Econometric Theory, **40**, 2024, 511–557.
doi:10.1017/S0266466622000494

TESTING FOR STRICT STATIONARITY VIA THE DISCRETE FOURIER TRANSFORM

Zhonghao Fu
Fudan University
and
Shanghai Institute of International Finance and Economics

Shang Gao
Fudan University

Liangjun Su
Tsinghua University

Xia Wang
Renmin University of China

This paper proposes a model-free test for the strict stationarity of a potentially vector-valued time series using the discrete Fourier transform (DFT) approach. We show that the DFT of a residual process based on the empirical characteristic function weakly converges to a zero spectrum in the frequency domain for a strictly stationary time series and a nonzero spectrum otherwise. The proposed test is powerful against various types of nonstationarity including deterministic trends and smooth or abrupt structural changes. It does not require smoothed nonparametric estimation and, thus, can detect the Pitman sequence of local alternatives at the parametric rate $T^{-1/2}$, faster than all existing nonparametric tests. We also design a class of derivative tests based on the characteristic function to test the stationarity in various moments. Monte Carlo studies demonstrate that our test has reasonarble size and excellent power. Our empirical application of exchange rates strongly suggests that both nominal and real exchange rate returns are nonstationary, which the augmented Dickey–Fuller and Kwiatkowski–Phillips–Schmidt–Shin tests may overlook.

1. INTRODUCTION

Stationarity is a fundamental assumption in many time series applications that generally enables convenient applications of statistical analysis, such as parameter estimation, inference, and forecasting. The stationarity assumption enables extracting or investigating the properties of a time series process that is stable over time.

Fu acknowledges financial support from the National Science Foundation of China (NSFC) under Nos. 71903032 and 72121002. Gao acknowledges financial support from the NSFC under No. 71973030. Su gratefully acknowledges the NSFC for financial support under No. 72133002. Wang acknowledges financial support from the NSFC under No. 71873151. Address correspondence to Xia Wang, School of Economics, Renmin University of China, Beijing, China; e-mail: wxia@ruc.edu.cn.

Various time series models are based on certain stationarity assumptions, such as the stationary autoregressive moving average (ARMA), generalized autoregressive conditional heteroskedasticity (GARCH), threshold autoregressive, and Markov regime-switch models. Furthermore, the strict stationarity assumption has been adopted in various nonparametric and semiparametric approaches, such as kernel and local polynomial estimation (Li and Racine, 2006). The strict stationarity assumption is also vital for tests involving the full distribution; for example, see Hong and White's (1995) test for correct model specification and tests for conditional independence by Su and White (2007) and Wang and Hong (2018).

However, indiscriminately imposed stationarity restrictions on time series models can lead to challenges for empirical studies. The prevalence of structural changes in the real world, such as policy switches, technology progress, or institutional changes, leads to abrupt or smooth structural changes in time series sequences. Empirical studies typically assume that a differenced time series satisfies certain stationarity conditions. However, there may still be a nonexplosive trend that cannot be eliminated by differencing. Failure to acknowledge such nonstationarity *a priori* while continuing to impose the stationarity assumption will result in a lack of consistency in inference and estimation. In the forecasting literature, density forecasts can provide more insight than point forecasts for macroeconomic risk management (see, e.g., Diebold, Gunther, and Tay, 1998; Diebold, Hahn, and Tay, 1999). However, owing to the instability of macroeconomic and financial time series (Rossi, 2013), density forecasts may deliver suboptimal predictions in case of time-varying underlying density (e.g., Rossi and Sekhposyan, 2013; González-Rivera and Sun, 2017). In financial risk management, the strict stationarity of financial returns is a critical condition for assessing the probability of extreme events (e.g., Koedijk, Schafgans, and de Vries, 1990; Quintos, Fan, and Phillips, 2001; Lin and Kao, 2008). In addition, studying the nonlinear dependence of financial variables requires testing the constancy of copulas, which may change with variation in the joint distribution over time (Busetti and Harvey, 2011; Manner, Stark, and Wied, 2019). It is therefore essential to test for stationarity in an underlying time series process before proposing a concrete analytical tool.

There is extensive literature on stationarity testing, most commonly investigating two types: the "strict stationarity", by which all finite-dimensional distributions do not depend on time, and "pth-order stationarity", which suggests the existence and time invariance of moments up to pth order. When $p = 2$, it indicates weak stationarity where the mean and covariance structure of a time series exist and do not depend on time. Related studies focus on testing whether economic processes behave like a random walk or are stationary around a certain trend. Unit-root tests—including the Dickey–Fuller (Dickey and Fuller, 1979), augmented Dickey–Fuller (ADF) (Dickey and Fuller, 1981), and Phillips and Perron's (1988) tests—are typical tests whose null hypotheses are unit roots against (trend) weak stationarity alternative hypotheses. Tests designed to examine the null of trending stationarity against the unit-root alternative constitute another test type and include the Kwiatkowski et al.'s (1992; hereafter KPSS) test and its variants (Leybourne and McCabe, 1994; Xiao, 2001; Busetti and Taylor, 2003, 2004; Hobijn, Franses, and Ooms, 2004; Cavaliere and Taylor, 2005; Xiao and Lima, 2007; Cavaliere and Taylor, 2009).

Investigating stationarity in the first two moments is sufficient for linear time series models to produce consistent estimation and valid inference. Nevertheless, many nonlinear time series models and distribution-based approaches require strict stationarity. For example, higher-order moments, including skewness and kurtosis, have economic interpretations in financial time series analysis. Tests for weak stationarity cannot provide valuable information on the stability of such higher-order moments. However, to the best of our knowledge, there are relatively few studies on strict stationarity tests. Kapetanios (2009) tests strict stationarity by using a nonparametric marginal density estimator, whereas Busetti and Harvey (2010) formulate tests based on weighted quantile indicators. Francq and Zakoïan (2012) propose a strict stationarity test under the GARCH framework. Hong, Wang, and Wang (2017) develop a model-free test by estimating a nonparametric time-varying characteristic function (CF) and comparing it with the empirical characteristic function (ECF). Additionally, Guo, Li, and Li (2019) propose a strict stationarity test under the double autoregressive (DAR) framework. Although existing studies have addressed strict stationarity testing, the tests they discuss exhibit certain undesired features. For smoothed nonparametric tests, such as those in Kapetanios (2009) and Hong et al. (2017), the power is affected by the choice of tuning parameters. Tests under certain model specifications, such as those of Francq and Zakoïan (2012) and Guo et al. (2019), can offer misleading results if the model is misspecified.

A related strand of literature explores the estimation and testing of structural breaks in joint distributions. Structural breaks may be a source of nonstationarity in a distribution, and strict stationarity can be tested using tests by Inoue (2001) based on empirical distribution functions (EDFs) and by Fu, Hong, and Wang (2022a) based on ECFs. However, although these tests have power for various types of nonstationarity, the asymptotic theories developed for a finite number of structural breaks in the distribution could be invalid under certain types. In addition, as all these tests require trimming the observations at the boundary regions of the sample, nonstationarity that exists only in the boundary regions may be missed.

Motivated by the importance of the strict stationarity test and the fact that existing approaches have various undesired features, we develop a novel test based on the discrete Fourier transform (DFT), a useful tool applied in a second-order stationarity test by Dwivedi and Subba Rao (2011) and Jentsch and Subba Rao (2015), and a structural change test in factor models by Fu, Hong, and Wang (2022b). We construct the test by comparing the DFT of a residual process from a generalized regression with a zero spectrum. As the CF has a one-to-one correspondence with the cumulative distribution function (CDF), we estimate the CF using the ECF; if the underlying time series process is strictly stationary, then the ECF should be consistent for the true CF. Thus, the DFT of the estimated residual process will converge to a zero spectrum in the frequency domain. In contrast, if the time series process is nonstationary, the ECF is no longer a consistent estimator for a time-varying CF. Therefore, the time-varying information could be extracted from the estimated residual process, and the DFT will converge to a nonzero spectrum. Based on this intuition, we construct a Cramér–von Mises (CvM)-type test statistic that compares the DFT of the estimated residuals and a zero spectrum under the null hypothesis of strict stationarity. This idea of using CF for hypothesis testing is similar to that of Su and White (2007), Chen and Hong (2010), Hong et al. (2017), Wang and Hong (2018), and Fu and Hong (2019). Our test has several advantages. First, unlike some existing strict stationarity tests, which are useful only under specific parametric assumptions, our DFT test is model-free and can detect various types of nonstationarity. The second advantage is that, unlike the existing nonparametric kernel-based strict stationarity tests, which depend on the choice of bandwidth, our DFT test is free of tuning parameters; choosing a smoothing or trimming parameter when computing our test statistic is not necessary. This method can detect a class of local alternatives that converge to the null hypothesis at a parametric convergence rate of $T^{-1/2}$, which is faster than the existing nonparametric tests. Third, compared with existing tests designed to test distributional breaks, our test does not require trimming data and can detect nonstationarity in the boundary regions. As a final advantage of our test, the asymptotic theory is relatively general for various types of nonstationarity, and only very weak conditions are required for the alternatives.

The remainder of this paper is organized as follows. Section 2 introduces our hypotheses of interest and proposes the CvM-type test statistic. In Section 3, we derive the asymptotic distribution and investigate the asymptotic power properties of our test. Section 4 proposes a pth-order moment stationarity test using the DFT. In Section 5, we study the finite-sample performance of our test using Monte Carlo simulation; we apply our test to the exchange rate market in Section 6. Section 7 concludes the paper. Proofs of the main results are relegated to the Mathematical Appendix. Further technical analysis, simulations, and application results are reported in the Supplementary Material.

Notation. We use $\mathbf{i}$ to denote the imaginary number, $\sqrt{-1} = \mathbf{i}$. For an $m \times n$ complex-valued matrix $A = (a_{ij})$, we use a_{ij} to denote the (i,j)th element, $A^* = (a_{ji}^*)$ to denote its complex conjugate, and $A' = (a_{ji})$ to denote its transpose. Let $||A|| \equiv (\sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|^2)^{1/2}$ denote the Frobenius norm of A, where $|\cdot|$ denotes the modulus of a complex number and $\equiv$ signifies a definitional relationship. We use $\xrightarrow{p}$, $\xrightarrow{d}$, and $\Rightarrow$ to denote convergence in probability, convergence in distribution, and weak convergence, respectively.

2. HYPOTHESES AND TEST STATISTIC

2.1. Hypotheses of Interest

Let $\{X_t\}_{t=1}^{\infty}$ be a d-dimensional time series process, where $d \geq 1$ denotes a fixed constant. For a given collection of time indices $(t_1, \ldots, t_m)$, we denote the joint

https://doi.org/10.1017/S0266466622000494 Published online by Cambridge University Press

CDF of $\{X_{t_1},\ldots,X_{t_m}\}$ as

$$F_{X_{t_1},\ldots,X_{t_m}}(x_1,\ldots,x_m) = P(X_{t_1} \le x_1,\ldots,X_{t_m} \le x_m).$$

If $\{X_t\}_{t=1}^{\infty}$ is strictly stationary, then $F_{X_{t_1},\ldots,X_{t_m}}(x_1,\ldots,x_m)$ depends on the time indices $(t_1,\ldots,t_m)$ only through the differences between them for any $m \ge 1$. To test whether a time series process is strictly stationary, we test the following null hypothesis:

$$\mathbb{H}_0 : F_{X_{t_1},\ldots,X_{t_m}}(x_1,\ldots,x_m) \;=\; F_{X_{t_1+k},\ldots,X_{t_m+k}}(x_1,\ldots,x_m) \tag{2.1}$$

for any collection of admissible time indices $(t_1,\ldots,t_m)$, realization $(x_1,\ldots,x_m)$ with $m \ge 1$, and integer $k \ge 1$. For a collection of pre-specified time indices $(t_1,\ldots,t_m)$, we assume $t_1 < t_2 < \cdots < t_m$, where a special case is $t_j = t+j-1$, for $j = 1,\ldots,m$. To simplify the expression of the joint distribution, we define the following dm-dimensional time series

$$Y_t \equiv (X_t', X_{t+t_2-t_1}', \ldots, X_{t+t_m-t_1}')',$$

where we follow Hong et al. (2017) to suppress the dependence of Y_t on $(t_1,t_2,\ldots,t_m)$ and m. The joint CDF of X_t, $X_{t+t_2-t_1}$, $\ldots$, and $X_{t+t_m-t_1}$ is identical to the CDF of Y_t. Therefore, the null and alternative hypotheses could be rewritten as

$$\mathbb{H}_0 : F_t(y) = P(Y_t \le y) \text{ does not depend on } t \text{ for all } y \in \mathbb{R}^{dm}, \tag{2.2}$$

and

$$\mathbb{H}_A : F_t(y) = P(Y_t \le y) \text{ depends on } t \text{ for some nonnegligible collection of } y \in \mathbb{R}^{dm}.$$

Note that strict stationarity requires that (2.1) hold for all $(t_1,t_2,\ldots,t_m)$ and all m. These values should be specified because some settings may only require certain aspects of strict stationarity. For example, to check m_0th-order stationarity of an AR(p) process,[1] one can let $m = m_0$. Another example is testing stationarity in a copula-based first-order Markov model, in which we can specify $m = 1$ to test whether the marginal distribution changes over time. For more examples, see the applications of strict stationarity in Joe (1997) and Chen and Fan (2006). Compared to the strict stationarity test of Kapetanios (2009), which focuses only on the marginal distribution, our test is more flexible and can test strict stationarity in a joint distribution.

To illustrate, we assume that m is fixed. Consider the CF of $\{Y_t\}$

$$\phi_t(u) \equiv E(e^{\mathbf{i}u'Y_t}) = \int e^{\mathbf{i}u'Y_t}\,\mathrm{d}F_t(y),$$

where $u \in \mathbb{R}^{dm}$ is a $dm \times 1$ nuisance parameter vector. Given that CF is unique for any particular distribution of Y_t, our test is equivalent to testing

[1] Here, the m_0th-order stationarity requires $F_{X_{t_1},\ldots,X_{t_{m_0}}}(x_1,\ldots,x_{m_0}) = F_{X_{t_1+k},\ldots,X_{t_{m_0}+k}}(x_1,\ldots,x_{m_0})$, which differs from the definition of stationarity up to the m_0th-order moment.

$$\mathbb{H}_0 : \phi_t(u) = \phi_0(u) \text{ for all } u \in \mathbb{R}^{dm} \tag{2.3}$$

and for some time-invariant CF $\phi_0(u)$ against

$$\mathbb{H}_A : \phi_t(u) \neq \phi(u) \text{ for some } u \in \mathbb{R}^{dm} \tag{2.4}$$

in a Borel set of positive measure and for any time-invariant CF $\phi(u)$. To achieve consistency of our test against various types of nonstationarity under $\mathbb{H}_A$, we must check whether (2.3) holds for all t and all $u \in \mathbb{R}^{dm}$ instead of a subset of $\mathbb{R}^{dm}$. This usually requires consistent nonparametric estimation for $\phi_t(u)$ (e.g., Hong et al., 2017). In this paper, we propose a novel approach that avoids undesired features in a smoothed nonparametric approach.

2.2. Discrete Fourier Transform

The idea of our test is to capture the time-varying behavior of $\phi_t(u)$ without directly estimating it. Unlike Hong et al. (2017), who rely on smoothed nonparametric estimation of $\phi_t(u)$, we use the DFT. To extract the potential time-varying features of $\phi_t(u)$, we consider the following generalized regression:

$$e^{\mathbf{i}u'Y_t} = \phi_t(u) + \varepsilon_t(u), \tag{2.5}$$

where $\varepsilon_t(u)$ is a generalized error process with $E[\varepsilon_t(u)] = 0$. For each fixed $u \in \mathbb{R}^{dm}$, (2.5) can be considered as a complex-valued time-varying coefficient model with a local constant (e.g., Robinson, 1989, 1991). The literature on the nonparametric estimation of $\phi_t(u)$ includes Ramsay (1991), Cai (2007), and Hong et al. (2017). In particular, the strict stationarity test of Hong et al. (2017) is based primarily on the comparison of the smoothed nonparametric consistent estimate of $\phi_t(u)$ with the ECF under the null. Their generalized Hausman-type test relies on nonparametric kernel estimation and can detect a class of local alternatives at a nonparametric rate. Additionally, the choice of bandwidth can significantly impact the smoothed nonparametric test results; two practitioners may draw opposing conclusions from the same dataset with different bandwidth choices. We propose our novel test based on the DFT to avoid these undesired features in a smoothed nonparametric test.

Specifically, we consider the following generalized residual process:

$$\begin{aligned}\hat{\varepsilon}_t(u) &= e^{\mathbf{i}u'Y_t} - \hat{\phi}_0(u) \\ &= \left[\phi_t(u) - \hat{\phi}_0(u)\right] + \varepsilon_t(u),\end{aligned}$$

where $\hat{\phi}_0(u) \equiv T^{-1}\sum_{t=1}^{T} e^{\mathbf{i}u'Y_t}$ is the ECF, and T is the sample size. Under $\mathbb{H}_0 : \phi_t(u) = \phi_0(u)$, for all t and u, the ECF is consistent for the CF. Thus, $\hat{\varepsilon}_t(u)$ is essentially a stationary zero-mean process in the frequency domain indexed by $u \in \mathbb{R}^{dm}$, and it weakly converges to a zero spectrum. However, under $\mathbb{H}_A : \phi_t(u) \neq \phi_0(u)$ for some t and u, $\hat{\varepsilon}_t(u)$ contains the time-varying feature of $\phi_t(u)$ because the ECF cannot consistently estimate a time-varying CF. To extract information from

$\{\hat{\varepsilon}_t(u)\}$ under $\mathbb{H}_A$, we propose a test based on the following empirical process:

$$\hat{A}(u,v) = \frac{1}{T}\sum_{t=1}^{T}\hat{\varepsilon}_t(u)e^{\mathbf{i}v2\pi t/T},$$

where $u \in \mathbb{R}^{dm}$ and $v \in \mathbb{R}$. Note that $\hat{A}(u,v)$ is the DFT of $\hat{\varepsilon}_t(u)$. Because the DFT converts a time series from the time domain to the frequency domain, $\hat{A}(u,v)$ contains complete time series information of $\hat{\varepsilon}_t(u)$. Hence, by comparing the spectrum of the estimated residual process with the null (zero) spectrum, we can infer whether the underlying time series is strictly stationary.

Note that $\hat{A}(u,v)$ can be written equivalently in terms of a demeaned Fourier transform of $e^{\mathbf{i}u'Y_t}$. Let $\bar{M}_t(v) \equiv e^{\mathbf{i}v2\pi t/T} - T^{-1}\sum_{s=1}^{T} e^{\mathbf{i}v2\pi s/T}$ be a demeaned Fourier process. Then, we can easily verify

$$\begin{aligned}\hat{A}(u,v) &= \frac{1}{T}\sum_{t=1}^{T}\bar{M}_t(v)\hat{\varepsilon}_t(u) = \frac{1}{T}\sum_{t=1}^{T}\bar{M}_t(v)e^{\mathbf{i}u'Y_t}\\ &= \hat{A}_1(u,v)+\hat{A}_2(u,v),\end{aligned} \tag{2.6}$$

where $\hat{A}_1(u,v) = T^{-1}\sum_{t=1}^{T}\bar{M}_t(v)\phi_t(u)$ and $\hat{A}_2(u,v) = T^{-1}\sum_{t=1}^{T}\bar{M}_t(v)\varepsilon_t(u)$.

The decomposition in (2.6) provides insight into the power of the test based on the use of $\hat{A}(u,v)$ against the alternatives. $\hat{A}_1(u,v)$ and $\hat{A}_2(u,v)$ are clearly the demeaned DFTs of the unknown CF $\phi_t(u)$ and the unobservable generalized error process $\varepsilon_t(u)$, respectively. Under $\mathbb{H}_0$, $\hat{A}_1(u,v) = 0$ for all $(u,v) \in \mathbb{R}^{dm+1}$, and the asymptotic behavior of DFT $\hat{A}(u,v)$ is dominated by $\hat{A}_2(u,v)$, which converges to a zero spectrum in the frequency domain at the typical $\sqrt{T}$-parametric rate. Under $\mathbb{H}_A$, $\hat{A}_1(u,v)$ converges to a nonzero spectrum $\mu(u,v)$ defined in Proposition 3.2, and $\hat{A}(u,v)$ is dominated by $\hat{A}_1(u,v)$, which captures the nonstationary features of $\phi_t(u)$. Thus, we can detect the nonstationarity of an unknown form in the distribution of $\{Y_t\}$.

Interestingly, DFT $\hat{A}(u,v)$ can be viewed as testing independence between Y_t and the rescaled time index t/T. Let $\hat{\psi}(u,v) = T^{-1}\sum_{t=1}^{T} e^{\mathbf{i}u'Y_t+\mathbf{i}v2\pi t/T}$, and $\hat{\lambda}(v) = T^{-1}\sum_{t=1}^{T} e^{\mathbf{i}v2\pi t/T}$, then it follows that

$$\hat{A}(u,v) = \hat{\psi}(u,v) - \hat{\phi}_0(u)\hat{\lambda}(v),$$

where $\hat{\psi}(u,v)$ can be viewed as a pseudo-joint ECF of Y_t and t/T and $\hat{\lambda}(v)$ as the pseudo-marginal ECF of the rescaled time index t/T in the sense that it follows the $U[0,1]$ distribution. Given that $\hat{\phi}_0(u)$ is the marginal ECF of Y_t, our DFT test can be viewed as testing the generalized distance covariance between the CF of Y_t and a pseudo-CF of the rescaled time index t/T. Such a construction is analogous to the distance-covariance approach proposed by Székely, Rizzo, and Bakirov (2007) for independent and identically distributed (i.i.d.) observations and by Zhou (2012) and Davis et al. (2018) for strictly stationary and weakly dependent observations. However, our asymptotic analysis differs substantially from those in Székely et al.

(2007), Zhou (2012), and Davis et al. (2018). The deterministic time index $\{t/T\}_{t=1}^{T}$ is not independent or stationary over time in these studies; hence, their results are not applicable.

Notably, our test can also be cast into the vast literature on model specification tests. Consider the spatial regression in (2.5). When $E(e^{\mathbf{i}u'Y_t})$ is correctly specified, we should have the following pseudo-conditional moment condition:

$$E[\varepsilon_t(u)|t/T] = 0 \text{ for all } u \in \mathbb{R}^{dm},$$

by pretending that t/T is a random variable. According to Bierens (1982, 1990), this conditional moment condition is equivalent to the following set of unconditional moment conditions:

$$E\left[\varepsilon_t(u)e^{\mathbf{i}v2\pi t/T}\right] = 0 \text{ for all } (u,v) \in \mathbb{R}^{dm+1},$$

where $e^{\mathbf{i}v2\pi t/T}$ can be interpreted as the choice of the "generically totally revealing (GTR)" functions, as defined by Stinchcombe and White (1998). As noted by an anonymous referee, $e^{\mathbf{i}v2\pi t/T}$ can be replaced by some other GTR functions, which can also serve as a foundation for alternative tests for strict stationarity.

2.3. Test Statistic

To examine the behavior of $\hat{A}(u,v)$ in the frequency domain, we construct the following CvM-type test statistic:

$$\hat{D} = T\int_{\mathbb{R}^{dm+1}} \left|\hat{A}(u,v)\right|^2 W(u,v)\,\mathrm{d}u\,\mathrm{d}v, \tag{2.7}$$

where $W(u,v) : \mathbb{R}^{dm+1} \to \mathbb{R}^+$ is a nonnegative symmetric weighting function.

Equation (2.7) implies $\hat{D}$ can capture the deviation of $\hat{A}(u,v)$ from a zero spectrum at all possible combinations of (u,v) in the frequency domain. Intuitively, u is a nuisance parameter for the CF. Checking all u ensures that the test statistic does not miss any information in the distribution of Y_t. v is the nuisance parameter introduced by the DFT. The time-varying behavior of $\phi_t(u)$ can be captured by various v values in the frequency domain. Thus, investigating all v ensures the consistency of $\hat{D}$ against various types of nonstationarity. The weighting function $W(u,v)$ enables assigning various weights to (u,v). Superficially, computing $\hat{D}$ requires numerical integration over $(u,v) \in \mathbb{R}^{dm+1}$, which can be time-consuming when dm is moderately large. A major advantage of our test is that, for certain appropriate choices of the weighting functions $W(u,v)$, we can avoid numerical integration to deliver a closed-form expression for $\hat{D}$ with nuisance parameters u and v integrated out.

In practice, we recommend choosing $W(u,v)$ in a product form $W(u,v) = W_1(u)W_2(v)$. Let

$$h_1(y,\tilde{y}) \equiv \int_{\mathbb{R}^{dm}} e^{\mathbf{i}u'(y-\tilde{y})}W_1(u)\,\mathrm{d}u \text{ and } h_2(\tau,\tilde{\tau}) \equiv \int_{\mathbb{R}} e^{\mathbf{i}v2\pi(\tau-\tilde{\tau})}W_2(v)\,\mathrm{d}v.$$

https://doi.org/10.1017/S0266466622000494 Published online by Cambridge University Press

Denote $h_{1st} \equiv h_1(Y_s, Y_t)$, $h_{2st} \equiv h_2(s/T, t/T)$, and $\tilde{h}_{\ell st} \equiv h_{\ell st} - \frac{1}{T}\sum_{t=1}^{T} h_{\ell st} - \frac{1}{T}\sum_{s=1}^{T} h_{\ell st} + \frac{1}{T^2}\sum_{s=1}^{T}\sum_{t=1}^{T} h_{\ell st}$, for $\ell = 1, 2$. Then it is easy to verify that[2]

$$\hat{D} = \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\tilde{h}_{1st}\tilde{h}_{2st}. \tag{2.8}$$

For example, if

$$W_1(u) = (2\pi\gamma_1^2)^{-\frac{dm}{2}} e^{-\frac{\|u\|^2}{2\gamma_1^2}} \text{ and } W_2(v) = (2\pi\gamma_2^2)^{-\frac{1}{2}} e^{-\frac{v^2}{2\gamma_2^2}}, \tag{2.9}$$

then $h_{1st} = e^{-\gamma_1^2\|Y_s - Y_t\|^2/2}$ and $h_{2st} = e^{-2\pi^2\gamma_2^2(s-t)^2/T^2}$, where γ_1 and γ_2 are scale parameters that determine the dispersion of weights assigned to $u \in \mathbb{R}^{dm}$ and $v \in \mathbb{R}$, respectively; if

$$W_1(u) = (2\gamma_1)^{-dm} e^{-\frac{1}{\gamma_1}\sum_{i=1}^{dm}|u_i|} \text{ and } W_2(v) = (2\gamma_2)^{-1} e^{-\frac{|v|}{\gamma_2}}, \tag{2.10}$$

then $h_{1st} = \prod_{i=1}^{dm}[1 + \gamma_1^2(Y_{is} - Y_{it})^2]^{-1}$ and $h_{2st} = [1 + \gamma_2^2 4\pi^2(s-t)^2/T^2]^{-1}$, where γ_1 and γ_2 are scale parameters that have analogous roles as in (2.9); if

$$W_1(u) = \left(c_{dm}\|u\|^{1+dm}\right)^{-1} \text{ and } W_2(v) = \pi^{-1}|v|^{-2}, \tag{2.11}$$

then $h_{1st} = \|Y_s - Y_t\|$ and $h_{2st} = 2\pi|s-t|/T$, where $c_{dm} = \pi^{(dm+1)/2}/\Gamma((dm+1)/2)$ and $\Gamma(\cdot)$ is the complete gamma function: $\Gamma(z) = \int_0^\infty s^{z-1}e^{-s}\mathrm{d}s$. Notice that the choice of weighting function in (2.11) is associated with those used in the distance covariance measure of Székely et al. (2007), Zhou (2012), and Su and Zheng (2017), among others. This indicates that the DFT is equivalent to testing independence between the distribution of Y_t and rescaled time index t/T in the sense that the latter follows the $U[0,1]$ distribution.

3. ASYMPTOTIC PROPERTIES

We now derive the asymptotic properties of $\hat{A}(u,v)$ and $\hat{D}$. Let C denote a generic constant that may vary across lines.

3.1. Assumptions

To establish the asymptotic theory, we first impose the following regularity conditions.

Assumption A.1. (i) $\{Y_t\}$ is a strong mixing process on $\mathbb{R}^{dm}$ with mixing coefficient $\alpha(\cdot)$ such that $\sum_{s=1}^{\infty} s^2\alpha(s) < \infty$ and $\sum_{s=1}^{\infty}\alpha(s)^{\delta/(2+\delta)} < \infty$ and (ii) $E(\|Y_t e^{\mathbf{i}u'Y_t}\|^{2+\delta}) < \infty$, for some $\delta > 0$.

[2]Note that $\hat{D} = T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T}\tilde{h}_{1st}h_{2st} = T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T}h_{1st}\tilde{h}_{2st}$.

https://doi.org/10.1017/S0266466622000494 Published online by Cambridge University Press

Assumption A.2. The weighting function $W(\cdot)$: $\mathbb{R}^{dm+1} \to \mathbb{R}^+$ is a nonnegative and symmetric function such that $\int_{\mathbb{R}^{dm+1}} W(u,v)\,\mathrm{d}u\,\mathrm{d}v < \infty$.

Assumption A.1(i) restricts $\{Y_t\}$ to be an α-mixing process satisfying certain mixing conditions. According to Fan and Yao (2003), a strictly stationary time series $\{Y_t, t = 0, \pm 1, \pm 2, \ldots\}$ is strong mixing when the α-mixing coefficient

$$\alpha(n) \equiv \sup_{A \in \mathcal{F}_{-\infty}^{0}, B \in \mathcal{F}_{n}^{\infty}} |P(A)P(B) - P(AB)| \to 0,$$

as $n \to \infty$, where $n = 1, 2, \ldots,$ and $\mathcal{F}_i^j$ denotes the σ-field of $\{Y_t, i \le t \le j\}$. The strong mixing condition has been widely used in the nonparametric time series literature. It is well known that a variety of time series processes, such as ARMA, bilinear, and GARCH processes, typically satisfy the strong mixing condition. Assumption A.1(ii) imposes a moment condition on $\{Y_t e^{\mathbf{i}u'Y_t}\}$, which is required to verify the tightness of certain empirical processes in the proof of Proposition 3.1. Note that we do not require the density of Y_t to exist because of the application of CF. Thus, the components of Y_t can be either continuous or discrete random variables or a mixture of the two. This is an improvement over the test by Kapetanios (2009), which requires a continuous probability density function that is violated for discretely valued time series. Furthermore, unlike the weak stationarity tests that usually require finite four-plus-order moment conditions, we impose no higher-order moment conditions on Y_t. This accounts for the main difference between strict and weak stationarity tests; the former are based on the entire distribution, allowing for the existence of only low-order moments, whereas the latter are based on the first two moments and require the existence of higher-order moments.

Assumption A.2 imposes a weak integrability condition on the weighting function $W(u,v)$ to guarantee that $\hat{D}$ is well-behaved asymptotically. When $W(u,v) = W_1(u)W_2(v)$, the first two examples given in the previous section (see (2.9) and (2.10) in particular) satisfy the integrability constraint. The last example in (2.11) does not satisfy Assumption A.2. Nevertheless, as discussed in Section S1 of the Supplementary Material, the integrability condition can be removed at the cost of a different proof strategy based on the V-statistic theory and Mercer theorem in functional analysis. Thus, the moment condition on $\{Y_t e^{\mathbf{i}u'Y_t}\}$ can also be replaced by that on h_{1st}, which is satisfied for all three choices of $W_1(\cdot)$ discussed in the last subsection.

As Hong et al. (2017) also consider tests for strict stationarity, it is interesting to compare Assumptions A.1 and A.2 with their assumptions. First, the strong mixing condition in Assumption A.1(i) is weaker than the β-mixing condition in Assumption 1 of Hong et al. (2017), even though they conjecture that they can relax their β-mixing condition to the α-mixing condition. Second, Hong et al.'s (2017) test does not require moment conditions because they rely on the local linear method to estimate the time-varying CF under the alternative, and they do not need to use the empirical process theory in their asymptotic analyses.

https://doi.org/10.1017/S0266466622000494 Published online by Cambridge University Press

In contrast, we impose some moment conditions on $\{Y_t e^{\mathbf{i}u'Y_t}\}$ in Assumption A.1(ii) to verify the tightness of the empirical process. As mentioned earlier, this moment condition can be relaxed using an alternative proof strategy. Third, although the condition on the weighting function is similar to that of Hong et al. (2017), it is weaker. The integrability of the weighting function is a sufficient condition for us to establish the limiting distribution of our test statistic, whereas Assumption 3 in Hong et al. (2017) requires that the weighting function has fourth-order finite moments. As mentioned above, our integrability condition can be removed by resorting to an alternative proof strategy. As pointed out by an anonymous referee, weighting functions for our proposed test are allowed to be nonintegrable, which is an improvement over Hong et al. (2017). Fourth, since Hong et al. (2017) must estimate the time-varying CF under the alternative, they have to impose smoothness conditions on the CF in their Assumption 4 for the power analysis. We need no such conditions as our approach does not require smoothed nonparametric estimation. Fifth, because Hong et al. (2017) require smoothed nonparametric estimation for the time-varying CF, they must specify the conditions on the kernel function and bandwidth; we need neither in our test.

3.2. Asymptotic Null Distributions

Let

$$\Gamma_1(u_1,u_2) \equiv \sum_{j=-\infty}^{\infty} \operatorname{cov}(e^{\mathbf{i}u_1'Y_t}, e^{-\mathbf{i}u_2'Y_{t-j}}) \text{ and}$$

$$\Gamma_2(v_1,v_2) \equiv \int_0^1 e^{\mathbf{i}v_1 2\pi\tau} e^{-\mathbf{i}v_2 2\pi\tau} \mathrm{d}\tau - \int_0^1 e^{\mathbf{i}v_1 2\pi\tau} \mathrm{d}\tau \int_0^1 e^{-\mathbf{i}v_2 2\pi\tau} \mathrm{d}\tau.$$

Clearly, $\Gamma_1(u_1,u_2)$ is a generalized long-run autocovariance of $e^{\mathbf{i}u'Y_t}$. $\Gamma_2(v_1,v_2)$ can be written as $\operatorname{cov}\left(e^{\mathbf{i}v_1 2\pi\tau}, e^{-\mathbf{i}v_2 2\pi\tau}\right)$, which represents the pseudo-covariance between $e^{\mathbf{i}v_1 2\pi\tau}$ and $e^{-\mathbf{i}v_2 2\pi\tau}$ in the sense that $\tau \sim U[0,1]$, the uniform distribution on the interval $[0,1]$.

Next, we provide the asymptotic distribution of $\hat{A}(u,v)$ under the null hypothesis.

PROPOSITION 3.1. *Suppose Assumption A.1 holds. Let $\mathbb{U} \equiv [-b,b]^{dm}$ and $\mathbb{V} \equiv [-c,c]$ be any bounded subsets of $\mathbb{R}^{dm}$ and $\mathbb{R}$, respectively, with $b > 0$ and $c > 0$. Then, under $\mathbb{H}_0$,*

$$\sqrt{T}\hat{A}(u,v) \Rightarrow S(u,v) \text{ on } \mathbb{W} \equiv \mathbb{U} \times \mathbb{V} \text{ as } T \to \infty,$$

where $S(u,v)$ is a complex-valued Gaussian process with a covariance kernel given by

$$\mathcal{K}_0(w_1,w_2) \equiv \Gamma_1(u_1,u_2)\Gamma_2(v_1,v_2),$$

where $w_l = \left(u_l', v_l\right)' \in \mathbb{W}$ for $l = 1,2$.

https://doi.org/10.1017/S0266466622000494 Published online by Cambridge University Press

Proposition 3.1 presents the asymptotic null distribution of $\hat{A}(u,v)$ scaled by $\sqrt{T}$, which is a complex-valued Gaussian process. The proof of Proposition 3.1 is primarily based on the central limit theorem for α-mixing processes; see, e.g., Theorem 5.20 in White (2001). Under $\mathbb{H}_0$, $\hat{A}_2(u,v)$ is the dominant term that converges to a complex-valued Gaussian process. Interestingly, when the underlying time series satisfies the null hypothesis, the covariance kernel of the limiting distribution of $\hat{A}_2(u,v)$ contains two components: the generalized long-run covariance of $\varepsilon_t(u_1)$ and $\varepsilon_t(u_2)$ and the pseudo-covariance introduced by DFT. Intuitively, under $\mathbb{H}_0$, the CF of Y_t is time-invariant such that Y_t is independent of time t/T. Hence, the variance of the DFT is asymptotically equivalent to the product of the generalized long-run autocovariance of $e^{\mathbf{i}u'Y_t}$ and the pseudo-covariance of $e^{\mathbf{i}v2\pi\tau}$.

The following theorem establishes the asymptotic distribution of the test statistic $\hat{D}$ under the null hypothesis.

THEOREM 3.1. *Suppose Assumptions A.1 and A.2 hold. Then, under $\mathbb{H}_0$,*

$$\hat{D} \xrightarrow{d} \int_{\mathbb{R}^{dm+1}} |S(u,v)|^2 W(u,v)\mathrm{d}u\mathrm{d}v \text{ as } T \to \infty,$$

where $S(u,v)$ is the complex-valued Gaussian process defined in Proposition 3.1.

Theorem 3.1 presents the asymptotic distribution of $\hat{D}$ under $\mathbb{H}_0$. This suggests that the integral of the square of $\sqrt{T}\hat{A}(u,v)$ weighted by $W(u,v)$ converges in distribution to an analogous integral of the complex-valued Gaussian process $S(u,v)$. We note that the weak convergence in Proposition 3.1 may not hold when $\|u\|$ or $|v|$ increases to infinity. Owing to the integrability of the weighting function $W(u,v)$, which assigns smaller weights as $\|u\|$ or $|v|$ increases, the impact caused by the nonconvergent region is asymptotically negligible. Consequently, $\hat{D}$ can check the deviation of $\hat{A}(u,v)$ from the zero spectrum over the entire frequency domain for $(u,v) \in \mathbb{R}^{dm+1}$. Nevertheless, $\hat{D}$ is not asymptotically pivotal because it depends on an unknown data generating process (DGP). In the following, we will propose a resampling method to obtain the asymptotic critical values. In response to an anonymous referee's question regarding the possibility of standardizing the statistic $\hat{D}$ to obtain an asymptotically pivotal test within our framework, we have illustrated this difficulty and discussed this issue in Section S2 of the Supplementary Material. This phenomenon does not pertain only to our test; it is true for a wide class of so-called nonsmoothing specification tests; see Chapter 13 of Li and Racine (2006). In any case, one should not interpret the nonasymptotic pivotal property of such a test as a serious drawback in comparison with the kernel-based smoothed nonparametric tests (e.g., Hong et al., 2017) simply because of the need to rely on resampling methods to make asymptotic inferences. It is well known that, even if many smooth tests are asymptotically pivotal under the null hypothesis, one cannot always rely on using the asymptotic critical values to make reliable inferences in finite samples, and resampling methods are frequently needed.

3.3. Asymptotic Power

In this subsection, we investigate the asymptotic properties of $\hat{D}$ under the alternative hypotheses, including both global and local properties.

Under the alternative, the process $\{Y_t\}$ is generally T-dependent; thus, we should write Y_{Tt} for Y_t. However, for notational simplicity, we continue to use the notation Y_t when there is no risk of confusion. Following the lead of Su and White (2010), we define the strong mixing coefficients $\alpha_T(j)$ as follows:

$$\begin{aligned}\alpha_T(j) \equiv \sup\nolimits_{1\le l\le T-j}\{&P(A\cap B)-P(A)P(B)|A\in\sigma(Y_{Tt}:1\le t\le l),\\ &B\in\sigma(Y_{Tt}:l+j\le t\le T)\},\ \text{for } j\le T-1;\\ \alpha_T(j)\equiv 0,\ &\text{for } j\ge T.\end{aligned}$$

The coefficient of strong mixing is defined by $\bar{\alpha}(j) \equiv \sup_{T\in\mathbb{N}}\alpha_T(j)$, for $j\in\mathbb{N}$, where $\mathbb{N}$ signifies the set of natural numbers. Note that $\bar{\alpha}(j)$ collapses to the usual strong mixing coefficient $\alpha(\cdot)$ under strict stationarity. We modify Assumption A.1 as follows.

Assumption A.3. (i) $\{Y_{Tt}\}$ is a double-array strong mixing process on $\mathbb{R}^{dm}$ with mixing coefficient $\bar{\alpha}(\cdot)$ such that $\sum_{s=1}^{\infty}\bar{\alpha}(s)<\infty$ and (ii) $\phi_t(u)=\phi(u,t/T)$ and $\mu_2\equiv\int_{\mathbb{R}^{dm+1}}|\mu(u,v)|^2W(u,v)\,\mathrm{d}u\,\mathrm{d}v>0$ where $\mu(u,v)\equiv\int_0^1\phi(u,\tau)e^{\mathbf{i}v2\pi\tau}\mathrm{d}\tau-\int_0^1\phi(u,\tau)\mathrm{d}\tau\int_0^1e^{\mathbf{i}v2\pi\tau}\mathrm{d}\tau$.

Assumption A.3(i) states that $\{Y_{Tt}\}$ is a double-array strong mixing process. Assumption A.3(ii) imposes conditions on the time-varying CF $\phi_t(u)$ to ensure nontrivial power under the global alternative. Note that we do not require $\phi(u,\cdot)$ to be continuous, as in Hong et al.'s (2017) test. This indicates that $\phi(u,\cdot)$ can be either discrete or continuous with respect to $\cdot$, another improvement over Hong et al.'s (2017) test. In addition, it is not necessary to impose any moment conditions on $\{Y_t\}$ because there is no need to establish the tightness condition when the global alternative is satisfied.

The following proposition shows that, under the global alternative hypothesis, $\hat{A}(u,v)$ converges to a nonzero spectrum.

PROPOSITION 3.2. *Suppose Assumption A.3 holds. Then, under $\mathbb{H}_A$,*

$$\hat{A}(u,v)\xrightarrow{p}\mu(u,v)\ \textit{as}\ T\to\infty,$$

where $\mu(u,v)$ can be expressed as $\operatorname{cov}\left[\phi(u,\tau),e^{\mathbf{i}v2\pi\tau}\right]$ *to signify that it is a pseudo-covariance between $\phi(u,\tau)$ and $e^{\mathbf{i}v2\pi\tau}$ in the sense that τ follows the $U[0,1]$ distribution.*

Proposition 3.2 shows that $\hat{A}(u,v)$ converges in probability to a nonzero spectrum in the frequency domain. This occurs because one can show that $\hat{A}_2(u,v)=O_P(T^{-1/2})$, and, by the Riemann summation approximation of integrals,

$$\hat{A}_1(u,v) = \frac{1}{T}\sum_{t=1}^{T}\phi_t(u)e^{\mathbf{i}v2\pi t/T} - \left[\frac{1}{T}\sum_{t=1}^{T}\phi_t(u)\right]\left[\frac{1}{T}\sum_{t=1}^{T}e^{\mathbf{i}v2\pi t/T}\right]$$
$$\to \int_0^1 \phi(u,\tau)e^{\mathbf{i}v2\pi\tau}\,\mathrm{d}\tau - \int_0^1 \phi(u,\tau)\,\mathrm{d}\tau \int_0^1 e^{\mathbf{i}v2\pi\tau}\,\mathrm{d}\tau \;=\; \mu(u,v).$$

Because, under $\mathbb{H}_A$, $\phi_t(u) = \phi(u,t/T)$ is a nonconstant function of time t/T, $\mathrm{cov}\left[\phi(u,\tau), e^{\mathbf{i}v2\pi\tau}\right]$ differs from 0 for some (u,v) in a Borel set of positive measures. This ensures the nontrivial power of our test under Assumption A.2(ii).

Next, we discuss the asymptotic global power of our test statistic $\hat{D}$.

THEOREM 3.2. *Suppose Assumptions A.2 and A.3 hold. Then, under $\mathbb{H}_A$, $P(\hat{D} > c_T) \to 1$ as $T \to \infty$ for any nonrandom sequence $c_T = o(T)$.*

Theorem 3.2 shows that our test $\hat{D}$ diverges to infinity in probability at rate T provided that $\mu_2 > 0$. It is consistent, then, against a variety of global alternatives that do not rely on specific DGPs, including abrupt and smooth changes in the distribution of Y_t. Compared with those parametric tests in Francq and Zakoïan (2012) and Guo et al. (2019), which are designed to have power under the GARCH and DAR models, respectively, our test has power against a larger class of alternatives. Compared with Hong et al.'s (2017) nonparametric kernel-based test, our test does not require the CF to be continuous in time and does not require the choice of a bandwidth parameter for nonparametric kernel estimation.

To study the asymptotic local power property of our test, we consider the following sequence of Pitman's local alternatives:

$$\mathbb{H}_A(\Delta_T): F_t(y) = F_0(y) + \Delta_T k_t(y) \text{ for all } (y,t), \tag{3.1}$$

where $F_0(y)$ is a time-invariant CDF, $k_t(y) = k(y,t/T)$ for some measurable function $k(\cdot,\cdot)$, and $\Delta_T \to 0$ measures the speed at which the time-varying CDF $F_t(\cdot)$ deviates away from the time-invariant CDF $F_0(\cdot)$. We then take the Fourier transform of (3.1) to obtain the following equivalent representation:

$$\mathbb{H}_A(\Delta_T): \phi_t(u) = \phi_0(u) + \Delta_T\theta_t(u) \text{ for all } (u,t), \tag{3.2}$$

where $\phi_0(u) = \int e^{\mathbf{i}u'y}\mathrm{d}F_0(y)$ and $\theta_t(u) \equiv \theta(u,t/T) \equiv \int e^{\mathbf{i}u'y}\mathrm{d}k_t(y)$ denote the Fourier transform of $F_0(y)$ and $k_t(y)$, respectively.

To obtain the asymptotic distribution of the DFT under $\mathbb{H}_A(\Delta_T)$, we add the following assumption.

Assumption A.4. (i) $\{Y_{Tt}\}$ is a strong mixing process on $\mathbb{R}^{dm}$ with mixing coefficient $\bar{\alpha}(\cdot)$ such that $\sum_{s=1}^{\infty} s^2\bar{\alpha}(s) < \infty$ and $\sum_{s=1}^{\infty}\bar{\alpha}(s)^{\delta/(2+\delta)} < \infty$; (ii) $\max_{1\le t\le T} E(\|Y_{Tt}e^{\mathbf{i}u'Y_{Tt}}\|^{2+\delta}) < \infty$ for some $\delta > 0$; and (iii) $\theta_t(u) = \theta(u,t/T)$, and $\theta(\cdot,\cdot)$ is Lipschitz continuous with respect to each of its arguments: $|\theta(u_1,v) - \theta(u_2,v)| \le C\|u_1 - u_2\|$ and $|\theta(u,v_1) - \theta(u,v_2)| \le C|v_1 - v_2|$.

Assumption A.4(i) and (ii) strengthens the mixing condition in Assumption A.3(i) and requires some moment conditions on $\{Y_{Tt}e^{\mathbf{i}u'Y_{Tt}}\}$ in order to verify the tightness condition for the proof of Theorem 3.3. The Lipschitz continuity condition in Assumption A.4(iii) significantly facilitates the verification process.

THEOREM 3.3. *Suppose Assumptions A.2 and A.4 hold. Then, under $\mathbb{H}_A(\Delta_T)$ with $\Delta_T = T^{-1/2}$, we have that*

$$\hat{D} \xrightarrow{d} \int_{\mathbb{R}^{dm+1}} |\zeta(u,v) + \mathcal{S}(u,v)|^2 W(u,v)\,\mathrm{d}u\,\mathrm{d}v,$$

where $\zeta(u,v) \equiv \operatorname{cov}[\theta(u,\tau), e^{\mathbf{i}v2\pi\tau}] = \int_0^1 \theta(u,\tau)e^{\mathbf{i}v2\pi\tau}\mathrm{d}\tau - \int_0^1 \theta(u,\tau)\mathrm{d}\tau \int_0^1 e^{\mathbf{i}v2\pi\tau}\mathrm{d}\tau$ is a pseudo-covariance between $\theta(u,\tau)$ and $e^{\mathbf{i}v2\pi\tau}$ in the sense that $\tau \sim U[0,1]$, and $\mathcal{S}(u,v)$ is a complex-valued Gaussian process with covariance kernel given by

$$\mathcal{K}_1(w_1,w_2) \equiv \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T} \operatorname{cov}(e^{\mathbf{i}u_1'Y_t}, e^{-\mathbf{i}u_2'Y_s})\bar{M}_t(v_1)\bar{M}_s(v_2)^*,$$

where $w_l = \left(u_l', v_l\right)' \in \mathbb{W}$ for $l = 1,2$.

Under the mixing condition in Assumption A.4(i), $\mathcal{K}_1(w_1,w_2)$ is well behaved for all $w_1, w_2 \in \mathbb{W}$. Theorem 3.3 indicates that our test $\hat{D}$ has nontrivial power against the class of local alternatives $\mathbb{H}_A(\Delta_T)$ with the parametric convergence rate $\Delta_T = T^{-1/2}$. The convergence rate is faster than the nonparametric rate of Kapetanios (2009). Although Kapetanios (2009) focuses on the univariate case, his test can clearly only detect local alternatives that converge to the null at the rate $T^{-1/2}h^{-d/2}$ when testing for strict stationarity of a d-dimensional time series using bandwidth h in kernel estimation. Such a nonparametric test suffers from the "curse of dimensionality", and, as d amplifies, the convergence rate can be slow owing to the use of nonparametric smoothing. Similarly, Hong et al.'s (2017) test can only detect a class of local alternatives that converge to the null at the rate of $T^{-1/2}h^{-1/4}$, which is also impacted adversely by bandwidth h. Our DFT method avoids the choice of the bandwidth parameter and the slow nonparametric convergence rate.

The strict stationarity test $\hat{D}$ proposed in this paper does not require the choice of any tuning parameter and can detect a class of local alternatives at a typical parametric convergence rate. The cost of using DFT to construct such a nonsmooth test statistic is that the test statistic is not asymptotically pivotal even under the null hypothesis so that the null limiting distribution of our statistic depends on the specific DGP. In practice, we need to rely on certain resampling methods to obtain the critical values or p-values. This should not be considered a drawback for our DFT-based nonsmooth test, however, as resampling methods are typically needed to obtain reliable inferences in finite samples even for a kernel-based smooth test.

3.4. Dependent Wild Bootstrap

Because our test is not asymptotically pivotal, in this section, we propose applying the dependent wild bootstrap (DWB) procedure of Shao (2010) to obtain bootstrap p-values. Also see Leucht and Neumann (2013) and Doukhan et al. (2015) for applications of DWB for hypotheses testing.

Our bootstrap procedure is as follows:

1. Given the observed sample $\{Y_t\}_{t=1}^T$, compute the test statistic $\hat{D}$ as (2.7).
2. Generate $\{\eta_t^\star\}_{t=1}^T$ according to the following AR(1) process:

$$\eta_t^\star = e^{-\frac{1}{b_T}}\eta_{t-1}^\star + \nu_t,$$

where $\nu_t \sim$ i.i.d. $N(0, 1-e^{-2/b_T})$, b_T is the block length such that $1/b_T + b_T/T = o(1)$, and $\eta_0^\star$ is drawn from the stationary distribution $N(0,1)$.
3. Denote the bootstrap empirical process as[3]

$$\hat{A}^\star(u,v) = \frac{1}{T}\sum_{t=1}^{T}\bar{M}_t(v)\hat{\varepsilon}_t(u)\eta_t^\star.$$

Calculate the bootstrap test statistic $\hat{D}^\star = T\int_{\mathbb{R}^{dm+1}}\left|\hat{A}^\star(u,v)\right|^2 W(u,v)\,\mathrm{d}u\,\mathrm{d}v$. Specifically, if the weighting function admits a product form such that $W(u,v) = W_1(u)W_2(v)$, then $\hat{D}^\star$ can be computed as $\hat{D}^\star = T^{-1}\sum_{t=1}^T\sum_{s=1}^T \tilde{h}_{1st}\tilde{h}_{2st}\eta_t^\star\eta_s^\star$, where $\tilde{h}_{1st}$ and $\tilde{h}_{2st}$ are defined as in Section 2.3.
4. Repeat Steps 2 and 3 for B times to obtain bootstrap test statistics $\{\hat{D}_b^\star\}_{b=1}^B$, where B is the number of bootstraps. Compute the bootstrap p-values by $p_B^\star = B^{-1}\sum_{b=1}^B \mathbf{1}(\hat{D}_b^\star > \hat{D})$.

In Step 2 above, we follow Doukhan et al. (2015) to generate $\{\eta_t^\star\}_{t=1}^T$ using a Gaussian process. By construction, $\{\eta_t^\star\}_{t=1}^T$ has zero mean, unit variance, and autocovariance function $A_T(\cdot)$ given by

$$A_T(t-s) \equiv E\left(\eta_t^\star\eta_s^\star\right) = e^{-\frac{|t-s|}{b_T}}.$$

Clearly, $A_T(t-s) \to 0$ as $|t-s|/b_T \to \infty$ and it is easy to verify that $\sum_{r=0}^{T-1}A_T(r) = O(b_T)$. Such results will be used frequently in the proof of Theorem 3.4.

Let $P^\star, E^\star$, $\mathrm{cov}^\star$, $\xrightarrow{d^\star}$, and $\xrightarrow{p^\star}$ denote the probability, expectation, covariance, convergence in distribution, and convergence in probability, respectively, under the bootstrap law by conditioning on the observed sample $\{Y_t\}_{t=1}^T$. To study the asymptotic behavior of $\hat{D}^\star$, we add the following assumption.

Assumption A.5. As $T \to \infty$, $1/b_T + b_T/T \to 0$.

Assumption A.5 imposes weak conditions on b_T, which plays a similar role as the block length in the block bootstrap. However, we do not use the block bootstrap,

[3]Note that both $\bar{M}_t(v)$ and $\hat{\varepsilon}_t(u)$ are demeaned so that each has sample mean 0.

https://doi.org/10.1017/S0266466622000494 Published online by Cambridge University Press

because the data generated from that procedure are typically not strictly stationary even with a strictly stationary original time series.

The following theorem establishes the asymptotic validity of the proposed resampling procedure.

THEOREM 3.4. *Suppose Assumptions A.2 and A.5 hold.*

(i) If Assumption A.4 is satisfied and $\{Y_t\}_{t=1}^T$ *satisfies the local alternative hypothesis* $\mathbb{H}_A(\Delta_T)$ *with* $\Delta_T = o(b_T^{-1/2})$, *we have*

$$\hat{D}^\star \xrightarrow{d^\star} \int_{\mathbb{R}^{dm+1}} |\mathcal{S}(u,v)|^2 W(u,v)\,\mathrm{d}u\,\mathrm{d}v \text{ in probability,}$$

where $\mathcal{S}(u,v)$ *is as defined in Theorem 3.3 that has the covariance kernel* $\mathcal{K}_1(w_1,w_2)$.

(ii) If either Assumption A.3 or A.4 is satisfied and $\{Y_t\}_{t=1}^T$ *satisfies* $\mathbb{H}_A(\Delta_T)$ *with* $\Delta_T \neq o(b_T^{-1/2})$ *including the special case of the global alternative with* $\Delta_T = 1$, *then* $\hat{D}^\star = O_{P^\star}(\Delta_T^2 b_T)$.

Theorem 3.4(i) reports the asymptotic distribution of $\hat{D}^\star$ under $\mathbb{H}_A(\Delta_T)$ with $\Delta_T = o(b_T^{-1/2})$. Here, we use "$\xrightarrow{d^\star}$ in probability" to denote the weak convergence in probability under the bootstrap law (see, e.g., Giné and Zinn, 1990). When $\mathbb{H}_0$ is satisfied (i.e., $\Delta_T = 0$) for the original sample $\{Y_t\}_{t=1}^T$, we observe that $\mathcal{K}_1(w_1,w_2)$ reduces to $\mathcal{K}_0(w_1,w_2)$, and the asymptotic distribution of $\hat{D}^\star$ coincides with the null limiting distribution of $\hat{D}$. This ensures the correct asymptotic size for the above DWB-based test. When $\Delta_T = T^{-1/2}$ as in the standard local power analysis (see Theorem 3.3), $\hat{D}^\star$ shares the same asymptotic variance as $\hat{D}$ but does not have the drifting term $\zeta(u,v)$ in its limit distribution. This indicates that our DWB-based test continues to offer nontrivial local power against local alternatives converging to the null at rate $T^{-1/2}$.

Theorem 3.4(ii) reports the asymptotic property of $\hat{D}^\star$ when Δ_T is large such that the data deviate more from the null hypothesis than in Theorem 3.4(i). In this case, it is easy to see from a modification of the proof of Theorem 3.4 that $\hat{D}$ diverges to infinity in probability at rate $\Delta_T^2 T$, which is faster than $\Delta_T^2 b_T$ under Assumption A.5, the largest possible rate at which $\hat{D}^\star$ diverges to infinity in probability. This ensures the consistency of our test. In particular, for global alternatives, we learn from Theorem 3.2 that $\hat{D}$ diverges to infinity in probability at rate T and from Theorem 3.4(ii) that $\hat{D}^\star$ diverges to infinity in probability at most at rate b_T. For the power performance of the DWB-based test, this suggests that a small value of b_T is preferred. Nevertheless, a too-small value of b_T may not yield good size performance when strong serial dependence is present in the data.

To implement the above DWB test, we must select the tuning parameter b_T. Although we focus on the nondata-driven case in the proof of the above theory, we recognize that, in practice, b_T should be chosen using a data-driven procedure. We propose adopting the minimum volatility (MV) method recently proposed

https://doi.org/10.1017/S0266466622000494 Published online by Cambridge University Press

by Rho and Shao (2019). Simulation studies demonstrate the good finite-sample performance of the DWB-based test.

4. TESTING THE *p*-TH ORDER STATIONARITY VIA THE DFT

Note that a joint CF can generate moments, if they exist, using an appropriate order of differentiation. In this section, we establish a sequence of derivative tests for the pth-order stationarity. A d-dimensional time series process $\{X_t\}_{t=1}^{\infty}$ is said to be stationary up to order p if, for any admissible collection of time indices $(t_1, t_2, \ldots, t_m)$ and any integer $k \geq 1$, the joint moments of $\{X_{t_1}, X_{t_2}, \ldots, X_{t_m}\}$ up to order p exist and equal those of $\{X_{t_1+k}, X_{t_2+k}, \ldots, X_{t_m+k}\}$. The null hypothesis of pth-order stationarity can then be formulated as

$$\mathbb{H}_0^{(p)} : E\left(X_{t_1,1}^{p_{1,1}} \cdots X_{t_1,d}^{p_{1,d}} X_{t_2,1}^{p_{2,1}} \cdots X_{t_m,d}^{p_{m,d}}\right) = E\left(X_{t_1+k,1}^{p_{1,1}} \cdots X_{t_1+k,d}^{p_{1,d}} X_{t_2+k,1}^{p_{2,1}} \cdots X_{t_m+k,d}^{p_{m,d}}\right)$$

for any integer $k \geq 1$ and all positive integers $p_{1,1}, \ldots, p_{m,d}$ such that $1 \leq \sum_{i=1}^{m} \sum_{j=1}^{d} p_{i,j} \leq p$. Equivalently, we can test the following:

$$\mathbb{H}_0^{(p)} : E\left(X_{t,1}^{p_{1,1}} \cdots X_{t,d}^{p_{1,d}} X_{t+t_2-t_1,1}^{p_{2,1}} \cdots X_{t+t_m-t_1,d}^{p_{m,d}}\right) \text{ does not depend on } t$$

for any pre-specified time indices $(t_1, t_2, \ldots, t_m)$ and any dm-dimensional nonnegative integer-valued vector

$$\tilde{p} = (p_{1,1}, \ldots, p_{1,d}, p_{2,1}, \ldots, p_{m,d})$$

satisfying $1 \leq \|\tilde{p}\|_1 \equiv \sum_{i=1}^{d} \sum_{j=1}^{m} p_{i,j} \leq p$, where $\|\cdot\|_1$ denotes the L_1 norm of $\cdot$. Given that

$$Y_t = (X_t', X_{t+t_2-t_1}', \ldots, X_{t+t_m-t_1}')',$$

we take the $\tilde{p}$th-order partial derivatives of the CF of Y_t, i.e., $\phi_t(u)$, with respect to $u = (u_1, u_2, \ldots, u_{dm})'$, and let $u_l = 0$, for $l = 1, \ldots, dm$:

$$\phi_t^{(\tilde{p})}(0) = \left. \frac{\partial^{\|\tilde{p}\|_1} \phi_t(u)}{\partial u_1^{p_{1,1}} \partial u_2^{p_{1,2}} \cdots \partial u_{dm}^{p_{m,d}}} \right|_{u=\mathbf{0}} = \mathbf{i}^{\|\tilde{p}\|_1} E\left(X_{t,1}^{p_{1,1}} \cdots X_{t,d}^{p_{1,d}} X_{t+t_2-t_1,1}^{p_{2,1}} \cdots X_{t+t_m-t_1,d}^{p_{m,d}}\right).$$

Because the derivatives of a CF uniquely determine their corresponding moments, testing the pth-order stationarity can be represented as testing

$$\mathbb{H}_0^{(p)} : \ \phi_t^{(\tilde{p})}(0) = \phi_0^{(\tilde{p})}(0) \ \text{ for all multi-indices } \tilde{p} \text{ with } 1 \leq \|\tilde{p}\|_1 \leq p$$

and for some constant $\phi_0^{(\tilde{p})}(0)$, against

$$\mathbb{H}_A^{(p)} : \ \phi_t^{(\tilde{p})}(0) \neq \phi^{(\tilde{p})}(0) \ \text{ for some multi-indices } \tilde{p} \text{ with } 1 \leq \|\tilde{p}\|_1 \leq p,$$

and for all constants $\phi^{(\tilde{p})}(0)$. For notational simplicity, we suppress the dependence of $\phi_t^{(\tilde{p})}(0)$ on the argument 0 and denote $\phi_t^{(\tilde{p})} \equiv \phi_t^{(\tilde{p})}(0)$.

https://doi.org/10.1017/S0266466622000494 Published online by Cambridge University Press

Let $\varphi_t^{(\tilde{p})} \equiv \left.\frac{\partial^{\|\tilde{p}\|_1} e^{\mathbf{i}u'Y_t}}{\partial u_1^{p_{1,1}} \partial u_2^{p_{1,2}} \cdots \partial u_{dm}^{p_{m,d}}}\right|_{u=\mathbf{0}} = \mathbf{i}^{\|\tilde{p}\|_1} X_{t,1}^{p_{1,1}} \cdots X_{t,d}^{p_{1,d}} X_{t+t_2-t_1,1}^{p_{2,1}} \cdots X_{t+t_m-t_1,d}^{p_{m,d}}$. Then, following analogous reasoning for (2.5), we consider the following pseudo-regression:

$$\varphi_t^{(\tilde{p})} = \phi_t^{(\tilde{p})} + \varepsilon_t^{(\tilde{p})}, \tag{4.1}$$

where $\varepsilon_t^{(\tilde{p})}$ is an error term with $E(\varepsilon_t^{(\tilde{p})}) = 0$. Intuitively, if the time series $\{X_t\}_{t=1}^T$ is pth-order stationary, the sample analog $\hat{\phi}_0^{(\tilde{p})} \equiv T^{-1} \sum_{t=1}^T X_{t,1}^{p_{1,1}} \cdots X_{t,d}^{p_{1,d}} X_{t+t_2-t_1,1}^{p_{2,1}} \cdots X_{t+t_m-t_1,d}^{p_{m,d}}$ will be consistent for $\phi_t^{(\tilde{p})} = \phi_0^{(\tilde{p})}$. However, under $\mathbb{H}_A^{(p)}$, some time-varying information of the joint pth-order moments will be captured by the estimated residuals in (4.1); that is, $\hat{\varepsilon}_t^{(\tilde{p})} \equiv \varphi_t^{(\tilde{p})} - \hat{\phi}_0^{(\tilde{p})}$. Hence, we can use the DFT of the residuals to test $\mathbb{H}_0^{(p)}$ against $\mathbb{H}_A^{(p)}$:

$$\begin{aligned}
\hat{A}^{(\tilde{p})}(v) &= \frac{1}{T}\sum_{t=1}^{T} \hat{\varepsilon}_t^{(\tilde{p})} e^{\mathbf{i}v2\pi t/T} \\
&= \frac{1}{T}\sum_{t=1}^{T} \phi_t^{(\tilde{p})} \bar{M}_t(v) + \frac{1}{T}\sum_{t=1}^{T} \varepsilon_t^{(\tilde{p})} \bar{M}_t(v) \\
&\equiv \hat{A}_1^{(\tilde{p})}(v) + \hat{A}_2^{(\tilde{p})}(v).
\end{aligned} \tag{4.2}$$

Following analogous arguments in Section 2, the asymptotic behavior of the DFT $\hat{A}^{(\tilde{p})}(v)$ is dominated by $\hat{A}_2^{(\tilde{p})}(v)$ under $\mathbb{H}_0^{(p)}$, which weakly converges to a zero spectrum in the frequency domain at the $\sqrt{T}$-parametric rate. However, under $\mathbb{H}_A^{(p)}$, $\hat{A}_1^{(\tilde{p})}(v)$ will converge to the following nonzero spectrum:

$$\mu^{(\tilde{p})}(v) \equiv \int_0^1 \phi^{(\tilde{p})}(\tau) e^{\mathbf{i}v2\pi\tau} \mathrm{d}\tau - \int_0^1 \phi^{(\tilde{p})}(\tau)\mathrm{d}\tau \int_0^1 e^{\mathbf{i}v2\pi\tau} \mathrm{d}\tau,$$

where we let $\phi_t^{(\tilde{p})} \equiv \phi^{(\tilde{p})}(t/T)$ be an integrable function of t/T. This implies that the DFT $\hat{A}^{(\tilde{p})}(v)$ is equivalent to a pseudo-covariance of $\phi_t^{(\tilde{p})}$ and the Fourier basis function of time t/T in the sense that it follows the $U[0,1]$ distribution.

To examine the behavior of the DFT $\hat{A}^{(\tilde{p})}(v)$ at each $v \in \mathbb{R}$, we construct the following CvM-type test statistic to test the pth-order stationarity:

$$\hat{D}^{(\tilde{p})} = T \int_{\mathbb{R}} \left|\hat{A}^{(\tilde{p})}(v)\right|^2 \tilde{W}(v)\,\mathrm{d}v, \tag{4.3}$$

where $\tilde{W}(\cdot) : \mathbb{R} \to \mathbb{R}^+$ is a nonnegative and symmetric function such that $\int_{\mathbb{R}} \tilde{W}(v)\mathrm{d}v < \infty$. Under conditions analogous to Assumptions A.1 and A.2, together with certain moment restrictions, we can show that under $\mathbb{H}_0^{(p)}$,

$$\hat{D}^{(\tilde{p})} \xrightarrow{d} \int_{\mathbb{R}} \left|S^{(\tilde{p})}(v)\right|^2 \tilde{W}(v)\,\mathrm{d}v,$$

https://doi.org/10.1017/S0266466622000494 Published online by Cambridge University Press

where $S^{(\tilde{p})}(v)$ is a complex-valued Gaussian process with a covariance kernel

$$\mathcal{K}^{(\tilde{p})}(v_1,v_2) \equiv E[S^{(\tilde{p})}(v_1)S^{(\tilde{p})}(v_2)^*] = \Gamma_1^{(\tilde{p})}\Gamma_2(v_1,v_2).$$

Here, $\Gamma_1^{(\tilde{p})} \equiv \sum_{j=-\infty}^{\infty} \operatorname{cov}(\varphi_t^{(\tilde{p})}, \varphi_{t-j}^{(\tilde{p})})$ is a long-run autocovariance, and $\Gamma_2(v_1,v_2)$ is as defined in Section 3.2. Under $\mathbb{H}_A^{(\tilde{p})}$, we can show that the test $\hat{D}^{(\tilde{p})}$ can detect a class of local alternatives that converges to the null hypothesis at the parametric rate $T^{-1/2}$.

Clearly, our derivative test statistic is not asymptotically pivotal. We can apply the DWB, as in Section 3.4, to obtain the bootstrap critical values. The validity of this resampling approach can be established similarly to Theorem 3.4 with certain moment conditions on X_t.

The pth-order stationarity test mentioned above has several frequently used forms. For instance, when $d = 1$ and $m = 1$, we have $\tilde{p} = p$. Thus, the test statistic $\hat{D}^{(\tilde{p})}$ is equivalent to testing pth-order stationarity in a univariate time series process $\{X_t\}$. Specifically, when $p = 2$, $m = 1$, and $d \geq 1$, we can use the proposed derivative test to determine whether the d-dimensional time series $\{X_t\}$ is weakly stationary. Note that weak stationarity requires that both the first two moments and the covariance structure of a time series process be constant over time. Because the derivative test generally imposes restraints on the joint moment of a pre-specified collection of time indices $(t_1, t_2, \ldots, t_m)$, both crude moments and product moments are required to be independent of time under the null hypothesis. As introduced in Section 1, there has been a vast literature on weak stationarity tests, including the DF (Dickey and Fuller, 1979), ADF (Dickey and Fuller, 1981), PP (Phillips and Perron, 1988), KPSS (Kwiatkowski et al., 1992), and LMC (Leybourne and McCabe, 1994) tests, as well as those in Xiao (2001), Hobijn et al. (2004), and Xiao and Lima (2007). These tests are typically based on a predefined linear time series model, such as an AR(1) process. Our pth-order stationarity test, on the other hand, does not rely on a correctly specified time series model, and has power against all types of violation of weak stationarity, including abrupt shifts or smooth changes in mean, variance, autocovariance, and higher-order moments. Furthermore, our test can detect a class of local alternatives that converge to the null of pth-order stationarity at the parametric rate $T^{-1/2}$, which is faster than the derivative tests of Hong et al. (2017). In Section S3.3 of the Supplementary Material, we provide simulation evidence to show the finite-sample performance of the proposed pth-order stationarity test.

5. MONTE CARLO SIMULATIONS

In this section, we present a simulation study to assess the finite-sample properties of our test. Furthermore, we compare the proposed approach with existing strict stationarity tests, including Inoue (2001), Kapetanios (2009), and Hong et al. (2017). Further simulation results, including an examination of the impact of b_T on our DWB-based test, a comparison with the model-based test of Francq

https://doi.org/10.1017/S0266466622000494 Published online by Cambridge University Press

and Zakoïan (2012) for GARCH(1,1) processes, and tests for the second-order stationarity, are reported in Section S3 of the Supplementary Material.

5.1. Data Generating Processes

To examine the size performance, we consider the following DGPs.

$$
\begin{aligned}
\text{DGP.S1}: \ & Y_t = 0.5Y_{t-1} + \varepsilon_t; \\
\text{DGP.S2}: \ & Y_t = \beta_t Y_{t-1} + \varepsilon_t, \ \beta_t = 0.5\beta_{t-1} + \eta_t; \\
\text{DGP.S3}: \ & Y_{t,1} = 0.4Y_{t-1,1} + \varepsilon_t, \\
& Y_{t,2} = 1 + 0.5Y_{t,1} + \epsilon_t, \ \epsilon_t = \sqrt{h_t}\iota_t, \ h_t = 0.2 + 0.5\epsilon_{t-1}^2; \\
\text{DGP.S4}: \ & Y_{t,1} = 1 + 0.3Y_{t-1,1} + \varepsilon_t + 0.2\varepsilon_{t-1}, \\
& Y_{t,2} = \sqrt{h_t}\iota_t, \ h_t = 0.3 + 0.2Y_{t-1,2}^2, \\
& Y_{t,3} = 0.8 + 0.1Y_{t,1} + Y_{t,2} + \eta_t; \\
\text{DGP.S5}: \ & \begin{pmatrix} Y_{t,1} \\ Y_{t,2} \\ Y_{t,3} \\ Y_{t,4} \end{pmatrix} = \begin{pmatrix} 0.1 \\ 0.3 \\ 0.1 \\ 0.7 \end{pmatrix} + \begin{pmatrix} 0.3 & 0.1 & 0 & 0 \\ 0 & 0.2 & 0 & 0 \\ 0 & 0.2 & 0.3 & 0 \\ 0 & 0.3 & 0 & 0.4 \end{pmatrix} \begin{pmatrix} Y_{t-1,1} \\ Y_{t-1,2} \\ Y_{t-1,3} \\ Y_{t-1,4} \end{pmatrix} + \varsigma_t;
\end{aligned}
$$

where

$$
\begin{aligned}
& \varepsilon_t \sim \text{i.i.d. } N(0,1); \\
& \eta_t \sim \text{i.i.d. } N(0,0.5^2); \\
& \iota_t \sim \text{i.i.d. } N(0,1); \ \text{and} \\
& \varsigma_t \sim \text{i.i.d. } N(0,\Sigma_\mathbf{1}), \ \Sigma_\mathbf{1} = \begin{pmatrix} 1.1 & 0.1 & 0.2 & 0.2 \\ 0.1 & 1.1 & 0.1 & 0.1 \\ 0.2 & 0.1 & 1.1 & 0.1 \\ 0.2 & 0.1 & 0.1 & 1.1 \end{pmatrix}.
\end{aligned}
$$

These DGPs cover various univariate and multivariate time series models. Among these, DGP.S1 is a stationary AR(1) process, and DGP.S2 is a random coefficient autoregressive model with the coefficient being a stationary AR(1) process. DGPs.S1 and S2 are designed to investigate how our test performs under univariate serially dependent linear time series models with a deterministic and random autoregressive coefficient, respectively. These two DGPs are considered by Hong et al. (2017). DGP.S3 is a bivariate time series following an AR(1) process and an AR(1)–ARCH(1) process, respectively. DGP.S4 is a three-dimensional time series in which the first variable follows an ARMA(1, 1) process, the second follows an ARCH(1) process, and the third is a linear regression with the first two variables as regressors. DGP.S5 is a four-dimensional VAR(1) process. DGPs.S3–S5 allow examination of how our test performs under various specifications of multivariate time series processes. They exhibit not only autocorrelation but also cross-correlation among the different dimensions. It is easy to apply results in

Davidson (1994), Doukhan (1994), and Carrasco and Chen (2002) to verify that the mixing conditions in Assumption A.1 are satisfied for all the above DGPs.

To examine the power performance of our test in finite samples, we consider the following DGPs:

$$\text{DGP.P1}: Y_t = Y_{t-1} + \varepsilon_t;$$
$$\text{DGP.P2}: Y_t = \varepsilon_t \mathbf{1}(t \le 0.5T) + \eta_t \mathbf{1}(t > 0.5T);$$
$$\text{DGP.P3}: Y_t = (1+\sqrt{2}\varepsilon_t)\mathbf{1}(t \le 0.5T) + \varepsilon_t^2 \mathbf{1}(t > 0.5T);$$
$$\text{DGP.P4}: Y_t = \sin(2\pi t/T) + \theta_t,\ \theta_t = 0.5\theta_{t-1} + \varepsilon_t;$$
$$\text{DGP.P5}: Y_{t,1} = 0.5Y_{t-1,1} + \varepsilon_t;$$
$$Y_{t,2} = \begin{cases} 1 - 0.5Y_{t,1} + \iota_t, & \text{if } t \le 0.3T; \\ 2.5 + Y_{t,1} + \iota_t, & \text{if } 0.3T < t \le 0.4T; \\ 1.5 - Y_{t,1} + \iota_t, & \text{if } 0.4T < t \le 0.5T; \\ 1 + 0.5Y_{t,1} + \iota_t, & \text{if } 0.5T < t \le 0.7T; \\ -0.6 - 0.3Y_{t,1} + \iota_t, & \text{if } t > 0.7T; \end{cases}$$
$$\text{DGP.P6}: \begin{pmatrix} Y_{t,1} \\ Y_{t,2} \\ Y_{t,3} \end{pmatrix} = \begin{pmatrix} 0.02 \\ 0.05 \\ 0 \end{pmatrix} + \begin{pmatrix} 0.3 & 0.3 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} Y_{t-1,1} \\ Y_{t-1,2} \\ Y_{t-1,3} \end{pmatrix} + \begin{pmatrix} \vartheta_{t,1} \\ \vartheta_{t,2} \\ \vartheta_{t,3} \end{pmatrix} + \begin{pmatrix} 0.5\vartheta_{t-1,1} \\ 0.2\vartheta_{t-1,2} \\ 0.1\vartheta_{t-1,3} \end{pmatrix};$$

where

$$\varepsilon_t \sim \text{i.i.d. } N(0,1);$$
$$\eta_t \sim \text{i.i.d. } N(0,2);$$
$$\iota_t \sim \text{i.i.d. } N(0,1);\ \text{and}$$
$$\vartheta_t \equiv \begin{pmatrix} \vartheta_{t,1} \\ \vartheta_{t,2} \\ \vartheta_{t,3} \end{pmatrix} \sim \text{i.i.d. } N(0, \Sigma_2),\ \text{with } \Sigma_2 = \begin{pmatrix} 0.5 & 0.1 & 0.2 \\ 0.1 & 1.1 & 0.1 \\ 0.2 & 0.1 & 1 \end{pmatrix}.$$

We consider various linear and nonlinear nonstationary time series with different dimensions. Among these, DGP.P1 is a random walk process. DGP.P2 has an abrupt structural break in variance, whereas the other moments are constant over time. Under DGP.P3, the first two moments remain constant over time, but higher-order moments undergo a single abrupt break. DGP.P4 exhibits smooth structural changes in mean characterized by a deterministic function of t/T. DGP.P5 is a bivariate time series process in which the first variable is weakly stationary, whereas the second variable has multiple structural breaks in mean. DGP.P6 is a three-dimensional VARMA$(1,1)$ process, with the last two variables being unit-root processes. Using the designs in DGPs.P5 and P6, we can assess the power performance of our test for multivariate time series. Specifically, these DGPs

https://doi.org/10.1017/S0266466622000494 Published online by Cambridge University Press

contain both univariate stationary and nonstationary processes, which can examine whether our test can capture nonstationarity coupled with stationary noise.

5.2. Finite-Sample Performance

In this subsection, we report the finite-sample performance of our test under the considered DGPs. For each DGP, we simulate 1,000 datasets using the sample sizes of $T = 100, 300$, and 500. The number of bootstrap iterations is set at $B = 500$. We consider the standard joint normal and joint Laplace weighting functions for our test. Specifically, we use (2.9) and (2.10) with $\gamma_1 = \gamma_2 = 1$ for the normal and Laplace weighting functions, respectively.

Because our test is not asymptotically pivotal, we implement the DWB given in Section 3.4 to obtain the critical values. We choose the block length b_T based on Rho and Shao's (2019) MV method: (i) choose a group of block length candidates $l_1, \ldots, l_k$; (ii) for each block length candidate l_i $(i = 1, \ldots, k)$, calculate the DWB statistic $\hat{D}_i^\star$; (iii) repeat B times to obtain a collection of $\{\hat{D}_{i,b}^\star\}_{b=1}^B$ for each l_i; (iv) let $\hat{F}_i$ be the EDF of $\{\hat{D}_{i,b}^\star\}_{b=1}^B$, i.e., $\hat{F}_i(z) = B^{-1}\sum_{b=1}^B \mathbf{1}(D_{i,b}^\star \leq z)$; for $i = 1, \ldots, k-1$, calculate the squared distance between $\hat{F}_i$ and $\hat{F}_{i+1} : H_i = \sum_{z \in \mathbb{Z}} |\hat{F}_i(z) - \hat{F}_{i+1}(z)|^2$, where $\mathbb{Z}$ is a set of *a priori* chosen grids for z; and (v) note that the optimal block length is $b_T = l_{\hat{\imath}}$, where $\hat{\imath} = \arg\min_{i=1,\ldots,k-1} H_i$. We set the candidate block lengths as $l = \lfloor \ln(\ln T) \rfloor, \lfloor \ln(\ln T) \rfloor + 1, \ldots, l_{\max}$ with the upper bound $l_{\max} = \left\lfloor 5(T/100)^{0.6} \right\rfloor$, where $\lfloor \cdot \rfloor$ denotes the integer part of $\cdot$. We set the starting value of the candidate block lengths to $\lfloor \ln(\ln T) \rfloor$ to ensure that all the considered block lengths satisfy Assumption A.5.

As mentioned above, we compare our test with existing tests for strict stationarity, including Inoue's (2001) weighted CvM test, Kapetanios' (2009) test, and Hong et al.'s (2017) test. Since both Kapetanios' (2009) and Hong et al.'s (2017) tests involve nonparametric smoothing, we follow Hong et al. (2017) and use the Epanechnikov kernel and set the bandwidth $h = (1/\sqrt{12})T^{-1/5}$. For Hong et al.'s (2017) test, we use the standard normal weighting function for integration over the nuisance parameters. In addition, the test statistic of Hong et al. (2017) involves a long-run variance estimator. We follow Hong et al. (2017) to set the lag order to $p_T = \min\left\{ \left\lfloor \left(\frac{3T}{2}\right)^{1/3} \left(\frac{2\hat{\rho}}{1-\hat{\rho}^2}\right)^{2/3} \right\rfloor, \left\lfloor 8\left(\frac{T}{100}\right)^{1/3} \right\rfloor \right\}$ with $\hat{\rho}$ being the estimator of first-order autocorrelation of Y_t. Other settings for Kapetanios' (2009) test are the same as those in Kapetanios (2009). In addition, all three tests require a moving block bootstrap to obtain the critical values. We adopt Politis and White's (2004) automatic block-length selection procedure for determining the block length.

Table 1 reports the empirical rejection rates at the 5% and 10% significance levels under DGPs.S1–S5. Overall, the empirical rejection rates of our tests using both the normal and Laplace weighting functions tend to converge to the corresponding nominal levels as the sample size increases. Although there exists reasonably acceptable under-rejection for $\hat{D}^L$ at the 5% level, it improves as the sample size increases. Hong et al.'s (2017) nonparametric test statistic $\hat{H}$ performs

TABLE 1. Size of strict stationarity tests under DGPs.S1–S5

		$\hat{D}^N$		$\hat{D}^L$		$\hat{H}$		$\hat{I}$		$\hat{K}$	
		5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
DGP.S1	$T=100$	0.047	0.135	0.033	0.112	0.028	0.082	0.080	0.165	0.064	0.133
	$T=300$	0.053	0.119	0.045	0.107	0.044	0.106	0.071	0.138	0.052	0.106
	$T=500$	0.050	0.100	0.040	0.091	0.040	0.092	0.058	0.128	0.050	0.098
DGP.S2	$T=100$	0.032	0.125	0.021	0.099	0.192	0.270	0.085	0.175	0.049	0.089
	$T=300$	0.029	0.094	0.023	0.078	0.220	0.288	0.062	0.132	0.042	0.086
	$T=500$	0.038	0.106	0.023	0.088	0.199	0.266	0.067	0.147	0.038	0.071
DGP.S3	$T=100$	0.042	0.127	0.025	0.103	0.070	0.130	0.072	0.153	0.032	0.079
	$T=300$	0.036	0.106	0.027	0.093	0.048	0.108	0.066	0.118	0.037	0.104
	$T=500$	0.047	0.096	0.035	0.088	0.050	0.098	0.058	0.122	0.044	0.103
DGP.S4	$T=100$	0.039	0.131	0.028	0.110	0.062	0.152	0.068	0.171	0.040	0.102
	$T=300$	0.047	0.136	0.043	0.104	0.058	0.096	0.059	0.117	0.048	0.102
	$T=500$	0.044	0.097	0.036	0.085	0.045	0.091	0.055	0.110	0.042	0.100
DGP.S5	$T=100$	0.026	0.122	0.013	0.083	0.044	0.092	0.059	0.153	0.003	0.010
	$T=300$	0.043	0.111	0.026	0.097	0.038	0.112	0.052	0.101	0.007	0.019
	$T=500$	0.042	0.114	0.025	0.098	0.049	0.107	0.059	0.116	0.014	0.029

Notes: (i) $\hat{D}^N$ and $\hat{D}^L$ denote DFT tests with normal weighting and Laplace weighting function, respectively; (ii) $\hat{H}$, $\hat{I}$, and $\hat{K}$ are the strict stationarity tests of Hong et al. (2017), Inoue (2001), and Kapetanios (2009), respectively; and (iii) for each test, the number of repetitions is 1,000 and the number of bootstrap samples is 500.

reasonably well under most DGPs except for a certain size distortion under DGP.S2, which is a random coefficient model. Regarding Inoue's (2001) CvM test statistic $\hat{I}$, there is an obvious oversize issue for most DGPs, particularly when the sample size T is small, implying that $\hat{I}$ tends to over-reject a stationary null hypothesis. Kapetanios' (2009) nonparametric test statistic $\hat{K}$ performs reasonably well in most cases except DGP.S5, in which the dimension of the data is relatively high, and a serious under-size distortion occurs.

Table 2 demonstrates the power performance of the tests under DGPs.P1–P6 at the 5% and 10% significance levels with sample sizes of $T = 100$, 300, and 500. Overall, our tests deliver a robust power performance against all proposed DGPs. As the sample size increases, the rejection rates of our tests under both normal and Laplace weighting functions converge to 1. Note that $\hat{D}^N$ and $\hat{D}^L$ yield similar simulation results for size and power experiments, indicating that the choice of weighting function does not substantially impact the finite-sample performance of our tests. Intuitively, as long as nonzero weights are assigned to the frequencies (u,v) that exhibit structural breaks, our test has power. Hong et al.'s (2017) nonparametric test statistic $\hat{H}$ is also quite powerful against all nonstationary DGPs

Table 2. Power of strict stationarity tests under DGPs.P1–P6

		$\hat{D}^N$		$\hat{D}^L$		$\hat{H}$		$\hat{I}$		$\hat{K}$	
		5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
DGP.P1	$T = 100$	0.851	0.943	0.828	0.933	0.854	0.910	0.442	0.682	0.288	0.465
	$T = 300$	0.949	0.973	0.960	0.983	0.992	0.994	0.602	0.744	0.389	0.567
	$T = 500$	0.957	0.980	0.962	0.986	0.998	0.999	0.666	0.777	0.484	0.632
DGP.P2	$T = 100$	0.654	0.856	0.487	0.754	0.562	0.748	0.319	0.548	0.070	0.158
	$T = 300$	1.000	1.000	0.999	1.000	0.996	0.998	0.967	0.995	0.208	0.349
	$T = 500$	1.000	1.000	1.000	1.000	0.999	1.000	1.000	1.000	0.329	0.460
DGP.P3	$T = 100$	0.395	0.614	0.491	0.720	0.280	0.400	0.427	0.620	0.153	0.274
	$T = 300$	0.974	0.996	0.999	1.000	0.708	0.816	0.979	0.998	0.332	0.494
	$T = 500$	1.000	1.000	1.000	1.000	0.899	0.940	1.000	1.000	0.561	0.716
DGP.P4	$T = 100$	0.801	0.920	0.747	0.904	0.324	0.580	0.241	0.689	0.029	0.100
	$T = 300$	0.995	0.999	0.995	0.999	0.984	0.998	0.462	0.790	0.108	0.194
	$T = 500$	1.000	1.000	1.000	1.000	1.000	1.000	0.969	1.000	0.265	0.454
DGP.P5	$T = 100$	0.936	0.994	0.896	0.994	0.748	0.906	0.466	0.865	0.074	0.151
	$T = 300$	1.000	1.000	1.000	1.000	1.000	1.000	0.977	1.000	0.073	0.152
	$T = 500$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.069	0.144
DGP.P6	$T = 100$	0.976	0.993	0.978	0.998	0.990	0.996	0.398	0.699	0.059	0.123
	$T = 300$	0.995	1.000	0.999	1.000	0.997	0.999	0.648	0.850	0.105	0.185
	$T = 500$	1.000	1.000	1.000	1.000	1.000	1.000	0.758	0.908	0.124	0.223

Notes: (i) $\hat{D}^N$ and $\hat{D}^L$ denote DFT tests with normal weighting and Laplace weighting function, respectively; (ii) $\hat{H}$, $\hat{I}$, and $\hat{K}$ are the strict stationarity tests of Hong et al. (2017), Inoue (2001), and Kapetanios (2009), respectively; and (iii) for each test, the number of repetitions is 1,000 and the number of bootstrap samples is 500.

we investigate. Under DGPs.P3–P5, our tests outperform $\hat{H}$, particularly when T is small. Intuitively, this is because our tests can detect a class of local alternatives that converge to the null at a faster rate. The power performance of Inoue's (2001) test statistic $\hat{I}$ is not as comparable to our tests and $\hat{H}$, especially when the sample size is small. Finally, Kapetanios' (2009) nonparametric test statistic $\hat{K}$ lacks any reasonable power for DGP.P6, which considers a multivariate time series process, and it has relatively low power against smooth structural changes and multiple structural breaks described by DGPs.P4 and P5.

6. AN EMPIRICAL APPLICATION TO THE EXCHANGE RATE RETURNS

The foreign exchange market is among the most important financial markets and has drawn substantial attention over the last four decades. Modeling and forecasting exchange rates are vital for decision-makers at both micro and macro levels. Obtaining reasonable estimators and accurate forecasts requires first examining the stationarity of exchange rate returns. Most empirical studies conclude that the floating exchange rates follow unit-root processes, and their returns are assumed to be stationary (see, e.g., Meese and Rogoff, 1988; Lothian and Taylor, 1996; Wu, 1996; Hegwood and Papell, 1998; Sollis, Leybourne, and Newbold, 2002). Among the many approaches proposed in the literature, the most popular ones are the ADF (Dickey and Fuller, 1981) and KPSS (Kwiatkowski et al., 1992) tests. However, the ADF and KPSS tests are explicitly designed for the unit-root process, which constitutes only one case of nonstationarity. Consequently, these tests may miss other types of nonstationarity.

In this section, we apply our test to investigate the stationarity of nominal and real exchange rate returns. We check the stationarity of four exchange rates: Great Britain Pound (*GBP*), Canadian Dollar (*CAD*), Japanese Yen (*JPY*), and Euro (*EUR*). The data are measured using the rates of *GBP*, *CAD*, *JPY*, and *EUR* to one U.S. Dollar, respectively. We use monthly data from January 1971 to April 2021 with 603 observations for *GBP*, *CAD*, and *JPY*. Since the Euro was introduced in January 1999, we have only 266 observations for *EUR*. The monthly nominal exchange rate returns are measured using the log difference of the end-of-period exchange rates. The real exchange rate is constructed as $r_{i,t} = e_{i,t} + \tilde{p}_t - p_{i,t}, i = 1,\ldots,4,$ with $e_{i,t}, \tilde{p}_t$, and $p_{i,t}$ being the nominal exchange rate returns, foreign price level, and domestic price level, respectively. Following Mark (1990), Wu (1996), Papell (1997), and other typical practices, we use the Consumer Price Indices to measure the price levels. All data are collected from the website of the U.S. Federal Reserve Bank of St. Louis.

We examine the stationarity of exchange rate returns using our test and several others. Specifically, we consider Dickey and Fuller's (1981) ADF test for the null hypothesis of the unit-root process, Kwiatkowski et al.'s (1992) KPSS test for the null hypothesis of trending stationarity, Inoue's (2001) weighted CvM statistic $\hat{I}$, Kapetanios' (2009) test $\hat{K}$, and Hong et al.'s (2017) test $\hat{H}$ for strict stationarity. As in the simulation studies, we consider the normal and Laplace weighting functions for our test. The tests of Kapetanios (2009) and Hong et al. (2017) are based on smoothed nonparametric regression, which involves the kernel function and bandwidth. We choose the Epanechnikov kernel function and the bandwidth $h = ch_0$ with $h_0 = (2.35/\sqrt{12})T^{-1/5}$ being the Silverman's rule-of-thumb bandwidth. By choosing the tuning parameter $c = 0.5, 1, 2$, and 3, we examine the effect of different bandwidths on Kapetanios' (2009) and Hong et al.'s (2017) tests. Other settings, including the block length and the lag order in calculating the long-run variance, are the same as those used in the simulation studies. For tests involving the bootstrap procedure, we set the bootstrap replication number $B = 1{,}000$. We also check the stationarity of these series using the ADF and the KPSS tests. The p-values of the ADF test are all below 0.001, and the p-values of the KPSS test are all above 0.100 for all the exchange rate returns investigated. If we use either the ADF or the KPSS test to check the stationarity of the underlying time series,

https://doi.org/10.1017/S0266466622000494 Published online by Cambridge University Press

TABLE 3. Stationarity tests for exchange rate returns

	$\hat{D}^N$	$\hat{D}^L$	$\hat{H}_{0.5h_0}$	$\hat{H}_{h_0}$	$\hat{H}_{2h_0}$	$\hat{H}_{3h_0}$	$\hat{I}$	$\hat{K}_{0.5h_0}$	$\hat{K}_{h_0}$	$\hat{K}_{2h_0}$	$\hat{K}_{3h_0}$
Nominal-univariate											
GBP	0.038	0.039	0.004	0.020	0.096	0.097	0.248	0.519	0.415	0.228	0.137
CAD	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.166	0.004	0.000	0.000
JAY	0.076	0.061	0.006	0.027	0.078	0.131	0.081	0.761	0.024	0.033	0.105
EUR	0.083	0.088	0.018	0.035	0.105	0.123	0.197	0.393	0.330	0.342	0.206
Real-univariate											
GBP	0.024	0.025	0.020	0.010	0.053	0.097	0.033	0.332	0.284	0.061	0.082
CAD	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.006	0.007	0.001	0.000
JAY	0.074	0.058	0.003	0.012	0.024	0.068	0.067	0.596	0.651	0.274	0.093
EUR	0.082	0.073	0.022	0.032	0.074	0.111	0.218	0.004	0.087	0.035	0.049
Nominal-bivariate											
(*GBP*, *CAD*)	0.007	0.004	0.000	0.000	0.000	0.000	0.003	0.726	0.608	0.163	0.211
(*GBP*, *JAY*)	0.017	0.011	0.000	0.000	0.004	0.012	0.048	0.237	0.090	0.054	0.024
(*GBP*, *EUR*)	0.258	0.191	0.053	0.109	0.299	0.400	0.323	0.065	0.191	0.141	0.142
(*CAD*, *JAY*)	0.003	0.003	0.018	0.000	0.000	0.000	0.000	0.230	0.517	0.205	0.059
(*CAD*, *EUR*)	0.044	0.035	0.007	0.026	0.122	0.181	0.496	0.614	0.819	0.841	0.979
(*JAY*, *EUR*)	0.084	0.077	0.001	0.016	0.070	0.101	0.208	0.910	0.196	0.452	0.056
Real-bivariate											
(*GBP*, *CAD*)	0.004	0.003	0.000	0.000	0.000	0.000	0.004	0.312	0.411	0.064	0.047
(*GBP*, *JAY*)	0.028	0.033	0.000	0.000	0.001	0.005	0.013	0.162	0.136	0.058	0.080
(*GBP*, *EUR*)	0.341	0.256	0.067	0.091	0.219	0.303	0.390	0.529	0.405	0.011	0.009
(*CAD*, *JAY*)	0.010	0.003	0.000	0.000	0.000	0.000	0.000	0.180	0.389	0.104	0.285
(*CAD*, *EUR*)	0.098	0.084	0.009	0.022	0.085	0.146	0.470	0.253	0.296	0.879	0.895
(*JAY*, *EUR*)	0.212	0.113	0.003	0.015	0.050	0.068	0.195	0.235	0.108	0.108	0.040

Notes: (i) Numbers in main entries are p-values; (ii) $\hat{D}^N$ and $\hat{D}^L$ denote DFT tests with normal and Laplace weighting function, respectively; (iii) $\hat{H}_{ch_0}$ denotes Hong et al.'s (2017) test with the bandwidth $h = ch_0$ with $c = 0.5, 1, 2,$ and 3, respectively; (iv) $\hat{I}$ denotes Inoue's (2001) weighted CvM statistic; (v) $\hat{K}_{ch_0}$ denotes Kapetanios' (2009) test using the bandwidth $h = ch_0$ with $c = 0.5, 1, 2,$ and 3, respectively; and (vi) for the tests involving bootstrap, the number of bootstrap samples is 1,000.

we regard these series as stationary processes. To save space, we do not report the results of the ADF and KPSS tests.

Table 3 reports the p-values of various tests for the nominal and real exchange rate returns. We consider both the univariate and the bivariate cases that combine two exchange rate returns as a random vector. The top two panels of Table 3 report the results for the univariate case. We note that our test, with normal and Laplace weighting functions, rejects the null hypothesis of strict stationarity for *GBP* and *CAD* at the 5% significance level for both the nominal and real exchange rate returns, but rejects the null hypothesis for *JAY* and *EUR* at the 10% significance level for both the nominal and real exchange rate returns. However, Kapetanios' (2009) and Hong et al.'s (2017) tests are sensitive to the choice of bandwidth. For example, Hong et al.'s (2017) test cannot reject the null hypothesis for the nominal exchange rate returns of *JAY* and *EUR* when we use the bandwidth $h = ch_0$ with $c = 3$ at the 10% significance level; however, it can reject the null hypothesis at the 5% level if we choose $c = 0.5$ or 1. Therefore, different choices of bandwidth yield mixed results for the tests of Kapetanios (2009) and Hong et al. (2017). Inoue's (2001) results fail to reject the null hypothesis at the 10% significance level for nominal returns of *GBP* and *EUR* and real returns of *EUR*. This result is consistent with our theoretical conclusion that our test is more powerful than the existing tests, including that of Inoue (2001).

The bottom two panels of Table 3 report the results for the multivariate case. Our DFT tests with both normal and Laplace weighting functions reject the null hypothesis of strict stationarity at the 10% significance level for all combinations except those of nominal exchange rate returns of (GBP, EUR) and real exchange rate returns of (GBP, EUR) and (JAY, EUR). However, Hong et al.'s (2017) results are mixed for the combinations of nominal exchange rate returns of (GBP, EUR) and (JAY, EUR) and real exchange rate returns of (GBP, EUR) and (CAD, EUR). Inoue's (2001) test fails to reject the null hypothesis at the 10% significance level for the nominal and real exchange rate returns of (GBP, EUR), (CAD, EUR), and (JAY, EUR). In addition, we note that Kapetanios' (2009) test is extremely sensitive to the choice of bandwidth, and most of the results cannot reject the null hypothesis. The low power of Kapetanios' (2009) test may result from the slow convergence rate $T^{-1/2}h^{-d/2}$, which is severely affected by the dimension of the underlying time series process and, hence, suffers from the curse of dimensionality problem.

7. CONCLUSION

Strict stationarity is a fundamental modeling assumption for time series analysis. This paper proposes a model-free test for strict stationarity based on a DFT approach with the basic idea of estimating a CF using the ECF. If the underlying time series is nonstationary, the estimated residuals will contain such time-varying information. We construct the DFT of the estimated generalized residuals and infer the existence of nonstationarity by examining the corresponding spectrum at each frequency. The test is powerful against a class of local alternatives. Specifically, the DFT of the generalized residuals converges to a Gaussian process under the null hypothesis, and the test statistic can detect local alternatives converging to the null at the parametric rate $T^{-1/2}$. In addition, using an appropriate choice of the weighting function can avoid high-dimensional numerical integration when computing the test statistic. Compared with strict stationarity tests based on smoothed nonparametric estimation, our test avoids the choice of tuning parameters and has a faster convergence rate. Furthermore, we examine the finite sample property

https://doi.org/10.1017/S0266466622000494 Published online by Cambridge University Press

of the test via Monte Carlo simulations. It shows that our test has a reasonable size performance and nontrivial power against various forms of nonstationarity. Moreover, employing our test on nominal and real exchange rate returns and their bivariate combinations reveals that the exchange rate returns are mostly nonstationary, in contrast to the conclusion of weak stationarity tests, e.g., the ADF and KPSS tests. Our test results are robust to reject the strict stationarity null hypothesis under various weighting functions. In contrast, the nonparametric strict stationarity tests are susceptible to the choice of bandwidth and usually fail to reject the null hypothesis of strict stationarity.

MATHEMATICAL APPENDIX

A.1. Proof of Proposition 3.1

Under $\mathbb{H}_0 : \phi_t(u) = \phi_0(u)$ for all $u \in \mathbb{R}^{dm}$ and all t, $\hat{A}_1(u,v) = 0$ and $\sqrt{T}\hat{A}(u,v) = \sqrt{T}\hat{A}_2(u,v)$. We only need to consider the asymptotic distribution of $\hat{S}(u,v) \equiv \sqrt{T}\hat{A}_2(u,v)$. First, we calculate the first two moments of $\hat{S}(u,v)$. For each fixed $(u',v)' \in \mathbb{R}^{dm+1}$, we have

$$E[\hat{S}(u,v)] = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\bar{M}_t(v)E[\varepsilon_t(u)] = 0.$$

For the second moment, we make the following decomposition:

$$\begin{aligned}
E[\hat{S}(u,v)\hat{S}(u,v)^*] &= \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}E[\varepsilon_t(u)\varepsilon_s(u)^*]\bar{M}_t(v)\bar{M}_s(v)^* \\
&= \frac{1}{T}\sum_{t=1}^{T}E[\varepsilon_t(u)\varepsilon_t(u)^*]|\bar{M}_t(v)|^2 \\
&\quad + \frac{1}{T}\sum_{t=2}^{T}\sum_{s=1}^{t-1}E[\varepsilon_t(u)\varepsilon_s(u)^*]\bar{M}_t(v)\bar{M}_s(v)^* \\
&\quad + \frac{1}{T}\sum_{t=1}^{T-1}\sum_{s=t+1}^{T}E[\varepsilon_t(u)\varepsilon_s(u)^*]\bar{M}_t(v)\bar{M}_s(v)^* \\
&\equiv V_1(u,v) + V_2(u,v) + V_3(u,v), \text{ say.}
\end{aligned}$$

Consider $V_1(u,v)$ first. Noting that $\varepsilon_t(u) = e^{\mathbf{i}u'Y_t} - \phi_t(u)$ and $\phi_t(u) = \phi_0(u)$ under $\mathbb{H}_0$, we can readily show that

$$E[\varepsilon_t(u)\varepsilon_t(u)^*] = E[1 - \phi_0(u)e^{-\mathbf{i}u'Y_t} - e^{\mathbf{i}u'Y_t}\phi_0(u)^* + |\phi_0(u)|^2] = 1 - |\phi_0(u)|^2.$$

Then,

$$
\begin{aligned}
V_1(u,v) &= E[\varepsilon_t(u)\varepsilon_t(u)^*]\frac{1}{T}\sum_{t=1}^{T}\left(1-2e^{\mathbf{i}v2\pi t/T}\frac{1}{T}\sum_{t=1}^{T}e^{-\mathbf{i}v2\pi t/T}+\left|\frac{1}{T}\sum_{t=1}^{T}e^{\mathbf{i}v2\pi t/T}\right|^2\right)\\
&= E[\varepsilon_t(u)\varepsilon_t(u)^*]\left(1-\left|\frac{1}{T}\sum_{t=1}^{T}e^{\mathbf{i}v2\pi t/T}\right|^2\right)\\
&\to E[\varepsilon_t(u)\varepsilon_t(u)^*]\left(1-\left|\int_0^1 e^{\mathbf{i}v2\pi\tau}\mathrm{d}\tau\right|^2\right).
\end{aligned}
$$

For $V_2(u,v)$, we have that under $\mathbb{H}_0$,

$$
\begin{aligned}
V_2(u,v) &= \sum_{j=1}^{T-1}\frac{1}{T}\sum_{t=j+1}^{T}E[\varepsilon_t(u)\varepsilon_{t-j}(u)^*]\bar{M}_t(v)\bar{M}_{t-j}(v)^*\\
&= \sum_{j=1}^{T-1}E[\varepsilon_t(u)\varepsilon_{t-j}(u)^*]\frac{1}{T}\sum_{t=j+1}^{T}\left(e^{\mathbf{i}v2\pi j/T}-e^{\mathbf{i}v2\pi t/T}\frac{1}{T}\sum_{t=1}^{T}e^{-\mathbf{i}v2\pi t/T}\right.\\
&\quad\left.-\,e^{-\mathbf{i}v2\pi(t-j)/T}\frac{1}{T}\sum_{t=1}^{T}e^{\mathbf{i}v2\pi t/T}+\left|\frac{1}{T}\sum_{t=1}^{T}e^{\mathbf{i}v2\pi t/T}\right|^2\right)\\
&= \sum_{j=1}^{T-1}E[\varepsilon_t(u)\varepsilon_{t-j}(u)^*]\left(\frac{T-j}{T}e^{\mathbf{i}v2\pi j/T}-\frac{1}{T}\sum_{s=1}^{T-j}e^{\mathbf{i}v2\pi(s+j)/T}\frac{1}{T}\sum_{t=1}^{T}e^{-\mathbf{i}v2\pi t/T}\right.\\
&\quad\left.-\frac{1}{T}\sum_{s=1}^{T-j}e^{-\mathbf{i}v2\pi s/T}\frac{1}{T}\sum_{t=1}^{T}e^{\mathbf{i}v2\pi t/T}+\frac{T-j}{T}\left|\frac{1}{T}\sum_{t=1}^{T}e^{\mathbf{i}v2\pi t/T}\right|^2\right)\\
&= \sum_{j=1}^{T-1}E[\varepsilon_t(u)\varepsilon_{t-j}(u)^*]\frac{T-j}{T}\left(e^{\mathbf{i}v2\pi j/T}-e^{\mathbf{i}v2\pi j/T}\frac{1}{T-j}\sum_{s=1}^{T-j}e^{\mathbf{i}v2\pi s/T}\frac{1}{T}\sum_{t=1}^{T}e^{-\mathbf{i}v2\pi t/T}\right.\\
&\quad\left.-\frac{1}{T-j}\sum_{s=1}^{T-j}e^{-\mathbf{i}v2\pi s/T}\frac{1}{T}\sum_{t=1}^{T}e^{\mathbf{i}v2\pi t/T}+\left|\frac{1}{T}\sum_{t=1}^{T}e^{\mathbf{i}v2\pi t/T}\right|^2\right)\\
&= \sum_{j=1}^{T-1}E[\varepsilon_t(u)\varepsilon_{t-j}(u)^*]\left(1-\frac{j}{T}\right)e^{\mathbf{i}v2\pi j/T}\left(1-\left|\int_0^1 e^{\mathbf{i}v2\pi\tau}\mathrm{d}\tau\right|^2\right)+o(1)\\
&\to \sum_{j=1}^{\infty}E[\varepsilon_t(u)\varepsilon_{t-j}(u)^*]\left(1-\left|\int_0^1 e^{\mathbf{i}v2\pi\tau}\mathrm{d}\tau\right|^2\right)
\end{aligned}
$$

by the dominated convergence theorem (DCT). Analogously,

$$V_3(u,v) = \sum_{j=-(T-1)}^{-1} E[\varepsilon_t(u)\varepsilon_{t-j}(u)^*]\left(1-\frac{j}{T}\right)e^{\mathbf{i}v2\pi j/T}\left(1-\left|\int_0^1 e^{\mathbf{i}v2\pi\tau}\mathrm{d}\tau\right|^2\right)+o(1)$$
$$\to \sum_{j=-\infty}^{-1} E[\varepsilon_t(u)\varepsilon_{t-j}(u)^*]\left(1-\left|\int_0^1 e^{\mathbf{i}v2\pi\tau}\mathrm{d}\tau\right|^2\right).$$

Therefore, it follows that

$$\begin{aligned}E[\hat{S}(u,v)\hat{S}(u,v)^*] &= \sum_{j=-(T-1)}^{T-1} E[\varepsilon_t(u)\varepsilon_{t-j}(u)^*]\left(1-\left|\int_0^1 e^{\mathbf{i}v2\pi\tau}\mathrm{d}\tau\right|^2\right)+o(1)\\ &\to \sum_{j=-\infty}^{\infty}\operatorname{cov}(e^{\mathbf{i}u'Y_t},e^{-\mathbf{i}u'Y_{t-j}})\left(1-\left|\int_0^1 e^{\mathbf{i}v2\pi\tau}\mathrm{d}\tau\right|^2\right)\\ &= \Gamma_1(u,u)\Gamma_2(v,v),\end{aligned}$$

where $\Gamma_1(u,u) \equiv \sum_{j=-\infty}^{\infty}\operatorname{cov}(e^{\mathbf{i}u'Y_t},e^{-\mathbf{i}u'Y_{t-j}})$ is a generalized long-run variance of $e^{\mathbf{i}u'Y_t}$ and $\Gamma_2(v,v) \equiv \operatorname{cov}\left(e^{\mathbf{i}v2\pi\tau},e^{-\mathbf{i}v2\pi\tau}\right) = 1-\left|\int_0^1 e^{\mathbf{i}v2\pi\tau}\mathrm{d}\tau\right|^2$ is the pseudo-covariance of the Fourier basis function of time in the sense that τ follows $U[0,1]$ distribution.

By the triangle inequality and the fact that $|e^{\mathbf{i}u'Y_t}| = |e^{\mathbf{i}v2\pi t/T}| = 1$ for all t,

$$\begin{aligned}\max_t|\varepsilon_t(u)\bar{M}_t(v)| &\le \max_t\left[\left|e^{\mathbf{i}u'Y_t}\right|+|\phi_t(u)|\right]\left[\max_t\left|e^{\mathbf{i}v2\pi t/T}\right|+\frac{1}{T}\sum_{t=1}^{T}\left|e^{\mathbf{i}v2\pi t/T}\right|\right]\\ &\le 2\times 2 = 4.\end{aligned}$$

Then we can apply Theorem 5.20 of White (2001) with $r=\infty$ to obtain that for each fixed $(u,v)\in\mathbb{W}\equiv\mathbb{U}\times\mathbb{V}$,

$$\hat{S}(u,v)\xrightarrow{d}N\left[0,\Gamma_1(u,u)\Gamma_2(v,v)\right].$$

Now, we show that $\hat{S}(u,v)$ is asymptotically tight on $\mathbb{U}\times\mathbb{V}$. For any $(u_1,v_1),(u_2,v_2)\in\mathbb{W}$, we apply the mean value theorem to obtain

$$\sum_{t=1}^{T}[\varepsilon_t(u_1)-\varepsilon_t(u_2)]\bar{M}_t(v_1) = \sum_{t=1}^{T}\Upsilon_t(\bar{u})'(u_1-u_2)\bar{M}_t(v_1) \text{ and}$$
$$\sum_{t=1}^{T}\left[\bar{M}_t(v_1)-\bar{M}_t(v_2)\right]\varepsilon_t(u_2) = \sum_{t=1}^{T}\Psi_t(\bar{v})(v_1-v_2)\varepsilon_t(u_2),$$

https://doi.org/10.1017/S0266466622000494 Published online by Cambridge University Press

where $\bar{u}$ lies between u_1 and u_2, $\bar{v}$ lies between v_1 and v_2,

$$\Upsilon_t(u) = \frac{\mathrm{d}\varepsilon_t(u)}{\mathrm{d}u} = \mathbf{i}\left[Y_t e^{\mathbf{i}u'Y_t} - E(Y_t e^{\mathbf{i}u'Y_t})\right], \text{ and}$$

$$\Psi_t(v) = \frac{\mathrm{d}\bar{M}_t(v)}{\mathrm{d}v} = \mathbf{i}2\pi\left[\frac{t}{T}e^{\mathbf{i}v2\pi t/T} - \frac{1}{T}\sum_{t=1}^{T}\frac{t}{T}e^{\mathbf{i}v2\pi t/T}\right].$$

Then, it follows

$$\begin{aligned}
& E\left[\left|\hat{S}(u_1,v_1) - \hat{S}(u_2,v_2)\right|^2\right] \\
&= E\left[\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}[\varepsilon_t(u_1)-\varepsilon_t(u_2)]\bar{M}_t(v_1) + \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_t(u_2)\left[\bar{M}_t(v_1)-\bar{M}_t(v_2)\right]\right|^2\right] \\
&\leq 2E\left[\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}[\varepsilon_t(u_1)-\varepsilon_t(u_2)]\bar{M}_t(v_1)\right|^2\right] + 2E\left[\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_t(u_2)\left[\bar{M}_t(v_1)-\bar{M}_t(v_2)\right]\right|^2\right] \\
&\leq 2E\left[\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Upsilon_t(\bar{u})\bar{M}_t(v_1)\right\|^2\right]\|u_1-u_2\|^2 + 2E\left[\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_t(u_2)\Psi_t(\bar{v})\right|^2\right](v_1-v_2)^2 \\
&\leq \frac{2}{T}\sum_{s,t=1}^{T}\left\|\mathrm{cov}\left[Y_t e^{\mathbf{i}\bar{u}'Y_t}, Y_s e^{-\mathbf{i}\bar{u}'Y_s}\right]\bar{M}_t(v_1)\bar{M}_s(v_1)^*\right\|\|u_1-u_2\|^2 \\
&\quad + \frac{2}{T}\sum_{s,t=1}^{T}\left|\mathrm{cov}\left[e^{\mathbf{i}u'Y_t}, e^{-\mathbf{i}u'Y_s}\right]\Psi_t(\bar{v})\Psi_s(\bar{v})^*\right|(v_1-v_2)^2 \\
&\leq \frac{8}{T}\sum_{s,t=1}^{T}\left\|\mathrm{cov}\left[Y_t e^{\mathbf{i}\bar{u}'Y_t}, Y_s e^{-\mathbf{i}\bar{u}'Y_s}\right]\right\|||u_1-u_2||^2 + \frac{32\pi^2}{T}\sum_{s,t=1}^{T}\left|\mathrm{cov}\left[e^{\mathbf{i}u'Y_t}, e^{-\mathbf{i}u'Y_s}\right]\right|(v_1-v_2)^2 \\
&\leq C(\|u_1-u_2\|^2 + |v_1-v_2|^2),
\end{aligned}$$

by the mixing condition in Assumption A.1 and the fact that $\max_t \sup_v |\bar{M}_t(v)| \leq 2$ and $\max_t \sup_v |\Psi_t(v)| \leq 4\pi$. Thus, $\hat{S}(u,v)$ is asymptotically tight on $\mathbb{W}$, and by Theorem 13.5 of Billingsley (1999), it follows

$\hat{S}(u,v) \Rightarrow S(u,v)$ on $\mathbb{W}$,

where $S(u,v)$ is a complex-valued Gaussian process with covariance kernel $\mathcal{K}_0(w_1,w_2) \equiv \Gamma_1(u_1,u_2)\Gamma_2(v_1,v_2)$. ∎

A.2. Proof of Theorem 3.1

Recall that $\mathbb{U} = [-b,b]^{dm}$ and $\mathbb{V} = [-c,c]$, where b and c are positive constants. Recall that $\mathbb{W} = \mathbb{U}\times\mathbb{V}$. By the continuous mapping theorem (CMT), for any fixed $\mathbb{W}$,

$$\hat{D}_{\mathbb{W}} \equiv \int_{\mathbb{W}}\left|\hat{S}(u,v)\right|^2 W(u,v)\,\mathrm{d}u\,\mathrm{d}v \xrightarrow{d} \int_{\mathbb{W}}|S(u,v)|^2 W(u,v)\,\mathrm{d}u\,\mathrm{d}v. \quad \textbf{(A.2.1)}$$

It remains to show that

$$\hat{D} \xrightarrow{d} \int_{\mathbb{R}^{dm+1}} |S(u,v)|^2 W(u,v)\,\mathrm{d}u\,\mathrm{d}v.$$

To proceed, we first check the uniform integrability (UI) of $\{\hat{D}_{\mathbb{W}}\}$ by verifying the sufficient condition that $E\left|\hat{D}_{\mathbb{W}}\right|^r \le C$ for some $r > 1$ and $C < \infty$. Let $c_{\mathbb{W}} = \int_{\mathbb{W}} W(u,v)\,\mathrm{d}u\,\mathrm{d}v$ and $\bar{W}(u,v) = W(u,v)/c_{\mathbb{W}}$. By the moment bounds for stationary mixing sequences (e.g., Yokoyama, 1980; Yang, 2007) and the fact that $|\varepsilon_t(u)\bar{M}_t(v)| \le 4$, we have

$$E\left[\left|\hat{S}(u,v)\right|^{2r}\right] \le CE|\varepsilon_t(u)\bar{M}_t(v)|^r \le C4^r.$$

Then, by the Jensen inequality and Fubini theorem, for any $r > 1,$ we have

$$\begin{aligned}
E\left|\hat{D}_{\mathbb{W}}\right|^r &= c_{\mathbb{W}}^r E\left[\int_{\mathbb{W}} \left|\hat{S}(u,v)\right|^2 \bar{W}(u,v)\,\mathrm{d}u\,\mathrm{d}v\right]^r \\
&\le c_{\mathbb{W}}^r E\left[\int_{\mathbb{W}} \left|\hat{S}(u,v)\right|^{2r} \bar{W}(u,v)\,\mathrm{d}u\,\mathrm{d}v\right] \\
&= c_{\mathbb{W}}^r \int_{\mathbb{W}} E\left[\left|\hat{S}(u,v)\right|^{2r}\right] \bar{W}(u,v)\,\mathrm{d}u\,\mathrm{d}v \\
&\le c_{\mathbb{W}}^r \int_{\mathbb{W}} C4^r \bar{W}(u,v)\,\mathrm{d}u\,\mathrm{d}v = C4^r c_{\mathbb{W}}^r < \infty.
\end{aligned}$$

Thus, $\{\hat{D}_{\mathbb{W}}\}$ is uniformly integrable. This, in conjunction with (A.2.1), implies that

$$E\left[\hat{D}_{\mathbb{W}}\right] \to \int_{\mathbb{W}} E|S(u,v)|^2 W(u,v)\,\mathrm{d}u\,\mathrm{d}v. \tag{A.2.2}$$

Similarly, $\left\{\left|\hat{S}(u,v)\right|^2\right\}$ also satisfies the UI condition for each fixed (u,v). This, along with the result in Proposition 3.1, implies that $E\left|\hat{S}(u,v)\right|^2 \to E|S(u,v)|^2$ for each fixed (u,v). Under Assumption A.2, for any $\epsilon > 0$, there exists a bounded subset set $\mathbb{W} = \mathbb{U} \times \mathbb{V}$ large enough such that

$$\int_{\mathbb{W}^c} E|S(u,v)|^2 W(u,v)\,\mathrm{d}u\,\mathrm{d}v < \frac{\epsilon^2}{2},$$

where $\mathbb{W}^c$ is the complementary set of $\mathbb{W}$ in $\mathbb{R}^{dm+1}$. Then, it follows that

$$\begin{aligned}
E\left[\int_{\mathbb{W}^c} \left|\hat{S}(u,v)\right|^2 W(u,v)\,\mathrm{d}u\,\mathrm{d}v\right] &= \int_{\mathbb{W}^c} E\left|\hat{S}(u,v)\right|^2 W(u,v)\,\mathrm{d}u\,\mathrm{d}v \\
&\to \int_{\mathbb{W}^c} E|S(u,v)|^2 W(u,v)\,\mathrm{d}u\,\mathrm{d}v < \frac{\epsilon^2}{2}.
\end{aligned}$$

For this $\mathbb{W}$, define

$$\begin{aligned}
\hat{D}_1 &= \int_{\mathbb{W}^c} \left|\hat{S}(u,v)\right|^2 W(u,v)\,\mathrm{d}u\,\mathrm{d}v, \\
\hat{D}_2 &= \int_{\mathbb{W}} \left|\hat{S}(u,v)\right|^2 W(u,v)\,\mathrm{d}u\,\mathrm{d}v,
\end{aligned}$$

$$D_1 = \int_{\mathbb{W}^c} |S(u,v)|^2 W(u,v)\,\mathrm{d}u\,\mathrm{d}v,$$
$$D_2 = \int_{\mathbb{W}} |S(u,v)|^2 W(u,v)\,\mathrm{d}u\,\mathrm{d}v.$$

Then, $\hat{D}_2 \xrightarrow{d} D_2, E(D_1) < \epsilon^2/2$, and there exists sufficiently large T_0 such that $E(\hat{D}_1) < \epsilon^2$ for $T \geq T_0$. We want to show that $\hat{D} = \hat{D}_1 + \hat{D}_2 \xrightarrow{d} D = D_1 + D_2$. This follows because for any $x \in \mathbb{R}$ and arbitrarily small $\epsilon > 0$, we have

$$\begin{aligned}
P(D_1 + D_2 \leq x - \epsilon) - \epsilon &\leq P(D_2 \leq x - \epsilon) - \epsilon \\
&= \liminf_{T\to\infty} P(\hat{D}_2 \leq x - \epsilon) - \epsilon \\
&= \liminf_{T\to\infty} [P(\hat{D}_1 + \hat{D}_2 \leq x) + P(\hat{D}_1 \geq \epsilon)] - \epsilon \\
&\leq \liminf_{T\to\infty} P(\hat{D}_1 + \hat{D}_2 \leq x) + \epsilon - \epsilon \\
&\leq \limsup_{T\to\infty} P(\hat{D}_1 + \hat{D}_2 \leq x) \\
&\leq \limsup_{T\to\infty} P(\hat{D}_2 \leq x) \\
&\leq P(D_2 \leq x) \\
&\leq P(D_1 + D_2 \leq x + \epsilon) + P(D_1 \geq \epsilon) \\
&\leq P(D \leq x + \epsilon) + \epsilon/2.
\end{aligned}$$

That is, $P(D \leq x - \epsilon) - \epsilon \leq \limsup_{T\to\infty} P(\hat{D} \leq x) \leq P(D \leq x + \epsilon) + \epsilon/2$ for any $\epsilon > 0$. This implies that

$$\hat{D} \xrightarrow{d} \int_{\mathbb{R}^{dm+1}} |S(u,v)|^2 W(u,v)\,\mathrm{d}u\,\mathrm{d}v. \qquad \blacksquare$$

A.3. Proof of Proposition 3.2

Noting that $\phi_t(u) = \phi(u, t/T)$ under $\mathbb{H}_A$, we have by the Riemann summation approximation of integrals,

$$\begin{aligned}
\hat{A}_1(u,v) &= \frac{1}{T}\sum_{t=1}^{T} \phi_t(u)\bar{M}_t(u) \\
&= \frac{1}{T}\sum_{t=1}^{T} \phi_t(u) e^{\mathbf{i}v2\pi t/T} - \frac{1}{T}\sum_{t=1}^{T} \phi_t(u) \frac{1}{T}\sum_{t=1}^{T} e^{\mathbf{i}v2\pi t/T} \\
&\to \int_0^1 \phi(u,\tau) e^{\mathbf{i}v2\pi\tau}\mathrm{d}\tau - \int_0^1 \phi(u,\tau)\mathrm{d}\tau \int_0^1 e^{\mathbf{i}v2\pi\tau}\mathrm{d}\tau \;\equiv\; \mu(u,v).
\end{aligned}$$

For $\hat{A}_2(u,v)$, we have $E[\hat{A}_2(u,v)] = 0$, for all $(u,v) \in \mathbb{W}$, and by the Davydov inequality for strong mixing processes and Assumption A.3(i),

https://doi.org/10.1017/S0266466622000494 Published online by Cambridge University Press

$$E\left|\hat{A}_2(u,v)\right|^2 = E\left|\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t(u)\bar{M}_t(v)\right|^2 \le \frac{4}{T^2}\sum_{t,s=1}^{T}|E[\varepsilon_t(u)\varepsilon_s(u)]|$$
$$\le \frac{C}{T^2}\sum_{t,s=1}^{T}\bar{\alpha}(|t-s|) = O(T^{-1}).$$

It follows that $\hat{A}_2(u,v) = O_P(T^{-1/2})$ for any $(u,v) \in \mathbb{W}$. Then, the conclusion follows. ■

A.4. Proof of Theorem 3.2

Note that

$$\begin{aligned} T^{-1}\hat{D} &= \int_{\mathbb{R}^{dm+1}}\left|\hat{A}(u,v)\right|^2 W(u,v)\,\mathrm{d}u\,\mathrm{d}v \\ &= \int_{\mathbb{R}^{dm+1}}\left|\hat{A}_1(u,v)\right|^2 W(u,v)\,\mathrm{d}u\,\mathrm{d}v + \int_{\mathbb{R}^{dm+1}}\left|\hat{A}_2(u,v)\right|^2 W(u,v)\,\mathrm{d}u\,\mathrm{d}v \\ &\quad + 2\mathrm{Re}\left(\int_{\mathbb{R}^{dm+1}}\hat{A}_1(u,v)\hat{A}_2(u,v)^* W(u,v)\,\mathrm{d}u\,\mathrm{d}v\right) \\ &\equiv \hat{D}_I + \hat{D}_{II} + 2\hat{D}_{III}, \end{aligned}$$

where Re(A) denotes the real part of A. By the Riemann summation approximation of integrals,

$$\hat{D}_I = \int_{\mathbb{R}^{dm+1}}\left|\hat{A}_1(u,v)\right|^2 W(u,v)\,\mathrm{d}u\,\mathrm{d}v \to \int_{\mathbb{R}^{dm+1}}|\mu(u,v)|^2 W(u,v)\,\mathrm{d}u\,\mathrm{d}v \equiv \mu_2 > 0.$$

By straightforward moment calculations, we can show that $E(\hat{D}_{II}) = O(T^{-1})$. Then, $\hat{D}_{II} = O_P(T^{-1})$ by the Markov inequality. In addition, $|\hat{D}_{III}| \le \{\hat{D}_I\hat{D}_{II}\}^{1/2} = O_P(T^{-1/2})$ by the Cauchy–Schwarz inequality. Consequently, we have shown that $T^{-1}\hat{D}$ is bounded away from 0 in probability by a positive constant μ_2 under Assumption A.3(ii). Then, the conclusion follows. ■

A.5. Proof of Theorem 3.3

Under $\mathbb{H}_A(\Delta_T): \phi_t(u) = \phi_0(u) + \Delta_T\theta_t(u)$ with $\Delta_T = T^{-1/2}$ and $\theta_t(u) = \theta(u,t/T)$, we have

$$\begin{aligned} \sqrt{T}\hat{A}_1(u,v) &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[\phi_0(u) + \Delta_T\theta_t(u)\right]\bar{M}_t(v) = \frac{1}{T}\sum_{t=1}^{T}\theta(u,t/T)\bar{M}_t(v) \\ &\to \zeta(u,v), \end{aligned}$$

where $\zeta(u,v) \equiv \mathrm{cov}[\theta(u,\tau), e^{\mathbf{i}v2\pi\tau}] = \int_0^1\theta(u,\tau)e^{\mathbf{i}v2\pi\tau}\mathrm{d}\tau - \int_0^1\theta(u,\tau)\mathrm{d}\tau\int_0^1 e^{\mathbf{i}v2\pi\tau}\mathrm{d}\tau$ is a pseudo-covariance between $\theta(u,\tau)$ and $e^{\mathbf{i}v2\pi\tau}$ in the sense that τ follows $U[0,1]$.

Now, consider $\hat{A}_2(u,v)$. As in the proof of Proposition 3.1, by letting $\sqrt{T}\hat{A}_2(u,v) = \hat{S}(u,v)$, we have that for each fixed $(u',v)' \in \mathbb{R}^{dm+1}$,

$$E[\hat{S}(u,v)] = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\bar{M}_t(v)E[\varepsilon_t(u)] = 0,$$

$$E[\hat{S}(u,v)\hat{S}(u,v)^*] = \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\text{cov}(e^{\mathbf{i}u'Y_t}, e^{-\mathbf{i}u'Y_s})\bar{M}_t(v)\bar{M}_s(v)^*,$$

and $\max_t \left|\bar{M}_t(v)\varepsilon_t(u)\right| \leq 4$. Then, by Assumption A.3 and Theorem 5.20 of White (2001), we have

$$\hat{S}(u,v) \xrightarrow{d} N[0, \mathcal{K}_1(w,w)],$$

where $\mathcal{K}_1(w,w) = \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\text{cov}(e^{\mathbf{i}u'Y_t}, e^{-\mathbf{i}u'Y_s})\bar{M}_t(v)\bar{M}_s(v)^*$ with $w = (u',v)'$. Following analogous arguments in the proof of Proposition 3.1, we can also show that $\hat{S}(u,v)$ is asymptotically tight on $\mathbb{U}\times\mathbb{V}$ under $\mathbb{H}_A(\Delta_T)$. Then,

$$\hat{S}(u,v) \Rightarrow \mathcal{S}(u,v),$$

where $\mathcal{S}(u,v)$ is a complex-valued Gaussian process with covariance kernel now given by

$$\mathcal{K}_1(w_1,w_2) \equiv \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\text{cov}(e^{\mathbf{i}u_1'Y_t}, e^{-\mathbf{i}u_2'Y_s})\bar{M}_t(v_1)\bar{M}_s(v_2)^*,$$

where $w_l = \left(u_l', v_l\right)' \in \mathbb{W}$, for $l = 1,2$. Note that $\mathcal{K}_1(w_1,w_2)$ is well defined under Assumption A.3 and $\mathcal{K}_1(w_1,w_2) = \mathcal{K}_0(w_1,w_2)$ under $\mathbb{H}_0$. By the CMT and analogous arguments in the proof of Theorem 3.1, we obtain the desired results. ■

A.6. Proof of Theorem 3.4

(i) We show that under $\mathbb{H}_A(\Delta_T)$ with $\Delta_T = o(b_T^{-1/2})$, $\hat{D}^\star \xrightarrow{d^\star} \int_{\mathbb{R}^{dm+1}}|\mathcal{S}(u,v)|^2\, W(u,v)\,\mathrm{d}u\,\mathrm{d}v$ in probability. Letting $\hat{S}(u,v)^\star = \sqrt{T}\hat{A}(u,v)^\star$, we prove this result in three steps: (i1) we establish the finite-dimensional convergence of $\{\hat{S}(u,v)^\star\}$ for any $(u,v) \in \mathbb{W} = \mathbb{U}\times\mathbb{V}$; (i2) we establish the asymptotic tightness of $\{\hat{S}(u,v)^\star\}$ on $\mathbb{W}$; and (i3) we establish the convergence in distribution of $\hat{D}^\star$ in probability.

Step (i1): We establish the finite-dimensional convergence of $\{\hat{S}(u,v)^\star\}$. Recall that $\bar{M}_t(v) \equiv e^{\mathbf{i}v2\pi t/T} - \frac{1}{T}\sum_{s=1}^{T}e^{\mathbf{i}v2\pi s/T}$, $\hat{\varepsilon}_t(u) = e^{\mathbf{i}u'Y_t} - \hat{\phi}_0(u)$, and $\varepsilon_t(u) = e^{\mathbf{i}u'Y_t} - E(e^{\mathbf{i}u'Y_t}) = e^{\mathbf{i}u'Y_t} - \phi_t(u)$. Let $w = \left(u',v\right)'$ and $w_l = \left(u_l',v_l\right)'$, for $l = 1,\ldots,k$, where (u,v), $(u_l,v_l) \in \mathbb{W}$. Let $\mathcal{K}(w_1,w_2)^\star \equiv \text{cov}^\star(\hat{S}(w_1)^\star, \hat{S}(w_2)^\star)$. Then,

$$\hat{S}(w)^\star = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\bar{M}_t(v)[e^{\mathbf{i}u'Y_t} - \hat{\phi}_0(u)]\eta_t^\star \text{ and}$$

$$\mathcal{K}(w_1,w_2)^\star = \frac{1}{T}\sum_{t,s=1}^{T}\bar{M}_t(v_1)\bar{M}_s(v_2)^*[e^{\mathbf{i}u_1'Y_t} - \hat{\phi}_0(u_1)][e^{-\mathbf{i}u_2'Y_s} - \hat{\phi}_0(-u_2)]A_T(t-s).$$

https://doi.org/10.1017/S0266466622000494 Published online by Cambridge University Press

Using $e^{\mathbf{i}u'Y_t} = \varepsilon_t(u) + \phi_t(u)$, we make the following decomposition for $\mathcal{K}(w_1,w_2)^\star$:

$$\begin{aligned}\mathcal{K}(w_1,w_2)^\star &= \frac{1}{T}\sum_{t,s=1}^{T} \bar{M}_t(v_1)\bar{M}_s(v_2)^*\varepsilon_t(u_1)\varepsilon_s(u_2)^* A_T(t-s)\\ &\quad + \frac{1}{T}\sum_{t,s=1}^{T} \bar{M}_t(v_1)\bar{M}_s(v_2)^*[\phi_t(u_1)-\hat{\phi}_0(u_1)][\phi_s(u_2)-\hat{\phi}_0(u_2)]^* A_T(t-s)\\ &\quad + \frac{1}{T}\sum_{t,s=1}^{T} \bar{M}_t(v_1)\bar{M}_s(v_2)^*\varepsilon_t(u_1)[\phi_s(u_2)-\hat{\phi}_0(u_2)]^* A_T(t-s)\\ &\quad + \frac{1}{T}\sum_{t,s=1}^{T} \bar{M}_t(v_1)\bar{M}_s(v_2)^*[\phi_t(u_1)-\hat{\phi}_0(u_1)]\varepsilon_s(u_2)^* A_T(t-s)\\ &\equiv \sum_{\ell=1}^{4}\mathcal{K}_l(w_1,w_2)^\star. \qquad \textbf{(A.6.1)}\end{aligned}$$

First, we prove that $\mathcal{K}_1(w_1,w_2)^\star \xrightarrow{p} \mathcal{K}_1(w_1,w_2)$ by showing that $E[\mathcal{K}_1(w_1,w_2)^\star] \to \mathcal{K}_1(w_1,w_2)$ and $\mathrm{Var}[\mathcal{K}_1(w_1,w_2)^\star] = o(1)$. Noting that $A_T(0) = 1$, we have

$$\begin{aligned}E[\mathcal{K}_1(w_1,w_2)^\star] &= \frac{1}{T}\sum_{t,s=1}^{T} \bar{M}_t(v_1)\bar{M}_s(v_2)^* E[\varepsilon_t(u_1)\varepsilon_s(u_2)^*] A_T(t-s)\\ &= \frac{1}{T}\sum_{t,s=1}^{T} \bar{M}_t(v_1)\bar{M}_s(v_2)^* E[\varepsilon_t(u_1)\varepsilon_s(u_2)^*]\\ &\quad + \frac{1}{T}\sum_{t=1}^{T-1}\sum_{s=t+1}^{T} \bar{M}_t(v_1)\bar{M}_s(v_2)^* E[\varepsilon_t(u_1)\varepsilon_s(u_2)^*]\{A_T(t-s)-1\}\\ &\quad + \frac{1}{T}\sum_{t=2}^{T}\sum_{s=1}^{t-1} \bar{M}_t(v_1)\bar{M}_s(v_2)^* E[\varepsilon_t(u_1)\varepsilon_s(u_2)^*]\{A_T(t-s)-1\}\\ &\equiv I_{1,1}(w_1,w_2) + I_{1,2}(w_1,w_2) + I_{1,3}(w_1,w_2).\end{aligned}$$

For $I_{1,1}(w_1,w_2)$, we have $I_{1,1}(w_1,w_2) = \frac{1}{T}\sum_{s,t=1}^{T} \bar{M}_t(v_1)\bar{M}_s(v_2)^* E[\varepsilon_t(u_1)\varepsilon_s(u_2)^*] \to \mathcal{K}_1(w_1,w_2)$. Pick up a sequence $\{p_T\}$ such that $1/p_T + p_T/b_T = o(1)$ as $T \to \infty$. For $I_{1,2}(w_1,w_2)$, we use the fact that $\sup_v \max_t \left|\bar{M}_t(v)\right| \le 2$ to obtain

$$\begin{aligned}\left|I_{1,2}(w_1,w_2)\right| &= \left|\frac{1}{T}\sum_{t=1}^{T-1}\sum_{j=1}^{T-t} \bar{M}_t(v_1)\bar{M}_{t+j}(v_2)^* E[\varepsilon_t(u_1)\varepsilon_{t+j}(u_2)^*]\{A_T(j)-1\}\right|\\ &\le \frac{4}{T}\sum_{t=1}^{T-1}\sum_{j=1}^{p_T} \left|E[\varepsilon_t(u_1)\varepsilon_{t+j}(u_2)^*]\right| |A_T(j)-1|\\ &\quad + \frac{4}{T}\sum_{t=1}^{T-1}\sum_{j=p_T+1}^{T} \left|E[\varepsilon_t(u_1)\varepsilon_{t+j}(u_2)^*]\right| |A_T(j)-1|.\end{aligned}$$

https://doi.org/10.1017/S0266466622000494 Published online by Cambridge University Press

The first term on the right-hand side (rhs) of the last inequality is $o(1)$ by the DCT by noticing that $\frac{1}{T}\sum_{t=1}^{T-1}\sum_{j=1}^{T}\left|E[\varepsilon_t(u_1)\varepsilon_{t+j}(u_2)^*]\right| = O(1)$ and $A_T(j)-1 = o(1)$ for any $|j| \le p_T = o(b_T)$; the second term on the rhs is bounded above by $\frac{4}{T}\sum_{t=1}^{T-1}\sum_{j=p_T+1}^{T}\left|E[\varepsilon_t(u_1)\varepsilon_{t+j}(u_2)^*]\right| = o(1)$ as $p_T \to \infty$. Then, $I_{1,2}(w_1,w_2) = o(1)$. Similarly, $I_{1,3}(w_1,w_2) = o(1)$. It follows that

$$E[\mathcal{K}_1(w_1,w_2)^\star] \to \mathcal{K}_1(w_1,w_2).$$

Let $\mathrm{cum}(Z_1,Z_2,Z_3,Z_4)= E\left[Z_1Z_2Z_3Z_4\right] - E[Z_1Z_2]E\left[Z_3Z_4\right] - E\left[Z_1Z_3\right]E[Z_2Z_4] - E[Z_1Z_4]E\left[Z_2Z_3\right]$. Then,

$$\begin{aligned}
&\mathrm{Var}\left(\mathcal{K}_1(w_1,w_2)^\star\right)\\
&= \frac{1}{T^2}\sum_{t,s,r,q=1}^{T}\bar{M}_t(v_1)\bar{M}_s(v_2)^*\bar{M}_r(v_1)\bar{M}_q(v_2)^*\{E[\varepsilon_t(u_1)\varepsilon_s(u_2)^*\varepsilon_r(u_1)\varepsilon_q(u_2)^*]\\
&\quad - E[\varepsilon_t(u_1)\varepsilon_s(u_2)^*]E[\varepsilon_r(u_1)\varepsilon_q(u_2)^*]\}A_T(t-s)A_T(r-q)\\
&= \frac{1}{T^2}\sum_{t,s,r,q=1}^{T}\bar{M}_t(v_1)\bar{M}_s(v_2)^*\bar{M}_r(v_1)\bar{M}_q(v_2)^*\mathrm{cum}(\varepsilon_t(u_1),\varepsilon_s(u_2)^*,\varepsilon_r(u_1),\varepsilon_q(u_2)^*)\\
&\qquad \times A_T(t-s)A_T(r-q)\\
&\quad + \frac{1}{T^2}\sum_{t,s,r,q=1}^{T}\bar{M}_t(v_1)\bar{M}_s(v_2)^*\bar{M}_r(v_1)\bar{M}_q(v_2)^*E[\varepsilon_t(u_1)\varepsilon_r(u_1)]E[\varepsilon_s(u_2)^*\varepsilon_q(u_2)^*]\\
&\qquad \times A_T(t-s)A_T(r-q)\\
&\quad + \frac{1}{T^2}\sum_{t,s,r,q=1}^{T}\bar{M}_t(v_1)\bar{M}_s(v_2)^*\bar{M}_r(v_1)\bar{M}_q(v_2)^*E[\varepsilon_t(u_1)\varepsilon_q(u_2)^*]E[\varepsilon_s(u_2)^*\varepsilon_r(u_1)]\\
&\qquad \times A_T(t-s)A_T(r-q)\\
&\equiv I_{1,4}(w_1,w_2)+I_{1,5}(w_1,w_2)+I_{1,6}(w_1,w_2).
\end{aligned}$$

Let $1 \le t \le s \le r \le q \le T$ and $\kappa = \max\{s-t, r-s, q-r\}$. By the uniform boundedness of $\varepsilon_t(u)$ and Davydov inequality for strong mixing processes, it is standard to show that under Assumption A.3(i),

$$\left|\mathrm{cum}(\varepsilon_t(u_1),\varepsilon_s(u_2)^*,\varepsilon_r(u_1),\varepsilon_q(u_2)^*)\right| \le C\bar{\alpha}(\kappa).$$

It follows that

$$\left|I_{1,4}(w_1,w_2)\right| \le \frac{C}{T^2}\sum_{t,s,r,q=1}^{T}\left|\mathrm{cum}(\varepsilon_t(u_1),\varepsilon_s(u_2)^*,\varepsilon_r(u_1),\varepsilon_q(u_2)^*)\right| \le \frac{C}{T}\sum_{\kappa=1}^{T}\kappa^2\bar{\alpha}(\kappa) = O(T^{-1}).$$

Next,

$$
\begin{aligned}
&\left|I_{1,5}(w_1,w_2)\right| \\
&\le \frac{C}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{r=1}^{T}\sum_{q=1}^{T}|E[\varepsilon_t(u_1)\varepsilon_r(u_1)]|\left|E[\varepsilon_s(u_2)^*\varepsilon_q(u_2)^*]\right|A_T(t-s)A_T(r-q) \\
&\le \frac{C}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{|j|=0}^{T-1}\sum_{|k|=0}^{T-1}\left|E[\varepsilon_t(u_1)\varepsilon_{t+j}(u_1)]\right|\left|E[\varepsilon_s(u_2)^*\varepsilon_{s+k}(u_2)^*]\right|A_T(t-s)A_T(t-s+j-k) \\
&\le \left(\max_{u\in\{u_1,u_2\}}\max_t\sum_{|j|=0}^{T-1}\left|E[\varepsilon_t(u)\varepsilon_{t+j}(u)]\right|\right)^2\frac{C}{T}\sum_{l=0}^{T-1}A_T(l) \\
&= O(1)\,O(T^{-1}b_T) = O(T^{-1}b_T),
\end{aligned}
$$

where the next to last inequality follows from the Davydov inequality and the fact that $\sum_{l=0}^{T-1}A_T(l) = O(b_T)$. Analogously, $\left|I_{1,6}(w_1,w_2)\right| = O(T^{-1}b_T)$. Hence, $\mathrm{Var}\left(\mathcal{K}(w_1,w_2)^\star\right) = O(T^{-1}b_T) = o(1)$ under Assumption A.5 and

$$\mathcal{K}_1(w_1,w_2)^\star \xrightarrow{p} \mathcal{K}_1(w_1,w_2).$$

Next, we study $\mathcal{K}_2(w_1,w_2)^\star$. Under $\mathbb{H}_A(\Delta_T): \phi_t(u) = \phi_0(u) + \Delta_T\theta_t(u)$ with $\Delta_T = o(b_T^{-1/2})$, we have

$$
\begin{aligned}
&\mathcal{K}_2(w_1,w_2)^\star \\
&= \frac{1}{T}\sum_{t,s=1}^{T}\bar{M}_t(v_1)\bar{M}_s(v_2)^*[\phi_t(u_1)-\hat{\phi}_0(u_1)][\phi_s(u_2)-\hat{\phi}_0(u_2)]^*A_T(t-s) \\
&= \frac{1}{T}\sum_{t,s=1}^{T}\bar{M}_t(v_1)\bar{M}_s(v_2)^*\{[\phi_0(u_1)-\hat{\phi}_0(u_1)][\phi_0(u_2)-\hat{\phi}_0(u_2)]^* + \Delta_T^2\theta_t(u_1)\theta_s(u_2)^* \\
&\quad + \Delta_T[\phi_0(u_1)-\hat{\phi}_0(u_1)]\theta_s(u_2)^* + \Delta_T\theta_t(u_1)[\phi_0(u_2)-\hat{\phi}_0(u_2)]^*\}A_T(t-s) \\
&\equiv \sum_{\ell=1}^{4}\mathcal{K}_{2,\ell}(w_1,w_2)^\star.
\end{aligned}
$$

Let $\hat{\delta}(u) = \hat{\phi}_0(u) - \phi_0(u)$. It is easy to show that $\hat{\delta}(u) = O_P(T^{-1/2}+\Delta_T)$ for each u under $\mathbb{H}_A(\Delta_T)$. Then,

$$
\begin{aligned}
\left|\mathcal{K}_{2,1}(w_1,w_2)^\star\right| &\le \left|\hat{\delta}(u_1)\right|\left|\hat{\delta}(u_2)\right|\frac{1}{T}\sum_{t,s=1}^{T}\left|\bar{M}_t(v_1)\bar{M}_s(v_2)^*\right|A_T(t-s) \\
&= O_P\left(T^{-1}+\Delta_T^2\right)\sum_{j=0}^{T-1}A_T(j) = O_P\left((T^{-1}+\Delta_T^2)b_T\right)
\end{aligned}
$$

https://doi.org/10.1017/S0266466622000494 Published online by Cambridge University Press

and

$$\left|\mathcal{K}_{2,3}(w_1,w_2)^\star\right| \le C\left|\hat{\delta}(u_1)\right| \frac{\Delta_T}{T}\sum_{t,s=1}^{T} A_T(t-s) = O_P((T^{-1/2}+\Delta_T)\Delta_T b_T).$$

Similarly, $\left|\mathcal{K}_{2,4}(w_1,w_2)^\star\right| = O_P((T^{-1/2}+\Delta_T)\Delta_T b_T)$. For $\mathcal{K}_{2,1}(w_1,w_2)^\star$, we have

$$\left|\mathcal{K}_{2,2}(w_1,w_2)^\star\right| \le \frac{C\Delta_T^2}{T}\sum_{t,s=1}^{T} A_T(t-s) = O(\Delta_T^2 b_T).$$

Consequently, we have shown that $\mathcal{K}_2(w_1,w_2)^\star = O_P\left((T^{-1}+\Delta_T^2+T^{-1/2}\Delta_T)b_T\right) = o_P(1)$.

Next, we study $\mathcal{K}_3(w_1,w_2)^\star$ and $\mathcal{K}_4(w_1,w_2)^\star$. For $\mathcal{K}_3(w_1,w_2)^\star$, we make the following decomposition:

$$\begin{aligned}
\mathcal{K}_3(w_1,w_2)^\star &= \frac{1}{T}\sum_{t,s=1}^{T} \bar{M}_t(v_1)\bar{M}_s(v_2)^*\varepsilon_t(u_1)[\phi_s(u_2)-\hat{\phi}_0(u_2)]^* A_T(t-s) \\
&= \frac{[\phi_0(u_2)-\hat{\phi}_0(u_2)]^*}{T}\sum_{t,s=1}^{T} \bar{M}_t(v_1)\bar{M}_s(v_2)^*\varepsilon_t(u_1)A_T(t-s) \\
&\quad + \frac{\Delta_T}{T}\sum_{t,s=1}^{T} \bar{M}_t(v_1)\bar{M}_s(v_2)^*\varepsilon_t(u_1)\theta_s(u_2)^* A_T(t-s) \\
&\equiv [\phi_0(u_2)-\hat{\phi}_0(u_2)]^*\mathcal{K}_{3,1}(w_1,w_2)^\star + \mathcal{K}_{3,2}(w_1,w_2)^\star.
\end{aligned}$$

It is straightforward to show that $E\left[\mathcal{K}_{3,\ell}(w_1,w_2)^\star\right]=0$, for $\ell=1,2$,

$$\mathrm{Var}[\mathcal{K}_{3,1}(w_1,w_2)^\star] \le \frac{C}{T^2}\sum_{t,s,r,q=1}^{T} |E[\varepsilon_t(u_1)\varepsilon_r(u_1)]|A_T(t-s)A_T(r-q) = O\left(T^{-1}b_T^2\right),$$

$$\mathrm{Var}[\mathcal{K}_{3,2}(w_1,w_2)^\star] \le \frac{C\Delta_T^2}{T^2}\sum_{t,s,r,q=1}^{T} |E[\varepsilon_t(u_1)\varepsilon_r(u_1)]|A_T(t-s)A_T(r-q) = O\left(\Delta_T^2 T^{-1}b_T^2\right).$$

It follows that $\mathcal{K}_3(w_1,w_2)^\star = O_P\left(\Delta_T T^{-1/2}b_T + T^{-1}b_T\right) = o_P(1)$ under Assumption A.5. Analogously, we can show that $\mathcal{K}_4(w_1,w_2)^\star = O_P\left(\Delta_T T^{-1/2}b_T + T^{-1}b_T\right) = o_P(1)$.

In sum, we have shown that

$$\mathcal{K}(w_1,w_2)^\star \equiv \mathrm{cov}^\star\left(\hat{S}(w_1)^\star,\hat{S}(w_2)^\star\right) \xrightarrow{p} \mathcal{K}_1(w_1,w_2) \text{ under } H_A(\Delta_T) \text{ with } \Delta_T = o(b_T^{-1/2}).$$

Now, noticing that the finite-dimensional distributions of the empirical process $\{\hat{S}(w)^\star\}$ are centered Gaussian by construction (cf. Doukhan et al., 2015), the above convergence of the covariance kernel in conjunction with the Cramér–Wold device implies that for arbitrary $w_1,\ldots,w_k \in \mathbb{W}$ and $k \in \mathbb{N}$,

$$\left(\hat{S}(w_1)^\star,\ldots,\hat{S}(w_k)^\star\right) \xrightarrow{d^\star} (S(w_1),\ldots,S(w_k)) \text{ in probability,}$$

where $\{S(w)\}$ is the limit Gaussian process with covariance kernel $\mathcal{K}_1(w_1, w_2)$. It is worth mentioning that if $\{\eta_t^\star\}$ is generated as a non-Gaussian process, then we can follow the proof of Theorem 3.1 in Shao (2010) to establish the above claim.

Step (i2): We establish the asymptotic tightness of $\{\hat{S}(w)^\star\}$. We aim to prove that, for any $\epsilon_1, \epsilon_2 > 0$, there exists $\varpi > 0$ such that

$$P\left(P^\star\left(\sup_{\|w_1-w_2\|\le\varpi}\left|\hat{S}(w_1)^\star - \hat{S}(w_2)^\star\right| \ge \epsilon_1\right) \le \epsilon_2\right) \to 1.$$

Note that $\hat{S}(w)^\star = \frac{1}{\sqrt{T}}\sum_{t=1}^T \bar{M}_t(v)[e^{\mathbf{i}u'Y_t} - \hat{\phi}_0(u)]\eta_t^\star$. Then,

$$\begin{aligned}
&E^\star\left[\left|\hat{S}(w_1)^\star - \hat{S}(w_2)^\star\right|^2\right] \\
&= E^\star\left[\left|\frac{1}{\sqrt{T}}\sum_{t=1}^T \bar{M}_t(v_1)[e^{\mathbf{i}u_1'Y_t} - \hat{\phi}_0(u_1)]\eta_t^\star - \frac{1}{\sqrt{T}}\sum_{t=1}^T \bar{M}_t(v_2)[e^{\mathbf{i}u_2'Y_t} - \hat{\phi}_0(u_2)]\eta_t^\star\right|^2\right] \\
&= E^\star\left\{\left|\frac{1}{\sqrt{T}}\sum_{t=1}^T\left[e^{\mathbf{i}u_1'Y_t} - e^{\mathbf{i}u_2'Y_t} - \hat{\phi}_0(u_1) + \hat{\phi}_0(u_2)\right]\bar{M}_t(v_1)\eta_t^\star\right.\right. \\
&\qquad\left.\left. + \frac{1}{\sqrt{T}}\sum_{t=1}^T[e^{\mathbf{i}u_2'Y_t} - \hat{\phi}_0(u_2)]\left[\bar{M}_t(v_1) - \bar{M}_t(v_2)\right]\eta_t^\star\right|^2\right\} \\
&\le 3E^\star\left[\left|\frac{1}{\sqrt{T}}\sum_{t=1}^T[\varepsilon_t(u_1) - \varepsilon_t(u_2)]\bar{M}_t(v_1)\eta_t^\star\right|^2\right] \\
&\qquad + 3E^\star\left[\left|\frac{1}{\sqrt{T}}\sum_{t=1}^T\left[\hat{\phi}_0(u_1) - \phi_t(u_1) - \hat{\phi}_0(u_2) + \phi_t(u_2)\right]\bar{M}_t(v_1)\eta_t^\star\right|^2\right] \\
&\qquad + 3E^\star\left[\left|\frac{1}{\sqrt{T}}\sum_{t=1}^T[e^{\mathbf{i}u_2'Y_t} - \hat{\phi}_0(u_2)]\left[\bar{M}_t(v_1) - \bar{M}_t(v_2)\right]\eta_t^\star\right|^2\right] \\
&\equiv 3II_1 + 3II_2 + 3II_3.
\end{aligned}$$

By Taylor expansions,

$$\begin{aligned}
E(II_1) &= EE^\star\left[\left|\frac{1}{\sqrt{T}}\sum_{t=1}^T[\varepsilon_t(u_1) - \varepsilon_t(u_2)]\bar{M}_t(v_1)\eta_t^\star\right|^2\right] \\
&\le EE^\star\left[\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^T \Upsilon_t(\bar{u})\bar{M}_t(v_1)\eta_t^\star\right\|^2\right]\|u_1 - u_2\|^2
\end{aligned}$$

$$= \frac{1}{T}\sum_{s,t=1}^{T} \mathrm{cov}(Y_t e^{\mathbf{i}\bar{u}'Y_t}, Y_s e^{-\mathbf{i}\bar{u}'Y_s})\bar{M}_t(v_1)\bar{M}_s(v_1)^* A_T(t-s)\,\|u_1-u_2\|^2$$
$$\leq C\|u_1-u_2\|^2,$$

and

$$II_3 = E^\star\left[\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}[e^{\mathbf{i}u_2'Y_t}-\hat{\phi}_0(u_2)]\left[\bar{M}_t(v_1)-\bar{M}_t(v_2)\right]\eta_t^\star\right|^2\right]$$
$$\leq E^\star\left[\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}[e^{\mathbf{i}u_2'Y_t}-\hat{\phi}_0(u_2)]\Psi_t(\bar{v})\eta_t^\star\right|^2\right](v_1-v_2)^2,$$

where $\Upsilon_t(\cdot),\Psi_t(\cdot),\ \bar{u}$, and $\bar{v}$ are as defined in the proof of Proposition 3.1. Note that

$$E^\star\left[\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}[e^{\mathbf{i}u_2'Y_t}-\hat{\phi}_0(u_2)]\Psi_t(\bar{v})\eta_t^\star\right|^2\right] \leq 3E^\star\left[\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}[e^{\mathbf{i}u_2'Y_t}-\phi_t(u_2)]\Psi_t(\bar{v})\eta_t^\star\right|^2\right]$$
$$+3E^\star\left[\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}[\phi_0(u_2)-\hat{\phi}_0(u_2)]\Psi_t(\bar{v})\eta_t^\star\right|^2\right]$$
$$+3E^\star\left[\left|\frac{\Delta_T}{\sqrt{T}}\sum_{t=1}^{T}\theta_t(u_2)\,\Psi_t(\bar{v})\eta_t^\star\right|^2\right]$$
$$\equiv 3II_{3,1}+3II_{3,2}+3II_{3,3}.$$

By straightforward (moment) calculations,

$$E\left|II_{3,1}\right| = \frac{1}{T}\sum_{s,t=1}^{T}\mathrm{cov}(e^{\mathbf{i}u_2'Y_t}, e^{-\mathbf{i}u_2'Y_s})\Psi_t(\bar{v})\Psi_s(\bar{v})^* A_T(t-s) = O(1),$$

$$E\left|II_{3,3}\right| = \frac{\Delta_T^2}{T}\sum_{s,t=1}^{T}\theta_t(u_2)\,\theta_s(u_2)\,\Psi_t(\bar{v})\Psi_s(\bar{v})^* A_T(t-s) = O\left(\Delta_T^2 b_T\right),\ \text{and}$$

$$II_{3,2} = \frac{\left|\phi_0(u_2)-\hat{\phi}_0(u_2)\right|^2}{T}\sum_{t,s=1}^{T}\Psi_t(\bar{v})\Psi_s(\bar{v})A_T(t-s) = O\left((T^{-1}+\Delta_T^2)b_T\right).$$

Then, $II_3 = O_P(1)(v_1-v_2)^2$.

To study II_2, let $\hat{\delta}_s(u_1,u_2) = \hat{\phi}_0(u_1)-\phi_s(u_1)-\hat{\phi}_0(u_2)+\phi_s(u_2)$ .Then, under $\mathbb{H}_A(\Delta_T)$, we have, uniformly in s,

https://doi.org/10.1017/S0266466622000494 Published online by Cambridge University Press

$$\begin{aligned}\left|\hat{\delta}_s(u_1,u_2)\right| &= \left|\frac{1}{T}\sum_{t=1}^{T}[\varepsilon_t(u_1)+\phi_t(u_1)] - \frac{1}{T}\sum_{t=1}^{T}[\varepsilon_t(u_2)+\phi_t(u_2)] - [\phi_s(u_1)-\phi_s(u_2)]\right| \\ &\leq \left|\frac{1}{T}\sum_{t=1}^{T}[\varepsilon_t(u_1)-\varepsilon_t(u_2)]\right| + \left|\frac{1}{T}\sum_{t=1}^{T}[\phi_t(u_1)-\phi_t(u_2)]\right| + |\phi_s(u_1)-\phi_s(u_2)| \\ &= \left|\frac{1}{T}\sum_{t=1}^{T}\mathbf{i}\Upsilon_t(\bar{u})'(u_1-u_2)\right| + \left|\frac{\Delta_T}{T}\sum_{t=1}^{T}[\theta_t(u_1)-\theta_t(u_2)]\right| + \Delta_T|\theta_s(u_1)-\theta_s(u_2)| \\ &\leq O_P(T^{-1/2}+\Delta_T)\|u_1-u_2\|,\end{aligned}$$

where we use the fact that $\frac{1}{T}\sum_{t=1}^{T}\Upsilon_t(\bar{u}) = O_P(T^{-1/2})$ and $\theta_t(\cdot)$ is Lipschitz continuous. Then,

$$\begin{aligned}|II_2| &= E^\star\left[\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[\hat{\phi}_0(u_1)-\phi_t(u_1)-\hat{\phi}_0(u_2)+\phi_t(u_2)\right]\bar{M}_t(v_1)\eta_t^\star\right|^2\right] \\ &= \frac{1}{T}\sum_{t,s=1}^{T}\hat{\delta}_t(u_1,u_2)\hat{\delta}_s(u_1,u_2)\bar{M}_t(v_1)\bar{M}_s(v_1)A_T(t-s) \\ &\leq \max_t\left|\hat{\delta}_t(u_1,u_2)\right|^2\frac{C}{T}\sum_{t,s=1}^{T}A_T(t-s) = O_P\left((T^{-1}+\Delta_T^2)b_T\right)\|u_1-u_2\|^2.\end{aligned}$$

In sum, we have shown that

$$E^\star\left[\left|\hat{S}(w_1)^\star - \hat{S}(w_2)^\star\right|^2\right] = O_P(1)\left[\|u_1-u_2\|^2+(v_1-v_2)^2\right] = O_P(1)\|w_1-w_2\|^2,$$

which implies the asymptotic tightness of $\{\hat{S}(w)^\star\}$ in probability.

Step (i3): We establish the convergence in distribution of $\hat{D}^\star$. Given the results in Steps (i1) and (i2), the remaining arguments essentially follow from those in the proof of Theorem 3.1.

(ii) Now consider the case where the alternative $\mathbb{H}_A(\Delta_T)$ is satisfied with $\Delta_T \neq o(b_T^{-1/2})$ for the original data. In this case, the decomposition in (A.6.1) continues to be valid, and so does the study for $\mathcal{K}_1(w_1,w_2)^\star$ in Step (i1) above: $\mathcal{K}_1(w_1,w_2)^\star \xrightarrow{p} \mathcal{K}_1(w_1,w_2)$. Following the analysis of $\mathcal{K}_l(w_1,w_2)^\star$, for $\ell = 2,3,4$ in Step (i1), now we can establish only that

$$\mathcal{K}_2(w_1,w_2)^\star = O_P\left(\Delta_T^2 b_T\right) \text{ and } \mathcal{K}_\ell(w_1,w_2)^\star = O_P(\Delta_T T^{-1/2} b_T) \text{ for } \ell = 3,4.$$

Consequently, $\mathcal{K}(w_1,w_2)^\star \equiv \mathrm{cov}^\star(\hat{S}(w_1)^\star,\hat{S}(w_2)^\star) = O_P\left(1+\Delta_T^2 b_T\right)$. This implies that $\hat{S}(w)^\star$ diverges to infinity at most at rate $\Delta_T b_T^{1/2}$ if $\Delta_T b_T^{1/2} \gg 1$. In the latter case, it is easy to find that the leading term in $\hat{S}(w)^\star$ is given by

$$\hat{S}_1(w)^\star \equiv \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\bar{M}_t(v)[\phi_t(u)-\hat{\phi}_0(u)]\eta_t^\star = O_{P^\star}(\Delta_T b_T^{1/2}),$$

where the order holds uniformly in w. As a result, the leading term in the expansion of $\hat{D}^{\star}$ is given by

$$\hat{D}_1^{\star} \equiv \int_{\mathbb{R}^{dm+1}} \left| \hat{S}_1(u,v)^{\star} \right|^2 W(u,v)\,\mathrm{d}u\,\mathrm{d}v = O_{P^{\star}}(\Delta_T^2 b_T),$$

which diverges to infinity at most at rate $\Delta_T^2 b_T$ when $\Delta_T b_T^{1/2} \gg 1$.

Let $\bar{O}$ and $\bar{O}_{P^{\star}}$ denote exact (probability) order such that $a_T = \bar{O}(c_T)$ signifies both $a_T/c_T = O(1)$ and $c_T/a_T = O(1)$. When $\Delta_T b_T^{1/2} = \bar{O}(1)$, it is easy to show that $\hat{D}^{\star} = \bar{O}_{P^{\star}}(1)$. This completes the proof of the theorem. ■

SUPPLEMENTARY MATERIAL

Fu, Z., S. Gao, L. Su and X. Wang (2022) Supplement to "Testing for Strict Stationarity via the Discrete Fourier Transform". *Econometric Theory Supplementary Material*. To view, please visit: https://doi.org/10.1017/S0266466622000494

REFERENCES

Bierens, H. (1982) Consistent model specification tests. *Journal of Econometrics* 20, 105–134.

Bierens, H. (1990) A consistent conditional moment test of functional form. *Econometrica* 58, 1443–1458.

Billingsley, P. (1999) *Convergence of Probability Measures*. Wiley.

Busetti, F. & A. Harvey (2010) Tests of strict stationarity based on quantile indicators. *Journal of Time Series Analysis* 31, 435–450.

Busetti, F. & A. Harvey (2011) When is a copula constant? A test for changing relationships. *Journal of Financial Econometrics* 9, 106–131.

Busetti, F. & A.M.R. Taylor (2003) Variance shifts, structural breaks, and stationarity tests. *Journal of Business & Economic Statistics* 21, 510–531.

Busetti, F. & A.M.R. Taylor (2004) Tests of stationarity against a change in persistence. *Journal of Econometrics* 123, 33–66.

Cai, Z. (2007) Trending time-varying coefficient time series models with serially correlated errors. *Journal of Econometrics* 136, 163–188.

Carrasco, M. & X. Chen (2002) Mixing and moment properties of various GARCH and stochastic volatility models. *Econometric Theory* 18, 17–39.

Cavaliere, G. & A.M.R. Taylor (2005) Stationarity tests under time-varying second moments. *Econometric Theory* 21, 1112–1129.

Cavaliere, G. & A.M.R. Taylor (2009) A note on testing covariance stationarity. *Econometric Reviews* 28, 364–371.

Chen, B. & Y. Hong (2010) Characteristic function-based testing for multifactor continuous-time Markov models via nonparametric regression. *Econometric Theory* 26, 1115–1179.

Chen, X. & Y. Fan (2006) Estimation of copula-based semiparametric time series models. *Journal of Econometrics* 130, 307–335.

Davidson, J. (1994) *Stochastic Limit Theory*. Oxford University Press.

Davis, R.A., M. Matsui, T. Mikosch, & P. Wan (2018) Applications of distance correlation to time series. *Bernoulli* 24, 3087–3116.

Dickey, D.A. & W.A. Fuller (1979) Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association* 74, 427–431.

Dickey, D.A. & W.A. Fuller (1981) Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica* 49, 1057–1072.
Diebold, F.X., T.A. Gunther, & A.S. Tay (1998) Evaluating density forecasts with applications to financial risk management. *International Economic Review* 39, 863–883.
Diebold, F.X., J. Hahn, & A.S. Tay (1999) Multivariate density forecast evaluation and calibration in financial risk management: High-frequency returns on foreign exchange. *The Review of Economics and Statistics* 81, 661–673.
Doukhan, P. (1994) *Mixing: Properties and Examples*. Springer.
Doukhan, P., G. Lang, A. Leucht, & M.H. Neumann (2015) Dependent wild bootstrap for the empirical process. *Journal of Time Series Analysis* 36, 290–314.
Dwivedi, Y. & S. Subba Rao (2011) A test for second-order stationarity of a time series based on the discrete Fourier transform. *Journal of Time Series Analysis* 32, 68–91.
Fan, J. & Q. Yao (2003) *Nonlinear Time Series: Nonparametric and Parametric Methods*. Springer.
Francq, C. & J.-M. Zakoïan (2012) Strict stationarity testing and estimation of explosive and stationary generalized autoregressive conditional heteroscedasticity models. *Econometrica* 80, 821–861.
Fu, Z. & Y. Hong (2019) A model-free consistent test for structural change in regression possibly with endogeneity. *Journal of Econometrics* 211, 206–242.
Fu, Z., Y. Hong and X. Wang (2022a) On multiple structural breaks in distribution: An empirical characteristic function approach. *Econometric Theory*, published online 25 March 2022. doi:10.1017/S026646662200010X.
Fu, Z., Y. Hong, & X. Wang (2022b) Testing for structural changes in large dimensional factor models via discrete Fourier transform. *Journal of Econometrics*, published online 6 August 2022. doi:10.1016/j.jeconom.2022.06.005.
Giné, E. & J. Zinn (1990) Bootstrapping general empirical measures. *The Annals of Probability* 18, 851–869.
González-Rivera, G. & Y. Sun (2017) Density forecast evaluation in unstable environments. *International Journal of Forecasting* 33, 416–432.
Guo, S., D. Li, & M. Li (2019) Strict stationarity testing and GLAD estimation of double autoregressive models. *Journal of Econometrics* 211, 319–337.
Hegwood, N.D. & D.H. Papell (1998) Quasi purchasing power parity. *International Journal of Finance & Economics* 3, 279–289.
Hobijn, B., P.H. Franses, & M. Ooms (2004) Generalizations of the KPSS-test for stationarity. *Statistica Neerlandica* 58, 483–502.
Hong, Y., X. Wang, & S. Wang (2017) Testing strict stationarity with applications to macroeconomic time series. *International Economic Review* 58, 1227–1277.
Hong, Y. & H. White (1995) Consistent specification testing via nonparametric series regression. *Econometrica* 63, 1133–1159.
Inoue, A. (2001) Testing for distributional change in time series. *Econometric Theory* 17, 156–187.
Jentsch, C. & S. Subba Rao (2015) A test for second order stationarity of a multivariate time series. *Journal of Econometrics* 185, 124–161.
Joe, H. (1997) *Multivariate Models and Multivariate Dependence Concepts*. CRC Press.
Kapetanios, G. (2009) Testing for strict stationarity in financial variables. *Journal of Banking and Finance* 33, 2346–2362.
Koedijk, K.G., M.M.A. Schafgans, & C.G. de Vries (1990) The tail index of exchange rate returns. *Journal of International Economics* 29, 93–108.
Kwiatkowski, D., P.C.B. Phillips, P. Schmidt, & Y. Shin (1992) Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics* 54, 159–178.
Leucht, A. & M.H. Neumann (2013) Dependent wild bootstrap for degenerate U- and V-statistics. *Journal of Multivariate Analysis* 117, 257–280.
Leybourne, S.J. & B.P.M. McCabe (1994) A consistent test for a unit root. *Journal of Business & Economic Statistics* 12, 157–166.

Li, Q. & J.S. Racine (2006) *Nonparametric Econometrics: Theory and Practice*. Princeton University Press.

Lin, C.-H. & T.-C. Kao (2008) Multiple structural changes in the tail behavior: Evidence from stock index futures returns. *Nonlinear Analysis: Real World Applications* 9, 1702–1713.

Lothian, J.R. & M.P. Taylor (1996) Real exchange rate behavior: The recent float from the perspective of the past two centuries. *Journal of Political Economy* 104, 488–509.

Manner, H., F. Stark, & D. Wied (2019) Testing for structural breaks in factor copula models. *Journal of Econometrics* 208, 324–345.

Mark, N.C. (1990) Real and nominal exchange rates in the long run: An empirical investigation. *Journal of International Economics* 28, 115–136.

Meese, R. & K. Rogoff (1988) Was it real? The exchange rate-interest differential relation over the modern floating-rate period. *The Journal of Finance* 43, 933–948.

Papell, D.H. (1997) Searching for stationarity: Purchasing power parity under the current float. *Journal of International Economics* 43, 313–332.

Phillips, P.C.B. & P. Perron (1988) Testing for a unit root in time series regression. *Biometrika* 75, 335–346.

Politis, D.N. & H. White (2004) Automatic block-length selection for the dependent bootstrap. *Econometric Reviews* 23, 53–70.

Quintos, C., Z. Fan, & P.C.B. Phillips (2001) Structural change tests in tail behavior and the Asian crisis. *The Review of Economic Studies* 68, 633–663.

Ramsay, J.O. (1991) Kernel smoothing approaches to nonparametric item characteristic curve estimation. *Psychometrika* 56, 611–630.

Rho, Y. & X. Shao (2019) Bootstrap-assisted unit root testing with piecewise locally stationary errors. *Econometric Theory* 35, 142–166.

Robinson, P.M. (1989) Nonparametric estimation of time-varying parameters. In P. Hackl (ed.), *Statistical Analysis and Forecasting of Economic Structural Change*, pp. 253–264. Springer.

Robinson, P.M. (1991) Time-varying nonlinear regression. In P. Hackl & A. H. Westlund (eds.), *Economic Structural Change*, pp. 179–190. Springer.

Rossi, B. (2013) Exchange rate predictability. *Journal of Economic Literature* 51, 1063–1119.

Rossi, B. & T. Sekhposyan (2013) Conditional predictive density evaluation in the presence of instabilities. *Journal of Econometrics* 177, 199–212.

Shao, X. (2010) The dependent wild bootstrap. *Journal of the American Statistical Association* 105, 218–235.

Sollis, R., S. Leybourne, & P. Newbold (2002) Tests for symmetric and asymmetric nonlinear mean reversion in real exchange rates. *Journal of Money, Credit and Banking* 34, 686–700.

Stinchcombe, M.B. & H. White (1998) Consistent specification testing with nuisance parameters present only under the alternative. *Econometric Theory* 14, 295–325.

Su, L. & H. White (2007) A consistent characteristic function-based test for conditional independence. *Journal of Econometrics* 141, 807–834.

Su, L. & H. White (2010) Testing structural change in partially linear models. *Econometric Theory* 26, 1761–1806.

Su, L. & X. Zheng (2017) A martingale-difference-divergence-based test for specification. *Economics Letters* 156, 162–167.

Székely, G.J., M.L. Rizzo, & N.K. Bakirov (2007) Measuring and testing dependence by correlation of distances. *Annals of Statistics* 35, 2769–2794.

Wang, X. & Y. Hong (2018) Characteristic function based testing for conditional independence: A nonparametric regression approach. *Econometric Theory* 34, 815–849.

White, H. (2001) *Asymptotic Theory for Econometricians*, Revised Edition. Academic Press.

Wu, Y. (1996) Are real exchange rates nonstationary? Evidence from a panel-data test. *Journal of Money, Credit and Banking* 28, 54–63.

Xiao, Z. (2001) Testing the null hypothesis of stationarity against an autoregressive unit root alternative. *Journal of Time Series Analysis* 22, 87–105.

Xiao, Z. & L.R. Lima (2007) Testing covariance stationarity. *Econometric Reviews* 26, 643–667.

Yang, S.C. (2007) Maximal moment inequality for partial sums of strong mixing sequences and application. *Acta Mathematica Sinica, English Series* 23, 1013–1024.

Yokoyama, R. (1980) Moment bounds for stationary mixing sequences. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 52, 45–57.

Zhou, Z. (2012) Measuring nonlinear dependence in time-series, a distance correlation approach. *Journal of Time Series Analysis* 33, 438–457.