INTERNATIONAL ECONOMIC REVIEW
Vol. 58, No. 4, November 2017

TESTING STRICT STATIONARITY WITH APPLICATIONS TO MACROECONOMIC TIME SERIES*

By Yongmiao Hong, Xia Wang, and Shouyang Wang[1]

Cornell University, U.S.A. and Xiamen University, China; Sun Yat-Sen University, China; Academy of Mathematics and Systems Science, Chinese Academy of Sciences, China and University of Chinese Academy of Sciences, China

We propose a model-free test for strict stationarity. The idea is to estimate a nonparametric time-varying characteristic function and compare it with the empirical characteristic function based on the whole sample. We also propose several derivative tests to check time-invariant moments, weak stationarity, and pth order stationarity. Monte Carlo studies demonstrate excellent power of our tests. We apply our tests to various macroeconomic time series and find overwhelming evidence against strict and weak stationarity for both level and first-differenced series. This suggests that the conventional time series econometric modeling strategies may have room to be improved by accommodating these time-varying features.

1. INTRODUCTION

A standard modeling strategy in time series econometrics is to first remove a trend from or take a (log-) difference of an obviously nonstationary economic time series and then employ a stationary, linear, or nonlinear time series model (e.g., vector autoregression, autoregressive threshold, and Markov regime switching) for the detrended or differenced series. Another widespread modeling method for nonstationary economic time series is unit root and cointegration. The popular unit root models assume that the stochastic shocks to the nonstationary series are stationary, whereas cointegration models assume the individual series are first-order integrated while their linear combination is stationary (Granger, 1981; Engle and Granger, 1987; Johansen, 1991). All above modeling strategies rely on the assumption of stationarity for the differenced series together with constant model parameters.

In time series econometrics, the assumption of stationarity provides a feasible way to infer the dynamics of a time series by combining realizations over different time periods together. This assumption simplifies modeling mechanisms and allows for elegant development of time series econometric methodology. It plays a crucial role in time series inference and forecasting. For example, many nonparametric and semiparametric estimators, such as kernel and local polynomial estimators, are all constructed based on the implication of the strict stationary assumption (Pagan and Schwert, 1990; Pagan and Ullah, 1999; Li and Racine, 2006). In addition, strict stationarity is also an indispensable assumption for the tests involving the entire probability structure (e.g., Rosenblatt, 1975; Robinson, 1991; Hong and White, 2005; Su and White, 2007, 2008).

Although stationarity is widely assumed in time series analysis, it may not be realistic when one bases inferences on observations over a long period. A main driving force for economic

*Manuscript received November 2014; revised May 2016.

[1] Xia Wang acknowledges financial supports from the National Science Foundation of China (No. 71401160), the Project of Ministry of Education of Humanities and Social Sciences China (No. 14YJC790120), and Fujian Provincial Key Laboratory of Statistics (Xiamen University, No. 2016003). We thank the editor Jesús Fernández-Villaverde, two referees, and the participants of 2014 Econometric Society China Meeting, 2015 Econometrics Conference in Honor of Aman Ullah, and Vanderbilt University for their comments and suggestions. All remaining errors are solely our own. Please address correspondence to: Xia Wang, Lingnan (University) College, Sun Yat-Sen University, Guangzhou 510275, China. Phone: +86 20 8411 1191. E-mail: *wangxia25@mail.sysu.edu.cn*.

structural changes is the "shock" induced by institutional changes, such as the changes of exchange rate systems from a fixed exchange rate mechanism to a floating exchange rate mechanism or the introduction of the Euro. Changes induced by policy switch, preference change, and technology progress can also cause structural changes. These changes may make the characteristics of time series, such as moments and distributions to be time varying. For example, Stock and Watson (1996) empirically check the stability of 76 univariate autoregressions and 5,700 bivariate relationships constituted by 76 representative U.S. monthly postwar macroeconomic time series and document significant evidence of instability in both univariate and bivariate autoregressive models. If the stationarity condition fails, inference and forecasting based on such an assumption may lead to misleading conclusions. Numerous empirical studies (e.g., Goyal and Welch, 2008; Dangl and Halling, 2012) document overwhelmingly significant in-sample evidence on predictability of asset returns but rather poor out-of-sample forecasts of various econometric models, which is likely due to the structural instability of the underlying economic processes. Obviously, if the underlying time series is nonstationary but we still build models under the framework of stationarity, the resulting economic analysis and inferences would be misleading.

Motivated by the important role of stationarity, numerous studies are devoted to test this fundamental property. Although one can suspect before running a test that the null of stationarity will be rejected in most cases, it is still necessary to develop some formal econometric tools to check this assumption rigorously. On the one hand, a formal test procedure may derive more reliable results than intuitive judgment; on the other hand, when the null of stationarity is rejected, one may like to gauge possible reasons of rejection, which may provide some ways to derive the stationary process. For example, if the rejections are due to the smooth and/or abrupt structural changes of the time series, one may shorten the time periods to get a stationary time series. There are several notions for stationarity. In the time series literature, a time series $\{Y_t\}$ is called a strictly stationary process if all its finite-dimensional joint distributions are time invariant, whereas it is called weakly stationary if the first two moments are time invariant. Moreover, $\{Y_t\}$ is called a pth-order stationary process if its first pth-order joint product moments are time invariant. Almost all existing tests for stationarity focus on weak stationarity instead of strict stationarity. More specifically, the existing literature mainly focuses on the unit root process, which is a special form of nonstationarity. They either test trend stationarity against unit root (e.g., Kwiatkowski et al., 1992; Bierens, 1993; Bierens and Guo, 1993; Xiao and Lima, 2007) or test unit root against trend stationarity (e.g., Dickey and Fuller, 1979, 1981; Phillips, 1987; Phillips and Perron, 1988). Recently, Dette et al. (2011) and Preuß et al. (2013) propose some model-free tests for weak stationarity by measuring the L^2 and Kolmogorov–Smirnov type distances between a time-varying spectral density estimator and a spectral density estimator based on the whole sample. Although the existing studies have achieved fruitful achievements on testing weak stationarity, the literature on testing strict stationarity is still deficient. In linear time series analysis, weak stationarity is a most suitable concept in most cases. However, when we turn to a nonlinear time series framework, the concept of strict stationarity is more useful, as the first two moments are insufficient to characterize the full dynamics and nonlinear features of a time series. With the development and prevalence of nonlinear modeling, strict stationarity becomes a maintained assumption. Testing strict stationarity is more challenging. To our knowledge, there are only a few papers that consider testing strict stationarity in the literature. Kapetanios (2009) tests the strict stationarity assumption by examining the time-invariance property of a nonparametric marginal density estimator based on a recursive method. Busetti and Harvey (2010) propose a test for strict stationarity based on a quantile indicator. To execute their test, one should compute the test statistic for each quantile $\tau \in (0, 1)$. Recently, Francq and Zakoïan (2012) propose a test for strict stationarity within the Generalized Autoregressive Conditional Heteroskedasticity (GARCH) framework.

In this article, we propose a model-free test for strict stationarity of a possibly vector-valued time series. Assuming that the finite-dimensional distribution of a time series is a smooth function of time, we estimate the time-varying characteristic function by local smoothing and compare it with the empirical characteristic function based on the whole sample. By construction, our test is most powerful against smooth or evolutionary structural changes in distribution. This is in contrast to most tests of structural breaks in the econometric literature, which only focus on abrupt structural breaks. Smooth structural changes are more realistic in reality. As Hansen (2001) points out, "it may seem unlikely that a structural break could be immediate and might seem more reasonable to allow a structural change to take a period of time to take effect." In addition to smooth structural changes in finite-dimensional distributions of the time series, our test is also able to detect a finite number of abrupt changes (i.e., structural breaks) in distribution, with possibly unknown break dates or the number of breaks. Our test complements the existing tests for strict stationarity and has a number of appealing features.

First of all, our test is constructed by examining the time invariance property of the finite-dimensional distribution of a time series instead of only its marginal distribution. Given a vector valued time series $\{Y_t\}$ and any finite time points $(t_1, t_2, \dots, t_m)$, we check whether the joint distribution of $(Y_{t_1}, Y_{t_2}, \dots, Y_{t_m})$ is time invariant. A special case of $m = 1$ checks the time invariant property of the marginal distribution, which is in spirit similar to the null hypothesis of Kapetanios (2009) and Busetti and Harvey (2010). By choosing m larger than 1, our test then checks the time invariance property of a joint distribution and hence could capture a wider range of deviations from strict stationarity, such as those with a time-varying joint distribution but time-invariant marginal distributions.

Second, our test could detect a class of local alternatives of smooth distributional changes with convergence rate $T^{-1/2}h^{-1/4}$ (where h is a bandwidth), which depends on neither the dimension of Y_t nor the number of time points checked. Unlike Kapetanios' (2009) test, which is based on the marginal density estimator using a recursive method, i.e., estimating the density $f_t(y)$ at Y_t using the first t observations, we first take the Fourier transform of a joint finite-dimensional distribution of the time series and then estimate the time-varying characteristic function by local smoothing. Thanks to the use of the characteristic function and local regression method, our test could detect a class of local alternatives with convergence rate $T^{-1/2}h^{-1/4}$ and thus avoids the potential "curse of dimensionality" problem associated with the dimension of Y_t and the number of time points checked.

Third, unlike Francq and Zakoïan's (2012) test for strict stationarity, we do not need to impose any parametric restrictions on the form of alternatives. As a result, our test not only avoids the misspecification problem but also is powerful in detecting various kinds of nonstationarity, such as abrupt and smooth changes in various moments, as well as unit root and trend stationarity.

Fourth, by using a multivariate normal weighting function, we avoid the intractable high dimensional integrations. Unlike Székely et al.'s (2007) nonintegrable weighting function, the normal density function satisfies our regularity conditions, particularly the integrability condition. Simulation studies show that our test with the normal weighting function performs reasonably well.

Fifth, our test is flexible in gauging possible sources of nonstationarity. By differentiating our test for strict stationarity with respect to auxiliary parameters at the origin up to various orders, we obtain a class of derivative tests for the time invariant property of various moments. The derivative tests could be used to test weak stationarity and, more generally, the pth order stationarity.

Finally, all the proposed test statistics follow a convenient null asymptotic $N(0, 1)$ distribution. The limiting distributions of most existing tests for stationarity are nonstandard, and one has to simulate critical values in practice. In contrast, it is easy to implement our tests by comparing the test statistics with the one-sided $N(0, 1)$ critical value z_α at significant level α.

In an empirical study, we apply our tests to various macroeconomic time series. We document overwhelming evidence of nonstationarity for both level and differenced series. And, our derivative tests document significant evidence of smoothly changing means and/or variances for the differenced series. These new findings suggest that the current time series modeling strategies, namely, modeling nonstationary economic time series by fitting a stationary linear or nonlinear model (e.g., VAR, threshold, regime-switching) to the differenced series or by using unit root and cointegration models with stationary disturbances, may have room to be improved by accommodating time-varying features. New time series models with smooth structural changes may be considered. An example is the class of locally stationary time series model (e.g., Dahlhaus, 1996) and its extensions with time-varying parameters.

The article is organized as follows. In Section 2, we formalize the strict stationarity property and state the hypotheses of interest. In Section 3, we construct our test statistic for strict stationarity. We derive the asymptotic distribution of our test in Section 4 and investigate its asymptotic power properties in Section 5. Section 6 develops a class of derivative tests for the time invariance property of various moments, including weak stationarity and pth order stationarity. In Section 7, we study the finite sample performance of our tests via simulation. Section 8 provides an empirical application to macroeconomic time series. Conclusions are provided in Section 9. All proofs are relegated to the Appendix.

2. STRICT STATIONARITY AND HYPOTHESES OF INTEREST

Let $\{Y_t\}_{t=1}^T$ be a d-dimensional stochastic time series, where $d \geq 1$. The components of Y_t could be either continuous or discrete random variables or a mixture of them. The process $\{Y_t\}$ is strict stationary if, for any admissible time indices $(t_1, t_2, \ldots, t_m)$ and any integer $k > 0$, the joint distribution of $\{Y_{t_1}, Y_{t_2}, \ldots, Y_{t_m}\}$ is the same as that of $\{Y_{t_1+k}, Y_{t_2+k}, \ldots, Y_{t_m+k}\}$. Different from the concept of weak stationarity, strict stationarity highlights the time invariance property of the entire joint distribution instead of only the first two moments of the time series.

Denote the joint cumulative distribution function (CDF) of $(Y_{t_1}, Y_{t_2}, \ldots, Y_{t_m})$ as follows:

$$F_{Y_{t_1},Y_{t_2},\ldots,Y_{t_m}}(y_1, y_2, \ldots, y_m) = P(Y_{t_1} \leq y_1, Y_{t_2} \leq y_2, \ldots, Y_{t_m} \leq y_m).$$

Then the null hypothesis of strict stationarity could be written as

$$\mathbb{H}_0 : F_{Y_{t_1},Y_{t_2},\ldots,Y_{t_m}}(y_1, y_2, \ldots, y_m) = F_{Y_{t_1+k},Y_{t_2+k},\ldots,Y_{t_m+k}}(y_1, y_2, \ldots, y_m) \tag{1}$$

for any admissible time indices $(t_1, t_2, \ldots, t_m)$, any sequence $(y_1, y_2, \ldots, y_m)$ with grouping size $m = 1, 2, \ldots$, and any integer $k > 0$. The alternative hypothesis is

$$\mathbb{H}_A : F_{Y_{t_1},Y_{t_2},\ldots,Y_{t_m}}(y_1, y_2, \ldots, y_m) \neq F_{Y_{t_1+k},Y_{t_2+k},\ldots,Y_{t_m+k}}(y_1, y_2, \ldots, y_m) \tag{2}$$

for some time indices $(t_1, t_2, \ldots, t_m)$, some sequence $(y_1, y_2, \ldots, y_m)$, and some integer $k > 0$.

Our approach is applicable to all admissible choices of time points $(t_1, t_2, \ldots, t_m)$. Without loss of generality, we assume $t_1 < t_2 < \cdots < t_m$. A special case is $(t_1, t_2, \ldots, t_m) = (t, t+1, \ldots, t+m-1)$. For any prespecified time points $(t_1, t_2, \ldots, t_m)$, put

$$X_t = \left(Y_t', Y_{t+t_2-t_1}', \ldots, Y_{t+t_m-t_1}'\right)'.$$

Then, the null hypothesis (1) could be rewritten using the CDF of X_t:

$$\mathbb{H}_0 : F_t(x) = P(X_t \leq x) \text{ does not depend on } t. \tag{3}$$

Thus, if we reject $\mathbb{H}_0$ for some time indices $(t_1, t_2, \ldots, t_m)$, we can conclude that the strict stationarity assumption fails. From a theoretical point of view, if only a finite number of time points is checked, we may miss some nonstationary processes with time-varying joint distributions. However, if the set of time indices considered is increased, the probability of making such a mistake will become smaller. In addition, the choice of time indices makes our test able to capture a broad range of nonstationarity. For example, the case of $m = 1$ yields the time invariance property of the marginal distribution, which is the null hypothesis of Kapetanios' (2009) strict stationarity test. As is well known, a time invariant marginal distribution of $\{Y_t\}$ is a necessary but not sufficient condition of strict stationarity. It is possible that the marginal distribution of $\{Y_t\}$ is time invariant, but the joint distribution of (Y_{t_1}, Y_{t_2}) is time varying. In this case, Kapetanios' (2009) test will have no power (see DGP.P7 in Section 7 for an example).

As the Fourier transform of the distribution function $F_t(x)$, the characteristic function

$$\phi_t(u) = E(e^{\mathbf{i}u'X_t}) = \int e^{\mathbf{i}u'X_t} dF_t(x) \tag{4}$$

contains the same information on the distribution of X_t. Therefore, we can use the characteristic function to equivalently represent the hypotheses $\mathbb{H}_0$ and $\mathbb{H}_A$ as follows:

$$\mathbb{H}_0 : \phi_t(u) = \psi_0(u) \quad \text{for any } u \in \mathbb{R}^{dm}, \tag{5}$$

$$\mathbb{H}_A : \phi_t(u) \neq \psi_0(u) \quad \text{for some } u \in \mathbb{R}^{dm}, \tag{6}$$

where $\psi_0(u) = E(e^{\mathbf{i}u'X_t})$ is an unknown time invariant characteristic function.

It is important to emphasize that we must check (5) for all possible values of $u \in \mathbb{R}^{dm}$ instead of only a subset of $\mathbb{R}^{dm}$. This is quite involved, but it ensures consistency of our test against $\mathbb{H}_A$. Moreover, by differentiating the characteristic function with respect to u, we can derive a class of test statistics for the time invariance property of various moments, including tests for weak stationarity and, more generally, the pth order stationarity for any positive integer $p \leq m$.

3. NONPARAMETRIC TESTING

3.1. *Nonparametric Estimation.* Given an observed possibly vector-valued sample $\{Y_t\}_{t=1}^T$ of size T, we can estimate the characteristic function of $X_t = (Y_t', Y_{t+t_2-t_1}', \ldots, Y_{t+t_m-t_1}')'$. Under $\mathbb{H}_0$, $\phi_t(u) = \psi_0(u)$ does not change over time, and it can be consistently estimated by the empirical characteristic function based on the entire sample:

$$\hat{\psi}_0(u) = \frac{1}{T}\sum_{t=1}^{T} e^{\mathbf{i}u'X_t}, \quad u \in \mathbb{R}^{dm}.$$

Since the time index of Y_t runs from 1 to T, the sum in $\hat{\psi}_0(u)$ is actually from 1 to $T - t_m$. However, for a finite constant t_m, we simply denote the upper bound of the sum as T, which does not affect our subsequent asymptotic analysis. Under $\mathbb{H}_A$, $\phi_t(u)$ is a function of u and t. Throughout, we assume

$$\phi_t(u) = \phi(u, t/T),$$

where $\phi(u, \cdot)$ is an unknown smooth function of t/T except for a finite number of discontinuity points, which imply a finite number of structural breaks in distribution. We note that the assumption that $\phi_t(u)$ is a function of ratio $t/T \in [0, 1]$ instead of time index t is a commonly used scaling scheme in the literature (e.g., Robinson, 1989; Phillips and Hansen, 1990; Cai, 2007). The intuitive explanation of this requirement is that the increasingly intense sampling of data points ensures consistent estimation of $\phi_t(u)$ at some fixed point t/T by increasing the amount of data on which it depends. For more discussion, one could refer to Robinson (1989, 1991).

To avoid model misspecification and allow our test to capture various deviations from $\mathbb{H}_0$, we use a nonparametric smoothing method to estimate $\phi_t(u)$. Specifically, we consider the following generalized time-varying regression model:

$$e^{\mathbf{i}u'X_t} = \phi_t(u) + \varepsilon_t(u), \tag{7}$$

where the generalized disturbance $\{\varepsilon_t(u) = e^{\mathbf{i}u'X_t} - \phi_t(u)\}$, which is a strictly stationary process under the null hypothesis with $E[\varepsilon_t(u)] = 0$ and a generalized long-run variance $\Omega(u, v) = \sum_{k=-\infty}^{\infty} \sigma_k(u, v)$, where $\sigma_k(u, v) = E[\varepsilon_t(u)\varepsilon_{t+k}(v)^*]$.

In fact, the representation in (7) can be viewed as a generalized form of a time-varying coefficient model with a local constant. The time-varying coefficient model was first introduced by Robinson (1989, 1991), and its nonparametric estimation was investigated by Robinson (1989, 1991), Orbe et al. (2000, 2005), and Cai (2007). The function $\phi(u, t/T)$ could be consistently estimated by various nonparametric methods such as kernel and local polynomial smoothing. Here, we use local linear smoothing. As shown in Cai (2007), although the kernel estimator shares the same asymptotic properties with the local linear estimator in the interior region of time, the latter converges faster than the former in the boundary regions and therefore avoids the well-known boundary problem in nonparametric estimation. According to Cai (2007), the local linear estimation of $\hat{\phi}_t(u)$ is given by

$$\hat{\phi}_t(u) \equiv \hat{\alpha}_0 = \frac{S_{T,2}\Gamma_{T,0} - S_{T,1}\Gamma_{T,1}}{S_{T,0}S_{T,2} - S_{T,1}^2}, \tag{8}$$

where $S_{T,k}(t) = T^{-1}\sum_{s=1}^{T}(\frac{s-t}{T})^k K_h(\frac{s-t}{T})$, $\Gamma_{T,k}(t) = T^{-1}\sum_{s=1}^{T}(\frac{s-t}{T})^k K_h(\frac{s-t}{T})e^{\mathbf{i}u'X_s}$, and $K_h(x) = h^{-1}K(x/h)$ with kernel $K : \mathbb{R} \to \mathbb{R}^+$ and bandwidth $h = h(T)$.

We can easily show that, uniformly in t/T in the interior region $[h, 1-h]$, the local linear estimator of $\phi_t(u)$ could be written as follows:

$$\hat{\phi}_t(u) = \frac{1}{T}\sum_{s=1}^{T} K_h\left(\frac{s-t}{T}\right) e^{\mathbf{i}u'X_s}[1 + o_P(1)], \tag{9}$$

and uniformly in $t/T = ch$ $(0 < c < 1)$ in the left boundary region, we have

$$\hat{\phi}_t(u) = \left[\frac{\mu_{2,c}}{\mu_{0,c}\mu_{2,c} - \mu_{1,c}^2}\Gamma_{T,0}(t) - \frac{\mu_{1,c}}{h\left(\mu_{0,c}\mu_{2,c} - \mu_{1,c}^2\right)}\Gamma_{T,1}(t)\right][1 + o_P(1)], \tag{10}$$

where $\mu_{k,c} = \int_{-c}^{1} u^k K(u)du$. A similar result holds for $t/T = 1 - ch$ in the right boundary region.

Equations (9) and (10) provide the local linear estimators of $\phi_t(u)$ in the interior and boundary regions, respectively. From Equations (9) and (10) we can further show that, although the convergence rate of $\hat{\phi}_t(u)$ in the boundary regions is the same as that in the interior region, the scales of asymptotic biases are different and the asymptotic variance at a boundary point is larger since fewer observations are available. To make the behavior of $\hat{\phi}_t(u)$ in the boundary regions similar to that in the interior region, we reflect the data in the boundary regions to obtain a pseudo data $X_t = X_{-t}$ for $-[Th] \le t \le -2$ and $X_t = X_{2T-t}$ for $T + 1 \le t \le T + [Th]$, where $[Th]$ denotes the integer part of Th. This reflection method was first introduced by Hall and Wehrly (1991) in nonparametric density estimation and was applied to estimate time-varying regression models by Chen and Hong (2012). By reflection, data points in the boundary regions become symmetric. Given a symmetric kernel with support on $[-1, 1]$, the estimator $\hat{\phi}_t(u)$ given by Equations (9) and (10) could be rewritten as

$$\hat{\phi}_t(u) = \frac{1}{T}\sum_{s=t-[Th]}^{t+[Th]} K_h\left(\frac{s-t}{T}\right) e^{\mathbf{i}u'X_s}[1 + o_P(1)], \text{ uniformly in } t \in \{1, 2, \ldots, T\}. \tag{11}$$

We emphasize that $\hat{\phi}_t(u)$ only involves smoothing over the normalized time t/T. As a result, neither the dimension d of Y_t nor the number m of time points affects the convergence rate of the estimator. This differs from Kapetanios' (2009) recursive estimator of the marginal density, which, in the present context, would involve smoothing over X_t and thus leads to the severe "curse of dimensionality" problem due to a large dimensionality of Y_t and/or a large number m of time points. The use of characteristic function and local linear regression allows us to avoid the notorious "curse of dimensionality" problem associated with d and m.

3.2. *Nonparametric Test.* Under $\mathbb{H}_0$, we have $\phi_t(u) = \psi_0(u)$ for all u and t. Therefore, we can test $\mathbb{H}_0$ by measuring the distance between $\hat{\phi}_t(u)$ and $\hat{\psi}_0(u)$ via the following sample quadratic form:

$$\hat{Q} = \frac{1}{T}\sum_{t=1}^{T}\int \left|\hat{\phi}_t(u) - \hat{\psi}_0(u)\right|^2 W(u)du, \tag{12}$$

where $W : \mathbb{R}^{dm} \to \mathbb{R}^+$ is a nonnegative symmetric weighting function of u. The introduction of $W(u)$ allows us to consider many points for u. Since X_t is a $dm \times 1$ vector, where the dimension d of Y_t and the number m of time points can be large, numerical integration in u would be computationally costly. To avoid intractable high dimensional numerical integration, one could employ any nonnegative weighting function with countable discontinuity points. One special case of discontinuous weighting functions is a discrete multivariate probability mass function. However, a discontinuous weighting function may adversely affect the power of the test. In this article, we will advocate using a continuous weighting function that can avoid numerical integration.

Our test statistic is a standardized version of (12):

$$\widehat{SQ} = (Th^{1/2}\hat{Q} - \hat{B})/\sqrt{\hat{V}}, \tag{13}$$

where the centering and scaling factors are

$$\hat{B} = h^{-1/2}\int |\hat{\Omega}(u,u)|W(u)du \int K^2(\tau)d\tau, \tag{14}$$

$$\hat{V} = 2\int\int |\hat{\Omega}(u,v)|^2 W(u)W(v)dudv \int \left[\int K(\tau)K(\tau+\eta)d\tau\right]^2 d\eta, \tag{15}$$

with $\hat{\Omega}(u,v)$ being a consistent estimator of the generalized long-run variance–covariance $\Omega(u,v) = \sum_{j=-\infty}^{\infty} E[\varepsilon_t(u)\varepsilon_{t+j}(v)^*]$. If the generalized disturbance $\varepsilon_t(u)$ is an i.i.d. or martingale difference sequence (m.d.s.) for each u, the centering and scaling factors $\hat{B}$ and $\hat{V}$ could be further simplified by replacing the generalized long-run covariances $|\Omega(u,u)|$ and $|\Omega(u,v)|$ with the generalized covariances $\sigma^2(u,u) = E[|\varepsilon_t(u)|^2] = 1 - |\psi_0(u)|^2$ and $\sigma^2(u,v) = |\psi_0(u+v) - \psi_0(u)\psi_0(v)|$, respectively.

Note that the generalized long-run variance appears in both centering factor $\hat{B}$ and scaling factor $\hat{V}$. In the conventional nonparametric regression with random covariate X_t, relaxing the i.i.d. or m.d.s. assumption to serial dependence does not change the asymptotic properties of a nonparametric estimator. However, this is not the case for a nonparametric smooth time-varying model (e.g., Cai, 2007). Intuitively speaking, the local linear estimator at the point x uses the data falling into the local interval $[x-h, x+h]$ in a space domain. For a random variable X_t, the observations whose values fall into the small interval are generally far away from each other in time and thus the dependence among the observations in the local interval is much weaker. However, in a nonparametric smooth time-varying model, the regressor is t/T. Thus, the local interval $[t/T - h, t/T + h]$ in the time domain contains observations with rather strong dependence. Consequently, the existence of serial dependence affects the asymptotic behavior of $\hat{\phi}_t(u)$.

To implement our test $\widehat{SQ}$, it is desired to provide a simple consistent estimator of the generalized long-run variance $\Omega(u,v)$. We extend Newey and West (1987) and Andrews' (1991) long-run variance estimator by replacing the conventional residuals with our generalized residuals $\varepsilon_t(u)$ to obtain an estimator of $\Omega(u, v)$. Define the jth-order sample generalized covariance function

$$\hat{\sigma}_j(u, v) = \begin{cases} \frac{1}{T}\sum_{t=j+1}^{T} \hat{\varepsilon}_t(u)\hat{\varepsilon}_{t-j}(v)^*, \ j = 0, 1, \ldots, p_T, \\ \frac{1}{T}\sum_{t=1-j}^{T} \hat{\varepsilon}_t(u)\hat{\varepsilon}_{t-j}(v)^*, \ j = -1, -2, \ldots, -p_T, \end{cases}$$

where the estimated generalized residuals $\hat{\varepsilon}_s(u) = e^{\mathbf{i}u'X_s} - \hat{\psi}_0(u)$. Then, $\Omega(u, v)$ could be estimated by

$$\hat{\Omega}(u, v) = \sum_{j=-p_T}^{p_T} k_l(j/p_T)\hat{\sigma}_j(u, v), \tag{16}$$

where the truncation lag order $p_T = p(T)$ satisfies $p_T h^{1/2} \to \infty$, $p_T/(Th) \to 0$ as $T \to \infty$. The kernel k_l in (16) is a symmetric, bounded, and square-integrable weighting function for lag order j. A simple example of k_l is the Bartlett kernel $k_l(z) = (1 - |z|)\mathbf{1}(|z| \leq 1)$, where $\mathbf{1}(\cdot)$ is the indicator function that takes value 1 if $|z| \leq 1$ and 0 otherwise.

As kindly pointed out by a referee, we could propose a combined test for strict stationarity. Denote $\hat{Q}_m$ as the quadratic form specified for the time indices $(t_1, t_2, \ldots, t_m)$ with grouping size m. The formula of $\hat{Q}_m$ is the same as that of $\hat{Q}$ given by Equation (12). Here, we explicit the index m to show that this distance is specified for the grouping size m. Now, we could define the combined test statistic $\hat{Q}^c = \sum_{m=1}^{M} \hat{Q}_m$ for some given maximum grouping size M. We could show that $\widehat{SQ}^c = [Th^{1/2}\hat{Q}^c - \hat{B}^c]/\sqrt{\hat{V}^c}$ is asymptotically $N(0, 1)$ with $\hat{B}^c = \sum_{m=1}^{M} \hat{B}_m$ and $\hat{V}^c$ being the asymptotic variance of $Th^{1/2}\hat{Q}^c$. However, due to the mutual dependence of $\hat{Q}^c$ over different m, the expression for $\hat{V}^c$ is rather tedious, and we suggest using the bootstrap method to approximate the limiting distribution of $Th^{1/2}\hat{Q}^c$.

3.3. *Test Statistic with a Product Normal PDF Weighting Function.* As mentioned before, the test statistic (13) involves dm-dimensional integration, which is usually calculated using numerical integration or approximated by simulation methods. When the dimension d of Y_t or the number m of time points is large, the aforementioned methods would be tremendously computationally costly. One can approximate integration by using a finite number of grid points for u. This could reduce the computational cost but may lead to power loss. Another way to avoid high dimensional numerical integration is to integrate (13) out analytically by choosing some suitable weighting function. In this article, we consider the following weighting function:

$$W(u) = (2\pi b)^{-\frac{dm}{2}} \exp\left(-\frac{|u|^2}{2b}\right), \tag{17}$$

where $b > 0$ is a scaling parameter, u is a dm-dimensional vector, and $|u|$ is the Euclidean norm of u in $\mathbb{R}^{dm}$. A special case of $b = 1$ is the product of the $N(0, 1)$ density, i.e., $W(u) = \prod_{i=1}^{dm} w(u_i)$ with $w(u_i) = (2\pi)^{-1/2}\exp(-u_i^2/2)$. Unlike Székely et al.'s (2007) weighting function, which is nonintegrable and does not have a finite moment, our weighting function in Equation (17) has some appealing features. It is integrable and has finite-order moments with $\int_{\mathbb{R}^{dm}} |u|^k W(u)du < \infty$ for any positive integer k, thus satisfying our regularity conditions. Most importantly, with this weighting function, the quadratic form (12) could be written as

$$\hat{Q}_W = \frac{1}{T^2}\sum_{s,l=-[Th]}^{T+[Th]} A(s, l)\exp\left(-\frac{b}{2}|X_s - X_l|^2\right), \tag{18}$$

where $A(s,l)=\frac{1}{T}\sum_{t=1}^{T}K_h(s,t)K_h(l,t)$ with

$$K_h(s,t)=\begin{cases}K_h\left(\frac{s-t}{T}\right)-1, & \text{if } 1\le s\le T,\\ K_h\left(\frac{s-t}{T}\right), & \text{if } s\le -2 \text{ or } s\ge T+1.\end{cases}$$

The test statistic (18) could be regarded as a weighted generalized distance between two realizations of the random vector X_t. Intuitively, the generalization from the distance $|X_s-X_l|$ to the exponential term allows us to capture the time-varying behavior of all higher order moments of X_t.

By employing the weighting function given by Equation (17), the standardized version test statistic (13) becomes

$$\widehat{SQ}_W=\left(Th^{1/2}\hat{Q}_W-\hat{B}_W\right)\bigg/\sqrt{\hat{V}_W}\,,$$

where $\hat{Q}_W$ is given by Equation (18) and

$$\hat{B}_W=h^{-1/2}\sum_{j=-p_T}^{p_T}k_l\left(\frac{j}{p_T}\right)\hat{\sigma}(j)\int K^2(\tau)d\tau \tag{19}$$

$$\hat{V}_W=2\sum_{i,j=-p_T}^{p_T}k_l\left(\frac{i}{p_T}\right)k_l\left(\frac{j}{p_T}\right)\hat{\sigma}(i,j)\int\left[\int K(\tau)K(\tau+\eta)d\tau\right]^2d\eta \tag{20}$$

with

$$\hat{\sigma}(j)=\frac{1}{T}\sum_{t=1+|j|}^{T}\hat{\delta}(t,t-j),\qquad \hat{\sigma}(i,j)=\frac{1}{T^2}\sum_{s=1+|i|}^{T}\sum_{t=1+|j|}^{T}\hat{\delta}(s,t)\hat{\delta}(s-i,t-j),$$

where $\hat{\delta}(i,j)=e^{-\frac{b}{2}|X_i-X_j|^2}-\frac{1}{T}\sum_{l_1=1}^{T}e^{-\frac{b}{2}|X_i-X_{l_1}|^2}-\frac{1}{T}\sum_{l_2=1}^{T}e^{-\frac{b}{2}|X_j-X_{l_2}|^2}+\frac{1}{T^2}\sum_{l_1,l_2=1}^{T}e^{-\frac{b}{2}|X_{l_1}-X_{l_2}|^2}$.

As a most appealing feature, $\widehat{SQ}_W$ avoids high dimensional numerical integration, which would otherwise be rather computationally tedious. The test statistic as well as the centering and scaling factors only involve summations, making our test quite convenient to use in practice. Simulation studies show that $\widehat{SQ}_W$ has excellent power in detecting various alternatives (cf. Section 7).

In addition to the normal weighting function, the product Laplace(0, b) PDF could also avoid numerical integration. By using the product Laplace(0, b) weighting function:

$$W(u)=(2b)^{-dm}\exp\left(-\frac{\sum_{i=1}^{dm}|u_i|}{b}\right),$$

our quadratic form test statistic given by (12) would become

$$\hat{Q}_W=\frac{1}{T^2}\sum_{s,l=-[Th]}^{T+[Th]}A(s,l)\prod_{i=1}^{dm}\frac{1}{1+b^2|X_{s,i}-X_{l,i}|^2},$$

where $X_{t,i}$ is the ith element of X_t. We note that Székely et al.'s (2007) nonintegrable weighting function also yields an analytic expression for the quadratic form $\hat{Q}$ in Equation (12), but it does not satisfy our regular conditions.

4. ASYMPTOTIC DISTRIBUTION

We will derive the asymptotic distribution of our test $\widehat{SQ}$ under $\mathbb{H}_0$. We first impose some regularity assumptions.

ASSUMPTION 1. *The stochastic process $\{Y_t\}$ is absolutely regular on $\mathbb{R}^d$ with β-mixing coefficients satisfying $\sum_{j=1}^{\infty} j^2\beta(j)^{\delta/(1+\delta)} < C$ for some $0 < \delta < 1$.*

ASSUMPTION 2. *The kernel function $K : \mathbb{R} \to \mathbb{R}^+$ is symmetric, bounded, and twice continuously differentiable over a compact support, say $[-1, 1]$, with $\int_{-1}^{1} K(u)du = 1$, $\int_{-1}^{1} uK(u)du = 0$, and $\int_{-1}^{1} u^r K(u)du = C_r < \infty$ for $r \geq 2$.*

ASSUMPTION 3. *$W : \mathbb{R}^{dm} \to \mathbb{R}^+$ is nonnegative symmetric integrable function with $\int_{\mathbb{R}^{dm}} |u|^4 W(u)du < \infty$.*

These assumptions are commonly used in the literature. The β-mixing condition in Assumption 1 restricts the degree of temporal dependence in $\{Y_t\}$ and is generally adopted in the nonparametric time series literature such as Hjellvik et al. (1998) and Chen and Hong (2010, 2012). In general, the mixing coefficients are defined to measure the strength of dependence for the two segments of a time series that are apart from each other in time. For a time series $\{Y_t, t = 0, \pm 1, \pm 2, \cdots\}$ and $n = 1, 2, \ldots$, define the β-mixing coefficient $\beta(n) = E\{\sup_{B \in \mathcal{F}_n^{\infty}} |P(B) - P(B|Y_0, Y_{-1}, \cdots)|\}$, where $\mathcal{F}_i^j$ denotes the σ-algebra generated by $\{Y_t, i \leq t \leq j\}$. If $\beta(n) \to 0$ as $n \to \infty$, then the process $\{Y_t\}$ is said to be β-mixing. A variety of time series processes, such as autoregressive moving average (ARMA), bilinear, and autoregressive conditional heteroskedastic (ARCH) process, satisfy the β-mixing condition (Fan and Li, 1999). The β-mixing assumption simplifies the proofs of theorems in this article but could be relaxed to the α-mixing condition or the local stationarity assumption (see, e.g., Zhang and Wu, 2011, 2012). In addition, we do not impose any continuity assumption on the distribution of Y_t. Therefore, the components of Y_t can be either continuous or discrete random variables or a mixture of them.

Assumption 2 imposes regularity conditions on the kernel function. The familiar positive and bounded kernels, such as the Epanechnikov, Quartic, and Uniform kernels, are included. However, it excludes the Gaussian kernel, which has unbounded support. Assumption 3 imposes some mild conditions on the weighting function $W(u)$. These conditions ensure the existence of the integral in (13). Many functions satisfy these conditions, an example being PDFs with finite fourth-order moments. In particular, Assumption 3 allows the normal weighting function given by Equation (17) but rules out Székely et al.'s (2007) nonintegrable weighting function.

Under these regularity conditions, we now state the asymptotic distribution of $\widehat{SQ}$ under $\mathbb{H}_0$.

THEOREM 1. *Suppose Assumptions 1–3 hold and $h = cT^{-\lambda}$ for $0 < \lambda < \frac{2}{3}$, where $0 < c < \infty$. Then $\widehat{SQ} \xrightarrow{d} N(0, 1)$ under $\mathbb{H}_0$ as $T \to \infty$.*

Our test statistic is based on a sample quadratic form, which measures the distance between the local linear estimator $\hat{\phi}_t(u)$ of the time-varying characteristic function and the empirical characteristic function $\hat{\psi}_0(u)$ based on the entire sample. Under $\mathbb{H}_0$, $\hat{\psi}_0(u)$ converges to the true characteristic function $\psi_0(u)$ with rate $\sqrt{T}$, which is faster than the convergence rate of the local linear estimator $\hat{\phi}_t(u)$. Consequently, the limiting behavior of $\widehat{SQ}$ is solely determined by $\hat{\phi}_t(u)$. In fact, the quadratic form statistic (after demeaned) yields a dominant degenerate U-statistic, which determines the asymptotic distribution of our test. We emphasize that, by allowing serial dependence in DGP, both the centering factor and scaling factor involve the generalized long-run variance. This differs from most of the existing nonparametric based tests, where relaxing i.i.d. or m.d.s. assumptions to some mixing conditions does not affect the limiting distribution of the tests. Intuitively speaking, the smoothing variable in our test is the time ratio t/T instead of an exogenous random variable. As a result, the observations with strong dependence show up in a local interval of time.

Theorem 1 allows the choice of a wide range of admissible rates for bandwidth h, including the optimal bandwidth $h \propto T^{-1/5}$ in terms of minimizing the integrated mean square error (MSE) of the nonparametric estimation for $\phi_t(u)$. In practice, one could choose h via some simple rules of thumb. In our simulation, we will use the bandwidth $h = (1/\sqrt{12})T^{-1/5}$, where $1/\sqrt{12}$ is the standard deviation of $U(0, 1)$, which could be viewed as the limiting distribution of grid points $\{t/T, t = 1, 2, \ldots, T\}$ as $T \to \infty$. We will examine the impact of using rule of thumb bandwidths with various tuning parameters. It is highly desirable to use data-driven methods to choose h in practice. Although the bandwidth based on cross-validation is asymptotically optimal for estimation of $\phi_t(u)$ in terms of the MSE, it is not optimal for our test. For testing problems, a central concern is Type I and Type II errors. In different but related contexts, Gao and Gijbels (2008) and Sun et al. (2008) propose some novel methods to choose a data-driven bandwidth by considering a trade-off between Type I and Type II errors. Specifically, based on the Edgeworth expansions of the asymptotic distribution of a test under a class of local alternatives, Gao and Gijbels (2008) choose h to maximize the power of their test subject to a control of Type I error, and Sun et al. (2008) choose h to minimize a weighted average of Type I and Type II errors, where the weight depends on the relative preference between Type I and Type II errors. It is possible to extend these approaches to our test. Nevertheless, the analytical expressions for the leading terms of the size and power functions or the two type errors of our test are rather involved. This is beyond the scope of this article and will be pursued in subsequent work.

Our quadratic form test statistic is nonnegative and hence is a one-sided test. It is asymptotically pivotal and has a convenient asymptotic $N(0, 1)$ distribution under $\mathbb{H}_0$. Consequently, we can compare the test statistic $\widehat{SQ}$ with the one-sided $N(0, 1)$ critical value z_α at significant level α, and reject $\mathbb{H}_0$ when $\widehat{SQ} > z_\alpha$.

5. ASYMPTOTIC POWER

To study the asymptotic power of $\widehat{SQ}$ under $\mathbb{H}_A$, we impose the following assumption.

ASSUMPTION 4. *Let $\phi_t(u) = E(e^{\mathbf{i}u'X_t})$ be the characteristic function of $X_t = (Y_t, Y_{t+t_2-t_1}, \ldots, Y_{t+t_m-t_1})'$ at time t for time indices $(t_1, t_2, \cdots t_m)$ in ascending order. (i) For each $u \in \mathbb{R}^{dm}$, $\phi_t(u) = \phi(u, \frac{t}{T})$ is continuous with respect to $\frac{t}{T}$ except for a finite number of discontinuity points $\tau \in [0, 1]$ satisfying $\sup_{\tau \in (0,1), u \in \mathbb{R}^{dm}} || \lim_{s \to \tau^+} \phi(u, s) - \lim_{s \to \tau^-} \phi(u, s)|| \leq C$. (ii) There exists a positive constant $c_\phi > 0$, such that $p \lim_{T \to \infty} \inf_{\tilde{\phi}(u) \in \Phi(u)} \frac{1}{T} \sum_{t=1}^{T} |\phi_t(u) - \tilde{\phi}(u)|^2 \geq c_\phi$ for u in some set with positive Lebesgue measure, where $\Phi(u)$ is a space that contains all possible cases of time invariant characteristic function.*

Assumption 4(i) allows the distribution of X_t to change smoothly over time or abruptly at discontinuity time points. No moment condition on X_t (or Y_t) is imposed, and X_t or Y_t need not to be continuous random variables or random vectors. This differs from Kapetanios (2009). Assumption 4(ii) is intuitively clear: In the space of characteristic function, we cannot find any time invariant characteristic function $\tilde{\phi}(u)$ such that $\tilde{\phi}(u)$ converges to the true characteristic function $\phi_t(u)$ for any u in the sense of MSE.

We first investigate the asymptotic power property of our test under $\mathbb{H}_A$.

THEOREM 2. *Suppose Assumptions 1–4 hold, and $h = cT^{-\lambda}$ for $0 < \lambda < \frac{2}{3}$ and $0 < c < \infty$. Then for any sequence of nonstochastic constants $\{M_T = o(Th^{1/2})\}$, $P(\widehat{SQ} > M_T) \to 1$ under $\mathbb{H}_A$ as $T \to \infty$.*

Theorem 2 shows that our test $\widehat{SQ}$ is consistent against any alternatives to $\mathbb{H}_0$ at any given significance level, subject to the regularity conditions imposed in Assumption 4. Note that we do not impose any restrictive auxiliary assumptions on the form of $\mathbb{H}_A$ except for the smoothness condition on the time-varying characteristic function $\phi(u, t/T)$. In particular, our model-free test is powerful in capturing various forms of nonstationarity, such as smooth and abrupt changes in the distribution of the time series $\{Y_t\}$.

To gain more insight into the asymptotic power of $\widehat{SQ}$, we now consider a class of local alternatives

$$\mathbb{H}_2(a_T):\ F_t(x) = F_0(x) + a_T q\left(x, \frac{t}{T}\right), \tag{21}$$

where $F_0(\cdot)$ is a time-invariant CDF, and $q(x, \frac{t}{T})$ is twice continuously differentiable with respect to t/T satisfying $q(x, \cdot) \neq 0$ on an interval of nonzero Borel measure and $\lim_{x\to\infty} q(x, t/T) = 0$. The term $a_T q(x, t/T)$ characterizes the departure of the time-varying CDF $F_t(\cdot)$ from the time invariant CDF $F_0(\cdot)$, and the rate a_T is the speed at which the deviation vanishes to 0 as the sample size $T \to \infty$. By taking the Fourier transform of Equation (21), we obtain

$$\phi_t(u) = \psi_0(u) + a_T \delta\left(u, \frac{t}{T}\right),$$

where $\delta(u, t/T) = \int e^{\mathbf{i}u'x} dq(x, t/T)$ is the Fourier transform of $q(x, t/T)$ and satisfies

$$\gamma \equiv \int \int_0^1 |\delta(u, \tau)|^2 W(u) d\tau du - \int \left[\int_0^1 \delta(u, \tau) d\tau\right]^2 W(u) du < \infty.$$

THEOREM 3. *Suppose Assumptions 1–4 and $\mathbb{H}_2(a_T)$ hold with $a_T = T^{-1/2}h^{-1/4}$ and bandwidth $h = cT^{-\lambda}$ for $0 < \lambda < \frac{2}{3}$ and $0 < c < \infty$. Then, as $T \to \infty$, the power of the test*

$$P[\widehat{SQ} \geq z_\alpha | \mathbb{H}_2(a_T)] \to 1 - \Phi(z_\alpha - \gamma/\sqrt{V}),$$

where $\Phi(\cdot)$ is the $N(0,1)$ CDF, z_α is the one-sided critical value of $N(0,1)$ at significance level α, and

$$V = \int \int |\Omega(u, v)|^2 W(u) W(v) du dv \int \left[\int \int K(\tau) K(\tau + \eta) d\tau\right]^2 d\eta.$$

Theorem 3 shows that our test has nontrivial power against the class of local alternatives $\mathbb{H}_2(a_T)$ with $a_T = T^{-1/2}h^{-1/4}$. The convergence rate $T^{-1/2}h^{-1/4}$ is slightly slower than the parametric rate $T^{-1/2}$. For example, if $h = T^{-1/5}$, then $T^{-1/2}h^{-1/4} = T^{-9/20}$, which is only slightly slower than $T^{-1/2}$. We note that the convergence rate is not affected by the dimension d of time series $\{Y_t\}$ and the number m of the prespecified time indices $(t_1, t_2, \ldots, t_m)$. In contrast, even though Kapetanios' (2009) "recursive" density estimator-based test only checks the time invariance property of marginal density of $\{Y_t\}$, it still suffers from a severe "curse of dimensionality" problem if the dimension d of Y_t is large. More importantly, we can show the convergence rate of Kapetanios' (2009) test would be $T^{-1/2}h^{-d/2}$, which is slower than our test and so is less asymptotically efficient for all $d \geq 1$. Thanks to the use of the characteristic function, our test is free of the notorious "curse of dimensionality" problem associated with the dimension d of Y_t and the number m of time points $(t_1, \ldots, t_m)$. More specifically, we use a regression approach, i.e., estimating the characteristic function (i.e., the Fourier transform of the probability density), whose convergence rate does not depend on the dimension dm. Moreover, unlike Kapetanios' (2009) test, which is only applicable to continuous random variables, we do not impose any continuity assumption on the distribution of Y_t. For our test, Y_t can be a discrete or continuous random vector or a mixture of them.

6. TESTING pTH-ORDER STATIONARITY

Another commonly used concept of stationarity in time series analysis is the pth-order stationarity. The stochastic vector-valued time series $\{Y_t\}$ is said to be stationary up to order p if, for any admissible time indices $(t_1, t_2, \ldots, t_m)$ and any integer $k > 0$, all the joint moments of $\{Y_{t_1}, Y_{t_2}, \ldots, Y_{t_m}\}$ up to order p exist and are equal to those of $\{Y_{t_1+k}, Y_{t_2+k}, \ldots, Y_{t_m+k}\}$, respectively. Denote $Y_t = (Y_{t,1}, Y_{t,2} \cdots, Y_{t,d})$. Then, we can represent the null hypothesis of pth-order stationarity as follows:

$$\mathbb{H}_0^{(p)} : E\left(Y_{t,1}^{p_{1,1}} \cdots Y_{t,d}^{p_{1,d}} Y_{t+t_2-t_1,1}^{p_{2,1}} \cdots Y_{t+t_m-t_1,d}^{p_{m,d}}\right) \text{ does not depend on } t \tag{22}$$

for any prespecified time indices $(t_1, t_2, \ldots, t_m)$ and any dm-dimensional integer $\tilde{p} = (p_{1,1}, \ldots, p_{1,d}, p_{2,1}, \ldots, p_{m,d})$ satisfying $\sum_{i=1}^m \sum_{j=1}^d p_{i,j} \leq p$. A special case of $p = 2$ yields the well-known concept of weak stationarity.

As the characteristic function can be differentiated to obtain various moments (if they exists), we may construct a class of derivative tests to check the pth-order stationarity. Suppose the pth-order moment of Y_t exists. For the characteristic function $\phi_t(u)$ in Equation (4), taking the $\tilde{p}$th-order partial derivative with respect to $u = (u_1, u_2, \ldots, u_{dm})$ at the origin, we have

$$\phi_t^{(\tilde{p})} = \frac{\partial^{\sum_{i=1}^m \sum_{j=1}^d p_{i,j}} \phi_t(u)}{\partial u_1^{p_{1,1}} u_2^{p_{1,2}} \cdots u_{dm}^{p_{m,d}}} |_{u=\mathbf{0}} = \mathbf{i}^{\sum_{i=1}^m \sum_{j=1}^d p_{i,j}} E\left(Y_{t,1}^{p_{1,1}} \cdots Y_{t,d}^{p_{1,d}} Y_{t+t_2-t_1,1}^{p_{2,1}} \cdots Y_{t+t_m-t_1,d}^{p_{m,d}}\right). \tag{23}$$

Therefore, the null hypothesis of (22) and its alternative could be equivalently represented as follows:

$$\mathbb{H}_0^{(p)} : \phi_t^{(\tilde{p})} = \psi_0^{(\tilde{p})} \text{ for any nonnegative integers with } \sum_{i=1}^m \sum_{j=1}^d p_{i,j} \leq p. \tag{24}$$

$$\mathbb{H}_A^{(p)} : \phi_t^{(\tilde{p})} \neq \psi_0^{(\tilde{p})} \text{ for some nonnegative integers with } \sum_{i=1}^m \sum_{j=1}^d p_{i,j} \leq p. \tag{25}$$

Similar to the construction of the test for strict stationarity, we can use the following quadratic form to test the time-invariance property of $\phi_t^{(\tilde{p})}$:

$$\hat{Q}^{(\tilde{p})} = \frac{1}{T} \sum_{t=1}^T \left|\hat{\phi}_t^{(\tilde{p})} - \hat{\psi}_0^{(\tilde{p})}\right|^2, \tag{26}$$

where $\hat{\phi}_t^{(\tilde{p})} = \frac{\partial^{\sum_{i=1}^m \sum_{j=1}^d p_{i,j}} \hat{\phi}_t(u)}{\partial u_1^{p_{1,1}} u_2^{p_{1,2}} \cdots u_{dm}^{p_{m,d}}} |_{u=0}$ and $\hat{\psi}_0^{(\tilde{p})} = \frac{1}{T} \sum_{t=1}^T Y_{t,1}^{p_{1,1}} \cdots Y_{t,d}^{p_{1,d}} Y_{t+t_2-t_1,1}^{p_{2,1}} \cdots Y_{t+t_m-t_1,d}^{p_{m,d}}$.

Following the proof of Theorem 1, we can show that, under the null hypothesis of $\mathbb{H}_0^{(p)}$ and some regular conditions such as the existence of Y_{it}'s $4p$th moment, the standardized version of $\hat{Q}^{(\tilde{p})}$ follows a standard normal distribution:

$$\widehat{SQ}^{(\tilde{p})} = \frac{Th^{1/2}\hat{Q}^{(\tilde{p})} - \hat{B}^{(\tilde{p})}}{\sqrt{\hat{V}^{(\tilde{p})}}} \xrightarrow{d} N(0, 1), \tag{27}$$

where

$$\hat{B}^{(\tilde{p})} = h^{-1/2}\hat{\Omega}(\tilde{p}) \int K^2(\tau) d\tau,$$

$$\hat{V}^{(\tilde{p})} = 2\hat{\Omega}(\tilde{p})^2 \int \left[\int K(\tau)K(\tau + \eta) d\tau\right]^2 d\eta,$$

and $\hat{\Omega}(\tilde{p})$ is a consistent estimator of the long-run variance $\Omega(\tilde{p}) = \sum_{j=-\infty}^{\infty} E[\varepsilon_t^{(\tilde{p})}\varepsilon_{t+j}^{(\tilde{p})}]$ with $\varepsilon_t^{(\tilde{p})} = Y_{t,1}^{p_{1,1}} \cdots Y_{t,d}^{p_{1,d}} Y_{t+t_2-t_1,1}^{p_{2,1}} \cdots Y_{t+t_m-t_1,d}^{p_{m,d}} - \phi_0^{(\tilde{p})}$. Thus, we could check the null hypothesis of pth-order stationarity by comparing the test statistic $\widehat{SQ}^{(\tilde{p})}$ with the one-sided $N(0,1)$ critical value z_α at significant level α and reject $\mathbb{H}_0^{(p)}$ if $\widehat{SQ}^{(\tilde{p})} > z_\alpha$ for some $\tilde{p}$. We note that the existence of $4p$th-order moment may be a sufficient but not necessary condition. However, the existence of the pth-order moment is necessary, because if the pth moment does not exist, it is meaningless to estimate this moment and construct the test statistic $\widehat{SQ}^{(p)}$. In addition, one can also consider the following compositive test statistic for pth-order stationarity:

$$\hat{Q}(p) = \sum_{\{\tilde{p}\in\mathbb{N}:\, \tilde{p}\, i_{dm} \le p\}} \hat{Q}^{(\tilde{p})}, \tag{28}$$

where i_{dm} is a dm-dimensional vector with elements all equal to 1. We could show that $\widehat{SQ}(p) = [Th^{1/2}\hat{Q}(p) - \hat{B}(p)]/\sqrt{V(p)}$ is asymptotically $N(0,1)$, with $\hat{B}(p) = \sum_{\{\tilde{p}\in\mathbb{N}:\,\tilde{p}\, i_{dm}\le p\}} \hat{B}^{(\tilde{p})}$ and $V(p)$ being the asymptotic variance of $Th^{1/2}\hat{Q}(p)$. However, due to the mutual dependence of $\hat{Q}^{(\tilde{p})}$ over different $\tilde{p}$, the expression of $V(p)$ is rather tedious, and we suggest using the bootstrap method to approximate the limiting distribution of $Th^{1/2}\hat{Q}(p)$.

We now consider a few special cases. First, let $d = 1$, $m = 1$, and $\tilde{p} = p$. Then $\widehat{SQ}^{(\tilde{p})}$ in Equation (27) tests the null hypothesis: $E(Y_t^p) = \mu_p$ for any t, i.e., the pth moment of random variable Y_t is time invariant. And the choice of $p = 1, 2, 3, 4$ yields the nonparametric tests for the time invariance property of the first four moments, respectively. More generally, by replacing μ_p with $f(\theta, t/T)$ for some known parametric function $f(\cdot,\cdot)$ but unknown parameter vector θ, we could extend our test to check whether the pth moment of Y_t follows a specific trend. By replacing $\hat{\psi}_0^{(p)} = \hat{\mu}_p$ with $f(\hat{\theta}, t/T)$, the test statistic (26) could be modified as follows:

$$\hat{Q}^{(p)} = \frac{1}{T}\sum_{t=1}^{T} \left|\hat{\phi}_t^{(p)} - f(\hat{\theta}, t/T)\right|^2, \tag{29}$$

where $\hat{\theta}$ is a $\sqrt{T}$ consistent estimator of θ. The case of (29) with $p = 1$ is Zhang and Wu's (2011) test for a parametric assumption of trend.

Next, consider $d = 2$, $m = 1$, and $\tilde{p} = (s, l)$. Then $\widehat{SQ}^{(\tilde{p})}$ in Equation (27) tests the null hypothesis: $E(Y_{t,1}^s Y_{t,2}^l) = \mu_{sl}$ for any t, i.e., the relationship between $Y_{t,1}^s$ and $Y_{t,2}^l$ is time invariant. A special case of $(l, s) = (1, 1)$ tests the time invariance relationship between $Y_{t,1}$ and $Y_{t,2}$. Suppose there exists a regression: $Y_{t,1} = \beta(t/T)Y_{t,2} + \varepsilon_t$. Then $\widehat{SQ}^{(\tilde{p})}$ with $\tilde{p} = (1, 1)$ tests whether $\beta(t/T) = \beta$.

A special case of $p = 2$ in Equation (28) yields a test statistic for weak stationarity. As mentioned in the introduction section, the existing literature provides various tests for weak stationarity, such as Dickey and Fuller (1979), Kwiatkowski et al. (1992), and Xiao and Lima (2007), among others. Compared with the existing tests, which check for a specific form of nonstationarity, our derivative tests are powerful in capturing any violation of weak stationarity, including smooth and abrupt structural changes in mean, variance, and autocovariance. Dette et al. (2011) propose a model-free test for weak stationary by measuring the L^2 distance between a time-varying spectral density estimator and the empirical spectral density. However, it can only detect local alternatives with a rate of $T^{-1/4}$, which is slower than our rate $T^{-1/2}h^{-1/4}$ given $Th \to \infty$. Preuß et al. (2013) improve the rate to g_T with $g_T T^{1/2} \to \infty$ by using the Kolmogorov–Simirnov type distance but at the expense of sacrificing asymptotic normality.

7. MONTE CARLO STUDY

We now conduct a Monte Carlo study to assess the finite sample performance of our test for strict stationarity in comparison to Kapetanios' (2009) test for strict stationarity. In addition, we also evaluate the finite sample performance of our derivative tests $\widehat{SQ}^{(1)}$, $\widehat{SQ}^{(2)}$, and $\widehat{SQ}(2)$, which examine the time invariance property of the first- and second-order moments and weak stationarity, respectively; we compare these derivative tests with the well-known Kwiatkowski et al. (1992) *KPSS* test and Preuß et al.'s (2013) test for weak stationarity.

To examine the sizes and powers of $\widehat{SQ}_W$, $\widehat{SQ}^{(1)}$, $\widehat{SQ}^{(2)}$, and $\widehat{SQ}(2)$ in finite samples, we consider the following DGPs:

$$
\begin{aligned}
&\text{DGP.S1}: Y_t = \varepsilon_t;\\
&\text{DGP.S2}: Y_t = 0.5Y_{t-1} + \varepsilon_t;\\
&\text{DGP.S3}: Y_t = \sqrt{h_t}\varepsilon_t, h_t = 0.2 + 0.3Y_{t-1}^2;\\
&\text{DGP.S4}: Y_t = \beta_t Y_{t-1} + \varepsilon_t, \beta_t = 0.5\beta_{t-1} + \eta_t, \{\eta_t\} \sim \text{i.i.d.}N(0, 0.1^2);\\
&\text{DGP.S5}: (1-0.8L)\Delta^{0.1} Y_t = \varepsilon_t;\\
&\text{DGP.S6}: Y_t = \eta_t, \{\eta_t\} \sim \text{Cauchy}(0,1);\\
&\text{DGP.S7}: Y_t = \sqrt{h_t}\varepsilon_t, h_t = 0.1 + 0.7\varepsilon_{t-1}^2 + 0.3h_{t-1};\\
&\text{DGP.P1}: Y_t = Y_{t-1} + \varepsilon_t;\\
&\text{DGP.P2}: Y_t = 0.2t + 0.5Y_{t-1} + \varepsilon_t;\\
&\text{DGP.P3}: Y_t = \varepsilon_t \mathbf{1}(t \le 0.5T) + (1+\varepsilon_t)\mathbf{1}(t > 0.5T);\\
&\text{DGP.P4}: Y_t = \varepsilon_t \mathbf{1}(t \le 0.5T) + 2\varepsilon_t \mathbf{1}(t > 0.5T);\\
&\text{DGP.P5}: Y_t = \exp(t/T - 0.5)\varepsilon_t;\\
&\text{DGP.P6}: Y_t = (1+\sqrt{2}\varepsilon_t)\mathbf{1}(t \le 0.4T) + \varepsilon_t^2 \mathbf{1}(t > 0.4T);\\
&\text{DGP.P7}: Y_t = \rho_t Y_{t-1} + \varepsilon_t, \varepsilon_t \sim N\left(0, 1-\rho_t^2\right), \rho_t = 0.5\sin(2\pi t/T);\\
&\text{DGP.P8}: Y_t = \alpha(t/T)Y_{t-1} + \varepsilon_t, \alpha(t/T) = \{1+\exp[-(t/T-0.5)/0.1]\}^{-1};\\
&\text{DGP.P9}: (1-0.5L)\Delta^{0.8} Y_t = \varepsilon_t;\\
&\text{DGP.P10}: Y_t = \sin 2\pi\left(\frac{t}{180} + \upsilon_t\right) + \phi_t, \phi_t = 0.5\phi_{t-1} + \varepsilon_t, \upsilon_t \sim U(-0.1, 0.1);
\end{aligned}
$$

where $\{\varepsilon_t\}$ is an i.i.d.$N(0,1)$ sequence except in DGP.P7. DGP.S1–S7 satisfy the null hypothesis of strict stationarity and could be used to study the size of our tests under DGPs with independent and dependent observations. Specifically, DGP.S2–S3 examine the performance of our tests under serial dependence and conditional heteroskedasticity, whereas DGP.S4–S5 are the strict stationary time-varying parameter DGP and the fractional stationary process. DGP.S6–S7 are stable and IGARCH processes, which are strict stationary but not weak stationary processes, as the first two moments do not exist. DGP.P1–P10 describe various nonstationarties and allow us to examine the power. Among them, DGP.P1–P2 are the well-known unit root and trend stationary processes. DGP.P3 has an abrupt break in mean, whereas DGP.P4 and P5 have abrupt and smooth changes in variance, respectively. DGP.P6 has an abrupt change in distribution but has the time-invariant mean and variance, and so it satisfies the weak stationarity assumption. DGP.P7 has the time-invariant marginal density, but the joint density of (Y_t, Y_{t-1}) changes smoothly because their correlation is time varying. DGP.P8–P10 are the local stationary process, the nonstationary but ergodic fractional process, and the periodic process, respectively.

For each DGP, we simulate 1000 data sets with the sample size $T = 100, 200, 500, 1000$ respectively. We use three kernels, i.e., the Uniform kernel $k(u) = \frac{1}{2}\mathbf{1}(|u| \le 1)$, the Epanechnikov kernel $k(u) = \frac{3}{4}(1-u^2)\mathbf{1}(|u| \le 1)$, and the Quartic kernel $k(u) = \frac{15}{16}(1-u^2)^2\mathbf{1}(|u| \le 1)$. To save space, we only report the results based on the Epanechnikov kernel. The results for the other two kernels are available from the authors upon request. Our simulation study shows that the choice of kernel function has little impact on the performance of our test. The number m of time points is set to 1 and 2, respectively, for each DGP. When $m = 1$, we examine the time-invariance property of the marginal density, which is similar to Kapetanios' (2009) test, whereas $m = 2$ allows us to check the time-varying joint distribution of (Y_t, Y_{t-1}), which Kapetanios' (2009) test will miss. Following Chen and Hong (2012), we use the simple rule-of-thumb bandwidth $h = (1/\sqrt{12})T^{-1/5}$, which attains the optimal rate for local smoothing estimation. We have also tried the rule-of-thumb bandwidths with different tuning parameters. Our results show that the sizes and powers of our tests are a little bit sensitive to the bandwidth by using asymptotic critical values. However, the block bootstrap described later alleviates the sensitivity of our tests to the choice of bandwidth and results in quite similar conclusions reported here. To save space, we do not report the results with different tuning parameters, which are available from the authors upon request.

Throughout our simulation study, we use product normal PDF weighting function with the extra parameter $b = 1$ for our tests of strict stationarity. To check the performance of our tests for strict stationarity with respect to b, we simulate the results for $b = 1$ and $b = 4$, respectively. The results are reported in Tables A.1 and A.2 in the Appendix. From the tables, we know our tests have reasonable sizes for both the case of $b = 1$ and $b = 4$. The power of our tests with $b = 4$ is quite similar to the case of $b = 1$ for most DGPs, showing that the influence of the extra parameter b is trivial. Besides, all our DGPs S1–S7 and P1–P10 are based on Gaussian errors. To check the performance of our tests for heavy tailed and skewed error distributions, we consider the error terms $\{\varepsilon_t\}$ that follow Student's t distribution with five degrees of freedom and the noncentral t distribution with five degrees of freedom and noncentrality parameter 1 for our DGP.S1–S2 and P1–P5. Tables A.3 and A.4 in the Appendix report the empirical reject rates for our tests of strict stationarity and weak stationarity under DGP.S1–S2 and P1–P5. The results are quite similar to those of Gaussian errors, showing that the performance of our tests is robust to the distributions of error terms.

Since our tests involve estimation of a generalized long-run variance, a lag order p_T should be selected. We follow Xiao and Lima (2007) to use Lima and Xiao's (2010) partially data-dependent lag order, which is Andrews' (1991) data-dependent plug-in lag order coupled with an upper bound:

$$p_T = \min\left\{\left[\left(\frac{3T}{2}\right)^{1/3}\left(\frac{2\hat{\rho}}{1-\hat{\rho}^2}\right)^{2/3}\right], \left[8\left(\frac{T}{100}\right)^{1/3}\right]\right\}, \tag{30}$$

where $[A]$ denotes the integer part of A, and $\hat{\rho}$ is the estimator of the first-order autocorrelation of Y_t.

Kapetanios (2009) proposes the following test for strict stationarity:

$$T_K = \max_{y\in\mathbb{Y}} \max_{t=T_0,\dots,T-1} \left\{\sqrt{Th}\left[\hat{f}_t(y) - \hat{f}_T(y)\right] \Big/ \sqrt{\hat{V}_t(y)}\right\}, \tag{31}$$

where $\mathbb{Y}$ is a set of a priori chosen grids for y, T_0 is the start period, $\hat{f}_t(y)$ is the kernel density estimator based on the first t observations $\{Y_s\}_{s=1}^t$ at y, and $\hat{V}_t(y)$ is the estimated variance of $\sqrt{Th}[\hat{f}_t(y) - \hat{f}_T(y)]$. Instead of deriving the asymptotic distribution of T_K, Kapetanios (2009) proposes using a nonparametric bootstrap approximation. We follow Kapetanios (2009) to choose the Gaussian kernel in estimating the density $\hat{f}_t(y)$, setting $T_0 = T/5$ and $\mathbb{Y}$ being 20 evenly spaced quantiles starting with the lowest decile and ending with the highest. To make Kapetanios' (2009) test and ours comparable, we also choose $h = (1/\sqrt{12})T^{-1/5}$ for Kapetanios' (2009) test.

Preuß et al.'s (2013) test is based on an estimate of a Kolmogorov–Smirnov type distance between the true time-varying spectral density and its best approximation through a stationary spectral density. As documented by the simulation study of Preuß et al. (2013), the test is robust to the number of groups N. Thus, we divide T observations into $N = 10$ groups with $M = T/N$ numbers in each group. The lag order p used in the bootstrap procedure is selected by the Akaike Information Criterion (AIC) proposed by this article with the maximum lag order $\bar{p} = 10$.

Besides asymptotic approximation, we also consider a nonparametric block bootstrap proposed by Künsch (1989) for our tests. The bootstrap kernel and resampling bandwidth are set to be the Gaussian kernel and $h_b = h$. The block $b_T = \max\{p_T, \bar{p}_T\}$, where $\bar{p}_T$ is the same as p_T given by Equation (30) except replacing $\hat{\rho}$ by the estimated first-order autocorrelation of Y_t^2. Given a set of observations $\{Y_t\}_{t=1}^T$ and the size of block b_T, we draw a bootstrap resample $\{Y_t^*\}_{t=1}^T$ as follows: (i) define the b_T-dimensional vector $X_t = (Y_t, Y_{t-1}, \ldots, Y_{t-b_T+1})$ and follow Künsch (1989) to resample with replacement from the block data $\{X_t\}_{t=1}^{T-b_T+1}$ to form pseudo-data $\{X_t^*\}_{t=1}^L$ satisfying $T = [Lb_T]$; (ii) denote the first T elements of $\{X_t^*\}_{t=1}^L$ as the bootstrap resample $\{Y_t^*\}_{t=1}^T$; (iii) repeat steps (i) and (ii) B times given each sample $\{Y_t\}_{t=1}^T$. Due to the tedious expression of the variance of $Th^{1/2}\widehat{Q}(2)$, we only report bootstrap results for our weak stationarity test. Since the bootstrap procedure is rather time consuming, we generate 500 data sets and set bootstrap replication number $B = 100$ for Kapetanios' (2009) test and ours.

We first examine strict stationarity. Table 1 reports the sizes of the tests for strict stationarity under DGP.S1–S7 at the 10% and 5% levels, using asymptotic and bootstrap critical values, respectively. Our test for strict stationarity has reasonable sizes using both asymptotic and bootstrap critical values under DGP.S1–S7. With asymptotic critical values, our test tends to underreject for DGP.S1 when $m = 1$ and overreject for DGP.S3–S4, S6–S7. The nonparametric block bootstrap can reduce underrejection and overrejection remarkably. As Kapetanios (2009) does not provide any limiting distribution for his test, we only examine its bootstrap performance. The size of Kapetanios' (2009) test is generally well behaved except for suffering from underrejection for DGP.S6. As kindly pointed out by one referee, we also consider a near unit root process. Both our test and Kapetanios' (2009) test suffer from severe overrejection under the near unit root process, and the empirical rejection rates increase as the sample size increases. That is, both our test and Kapetanios' (2009) test falsely regard the near unit root as a nonstationary process. This is apparently a drawback of our tests. We will pursue the important issue of distinguishing the near unit root process from the nonstationary process in a subsequent study.

Table 2 reports the powers of the tests for strict stationarity under DGP.P1–P10 at the 10% and 5% significance levels, using asymptotic and bootstrap critical values, respectively. Our $\widehat{SQ}_W$ tests for strict stationarity with $m = 1$ and $m = 2$, respectively, are powerful in detecting all these 10 DGPs except DGP.P7 and achieve unit power in most cases quickly as T increases. The results based on bootstrap critical values are quite similar to those based on asymptotic critical values. The power of Kapetanios' (2009) test is relative low, especially for the time-varying mean or variance processes in DGP.P3–P5 and the periodic process given by DGP.P10. This is consistent with our theoretical results that our test is asymptotically more powerful than Kapetanios' (2009) test. Moreover, it is interesting to see that both Kapetanios' (2009) test and our $\widehat{SQ}_W$ test with $m = 1$ have little power against DGP.P7, but our test with $m = 2$ is powerful against DGP.P7. This is because the marginal density of $\{Y_t\}$ under DGP.P7 follows a time invariant $N(0, 1)$, but (Y_t, Y_{t-1}) jointly follows a bivariate normal distribution with time-varying correlation ρ_t. Since the time-varying relationship is widespread in economics, it may not be a good idea to only examine the time invariance property of the marginal distribution of a time series, as does Kapetanios' (2009) test. For our test, if it cannot reject the null hypothesis with $m = 1$, we can consider $m > 1$ to test the time-varying joint distribution of a time series.

We now turn to examine weak stationarity. Table 3 reports the sizes of our derivative tests, the *KPSS* test, and Preuß et al.'s (2013) test under DGP.S1–S7 at the 10% and 5% levels,

TABLE 1
SIZE OF STRICT STATIONARITY TESTS UNDER DGP.S1–S7

		$\widehat{SQ}^{AS}_{W,m=1}$		$\widehat{SQ}^{BS}_{W,m=1}$		$\widehat{SQ}^{AS}_{W,m=2}$		$\widehat{SQ}^{BS}_{W,m=2}$		T_K	
		5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
DGP.S1	$T = 100$	0.034	0.056	0.042	0.084	0.057	0.083	0.052	0.092	0.060	0.104
	$T = 200$	0.044	0.073	0.058	0.088	0.071	0.109	0.048	0.112	0.052	0.100
	$T = 500$	0.040	0.074	0.052	0.100	0.079	0.109	0.054	0.084	0.046	0.086
	$T = 1000$	0.037	0.055	0.036	0.084	0.070	0.107	0.036	0.082	0.042	0.084
DGP.S2	$T = 100$	0.042	0.090	0.036	0.086	0.035	0.066	0.038	0.088	0.092	0.154
	$T = 200$	0.043	0.099	0.042	0.080	0.039	0.079	0.036	0.086	0.076	0.144
	$T = 500$	0.068	0.117	0.054	0.106	0.050	0.089	0.048	0.098	0.078	0.150
	$T = 1000$	0.058	0.098	0.044	0.112	0.047	0.083	0.038	0.086	0.076	0.134
DGP.S3	$T = 100$	0.088	0.132	0.054	0.124	0.078	0.138	0.056	0.130	0.074	0.122
	$T = 200$	0.093	0.145	0.058	0.116	0.121	0.167	0.064	0.122	0.050	0.116
	$T = 500$	0.088	0.144	0.058	0.104	0.130	0.187	0.054	0.116	0.052	0.106
	$T = 1000$	0.088	0.140	0.050	0.094	0.156	0.229	0.050	0.100	0.048	0.084
DGP.S4	$T = 100$	0.042	0.064	0.066	0.138	0.077	0.108	0.064	0.114	0.072	0.130
	$T = 200$	0.050	0.090	0.082	0.136	0.076	0.110	0.060	0.114	0.064	0.104
	$T = 500$	0.039	0.065	0.050	0.104	0.093	0.130	0.054	0.108	0.056	0.096
	$T = 1000$	0.036	0.070	0.042	0.112	0.090	0.136	0.048	0.102	0.054	0.112
DGP.S5	$T = 100$	0.010	0.064	0.092	0.168	0.010	0.052	0.100	0.188	0.076	0.140
	$T = 200$	0.051	0.088	0.066	0.158	0.039	0.088	0.086	0.160	0.080	0.140
	$T = 500$	0.076	0.147	0.074	0.156	0.078	0.138	0.090	0.168	0.106	0.156
	$T = 1000$	0.065	0.144	0.060	0.134	0.083	0.115	0.084	0.144	0.064	0.114
DGP.S6	$T = 100$	0.040	0.072	0.066	0.144	0.206	0.274	0.100	0.170	0.012	0.024
	$T = 200$	0.077	0.126	0.068	0.114	0.285	0.374	0.064	0.126	0.014	0.024
	$T = 500$	0.092	0.160	0.074	0.106	0.361	0.434	0.070	0.120	0.004	0.010
	$T = 1000$	0.094	0.154	0.068	0.112	0.374	0.471	0.068	0.134	0.006	0.020
DGP.S7	$T = 100$	0.134	0.212	0.092	0.174	0.216	0.316	0.086	0.176	0.060	0.124
	$T = 200$	0.151	0.236	0.074	0.138	0.263	0.334	0.082	0.144	0.086	0.138
	$T = 500$	0.213	0.300	0.066	0.128	0.337	0.442	0.076	0.138	0.050	0.116
	$T = 1000$	0.210	0.301	0.056	0.116	0.362	0.469	0.062	0.114	0.058	0.116

NOTES: (i) $\widehat{SQ}^{AS}_{W}$ and $\widehat{SQ}^{BS}_{W}$ denote the results of $\widehat{SQ}_{W}$ using asymptotic and bootstrap critical values, respectively; (ii) T_K denotes Kapetanios' (2009) test; (iii) the results using asymptotic critical values are based on 1000 iterations, whereas the bootstrap results are based on 500 iterations.

using asymptotic and bootstrap critical values, respectively. Note that DGP.S6–S7 are weakly nonstationary processes. More specifically, any moment of DGP.S6 does not exist while the second-order moment of DGP.S7 does not exist. Now, we consider the results for DGP.S1–S5. We observe that the empirical sizes of $\widehat{SQ}^{(1)}$ are close to nominal sizes using both asymptotic and bootstrap critical values for DGP.S1–S4, whereas the results of $\widehat{SQ}^{(1)}$ for DGP.S5 suffer from overrejection slightly. The $\widehat{SQ}^{(2)}$ test using asymptotic critical values suffers from slight underrejection for DGP.S1 and a bit severe overrejection for DGP.S3, but the bootstrap procedure delivers reasonable sizes. The size of our weak stationarity test $\widehat{SQ}(2)$ is close to the nominal level for DGP.S1–S4 and suffers from overrejection slightly for DGP.S5. From the results of DGP.S6–S7, we observe that our test for weak stationarity is powerless for these two DGPs, which is consistent with our theory. The *KPSS* test has reasonable size except for suffering from a bit of overrejection for DGP.S2 and DGP.S4–S5, and it is also powerless in detecting DGP.S6–S7. In contrast, the Preuß et al.'s (2013) test only has reasonable size for DGP.S1–S2 and suffers from severe overrejection for DGP.S3–S5. However, Preuß et al.'s (2013) test is powerful in capturing DGP.S6–S7.

Table 4 reports the powers of the *KPSS* test, Preuß et al.'s (2013) test, and our derivative tests $\widehat{SQ}^{(i)}$, $i = 1, 2$ and $\widehat{SQ}(2)$ under DGP.P1-P10 at the 10% and 5% significance levels, using

TABLE 2
POWER OF STRICT STATIONARITY TESTS UNDER DGP.P1–P10

		$\widehat{SQ}^{AS}_{W,m=1}$		$\widehat{SQ}^{BS}_{W,m=1}$		$\widehat{SQ}^{AS}_{W,m=2}$		$\widehat{SQ}^{BS}_{W,m=2}$		T_K	
		5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
DGP.P1	$T = 100$	0.821	0.908	0.850	0.916	0.798	0.889	0.830	0.894	0.724	0.814
	$T = 200$	0.985	0.991	0.970	0.992	0.982	0.990	0.962	0.982	0.888	0.932
	$T = 500$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.990	0.994
	$T = 1000$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
DGP.P2	$T = 100$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	$T = 200$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	$T = 500$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	$T = 1000$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
DGP.P3	$T = 100$	0.941	0.964	0.918	0.970	0.948	0.970	0.868	0.946	0.146	0.232
	$T = 200$	1.00	1.00	0.998	0.998	1.00	1.00	1.00	1.00	0.208	0.336
	$T = 500$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.444	0.652
	$T = 1000$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.790	0.914
DGP.P4	$T = 100$	0.440	0.569	0.500	0.676	0.754	0.851	0.638	0.804	0.122	0.194
	$T = 200$	0.921	0.960	0.934	0.970	0.989	0.996	0.966	0.986	0.134	0.220
	$T = 500$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.228	0.370
	$T = 1000$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.444	0.564
DGP.P5	$T = 100$	0.297	0.431	0.362	0.528	0.567	0.659	0.470	0.604	0.124	0.194
	$T = 200$	0.739	0.816	0.762	0.884	0.925	0.950	0.844	0.920	0.128	0.200
	$T = 500$	0.999	0.999	1.00	1.00	1.00	1.00	1.00	1.00	0.168	0.278
	$T = 1000$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.364	0.474
DGP.P6	$T = 100$	0.243	0.313	0.256	0.400	0.359	0.456	0.210	0.330	0.222	0.362
	$T = 200$	0.492	0.578	0.500	0.632	0.627	0.703	0.426	0.556	0.434	0.602
	$T = 500$	0.899	0.940	0.916	0.956	0.960	0.972	0.866	0.924	0.788	0.888
	$T = 1000$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.956	0.982
DGP.P7	$T = 100$	0.118	0.175	0.112	0.184	0.256	0.338	0.172	0.258	0.076	0.134
	$T = 200$	0.141	0.194	0.128	0.212	0.396	0.498	0.262	0.398	0.064	0.138
	$T = 500$	0.175	0.246	0.166	0.240	0.818	0.883	0.680	0.792	0.066	0.104
	$T = 1000$	0.208	0.286	0.228	0.308	0.994	0.997	0.952	0.986	0.062	0.120
DGP.P8	$T = 100$	0.548	0.676	0.510	0.658	0.506	0.661	0.532	0.692	0.230	0.326
	$T = 200$	0.817	0.902	0.778	0.854	0.835	0.912	0.810	0.896	0.310	0.426
	$T = 500$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.666	0.784
	$T = 1000$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.906	0.946
DGP.P9	$T = 100$	0.844	0.925	0.498	0.684	0.761	0.878	0.490	0.678	0.488	0.578
	$T = 200$	0.967	0.992	0.872	0.928	0.956	0.972	0.866	0.902	0.658	0.744
	$T = 500$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.824	0.896
	$T = 1000$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.870	0.934
DGP.P10	$T = 100$	0.898	0.954	0.872	0.954	0.815	0.894	0.772	0.892	0.152	0.248
	$T = 200$	1.00	1.00	1.00	1.00	0.998	1.00	0.998	1.00	0.326	0.476
	$T = 500$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.084	0.140
	$T = 1000$	0.871	0.934	0.744	0.852	0.837	0.894	0.648	0.734	0.092	0.144

NOTES: (i) $\widehat{SQ}^{AS}_W$ and $\widehat{SQ}^{BS}_W$ denote the results of $\widehat{SQ}_W$ using asymptotic and bootstrap critical values, respectively; (ii) T_K denotes Kapetanios' (2009) test; (iii) the results using asymptotic critical values are based on 1000 iterations, whereas the bootstrap results are based on 500 iterations.

asymptotic and bootstrap critical values, respectively. We observe that $\widehat{SQ}^{(1)}$ is powerful in capturing changes in mean given by DGP.P1–P3 and DGP.P8–P10, and it is robust to changes in higher order moments in DGP.P4–P7. Similarly, $\widehat{SQ}^{(2)}$ is powerful in capturing changes in variance given by DGP.P1–P2, DGP.P4–P5, and DGP.P8–P9, and it is robust to changes in other moments. Interestingly, both of our derivative tests for time-varying mean and variance are powerful against the unit root and trend stationary processes given by DGP.P1–P2. Moreover,

TABLE 3

SIZE OF WEAK STATIONARITY TESTS UNDER DGP.S1–S7

		$\widehat{SQ}^{(1)}_{AS}$		$\widehat{SQ}^{(1)}_{BS}$		$\widehat{SQ}^{(2)}_{AS}$		$\widehat{SQ}^{(2)}_{BS}$		$\widehat{SQ}(2)$		*KPSS*		*PVD*	
		5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
DGP.S1	$T = 100$	0.044	0.067	0.060	0.110	0.023	0.053	0.068	0.112	0.060	0.118	0.045	0.095	0.048	0.096
	$T = 200$	0.048	0.078	0.064	0.120	0.033	0.062	0.050	0.098	0.050	0.100	0.057	0.098	0.036	0.084
	$T = 500$	0.048	0.080	0.040	0.092	0.037	0.062	0.044	0.110	0.058	0.098	0.057	0.103	0.056	0.098
	$T = 1000$	0.043	0.074	0.040	0.088	0.048	0.085	0.046	0.122	0.048	0.104	0.051	0.095	0.066	0.114
DGP.S2	$T = 100$	0.051	0.105	0.054	0.110	0.015	0.033	0.050	0.096	0.066	0.138	0.082	0.172	0.046	0.072
	$T = 200$	0.041	0.102	0.040	0.102	0.028	0.061	0.058	0.098	0.064	0.102	0.074	0.151	0.056	0.108
	$T = 500$	0.067	0.103	0.042	0.110	0.035	0.069	0.046	0.098	0.042	0.114	0.064	0.135	0.062	0.110
	$T = 1000$	0.062	0.116	0.036	0.088	0.032	0.083	0.048	0.102	0.056	0.102	0.069	0.124	0.048	0.098
DGP.S3	$T = 100$	0.036	0.067	0.044	0.094	0.183	0.252	0.088	0.156	0.072	0.140	0.055	0.101	0.146	0.218
	$T = 200$	0.043	0.075	0.038	0.092	0.238	0.323	0.068	0.154	0.066	0.132	0.063	0.116	0.168	0.250
	$T = 500$	0.050	0.070	0.040	0.104	0.269	0.362	0.066	0.152	0.068	0.142	0.050	0.093	0.202	0.270
	$T = 1000$	0.040	0.071	0.048	0.086	0.340	0.423	0.060	0.126	0.056	0.120	0.043	0.089	0.252	0.334
DGP.S4	$T = 100$	0.042	0.069	0.056	0.116	0.042	0.078	0.068	0.118	0.052	0.110	0.051	0.100	0.052	0.094
	$T = 200$	0.047	0.090	0.060	0.134	0.050	0.087	0.078	0.126	0.042	0.112	0.068	0.118	0.054	0.100
	$T = 500$	0.045	0.066	0.052	0.094	0.053	0.090	0.070	0.120	0.058	0.130	0.047	0.103	0.048	0.106
	$T = 1000$	0.046	0.072	0.046	0.096	0.064	0.092	0.082	0.126	0.060	0.126	0.047	0.097	0.064	0.122
DGP.S5	$T = 100$	0.000	0.000	0.046	0.106	0.014	0.044	0.060	0.138	0.090	0.204	0.033	0.155	0.222	0.278
	$T = 200$	0.011	0.023	0.114	0.204	0.031	0.078	0.054	0.120	0.102	0.180	0.051	0.127	0.280	0.342
	$T = 500$	0.046	0.094	0.138	0.224	0.070	0.124	0.058	0.146	0.088	0.176	0.074	0.173	0.314	0.388
	$T = 1000$	0.121	0.180	0.200	0.308	0.057	0.122	0.048	0.118	0.062	0.162	0.087	0.165	0.304	0.402
DGP.S6	$T = 100$	0.018	0.038	0.046	0.084	0.014	0.024	0.034	0.084	0.042	0.078	0.022	0.068	0.864	0.896
	$T = 200$	0.043	0.055	0.052	0.104	0.023	0.048	0.060	0.106	0.058	0.112	0.032	0.090	0.950	0.960
	$T = 500$	0.031	0.050	0.044	0.074	0.042	0.061	0.056	0.100	0.062	0.094	0.024	0.071	0.990	0.990
	$T = 1000$	0.024	0.046	0.034	0.066	0.027	0.048	0.046	0.072	0.042	0.086	0.029	0.068	0.992	0.992
DGP.S7	$T = 100$	0.036	0.064	0.044	0.086	0.230	0.314	0.090	0.162	0.070	0.180	0.046	0.109	0.340	0.456
	$T = 200$	0.050	0.067	0.032	0.086	0.279	0.354	0.084	0.168	0.074	0.146	0.057	0.101	0.474	0.596
	$T = 500$	0.051	0.070	0.044	0.088	0.326	0.411	0.046	0.118	0.046	0.132	0.062	0.111	0.550	0.658
	$T = 1000$	0.048	0.083	0.048	0.094	0.363	0.442	0.056	0.120	0.050	0.134	0.050	0.098	0.618	0.728

NOTES: (i) $\widehat{SQ}^{(i)}_{AS}$ and $\widehat{SQ}^{(i)}_{BS}$ denote the results of $\widehat{SQ}^{(i)}$ using asymptotic and bootstrap critical values, respectively; (ii) $\widehat{SQ}(2)$ denotes the results of our weakly stationarity test using bootstrap critical values; (iii) *KPSS* denotes Kwiatkowski et al.'s (1992) test; (iv) *PVD* denotes Preuß et al.'s (2013) test for weak stationarity; and (v) the results using asymptotic critical values are based on 1000 iterations, whereas the bootstrap results are based on 500 iterations.

TABLE 4
POWER OF WEAK STATIONARITY TESTS UNDER DGP.P1–P10

		$\widehat{SQ}_{AS}^{(1)}$		$\widehat{SQ}_{BS}^{(1)}$		$\widehat{SQ}_{AS}^{(2)}$		$\widehat{SQ}_{BS}^{(2)}$		$\widehat{SQ}(2)$		$KPSS$		PVD	
		5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
DGP.P1	$T = 100$	0.789	0.896	0.814	0.856	0.248	0.430	0.300	0.392	0.552	0.772	0.569	0.718	0.124	0.134
	$T = 200$	0.984	0.990	0.934	0.972	0.720	0.807	0.548	0.650	0.832	0.936	0.801	0.880	0.118	0.136
	$T = 500$	1.00	1.00	0.990	0.994	0.967	0.976	0.886	0.926	0.990	0.996	0.950	0.978	0.106	0.120
	$T = 1000$	1.00	1.00	1.00	1.00	0.994	0.998	0.960	0.980	1.00	1.00	0.992	0.999	0.124	0.132
DGP.P2	$T = 100$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.037	0.103	0.046	0.076
	$T = 200$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.042	0.097	0.280	0.378
	$T = 500$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.034	0.093	0.526	0.608
	$T = 1000$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.054	0.097	0.604	0.720
DGP.P3	$T = 100$	0.967	0.981	0.946	0.978	0.019	0.036	0.042	0.078	0.546	0.742	0.142	0.297	0.022	0.048
	$T = 200$	1.00	1.00	1.00	1.00	0.016	0.038	0.048	0.102	0.958	0.988	0.382	0.604	0.012	0.032
	$T = 500$	1.00	1.00	1.00	1.00	0.023	0.043	0.046	0.098	1.00	1.00	0.925	0.973	0.000	0.004
	$T = 1000$	1.00	1.00	1.00	1.00	0.036	0.059	0.042	0.092	1.00	1.00	0.999	0.999	0.000	0.000
DGP.P4	$T = 100$	0.055	0.087	0.076	0.134	0.875	0.972	0.874	0.946	0.774	0.902	0.055	0.101	0.918	0.942
	$T = 200$	0.055	0.088	0.080	0.126	1.00	1.00	1.00	1.00	0.998	0.100	0.065	0.113	0.998	1.00
	$T = 500$	0.066	0.105	0.068	0.124	1.00	1.00	1.00	1.00	1.00	1.00	0.060	0.124	1.00	1.00
	$T = 1000$	0.059	0.078	0.060	0.090	1.00	1.00	1.00	1.00	1.00	1.00	0.065	0.114	1.00	1.00
DGP.P5	$T = 100$	0.059	0.093	0.078	0.134	0.798	0.855	0.828	0.898	0.728	0.848	0.053	0.100	0.796	0.876
	$T = 200$	0.069	0.103	0.094	0.156	0.995	0.996	0.996	0.996	0.986	0.994	0.066	0.113	0.988	0.998
	$T = 500$	0.075	0.119	0.070	0.116	1.00	1.00	1.00	1.00	1.00	1.00	0.071	0.124	1.00	1.00
	$T = 1000$	0.071	0.090	0.066	0.114	1.00	1.00	1.00	1.00	1.00	1.00	0.067	0.103	1.00	1.00
DGP.P6	$T = 100$	0.028	0.041	0.032	0.074	0.054	0.083	0.098	0.174	0.040	0.104	0.035	0.082	0.156	0.212
	$T = 200$	0.041	0.059	0.064	0.126	0.057	0.083	0.094	0.144	0.052	0.102	0.049	0.100	0.102	0.168
	$T = 500$	0.053	0.087	0.070	0.124	0.063	0.104	0.106	0.162	0.048	0.118	0.049	0.097	0.052	0.100
	$T = 1000$	0.053	0.108	0.070	0.108	0.078	0.105	0.074	0.128	0.046	0.084	0.057	0.109	0.018	0.032
DGP.P7	$T = 100$	0.109	0.156	0.126	0.192	0.069	0.116	0.074	0.142	0.424	0.578	0.087	0.160	0.168	0.258
	$T = 200$	0.138	0.199	0.166	0.252	0.086	0.139	0.082	0.166	0.786	0.878	0.108	0.186	0.256	0.418
	$T = 500$	0.170	0.226	0.158	0.248	0.125	0.178	0.100	0.194	0.990	0.998	0.105	0.175	0.674	0.838
	$T = 1000$	0.222	0.282	0.216	0.290	0.144	0.208	0.136	0.214	1.00	1.00	0.111	0.187	0.760	0.892

(*Continued*)

TABLE 4
CONTINUED

		$\widehat{SQ}_{AS}^{(1)}$		$\widehat{SQ}_{BS}^{(1)}$		$\widehat{SQ}_{AS}^{(2)}$		$\widehat{SQ}_{BS}^{(2)}$		$\widehat{SQ}(2)$		*KPSS*		*PVD*	
		5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
DGP.P8	$T = 100$	0.523	0.648	0.620	0.710	0.332	0.463	0.448	0.684	0.588	0.818	0.277	0.496	0.122	0.188
	$T = 200$	0.724	0.791	0.712	0.772	0.704	0.798	0.728	0.840	0.872	0.964	0.425	0.561	0.602	0.690
	$T = 500$	0.840	0.870	0.782	0.832	0.989	0.994	0.974	0.992	0.998	1.00	0.503	0.594	0.960	0.964
	$T = 1000$	0.881	0.899	0.818	0.856	0.540	0.629	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
DGP.P9	$T = 100$	0.814	0.916	0.850	0.902	0.168	0.292	0.258	0.392	0.530	0.766	0.257	0.467	0.022	0.058
	$T = 200$	0.963	0.990	0.928	0.962	0.466	0.626	0.428	0.562	0.860	0.930	0.545	0.692	0.066	0.114
	$T = 500$	0.998	0.998	0.996	0.998	0.853	0.888	0.746	0.814	0.986	0.992	0.721	0.844	0.284	0.358
	$T = 1000$	1.00	1.00	1.00	1.00	0.930	0.966	0.800	0.872	1.00	1.00	0.595	0.745	0.548	0.636
DGP.P10	$T = 100$	0.974	0.996	0.990	0.996	0.122	0.180	0.202	0.296	0.760	0.910	0.295	0.465	0.016	0.036
	$T = 200$	1.00	1.00	1.00	1.00	0.069	0.161	0.116	0.236	0.988	0.998	1.00	1.00	0.018	0.046
	$T = 500$	1.00	1.00	1.00	1.00	0.024	0.062	0.026	0.062	1.00	1.00	0.035	0.354	0.004	0.038
	$T = 1000$	1.00	1.00	0.690	0.968	0.008	0.029	0.012	0.026	0.342	0.584	0.000	0.000	0.172	0.282

NOTES: (i) $\widehat{SQ}_{AS}^{(i)}$ and $\widehat{SQ}_{BS}^{(i)}$ denote the results of $\widehat{SQ}^{(i)}$ using asymptotic and bootstrap critical values respectively; (ii) $\widehat{SQ}(2)$ denotes the results of our weakly stationarity test using bootstrap critical values; (iii) *KPSS* denotes Kwiatkowski et al.'s (1992) test; (iv) *PVD* denotes Preuß et al.'s (2013) test for weak stationarity; (v) the results using asymptotic critical values are based on 1000 iterations, whereas the bootstrap results are based on 500 iterations.

our weak stationarity test $\widehat{SQ}(2)$ is powerful in capturing various violations of weak stationarity given by DGP.P1–P5 and P7–P10 and is robust to DGP.P6, which is weakly stationary but not strictly stationary. The *KPSS* test is only powerful in detecting the unit root process given by DGP.P1, the abrupt change in mean given by DGP.P3, the local stationary process given by DGP.P8, and the nonstationary but ergodic process given by DGP.P9. It is not difficult to understand the powerfulness of the *KPSS* test in capturing the unit root process, as it tests the null hypothesis of weak stationarity against a unit root process. However, if there exists abrupt change in mean, as shown in DGP.P3, the *KPSS* test may falsely regard it as a unit root process. In contrast, our derivative test $\widehat{SQ}^{(1)}$ can identify structural breaks in mean and does not suffer from making such a mistake. In comparison, the power of Preuß et al.'s (2013) test is relatively lower than our weak stationary test for most cases. Specifically, Preuß et al.'s (2013) test has little power against the nonstationary processes DGP.P1, P3, and P10, which are a unit root process, an abrupt break in mean process, and a periodic process, respectively, whereas our test is powerful to detect these three forms of nonstatinarity.

8. ARE DIFFERENCED MACROECONOMIC TIME SERIES STATIONARY?

Although the assumption of strict stationarity plays a fundamental role in econometric modeling, inference, and forecasting, the existing literature either accepts the strict stationarity assumption directly without using any formal test or relies on the results of such popular tests as *ADF* and *KPSS*. However, as shown in our simulation studies, the results of the *ADF* and *KPSS* tests need to be interpreted with caution. For example, they may mistake a nonstationary time series with changes in mean as a unit root process. In contrast, our tests can avoid such a mistake. In this section, we will use our tests to examine the stationarity assumption for various macroeconomic time series.

In this study, we mainly focus on two fundamental questions: (1) Are macroeconomic time series stationary? (2) If the level series are nonstationary, then are the differenced (differenced or log-differenced) series stationary? To answer these questions, we check stationarity for 14 representative time series, including output (nominal GDP, real GDP, nominal GNP, real GNP, industrial production), price (CPI, GDP, price deflator), money (M1, M2), consumption (real personal consumption), employment (unemployment rate), income (real per capita income), and interest rate (10-year Treasury Constant Maturity Rate, Effective Federal Funds Rate). Besides, we also test stationarity for the other 50 series. The results are reported in the Appendix. The data we used are U.S. monthly or quarterly data. All the data are collected from the website of the Federal Reserve Bank of St. Louis, spanning from the earliest available sample to 2014M3/2014Q1 or the most up-to-date data until May 2014. Most of the data have been seasonally adjusted.

In addition to our $\widehat{SQ}_W$ test for strict stationarity, we also consider some other popular tests, namely, Dickey and Fuller's (1979, 1981) *ADF* test, Kwiatkowski et al.'s (1992) *KPSS* test, Kapetanios' (2009) strict stationarity test, our derivative tests $\widehat{SQ}^{(i)}$, $i = 1, 2$, for time-invariant properties of the first two moments, and our weak stationary test $\widehat{SQ}(2)$. The first two of them are well-known tests for weak stationarity versus a unit root process. All the data have been standardized to have zero mean and unit variance before applying these tests. The smoothing parameters, kernel functions, and other presettings for the tests of Kapetanios (2009) and our tests $\widehat{SQ}_W$, $\widehat{SQ}^{(i)}$, $i = 1, 2$, and $\widehat{SQ}(2)$ are all the same as those used in the simulation study.

We first check stationarity for the level series or the log transformed series. The results of all the tests show that all time series and their log transformed series examined except the unemployment rate are nonstationary. It is easy to understand the results for output, price, money, consumption, and income, as these series show significant upward trends. The results of the nominal interest rates are consistent with the nonstationarity argument in the literature (Fama, 1975, 1976, 1977; Fama and Gibbons, 1982; Mankiw and Miron, 1986). For space, we do not report the results for these 13 level series with the same conclusions for all tests, which are available from the authors upon request. The last row in Table 5 reports the results for the unemployment rate. Although the *ADF* test rejects the null of unit root, the *KPSS* test and the tests of Kapetanios (2009) and ours for strict stationarity all document evidence of nonstationarity of the unemployment rate at the 5% level. Furthermore, our derivative tests find a time-varying mean (but not time-varying variance) for the unemployment rate. Since a unit root process has both time-varying mean and variance, the results of our derivative tests suggest that the unemployment rate is a process with structural changes in mean instead of a unit root process. Montgomery et al. (1998) compare forecasting performance for various time series models by assuming the unit root property of the U.S. unemployment rate. Given our evidence that the unemployment rate is not a unit root process, there may exist room for improving forecasts of the unemployment rate using time series models that do not assume a unit root process.

We now check stationarity for the corresponding differenced or log-differenced series of 13 representative time series. Specifically, we take log differences for output, price, money, consumption, and income to obtain growth rate series, which are widely used in econometric modeling. Since the Federal Reserve Broad (Fed) usually adjusts its target interest rate by a multiple of 25 basis points, not by a certain percentage of the interest level, the related literature (Bae and de Jone, 2007) usually assumes their difference instead of log difference to be a stationary process. Hence, we test stationarity for the difference of interest rates.

Table 5 reports the results of tests for the differenced series of 13 representative variables. At the 5% level, our $\widehat{SQ}$ test rejects the null of strict stationarity for all the differenced series of 13 representative variables, using both asymptotic and bootstrap critical values. Our $\widehat{SQ}(2)$ also rejects the null of weak stationarity for all these 13 variables at the 5% level. Furthermore, our derivative test $\widehat{SQ}^{(1)}$ documents the structural changes in mean for the growth rates of nominal output, consumption, prices, and money supply, whereas the derivative test $\widehat{SQ}^{(2)}$ reveals significant evidence of a time-varying variance for almost all the variables except for the growth rate of prices and the broad money supply M2. This finding is consistent with Kim and Nelson (1999), McConnell and Perez-Quiros (2000), Stock and Watson (2009), and Breitung and Eickmeier's (2011) conclusion of "Great Moderation," which means that the volatility of many U.S. variables including output and consumption have declined substantially since the 1980s. According to the results of our strict stationarity test and the derivative tests, all the 13 differenced series are neither strictly nor weakly stationary. Stock and Watson (1996) check parameter stability of univariate linear autoregressions for 76 U.S. series by using various stability tests and the results evidently depend on the used stability tests. Unlike the tests used by Stock and Watson (1996), all of our tests are model free, and they check the stability of distribution, mean, and variance of the underlying time series directly.

For comparison, Kapetanios' (2009) test cannot reject the null of strict stationarity for most of the differenced series. This is consistent with the results of our simulation studies that Kapetanios' (2009) test has low power to detect structural changes in mean and variance. Moreover, the *ADF* test rejects the null hypothesis of unit root for all these differenced series and accepts weak stationarity. The results of the *KPSS* test are mixed, which rejects the null of weak stationarity for seven variables and cannot reject the null of weak stationarity for others. The failure of the *ADF* and *KPSS* tests in capturing nonstationarity and the contradictory results of these two tests may be due to the existence of structural changes in mean and/or variance, which is neither covered by the null nor included by the alternative of the *ADF* and *KPSS* tests.

The new empirical findings in this section have important implications. In particular, our results suggest that the usual practice of time series econometric modeling, namely, modeling nonstationary time series by using unit root/cointegration models with constant model parameters and weakly stationary shocks or by taking first differences and then considering a weakly stationary model for the differenced series, may have room to be improved by accommodating the documented time-varying features of macroeconomic time series. For

TABLE 5
STATIONARITY TESTS FOR MACROECONOMIC TIME SERIES

	Tran.	Period	$\widehat{SQ}_W^{AS}$	$\widehat{SQ}_W^{BS}$	$\widehat{SQ}_{AS}^{(1)}$	$\widehat{SQ}_{BS}^{(1)}$	$\widehat{SQ}_{AS}^{(2)}$	$\widehat{SQ}_{BS}^{(2)}$	$\widehat{SQ}(2)$	T_K	$KPSS$	ADF
Nominal GDP	3	1947Q1–2014Q1	0.000	0.000	0.000	0.000	0.004	0.015	0.010	0.360	< 0.010	< 0.001
Nominal GNP	3	1947Q1–2013Q4	0.000	0.000	0.000	0.000	0.006	0.025	0.015	0.430	< 0.010	< 0.001
Real GDP	3	1947Q1–2014Q1	0.008	0.045	0.523	0.500	0.000	0.000	0.040	0.190	> 0.100	< 0.001
Real GNP	3	1947Q1–2013Q4	0.014	0.035	0.585	0.635	0.001	0.020	0.045	0.260	> 0.100	< 0.001
Industrial Production	3	1919M1–2014M3	0.000	0.000	0.914	0.980	0.000	0.000	0.000	0.005	> 0.100	< 0.001
Real Personal Consumption	3	1947Q1–2014Q1	0.000	0.000	0.038	0.045	0.004	0.010	0.035	0.355	> 0.100	< 0.001
Real Per Capita Income	3	1947Q1–2014Q1	0.001	0.010	0.280	0.230	0.004	0.010	0.015	0.285	> 0.100	< 0.001
CPI	3	1955M1–2013M12	0.000	0.000	0.000	0.000	0.050	0.115	0.010	0.000	< 0.010	0.006
GDP Price Deflator	3	1947Q1–2014Q1	0.000	0.005	0.000	0.000	0.105	0.150	0.050	0.015	< 0.010	0.043
M1	3	1959M1–2014M3	0.000	0.000	0.000	0.000	0.000	0.005	0.010	0.305	< 0.010	< 0.001
M2	3	1959M1–2014M3	0.000	0.000	0.000	0.000	0.486	0.415	0.040	0.015	< 0.010	< 0.001
10-Year Treasury Maturity Rate	2	1953M4–2014M3	0.000	0.000	0.731	0.720	0.000	0.000	0.000	0.000	> 0.100	< 0.001
Effective Federal Funds Rate	2	1954M7–2014M3	0.000	0.015	0.927	0.955	0.000	0.015	0.000	0.720	> 0.100	< 0.001
Unemployment Rate	1	1948M1–2014M4	0.000	0.000	0.000	0.000	0.226	0.470	0.020	0.000	< 0.010	0.003

NOTES: (i) Numbers in the main entries are the p-values; (ii) the data transformation codes are: 1. level of the series; 2. first difference; 3. first difference of the logarithm; (iii) $\widehat{SQ}_W^{AS}$ and $\widehat{SQ}_W^{BS}$ denote the results of the $\widehat{SQ}_W$ test using asymptotic and bootstrap critical values, respectively; T_K, $KPSS$, and ADF denote Kapetanios' (2009) test, Kwiatkowski et al.'s (1992) test, and Dickey and Fuller's (1979) test, respectively; $\widehat{SQ}_{AS}^{(i)}$ and $\widehat{SQ}_{BS}^{(i)}$ denote the results of the $\widehat{SQ}^{(i)}$ test using asymptotic and bootstrap critical values, respectively; and (iv) p-values of the bootstrap results are based on 200 bootstrap iterations.

example, a locally stationary time series with time-varying parameters (e.g., Dahlhaus, 1996) may be considered for the differenced series. This may provide more reliable inference and forecasting.

9. CONCLUDING REMARKS

Strict stationarity is one of the most fundamental assumptions in time series econometrics, especially in nonlinear and nonparametric modeling, testing, and forecasting. This article contributes to the existing literature by proposing a model-free test for strict stationarity of a possibly vector-valued time series. The idea is to nonparametrically estimate the time-varying characteristic function and compare it with the empirical characteristic function based on the entire sample via a weighted distance measure. By employing a multivariate standard normal weighting function, we can avoid intractable high dimensional numerical integration. Furthermore, by taking the appropriate order of partial derivatives, we can gauge possible sources of nonstationarity and construct tests for weak stationarity and, more generally, the pth-order stationarity. All of the proposed tests have a convenient null asymptotic one-sided $N(0,1)$ distribution. Monte Carlo studies show that our tests have reasonable size and excellent power in detecting various nonstationarties, including abrupt and smooth structural changes as well as unit root and trend stationarity. In an empirical application to macroeconomic time series, we find significant evidence against strict and weak stationarity for both level and first differenced series. This suggests that the usual practice of modeling nonstationary time series by using unit root and cointegration models with constant model parameters and stationary shocks or by taking the first-order difference and then considering a stationary model (e.g., VAR, autoregressive threshold, Markov regime switching) for the differenced series may fail to accommodate time-varying features of macroeconomic time series. New time series models that can accommodate the stylized facts documented in this article, such as locally stationary time series models with time-varying parameters, may be pursued to obtain more reliable inference and forecasting.

APPENDIX

This appendix contains three parts. Section A provides the detailed proofs of Theorems 1–3, Section B reports the simulation results that are not reported in our article, and Section C reports the empirical results which are not reported in our article.

A. *Mathematical Proofs of Theorems 1–3.* Throughout the appendix, $C \in (0,\infty)$ is a generic bounded constant that may vary from case to case, $\xi_t = (X_t', t/T)'$, A^* denotes the conjugate of A, and $Re(A)$ denotes the real part of A.

PROOF OF THEOREM 1. Under $\mathbb{H}_0 : \phi_t(u) = \psi_0(u)$, we can decompose $Th^{1/2}\hat{Q}$ as follows:

$$
\begin{aligned}
Th^{1/2}\hat{Q} &= h^{1/2}\sum_{t=1}^{T}\int \left|\hat{\phi}_t(u) - \hat{\psi}_0(u)\right|^2 W(u)du \\
&= h^{1/2}\sum_{t=1}^{T}\int \left|\left[\hat{\phi}_t(u) - \phi_t(u)\right] - \left[\hat{\psi}_0(u) - \psi_0(u)\right]\right|^2 W(u)du \\
&= h^{1/2}\sum_{t=1}^{T}\int \left|\hat{\phi}_t(u) - \phi_t(u)\right|^2 W(u)du + h^{1/2}\sum_{t=1}^{T}\int \left|\hat{\psi}_0(u) - \psi_0(u)\right|^2 W(u)du \\
&\quad -2h^{1/2}\sum_{t=1}^{T}\int Re\left[\left(\hat{\phi}_t(u) - \phi_t(u)\right)\left(\hat{\psi}_0(u) - \psi_0(u)\right)^*\right] W(u)du \\
&= Q_1 + Q_2 - 2Q_3, \quad \text{say.}
\end{aligned}
$$

To show Theorem 1, it is sufficient to show Propositions A.1 to A.5 as follows.

PROPOSITION A.1. *Under the condition of Theorem 1, $Q_1 = B_1 + \tilde{U} + o_P(1)$, where*

$$B_1 = h^{-1/2} \int E[|\varepsilon_s(u)|^2] W(u) du \int K^2(\tau) d\tau,$$

$$\tilde{U} = \frac{1}{Th^{1/2}} \sum_{1 \le s \ne r \le T} U(\xi_s, \xi_r)$$

$$= \frac{1}{Th^{1/2}} \sum_{1 \le s \ne r \le T} \int \int K(\tau) K\left(\tau + \frac{s-r}{Th}\right) \varepsilon_s(u) \varepsilon_r^*(u) W(u) du d\tau.$$

PROPOSITION A.2. *Under the conditions of Theorem 1, $Q_2 = o_P(1)$.*

PROPOSITION A.3. *Under the conditions of Theorem 1, $Q_3 = o_P(1)$.*

PROPOSITION A.4. *Under the conditions of Theorem 1, $[\tilde{U} - B_2]/\sqrt{V} \xrightarrow{d} N(0,1)$, where*

$$B_2 = h^{-1/2} \sum_{k=1}^{\infty} \int [\sigma_k(u,u) + \sigma_{-k}(u,u)] W(u) du \int K^2(\tau) d\tau$$

$$V = 2 \int \int |\Omega(u,v)|^2 W(u) W(v) du dv \int \left[\int K(\tau) K(\tau + \eta) d\tau \right]^2 d\eta.$$

Based on Propositions A.1 to A.4, we can obtain the asymptotic centering factor $B = B_1 + B_2$ and scaling factor V as well as the asymptotic distribution of our test statistic. As our test statistic is obtained by replacing the asymptotic centering factor B and scaling factor V by their estimators $\hat{B}$ and $\hat{V}$, we should show that replacing B and V by their consistent estimators $\hat{B}$ and $\hat{V}$ does not affect the limiting distribution of our test statistic.

PROPOSITION A.5. *Under the conditions of Theorem 1, $\hat{B} - B = o_P(1)$ and $\hat{V} - V = o_P(1)$.*

The proof of Theorem 1 will be completed provided Propositions A.1–A.5 are proven, which we turn to next.

PROOF OF PROPOSITION A.1. Given any $(u, t/T)$, define $\bar{\phi}_t(u) = E[\hat{\phi}_t(u)]$. We first decompose Q_1 as follows:

$$\begin{aligned} Q_1 &= h^{1/2} \sum_{t=1}^{T} \int \left|\hat{\phi}_t(u) - \phi_t(u)\right|^2 W(u) du \\ &= h^{1/2} \sum_{t=1}^{T} \int \left|\hat{\phi}_t(u) - \bar{\phi}_t(u)\right|^2 W(u) du + h^{1/2} \sum_{t=1}^{T} \int \left|\bar{\phi}_t(u) - \phi_t(u)\right|^2 W(u) du \\ &\quad + 2h^{1/2} \sum_{t=1}^{T} \int Re\left\{[\hat{\phi}_t(u) - \bar{\phi}_t(u)][\bar{\phi}_t(u) - \phi_t(u)]^*\right\} W(u) du \\ &= Q_{11} + Q_{12} + Q_{13}, \text{ say.} \end{aligned} \tag{A.1}$$

The proof of Proposition A.1 consists of the proofs of Lemmas A.1–A.3 below.

LEMMA A.1. *Let* Q_{11} *be defined as in* (A.1). *Then* $Q_{11} = B_1 + \tilde{U} + o_P(1)$.

LEMMA A.2. *Let* Q_{12} *be defined as in* (A.1). *Then* $Q_{12} = o_P(1)$.

LEMMA A.3. *Let* Q_{13} *be defined as in* (A.1). *Then* $Q_{13} = o_P(1)$.

PROOF OF LEMMA A.1. We decompose Q_{11} as follows:

$$
\begin{aligned}
Q_{11} &= h^{1/2}\sum_{t=1}^{T}\int \left|\hat{\phi}_t(u)-\bar{\phi}_t(u)\right|^2 W(u)du = h^{1/2}\sum_{t=1}^{T}\int \left|\sum_{s=1}^{T}\frac{1}{Th}K\left(\frac{s-t}{Th}\right)\varepsilon_s(u)\right|^2 W(u)du \\
&\quad + h^{1/2}\sum_{t=1}^{[Th]}\int \left|\sum_{s=1}^{T}\frac{1}{Th}K\left(\frac{s+t}{Th}\right)\varepsilon_s(u)\right|^2 W(u)du \\
&\quad + h^{1/2}\sum_{t=1}^{[Th]}\int \frac{1}{T^2h^2}\sum_{s=1}^{T}\sum_{r=1}^{T}K\left(\frac{s-t}{Th}\right)K\left(\frac{r+t}{Th}\right)Re\left[\varepsilon_s(u)\varepsilon_r(u)^* + \varepsilon_s(u)^*\varepsilon_r(u)\right]W(u)du \\
&\quad + h^{1/2}\sum_{t=T-[Th]}^{T}\int \left|\sum_{s=1}^{T}\frac{1}{Th}K\left(\frac{s+t-2T}{Th}\right)\varepsilon_s(u)\right|^2 W(u)du \\
&\quad + h^{1/2}\sum_{t=T-[Th]}^{T}\int \frac{1}{T^2h^2}\sum_{s=1}^{T}\sum_{r=1}^{T}K\left(\frac{s-t}{Th}\right)K\left(\frac{r+t-2T}{Th}\right)Re\left[\varepsilon_s(u)\varepsilon_r(u)^*\right. \\
&\quad \left. + \varepsilon_s(u)^*\varepsilon_r(u)\right]W(u)du + o_P(1) \\
&= A + R_1 + R_2 + R_3 + R_4, \text{ say.}
\end{aligned}
$$

We further decompose the first term as follows:

$$
\begin{aligned}
A &= \frac{1}{T^2h^{3/2}}\sum_{t=1}^{T}K^2(0)\int |\varepsilon_t(u)|^2 W(u)du + \frac{2}{T^2h^{3/2}}\sum_{s\neq t}K(0)K\left(\frac{s-t}{Th}\right)\int \varepsilon_s(u)\varepsilon_t(u)^* W(u)du \\
&\quad + \frac{1}{T^2h^{3/2}}\sum_{s\neq t}K^2\left(\frac{s-t}{Th}\right)\int |\varepsilon_s(u)|^2 W(u)du \\
&\quad + \frac{1}{T^2h^{3/2}}\sum_{s\neq r\neq t}K\left(\frac{s-t}{Th}\right)K\left(\frac{r-t}{Th}\right)\int \varepsilon_r(u)\varepsilon_s(u)^* W(u)du \\
&= R_5 + R_6 + A_1 + A_2, \text{ say.}
\end{aligned}
$$

Now, we consider the A_1 term.

$$
\begin{aligned}
A_1 &= \frac{2}{T^2h^{3/2}}\sum_{t=1}^{T-1}\sum_{j=1}^{T-t}K^2\left(\frac{j}{Th}\right)\int |\varepsilon_t(u)|^2 W(u)du \\
&= \frac{2}{Th^{3/2}}\sum_{j=1}^{T-1}\left(1-\frac{j}{T}\right)K^2\left(\frac{j}{Th}\right)\int E\left[|\varepsilon_t(u)|^2\right]W(u)du
\end{aligned}
$$

$$+\left[\frac{2}{T^2h^{3/2}}\sum_{t=1}^{T-1}\sum_{j=1}^{T-t}K^2\left(\frac{j}{Th}\right)\int|\varepsilon_t(u)|^2W(u)du\right.$$

$$\left.-\frac{2}{Th^{3/2}}\sum_{j=1}^{T-1}\left(1-\frac{j}{T}\right)K^2\left(\frac{j}{Th}\right)\int E\left[|\varepsilon_t(u)|^2\right]W(u)du\right]$$

$$=A_1^{(1)}+A_1^{(2)}.$$

By the Riemann sum approximation of an integral, we have

$$A_1^{(1)}=h^{-1/2}\int E\left[|\varepsilon_t(u)|^2\right]W(u)du\int K^2(\tau)d\tau+o(1)=B_1+o(1).$$

It is easy to show $E[A_1^{(2)}]=0$ and $E[A_1^{(2)}]^2=O(T^{-2}h^{-3+1/(1+\delta)})$. Then, $A_1^{(2)}=o_P(1)$ follows from Chebyshev's inequality.

Then, we consider the A_2 term. By the Riemann approximation of an integral, we have

$$\begin{aligned}A_2&=\frac{1}{T^2h^{3/2}}\sum_{s\neq r\neq t}K\left(\frac{s-t}{Th}\right)K\left(\frac{s-t}{Th}+\frac{r-s}{Th}\right)\int\varepsilon_s(u)\varepsilon_r(u)^*W(u)du\\&=\frac{1}{Th^{1/2}}\sum_{s\neq r}\int K(\tau)K\left(\tau+\frac{s-r}{Th}\right)d\tau\int\varepsilon_s(u)\varepsilon_r(u)^*W(u)du+o_P(1)\\&=\tilde{U}+o_P(1).\end{aligned}$$

The proof of Lemma A.1 will be completed provided $R_i=o_P(1),\ i=1,2,\ldots,6$. The decompositions and proofs of R_1 and R_3 are quite similar to those of term A except the time index changes from $t=1,\cdots T$ to $t=1,\ldots,[Th]$ and $t=T-[Th],\ldots,T$. Thus, $R_1=h\cdot O_P(A)=o_P(1)$ and $R_3=h\cdot O_P(A)=o_P(1)$.

R_2 could be decomposed as

$$\begin{aligned}R_2&=2T^{-2}h^{-3/2}\sum_{t=1}^{[Th]}\sum_{s=1}^{T}\int K\left(\frac{s-t}{Th}\right)K\left(\frac{s+t}{Th}\right)|\varepsilon_s(u)|^2W(u)du\\&\quad+T^{-2}h^{-3/2}\sum_{t=1}^{[Th]}\sum_{s\neq r}\int K\left(\frac{s-t}{Th}\right)K\left(\frac{r+t}{Th}\right)\left[\varepsilon_s(u)\varepsilon_r(u)^*+\varepsilon_s(u)^*\varepsilon_r(u)\right]W(u)du\\&=R_{21}+R_{22},\ \text{say}.\end{aligned}$$

We can easily show $E[R_{21}]=2h^{1/2}\int\Omega(u,u)W(u)du\int\int K(\tau)K(\tau+2\eta)d\tau d\eta=O(h^{1/2})$ and $E[R_{21}^2]=o(h)$. Thus, $R_{21}=o_P(1)$. R_{22} is quite similar to term A_2 except for the range of the time index t changing from $t=1,\cdots T$ to $t=1,\ldots,[Th]$ and thus $R_{22}=o_P(1)$. By a similar argument, we know $R_4=o_P(1)$.

Now, we consider the term R_5:

$$\begin{aligned}R_5&=\frac{K^2(0)}{T^2h^{3/2}}\int\sum_{t=1}^{T}E|\varepsilon_t(u)|^2W(u)du+\frac{K^2(0)}{T^2h^{3/2}}\int\sum_{t=1}^{T}\left[|\varepsilon_t(u)|^2-E|\varepsilon_t(u)|^2\right]W(u)du\\&=R_{51}+R_{52},\quad\text{say}.\end{aligned}$$

Given Assumptions 2 and 3, we obtain the constant $R_{51} = O(T^{-1}h^{-3/2}) = o(1)$. Since $E(R_{52}) = 0$ and

$$\begin{aligned}\operatorname{var}(R_{52}) &\le K^4(0)T^{-4}h^{-3}\sum_{t=1}^{T}\operatorname{var}\left(\int |\varepsilon_t(u)|^2 W(u)du\right)\\ &\quad + K^4(0)T^{-3}h^{-3}\sum_{t=1}^{T-1}\left|\operatorname{cov}\left(\int |\varepsilon_1(u)|^2 W(u)du, \int |\varepsilon_{1+t}(u)|^2 W(u)du\right)\right|\\ &= O(T^{-3}h^{-3}).\end{aligned}$$

Then, by Chebyshev's inequality, we know $R_5 = o_p(1)$. Finally, define

$$\Psi_1(\xi_s, \xi_t) = \int K(0)K\left(\frac{s-t}{Th}\right)[\varepsilon_s(u)\varepsilon_t(u)^* + \varepsilon_t(u)\varepsilon_s(u)^*]\, W(u)du.$$

We have $R_6 = \frac{4}{T^2h^{3/2}}\sum_{1\le s<r\le T}\Psi_1(\xi_s, \xi_t)$, and $\int \Psi_1(\xi_s, \xi_t)dP(\xi_s) = 0$. Therefore,

$$E|R_6|^2 \le \frac{C}{T^2h^3}\left[E|\Psi_1(\xi_s, \xi_t)|^{2(1+\delta)}\right]^{1/(1+\delta)}\sum_{j=1}^{T-p} j\beta^{\delta/(1+\delta)}(j) = O(T^{-2}h^{-3+1/(1+\delta)}) = o(1).$$

According to Chebyshev's inequality, we have $R_6 = o_P(1)$. ■

Proof of Lemma A.2. Under $\mathbb{H}_0$, we have $\phi_t(u) = \psi_0(u)$. Therefore, the kth order derivative of $\phi_t(u)$ with respect to $\tilde{t} = t/T$: $\phi_t^{(k)}(u) = 0$ for any $u \in \mathbb{R}^{dm}$ and $k \ge 1$. Since the bias term

$$\bar{\phi}_t(u) - \phi_t(u) = \sum_{k=2}^{\infty}\frac{h^k}{k!}C_k\phi_t^{(k)}(u) = 0,$$

we know $Q_{12} = 0$. ■

Proof of Lemma A.3. $Q_{13} = 0$ because the bias term $\bar{\phi}_t(u) - \phi_t(u) = 0$. ■

Proof of Proposition A.2. We decompose Q_2 as follows:

$$\begin{aligned}Q_2 &= Th^{1/2}\int |\hat{\psi}_0(u) - \psi_0(u)|^2 W(u)du = Th^{1/2}\int\left|\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t(u)\right|^2 W(u)du\\ &= T^{-1}h^{1/2}\sum_{t=1}^{T}\int |\varepsilon_t(u)|^2 W(u)du + T^{-1}h^{1/2}\sum_{1\le t\ne s\le T}\int \varepsilon_t(u)\varepsilon_s(u)^* W(u)du\\ &= Q_{21} + Q_{22}.\end{aligned}$$

Obviously, $E(Q_{21}) = O(h^{1/2})$ and $E(Q_{22}) = O(h^{1/2})$. Define $\Psi_2(\xi_t) = \int |\varepsilon_t(u)|^2 W(u)du$ and $\Psi_3(\xi_t, \xi_s) = \int \varepsilon_t(u)\varepsilon_s(u)^* W(u)du$. Since

$$\begin{aligned}E\left[Q_{21}^2\right] &\le T^{-2}h\sum_{t=1}^{T}\operatorname{var}[\Psi_2(\xi_t)] + 2T^{-1}h\sum_{j=1}^{T-1}|\operatorname{cov}[\Psi_2(\xi_1), \Psi_2(\xi_{1+j})]|\\ &\le CT^{-1}h + CT^{-1}h\sum_{j=1}^{T-1} j^2\beta(j)^{\delta/(1+\delta)} = O(T^{-1}h) = o(1),\end{aligned}$$

we have $Q_{21} = o_P(1)$ by Chebyshev's inequality. In addition, by Lemma A(ii) of Hjellvik et al. (1998), we have

$$E\left[Q_{22}^2\right] \le Ch\left[E|\Psi_3(\xi_t, \xi_s)|^{2(1+\delta)}\right]^{1/(1+\delta)} \sum_{j=1}^{T} j^2 \beta(j)^{\delta/(1+\delta)} = O(h).$$

Then, $Q_{22} = o_P(1)$ follows from Chebyshev's inequality. Therefore, we have finished the proof of Proposition A.2. ■

Proof of Proposition A.3. We decompose Q_3 as follows:

$$\begin{aligned}
Q_3 &= h^{1/2} \sum_{t=1}^{T} \int Re\{[\hat{\phi}_t(u) - \phi_t(u)][\hat{\psi}_0(u) - \psi_0(u)]^*\} W(u)du \\
&= h^{1/2} \sum_{t=1}^{T} \int Re\left\{\left[\sum_{s=1}^{T} \frac{1}{Th} K\left(\frac{s-t}{Th}\right) \varepsilon_s(u)\right]\left[\sum_{r=1}^{T} \frac{1}{T} \varepsilon_r(u)\right]^*\right\} W(u)du \\
&= T^{-2}h^{-1/2} \sum_{t=1}^{T} K(0) \int |\varepsilon_t(u)|^2 W(u)du + T^{-2}h^{-1/2} \sum_{r \ne t} K(0) \int Re\left[\varepsilon_t(u)\varepsilon_r(u)^*\right] W(u)du \\
&\quad + T^{-2}h^{-1/2} \sum_{s \ne t} K\left(\frac{s-t}{Th}\right) \int Re\left[\varepsilon_s(u)\varepsilon_t(u)^*\right] W(u)du \\
&\quad + T^{-2}h^{-1/2} \sum_{s \ne t} K\left(\frac{s-t}{Th}\right) \int |\varepsilon_s(u)|^2 W(u)du \\
&\quad + T^{-2}h^{-1/2} \sum_{s \ne r \ne t} K\left(\frac{s-t}{Th}\right) \int Re\left[\varepsilon_s(u)\varepsilon_r(u)^*\right] W(u)du \\
&= Q_{31} + Q_{32} + Q_{33} + Q_{34} + Q_{35}, \text{ say,}
\end{aligned}$$

where we have used the fact that $\bar{\phi}_t(u) - \phi_t(u) = 0$ and $\bar{\phi}_0(u) - \psi_0(u) = 0$ under $\mathbb{H}_0$. We can easily show $E(Q_{31}) = O(T^{-1}h^{-1/2}) = o(1)$, $E(Q_{32}) = O(T^{-1}h^{-1/2}) = o(1)$, $E(Q_{33}) = O(T^{-1}h^{1/2}) = o(1)$, $E(Q_{34}) = O(h^{1/2}) = o(1)$ and $E(Q_{35}) = O(h^{1/2}) = o(1)$. It is also straightforward to show $E[Q_{31}^2] = O(T^{-3}h^{-1}) = o(1)$ and $E[Q_{32}^2] = O(T^{-2}h^{-1}) = o(1)$. By Lemma A(ii) of Hjellvik et al. (1998), we have $E[Q_{33}^2] = O(T^{-2}h^{1/(1+\delta)-1}) = o(1)$ and $E[Q_{34}^2] = O(T^{-2}h^{1/(1+\delta)-1}) = o(1)$. Therefore, we have $Q_{31} = o_P(1)$, $Q_{32} = o_P(1)$, $Q_{33} = o_P(1)$, and $Q_{34} = o_P(1)$ by Chebyshev's inequality.

Finally, we prove $Q_{35} = o_P(1)$. Define

$$\begin{aligned}
\Psi_4(\xi_t, \xi_s, \xi_r) &= K\left(\frac{s-t}{Th}\right) \int Re\left[\varepsilon_s(u)\varepsilon_r(u)^*\right] W(u)du + K\left(\frac{r-s}{Th}\right) \int Re\left[\varepsilon_r(u)\varepsilon_t(u)^*\right] W(u)du \\
&\quad + K\left(\frac{t-r}{Th}\right) \int Re\left[\varepsilon_t(u)\varepsilon_s(u)^*\right] W(u)du, \\
\Psi_4(\xi_s, \xi_r) &= \int \Psi_4(\xi_t, \xi_s, \xi_r) dP(\xi_t) \\
&= \int\int K\left(\frac{s-t}{Th}\right) Re\left[\varepsilon_s(u)\varepsilon_r(u)^*\right] W(u)du dP(\xi_t) \\
&= h \int\int K(\tau) Re\left[\varepsilon_s(u)\varepsilon_r(u)^*\right] W(u)du d\tau.
\end{aligned}$$

Then, Q_{35} could be rewritten as

$$Q_{35} = \frac{1}{3T^2h^{1/2}} \sum_{s\neq r\neq t} [\Psi_4(\xi_t, \xi_s, \xi_r) - \Psi_4(\xi_t, \xi_s) - \Psi_4(\xi_s, \xi_r) - \Psi_4(\xi_t, \xi_r)] - \frac{2}{T^2h^{1/2}} \sum_{s\neq r} \Psi_4(\xi_s, \xi_r)$$
$$+ \frac{1}{Th^{1/2}} \sum_{s\neq r} \Psi_4(\xi_s, \xi_r) = Q_{35}^{(1)} + Q_{35}^{(2)} + Q_{35}^{(3)}, \text{ say.}$$

We note that $M = \int |\Psi_4(\xi_t, \xi_s, \xi_r)|^2 dP = O(h)$, where P denotes any one of the probability measures in the set $\{P(\xi_t, \xi_s, \xi_r), P(\xi_t)P(\xi_s, \xi_r), P(\xi_t, \xi_r)P(\xi_s), P(\xi_t)P(\xi_s)P(\xi_r)\}$. By Lemma A(i) of Hjellvik et al. (1998), we have

$$E\left[Q_{35}^{(1)}\right]^2 \leq CT^{-1}h^{-1}M^{1/(1+\delta)} \sum_{j=1}^{\infty} j^2\beta(j)^{\delta/(1+\delta)} = O(T^{-1}h^{1/(1+\delta)-1}) = o(1).$$

Thus, $Q_{35}^{(1)} = o_P(1)$. $Q_{35}^{(2)} = o_P(1)$ provided $Q_{35}^{(3)} = o_P(1)$ holds, which we turn to next. We can easily show $Q_{35}^{(3)}$ is a degenerate second-order U statistic satisfying $E[Q_{35}^{(3)}] = O(h^{1/2})$ and $E[Q_{35}^{(3)}]^2 = O(h)$. By Chebyshev's inequality, we have $Q_{35}^{(3)} = o_P(1)$. Therefore, we have finished the proof of Proposition A.3. ■

PROOF OF PROPOSITION A.4. Consider the U statistic:

$$\tilde{U} = \frac{1}{Th^{1/2}} \sum_{r\neq s} U(\xi_s, \xi_r) = \frac{1}{Th^{1/2}} \sum_{r\neq s} \int K(\tau)K\left(\tau + \frac{s-r}{Th}\right) d\tau \int \varepsilon_s(u)\varepsilon_r(u)^* W(u) du.$$

Because $E[U(\xi_s, \xi)] = E[U(\xi', \xi_r)] = 0$ for any given ξ and ξ', $\tilde{U}$ is a second-order degenerate U -statistic.

Since $Th \to \infty$, there exists $p_T \to \infty$ such that $p_T = o(Th)$ and $\sum_{k=P_T}^{\infty} k^2\beta(k) \leq Cp_T^{-1}$. Denote $\mathbb{S}_1 = \{(s, r) : 1 \leq |s-r| \leq p_T, 1 \leq r \neq s \leq T\}$ and $\mathbb{S}_2 = \{(s, r) : 1 \leq r \neq s \leq T\} - \mathbb{S}_1$. Then

$$\begin{aligned}\tilde{U} &= \frac{1}{Th^{1/2}} \sum_{\mathbb{S}_1} U(\xi_s, \xi_r) + \frac{1}{Th^{1/2}} \sum_{\mathbb{S}_2} U(\xi_s, \xi_r) \\ &= U_1 + U_2,\end{aligned}$$

$$\begin{aligned}U_1 &= \frac{1}{Th^{1/2}} \sum_{\mathbb{S}_1} \int K(\tau)K\left(\tau + \frac{s-r}{Th}\right) d\tau \int \varepsilon_s(u)\varepsilon_r(u)^* W(u) du \\ &= \frac{1}{Th^{1/2}} \sum_{\mathbb{S}_1} \int K^2(\tau) d\tau \int \varepsilon_s(u)\varepsilon_r(u)^* W(u) du \\ &\quad + \frac{1}{Th^{1/2}} \sum_{\mathbb{S}_1} \int K(\tau)\left[K\left(\tau + \frac{s-r}{Th}\right) - K(\tau)\right] d\tau \int \varepsilon_s(u)\varepsilon_r(u)^* W(u) du \\ &= U_{11} + U_{12}.\end{aligned}$$

Now, we consider U_{11}. The asymptotic mean of U_{11} is

$$E[U_{11}] = \frac{1}{Th^{1/2}} \sum_{r=1}^{T-1} \sum_{s=r+1}^{r+p_T} \int E\left[\varepsilon_s(u)\varepsilon_r(u)^* + \varepsilon_s(u)^*\varepsilon_r(u)\right] W(u)du \int K^2(\tau)d\tau$$

$$= h^{-1/2} \sum_{k=1}^{p_T} \int E\left[\varepsilon_t(u)\varepsilon_{t+k}(u)^* + \varepsilon_t(u)^*\varepsilon_{t+k}(u)\right] W(u)du \int K^2(\tau)d\tau$$

$$= h^{-1/2} \int \sum_{k=1}^{\infty} [\sigma_k(u,u) + \sigma_{-k}(u,u)] W(u)du \int K^2(\tau)d\tau = B_2$$

as $p_T \to \infty$. The asymptotic variance of U_{11} is

$$\mathrm{var}(U_{11}) = E[U_{11}^2] - E[U_{11}]^2 = \frac{1}{T^2h} \sum_{\mathbb{S}_1(s,r)} \sum_{\mathbb{S}_1(k,l)} \int\int \big\{E\left[\varepsilon_s(u)\varepsilon_r(u)^*\varepsilon_k(v)^*\varepsilon_l(v)\right]$$

$$- E\left[\varepsilon_s(u)\varepsilon_r(u)^*\right] E\left[\varepsilon_k(v)^*\varepsilon_l(v)\right]\big\} W(u)W(v)dudv \left[\int K^2(\tau)d\tau\right]^2.$$

Denote

$$h(\xi_s, \xi_r) = \int \big\{\varepsilon_s(u)\varepsilon_r(u)^* - E\left[\varepsilon_s(u)\varepsilon_r(u)^*\right]\big\} W(u)du,$$

then, we can write

$$\tilde{U}_{11} = [U_{11} - E(U_{11})]^2$$

$$= \frac{4}{T^2h} \sum_{r=1}^{T-1} \sum_{s=r+1}^{\min\{T,p_T+r\}} \sum_{l=1}^{T-1} \sum_{k=l+1}^{\min\{T,p_T+l\}} Re\{h(\xi_s,\xi_r)h(\xi_k,\xi_l)\} \left[\int K^2(\tau)d\tau\right]^2.$$

It is obvious that $\mathrm{var}(U_{11}) = E[\tilde{U}_{11}]$. Let $i_{(1)}, i_{(2)}, i_{(3)}, i_{(4)}$ be the permutation of s, r, k, l in ascending order. Then, by Lemma 1 of Yoshihara (1976), we have

$$|E[h(\xi_s,\xi_r)h(\xi_k,\xi_l)]| \le 4M^{1/(1+\delta)}\beta(d)^{\delta/(1+\delta)},$$

where $d = \max\{i_{(4)} - i_{(3)}, i_{(3)} - i_{(2)}, i_{(2)} - i_{(1)}\}$ and

$$M = \max\left\{\int\int |h(\xi_s,\xi_r)|^{2+\delta} dP(\xi_s)dP(\xi_r), \int |h(\xi_s,\xi_r)|^{2+\delta} dP(\xi_s,\xi_r)\right\}.$$

If $d = i_{(2)} - i_{(1)}$, we have

$$\sum_{r=1}^{T-1} \sum_{s=r+1}^{\min\{T,p_T+r\}} \sum_{l=1}^{T-1} \sum_{k=l+1}^{\min\{T,p_T+l\}} Re\left\{E[h(\xi_s,\xi_r)h(\xi_k,\xi_l)]\right\}$$

$$\le \sum_{i_{(1)}=1}^{T-3} \sum_{i_{(2)}=i_{(1)}+\max\{i_{(3)}-i_{(2)},i_{(2)}-i_{(1)}\}}^{\min\{T,p_T+r\}} \sum_{i_{(3)}=i_{(2)}+1}^{T-1} \sum_{i_4=i_{(3)}+1}^{\min\{T,p_T+l\}} 4M^{1/(1+\delta)}\beta(i_{(2)}-i_{(1)})^{\delta/(1+\delta)}$$

$$\le \sum_{i_{(1)}=1}^{T-3} \sum_{i_{(2)}=i_{(1)}+1}^{T-2} 4M^{1/(1+\delta)}(i_{(2)}-i_{(1)})^2\beta(i_{(2)}-i_{(1)})^{\delta/(1+\delta)}$$

$$\le 4TM^{1/(1+\delta)} \sum_{k=1}^{T} k^2\beta(k)^{\delta/(1+\delta)}.$$

Similarly, we can show that, if $d = i_{(4)} - i_{(3)}$, we have

$$\sum_{r=1}^{T-1} \sum_{s=r+1}^{\min\{T,p_T+r\}} \sum_{l=1}^{T-1} \sum_{k=l+1}^{\min\{T,p_T+l\}} Re\left\{E[h(\xi_s,\xi_r)h(\xi_k,\xi_l)]\right\} \leq 4TM^{1/(1+\delta)} \sum_{k=1}^{T} k^2 \beta(k)^{\delta/(1+\delta)}.$$

If $d = i_{(3)} - i_{(2)}$, we have

$$\begin{aligned}
&\sum_{r=1}^{T-1} \sum_{s=r+1}^{\min\{T,p_T+r\}} \sum_{l=1}^{T-1} \sum_{k=l+1}^{\min\{T,p_T+l\}} Re\left\{E[h(\xi_s,\xi_r)h(\xi_k,\xi_l)]\right\} \\
&\quad \leq \max \left\{ \sum_{i_{(1)}=1}^{T-3} \sum_{i_{(2)}=i_{(i)}+1}^{\min\{T-2,p_T+i_{(1)}\}} \sum_{i_{(3)}=i_{(2)}+\max\{i_{(2)}-i_{(1)},i_{(4)}-i_{(3)}\}}^{\min\{T-1,p_T+i_{(1)}\}} \sum_{i_{(4)}=i_{(3)}+1}^{\min\{T,p_T+i_{(2)}\}} 4M^{1/(1+\delta)}\beta(i_{(3)}-i_{(2)})^{\delta/(1+\delta)}, \right. \\
&\qquad \sum_{i_{(1)}=1}^{T-3} \sum_{i_{(2)}=i_{(i)}+1}^{\min\{T-2,p_T+i_{(1)}\}} \sum_{i_{(3)}=i_{(2)}+\max\{i_{(2)}-i_{(1)},i_{(4)}-i_{(3)}\}}^{T-1} \sum_{i_{(4)}=i_{(3)}+1}^{\min\{T,p_T+i_{(3)}\}} 4M^{1/(1+\delta)}\beta(i_{(3)}-i_{(2)})^{\delta/(1+\delta)}, \\
&\qquad \left. \sum_{i_{(1)}=1}^{T-3} \sum_{i_{(2)}=i_{(i)}+1}^{\min\{T-2,p_T+i_{(1)}\}} \sum_{i_{(3)}=i_{(2)}+\max\{i_{(2)}-i_{(1)},i_{(4)}-i_{(3)}\}}^{\min\{T-2,p_T+i_{(1)}\}} \sum_{i_{(4)}=i_{(3)}+1}^{\min\{T-2,p_T+i_{(1)}\}} 4M^{1/(1+\delta)}\beta(i_{(3)}-i_{(2)})^{\delta/(1+\delta)} \right\} \\
&\quad \leq 4p_T TM^{1/(1+\delta)} \sum_{k=1}^{T} k\beta(k)^{\delta/(1+\delta)}.
\end{aligned}$$

Therefore, we have

$$\mathrm{var}(U_{11}) \leq C\frac{p_T}{Th} M^{1/(1+\delta)} \sum_{k=1}^{T} k\beta(k)^{\delta/(1+\delta)} \left[\int K^2(\tau)d\tau\right]^2 = o(1).$$

Thus, $U_{11} = E[U_{11}] + o_P(1)$.

Now, we prove $U_{12} = o_P(1)$. By the Lipschitz condition, we have

$$\left| K\left(\tau + \frac{s-r}{Th}\right) - K(\tau) \right| \leq C\frac{|s-r|}{Th}.$$

Thus,

$$\begin{aligned}
E(U_{12}) &= \frac{2}{Th^{1/2}} \sum_{r=1}^{T-1} \sum_{s=r+1}^{\min\{T,r+p_T\}} \int K(\tau)\left[K\left(\tau+\frac{s-r}{Th}\right) - K(\tau)\right]d\tau \int Re\left\{E\left[\varepsilon_s(u)\varepsilon_r(u)^*\right]\right\} W(u)du \\
&\leq C\frac{1}{Th^{3/2}} \sum_{k=1}^{p_T} k\beta(k)^{\delta/(1+\delta)} = o(1)
\end{aligned}$$

and

$$\begin{aligned}
E\left[U_{12}^2\right] = \frac{4}{T^2h} \sum_{r=1}^{T-1} \sum_{s=r+1}^{\min\{T,r+p_T\}} \sum_{l=1}^{T-1} \sum_{k=l+1}^{\min\{T,l+p_T\}} \int K(\tau_1)\left[K\left(\tau_1+\frac{s-r}{Th}\right) - K(\tau_1)\right]d\tau_1 \\
\times \int K(\tau_2)\left[K\left(\tau_2+\frac{l-k}{Th}\right) - K(\tau_2)\right]d\tau_2 \int\int Re\left\{E\left[\varepsilon_s(u)\varepsilon_r(u)^*\varepsilon_l(v)^*\varepsilon_k(v)\right]\right\} W(u)W(v)dudv
\end{aligned}$$

$$\leq \frac{C}{T^2h}\sum_{r=1}^{T-1}\sum_{s=r+1}^{\min\{T,r+p_T\}}\sum_{l=1}^{T-1}\sum_{k=l+1}^{\min\{T,l+p_T\}}\frac{(s-r)}{Th}\frac{(l-k)}{Th}\int\int Re\left\{E\left[\varepsilon_s(u)\varepsilon_r(u)^*\varepsilon_l(v)^*\varepsilon_k(v)\right]\right\}W(u)W(v)dudv$$

$$= \frac{C}{T^2h}\sum_{r=1}^{T-1}\sum_{s=r+1}^{\min\{T,r+p_T\}}\sum_{l=1}^{T-1}\sum_{k=l+1}^{\min\{T,l+p_T\}}\frac{(s-r)}{Th}\frac{(l-k)}{Th}$$

$$\times\int\int Re\left\{E\left[\varepsilon_s(u)\varepsilon_r(u)^*\varepsilon_l(v)^*\varepsilon_k(v)\right]-E\left[\varepsilon_s(u)\varepsilon_r(u)^*\right]E\left[\varepsilon_l(v)^*\varepsilon_k(v)\right]\right\}W(u)W(v)dudv$$

$$+\frac{C}{T^2h}\sum_{r=1}^{T-1}\sum_{s=r+1}^{\min\{T,r+p_T\}}\sum_{l=1}^{T-1}\sum_{k=l+1}^{\min\{T,l+p_T\}}\frac{(s-r)}{Th}\frac{(l-k)}{Th}\int\int Re\left\{E\left[\varepsilon_s(u)\varepsilon_r(u)^*\right]E\left[\varepsilon_l(v)^*\varepsilon_k(v)\right]\right\}W(u)W(v)dudv$$

$$= V_1+V_2.$$

Following the analysis of $\text{var}(U_{11})$, we can show $V_1 = O(p_T^3T^{-3}h^{-3}) = o(1)$ and

$$V_2 \leq \tfrac{C}{T^2h^3}\sum_{k_1=1}^{p_T}k_1\beta(k)\sum_{k_2=1}^{p_T}k_2\beta(k) = O(T^{-2}h^{-3}) = o(1)$$

Hence, we have $U_{12} = o_P(1)$ by Chebyshev's inequality.

Next, we consider the U_2 term.

$$U_2 = \tfrac{2}{Th^{1/2}}\sum_{r=1}^{T-p_T-1}\sum_{s=r+p_T+1}^{T}\int K(\tau)K\left(\tau+\frac{s-r}{Th}\right)d\tau\int Re\left[\varepsilon_s(u)\varepsilon_r(u)^*\right]W(u)du.$$

Similarly as Cai (2007), we have

$$\left|E[U_2]\right| \leq CT^{-1}h^{-1/2}\sum_{k=1}^{T}\int K(\tau)K\left(\tau+\frac{k}{Th}\right)d\tau\sum_{l=p_T+1}^{T}\beta(l) = O(h^{1/2}p_T^{-1}) = o(1).$$

Now, we consider the variance of U_2:

$$\text{var}(U_2) = E(U_2^2)$$

$$= \frac{4}{T^2h}\sum_{s=1}^{T-p_T}\sum_{r=p_T+1}^{T-s}\sum_{l=1}^{T-p_T}\sum_{k=p_T+1}^{T-l}\int K(\tau_1)K\left(\tau_1+\frac{r}{Th}\right)d\tau_1\int K(\tau_2)K\left(\tau_2+\frac{k}{Th}\right)d\tau_2$$

$$\times\int\int Re\left\{E\left[\varepsilon_s(u)\varepsilon_{s+r}(u)^*\varepsilon_l(v)^*\varepsilon_{k+l}(v)\right]\right\}W(u)W(v)dudv$$

$$= \frac{4}{T^2h}\sum_{s=1}^{T-p_T}\sum_{r=p_T+1}^{T-s}\left[\int K(\tau)K\left(\tau+\frac{r}{Th}\right)d\tau\right]^2$$

$$\times\int\int Re\left\{E\left[\varepsilon_s(u)\varepsilon_{s+r}(u)^*\varepsilon_s(v)^*\varepsilon_{s+r}(v)\right]\right\}W(u)W(v)dudv$$

$$+\frac{4}{T^2h}\sum_{s=1}^{T-p_T}\sum_{l\neq s=1}^{T-p_T}\sum_{r=p_T+1}^{T-\max\{s,l\}}\int K(\tau_1)K\left(\tau_1+\frac{r}{Th}\right)d\tau_1\int K(\tau_2)K\left(\tau_2+\frac{r}{Th}\right)d\tau_2$$

$$\times\int\int Re\left\{E\left[\varepsilon_s(u)\varepsilon_{s+r}(u)^*\varepsilon_l(v)^*\varepsilon_{l+r}(v)\right]\right\}W(u)W(v)dudv$$

$$+\frac{4}{T^2h}\sum_{s=1}^{T-p_T}\sum_{r=p_T+1}^{T-s}\sum_{k\neq r=p_T+1}^{T-s}\int K(\tau_1)K\left(\tau_1+\frac{r}{Th}\right)d\tau_1\int K(\tau_2)K\left(\tau_2+\frac{k}{Th}\right)d\tau_2$$

$$\times \int \int Re\left\{E\left[\varepsilon_s(u)\varepsilon_{s+r}(u)^*\varepsilon_s(v)^*\varepsilon_{s+k}(v)\right]\right\} W(u)W(v)dudv$$

$$+ \frac{4}{T^2h}\sum_{s=1}^{T-p_T}\sum_{r=p_T+1}^{T-s}\sum_{l\neq s=1}^{T-p_T}\sum_{k\neq r=p_T+1}^{T-l}\int K(\tau_1)K\left(\tau_1+\frac{r}{Th}\right)d\tau_1\int K(\tau_2)K\left(\tau_2+\frac{k}{Th}\right)d\tau_2$$

$$\times \int \int Re\left\{E\left[\varepsilon_s(u)\varepsilon_{s+r}(u)^*\varepsilon_l(v)^*\varepsilon_{k+l}(v)\right]\right\} W(u)W(v)dudv$$

$$= V_3 + V_4 + V_5 + V_6.$$

We shall analyze each of these terms $\{V_i\}_{i=3}^6$ above to determine the asymptotic variance of the degenerate U statistic.

We first consider the V_3 term. Since

$$\begin{aligned}
&Re\left[\varepsilon_s(u)\varepsilon_s(v)\varepsilon_r(u)^*\varepsilon_r(v)^*\right]\\
&= Re\left\{\left[\varepsilon_s(u+v)-\phi_s(u)\varepsilon_s(v)-\phi_s(v)\varepsilon_s(u)+\phi_s(u,v)-\phi_s(u)\phi_s(v)\right]\right.\\
&\quad\left.\times\left[\varepsilon_r(u+v)-\phi_r(u)\varepsilon_r(v)-\phi_r(v)\varepsilon_r(u)+\phi_r(u,v)-\phi_r(u)\phi_r(v)\right]^*\right\}\\
&= Re\left\{\left[\phi_s(u,v)-\phi_s(u)\phi_s(v)\right]\left[\phi_r(u,v)-\phi_r(u)\phi_r(v)\right]^*\right\} + Re\left[\Upsilon_{s,r}(u,v)\right]\\
&= \left|\phi(u+v)-\phi(u)\phi(v)\right|^2 + Re\left[\Upsilon_{s,r}(u,v)\right],
\end{aligned}$$

where

$$\begin{aligned}
\Upsilon_{s,r}(u,v) &= \varepsilon_s(u+v)\varepsilon_r(u+v)^* + \phi_s(v)\phi_r(v)^*\varepsilon_s(u)\varepsilon_r(u)^* + \phi_s(u)\phi_r(u)^*\varepsilon_s(v)\varepsilon_r(v)^*\\
&\quad + \phi_s(u)\phi_r(v)^*\varepsilon_s(v)\varepsilon_r(u)^* + \phi_s(v)\phi_r(u)^*\varepsilon_s(u)\varepsilon_r(v)^* - \phi_r(u)^*\varepsilon_s(u+v)\varepsilon_r(v)^*\\
&\quad -\phi_r(v)^*\varepsilon_s(u+v)\varepsilon_r(u)^* - \phi_s(u)\varepsilon_s(v)\varepsilon_r(u+v)^* - \phi_s(v)\varepsilon_s(u)\varepsilon_r(u+v)^*,
\end{aligned}$$

we have

$$\begin{aligned}
V_3 &= \frac{4}{T^2h}\sum_{s=1}^{T-p_T}\sum_{r=p_T+1}^{T-s}\int\int\left|\psi_0(u+v)-\psi_0(u)\psi_0(v)\right|^2 W(u)W(v)dudv\left[\int K(\tau)K\left(\tau+\frac{r}{Th}\right)d\tau\right]^2\\
&\quad + \frac{4}{T^2h}\sum_{s=1}^{T-p_T}\sum_{r=p_T+1}^{T-s}\int\int Re\left\{E\left[\Upsilon_{s,s+r}(u,v)\right]\right\}W(u)W(v)dudv\left[\int K(\tau)K\left(\tau+\frac{r}{Th}\right)d\tau\right]^2\\
&= V_{31}+V_{32},
\end{aligned}$$

where

$$V_{31} = \frac{2(T-p_T)(T-p_T-1)}{T^2}\int\int\left|\psi_0(u+v)-\psi_0(u)\psi_0(v)\right|^2 W(u)W(v)dudv\int\left[\int K(\tau)K(\tau+\eta)\,d\tau\right]^2 d\eta + o(1)$$

by the Riemann sum approximation of an integral and

$$V_{32} \leq \frac{C}{Th}\sum_{r=p_T+1}^{T-1}\int\int Re\left\{E\left[\Upsilon_{1,1+r}(u,v)\right]\right\}W(u)W(v)dudv\left[\int K(\tau)K\left(\tau+\frac{r}{Th}\right)d\tau\right]^2 = O(T^{-1}h^{-1}p_T^{-1}) = o(1).$$

Now, we consider the V_4 term.

$$V_4 = \frac{4}{T^2 h} \sum_{r=1}^{T-p_T-1} \sum_{l \neq s=1}^{T-p_T-1} \sum_{s=r+p_T+1}^{T} \int K(\tau_1) K\left(\tau_1 + \frac{s-r}{Th}\right) d\tau_1 \int K(\tau_2) K\left(\tau_2 + \frac{s-l}{Th}\right) d\tau_2$$
$$\times \int\int Re\left\{E\left[\varepsilon_s(u)\varepsilon_r(u)^* \varepsilon_l(v)^* \varepsilon_s(v)\right]\right\} W(u)W(v)dudv$$
$$= \frac{4}{T^2 h} \sum_{r=1}^{T-p_T-1} \sum_{l \neq s=1}^{T-p_T-1} \sum_{s=r+p_T+1}^{T} \int K(\tau_1) K\left(\tau_1 + \frac{s-r}{Th}\right) d\tau_1 \int K(\tau_2) K\left(\tau_2 + \frac{s-l}{Th}\right) d\tau_2$$
$$\times \int\int Re\left\{\left[\psi_0(u+v) - \psi_0(u)\psi_0(v)\right] E\left[\varepsilon_r(u)^* \varepsilon_l(v)^*\right] + E\left[\varepsilon_s(u+v)\varepsilon_r(u)^* \varepsilon_l(v)^*\right]\right.$$
$$\left. -\psi_0(u) E\left[\varepsilon_s(v)\varepsilon_r(u)^* \varepsilon_l(v)^*\right] - \psi_0(v) E\left[\varepsilon_s(u)\varepsilon_r(u)^* \varepsilon_l(v)^*\right]\right\} W(u)W(v)dudv$$
$$= V_{41} + V_{42} + V_{43} + V_{44}.$$

The V_{41} term could be split into two terms as $\sum_{|r-l|\leq p_T}\sum_s$ and $\sum_{|r-l|>p_T}\sum_s$, denoted as V_{411} and V_{412}, respectively. It is straightforward to show $V_{412} = o(1)$ as $\sum_{k>p_T} E[\varepsilon_t(u)\varepsilon_{t+k}(v)^*] < p_T^{-1}$. By the Riemann sum approximation of the integral, we have

$$V_{411} = \frac{2(T-p_T)(T-p_T-1)}{T^2} \int \left[\int K(\tau) K(\tau+\eta)\, d\tau\right]^2 d\eta$$
$$\times \int\int \left[\psi_0(u+v) - \psi_0(u)\psi_0(v)\right] \sum_{k=1}^{p_T} E[\varepsilon_t(u)\varepsilon_{t+k}(v)^* + \varepsilon_t(u)^* \varepsilon_{t+k}(v)] W(u)W(v)dudv + o(1)$$
$$= 2\int \left[\int K(\tau)K(\tau+\eta)\, d\tau\right]^2 d\eta \int\int \left[\psi_0(u+v) - \psi_0(u)\psi_0(v)\right] \Omega_+(u,v) W(u)W(v)dudv + o(1)$$

with $\Omega_+(u,v) = \sum_{k=1}^{p_T} E[\varepsilon_t(u)\varepsilon_{t+k}(v)^* + \varepsilon_t(u)^*\varepsilon_{t+k}(v)]$.

Now, we consider the V_{42} term. Denote $d = \max\{|s-r|, |r-l|, |l-s|\}$. Then, V_{42} could be split into two terms with $d \leq p_T$ and $d > p_T$, denoted as V_{421} and V_{422}, respectively. It is straightforward to show $V_{421} = O(p_T/T)$ and $V_{422} = O(p_T^{-1})$, where we have use the inequality $\sup_{u,u'} E[\varepsilon_s(u)\varepsilon_r(u)\varepsilon_l(u'^{\delta/(1+\delta)}$, which is derived from Roussas and Ioannides (1987). Thus, $V_{42} = o(1)$. Following the analogous arguments, we can show $V_{43} = o(1)$ and $V_{44} = o(1)$.

Next, we consider the V_5 term. Following the analogous analysis of V_4, we can show

$$V_5 = 2\int \left[\int K(\tau)K(\tau+\eta)\, d\tau\right]^2 d\eta \int\int \left[\psi_0(u+v) - \psi_0(u)\psi_0(v)\right] \Omega_+(u,v) W(u)W(v)dudv$$
$$+ o(1).$$

Finally, we consider the V_6 term.

$$V_6 = \frac{4}{T^2 h} \sum_{r=1}^{T-p_T} \sum_{s=p_T+r}^{T} \sum_{l \neq r=1}^{T-p_T} \sum_{k \neq s=p_T+l}^{T} \int K(\tau_1) K\left(\tau_1 + \frac{s-r}{Th}\right) d\tau_1 \int K(\tau_2) K\left(\tau_2 + \frac{k-l}{Th}\right) d\tau_2$$
$$\times \int\int Re\left\{E\left[\varepsilon_s(u)\varepsilon_r(u)^* \varepsilon_l(v)^* \varepsilon_k(v)^*\right] - E\left[\varepsilon_s(u)\varepsilon_l(v)^*\right] E\left[\varepsilon_r(u)^* \varepsilon_k(v)^*\right]\right\} W(u)W(v)dudv$$
$$+ \frac{4}{T^2 h} \sum_{r=1}^{T-p_T} \sum_{s=p_T+r}^{T} \sum_{l \neq r=1}^{T-p_T} \sum_{k \neq s=p_T+l}^{T} \int K(\tau_1) K\left(\tau_1 + \frac{s-r}{Th}\right) d\tau_1 \int K(\tau_2) K\left(\tau_2 + \frac{k-l}{Th}\right) d\tau_2$$

$$\times \int \int Re \left\{ E\left[\varepsilon_r(u)\varepsilon_l(v)^*\right] E\left[\varepsilon_s(u)^*\varepsilon_k(v)^*\right] \right\} W(u)W(v)dudv$$
$$= V_{61} + V_{62}.$$

Following the analogous analysis of the $\mathrm{var}(U_{11})$, we can show $V_{61} = o(1)$. Now, we consider the V_{62} term:

$$V_{62} = \frac{4}{T^2h} \sum_{r=1}^{T-p_T} \sum_{s=p_T+r}^{T} \sum_{l\neq r=1}^{T-p_T} \sum_{k\neq s=p_T+l}^{T} \int K(\tau_1)K\left(\tau_1 + \frac{s-r}{Th}\right)d\tau_1 \int K(\tau_2)K\left(\tau_2 + \frac{k-l}{Th}\right)d\tau_2$$
$$\times \int \int Re\left\{\sigma_{|r-l|}(u,v)\sigma_{|s-k|}(u,v)^*\right\} W(u)W(v)dudv.$$

Since $\sum_{k>p_T} E[\varepsilon_t(u)\varepsilon_{t+k}(v)] < p_T^{-1}$ for any fixed constants (u, v), we can split V_{62} to four terms as $\sum_{|r-l|<p_T,|s-k|<p_T}(\cdots)$, $\sum_{|r-l|\geq p_T,|s-k|<p_T}(\cdots)$, $\sum_{|r-l|<p_T,|s-k|\geq p_T}(\cdots)$, $\sum_{|r-l|\geq p_T,|s-k|\geq p_T}(\cdots)$ and can further show the last three terms are all higher order terms. Therefore,

$$V_{62} = \frac{4}{T^2h} \sum_{r=1}^{T-p_T} \sum_{l\neq r,|l-r|<p_T}^{T-p_T} \sum_{s=p_T+r}^{T} \sum_{s=p_T+l,|s-k|<p_T}^{T} \int K(\tau_1)K\left(\tau_1 + \frac{s-r}{Th}\right)d\tau_1 \int K(\tau_2)K\left(\tau_2 + \frac{k-l}{Th}\right)d\tau_2$$
$$\times \int \int Re\left\{\sigma_{|r-l|}(u,v)\sigma_{|s-k|}(u,v)^*\right\} W(u)W(v)dudv + o(1)$$
$$= \frac{16}{T^2h} \sum_{r=1}^{T-p_T} \sum_{l'=1}^{\min\{|T-p_T-r|,p_T\}} \sum_{k=l'+r+p_T}^{T} \sum_{s'=\max\{r-k+p_T,1\}}^{\min\{|T-k|,p_T\}} \int \int \sigma_{l'}(u,v)\sigma_{s'}(u,v)W(u)W(v)dudv$$
$$\times \int K(\tau_1)K\left(\tau_1 + \frac{k-r+s'}{Th}\right)d\tau_1 \int K(\tau_2)K\left(\tau_2 + \frac{k-r-l'}{Th}\right)d\tau_2 + o(1)$$
$$= 2\int\int \left|\sum_{k=1}^{p_T}[\sigma_k(u,v) + \sigma_{-k}(u,v)]\right|^2 W(u)W(v)dudv \int \left[\int K(\tau)K(\tau+\eta)\,d\tau\right]^2 d\eta + o(1)$$
$$= 2\int\int \left|\Omega_+(u,v)\right|^2 W(u)W(v)dudv \int \left[\int K(\tau)K(\tau+\eta)\,d\tau\right]^2 d\eta + o(1).$$

Therefore, we get the asymptotic variance of U_2:

$$\mathrm{var}\left(U_2\right) = 2\int\int \left|\Omega(u,v)\right|^2 W(u)W(v)dudv \int \left[\int K(\tau)K(\tau+\eta)\,d\tau\right]^2 d\eta,$$

where $\Omega(u, v) = \sum_{k=-\infty}^{\infty} E[\varepsilon_t(u)\varepsilon_{t+k}(v)^*] < \infty$ is a generalized long-run variance.

Finally, we prove the asymptotic normality that $Z = [\mathrm{var}(U_2)]^{-1/2}U_2 \xrightarrow{d} N(0, 1)$. Following the proof of Lemma A.4 in Kim et al. (2011), we could show

$$E[Z^{2r+1}] = o(1), \quad E[Z^{2r}] = (2r-1)!!,$$

where r is a positive integer and $(2r-1)!!$ denotes the product of every odd number from 1 to $2r-1$. The proof is quite tedious and could be found in Lemma A.4 of Kim et al. (2011). Then, for all $u \in \mathbb{R}$, the moment generating function of Z is

$$M_Z(u) = E\left(e^{uZ}\right) = \sum_{k=1}^{\infty} \frac{u^k}{k!}E[Z^k] = \sum_{r=1}^{\infty} \frac{u^{2r}}{(2r)!}E[Z^{2r}] = \sum_{r=1}^{\infty} \frac{u^{2r}(2r-1)!!}{(2r)!}$$
$$= \sum_{r=1}^{\infty} \frac{1}{r!}\left(\frac{u^2}{2}\right)^r = e^{\frac{u^2}{2}} = M_{N(0,1)}(u).$$

By the uniqueness of moment generating function, we know $Z \sim N(0, 1)$. Therefore, we have established the asymptotic normality for U_2 and finish the proof of Proposition A.4. We now summarize our result that $\tilde{U}$ asymptotically follows a normal distribution with asymptotic mean and variance as follows:

$$E[\tilde{U}] = h^{-1/2} \int \sum_{k=1}^{\infty} [\sigma_k(u) + \sigma_{-k}(u)]\, W(u) du \int K^2(\tau) d\tau,$$

$$\operatorname{var}[\tilde{U}] = 2 \int \int \left|\Omega(u, v)\right|^2 W(u) W(v) du dv \int \left[\int K(\tau) K(\tau + \eta)\, d\tau\right]^2 d\eta.$$

■

PROOF OF PROPOSITION A.5. We should prove (i) $\hat{B} - B = o_P(1)$ and (ii) $\hat{V} - V = o_P(1)$. Since the proof of (i) and (ii) are similar, we focus on the proof of (i) here. Define $\tilde{\Omega}(u, u) = \sum_{i=-p_T}^{p_T} k_l(\frac{j}{p_T}) \tilde{\sigma}_j(u)$ with $\tilde{\sigma}_j(u) = \frac{1}{T} \sum_{t=j+1}^{T} \varepsilon_t(u) \varepsilon_{t-j}(u)^*$. Then we could decompose $\hat{B} - B$ as follows:

$$\begin{aligned}
\hat{B} - B &= h^{-1/2} \int \left[\tilde{\Omega}(u, u) - \Omega(u, u)\right] W(u) du \int K(\tau)^2 d\tau + h^{-1/2} \int \left[\hat{\Omega}(u, u) - \tilde{\Omega}(u, u)\right] W(u) du \int K(\tau)^2 d\tau \\
&= B_1 + B_2.
\end{aligned}$$

We first show $B_1 = o_P(1)$. Since

$$\begin{aligned}
E[B_1] &= h^{-1/2} \int \left[\sum_{j=-p_T}^{p_T} \frac{(T-j)}{T} k_l\left(\frac{j}{p_T}\right) \sigma_j(u) - \sum_{j=-\infty}^{\infty} \sigma_j(u)\right] W(u) du \int K(\tau)^2 d\tau \\
&= h^{-1/2} \int \left\{\sum_{|j| \le p_T} \left[k_l\left(\frac{j}{p_T}\right) - 1\right] \sigma_j(u) - \sum_{|j| \le p_T} \frac{j}{T} k_l\left(\frac{j}{p_T}\right) \sigma_j(u) - \sum_{|j| > p_T} \sigma_j(u)\right\} W(u) du \int K(\tau)^2 d\tau \\
&= B_1^{(1)} - B_1^{(2)} - B_1^{(3)}, \text{ say.}
\end{aligned}$$

Since the kernel function $k_l(\cdot)$ satisfies the Lipschitz condition, we have

$$B_1^{(1)} \le h^{-1/2} p_T^{-1} \int \sum_{|j| \le p_T} j \sigma_j(u) W(u) du \int K^2(\tau) d\tau = O\left(h^{-1/2} p_T^{-1}\right) = o(1)$$

by the β-mixing condition and the assumption that $p_T h^{1/2} \to \infty$. Similarly, we have $B_1^{(2)} = O(T^{-1} h^{-1/2}) = o(1)$ by the fact that $|k_l(\cdot)| \le 1$ and the β-mixing condition. In addition, $B_1^{(3)} = O(p_T^{-1} h^{-1/2}) = o(1)$ following from $\sum_{|j| > p_T} \sigma_j(u) \le p_T^{-1}$. Thus, we have $E[B_1] = o(1)$. Following the proof of Theorem 5A in Parzen (1957), we could show $E[B_1^2] = O(T^{-1} h^{-1}) = o(1)$. Hence we obtain $B_1 = o_P(1)$ by Chebyshev's inequality.

Now, we show $B_2 = o_P(1)$. Under the null hypothesis, B_2 could be decomposed as follows:

$$\begin{aligned}
B_2 &= h^{-1/2} \int \sum_{j=-p_T}^{p_T} \sum_{t=j+1}^{T} \frac{1}{T} k_l\left(\frac{j}{p_T}\right) \left[\hat{\varepsilon}_t(u) \hat{\varepsilon}_{t-j}(u)^* - \varepsilon_t(u) \varepsilon_{t-j}(u)^*\right] W(u) du \int K^2(\tau) d\tau \\
&= T^{-1} h^{-1/2} \int \sum_{j=-p_T}^{p_T} \sum_{t=j+1}^{T} k_l\left(\frac{j}{p_T}\right) \left|\hat{\psi}_0(u) - \psi_0(u)\right|^2 W(u) du \int K^2(\tau) d\tau \\
&\quad - T^{-1} h^{-1/2} \int \sum_{j=-p_T}^{p_T} \sum_{t=j+1}^{T} k_l\left(\frac{j}{p_T}\right) \left[\hat{\psi}_0(u) - \psi_0(u)\right] \left[e^{\mathbf{i} u X_t} - \psi_0(u)\right]^* W(u) du \int K^2(\tau) d\tau
\end{aligned}$$

$$-T^{-1}h^{-1/2}\int\sum_{j=-p_T}^{p_T}\sum_{t=j+1}^{T}k_l\left(\frac{j}{p_T}\right)\left[\hat{\psi}_0(u)-\psi_0(u)\right]\left[e^{\mathbf{i}uX_{t-j}}-\psi_0(u)\right]^*W(u)du\int K^2(\tau)d\tau$$
$$=B_{21}-B_{22}-B_{23},\text{ say.}$$

Now, we consider the B_{21} term:

$$B_{21}=T^{-3}h^{-1/2}\sum_{j=-p_T}^{p_T}\sum_{t=j+1}^{T}\sum_{r=1}^{T}\sum_{s=1}^{T}k_l\left(\frac{j}{p_T}\right)\int\varepsilon_r(u)\varepsilon_s(u)^*W(u)du\int K^2(\tau)d\tau.$$

It is straightforward to show $E|B_{21}|=O(T^{-1}h^{-1/2}p_T)$. Thus, $B_{21}=o_P(1)$ by Markov's inequality. In addition,

$$E\left|B_{22}\right|=T^{-2}h^{-1/2}\int\sum_{j=-p_T}^{p_T}\sum_{t=j+1}^{T}\sum_{s=1}^{T}k_l\left(\frac{j}{p_T}\right)E\left|\varepsilon_s(u)\varepsilon_t(u)^*\right|W(u)du\int K^2(\tau)d\tau$$
$$=O(T^{-1}h^{-1/2}p_T)=o(1).$$

Hence, we have $B_{22}=o_P(1)$ by Markov's inequality. Similarly, we have $B_{23}=o_P(1)$. Consequently, we have $B_2=o_P(1)$. Thus, we have shown $\hat{B}-B=o_P(1)$. ■

PROOF OF THEOREM 2. Under the alternative hypothesis, we can decompose $Th^{1/2}\hat{Q}$ as follows:

$$Th^{1/2}\hat{Q}=h^{1/2}\sum_{t=1}^{T}\int\left|\hat{\phi}_t(u)-\phi_t(u)\right|^2W(u)du+h^{1/2}\sum_{t=1}^{T}\int\left|\hat{\psi}_0(u)-\phi_t(u)\right|^2W(u)du$$
$$-2h^{1/2}\sum_{t=1}^{T}\int Re\left\{\left[\hat{\phi}_t(u)-\phi_t(u)\right]\left[\hat{\psi}_0(u)-\phi_t(u)\right]^*\right\}W(u)du$$
$$=Q_4+Q_5+Q_6,\text{ say.}$$

We first consider the term Q_4.

$$Q_4=h^{1/2}\sum_{t=1}^{T}\int\left|\sum_{s=t-[Th]}^{s=t+[Th]}\frac{1}{Th}K\left(\frac{s-t}{Th}\right)\varepsilon_s(u)\right|^2W(u)du+h^{1/2}\sum_{t=1}^{T}\int\left|\sum_{s=t-[Th]}^{s=t+[Th]}\frac{1}{Th}K\left(\frac{s-t}{Th}\right)\left[\phi_s(u)-\phi_t(u)\right]\right|^2W(u)du$$
$$+2h^{1/2}\sum_{t=1}^{T}\int Re\left\{\left[\sum_{s=t-[Th]}^{t+[Th]}\frac{1}{Th}K\left(\frac{s-t}{Th}\right)\varepsilon_s(u)\right]\left[\sum_{l=t-[Th]}^{t+[Th]}\frac{1}{Th}K\left(\frac{l-t}{Th}\right)(\phi_l(u)-\phi_t(u))\right]^*\right\}W(u)du$$
$$=Q_{41}+Q_{42}+Q_{43}.$$

Following analogous reasoning to the proof of Lemma 1, we can show $E(Q_{41})=O(h^{-1/2})$ and $\text{var}(Q_{41})=O(1)$. In addition,

$$|Q_{42}|\le\sup_{u\in\mathbb{R}^{dm},|s-t|\le Th}\left|\phi_s(u)-\phi_t(u)\right|^2h^{1/2}\sum_{t=1}^{T}\left|\frac{1}{Th}\sum_{s=t-[Th]}^{t+[Th]}K\left(\tfrac{s-t}{Th}\right)\right|^2\int W(u)du=o(Th^{1/2})$$

because $\sup_{u\in\mathbb{R}^{dm},|s-t|\le Th}|\phi_s(u)-\phi_t(u)|^2=o(1)$ if t/T belongs to continuity points and to C if t/T belongs to discontinuity points and the number of discontinuity points is finite by Assumption 4(i). Using a similar argument, we can show $Q_{43}=o_P(Th^{1/2})$. Thus, $Q_4=o_P(Th^{1/2})$.

We now show $Q_5 = O_P(Th^{1/2})$. Define $\bar{\psi}_0(u) = E[\hat{\psi}_0(u)]$ and $\tilde{\phi}_0(u) = \int_0^1 \phi(u,\tau)d\tau$, which is equal to $\psi_0(u)$ under $\mathbb{H}_0$, but not under $\mathbb{H}_A$. Then, we could decompose Q_5 as follows:

$$
\begin{aligned}
Q_5 &= h^{1/2}\sum_{t=1}^{T}\int \left|\hat{\psi}_0(u)-\bar{\psi}_0(u)\right|^2 W(u)du + h^{1/2}\sum_{t=1}^{T}\int \left|\bar{\psi}_0(u)-\tilde{\phi}_0(u)\right|^2 W(u)du \\
&\quad + h^{1/2}\sum_{t=1}^{T}\int \left|\tilde{\phi}_0(u)-\phi_t(u)\right|^2 W(u)du + 2h^{1/2}\sum_{t=1}^{T}\int Re\left\{\left[\hat{\psi}_0(u)-\bar{\psi}_0(u)\right]\left[\bar{\psi}_0(u)-\tilde{\phi}_0(u)\right]^*\right\} W(u)du \\
&\quad + 2h^{1/2}\sum_{t=1}^{T}\int Re\left\{\left[\hat{\psi}_0(u)-\bar{\psi}_0(u)\right]\left[\tilde{\phi}_0(u)-\phi_t(u)\right]^*\right\} W(u)du \\
&\quad + 2h^{1/2}\sum_{t=1}^{T}\int Re\left\{\left[\bar{\psi}_0(u)-\tilde{\phi}_0(u)\right]\left[\tilde{\phi}_0(u)-\phi_t(u)\right]^*\right\} W(u)du \\
&= Q_{51}+Q_{52}+Q_{53}+Q_{54}+Q_{55}+Q_{56}.
\end{aligned}
$$

Following analogous analysis of Q_{21}, we can show $Q_{51} = O_P(h^{1/2}) = o_P(1)$. Since $\bar{\psi}_0(u)$ is the Riemann sum approximation of $\tilde{\phi}_0(u)$ and by Assumption 4(i) the number of discontinuous points of $\phi_t(u) = \phi(u, t/T)$ is finite, we know $Q_{52} = O(T^{-1}h^{1/2}) = o(1)$. By Assumption 4(ii), we have that for sufficient large T,

$$
\begin{aligned}
T^{-1}h^{-1/2}Q_{53} &= \frac{1}{T}\sum_{t=1}^{T}\int \left|\tilde{\phi}_0(u)-\phi_t(u)\right|^2 W(u)du \\
&\geq \frac{1}{T}\sum_{t=1}^{T}\int \inf_{\tilde{\phi}(u)\in\Phi(u)} \left|\tilde{\phi}(u)-\phi_t(u)\right|^2 W(u)du \\
&\geq c_\phi C > 0,
\end{aligned}
$$

for some positive constant C. In addition, we can easily show that $Q_{54} = O_P(T^{-1/2}h^{1/2})$, $Q_{55} = O_P(T^{1/2}h^{1/2})$, and $Q_{56} = O_P(h^{1/2})$. Thus, $Q_5 = O_P(Th^{1/2})$. Using a similar argument, we can also show $Q_6 = o_P(Th^{1/2})$. Moreover, we can easily show $\hat{B} = O_P(h^{-1/2})$ and $\hat{V} = O_P(1)$ under the alternative hypothesis. It follows that

$$
\frac{\widehat{SQ}}{Th^{1/2}} = T^{-1}h^{-1/2}\hat{V}^{-1/2}(Q_4+Q_5+Q_6-\hat{B}) \geq \hat{V}^{-1/2}c_\phi C/2
$$

with probability approaching 1. Consequently, $P(\widehat{SQ} > M_T) \to 1$ for any $M_T = o(Th^{1/2})$ as $T \to \infty$. ■

PROOF OF THEOREM 3. Under the class of local alternatives $\mathbb{H}_2(a_T): \phi_t(u) = \psi_0(u) + a_T\delta(u, t/T)$, we can decompose $Th^{1/2}\hat{Q}$ as follows:

$$
\begin{aligned}
Th^{1/2}\hat{Q} &= h^{1/2}\sum_{t=1}^{T}\left|\left[\hat{\phi}_t(u)-\bar{\phi}_t(u)\right]-\left[\hat{\psi}_0(u)-\bar{\psi}_0(u)\right]+\left[\bar{\phi}_t(u)-\phi_t(u)\right]-\left[\bar{\psi}_0(u)-\psi_0(u)\right]+\left[\phi_t(u)-\psi_0(u)\right]\right|^2 W(u)du \\
&= h^{1/2}\sum_{t=1}^{T}\int \left|\left[\hat{\phi}_t(u)-\bar{\phi}_t(u)\right]-\left[\hat{\psi}_0(u)-\bar{\psi}_0(u)\right]\right|^2 W(u)du \\
&\quad + h^{1/2}\sum_{t=1}^{T}\int \left|\left[\bar{\phi}_t(u)-\phi_t(u)\right]-\left[\bar{\psi}_0(u)-\psi_0(u)\right]\right|^2 W(u)du + h^{1/2}\sum_{t=1}^{T}\int \left|\phi_t(u)-\psi_0(u)\right|^2 W(u)du
\end{aligned}
$$

$$
\begin{aligned}
&+ 2h^{1/2}\sum_{t=1}^{T}\int Re\left\{\left[(\hat{\phi}_t(u)-\bar{\phi}_t(u))-(\hat{\psi}_0(u)-\bar{\psi}_0(u))\right]\left[(\bar{\phi}_t(u)-\phi_t(u))-(\bar{\psi}_0(u)-\psi_0(u))\right]^*\right\}W(u)du\\
&+ 2h^{1/2}\sum_{t=1}^{T}\int Re\left\{\left[(\hat{\phi}_t(u)-\bar{\phi}_t(u))-(\hat{\psi}_0(u)-\bar{\psi}_0(u))\right](\phi_t(u)-\psi_0(u))^*\right\}W(u)du\\
&+ 2h^{1/2}\sum_{t=1}^{T}\int Re\left\{\left[(\bar{\phi}_t(u)-\phi_t(u))-(\bar{\psi}_0(u)-\psi_0(u))\right](\phi_t(u)-\psi_0(u))^*\right\}W(u)du\\
&= J_1+J_2+J_3+J_4+J_5+J_6,\ \text{say,}
\end{aligned}
$$

where J_1 determines the asymptotic centering factor and scaling factor under $\mathbb{H}_0$. Following the proof of Theorem 1, we can show that $(J_1 - B)\sqrt{V} \xrightarrow{d} N(0,1)$ as $T \to \infty$ under $\mathbb{H}_2(a_T)$.

Now, we consider the term J_2. We could further decompose J_2 as follows:

$$
\begin{aligned}
J_2 &= h^{1/2}\sum_{t=1}^{T}\int\left|\bar{\phi}_t(u)-\phi_t(u)\right|^2 W(u)du + h^{1/2}\sum_{t=1}^{T}\int\left|\bar{\psi}_0(u)-\psi_0(u)\right|^2 W(u)du\\
&\quad -2h^{1/2}\sum_{t=1}^{T}\int Re\left\{\left[\bar{\phi}_t(u)-\phi_t(u)\right]\left[\bar{\psi}_0(u)-\psi_0(u)\right]^*\right\}W(u)du\\
&= J_{21}+J_{22}+J_{23}.
\end{aligned}
$$

Since $\bar{\phi}_t(u)-\phi_t(u)=\frac{1}{2}a_Th^2C_k\frac{\partial^2\delta(u,\tau)}{\partial\tau^2}+O(a_Th^3)$ and $\bar{\psi}_0(u)-\psi_0(u)=\frac{a_T}{T}\sum_{t=1}^{T}\delta(u,t/T)\to a_T\int_0^1\delta(u,\tau)d\tau$ uniformly in $u\in\mathbb{R}^{dm}$, we have $J_{21}=O(a_T^2Th^{1/2}h^4)=O(h^4)$, $J_{23}=O(Th^{1/2}a_T^2h^2)=O(h^2)$, and $J_{22}\xrightarrow{p}\int[\int_0^1\delta(u,\tau)d\tau]^2W(u)du$.

By the weak law of large number, we have, as $n\to\infty$, $J_3\xrightarrow{p}\int\int_0^1|\delta(u,\tau)|^2W(u)dud\tau$.

Moreover, we decompose J_4 as follows:

$$
\begin{aligned}
J_4 &= 2T^{-1}h^{-1/2}K(0)\sum_{t=1}^{T}\int Re\left\{\varepsilon_t(u)\left[(\bar{\phi}_t(u)-\phi_t(u))-(\bar{\psi}_0(u)-\psi_0(u))\right]^*\right\}W(u)du\\
&\quad + 2T^{-1}h^{-1/2}\sum_{t\neq s}K\left(\frac{s-t}{Th}\right)\int Re\left\{\varepsilon_s(u)\left[(\bar{\phi}_t(u)-\phi_t(u))-(\bar{\psi}_0(u)-\psi_0(u))\right]^*\right\}W(u)du\\
&\quad -2T^{-1}h^{1/2}\sum_{t,s=1}^{T}\int Re\left\{\varepsilon_s(u)\left[(\bar{\phi}_t(u)-\phi_t(u))-(\bar{\psi}_0(u)-\psi_0(u))\right]^*\right\}W(u)du\\
&= J_{41}+J_{42}+J_{43}.
\end{aligned}
$$

It is straightforward to show $E(J_{41})=0$, $E(J_{42})=0$, $E(J_{43})=0$, and $E[J_{41}^2]=O(a_T^2T^{-1}h^{-1})=o(1)$, $E[J_{42}^2]=O(a_T^2h^{\delta/(1+\delta)})$, $E[J_{43}^2]=O(a_T^2h)=o(1)$. Hence, we have $J_{41}=o_P(1)$, $J_{42}=o_P(1)$, and $J_{43}=o_P(1)$ by Chebyshev's inequality.

By a similar argument, we can also show $J_5=o_P(1)$. Now, we consider the term J_6. Since $\bar{\phi}_t(u)-\phi_t(u)=\frac{1}{2}a_Th^2C_k\frac{\partial^2\delta(u,\tau)}{\partial\tau^2}+O(a_Th^3)$ and $\bar{\psi}_0(u)-\psi_0(u)=\frac{1}{T}\sum_{t=1}^{T}a_T\delta_t(u)$, we have $J_6\xrightarrow{p}-2\int[\int_0^1\delta(u,\tau)d\tau]^2W(u)du$. Therefore, we have $E[Th^{1/2}\hat{Q}]=B+\int\int_0^1|\delta(u,\tau)|^2W(u)d\tau du-\int[\int_0^1\delta(u,\tau)d\tau]^2W(u)du$. In addition, under $\mathbb{H}_2(a_T)$, the asymptotic variance of $\widehat{SQ}$, $\text{avar}(\widehat{SQ})=V$. Consequently, we obtain the desired result of Theorem 3. ■

TABLE A.1

SIZE OF STRICT STATIONARITY TESTS UNDER DGP.S1–S3 FOR DIFFERENT SCALE PARAMETER b

		$b=1$								$b=4$							
		$\widehat{SQ}_{W,m=1}^{AS}$		$\widehat{SQ}_{W,m=1}^{BS}$		$\widehat{SQ}_{W,m=2}^{AS}$		$\widehat{SQ}_{W,m=2}^{BS}$		$\widehat{SQ}_{W,m=1}^{AS}$		$\widehat{SQ}_{W,m=1}^{BS}$		$\widehat{SQ}_{W,m=2}^{AS}$		$\widehat{SQ}_{W,m=2}^{BS}$	
		5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
DGP.S1	$T=100$	0.034	0.056	0.042	0.084	0.057	0.083	0.052	0.092	.0015	0.041	0.040	0.102	0.019	0.036	0.042	0.092
	$T=200$	0.044	0.073	0.058	0.088	0.071	0.109	0.048	0.112	0.024	0.039	0.050	0.108	0.035	0.058	0.050	0.098
	$T=500$	0.040	0.074	0.052	0.100	0.079	0.109	0.054	0.084	0.023	0.040	0.028	0.086	0.025	0.039	0.042	0.092
	$T=1000$	0.037	0.055	0.036	0.084	0.070	0.107	0.036	0.082	0.029	0.049	0.052	0.092	0.024	0.047	0.030	0.082
DGP.S2	$T=100$	0.042	0.090	0.036	0.086	0.035	0.066	0.038	0.088	0.022	0.051	0.042	0.122	0.007	0.016	0.038	0.100
	$T=200$	0.043	0.099	0.042	0.080	0.039	0.079	0.036	0.086	0.028	0.053	0.042	0.090	0.010	0.027	0.038	0.092
	$T=500$	0.068	0.117	0.054	0.106	0.050	0.089	0.048	0.098	0.024	0.049	0.040	0.100	0.008	0.021	0.054	0.124
	$T=1000$	0.058	0.098	0.044	0.112	0.047	0.083	0.038	0.086	0.046	0.075	0.048	0.096	0.081	0.124	0.058	0.142
DGP.S3	$T=100$	0.088	0.132	0.054	0.124	0.078	0.138	0.056	0.130	0.062	0.108	0.106	0.018	0.060	0.089	0.076	0.146
	$T=200$	0.093	0.145	0.058	0.116	0.121	0.167	0.064	0.122	0.068	0.106	0.102	0.168	0.075	0.114	0.088	0.156
	$T=500$	0.088	0.144	0.058	0.104	0.130	0.187	0.054	0.116	0.078	0.114	0.094	0.146	0.078	0.133	0.078	0.146
	$T=1000$	0.088	0.140	0.050	0.094	0.156	0.229	0.050	0.100	0.056	0.106	0.082	0.152	0.080	0.124	0.058	0.142

NOTES: (i) $\widehat{SQ}_W^{AS}$ and $\widehat{SQ}_W^{BS}$ denote the results of $\widehat{SQ}_W$ using asymptotic and bootstrap critical values respectively; (ii) T_K denotes Kapetanios' (2009) test; (iii) the results using asymptotic critical values are based on 1000 iterations, whereas the bootstrap results are based on 500 iterations.

TABLE A.2

POWER OF STRICT STATIONARITY TESTS UNDER DGP.P1–P7 FOR DIFFERENT SCALE PARAMETER b

		$b=1$								$b=4$							
		$\widehat{SQ}_{W,m=1}^{AS}$		$\widehat{SQ}_{W,m=1}^{BS}$		$\widehat{SQ}_{W,m=2}^{AS}$		$\widehat{SQ}_{W,m=2}^{BS}$		$\widehat{SQ}_{W,m=1}^{AS}$		$\widehat{SQ}_{W,m=1}^{BS}$		$\widehat{SQ}_{W,m=2}^{AS}$		$\widehat{SQ}_{W,m=2}^{BS}$	
		5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
DGP.P1	$T=100$	0.821	0.908	0.850	0.916	0.798	0.889	0.830	0.894	0.744	0.843	0.792	0.862	0.647	0.766	0.760	0.838
	$T=200$	0.985	0.991	0.970	0.992	0.982	0.990	0.962	0.982	0.964	0.981	0.950	0.970	0.944	0.965	0.936	0.958
	$T=500$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	$T=1000$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
DGP.P2	$T=100$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	$T=200$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	$T=500$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	$T=1000$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
DGP.P3	$T=100$	0.941	0.964	0.918	0.970	0.948	0.970	0.868	0.946	0.664	0.771	0.758	0.862	0.474	0.623	0.598	0.744
	$T=200$	1.00	1.00	0.998	0.998	1.00	1.00	1.00	1.00	0.982	0.995	0.984	0.994	0.922	0.960	0.936	0.976
	$T=500$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	$T=1000$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
DGP.P4	$T=100$	0.440	0.569	0.500	0.676	0.754	0.851	0.638	0.804	0.436	0.535	0.566	0.708	0.537	0.638	0.610	0.740
	$T=200$	0.921	0.960	0.934	0.970	0.989	0.996	0.966	0.986	0.910	0.946	0.924	0.964	0.944	0.979	0.950	0.974
	$T=500$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	$T=1000$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
DGP.P5	$T=100$	0.297	0.431	0.362	0.528	0.567	0.659	0.470	0.604	0.276	0.371	0.374	0.526	0.361	0.434	0.400	0.544
	$T=200$	0.739	0.816	0.762	0.884	0.925	0.950	0.844	0.920	0.707	0.791	0.774	0.872	0.780	0.853	0.792	0.882
	$T=500$	0.999	0.999	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	$T=1000$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
DGP.P6	$T=100$	0.243	0.313	0.256	0.400	0.359	0.456	0.210	0.330	0.330	0.427	0.434	0.562	0.376	0.463	0.422	0.560
	$T=200$	0.492	0.578	0.500	0.632	0.627	0.703	0.426	0.556	0.761	0.826	0.816	0.886	0.791	0.840	0.792	0.860
	$T=500$	0.899	0.940	0.916	0.956	0.960	0.972	0.866	0.924	0.998	1.00	1.00	1.00	0.998	1.00	1.00	1.00
	$T=1000$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
DGP.P7	$T=100$	0.118	0.175	0.112	0.184	0.256	0.338	0.172	0.258	0.048	0.091	0.110	0.166	0.091	0.134	0.144	0.212
	$T=200$	0.141	0.194	0.128	0.212	0.396	0.498	0.262	0.398	0.066	0.118	0.108	0.188	0.210	0.301	0.262	0.388
	$T=500$	0.175	0.246	0.166	0.240	0.818	0.883	0.680	0.792	0.094	0.130	0.118	0.198	0.562	0.678	0.532	0.716
	$T=1000$	0.208	0.286	0.228	0.308	0.994	0.997	0.952	0.986	0.110	0.180	0.126	0.240	0.921	0.962	0.896	0.946

NOTES: (i) $\widehat{SQ}_W^{AS}$ and $\widehat{SQ}_W^{BS}$ denote the results of $\widehat{SQ}_W$ using asymptotic and bootstrap critical values respectively; (ii) T_K denotes Kapetanios' (2009) test; and (iii) the results using asymptotic critical values are based on 1000 iterations, whereas the bootstrap results are based on 500 iterations.

TABLE A.3

REJECT RATIOS OF STRICT STATIONARITY TESTS FOR HEAVY-TAILED AND SKEWED ERROR DISTRIBUTIONS

		Student's *t* Distribution								Noncentral $t(5, 1)$ Distribution							
		$\widehat{SQ}^{AS}_{W,m=1}$		$\widehat{SQ}^{BS}_{W,m=1}$		$\widehat{SQ}^{AS}_{W,m=2}$		$\widehat{SQ}^{BS}_{W,m=2}$		$\widehat{SQ}^{AS}_{W,m=1}$		$\widehat{SQ}^{BS}_{W,m=1}$		$\widehat{SQ}^{AS}_{W,m=2}$		$\widehat{SQ}^{BS}_{W,m=2}$	
		5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
DGP.S1	$T = 100$	0.018	0.040	0.048	0.100	0.022	0.044	0.040	0.092	0.034	0.070	0.076	0.114	0.098	0.126	0.066	0.124
	$T = 200$	0.032	0.054	0.054	0.110	0.042	0.074	0.048	0.102	0.036	0.068	0.046	0.106	0.120	0.158	0.076	0.132
	$T = 500$	0.012	0.032	0.024	0.068	0.066	0.118	0.034	0.072	0.028	0.042	0.036	0.084	0.090	0.138	0.040	0.084
	$T = 1000$	0.042	0.060	0.044	0.100	0.124	0.176	0.048	0.106	0.040	0.086	0.064	0.120	0.124	0.200	0.074	0.114
DGP.S2	$T = 100$	0.040	0.082	0.040	0.098	0.020	0.078	0.038	0.104	0.048	0.094	0.028	0.086	0.044	0.086	0.016	0.076
	$T = 200$	0.050	0.098	0.034	0.090	0.038	0.090	0.038	0.094	0.056	0.132	0.034	0.098	0.052	0.116	0.032	0.106
	$T = 500$	0.058	0.096	0.052	0.100	0.038	0.066	0.056	0.098	0.064	0.134	0.068	0.118	0.076	0.130	0.050	0.114
	$T = 1000$	0.044	0.104	0.050	0.096	0.045	0.080	0.036	0.104	0.064	0.080	0.030	0.090	0.056	0.074	0.038	0.086
DGP.P1	$T = 100$	0.808	0.898	0.824	0.894	0.774	0.878	0.820	0.888	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	$T = 200$	0.972	0.980	0.960	0.976	0.970	0.976	0.958	0.974	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	$T = 500$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	$T = 1000$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
DGP.P2	$T = 100$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	$T = 200$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	$T = 500$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	$T = 1000$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
DGP.P3	$T = 100$	0.846	0.892	0.860	0.932	0.846	0.890	0.766	0.872	0.828	0.884	0.854	0.914	0.826	0.888	0.752	0.868
	$T = 200$	0.996	0.996	0.996	0.998	0.994	0.996	0.990	0.994	0.994	0.996	0.990	0.996	0.996	0.998	0.980	0.996
	$T = 500$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	$T = 1000$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
DGP.P4	$T = 100$	0.384	0.496	0.472	0.618	0.666	0.742	0.570	0.708	0.726	0.814	0.774	0.862	0.838	0.888	0.766	0.868
	$T = 200$	0.878	0.928	0.898	0.956	0.968	0.980	0.940	0.972	0.980	0.990	0.982	0.990	0.996	0.996	0.986	0.992
	$T = 500$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	$T = 1000$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
DGP.P5	$T = 100$	0.274	0.368	0.348	0.478	0.526	0.604	0.416	0.566	0.546	0.650	0.624	0.762	0.718	0.776	0.628	0.756
	$T = 200$	0.638	0.746	0.698	0.810	0.862	0.908	0.778	0.862	0.934	0.962	0.940	0.972	0.978	0.984	0.948	0.978
	$T = 500$	0.994	0.994	0.994	0.996	0.998	1.00	0.996	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	$T = 1000$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

NOTES: (i) $\widehat{SQ}^{AS}_W$ and $\widehat{SQ}^{BS}_W$ denote the results of $\widehat{SQ}_W$ using asymptotic and bootstrap critical values respectively; (ii) T_K denotes Kapetanios' (2009) test; and (iii) the results are based on 500 iterations.

TABLE A.4

REJECT RATIOS OF WEAK STATIONARITY TESTS FOR HEAVY-TAILED AND SKEWED ERROR DISTRIBUTIONS

		Student's $t(5)$ Distribution										Noncentral $t(5,1)$ Distribution									
		$\widehat{SQ}_{AS}^{(1)}$		$\widehat{SQ}_{BS}^{(1)}$		$\widehat{SQ}_{AS}^{(2)}$		$\widehat{SQ}_{BS}^{(2)}$		$\widehat{SQ}(2)$		$\widehat{SQ}_{AS}^{(1)}$		$\widehat{SQ}_{BS}^{(1)}$		$\widehat{SQ}_{AS}^{(2)}$		$\widehat{SQ}_{BS}^{(2)}$		$\widehat{SQ}(2)$	
		5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
DGP.S1	$T=100$	0.018	0.038	0.044	0.104	0.014	0.034	0.036	0.080	0.052	0.094	0.030	0.062	0.058	0.126	0.022	0.046	0.046	0.100	0.038	0.114
	$T=200$	0.048	0.078	0.056	0.126	0.048	0.074	0.062	0.120	0.052	0.094	0.052	0.088	0.068	0.120	0.038	0.064	0.058	0.100	0.060	0.096
	$T=500$	0.030	0.064	0.032	0.084	0.050	0.082	0.082	0.136	0.052	0.122	0.040	0.062	0.050	0.090	0.038	0.068	0.066	0.106	0.052	0.130
	$T=1000$	0.058	0.084	0.058	0.104	0.032	0.068	0.052	0.096	0.058	0.126	0.062	0.100	0.068	0.110	0.026	0.052	0.042	0.088	0.054	0.094
DGP.S2	$T=100$	0.042	0.096	0.058	0.136	0.018	0.044	0.058	0.100	0.056	0.134	0.054	0.098	0.080	0.154	0.022	0.040	0.054	0.102	0.064	0.148
	$T=200$	0.050	0.090	0.058	0.096	0.024	0.050	0.054	0.124	0.062	0.130	0.046	0.108	0.056	0.110	0.024	0.054	0.054	0.104	0.068	0.116
	$T=500$	0.054	0.106	0.060	0.100	0.040	0.066	0.064	0.106	0.066	0.120	0.064	0.098	0.060	0.094	0.030	0.052	0.052	0.094	0.044	0.106
	$T=1000$	0.056	0.098	0.050	0.094	0.054	0.074	0.054	0.110	0.068	0.112	0.042	0.092	0.042	0.084	0.044	0.072	0.054	0.100	0.028	0.076
DGP.P1	$T=100$	0.770	0.892	0.828	0.896	0.236	0.444	0.304	0.488	0.520	0.784	1.00	1.00	1.00	1.00	1.00	1.00	0.998	1.00	0.984	1.00
	$T=200$	0.974	0.982	0.966	0.976	0.724	0.806	0.632	0.756	0.824	0.948	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	$T=500$	0.998	1.00	0.998	0.998	0.970	0.982	0.914	0.946	0.994	0.998	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	$T=1000$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
DGP.P2	$T=100$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	$T=200$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	$T=500$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	$T=1000$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
DGP.P3	$T=100$	0.798	0.866	0.780	0.882	0.016	0.030	0.036	0.074	0.258	0.442	0.738	0.810	0.738	0.846	0.030	0.050	0.048	0.110	0.218	0.364
	$T=200$	0.992	0.992	0.988	0.996	0.030	0.052	0.054	0.094	0.554	0.734	0.980	0.990	0.974	0.990	0.036	0.056	0.050	0.088	0.458	0.638
	$T=500$	1.00	1.00	1.00	1.00	0.046	0.074	0.058	0.124	0.936	0.978	1.00	1.00	1.00	1.00	0.034	0.064	0.048	0.114	0.782	0.892
	$T=1000$	1.00	1.00	1.00	1.00	0.044	0.058	0.042	0.074	1.00	1.00	1.00	1.00	1.00	1.00	0.038	0.054	0.050	0.082	0.944	0.978
DGP.P4	$T=100$	0.040	0.072	0.074	0.124	0.486	0.582	0.578	0.708	0.458	0.648	0.450	0.538	0.482	0.628	0.358	0.472	0.476	0.614	0.492	0.678
	$T=200$	0.064	0.092	0.074	0.122	0.824	0.868	0.870	0.916	0.854	0.932	0.830	0.888	0.844	0.918	0.672	0.766	0.736	0.828	0.802	0.906
	$T=500$	0.032	0.068	0.036	0.084	0.958	0.968	0.964	0.982	0.976	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	$T=1000$	0.072	0.108	0.078	0.130	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
DGP.P5	$T=100$	0.064	0.096	0.080	0.146	0.416	0.518	0.520	0.654	0.522	0.664	0.400	0.472	0.444	0.552	0.378	0.478	0.474	0.590	0.484	0.664
	$T=200$	0.080	0.108	0.088	0.136	0.744	0.816	0.794	0.880	0.772	0.882	0.740	0.816	0.750	0.832	0.634	0.726	0.704	0.804	0.728	0.850
	$T=500$	0.058	0.098	0.076	0.116	0.956	0.968	0.960	0.972	0.982	0.990	0.990	0.994	0.988	0.994	0.912	0.948	0.930	0.962	0.968	0.982
	$T=1000$	0.068	0.098	0.064	0.110	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

NOTES: (i) $\widehat{SQ}_{AS}^{(i)}$ and $\widehat{SQ}_{BS}^{(i)}$ denote the results of $\widehat{SQ}^{(i)}$ using asymptotic and bootstrap critical values respectively; (ii) $\widehat{SQ}(2)$ denotes the results of our weakly stationarity test using bootstrap critical values; and (iii) the results are based on 500 iterations.

TABLE A.5
STATIONARITY TESTS FOR MACROECONOMIC TIME SERIES

	Tran.	Period	$\widehat{SQ}_W^{AS}$	$\widehat{SQ}_W^{BS}$	$\widehat{SQ}_{AS}^{(1)}$	$\widehat{SQ}_{BS}^{(1)}$	$\widehat{SQ}_{AS}^{(2)}$	$\widehat{SQ}_{BS}^{(2)}$	T_K	$KPSS$	ADF
Final Sales: Goods (Chained)	3	1947Q1–2014Q1	0.134	0.345	0.894	0.960	0.021	0.100	0.505	> 0.100	< 0.001
Final Sales: Durable (Chained)	3	1947Q1–2014Q1	0.009	0.080	0.898	0.965	0.000	0.000	0.180	> 0.100	< 0.001
Final Sales: Nondurable (Chained)	3	1947Q1–2014Q1	0.098	0.060	0.773	0.660	0.142	0.130	0.930	> 0.100	< 0.001
Real Personal Consump.: Nondurable	3	1947Q1–2014Q1	0.001	0.015	0.288	0.320	0.000	0.015	0.125	> 0.100	< 0.001
Real Personal Consump.: Durable	3	1947Q1–2014Q1	0.066	0.075	0.881	0.850	0.000	0.000	0.545	> 0.100	< 0.001
Real Personal Consump.: Service	3	1947Q1–2014Q1	0.000	0.000	0.000	0.000	0.012	0.040	0.270	0.048	< 0.001
Real Exports of Goods & Services	3	1947Q1–2014Q1	0.000	0.000	0.173	0.180	0.000	0.000	0.180	0.048	< 0.001
Real Imports of Goods & Services	3	1947Q1–2014Q1	0.000	0.000	0.801	0.820	0.000	0.000	0.090	> 0.100	< 0.001
Real Gov. Consump. & Gross Inv.	3	1947Q1–2014Q1	0.002	0.135	0.228	0.465	0.034	0.215	0.120	> 0.100	< 0.001
Real Disp. Personal Income	3	1947Q1–2014Q1	0.000	0.000	0.026	0.030	0.040	0.035	0.500	> 0.100	< 0.001
Real Personal Income Exclud. Transfer	3	1947Q1–2014Q1	0.139	0.220	0.421	0.450	0.101	0.125	0.460	> 0.100	< 0.001
M3	3	1959M1–2013M12	0.000	0.000	0.000	0.000	0.387	0.365	0.010	< 0.010	< 0.001
Monetary Base: Total	3	1959M1–2014M4	0.000	0.000	0.027	0.055	0.168	0.215	0.970	> 0.100	< 0.001
St. Louis Adj. Monetary Base	3	1918M1–2014M4	0.000	0.000	0.000	0.015	0.010	0.075	0.015	0.050	< 0.001
Moody's Corp. Bond Yield: Aaa	3	1919M1–2014M4	0.000	0.000	0.148	0.140	0.000	0.000	0.815	> 0.100	< 0.001
Moody's Corp. Bond Yield: Baa	3	1919M1–2014M4	0.000	0.000	0.409	0.375	0.001	0.030	0.130	0.039	< 0.001
3-Mon. Treas. Bill: Secondary Market	2	1934M1–2014M4	0.000	0.000	0.929	0.970	0.000	0.000	0.895	> 0.100	< 0.001
1-Year Treas. Const. Maturity Rate	2	1953M4–2014M4	0.000	0.000	0.890	0.905	0.000	0.000	0.690	> 0.100	< 0.001
Japan/U.S. Exchange Rate	3	1971M1–2014M5	0.268	0.215	0.689	0.680	0.265	0.205	0.830	> 0.100	< 0.001
Denmark/U.S. Exchange Rate	3	1971M1–2014M5	0.051	0.100	0.375	0.345	0.428	0.460	0.005	> 0.100	< 0.001
U.S./Euro Exchange Rate	3	1999M1–2014M5	0.064	0.055	0.359	0.235	0.126	0.115	0.220	> 0.100	< 0.001
Ave. Duration Unemploy.	2	1948M1–2014M4	0.000	0.000	0.198	0.140	0.000	0.000	0.875	> 0.100	< 0.001
Ave. Weekly Hour: Manufacturing	1	1939M1–2014M4	0.000	0.000	0.062	0.220	0.000	0.000	0.000	0.024	< 0.001
Ave. Hourly Earning: Manufacturing	3	1939M1–2014M4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.019	< 0.001
Ave. Weekly Overtime Hours: Manuf.	1	1956M1–2014M4	0.000	0.000	0.000	0.000	0.001	0.045	0.000	< 0.010	0.031
Housing authorized: Total New Priv.	3	1960M1–2014M3	0.000	0.000	0.676	0.575	0.000	0.000	0.170	> 0.100	< 0.001
Housing start: Midwest	3	1959M1–2014M3	0.006	0.005	0.928	0.840	0.000	0.000	0.930	> 0.100	< 0.001
Housing start: South	3	1959M1–2014M3	0.667	0.635	0.858	0.870	0.081	0.060	0.720	> 0.100	< 0.001
Housing start: West	3	1959M1–2014M3	0.741	0.660	0.856	0.780	0.497	0.370	0.595	> 0.100	< 0.001
Housing start: Northeast	3	1959M1–2014M3	0.035	0.035	0.921	0.715	0.043	0.070	0.115	> 0.100	< 0.001
IP: Consumer Goods	3	1939M1–2014M3	0.000	0.000	0.175	0.220	0.000	0.000	0.290	> 0.100	< 0.001
IP: Durable Consumer Goods	3	1947M1–2014M3	0.002	0.070	0.911	0.955	0.000	0.015	0.005	> 0.100	< 0.001
IP: Nondurable Consumer Goods	3	1947M1–2014M3	0.000	0.000	0.000	0.000	0.126	0.135	0.635	> 0.100	< 0.001

(*Continued*)

TABLE A.5
CONTINUED

	Tran.	Period	$\widehat{SQ}_W^{AS}$	$\widehat{SQ}_W^{BS}$	$\widehat{SQ}_{AS}^{(1)}$	$\widehat{SQ}_{BS}^{(1)}$	$\widehat{SQ}_{AS}^{(2)}$	$\widehat{SQ}_{BS}^{(2)}$	T_K	$KPSS$	ADF
IP: Dur. Manufacturing	3	1972M1–2014M3	0.256	0.525	0.666	0.825	0.509	0.640	0.060	> 0.100	< 0.001
IP: Nondur. Manufacturing	3	1972M1–2014M3	0.013	0.040	0.462	0.555	0.077	0.250	0.170	> 0.100	< 0.001
IP: Final Products	3	1939M1–2014M3	0.000	0.005	0.281	0.540	0.000	0.005	0.000	> 0.100	< 0.001
IP: Material	3	1939M1–2014M3	0.000	0.000	0.359	0.385	0.000	0.010	0.050	> 0.100	< 0.001
IP: Dur. Material	3	1947M1–2014M3	0.000	0.015	0.936	0.985	0.000	0.000	0.070	> 0.100	< 0.001
IP: Nondur. Material	3	1954M1–2014M3	0.000	0.005	0.007	0.110	0.301	0.430	0.025	> 0.100	< 0.001
IP: Manufacturing	3	1972M1–2014M3	0.131	0.305	0.704	0.895	0.229	0.300	0.065	> 0.100	< 0.001
IP: Mining	3	1972M1–2014M3	0.001	0.010	0.798	0.715	0.267	0.160	0.490	> 0.100	< 0.001
Cap. Utili.: Manufacturing	2	1972M1–2014M3	0.000	0.000	0.000	0.000	0.501	0.710	0.000	0.032	< 0.001
IP: Electric and Gas Utilities	3	1972M1–2014M3	0.768	0.550	0.964	0.970	0.216	0.200	0.805	> 0.100	< 0.001
ISM Manufacturing: Inventories Index	1	1948M1–2014M3	0.001	0.010	0.111	0.255	0.000	0.000	0.000	0.036	< 0.001
All Employees: Total Nonfarm	3	1939M1–2014M4	0.000	0.000	0.030	0.135	0.000	0.000	0.015	> 0.100	< 0.001
Civ. Employ.: Pop. Ratio	2	1948M1–2014M4	0.000	0.000	0.375	0.485	0.000	0.000	0.010	0.053	< 0.001
Civ. Labor Force	3	1948M1–2014M4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	< 0.001	< 0.010
Civ. Labor Force: Part. Rate	2	1948M1–2014M4	0.000	0.000	0.003	0.000	0.000	0.000	0.010	< 0.010	< 0.001
Civ. Unemployment Rate	1	1948M1–2014M4	0.000	0.000	0.000	0.000	0.226	0.470	0.000	< 0.010	0.003
Change in Private Inventories	1	1947M1–2014M1	0.000	0.005	0.382	0.565	0.000	0.000	0.000	> 0.100	< 0.001

NOTES: (i) Numbers in the main entries are the p-values; (ii) the data transformation codes are: 1. level of the series; 2. first difference; 3. first difference of the logarithm; (iii) $\widehat{SQ}_W^{AS}$ and $\widehat{SQ}_W^{BS}$ denote the results of $\widehat{SQ}_W$ test using asymptotic and bootstrap critical values respectively; T_K, $KPSS$, and ADF denote Kapetanios' (2009) test, Kwiatkowski et al.'s (1992) test, and Dickey and Fuller's (1979) test, respectively; $\widehat{SQ}_{AS}^{(i)}$ and $\widehat{SQ}_{BS}^{(i)}$ denote the results of $\widehat{SQ}^{(i)}$ test using asymptotic and bootstrap critical values respectively; (iv) p-values of the bootstrap results are based on 200 bootstrap iterations.

B. *Simulation Results Not Reported in the Article.*
B.1. To check the performance of our strict stationarity test with respect to the extra parameter b, we consider different values for b, i.e. $b = 1$ and $b = 4$. Table A.1 reports the size performance of the test under DGP.S1–S3.

B.2. To check the performance of our strict stationarity test with respect to the extra parameter b, we consider different values for b, i.e. $b = 1$ and $b = 4$. Table A.2 reports the power performance of the test under DGP.P1–P7.

B.3. To check the performance of our strict stationarity tests for heavy tailed and skewed error distributions, we consider the error terms $\{\varepsilon_t\}$ that follow student's t distribution with 5 degrees of freedom and the noncentral t distribution with 5 degrees of freedom and noncentrality parameter 1 for DGP.S1–S2 and P1–P5. Table A.3 reports the simulated results.

B.4. To check the performance of our weak stationarity tests for heavy tailed and skewed error distributions, we consider the error terms $\{\varepsilon_t\}$ that follow student's t distribution with 5 degrees of freedom and the noncentral t distribution with 5 degrees of freedom and noncentrality parameter 1 for DGP.S1–S2 and P1–P5. Table A.4 reports the simulated results.

C. *Empirical Results Not Reported in the Article.* Table A.5 reports the simulated results.

REFERENCES

ANDREWS, D. W. K., "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica* 59 (1991), 817–58.

BAE, Y., AND R. M. DE JONE, "Money Demand Function Estimation by Nonlinear Cointegration," *Journal of Applied Econometrics* 22 (2007), 767–93.

BIERENS, H. J., "Higher Order Sample Autocorrelations and the Unit Root Hypothesis," *Journal of Econometrics* 57 (1993), 137–60.

———, AND S. GUO, "Testing Stationarity and Trend Stationarity against the Unit Root Hypothesis," *Econometric Reviews* 12 (1993), 1–32.

BREITUNG, J., AND S. EICKMEIER, "Testing for Structural Breaks in Dynamic Factor Models," *Journal of Econometrics* 163 (2011), 71–84.

BUSETTI, F., AND A. HARVEY, "Tests of Strict Stationarity Based on Quantile Indicators," *Journal of Time Series Analysis* 31 (2010), 435–50.

CAI, Z., "Trending Time-Varying Coefficient Time Series Models with Serially Correlated Errors," *Journal of Econometrics* 136 (2007), 163–88.

CHEN, B., AND Y. HONG, "Characteristic Function-Based Testing for Multifactor Continuous-Time Markov Models via Nonparametric Regression," *Econometric Theory* 26 (2010), 1115–79.

———, AND ———, "Testing for Smooth Structural Changes in Time Series Models via Nonparametric Regression," *Econometrica* 80 (2012), 1157–83.

DAHLHAUS, R., "On the Kullback-Leibler Information Divergence of Locally Stationary Processes," *Stochastic Process Application* 62 (1996), 139–68.

DANGL, T., AND M. HALLING, "Predictive Regressions with Time-varying Coefficients," *Journal of Financial Economics* 106 (2012), 157–81.

DETTE, H., P. PREUSS, AND M. VETTER, "A Measure of Stationarity in Locally Stationary Processes with Applications to Testing," *Journal of the American Statistical Association* 106 (2011), 1113–24.

DICKEY, D. A., AND W. A. FULLER, "Distribution of the Estimators for Autoregressive Time Series with a Unit Root," *Journal of the American Statistical Association* 74 (1979), 427–31.

———, AND ———, "Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root," *Econometrica* 49 (1981), 1057–72.

ENGLE, R. F., AND C. W. GRANGER, "Co-integration and Error Correction: Representation, Estimation, and Testing," *Econometrica* 55 (1987), 251–76.

FAMA, E. F., "Short Term Interest Rates as Predictors of Inflation," *American Economic Review* 65 (1975), 269–82.

———, "Inflation Uncertainty and Expected Returns on Treasury Bills," *Journal of Political Economy* 84 (1976), 427–48.

———, "Interest Rates and Inflation: The Message in the Entrails," *American Economic Review* 65 (1977), 487–96.

———, AND M. R. GIBBONS, "Inflation, Real Returns and Capital Investment," *Journal of Monetary Economics* 9 (1982), 297–323.

FAN, Y., AND Q. LI, "Root-n-Consistent Estimation of Partially Linear Time Series Models," *Journal of Nonparametric Statistics* 64 (1999), 865–90.

FRANCQ, C., AND J. ZAKOÏAN, "Strict Stationarity Testing and Estimation of Explosive and Stationary Generalized Autoregressive Conditional Heteroscedasticity Model," *Econometrica* 80 (2012), 821–61.

GAO, J., AND I. GIJBELS, "Bandwidth Selection in Nonparametric Kernel Testing," *Journal of the American Statistical Association* 103 (2008), 1584–94.

GOYAL, A., AND I. WELCH, "A Comprehensive Look at the Empirical Performance of Equity Premium Prediction," *Review of Financial Studies* 21 (2008), 1455–508.

GRANGER, C. W. J., "Some Properties of Time Series Data and Their Use in Econometric Model Specification," *Journal of Econometrics* 16 (1981), 121–30.

HALL, P., AND T. E. WEHRLY, "A Geometrical Method for Removing Edge Effects from Kernel-Type Nonparametric Regression Estimators," *Journal of the American Statistical Association* 86 (1991), 665–72.

HANSEN, B., "The New Econometrics of Structural Change: Dating Breaks in U.S. Labor Productivity," *Journal of Economic Perspectives* 15 (2001), 117–28.

HJELLVIK, V., Q. YAO, AND D. TJØSTHEIM, "Linearity Testing Using Local Polynomial Approximation," *Journal of Statistical Planning and Inference* 68 (1998), 295–321.

HONG, Y., AND H. WHITE, "Asymptotic Distribution Theory for Nonparametric Entropy Measures of Serial Dependence," *Econometrica* 73 (2005), 837–901.

JOHANSEN, S., "Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models," *Econometrica* 59 (1991), 1551–80.

KAPETANIOS, G., "Testing for Strict Stationarity in Financial Variables," *Journal of Banking & Finance* 33 (2009), 2346–62.

KIM, C. J., AND C. R. NELSON, "Has the US Economy Become More Stable? A Bayesian Approach Based on a Markov-Switching Model of the Business Cycle," *Review of Economics and Statistics* 81 (1999), 608–16.

KIM, T. Y., Z. LUO, AND C. KIM, "The Central Limit Theorem for Degenerate Variable U-Statistics Under Dependence," *Journal of Nonparametric Statistics* 23 (2011), 683–99.

KÜNSCH, H. R., "The Jackknife and the Bootstrap for General Stationary Observations," *Annals of Statistics*, 17 (1989), 1217–41.

KWIATKOWSKI, D., P. C. B. PHILLIPS, P. SCHMIDT, AND Y. SHIN, "Testing the Null Hypothesis of Stationarity Against the Alternative of a Unit Root," *Journal of Econometrics* 54 (1992), 159–178.

LI, Q., AND J. S. RACINE, *Nonparametric Econometrics: Theory and Practice* (Princeton, NJ: Princeton University Press, 2006).

LIMA, L., AND Z. XIAO, "Is There Long Memory in Financial Time Series?" *Applied Financial Economics* 20 (2010), 487–500.

MANKIW, N. G., AND J. A. MIRON, "The Changing Behavior of the Term Structure of Interest Rates," *Quarterly Journal of Economics* C1-2 (1986), 211–28.

MCCONNELL, M., AND G. PEREZ-QUIROS, "Output Fluctuations in the United States: What Has Changed Since the Early 1980's?" *American Economic Review* 90 (2000), 1464–76.

MONTGOMERY, A., V. ZARNOWITZ, R. TSAY, AND G. TIAO, "Forecasting the U.S. Unemployment Rate," *Journal of the American Statistical Association* 93 (1998), 478–93.

NEWEY, W. K., AND K. D. WEST, "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica* 55 (1987), 703–8.

ORBE, S., E. FERREIRA, AND J. RODRIGUEZ-PÓO, "A Nonparametric Method to Estimate Time Varying Coefficients Under Seasonal Constraints," *Journal of Nonparametric Statistics* 12 (2000), 779–806.

———, ———, AND ———, "Nonparametric Estimation of Time Varying Parameters Under Shape Restrictions," *Journal of Econometrics* 126 (2005), 53–77.

PAGAN, A. R., AND G. W. SCHWERT, "Alternative Models for Conditional Stock Volatility," *Journal of Econometrics* 45 (1990), 267–90.

———, AND A. ULLAH, *Nonparametric Econometrics* (New York: Cambridge University Press, 1999).

PARZEN, E., "On Consistent Covariance Estimates of the Spectrum of a Stationary Time Series," *Annals of Mathematical Statistics* 28 (1957), 329–48.

PHILLIPS, P. C. B., "Time Series Regression with a Unit Root," *Econometrica* 55 (1987), 277–301.

———, AND B. E. HANSEN, "Statistical Inference in Instrumental Variables Regression with $I(1)$ Processes," *Review of Economic Studies* 57 (1990), 99–125.

———, AND P. PERRON, "Testing for a Unit Root in Time Series Regression," *Biometrika* 75 (1988), 335–46.

PREUSS, P., M. VETTER, AND H. DETTE, "A Test for Stationarity Based on Empirical Processes," *Bernoulli* 19 (2013), 2715–49.

ROBINSON, P. M., "Nonparametric Estimation of Time-Varying Parameters," in P. Hackl, ed., *Statistical Analysis and Forecasting of Economic Structural Change* (Berlin: Springer, 1989), 253–64.

———, "Time-Varying Nonlinear Regression," in P. Hackl, ed., *Economic Structural Change Analysis and Forecasting* (Berlin: Springer, 1991), 179–90.

ROUSSAS, G. G., AND D. IOANNIDES, "Moment Inequalities for Mixing Sequences of Random Variables," *Stochastic Analysis and Applications* 5 (1987), 61–120.

SU, L., AND H. WHITE, "A Consistent Characteristic Function-Based Test for Conditional Independence," *Journal of Econometrics* 141 (2007), 807–34.

———, AND ———, "A Nonparametric Hellinger Metric Test for Conditional Independence," *Econometric Theory* 24 (2008), 829–64.

SUN, Y., P. C. B. PHILLIPS, AND S. JIN, "Optimal Bandwidth Selection in Heteroskedasticity-Autocorrelation Robust Testing," *Econometrica* 76 (2008), 175–94.

STOCK, J. H., AND M. W. WATSON, "Evidence on Structural Instability in Macroeconomic Time Series Relations," *Journal of Business and Economic Statistics* 14 (1996), 11–30.

———, AND ———, "Forecasting in Dynamic Factor Models Subject to Structural Instability," in J. Castle and N. Shephard, eds., *The Methodology and Practice of Econometrics, A Festschrift in Honour of Professor David F. Hendry* (Oxford: Oxford University Press, 2009), 173–204.

SZÉKELY, G., M. RIZZO, AND N. BAKIROV, "Measuring and Testing Dependence by Correlation of Distances," *Annals of Statistics* 35 (2007), 2769–94.

XIAO, Z., AND L. LIMA, "Testing Covariance Stationarity," *Econometric Reviews* 26 (2007), 643–67.

YOSHIHARA, K., "Limiting Behavior of U-Statistics for Stationary, Absolutely Regular Processes," *Probability Theory and Related Fields* 35 (1976), 237–52.

ZHANG, T., AND W. WU, "Testing Parametric Assumptions of Trends of a Nonstationary Time Series," *Biometrica* 98 (2011), 599–614.

———, AND ———, "Inference of Time Varying Regression Models," *Annals of Statistics* 40 (2012), 1376–402.