$T=64$	$T=128$	$T=256$	$T=512$	$T=256$	$T=512$	$T=256$	$T=512$
NI	10	53.5	88.2	99.5	100.0	91.9	99.7	37.3	65.4
	5	40.1	79.6	99.3	100.0	82.8	98.6	22.2	46.9
	1	15.6	53.2	92.3	100.0	‡	‡	5.4	18.2
NII	10	39.5	86.0	100.0	100.0	94.3	97.8	41.3	83.4
	5	25.3	79.2	100.0	100.0	68.1	83.0	25.4	65.8
	1	9.2	60.4	98.9	100.0	‡	‡	4.4	29.0
NIII	10	62.3	98.3	100.0	100.0	99.7	100.0	48.7	95.9
	5	47.8	97.1	100.0	100.0	89.9	96.4	31.4	84.8
	1	18.7	90.3	100.0	100.0	‡	‡	7.4	49.1
NIV	10	14.6	16.0	30.1	55.8	33.6	42.2	49.2	86.5
	5	8.3	9.4	21.3	49.0	20.7	27.7	37.9	76.8
	1	3.6	3.4	9.8	30.9	‡	‡	5.3	24.3

†Quoted from Table 4 in Dette *et al.* (2011).
‡Not available.

Table 4. Empirical power for model (16) at significance level α

T	α *(%)*	*Powers (%) for the following tests and hypotheses:*					
		Proposed test			*Dette et al. (2011)*†		
		$H_0: b=0$	$H_a: b=0.5$	$H_a: b=1$	$H_0: b=0$	$H_a: b=0.5$	$H_a: b=1$
256	10	9.6	31.3	90.1	13.8	20.0	46.3
	5	4.3	22.0	82.7	6.2	9.0	26.6
512	10	10.5	70.2	99.9	12.5	23.4	58.2
	5	4.8	58.2	99.8	5.3	11.7	43.0
1024	10	9.6	97.9	100.0	12.2	26.3	82.5
	5	6.0	96.3	100.0	3.8	14.0	69.0

†Quoted from Table 1 in Dette *et al.* (2011).

The process is stationary if and only if $b=0$. Intuitively, the deviation of the process from stationarity increases when b increases. 1000 independent time series of lengths $T=256$, 512, 1024 were generated. The empirical power of the test proposed is given in Table 4, quoting the results in Dette *et al.* (2011) for comparisons. When $b=0$, the type I error rates of our test are all close to that in Dette *et al.* (2011) and to the actual nominal levels, with that of Dette *et al.* (2011) being somewhat higher at the nominal level of 10%. When $b=0.5$ and $b=1$, the power of our test is much higher than that of the test in Dette *et al.* (2011).

To make further comparisons with existing methods, we considered four additional models given in Preuß *et al.* (2013) which compared their test with several existing tests including those in Paparoditis (2010), Dette *et al.* (2011) and Dwivedi and Subba Rao (2011). These models are as follows:

$$X_t = -0.9\sqrt{(t/T)}X_{t-1} + Z_t$$

(model NV);

$$X_t = \begin{cases} 0.5X_{t-1} + Z_t, & \text{for } 1 \leqslant t \leqslant T/2, \\ -0.5X_{t-1} + Z_t, & \text{for } T/2+1 < t \leqslant T \end{cases}$$

(model NVI);

$$X_t = 0.8 \cos\{1.5 - \cos(4\pi t/T)\} Z_{t-1} + Z_t$$

(model NVII);

$$X_t = 0.8 \cos\{1.5 - \cos(4\pi t/T)\} Z_{t-6} + Z_t$$

(model NVIII).

1000 independent time series of lengths $T = 64$, 128, 256 were generated for each model. The results of the test proposed are given in Table 5, in which we also quote the results reported in Preuß *et al.* (2013) for their test using '$M = 8$', the test of Dette *et al.* (2011), the test of Paparoditis (2010) and the test of Dwivedi and Subba Rao (2011) using a maximal lag of 5; see Tables 3 and 5, 6 and 7 in Preuß *et al.* (2013) respectively. In addition, Table 5 includes the results of Nason's

Table 5. Empirical power for models NV–NVIII at significance level α

T	α (%)	*Powers (%) for the following models and tests:*							
		NV	*NVI*	*NVII*	*NVIII*	*NV*	*NVI*	*NVII*	*NVIII*
		Proposed test				*Dette et al. (2011)*			
64	10	27.5	75.8	36.5	33.0	23.2	34.4	35.0	11.6
	5	13.0	62.0	24.6	20.7	18.8	25.0	24.4	5.6
128	10	43.8	99.4	65.9	62.3	33.0	55.2	58.4	33.6
	5	25.9	98.8	49.9	46.0	25.6	37.0	49.0	22.6
256	10	74.5	100.0	94.0	96.0	41.2	92.2	83.6	67.0
	5	60.3	100.0	88.8	91.2	28.2	74.6	74.0	53.2
		Dwivedi and Subba Rao (2011)				*Nason (2013a)*			
64	10	10.0	16.4	12.0	9.8	1.7	0.1	0.2	0.0
	5	5.6	8.2	7.2	4.6	0.9	0.0	0.0	0.0
128	10	11.4	20.8	20.6	16.2	11.7	9.9	8.4	5.9
	5	5.8	12.2	12.6	9.2	7.2	4.4	4.9	2.7
256	10	21.0	27.6	34.0	27.2	35.7	68.6	25.2	25.8
	5	12.8	17.4	23.4	17.4	25.6	42.7	14.5	15.8
		Paparoditis (2010)				*Preuß et al. (2013)*			
64	10	12.2	17.0	10.4	6.4	32.8	27.0	9.8	10.4
	5	5.0	7.8	5.8	3.4	18.6	16.8	4.6	5.2
128	10	26.2	19.8	21.8	14.0	54.6	46.6	15.4	13.0
	5	15.8	11.2	12.8	8.2	39.6	30.8	9.9	7.2
256	10	38.0	44.8	42.8	18.0	81.4	91.2	18.6	16.6
	5	24.8	29.8	28.8	10.2	67.2	74.2	11.0	9.8

test for these time series to have additional comparisons when T is relatively small. For model NV, the test of Preuß *et al.* (2013) had the highest power and the test proposed had the second highest overall. For models NVI–NVIII, the test proposed outperformed all the other five tests.

The aforementioned simulation studies used our default choice of R and M given in Section 2.5. We also experimented with alternative selections of R and M to see their sensitivity to the test proposed. In alternative 1, we set $R = \lfloor \log_2(T) - 2 \rfloor$ and $M = \lfloor T^{1/3} + 2 \rfloor$. This alternative allows us to evaluate the empirical rejection rates by using a slightly larger R and M in the neighbourhood. In alternative 2, we set $R = \lfloor 2T^{1/4} \rfloor$ and $M = \lfloor T^{2/5} \rfloor$. This alternative allows us to check the empirical rejection rates using R and M that have faster growing rates. The values of R and M under these alternative options are given in Table 6. The results given in

Table 6. Values of (R, M) in various options

Option	*(R, M) for the following values of T:*				
	$T = 64$	$T = 128$	$T = 256$	$T = 512$	$T = 1024$
Default	(2, 4)	(3, 5)	(4, 6)	(5, 8)	(6, 10)
Alternative 1	(4, 6)	(5, 7)	(6, 8)	(7, 10)	(8, 12)
Alternative 2	(5, 5)	(6, 6)	(8, 9)	(9, 12)	(11, 16)

Table 7. Empirical power of the proposed test at significance level α by using various R and M

T	*α (%)*	*Powers (%) for the following models:*						
		NI	*NII*	*NIII*	*NIV*	*Time varying AR model*		
						$b = 0$	$b = 0.5$	$b = 1$
Alternative 1: $R = \lfloor log_2(T) - 2 \rfloor$, $M = \lfloor T^{1/3} + 2 \rfloor$								
64	10	50.4	35.5	57.7	14.4	12.7	13.6	18.0
	5	36.2	23.5	43.3	9.3	6.7	7.2	8.5
128	10	87.0	84.9	98.0	16.7	11.1	16.8	44.9
	5	76.7	76.0	96.4	11.0	5.4	8.2	34.0
256	10	99.5	100.0	100.0	30.9	10.2	29.8	89.3
	5	99.1	100.0	100.0	22.9	5.2	21.5	82.7
512	10	100.0	100.0	100.0	61.1	9.5	69.1	99.9
	5	100.0	100.0	100.0	50.9	4.6	57.3	99.8
Alternative 2: $R = \lfloor 2T^{1/4} \rfloor$, $M = \lfloor T^{2/5} \rfloor$								
64	10	52.9	37.6	60.1	15.6	12.1	13.6	20.0
	5	39.4	24.9	45.6	9.3	7.1	7.3	8.3
128	10	88.0	85.3	98.1	16.9	11.5	17.5	46.1
	5	77.6	77.7	96.7	11.0	5.2	8.9	34.8
256	10	99.6	100.0	100.0	31.0	9.8	29.7	89.1
	5	99.1	99.9	100.0	22.8	5.3	20.8	82.1
512	10	100.0	100.0	100.0	60.7	10.7	68.1	99.9
	5	100.0	100.0	100.0	50.8	4.7	56.3	99.8

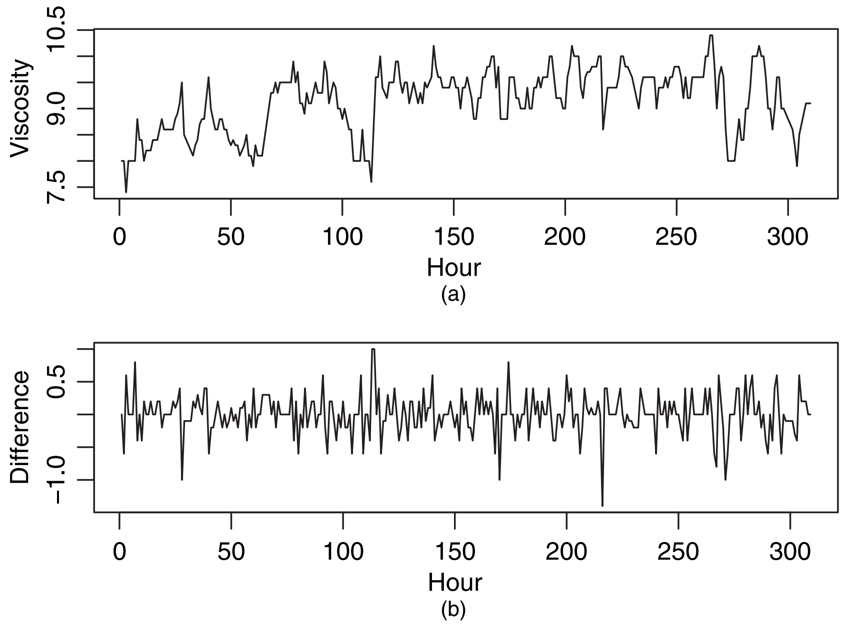

Fig. 2. Hourly chemical process viscosity readings: (a) original data; (b) data after first-order differencing

Table 7 illustrate that the test proposed worked well empirically under both alternative options as well. When $b=0$ in the time varying AR model which is a case for a stationary process, we observe that the type I error rates and power are close to the corresponding nominal levels for different sample sizes in both alternative options. Moreover, under these alternative options the type I error rates and power of the test proposed are quite close to those in the original setting (see Tables 3 and 4), especially when $T \geqslant 128$. Therefore, our test appears to be stable among reasonable choices of R and M.

3.3. *An example*

In this subsection, we illustrate an application of the proposed test to the chemical process viscosity readings data in Box *et al.* (1994). There are 310 hourly chemical process viscosity readings. In Box *et al.* (1994), the data were modelled via an integrated MA model with difference parameter 1 and MA parameter 1. Several analyses (Han and Tsung, 2006) in the quality control area were done by using this data set via control plots with the conclusion that there is a shift in the mean. We would like to see whether there is any change in the autocovariance structure of the time series. We first removed the trend by a differencing operation. Fig. 2 shows the original data and data after first-order differencing. We used $(R, M)=(5, 6)$ for the test. The kurtosis estimate $\hat{\kappa}_4 = 2.57$ indicates that the innovations might be slightly heavier tailed. The test statistic $\hat{D}_{M,R} = 0.085$ is much smaller than the critical value 3.47 at significance level 0.1. Therefore, it is plausible to assume that the differenced time series is stationary.

4. Concluding remarks

In this paper, we proposed a double-order selection test to check the second-order stationarity. The test is based on the deviations of autocovariances of the systematic samples from those of the whole time series. The systematic samples were defined via the class of Walsh functions so that the deviations of these different systematic samples are asymptotically independent. This test is fully non-parametric. The asymptotic null distribution of the test statistic was obtained and the consistency of the test was studied under both fixed and local alternatives. Simulation studies were conducted to examine the finite sample properties of the test. The results indicate that the empirical type I error rates of the test proposed are good for various models even when the length of time series is relatively small or moderate. Furthermore, the test appears generally to have very good power compared with many existing methods, suggesting that it is likely to be a versatile tool in practice. We currently do not have a data-driven procedure for choosing R and M as well as the bandwidths for the Daniell kernel that is used for estimating the nuisance parameter κ_4. It would be an interesting research problem for future study. Finally, the method of defining the test statistic through systematic samples obtained by applying Walsh functions to the original sample coupled with an order selection technique can be generalized to a broad class of problems. For example, the idea can be applied to data from high dimensional analysis-of-variance settings (Wang and Akritas, 2011) in which a large number of levels of some factors exist, or functional data analysis (Wang and Akritas, 2010) to derive more powerful tests.

Acknowledgements

We thank two Joint Editors, an Associate Editor and two referees for their helpful comments and suggestions that have led to a much improved version of this paper. S. Wang's research was partially supported by award KUS-CI-016-04, made by King Abdullah University of Science and Technology. H. Wang's research was partially supported by a grant from the Simons Foundation (246077). Part of the work was carried out while S. Wang was visiting the Australian National University supported by the Mathematical Sciences Research Visitor Programme.

Appendix A: Technical details

A.1. Lemma for the properties of discrete Walsh functions

Lemma 3. Let k, k_1 and k_2 be positive integers which are less than $T/2$.

(a) $|\Sigma_{t=1}^{T} \mathcal{W}_k(t)| \leqslant k+1$. If N_k divides T, then $\Sigma_{t=1}^{T} \mathcal{W}_k(t)=0$.
(b) For any k_1, $k_2 < T$ with $k_1 \neq k_2$, $|\Sigma_{t=1}^{T} \mathcal{W}_{k_1}(t)\mathcal{W}_{k_2}(t)| \leqslant k_1+k_2+1$.
(c) For $d=0, 1, \ldots, \lfloor T/(2N_k)\rfloor$, where $\lfloor T/N_k\rfloor$ is the minimum block size in $\boldsymbol{\mathcal{W}}_k$,

$$\sum_{t=d+1}^{T} \mathcal{W}_k(t)\mathcal{W}_k(t-d) = T-(2k+1)d.$$

(d) For $d=0, 1, \ldots$, and k_1, $k_2 < T$ with $k_1 \neq k_2$,

$$\left|\sum_{t=d+1}^{T} \mathcal{W}_{k_1}(t)\mathcal{W}_{k_2}(t-d)\right| \leqslant 2(k_1+k_2+1)(d+1).$$

Proof.

(a) The $W_k(x)$ for $x \in [0, 1)$ is a step function of k switches of values. Let $\tau_{k,0}=0$, $\tau_{k,k+1}=1$ and $0<\tau_{k,1}<\ldots<\tau_{k,k}<1$ be the k points of value changing such that $W_k(x)=(-1)^s$ if $x \in [\tau_{k,s}, \tau_{k,s+1})$ for $s=0, \ldots, k$. By the definition of a continuous Walsh function, it is clear that $\int_0^1 W_k(x)\,\mathrm{d}x=0$ and $\tau_{k,s+1}-\tau_{k,s} \geqslant N_k^{-1}$. Since $T>2k$, at least one integer $\lfloor T\tau_{k,s}\rfloor$ is in the interval $[T\tau_{k,s-1}, T\tau_{k,s})$. Let $\lceil a\rceil$ be the smallest integer not less than a. Since $\mathcal{W}_k(t)=W\{(t-1)/T\}$, $0<c_1<\ldots<c_k \leqslant T$, where $c_s=\lceil T\tau_{k,s}\rceil+1$, are the corresponding sign change points for the discrete Walsh function $\boldsymbol{\mathcal{W}}_k$. We have

$$T\int_0^1 W_k(x)\,\mathrm{d}x=\sum_{s=1}^{k+1} W_k(\tau_{k,s-1})(T\tau_{k,s}-T\tau_{k,s-1})$$

and

$$\sum_{t=1}^{T}\mathcal{W}_k(t)=\sum_{s=1}^{k+1} W_k(\tau_{k,s-1})(c_s-c_{s-1}),$$

where $c_0=0$ and $c_{k+1}=T+1$. Therefore,

$$\begin{aligned}\left|\sum_{t=1}^{T}\mathcal{W}_k(t)\right|&=\left|\sum_{t=1}^{T}\mathcal{W}_k(t)-T\int_0^1 W_k(x)\,\mathrm{d}x\right|\\&=\left|\sum_{s=1}^{k+1} W_k(\tau_{k,s-1})(c_k-c_{k-1})-\sum_{s=1}^{k+1} W_k(\tau_{k,s-1})(T\tau_{k,s}-T\tau_{k,s-1})\right|\\&\leqslant\left|\sum_{s=1}^{k+1} W_k(\tau_{k,s-1})(\lceil T\tau_{k,s}\rceil-\lceil T\tau_{k,s-1}\rceil)-\sum_{s=1}^{k+1} W_k(\tau_{k,s-1})(T\tau_{k,s}-T\tau_{k,s-1})\right|\\&\leqslant\sum_{s=1}^{k+1}\left|\lceil T\tau_{k,s}\rceil-T\tau_{k,s}-(\lceil T\tau_{k,s-1}\rceil-T\tau_{k,s-1})\right|\leqslant k+1.\end{aligned}$$

By the recursive definition of continuous Walsh functions, for any $s=1,\ldots,k$, $\tau_{k,s}=l_s/N_k$ where $1\leqslant l_s\leqslant N_k-1$ is an integer depending s and k. If N_k divides T, then $\lceil T\tau_{k,s}\rceil=T\tau_{k,s}$ and $\lceil T\tau_{k,s-1}\rceil=T\tau_{k,s-1}$. Therefore, $\Sigma_{t=1}^{T}\mathcal{W}_k(t)=0$.

(b) There are at most k_1+k_2 points where signs change in function $W_{k_1}(x)\,W_{k_2}(x)$, $x\in[0,1)$, and its discrete version $\mathcal{W}_{k_1}(t)\,\mathcal{W}_{k_2}(t), t=1,\ldots,T$. Similarly to the proof of part (a), we have

$$\left|\sum_{t=1}^{T}\mathcal{W}_{k_1}(t)\,\mathcal{W}_{k_2}(t)\right|=\left|\sum_{t=1}^{T}\mathcal{W}_{k_1}(t)\,\mathcal{W}_{k_2}(t)-T\int_0^1 W_{k_1}(x)\,W_{k_2}(x)\,\mathrm{d}x\right|\leqslant k_1+k_2+1.$$

(c) There are k sign change points $c_1,c_2,\ldots,c_k$ of $\boldsymbol{\mathcal{W}}_k$. Under the condition that d is no more than half of the minimum block size, there are only d values of $\mathcal{W}_k(t)\,\mathcal{W}_k(t-d)=-1$ that are different from $\mathcal{W}_k(t)\,\mathcal{W}_k(t)=1$ at $t=c_i,c_i+1,\ldots,c_i+d-1$ in the $(i+1)$th block, $i=1,2,\ldots,k$. Therefore,

$$\sum_{t=d+1}^{T}\mathcal{W}_k(t)\,\mathcal{W}_k(t-d)-T=\sum_{t=d+1}^{T}\mathcal{W}_k(t)\,\mathcal{W}_k(t-d)-\sum_{t=d+1}^{T}\mathcal{W}_k(t)\,\mathcal{W}_k(t)-d=-(2k+1)d.$$

Hence, we have $\Sigma_{t=d+1}^{T}\mathcal{W}_k(t)\,\mathcal{W}_k(t-d)=T-(2k+1)d$.

(d) In total, there are no more than k_1+k_2 sign changes of the values in $\mathcal{W}_{k_1}(t)\,\mathcal{W}_{k_2}(t-d)$, $t=d+1,d+2,\ldots,T$. Between a sign change point and the next sign change point, there are no more than d values of $\mathcal{W}_{k_1}(t)\,\mathcal{W}_{k_2}(t-d)=-1$ that are different from $\mathcal{W}_{k_1}(t)\,\mathcal{W}_{k_2}(t)=1$. Therefore,

$$\begin{aligned}\left|\sum_{t=d+1}^{T}\mathcal{W}_{k_1}(t)\,\mathcal{W}_{k_2}(t-d)\right|&=\left|\sum_{t=d+1}^{T}\mathcal{W}_{k_1}(t)\,\mathcal{W}_{k_2}(t-d)-\sum_{t=1}^{T}\mathcal{W}_{k_1}(t)\,\mathcal{W}_{k_2}(t)\right|+\left|\sum_{t=1}^{T}\mathcal{W}_{k_1}(t)\,\mathcal{W}_{k_2}(t)\right|\\&\leqslant 2(k_1+k_2+1)(d+1).\end{aligned}$$

A.2. Lemma for the proof of theorem 1

Lemma 4. Under the conditions in theorem 1, $d>0$ and $d/R\to\infty$ as $T\to\infty$, we have

$$\max_{k_1\leqslant k_2\leqslant M, i\leqslant j\leqslant R}\frac{1}{T}\sum_{\delta=-(T-1)}^{T-1}\left|V(\delta,i,j)\{1-I_d(\delta)\}\sum_{t_1=|\delta|+1}^{T}\mathcal{W}_{k_1}(t_1)\,\mathcal{W}_{k_2}(t_1-|\delta|)\right|\to 0,$$

where $V(\delta,i,j)=\gamma_\delta\gamma_{\delta-i+j}+\gamma_{\delta+i}\gamma_{\delta-j}+\kappa_4\sigma^4\,\Sigma_l\,\phi_l\phi_{l+i}\phi_{l+\delta}\phi_{l+\delta+j}$, and $I_d(t)=1$ if $|t|<d$ and $I_d(t)=0$ otherwise.

Proof. Clearly, the result is true when $d\geqslant T$. For convenience, we assume that $d<T$. Let $\phi_l=0$ if $l<0$. We always have

$$
\begin{aligned}
4\sum_{l}\phi_l\phi_{l+i}\phi_{l+\delta}\phi_{l+\delta+j} &\leqslant \sum_{l}(\phi_l^2+\phi_{l+i}^2)(\phi_{l+\delta}^2+\phi_{l+\delta+j}^2)\\
&=\sum_{l}(\phi_l^2\phi_{l+\delta}^2+\phi_l^2\phi_{l+\delta+j}^2+\phi_{l+i}^2\phi_{l+\delta}^2+\phi_{l+i}^2\phi_{l+\delta+j}^2)\\
&=\sum_{l}\phi_l^2\phi_{l+\delta}^2+\sum_{l}\phi_l^2\phi_{l+\delta+j}^2+\sum_{l}\phi_{l+i}^2\phi_{l+\delta}^2+\sum_{l}\phi_{l+i}^2\phi_{l+\delta+j}^2\\
&\leqslant \sum_{l}\phi_l^2\phi_{l+\delta}^2+\sum_{l}\phi_l^2\phi_{l+\delta+j}^2+\sum_{l}\phi_l^2\phi_{l+\delta-i}^2+\sum_{l}\phi_l^2\phi_{l+\delta-i+j}^2.
\end{aligned}
$$

Let $G_\phi=\sigma^4K_0^4|\kappa_4|/4$, where K_0 is given in assumption 1. Since

$$
\left|\sum_{t_1=|\delta|+1}^{T}\mathcal{W}_{k_1}(t_1)\mathcal{W}_{k_2}(t_1-|\delta|)\right|\leqslant T
$$

for $1\leqslant k_1\leqslant k_2$, it is readily seen that

$$
\begin{aligned}
\max_{k_1\leqslant k_2\leqslant M, i\leqslant j\leqslant R}\frac{1}{T}\sum_{\delta=-(T-1)}^{T-1}&\left|V(\delta,i,j)\{1-I_d(\delta)\}\sum_{t_1=|\delta|+1}^{T}\mathcal{W}_{k_1}(t_1)\mathcal{W}_{k_2}(t_1-|\delta|)\right|\\
&\leqslant \max_{i\leqslant j\leqslant R}\sum_{\delta=-(T-1)}^{T-1}|V(\delta,i,j)|\{1-I_d(\delta)\}=\max_{i\leqslant j\leqslant R}\sum_{|\delta|=d}^{T-1}|V(\delta,i,j)|\leqslant I_{\mathrm{a}}+I_{\mathrm{b}}+I_{\mathrm{c}},
\end{aligned}
$$

where $I_{\mathrm{a}}=\gamma_0(\Sigma_{\delta=-(T-1)}^{-d}+\Sigma_{\delta=d}^{T-1})|\gamma_\delta|$, $I_{\mathrm{b}}=\gamma_0(\Sigma_{\delta=-(T-1)}^{-d+R}+\Sigma_{\delta=d-R}^{T-1})|\gamma_\delta|$ and

$$
I_{\mathrm{c}}=\max_{i\leqslant j\leqslant R}G_\phi\sum_{d\leqslant|\delta|\leqslant T-1}\left(\sum_{l}\phi_l^2\phi_{l+\delta}^2+\sum_{l}\phi_l^2\phi_{l+\delta+j}^2+\sum_{l}\phi_l^2\phi_{l+\delta-i}^2+\sum_{l}\phi_l^2\phi_{l+\delta-i+j}^2\right).
$$

Next we show that $\max_{i\leqslant j\leqslant R}\Sigma_{d\leqslant|\delta|\leqslant T-1}\Sigma_l\phi_l^2\phi_{l+\delta-i+j}^2\to 0$ as $T\to\infty$. When T is sufficiently large so that d is much larger than R,

$$
\begin{aligned}
\max_{i\leqslant j\leqslant R}\sum_{|\delta|=d}^{T-1}\sum_{l}\phi_l^2\phi_{l+\delta+j-i}^2&=\max_{i\leqslant j\leqslant R}\sum_{l}\left(\phi_l^2\sum_{|\delta|=d}^{T-1}\phi_{l+\delta+j-i}^2\right)\\
&=\max_{i\leqslant j\leqslant R}\sum_{|l|\leqslant d/2}\left(\phi_l^2\sum_{|\delta|=d}^{T-1}\phi_{l+\delta+j-i}^2\right)+\max_{i\leqslant j\leqslant R}\sum_{|l|>d/2}\left(\phi_l^2\sum_{|\delta|=d}^{T-1}\phi_{l+\delta+j-i}^2\right)\\
&\leqslant \max_{i\leqslant j\leqslant R}K_0^2\sum_{|l|\leqslant d/2}\left\{\phi_l^2\sum_{|\delta|=d}^{\infty}v^2(l+\delta+j-i)\right\}+K_0^2\left(\sum_{k}\phi_k^2\right)\sum_{|l|>d/2}v^2(l)\\
&\leqslant \frac{4}{d-2R}\sum_{|l|\leqslant d/2}\phi_l^2+\frac{4}{d}\left(\sum_{l}\phi_l^2\right)K_0^2\to 0
\end{aligned}
$$

as $T\to\infty$, because $|l+\delta-i+j|\geqslant(d-2R)/2$ given that $|l|\leqslant d/2$, $\Sigma_l\phi_l^2<\infty$, $d\to\infty$, and $R/d\to 0$ as $T\to\infty$. Similarly, we have $\Sigma_{d\leqslant|\delta|\leqslant T-1}\Sigma_l\phi_l^2\phi_{l+\delta}^2\to 0$, $\max_{i\leqslant j\leqslant R}\sum_{d\leqslant|\delta|\leqslant T-1}\sum_l\phi_l^2\phi_{l+\delta+j}^2\to 0$ and $\max_{i\leqslant j\leqslant R}\Sigma_{d\leqslant|\delta|\leqslant T-1}\Sigma_l\phi_l^2\phi_{l+\delta-i}^2\to 0$ as $T\to\infty$. Therefore, $I_{\mathrm{c}}\to 0$ as $T\to\infty$. Since $\Sigma_\delta|\gamma_\delta|<\infty$ under assumption 1, it is easy to see that $I_{\mathrm{a}}\to 0$ and $I_{\mathrm{b}}\to 0$ as $d/R\to\infty$. Hence, $I_{\mathrm{a}}+I_{\mathrm{b}}+I_{\mathrm{c}}\to 0$ as $d/R\to\infty$ and the proof is completed.

A.3. Proof of theorem 1

Let $\tilde{\gamma}_h=T^{-1}\Sigma_{t=1}^{T}X_tX_{t+h}$. It is easily seen that, under assumptions 1 and 2,

$$
\begin{aligned}
\max_{1\leqslant k\leqslant M,\,h\leqslant R}T^{1/2}|(\hat{\gamma}_h^{(k)}-\hat{\gamma}_h)-(\hat{\gamma}_h^{(k)}-\tilde{\gamma}_h)|&=\max_{h\leqslant R}T^{-1/2}\left|\sum_{t=T-h}^{T}X_tX_{t+h}\right|\\
&\leqslant T^{-1/2}\sum_{t=T-R}^{T+R}X_t^2=O_p\{\log(T)T^{-1/2}\}.
\end{aligned}
$$

Furthermore, we have

$$
\max_{1\leqslant k_1\leqslant k_2\leqslant M,h\leqslant R}|T\operatorname{cov}(\hat{\gamma}_i^{(k_1)}-\hat{\gamma}_i,\hat{\gamma}_j^{(k_2)}-\hat{\gamma}_j)-T\operatorname{cov}(\hat{\gamma}_i^{(k_1)}-\tilde{\gamma}_i,\hat{\gamma}_j^{(k_2)}-\tilde{\gamma}_j)|=O_p\{\log(T)T^{-1/2}\}. \tag{17}
$$

Therefore, we can use $\hat{\gamma}_h^{(k)} - \tilde{\gamma}_h = (-1)^{k-1} T^{-1} \Sigma_{t=1}^T X_t X_{t+h} \mathcal{W}_k(t)$ to facilitate the development of asymptotic properties of $\hat{\gamma}_h^{(k)} - \tilde{\gamma}_h$. We have

$$\begin{aligned}(-1)^{k_1+k_2} \operatorname{cov}(\hat{\gamma}_i^{(k_1)} - \tilde{\gamma}_i, \hat{\gamma}_j^{(k_2)} - \tilde{\gamma}_j) &= \operatorname{cov}\left\{ \frac{1}{T} \sum_{t=1}^T X_t X_{t+i} \mathcal{W}_{k_1}(t), \frac{1}{T} \sum_{t=1}^T X_t X_{t+j} \mathcal{W}_{k_2}(t) \right\} \\ &= E\left\{ \frac{1}{T^2} \sum_{t_1=1}^T \sum_{t_2=1}^T X_{t_1} X_{t_1+i} X_{t_2} X_{t_2+j} \mathcal{W}_{k_1}(t_1) \mathcal{W}_{k_2}(t_2) \right\} \\ &\quad - E\left\{ \frac{1}{T} \sum_{t=1}^T X_t X_{t+i} \mathcal{W}_{k_1}(t) \right\} E\left\{ \frac{1}{T} \sum_{t=1}^T X_t X_{t+j} \mathcal{W}_{k_2}(t) \right\}.\end{aligned}$$

Brockwell and Davis (1991), page 227, showed that

$$E(X_{t_1} X_{t_1+i} X_{t_2} X_{t_2+j}) = \gamma_i \gamma_j + \gamma_{t_2-t_1} \gamma_{t_2-t_1-i+j} + \gamma_{t_2-t_1+i} \gamma_{t_2-t_1-j} + \kappa_4 \sigma^4 \sum_{l=-\infty}^{\infty} \phi_l \phi_{l+i} \phi_{l+t_2-t_1} \phi_{l+t_2-t_1+j}. \tag{18}$$

It is readily seen that

$$\sum_{t_1=1}^T \sum_{t_2=1}^T \gamma_i \gamma_j \mathcal{W}_{k_1}(t_1) \mathcal{W}_{k_2}(t_2) - E\left\{ \frac{1}{T} \sum_{t=1}^T X_t X_{t+i} \mathcal{W}_{k_1}(t) \right\} E\left\{ \frac{1}{T} \sum_{t=1}^T X_t X_{t+j} \mathcal{W}_{k_2}(t) \right\} = 0.$$

Therefore, by equation (18),

$$(-1)^{k_1+k_2} \operatorname{cov}(\hat{\gamma}_i^{(k_1)} - \tilde{\gamma}_i, \hat{\gamma}_j^{(k_2)} - \tilde{\gamma}_j) = \frac{1}{T^2} \sum_{t_1=1}^T \sum_{t_2=1}^T \{ V(t_2 - t_1, i, j) \mathcal{W}_{k_1}(t_1) \mathcal{W}_{k_2}(t_2) \},$$

where $V(\delta, i, j)$ is given in lemma 4. Interchanging the order of the double sum, with $\delta = t_2 - t_1$ we obtain

$$\begin{aligned}(-1)^{k_1+k_2} T \operatorname{cov}(\hat{\gamma}_i^{(k_1)} - \tilde{\gamma}_i, \hat{\gamma}_j^{(k_2)} - \tilde{\gamma}_j) &= \frac{1}{T} \left(\sum_{\delta=-T+1}^{0} \sum_{t_1=1-\delta}^{T} + \sum_{\delta=1}^{T-1} \sum_{t_1=1}^{T-\delta} \right) \left\{ V(\delta, i, j) \mathcal{W}_{k_1}(t_1) \mathcal{W}_{k_2}(t_1 + \delta) \right\} \\ &= \frac{1}{T} \left\{ \sum_{\delta=-T+1}^{0} V(\delta, i, j) \sum_{t_1=1-\delta}^{T} \mathcal{W}_{k_1}(t_1) \mathcal{W}_{k_2}(t_1 + \delta) \right. \\ &\quad \left. + \sum_{\delta=1}^{T-1} V(\delta, i, j) \sum_{t_1=1}^{T-\delta} \mathcal{W}_{k_1}(t_1) \mathcal{W}_{k_2}(t_1 + \delta) \right\} \\ &= \frac{1}{T} \left\{ \sum_{\delta=-T+1}^{0} V(\delta, i, j) \sum_{t_1=1+|\delta|}^{T} \mathcal{W}_{k_1}(t_1) \mathcal{W}_{k_2}(t_1 - |\delta|) \right. \\ &\quad \left. + \sum_{\delta=1}^{T-1} V(\delta, i, j) \sum_{t_2=1+|\delta|}^{T} \mathcal{W}_{k_1}(t_2 - |\delta|) \mathcal{W}_{k_2}(t_2) \right\} = \frac{1}{T} (I_1 + I_2 + I_3 + I_4),\end{aligned} \tag{19}$$

where

$$\begin{aligned} I_1 &= \sum_{\delta=-d}^{0} \{ V(\delta, i, j) \sum_{t_1=1+|\delta|}^{T} \mathcal{W}_{k_1}(t_1) \mathcal{W}_{k_2}(t_1 - |\delta|) \}, \\ I_2 &= \sum_{\delta=1}^{d} \{ V(\delta, i, j) \sum_{t_2=1+|\delta|}^{T} \mathcal{W}_{k_1}(t_2 - |\delta|) \mathcal{W}_{k_2}(t_2) \}, \\ I_3 &= \sum_{\delta=-(T-1)}^{-d-1} \{ V(\delta, i, j) \sum_{t_1=1+|\delta|}^{T} \mathcal{W}_{k_1}(t_1) \mathcal{W}_{k_2}(t_1 - |\delta|) \}, \\ I_4 &= \sum_{\delta=d+1}^{T-1} \{ V(\delta, i, j) \sum_{t_2=1+|\delta|}^{T} \mathcal{W}_{k_1}(t_2 - |\delta|) \mathcal{W}_{k_2}(t_2) \} \end{aligned}$$

and integer d is selected such that $d^2 M/T \to 0$ and $d/R \to \infty$ as $T \to \infty$. By lemma 4, $\max_{k_1 < k_2 \leqslant M,\, i \leqslant j \leqslant R} (I_3 + I_4)/T \to 0$. By lemma 3, part (d),

$$\max_{k_1<k_2\leqslant M,\, i\leqslant j\leqslant R}(|I_1|+|I_2|)\leqslant 2\sum_{\delta=-d}^{d}|V(\delta,i,j)(k_1+k_2+1)(|\delta|+1)|$$
$$\leqslant (k_1+k_2+1)(d+1)(d+2)V_{\sup}\leqslant 2M(d+1)(d+2)V_{\sup},$$

where $V_{\sup}=\sup\{|V(\delta,i,j)|,\delta=0,\pm1,\pm2,\ldots\}$, which is finite since $\Sigma_{\delta=-\infty}^{\infty}V(\delta,i,j)<\infty$ as shown in the proof of proposition 7.3.1 in Brockwell and Davis (1991). Hence, $\max_{k_1<k_2\leqslant M,\, i\leqslant j\leqslant R}(|I_1|+|I_2|)/T\to 0$ as $T\to\infty$, with equation (17), leading to

$$\max_{1\leqslant k_1<k_2\leqslant M,\, 0\leqslant i\leqslant j\leqslant R}|T\operatorname{cov}(\hat{\gamma}_i^{(k_1)}-\hat{\gamma}_i,\hat{\gamma}_j^{(k_2)}-\hat{\gamma}_j)|\to 0.$$

A.4. Proof of theorem 2

Similarly to equation (19), it is readily seen that, for $\delta=t_2-t_1$,

$$T\operatorname{cov}(\hat{\gamma}_i^{(k)}-\tilde{\gamma}_i,\hat{\gamma}_j^{(k)}-\tilde{\gamma}_j)=\frac{1}{T}\sum_{\delta=-(T-1)}^{T-1}\left\{V(\delta,i,j)\sum_{t_1=1+|\delta|}^{T}\mathcal{W}_k(t_1)\mathcal{W}_k(t_1-|\delta|)\right\}.$$

Let integer d be selected such that $d/T\to 0$ and $d/R\to\infty$ as $T\to\infty$. Then we have

$$T\operatorname{cov}(\hat{\gamma}_i^{(k)}-\tilde{\gamma}_i,\hat{\gamma}_j^{(k)}-\tilde{\gamma}_j)=\frac{1}{T}\sum_{\delta=-(T-1)}^{T-1}\left\{V(\delta,i,j)\,I_d(\delta)\sum_{t_1=1+|\delta|}^{T}W_k(t_1)\,W_k(t_1-|\delta|)\right\}$$
$$+\frac{1}{T}\sum_{\delta=-(T-1)}^{T-1}\left\{V(\delta,i,j)\{1-I_d(\delta)\}\sum_{t_1=1+|\delta|}^{T}W_k(t_1)\,W_k(t_1-|\delta|)\right\}.$$

By lemma 3, part (c),

$$\frac{1}{T}\sum_{\delta=-(T-1)}^{T-1}\left\{V(\delta,i,j)\,I_d(\delta)\sum_{t_1=1+|\delta|}^{T}W_k(t_1)\,W_k(t_1-|\delta|)\right\}=\sum_{\delta=-d}^{d}V(\delta,i,j)\left\{1-\frac{(2k+1)|\delta|}{T}\right\}.$$

Moreover, by lemma 4,

$$\frac{1}{T}\sum_{\delta=-(T-1)}^{T-1}\left[V(\delta,i,j)\{1-I_d(\delta)\}\sum_{t_1=1+|\delta|}^{T}W_k(t_1)\,W_k(t_1-|\delta|)\right]\to 0.$$

Therefore, we obtain

$$\lim_{T\to\infty}T\operatorname{cov}(\hat{\gamma}_i^{(k)}-\tilde{\gamma}_i,\hat{\gamma}_j^{(k)}-\tilde{\gamma}_j)=\lim_{T\to\infty}\sum_{\delta=-d}^{d}\left[V(\delta,i,j)\left\{1-\frac{(2k+1)|\delta|}{T}\right\}\right]=\sum_{\delta=-\infty}^{\infty}V(\delta,i,j).$$

By equation (7.3.5) in Brockwell and Davis (1991), the limit result immediately after that equation, and equation (17), we have

$$\lim_{T\to\infty}T\operatorname{cov}(\hat{\gamma}_i^{(k)}-\hat{\gamma}_i,\hat{\gamma}_j^{(k)}-\hat{\gamma}_j)=\kappa_4\gamma_i\gamma_j+\sum_{\delta=-\infty}^{\infty}(\gamma_\delta\gamma_{\delta-i+j}+\gamma_{\delta+i}\gamma_{\delta-j}).$$

A.5. Proof of lemma 1

Divide $\mathcal{T}$ into N_M blocks such that the lth block is

$$B_l=\{\lfloor(l-1)T/N_M\rfloor+1,\ldots,\lfloor lT/N_M\rfloor\}\qquad\text{for } l=1,\ldots,N_M.$$

Let $|B_l|$ denote the length of B_l so that it is either $\nu_{T,M}=\lfloor T/N_M\rfloor$ or $\lfloor T/N_M\rfloor+1$ depending on l. For each B_l, the values of $\mathcal{W}_k(t)$, $t\in B_l$, all take on only one value of either -1 or 1 depending on k and l and we call the value $\mathcal{W}_{k,l}$. We also divide B_l into two parts, $B_{l,1}=\{\lfloor(l-1)T/N_M\rfloor+1,\ldots,\lfloor(l-1)T/N_M\rfloor+\nu_{T,M,1}\}$ and $B_{l,2}=\{\lfloor(l-1)T/N_M\rfloor+\nu_{T,M,1}+1,\ldots,\lfloor lT/N_M\rfloor\}$, where $\nu_{T,M,1}=\lfloor T^{a_0}\rfloor$ if $|B_l|=\nu_{T,M}$ and $\nu_{T,M,1}=\lfloor T^{a_0}\rfloor+1$ if $|B_l|=\nu_{T,M}+1$ for some $a_0>0$ satisfying $T^{-a_0}M<\infty$ and $T^{2a_0}/\nu_{T,M}\to 0$. Intuitively, by this division scheme, blocks $B_{l-1,2}$ and $B_{l,2}$ are separated by block $B_{l,1}$ of minimum length $\lfloor T^{a_0}\rfloor$ which is a negligible part of block B_l. In addition, let $|B_{l,1}|$ and $|B_{l,2}|$ be the length of $B_{l,1}$ and $B_{l,2}$ respectively. It is easy to see that $|B_{l,2}| = \nu_{T,M} - \lfloor T^{a_0} \rfloor$.

Let $\tilde{\gamma}_{h,B_l} = |B_{l,2}|^{-1} \Sigma_{t\in B_l} X_t X_{t+h}$, $\tilde{\eta}_{h,B_l} = |B_{l,2}|^{-1} \Sigma_{t\in B_l} (X_t X_{t+h} - \gamma_h)$, $\tilde{\eta}_{h,B_{l,1}} = |B_{l,2}|^{-1} \Sigma_{t\in B_{l,1}} (X_t X_{t+h} - \gamma_h)$ and $\tilde{\eta}_{h,B_{l,2}} = |B_{l,2}|^{-1} \Sigma_{t\in B_{l,2}} (X_t X_{t+h} - \gamma_h)$. The corresponding vectors are $\tilde{\boldsymbol{\gamma}}_{r,B_l} = (\tilde{\gamma}_{0,B_l}, \tilde{\gamma}_{1,B_l}, \ldots, \tilde{\gamma}_{r-1,B_l})^{\mathrm{T}}$ $\tilde{\boldsymbol{\eta}}_{r,B_{l,1}} = (\tilde{\eta}_{0,B_{l,1}}, \tilde{\eta}_{1,B_{l,1}}, \ldots, \tilde{\eta}_{r-1,B_{l,1}})^{\mathrm{T}}$ and $\tilde{\boldsymbol{\eta}}_{r,B_{l,2}} = (\tilde{\eta}_{0,B_{l,2}}, \tilde{\eta}_{1,B_{l,2}}, \ldots, \tilde{\eta}_{r-1,B_{l,2}})^{\mathrm{T}}$. Actually, $\tilde{\boldsymbol{\eta}}_{R,B_{l,1}}$ and $\tilde{\boldsymbol{\eta}}_{R,B_{l,2}}$ are almost the same as $\boldsymbol{\eta}_l$ and $\boldsymbol{\eta}_l^0$ defined in equations (3.1) and (3.2) in Keenan (1997). In addition, similarly to $\boldsymbol{\zeta}_p^0$, $\boldsymbol{\zeta}_p$ and $\hat{\boldsymbol{\zeta}}_p$ in equation (3.4) in Keenan (1997), we define $\boldsymbol{\xi}^{1,k} = \Sigma_{l=1}^{N_M} \tilde{\boldsymbol{\eta}}_{R,B_{l,1}} \mathcal{W}_{k,l}$, $\boldsymbol{\xi}^{2,k} = \Sigma_{l=1}^{N_M} \tilde{\boldsymbol{\eta}}_{R,B_{l,2}} \mathcal{W}_{k,l}$ and $\boldsymbol{\xi}^{k,*} = \Sigma_{l=1}^{N_M} \boldsymbol{\eta}_l^* \mathcal{W}_{k,l}$, where $\boldsymbol{\eta}_l^*$ are IID random vectors of $\tilde{\boldsymbol{\eta}}_{R,B_{1,2}}$. The distributions of $\boldsymbol{\xi}^{k,*}$, $-\boldsymbol{\xi}^{k,*}$ and $(-1)^{k-1}\boldsymbol{\xi}^{k,*}$ are all the same, which do not depend on k or the sign, since $\boldsymbol{\xi}^{k,*}$ and $-\boldsymbol{\xi}^{k,*}$ are always the sum of $N_M/2$ random vectors of $\tilde{\boldsymbol{\eta}}_{R,B_{1,2}}$ and $N_M/2$ IID random vectors of $-\tilde{\boldsymbol{\eta}}_{R,B_{1,2}}$ respectively.

Considering the length of block $B_{1,2}$, we have $R = O[\log(|B_{1,2}|)/\log\{\log(|B_{1,2}|)\}]$ since $R = O[\log(T)/\log\{\log(T)\}]$, $M = O(T^{-1/3})$ and $|B_{1,2}| = O(T/M)$. Therefore, according to theorem 3.1 in Keenan (1997), for $B_{1,2}$ we have

$$\lim_{T\to\infty} \sup_{\mathbf{x}} |P(|B_{1,2}|^{1/2} \tilde{\boldsymbol{\eta}}_{R,B_{1,2}} \leqslant \mathbf{x}) - P\{(g_1, \ldots, g_R)^{\mathrm{T}} \leqslant \mathbf{x}\}| \to 0. \tag{20}$$

Note that $\tilde{\boldsymbol{\eta}}_{R,B_{1,2}} \mathcal{W}_{k,l}$ equals $\tilde{\boldsymbol{\eta}}_{R,B_{1,2}}$ or $-\tilde{\boldsymbol{\eta}}_{R,B_{1,2}}$ since $\mathcal{W}_{k,l}$ takes on only values 1 or -1. The result in equation (20) is still valid if $\tilde{\boldsymbol{\eta}}_{R,B_{1,2}}$ in the equation is replaced by $-\tilde{\boldsymbol{\eta}}_{R,B_{1,2}}$ or $\tilde{\boldsymbol{\eta}}_{R,B_{1,2}} \mathcal{W}_{k,l}$ for any k and l. Recall that $\boldsymbol{\xi}^{k,*}$ is the sum of N_M IID random vectors of either $\tilde{\boldsymbol{\eta}}_{R,B_{1,2}}$ or $-\tilde{\boldsymbol{\eta}}_{R,B_{1,2}}$. By equation (20), we have

$$\lim_{T\to\infty} \sup_{1\leqslant k\leqslant M,\mathbf{x}} \left| P\left\{ T^{1/2}\left(\frac{1}{N_M}\boldsymbol{\xi}^{k,*}\right) \leqslant \mathbf{x} \right\} - P\{(g_1, \ldots, g_R)^{\mathrm{T}} \leqslant \mathbf{x}\} \right| \to 0.$$

Owing to the strong mixing assumption in equation (5), by applying theorem 17.2.1 in Ibragimov and Linnik (1971) N_M times, we have

$$\max_{k\leqslant M} |F^{(k)}(\boldsymbol{\theta}_R) - F^{(k)*}(\boldsymbol{\theta}_R)| \leqslant C^* T^{(1-\varsigma)a_0}, \tag{21}$$

where $\varsigma > 1$ is given in equation (5), C^* is a constant, $F^{(k)}(\boldsymbol{\theta}_R) = E\{\exp(\mathrm{i} N_M^{-1/2} \boldsymbol{\theta}_R^{\mathrm{T}} \boldsymbol{\xi}^{2,k})\}$, $1 \leqslant k \leqslant M$, are the characteristic functions of random vectors $N_M^{-1/2}\boldsymbol{\xi}^{2,k}$, $F^{(k)*}(\boldsymbol{\theta}_R) = E\{\exp(\mathrm{i} N_M^{-1/2} \boldsymbol{\theta}_R^{\mathrm{T}} \boldsymbol{\xi}^{k,*})\}$, $\boldsymbol{\theta}_R$ is an R-dimensional vector of parameters and i here is the imaginary unit. The result in equation (21) is essentially the same as that in equation (3.23) of Keenan (1997). Thus, we have

$$\lim_{T\to\infty} \sup_{1\leqslant k\leqslant M,\mathbf{x}} \left| P\left\{ T^{1/2}\left(\frac{1}{N_M}\boldsymbol{\xi}^{2,k}\right) \leqslant \mathbf{x} \right\} - P\{(g_1, \ldots, g_R)^{\mathrm{T}} \leqslant \mathbf{x}\} \right| \to 0. \tag{22}$$

In addition,

$$\begin{aligned}
\hat{\gamma}_h^{(k)} - \tilde{\gamma}_h &= \frac{(-1)^{k-1}}{T} \sum_{t=1}^{T} X_t X_{t+h} \mathcal{W}_k(t) = \frac{(-1)^{k-1}}{T} \sum_{t=1}^{T} (X_t X_{t+h} - \gamma_h) \mathcal{W}_k(t) \\
&= \frac{(-1)^{k-1}|B_{l,2}|}{T} \sum_{l=1}^{N_M} (\tilde{\gamma}_{h,B_l} - \gamma_h) \mathcal{W}_{k,l} = \frac{(-1)^{k-1} C_{T,M}^*}{N_M} \sum_{l=1}^{N_M} \tilde{\eta}_{h,B_l} \mathcal{W}_{k,l} \\
&= \frac{(-1)^{k-1} C_{T,M}^*}{N_M} (\xi_h^{1,k} + \xi_h^{2,k}),
\end{aligned} \tag{23}$$

where $C_{T,M}^* = |B_{l,2}| N_M T^{-1}$ and $\xi_h^{1,k}$ and $\xi_h^{2,k}$ are the hth elements of vectors $\boldsymbol{\xi}^{1,k}$ and $\boldsymbol{\xi}^{2,k}$ respectively.

By equation (23), we have

$$\begin{aligned}
\max_{1\leqslant k\leqslant M, h\leqslant R} T^{1/2} \left| (\hat{\gamma}_h^{(k)} - \tilde{\gamma}_h) - \frac{(-1)^{k-1} C_{T,M}^*}{N_M} \xi_h^{2,k} \right| &= \max_{1\leqslant k\leqslant M, h\leqslant R} \left| \frac{T^{1/2} C_{T,M}^*}{N_M} \xi_h^{1,k} \right| \\
&= \max_{1\leqslant k\leqslant M, h\leqslant R} \left| \frac{T^{1/2} C_{T,M}^*}{N_M |B_{l,2}|} \sum_{l=1}^{N_M} \{\mathcal{W}_{k,l} \sum_{t\in B_{l,1}} (X_t X_{t+h} - \gamma_h)\} \right|
\end{aligned}$$

$$\leqslant \max_{1\leqslant k\leqslant M, h\leqslant R} \frac{\nu_{T,M,1} C^{*}_{T,M}}{|B_{l,2}|} \times \left| \frac{T^{1/2}}{N_M} \left[\sum_{l=1}^{N_M} \mathcal{W}_{k,l} \sum_{t\in B_{l,1}} \{ |B_{l,1}|^{-1} (X_t X_{t+h} - \gamma_h)\} \right] \right|$$

$$= \max_{1\leqslant k\leqslant M, h\leqslant R} \frac{\nu_{T,M,1} C^{*}_{T,M}}{|B_{l,2}|} \left| \frac{T^{1/2}}{N_M} \xi_h^{1,k,*} \right|, \tag{24}$$

where $\xi_h^{1,k,*} = \Sigma_{l=1}^{N_M} \Sigma_{t\in B_{l,1}} \{ |B_{l,1}|^{-1} (X_t X_{t+h} - \gamma_h) \mathcal{W}_{k,l} \}$. Let $\boldsymbol{\xi}^{1,k,*} = (\xi_0^{1,k,*}, \ldots, \xi_{R-1}^{1,k,*})^{\mathrm{T}}$. Note that $\boldsymbol{\xi}^{1,k,*}$ on $B_{l,1}$ is analogous to $\boldsymbol{\xi}^{2,k}$ on $B_{l,2}$ with the hth element $\Sigma_{l=1}^{N_M} \Sigma_{t\in B_{l,2}} \{ |B_{l,2}|^{-1} (X_t X_{t+h} - \gamma_h) \mathcal{W}_{k,l} \}$. With some arguments similar to those leading to equation (22) on $\boldsymbol{\xi}^{1,k,*}$, one can obtain that

$$\max_{1\leqslant k\leqslant M, h\leqslant R} \left| \frac{T^{1/2}}{N_M} \left[\sum_{l=1}^{N_M} \mathcal{W}_{k,l} \sum_{t\in B_{l,1}} \{ |B_{l,1}|^{-1} (X_t X_{t+h} - \gamma_h)\} \right] \right| = O_p(1).$$

By equation (24) and the fact that $\nu_{T,M,1} C^{*}_{T,M} |B_{l,2}|^{-1} \to 0$ as $T\to\infty$, we have

$$\max_{1\leqslant k\leqslant M, h\leqslant R} T^{1/2} \left| (\hat{\gamma}_h^{(k)} - \tilde{\gamma}_h) - \frac{(-1)^{k-1} C^{*}_{T,M}}{N_M} \xi_h^{2,k} \right| \to 0 \tag{25}$$

in probability as $T\to\infty$. By equations (22) and (25),

$$\lim_{T\to\infty} \sup_{1\leqslant k\leqslant M, \mathbf{x}} |P\{T^{1/2}(\hat{\boldsymbol{\gamma}}_R^{(k)} - \hat{\boldsymbol{\gamma}}_R) \leqslant \mathbf{x}\} - P\{(g_1, \ldots, g_R)^{\mathrm{T}} \leqslant \mathbf{x}\}| \to 0.$$

A.6. Proof of lemma 2

Since $Q_T\to\infty$ and $Q_T = O(T^{\lambda})$, where $\lambda\in(0, 0.5)$, $Q_T + \log(T)$ is bounded by $O(T^{\lambda_0})$ for some $\lambda_0 \in (\lambda, 0.5)$. Keenan (1997), page 65, pointed out that

$$P\left[\limsup_{T\to\infty} \max_{0\leqslant h\leqslant M_0} \left| \left\{ \frac{T}{\log(T)} \right\}^{1/2} (\hat{\gamma}_h - \gamma_h) \right| \leqslant A_0 \right] = 1,$$

where $M_0 = O[\{T/\log(T)\}^{1/2}]$ for some $A_0 > 0$. Similar results can also be obtained from theorem 1 in Xiao and Wu (2012). Clearly, we have

$$\max_{0\leqslant h\leqslant 2\{Q_T + \log(T)\}} |\hat{\gamma}_h - \gamma_h| = O_p[\{T/\log(T)\}^{-1/2}].$$

By the assumption $|\phi_l| \leqslant K_0\, v(l)$, we have $\gamma_h \leqslant (K_0 \Sigma_{l=0}^{\infty} \phi_l)/h$. Therefore,

$$\left| \sum_{\delta=Q_T}^{\infty} \gamma_\delta \gamma_{\delta-i+j} \right| \leqslant \left(K_0 \sum_{l=0}^{\infty} \phi_l \right)^2 \sum_{\delta=Q_T}^{\infty} (\delta - 2R)^{-2} = O(Q_T^{-1})$$

and

$$\left| \sum_{\delta=Q_T}^{\infty} \gamma_{\delta+i} \gamma_{\delta-j} \right| \leqslant \left(K_0 \sum_{l=0}^{\infty} \phi_l \right)^2 \sum_{\delta=Q_T}^{\infty} (\delta - R)^{-2} = O(Q_T^{-1}).$$

Hence, we obtain

$$\max_{0\leqslant i,j\leqslant R-1} |\hat{\Gamma}_{i+1,j+1} - \Gamma_{i+1,j+1}| \leqslant \max_{0\leqslant i,j\leqslant R-1} \left\{ 2 \left| \sum_{\delta=Q_T}^{\infty} (\gamma_\delta \gamma_{\delta-i+j} + \gamma_{\delta+i}\gamma_{\delta-j}) \right| + \left| \sum_{\delta=-Q_T}^{Q_T} \{ (\hat{\gamma}_\delta \hat{\gamma}_{\delta-i+j} + \hat{\gamma}_{\delta+i}\hat{\gamma}_{\delta-j}) - (\gamma_\delta \gamma_{\delta-i+j} + \gamma_{\delta+i}\gamma_{\delta-j}) \} \right| + |\hat{\kappa}_4 \hat{\gamma}_i \hat{\gamma}_j - \kappa_4 \gamma_i \gamma_j| \right\}$$

$$= O(Q_T^{-1}) + O_p[\{T/\log(T)\}^{-1/2} Q_T] + o_p(1) = o_p(1).$$

Since $A_{k,v} \to 1$ uniformly for all $1\leqslant k\leqslant M$ as $T\to\infty$, this inequality implies that $\max_{1\leqslant k\leqslant M,\, 1\leqslant i,\, j\leqslant R} |\hat{\Gamma}_{ij}^{(k)} - \Gamma_{ij}| \to 0$ in probability as $T\to\infty$.

A.7. A lemma for the proof of theorem 3

Lemma 5. Under the conditions in theorem 3, for a fixed number R_0, we have

$$\hat{D}^k_{R_0} \to \max_{1 \leqslant r \leqslant R_0} \left(\sum_{i=1}^{r} e^2_{k,i} - 2r \right)$$

in distribution uniformly for all $1 \leqslant k \leqslant M$ as $T \to \infty$.

Proof. Since both $\hat{\boldsymbol{\Gamma}}^{(k)}_r$ and $\boldsymbol{\Gamma}_r$ are symmetric, by the Cholesky decomposition, we obtain $\boldsymbol{\Gamma}_r = \mathbf{L}_r \mathbf{L}_r^{\mathrm{T}}$ and $\hat{\boldsymbol{\Gamma}}^{(k)}_r = \hat{\mathbf{L}}_{r,k} \hat{\mathbf{L}}^{\mathrm{T}}_{r,k}$, where $\mathbf{L}_r$ and $\hat{\mathbf{L}}_{r,k}$ are lower triangular matrices. By equation (8), $\hat{\mathbf{L}}_{r,k} \to \mathbf{L}_r$ in probability uniformly for all $1 \leqslant k \leqslant M$. Therefore, by Slutsky's theorem and the uniform convergence in equation (7), $T^{1/2} \hat{\mathbf{L}}^{-1}_{r,k} (\hat{\gamma}^{(k)}_r - \hat{\gamma}_r) \to N(0, \mathcal{I}_r)$ in distribution uniformly for all $1 \leqslant k \leqslant M$, where $\mathcal{I}_r$ is an $r \times r$ identity matrix. In addition, for each $r = 1, \ldots, R_0$, $\hat{\boldsymbol{\Gamma}}^{(k)}_r$ is the upper $r \times r$ submatrix of $\hat{\boldsymbol{\Gamma}}^{(k)}_{R_0}$, and hence matrix $\hat{\mathbf{L}}_{r,k}$ is the upper $r \times r$ submatrix of $\hat{\mathbf{L}}_{R_0,k}$. Therefore, by lemma 2,

$$\hat{D}^k_{R_0} = \max_{1 \leqslant r \leqslant R_0} \{ (T^{1/2} \hat{\mathbf{L}}^{-1}_{r,k} (\hat{\gamma}^{(k)}_r - \hat{\gamma}_r))^{\mathrm{T}} (T^{1/2} \hat{\mathbf{L}}^{-1}_{r,k} (\hat{\gamma}^{(k)}_r - \hat{\gamma})) - 2r \} \to \max_{1 \leqslant r \leqslant R_0} \left(\sum_{i=1}^{r} e^2_i - 2r \right)$$

in distribution uniformly for all $1 \leqslant k \leqslant M$, where $\{e_i, i = 1, 2, \ldots\}$ are independent standard normal random variables.

A.8. Proof of theorem 3

For a fixed number R_0, let $\Lambda_{R_0} = \max_{1 \leqslant r \leqslant R_0} (\Sigma^r_{i=1} e^2_{k,i} - 2r)$, $\Lambda_\infty = \sup_{r \geqslant 1} (\Sigma^r_{i=1} e^2_{k,i} - 2r)$, $\Lambda^k_{R_0,\infty} = \sup_{r > R_0} (\Sigma^r_{i=1} e^2_{k,i} - 2r)$ and

$$\hat{D}^k_{R_0,R} = \max_{R_0 < r \leqslant R} \{ T(\hat{\gamma}^{(k)}_r - \hat{\gamma}_r)^{\mathrm{T}} (\hat{\boldsymbol{\Gamma}}^{(k)}_r)^{-1} (\hat{\gamma}^{(k)}_r - \hat{\gamma}) - 2r \}.$$

For any given $x \in (-\infty, \infty)$, we have

$$\begin{aligned} |P(\hat{D}^k_R > x) - P(\Lambda_\infty > x)| &\leqslant |P(\hat{D}^k_R > x) - P(\hat{D}^k_{R_0} > x)| + |P(\hat{D}^k_{R_0} > x) - P(\Lambda_{R_0} > x)| \\ &\quad + |P(\Lambda_{R_0} > x) - P(\Lambda_\infty > x)| \\ &= |P(\hat{D}^k_{R_0,R} > x) - P(\{\hat{D}^k_{R_0,R} > x\} \cap \{\hat{D}^k_{R_0} > x\})| \\ &\quad + |P(\hat{D}^k_{R_0} > x) - P(\Lambda_{R_0} > x)| + |P(\Lambda_{R_0} > x) - P(\Lambda_\infty > x)| \\ &\leqslant P(\hat{D}^k_{R_0,R} > x) + |P(\hat{D}^k_{R_0} > x) - P(\Lambda_{R_0} > x)| + P(\Lambda^k_{R_0,\infty} > x). \end{aligned} \tag{26}$$

Since $E(e^2_{k,i}) = 1$ and $\mathrm{var}(e^2_{k,i}) = 2$ for any $k = 1, 2, \ldots$ and $i = 1, 2, \ldots$, by the strong law of large numbers,

$$\left(\sum_{i=1}^{r} e^2_{k,i} - 2r \right) \bigg/ r \to -1 \tag{27}$$

with probability 1, as $r \to \infty$. Clearly, $\Lambda^k_{R_0,\infty} = \sup_{r > R_0} (\Sigma^r_{i=1} e^2_{k,i} - 2r)$, $k = 1, 2, \ldots,$ are IID. Therefore, for any $\varepsilon > 0$, when $R_0 > R^*_{0,\varepsilon}$ which is sufficiently large and at least $2|x|$, by equation (27),

$$\begin{aligned} \sup_k P(\Lambda^k_{R_0,\infty} > x) = P(\Lambda^1_{R_0,\infty} > x) &= P\left\{ \sup_{r > R_0} \left(\sum_{i=1}^{r} e^2_{1,i} - 2r \right) > x \right\} \\ &= P\left[\sup_{r > R_0} \left\{ \left(\sum_{i=1}^{r} e^2_{1,i} - 2r \right) \bigg/ r \right\} > x/r \right] \\ &\leqslant P\left[\sup_{r > R_0} \left\{ \left(\sum_{i=1}^{r} e^2_{1,i} - 2r \right) \bigg/ r \right\} > -0.5 \right] < \varepsilon. \end{aligned} \tag{28}$$

By lemma 5 when T is sufficiently large,

$$\sup_{1 \leqslant k \leqslant M} |P(\hat{D}^k_{R_0} > x) - P(\Lambda_{R_0} > x)| < \varepsilon. \tag{29}$$

By the Sherman–Morrison–Woodbury formula, we have

$$(\hat{\mathbf{\Gamma}}_r^{(k)})^{-1} = \mathbf{\Gamma}_r^{-1} + \mathbf{\Gamma}_r^{-1}\mathcal{B}_r^{(k)}(\mathcal{B}_r^{(k)} + \mathcal{B}_r^{(k)}\mathbf{\Gamma}_r^{-1}\mathcal{B}_r^{(k)})^{-1}\mathcal{B}_r^{(k)}\mathbf{\Gamma}_r^{-1},$$

where $\boldsymbol{\mathcal{B}}_r^{(k)} = \hat{\mathbf{\Gamma}}_r^{(k)} - \mathbf{\Gamma}_r$. By equation (8), we know that $\max_{1\leqslant i,j\leqslant R} |\hat{\Gamma}_{ij}^{(k)} - \Gamma_{ij}| \to 0$ as $T \to \infty$. Therefore, $\sup_{1\leqslant k\leqslant M, 1\leqslant r\leqslant R} T(\hat{\gamma}_r^{(k)} - \hat{\gamma}_r)^{\mathrm{T}}\{\mathbf{\Gamma}_r^{-1} - (\hat{\mathbf{\Gamma}}_r^{(k)})^{-1}\}(\hat{\gamma}_r^{(k)} - \hat{\gamma}_r)$ is a negligible term. By definition, both $\hat{D}_{R_0,R}^k$ and $\max_{R_0<r\leqslant R} \Sigma_{i=1}^r (e_{k,i}^2 - 2r)$ are the results of the same continuous transformation from $T^{1/2}(\hat{\gamma}_R^{(k)} - \hat{\gamma}_R)$ and $(g_1, \ldots, g_R)^{\mathrm{T}}$ respectively. By equation (6) in lemma 1, the characteristic function of $\hat{D}_{R_0,R}^k$ converges to that of $\max_{R_0<r\leqslant R} \Sigma_{i=1}^r (e_{k,i}^2 - 2r)$ uniformly for all $1 \leqslant k \leqslant M$. Hence, we have

$$\max_{1\leqslant k\leqslant M} \left| P(\hat{D}_{R_0,R}^k > x) - P\left\{ \max_{R_0<r\leqslant R} \sum_{i=1}^r (e_{k,i}^2 - 2r) > x \right\} \right| \to 0 \tag{30}$$

as $T \to \infty$. By equation (28), we select $R_0 > R_{0,\varepsilon/2}^*$ such that

$$\max_{1\leqslant k\leqslant M} P\left\{ \max_{R_0<r\leqslant R} \sum_{i=1}^r (e_{k,i}^2 - 2r) > x \right\} \leqslant \sup_k P(\Lambda_{R_0,\infty}^k > x) < \frac{\varepsilon}{2}.$$

By equation (30), for any $\varepsilon > 0$, when $T > T_0 \geqslant \exp(R_{0,\varepsilon}^*)$ which is a sufficiently large number, we have

$$\begin{aligned} \max_{1\leqslant k\leqslant M} |P(\hat{D}_{R_0,R}^k > x)| &\leqslant \max_{1\leqslant k\leqslant M} \left| P(\hat{D}_{R_0,R}^k > x) - P\left\{ \max_{R_0<r\leqslant R} \sum_{i=1}^r (e_{k,i}^2 - 2r) > x \right\} \right| \\ &\quad + \max_{1\leqslant k\leqslant M} P\left\{ \max_{R_0<r\leqslant R} \sum_{i=1}^r (e_{k,i}^2 - 2r) > x \right\} < \varepsilon. \end{aligned} \tag{31}$$

By equations (26)–(29) and (31), when R_0 and T are sufficiently large, $\sup_{1\leqslant k\leqslant M} |P(\hat{D}_R^k > x) - P(\Lambda_\infty > x)| < 3\varepsilon$, which implies that $\hat{D}_R^k \to \Lambda_\infty$ in distribution as $T \to \infty$ uniformly for $1 \leqslant k \leqslant M$.

A.9. Proof of theorem 4

For any given $x \in (-\infty, \infty)$, we have

$$\begin{aligned} P\left(\sum_{i=1}^r e_{k,i}^2 - 2r \geqslant x \right) &= P\left\{ \sum_{i=1}^r e_{k,i}^2 - r \geqslant \frac{1}{2}\left(x + \frac{r}{2}\right) + \frac{1}{2}\left(x + \frac{r}{2} + r\right) \right\} \\ &\leqslant P\left[\sum_{i=1}^r e_{k,i}^2 - r \geqslant \frac{1}{2}\left(x + \frac{r}{2}\right) + \left\{ \left(x + \frac{r}{2}\right) r \right\}^{1/2} \right] \\ &= P\left[\sum_{i=1}^r e_{k,i}^2 - r \geqslant 2\left\{ \frac{1}{4}\left(x + \frac{r}{2}\right) \right\} + 2\left\{ \frac{1}{4}\left(x + \frac{r}{2}\right) r \right\}^{1/2} \right] \\ &= P\left\{ \sum_{i=1}^r e_{k,i}^2 - r \geqslant 2z + 2(zr)^{1/2} \right\} \end{aligned}$$

for any $r \geqslant -2x$, where $z = (x + r/2)/4$. Lemma 1 in Laurent and Massart (2000) shows that, for $z \geqslant 0$,

$$P\left\{ \sum_{i=1}^r e_{k,i}^2 - r \geqslant 2(zr)^{1/2} + 2z \right\} \leqslant \exp(-z).$$

Hence, we obtain $P(\Sigma_{i=1}^r e_{k,i}^2 - 2r \geqslant x) \leqslant \exp(-x/4 - r/8)$. In addition, when $r < -2x$, the result above is trivial because the probability is always bounded by 1. Therefore, we have

$$P\left\{ \sup_{r\geqslant 1} \left(\sum_{i=1}^r e_{k,i}^2 - 2r \right) \geqslant x \right\} \leqslant \sum_{r=1}^\infty P\left(\sum_{i=1}^r e_{k,i}^2 - 2r \geqslant x \right) = G_\infty \exp(-x/4),$$

where $G_\infty = \Sigma_{r=1}^\infty \exp(-r/8)$ is finite. For any M_0, we have

$$\begin{aligned} P\left[\sup_{k>M_0} \left\{ \sup_{r\geqslant 1} \left(\sum_{i=1}^r e_{k,i}^2 - 2r \right) - (k-1)^{1/2} \right\} > x \right] &\leqslant \sum_{k=M_0}^\infty P\left\{ \sup_{r\geqslant 1} \left(\sum_{i=1}^r e_{k,i}^2 - 2r \right) \geqslant x + (k-1)^{1/2} \right\} \\ &\leqslant \sum_{k=M_0}^\infty G_\infty \exp[-\{x + (k-1)^{1/2}\}/4] \\ &= G_\infty \exp(-x/4) \sum_{k=M_0}^\infty \exp\{-(k-1)^{1/2}/4\}, \end{aligned} \tag{32}$$

which can be arbitrarily close to 0 if M_0 is sufficiently large. By theorem 3, $\lim_{T\to\infty} P(\hat{D}_R^k \geqslant x) \leqslant G_\infty \exp(-x/4)$ uniformly for all $1 \leqslant k \leqslant M$. Similarly to equation (32), we also have

$$\begin{aligned}\lim_{T\to\infty} P[\max_{M_0<k\leqslant M}\{\hat{D}_R^k-(k-1)^{1/2}\}\geqslant x] &\leqslant \lim_{T\to\infty}\sum_{k=M_0}^{M} P\{\hat{D}_R^k\geqslant x+(k-1)^{1/2}\}\\ &\leqslant \sum_{k=M_0}^{\infty} G_\infty \exp[-\{x+(k-1)^{1/2}\}/4]\\ &= G_\infty \exp(-x/4)\sum_{k=M_0}^{\infty}\exp\{-(k-1)^{1/2}/4\}. \end{aligned} \tag{33}$$

By theorems 1 and 3, $\hat{D}_R^k, k=1,2,\ldots,M_0$, are asymptotically IID random variables. By Slutsky's theorem, we have

$$\hat{D}_{M_0,R} \to \max_{1\leqslant k\leqslant M_0}\left\{\sup_{r\geqslant 1}\left(\sum_{i=1}^{r} e_{k,i}^2-2r\right)-(k-1)^{1/2}\right\} \tag{34}$$

in distribution as $T\to\infty$. By equations (32)–(34), for any $x, \varepsilon>0$, there is some large number M_0 such that

$$\begin{aligned}&\left|P(\hat{D}_{M,R}>x)-P\left[\sup_{k\geqslant 1}\left\{\sup_{r\geqslant 1}\left(\sum_{i=1}^{r} e_{k,i}^2-2r\right)-(k-1)^{1/2}\right\}>x\right]\right|\\ &\quad\leqslant \left|P(\hat{D}_{M_0,R}>x)-P\left[\max_{1\leqslant k\leqslant M_0}\left\{\sup_{r\geqslant 1}\left(\sum_{i=1}^{r} e_{k,i}^2-2r\right)-(k-1)^{1/2}\right\}>x\right]\right|\\ &\quad\quad+P\left[\max_{M_0<k\leqslant M}\{\hat{D}_R^k-(k-1)^{1/2}\}>x\right]+P\left[\sup_{k>M_0}\left\{\sup_{r\geqslant 1}\left(\sum_{i=1}^{r} e_{k,i}^2-2r\right)-(k-1)^{1/2}\right\}>x\right]<\varepsilon\end{aligned}$$

when T is sufficiently large. Thus, when $T\to\infty$ we have

$$\hat{D}_{M,R}\to \sup_{k\geqslant 1}\left\{\sup_{r\geqslant 1}\left(\sum_{i=1}^{r} e_{k,i}^2-2r\right)-(k-1)^{1/2}\right\} \qquad \text{in distribution.}$$

A.10. A lemma for the proof of theorem 6

Lemma 6. If a process is a locally stationary process of form (10) satisfying assumption 2.1 of Dahlhaus (2009), for any positive integers $i\leqslant j$ and $k\geqslant 1$,

$$\lim_{T\to\infty} T\,\mathrm{cov}(\hat{\gamma}_i^{(k)}-\hat{\gamma}_i, \hat{\gamma}_j^{(k)}-\hat{\gamma}_j)<\infty.$$

Proof. By model (10), for any integer h, we have

$$\begin{aligned}&E(X_tX_{t+i}X_{t+h+i}X_{t+h+i+j})\\ &\quad=\sum_{l_1}\sum_{l_2}\sum_{l_3}\sum_{l_4}\phi_{l_1,t,T}\phi_{l_2+i,t+i,T}\phi_{l_3+h+i,t+h+i,T}\phi_{l_4+h+i+j,t+h+i+j,T}\,E(Z_{t-l_1}Z_{t-l_2}Z_{t-l_3}Z_{t-l_4})\\ &\quad=\kappa_4\sigma^4\sum_{l}\phi_{l,t,T}\phi_{l+i,t+i,T}\phi_{l+h+i,t+h+i,T}\phi_{l+h+i+j,t+h+i+j,T}\\ &\quad\quad+\sigma^4\sum_{l_1}\sum_{l_2}\phi_{l_1,t,T}\phi_{l_1+i,t+i,T}\phi_{l_2,t+h+i,T}\phi_{l_2+j,\,t+h+i+j,T}\\ &\quad\quad+\sigma^4\sum_{l_1}\sum_{l_2}\phi_{l_1,t,T}\phi_{l_2,t+i,T}\phi_{l_1+h+i,t+h+i,T}\phi_{l_2+h+j,t+h+i+j,T}\\ &\quad\quad+\sigma^4\sum_{l_1}\sum_{l_2}\phi_{l_1,t,T}\phi_{l_2,\,t+i,\,T}\phi_{l_2+h,t+h+i,T}\phi_{l_1+h+i+j,t+h+i+j,T}.\end{aligned}$$

Therefore,

$$\mathrm{cov}(\hat{\gamma}_i^{(k)}-\tilde{\gamma}_i, \hat{\gamma}_j^{(k)}-\tilde{\gamma}_j)=\mathrm{cov}\left\{\frac{1}{T}\sum_{t=1}^{T}X_tX_{t+i}\,\mathcal{W}_k(t), \frac{1}{T}\sum_{t=1}^{T}X_tX_{t+j}\,\mathcal{W}_k(t)\right\}$$

$$= E\left\{ \frac{1}{T^2} \sum_{t_1=1}^{T} \sum_{t_2=1}^{T} X_{t_1} X_{t_1+i} X_{t_2} X_{t_2+j} \mathcal{W}_k(t_1) \mathcal{W}_k(t_2) \right\} - E(\hat{\gamma}_i^{(k)} - \tilde{\gamma}_i)\, E(\tilde{\gamma}_j^{(k)} - \tilde{\gamma}_j)$$

$$= \frac{1}{T^2} \sum_{t_1=1}^{T} \sum_{t_2=1}^{T} V(t_1, t_2, i, j) \mathcal{W}_k(t_1) \mathcal{W}_k(t_2)$$

$$+ \frac{1}{T^2} \sum_{t_1=1}^{T} \sum_{t_2=1}^{T} \mathrm{cov}(X_{t_1}, X_{t_1+i}) \,\mathrm{cov}(X_{t_2}, X_{t_2+j}) \mathcal{W}_k(t_1) \mathcal{W}_k(t_2) - E(\hat{\gamma}_i^{(k)} - \tilde{\gamma}_i)\, E(\hat{\gamma}_j^{(k)} - \tilde{\gamma}_j)$$

$$= \frac{1}{T^2} \sum_{t_1=1}^{T} \sum_{t_2=1}^{T} V(t_1, t_2, i, j) \mathcal{W}_k(t_1) \mathcal{W}_k(t_2),$$

where

$$\begin{aligned} V(t_1, t_2, i, j) = {} & \kappa_4 \sigma^4 \sum_l \phi_{l,t_1,T} \phi_{l+i,t_1+i,T} \phi_{l+t_2-t_1,t_2,T} \phi_{l+t_2-t_1+j,\,t_2+j,\,T} \\ & + \sigma^4 \sum_{l_1} \sum_{l_2} \phi_{l_1,t_1,T} \phi_{l_2,\,t_1+i,\,T} \phi_{l_1+t_2-t_1,t_2,T} \phi_{l_2+t_2-t_1-i+j,t_2+j,T} \\ & + \sigma^4 \sum_{l_1} \sum_{l_2} \phi_{l_1,t_1,T} \phi_{l_2,\,t_1+i,\,T} \phi_{l_2+t_2-t_1-i,t_2,T} \phi_{l_1+t_2-t_1+j,t_2+j,T}. \end{aligned}$$

By the assumption in equation (11), it is readily seen that $|V(t_1, t_2, i, j)| \leqslant |V^*(t_2 - t_1, i, j)|$ which is $V(t_1, t_2, i, j)$ of stationary process $X_t^* = \Sigma_{l=-\infty}^{\infty} \phi_l^* Z_{t-l}$, where $\phi_l^* = K_0 |v(l)|$. Since $\{X_t^*\}$ satisfies the conditions of proposition 7.3.1 of Brockwell and Davis (1991), we have

$$\lim_{T\to\infty} |T\, \mathrm{cov}(\hat{\gamma}_i^{(k)} - \hat{\gamma}_i, \hat{\gamma}_j^{(k)} - \hat{\gamma}_j)| \leqslant \lim_{T\to\infty} \frac{1}{T} \sum_{t_1=1}^{T} \sum_{t_2=1}^{T} |V^*(t_2 - t_1, i, j) \mathcal{W}_k(t_1) \mathcal{W}_k(t_2)| < \infty.$$

A.11. Proof of theorem 5

Under a sequence of local alternatives in equation (14), by equation (13), we have

$$|E(\hat{\gamma}_{h_0}^{(k_0)} - \hat{\gamma}_{h_0})| = l_T \left| \int_0^1 c_{f,h_0}(u)\, W_{k_0}(u)\, \mathrm{d}u \right| + O(1/T) \tag{35}$$

and $|\int_0^1 c_{f,h_0}(u)\, W_{k_0}(u)\, \mathrm{d}u| > 0$ at some h_0 and k_0. By equation (36) in Dahlhaus and Polonik (2009), we have the classical periodogram $I_T(\omega_j) = \Sigma_{|k|<T} \hat{\gamma}_k \exp(-\mathrm{i}k\omega_j)$ for a locally stationary process, where $\omega_j = 2\pi j/T$ and $j = 0, \pm 1, \ldots$. By Parseval's theorem and Holder's inequality, for any integers h_1 and h_2, we have

$$\begin{aligned} |\hat{\Gamma}_{h_1,h_2}| &= |\hat{\kappa}_4 \hat{\gamma}_{h_1} \hat{\gamma}_{h_2} + \sum_{v=-Q_T}^{Q_T} (\hat{\gamma}_v \hat{\gamma}_{v-h_1+h_2} + \hat{\gamma}_{v+h_1} \hat{\gamma}_{v-h_2})| \\ &\leqslant \hat{\kappa}_4 \hat{\gamma}_0^2 + 2 \sum_{|l|<T} \hat{\gamma}_l^2 = \hat{\kappa}_4 \hat{\gamma}_0^2 + \frac{2\pi}{T} \sum_{|j| \leqslant \lfloor T/2 \rfloor} I_T^2(\omega_j) = O_p(1). \end{aligned}$$

Hence $\hat{\mathbf{\Gamma}}_{h_0}^{(k_0)}$ is $O_p(1)$. In addition, by equation (35) and lemma 6, we have $|T^{1/2}(\hat{\gamma}_{h_0}^{(k_0)} - \hat{\gamma}_{h_0})| \to \infty$ in probability at rate $T^{1/2} l_T$. Therefore,

$$\hat{D}_{M,R} \geqslant \lim_{T\to\infty} T(\hat{\gamma}_{h_0}^{(k_0)} - \hat{\gamma}_{h_0})^{\mathrm{T}} (\hat{\mathbf{\Gamma}}_{h_0}^{(k_0)})^{-1} (\hat{\gamma}_{h_0}^{(k_0)} - \hat{\gamma}_{h_0}) - 2h_0 - (k_0 - 1)^{1/2} \to \infty \qquad \text{in probability.}$$

References

Ahmed, N. and Rao, K. R. (1975) *Orthogonal Transformations for Digital Signal Processing*. New York: Springer.
Box, G. E. P. and Jenkins, G. M. (1970) *Time Series Analysis: Forecasting and Control*. San Francisco: Holden-Day.
Box, G., Jenkins, G. and Reinsel, G. (1994) *Time Series Analysis: Forecasting and Control*, 3rd edn. New York: Prentice Hall.
Brockwell, J. P. and Davis, A. R. (1991) *Time Series: Theory and Methods*, 2nd edn. New York: Springer.
Dahlhaus, R. (1997) Fitting time series models to nonstationary processes. *Ann. Statist.*, **25**, 1–37.

Dahlhaus, R. (2009) Local inference for locally stationary time series based on the empirical spectral measure. *J. Econmetr.*, **151**, 101–112.
Dahlhaus, R. and Polonik, W. (2009) Empirical spectral processes for locally stationary time series. *Bernoulli*, **15**, 1–39.
Davidson, J. (1994) *Stochastic Limit Theory: an Introduction for Econometricians*. Oxford: Oxford University Press.
Dette, H., Preuß, P. and Vetter, M. (2011) A measure of stationarity in locally stationary processes with applications to testing. *J. Am. Statist. Ass.*, **106**, 1113–1124.
Dwivedi, Y. and Subba Rao, S. (2011) A test for second order stationarity of a time series based on the discrete Fourier transform. *J. Time Ser. Anal.*, **32**, 68–91.
Eubank, R. L. and Hart, J. (1992) Testing goodness-of-fit in regression via order selection criteria. *Ann. Statist.*, **20**, 1412–1425.
Han, D. and Tsung, F. (2006) A reference-free cuscore chart for dynamic mean change detection and a unified framework for charting performance comparison. *J. Am. Statist. Ass.*, **101**, 368–386.
Hart, J. D. (1997) *Nonparametric Smoothing and Lack-of-fit Tests*. New York: Springer.
Ibragimov, I. A. and Linnik, Y. V. (1971) *Independent and Stationary Sequences of Random Variables*. Groningen: Wolters–Noordhoff.
Jirak, M. (2011) On the maximum of covariance estimators. *J. Multiv. Anal.*, **102**, 1032–1046.
Keenan, D. M. (1997) A central limit theorem for m(n) autocovariances. *J. Time Ser. Anal.*, **18**, 61–78.
Kim, J. D. (2000) An order selection criterion for testing goodness of fit. *J. Am. Statist. Ass.*, **95**, 829–835.
Kreiss, J. P. and Paparoditis, E. (2015) Bootstrapping locally stationary processes. *J. R. Statist. Soc.* B, **77**, in the press.
Laurent, B. and Massart, P. (2000) Adaptive estimation of a quadratic functional by model selection. *Ann. Statist.*, **28**, 1302–1338.
Lee, S., Ha, J., Na, O. and Na, S. (2003) The cusum test for parameter change in time series models. *Scand. J. Statist.*, **30**, 781–796.
Nason, G. (2013a) A test for second-order stationarity and approximate confidence intervals for localized autocovariances for locally stationary time series. *J. R. Statist. Soc.* B, **75**, 879–904.
Nason, G. (2013b) locits: tests of stationarity and localized autocovariance. *R Package, Version 1.4*.
Nason, G. P., von Sachs, R. and Kroisandt, G. (2000) Wavelet processes and adaptive estimation of the evolutionary wavelet spectrum. *J. R. Statist. Soc.* B, **62**, 271–292.
Ombao, H., von Sachs, R. and Guo, W. (2005) SLEX analysis of multivariate nonstationary time series. *J. Am. Statist. Ass.*, **100**, 519–531.
Paparoditis, E. (2009) Testing temporal constancy of the spectral structure of a time series. *Bernoulli*, **15**, 1190–1221.
Paparoditis, E. (2010) Validating stationarity assumptions in time series analysis by rolling local periodograms. *J. Am. Statist. Ass.*, **105**, 839–851.
Paparoditis, E. and Preuß, P. (2013) Local power properties of frequency domain based tests for stationarity. *Preprint*.
Preuß, P., Vetter, M. and Dette, H. (2013) A test for stationarity based on empirical processes. *Bernoulli*, **19**, 2715–2749.
Priestley, M. B. and Subba Rao, T. (1969) A test for non-stationarity of time-series. *J. R. Statist. Soc.* B, **31**, 140–149.
von Sachs, R. and Neumann, M. H. (2000) A wavelet-based test for stationarity. *J. Time Ser. Anal.*, **21**, 597–613.
Stoffer, D. (1987) Walsh-fourier analysis of discrete-valued time series. *J. Time Ser. Anal.*, **8**, 449–467.
Stoffer, D. (1991) Walsh-Fourier analysis and its statistical applications (with discussion). *J. Am. Statist. Ass.*, **86**, 462–483.
Wang, H. and Akritas, M. (2010) Rank test for heteroscedastic functional data. *J. Multiv. Anal.*, **101**, 1791–1805.
Wang, H. and Akritas, M. (2011) Asymptotically distribution free tests in heteroscedastic unbalanced high dimensional anova. *Statist. Sin.*, **21**, 1341–1377.
Xiao, H. and Wu, W. B. (2012) Covariance matrix estimation for stationary time series. *Ann. Statist.*, **40**, 466–493.