Online Supplementary Material for
"Testing for Strict Stationarity via the Discrete Fourier Transform"

Zhonghao Fu[a,b], Shang Gao[a], Liangjun Su[c], and Xia Wang[d]

[a] School of Economics, Fudan University

[b] Shanghai Institute of International Finance and Economics

[c] School of Economics and Management, Tsinghua University

[d] School of Economics, Renmin University of China

This online supplement is composed of three parts. Section S1 provides an alternative method to study the limiting null distribution of our test statistic. Section S2 provides some discussion on the asymptotic pivotality of our test. Section S3 provides some additional simulation and application results.

S1 An Alternative Proof for the Limiting Null Distribution

In Section 3.1, we conjectured that the integrability condition in Assumption A.2 can be discarded with an alternative proof using the theory for V-statistics and the Mercer theorem in functional analysis. In this section, we provide the proof of Theorem 3.1 by restricting our attention to the case where the weighting function $W(u,v)$ can be written as $W_1(w)\,W_2(v)$. Let $\sum_{s,t=1}^{T}=\sum_{s=1}^{T}\sum_{t=1}^{T}$, $\sum_{s,t,r=1}^{T}=\sum_{s=1}^{T}\sum_{t=1}^{T}\sum_{r=1}^{T}$, and $\sum_{s,t,r,m=1}^{T}=\sum_{s=1}^{T}\sum_{t=1}^{T}\sum_{r=1}^{T}\sum_{m=1}^{T}$. Let $\max_s=\max_{1\leq s\leq T}$.

Let $\{V_t\}$ denote a sequence of random variables that are uniformly distributed on $[0,1]$ and independent of the process $\{Y_t\}$. Let $h_{1ts}=h_1(Y_t,Y_s)$ and $h_{2ts}=h_2(V_t,V_s)$, where h_1 and h_2 are as defined in Section 2. We make the following assumptions.

Assumption S.1 (i) $\{Y_t\}$ is a strong mixing process on $\mathbb{R}^{dm}$ with the mixing coefficient $\alpha(\cdot)$ such that $\sum_{s=1}^{\infty}s\alpha(s)^{\delta/(2+\delta)}<\infty$; (ii) $\max_{1\leq t,s\leq T}E(|h_{1ts}|^{2+\delta})<\infty$ and $\max_{1\leq t,s\leq T}E(|h_{2ts}|^{2+\delta})<\infty$ for some $\delta>0$.

Assumption S.2 The weighting function $W(\cdot)$: $\mathbb{R}^{dm+1}\to\mathbb{R}^{+}$ is a nonnegative and symmetric function such that $W(u,v)=W_1(u)W_2(v)$.

Assumptions S.1 and S.2 parallel Assumption A.1 and A.2 in the paper. One major difference is that we do not need to impose integrability on $W(u,v)$. Instead, we assume that the weighting function can be written as the product of two functions. This assumption can be relaxed at the cost of more lengthy derivations on the kernel used in the Mercer theorem. The second major difference is that we do not need to impose any moment conditions on Y_t. Instead we impose some weak moment conditions on h_{1ts} and h_{2ts}. It is easy to see that the moment conditions in Assumption S.2(ii) are satisfied for the weighting functions discussed in the main text even if Y_t does not have any finite first moment. The symmetric properties of W_1 and W_2 ensure that both $h_1(\cdot,\cdot)$ and $h_2(\cdot,\cdot)$ are symmetric functions.

The following theorem presents an alternative representation of the main result in Theorem 3.1.

Theorem S1.1. *Suppose Assumptions S.1 and S.2 hold. Then under $\mathbb{H}_0$ we have*

$$\hat{D} \xrightarrow{d} \sum_{j=0}^{\infty} \lambda_j \mathcal{Z}_j^2,$$

where λ_j's are the positive eigenvalues defined in (S1.2) and (S1.3) below, and $\{\mathcal{Z}_j\}_{j=0}^{\infty}$ is a sequence of zero-mean Gaussian variables with covariance defined in (S1.5) below.

To prove Theorem S1.1, we state the following lemma of Sun and Chiang (1997) which will be used repeatedly.

Lemma S1.2. *Let $\{V_i, i \geq 1\}$ be a v-dimensional strong mixing process with mixing coefficient $\alpha(\cdot)$. Let $F_{i_1,\ldots,i_m}$, denote the distribution function of $(V_{i_1},\ldots,V_{i_m})$. For any integer $m>1$ and integers $(i_1,\ldots,i_m)$ such that $1 \leq i_1 < i_2 < \ldots < i_m$, let θ be a Borel measurable function such that max$\{\int |\theta(v_1,\ldots,v_m)|^{1+\widetilde{\eta}}\, dF_{i_1,\ldots,i_j}(v_1,\ldots,v_j)\, dF_{i_j+1,\ldots,i_m}(v_{j+1},\ldots,v_m), \int |\theta(v_1,\ldots,v_m)|^{1+\widetilde{\eta}}\, dF_{i_1,\ldots,i_m}\}$ $\leq M_n$ for some $\widetilde{\eta}>0$. Then $|\int \theta(v_1,\ldots,v_m)\, dF_{i_1,\ldots,i_m}(v_1,\ldots,v_m) - \int \theta(v_1,\ldots,v_m)\, dF_{i_1,\ldots,i_j}(v_1,\ldots,v_j)$ $dF_{i_j+1,\ldots,i_m}(v_{j+1},\ldots,v_m)| \leq 4M_n^{1/(1+\widetilde{\eta})}\alpha(i_{j+1}-i_j)^{\widetilde{\eta}/(1+\widetilde{\eta})}$.*

Proof of Theorem S1.1.

By (2.8), we have

$$\hat{D} = \frac{1}{T}\sum_{s,t=1}^{T} \tilde{h}_{1st}\tilde{h}_{2st} = \frac{1}{T}\sum_{s,t=1}^{T} \tilde{h}_{1st} h_{2st},$$

where $\tilde{h}_{\ell st} \equiv h_{\ell st} - \frac{1}{T}\sum_{t=1}^{T} h_{\ell st} - \frac{1}{T}\sum_{s=1}^{T} h_{\ell st} + \frac{1}{T^2}\sum_{s,t=1}^{T} h_{\ell st}$ for $\ell = 1,2$, $h_{1st} \equiv h_1(Y_s,Y_t) \equiv \int_{\mathbb{R}^{dm}} e^{\mathbf{i}u'(Y_s-Y_t)} W_1(u)\, \mathrm{d}u$, and $h_{2st} \equiv h_2(\frac{s}{T},\frac{t}{T}) \equiv \int_{\mathbb{R}} e^{\mathbf{i}v2\pi(s-t)/T} W_2(v)\, \mathrm{d}v$. In the last displayed equation, the second equality holds by the fact that $\frac{1}{T}\sum_{s=1}^{T}\tilde{h}_{1st}=\frac{1}{T}\sum_{t=1}^{T}\tilde{h}_{1st}=0$. Because $\tilde{h}_{1st}$ involves double demeaning operator, $\hat{D}$ is a fourth-order V-statistic. Define

$$\breve{h}_{1st}\equiv h_{1st}-\frac{1}{T}\sum_{t=1}^{T}E_t\left(h_{1st}\right)-\frac{1}{T}\sum_{s=1}^{T}E_s\left(h_{1st}\right)+\frac{1}{T^2}\sum_{s,t=1}^{T}E_tE_s\left(h_{1st}\right)$$

where $E_t\left(\cdot\right)$ denotes expectation with respect to t-indexed random variable only. Similarly, let

$$\breve{h}_{2st}\equiv h_2\left(\frac{s}{T},\frac{t}{T}\right)-\int_0^1 h_2\left(\frac{s}{T},\tau\right)\mathrm{d}\tau-\int_0^1 h_2\left(\tilde{\tau},\frac{t}{T}\right)\mathrm{d}\tilde{\tau}+\int_0^1\int_0^1 h_2\left(\tilde{\tau},\tau\right)\mathrm{d}\tilde{\tau}\mathrm{d}\tau.$$

We prove the theorem by showing that

(i) $\hat{D}=\hat{D}_1+o_p(1)$ with $\hat{D}_1=\frac{1}{T}\sum_{s,t=1}^{T}\breve{h}_{1st}\breve{h}_{2st}$,

(ii) $\hat{D}_1\overset{d}{\to}\sum_{j=0}^{\infty}\lambda_j\mathcal{Z}_j^2$.

Step 1. We prove that $\hat{D}=\hat{D}_1+o_p(1)$.

Then

$$\begin{aligned}
\tilde{h}_{1st} &= h_{1st}-\frac{1}{T}\sum_{t=1}^{T}h_{1st}-\frac{1}{T}\sum_{s=1}^{T}h_{1st}+\frac{1}{T^2}\sum_{s,t=1}^{T}h_{1st}\\
&= \breve{h}_{1st}-\frac{1}{T}\sum_{t=1}^{T}[h_{1st}-E_t\left(h_{1st}\right)]-\frac{1}{T}\sum_{s=1}^{T}[h_{1st}-E_s(h_{1st})]+\frac{1}{T^2}\sum_{s,t=1}^{T}[h_{1st}-E_tE_s\left(h_{1st}\right)]
\end{aligned}$$

and

$$\begin{aligned}
\tilde{h}_{2st} &= h_{2st}-\frac{1}{T}\sum_{t=1}^{T}h_{2st}-\frac{1}{T}\sum_{s=1}^{T}h_{2st}+\frac{1}{T^2}\sum_{s,t=1}^{T}h_{2st}\\
&= \breve{h}_{2st}-\left[\frac{1}{T}\sum_{t=1}^{T}h_{2st}-\int_0^1 h_2\left(\frac{s}{T},\tau\right)\mathrm{d}\tau\right]-\left[\frac{1}{T}\sum_{s=1}^{T}h_{2st}-\int_0^1 h_2\left(\tilde{\tau},\frac{t}{T}\right)\mathrm{d}\tilde{\tau}\right]\\
&\quad+\left[\frac{1}{T^2}\sum_{s,t=1}^{T}h_{2st}-\int_0^1\int_0^1 h_2\left(\tilde{\tau},\tau\right)\mathrm{d}\tilde{\tau}\mathrm{d}\tau\right]\\
&\equiv \breve{h}_{2st}-r_{1s}-r_{2t}+r_3.
\end{aligned}$$

Under the null of strict stationarity, it is easy to see that we can rewrite $\breve{h}_{1st}$ as $\breve{h}_{1st}=h_{1st}-E_t\left(h_{1st}\right)$ $-E_s\left(h_{1st}\right)+E_tE_s\left(h_{1st}\right)$ so that $E_t(\breve{h}_{1st})=E_s(\breve{h}_{1st})=0$. Noting that $\frac{1}{T}\sum_{s=1}^{T}\tilde{h}_{2st}=\frac{1}{T}\sum_{t=1}^{T}\tilde{h}_{2st}=$

0, we have

$$\begin{aligned}
\hat{D} &= \frac{1}{T}\sum_{s,t=1}^{T}\tilde{h}_{1st}\tilde{h}_{2st} = \frac{1}{T}\sum_{s,t=1}^{T}\breve{h}_{1st}\tilde{h}_{2st} \\
&= \frac{1}{T}\sum_{s,t=1}^{T}\breve{h}_{1st}\breve{h}_{2st} - \frac{1}{T}\sum_{s,t=1}^{T}\breve{h}_{1st}r_{1s} - \frac{1}{T}\sum_{s,t=1}^{T}\breve{h}_{1st}r_{2s} + \frac{r_3}{T}\sum_{s,t=1}^{T}\breve{h}_{1st} \\
&\equiv \hat{D}_1 - \hat{D}_2 - \hat{D}_3 + \hat{D}_4.
\end{aligned}$$

By the property of Riemann summation approximation to a definite integral, we can readily show that $r_{\ell s} = O\left(T^{-1}\right)$ for $\ell = 1, 2$ uniformly in s and $r_3 = O\left(T^{-1}\right)$. Now

$$\hat{D}_2 = \frac{1}{T}\sum_{s=1}^{T}\breve{h}_{1ss}r_{1s} + \frac{1}{T}\sum_{1\le s<t\le T}\breve{h}_{1st}r_{1s} + \frac{1}{T}\sum_{1\le t<s\le T}\breve{h}_{1st}r_{1s} \equiv \hat{D}_{2,1} + \hat{D}_{2,2} + \hat{D}_{2,3}.$$

Note that $\left|\hat{D}_{2,1}\right| \le \max_s |r_{1s}| \frac{1}{T}\sum_{s=1}^{T}\breve{h}_{1ss} = O(T^{-1})O_p(1) = O_p(T^{-1})$. Next, we study $\hat{D}_{2,2}$. Let $M_{1T} = \max\left\{\max_{1\le s<t\le T} E\left|h_{1st}\right|^{2+\delta}, E_t E_s \left|h_{1st}\right|^{2+\delta}\right\}$. By Lemma S1.2 with $\eta = 1+\delta$,

$$\begin{aligned}
\left|E\left(\hat{D}_{2,2}\right)\right| &= \left|\frac{1}{T}\sum_{1\le s<t\le T} E\left(\breve{h}_{1st}\right) r_{1s}\right| = \left|\frac{1}{T}\sum_{1\le s<t\le T}\left[E\left(h_{1st}\right) - E_t E_s\left(h_{1st}\right)\right] r_{1s}\right| \\
&\le \frac{4M_{1T}^{1/(2+\delta)}}{T}\sum_{1\le s<t\le T}\alpha\left(t-s\right)^{(1+\delta)/(2+\delta)}\left|r_{1s}\right| \\
&\lesssim \sum_{\tau=1}^{T-1}\alpha\left(\tau\right)^{(1+\delta)/(2+\delta)}\max_s\left|r_{1s}\right| = O\left(T^{-1}\right),
\end{aligned}$$

where we use the fact that $M_{1T} = O(1)$. Now, notice that

$$\begin{aligned}
E\left[\left(\hat{D}_{2,2}\right)^2\right] &= \frac{1}{T^2}\sum_{1\le s<t\le T}\sum_{1\le r<m\le T} E\left(\breve{h}_{1st}\breve{h}_{1rm}\right) r_{1s}r_{1r} \\
&= \frac{1}{T^2}\sum_{1\le s<t\le T} E\left(\breve{h}_{1st}^2\right) r_{1s}^2 + \frac{2}{T^2}\sum_{1\le s<t<r\le T} E\left(\breve{h}_{1st}\breve{h}_{1sr}\right) r_{1s}^2 \\
&\quad + \frac{1}{T^2}\sum_{1\le s<t\le T, 1\le r<m\le T, \#\{s,t,r,m\}=4} E\left(\breve{h}_{1st}\breve{h}_{1rm}\right) r_{1s}r_{1r} \\
&\equiv I_{1T} + I_{2T} + I_{3T},
\end{aligned}$$

where $\#A$ denotes the cardinality of set A. Let $M_{2T} \equiv \max\{\max_{1\le s<t<r\le T} E\left|\breve{h}_{1st}\breve{h}_{1sr}\right|^{1+\delta/2}, E_t E_s E_r$

$\left|\breve{h}_{1st}\breve{h}_{1sr}\right|^{1+\delta/2}\}$. It is easy to see that

$$I_{1T} = \frac{1}{T^2}\sum_{1\le s<t\le T} E\left(\breve{h}_{1st}^2\right) r_{1s}^2 \lesssim \max_s r_{1s}^2 = O(T^{-2}),$$

and

$$\begin{aligned} I_{2T} &= \frac{2}{T^2}\sum_{1\le s<t<r\le T} E\left(\breve{h}_{1st}\breve{h}_{1sr}\right) r_{1s}^2 \le \frac{8M_{2T}^{2/(2+\delta)}}{T^2}\sum_{1\le s<t<r\le T}\alpha\left(r-t\right)^{\delta/(2+\delta)}\max_s r_{1s}^2 \\ &\lesssim \sum_{\tau=1}^{T-1}\alpha\left(\tau\right)^{\delta/(2+\delta)}\max_s r_{1s}^2 = O\left(T^{-2}\right). \end{aligned}$$

For I_{3T}, we have

$$I_{3T} = T^{-2}\sum_{1\le t_1<t_2\le T, 1\le t_3<t_4\le T, \#\{t_1,t_2,t_3,t_4\}=4} \breve{h}_{1t_1t_2} r_{1t_1}\breve{h}_{1t_3t_4} r_{1t_3}. \tag{S1.1}$$

Let $1\le k_1<\ldots<k_4\le T$ be the permutation of $t_1,\ldots,t_4$ in ascending order and let d_c be the c-th largest difference among $k_{j+1}-k_j$, for $j=1,2$, and 3. Define

$$H\left(k_1,\ldots,k_4\right) = \breve{h}_{1t_1t_2} r_{1t_1}\breve{h}_{1t_3t_4} r_{1t_3}.$$

Let $M_{3T} = \max\left\{\max_{1\le s,t,r,m\le T,\#\{s,t,r,m\}=4} E\left|\breve{h}_{1st}\breve{h}_{1rm}\right|^{1+\delta/2},\ E_tE_sE_rE_m\left|\breve{h}_{1st}\breve{h}_{1rm}\right|^{1+\delta/2}\right\}$. Since $E_{k_1}\left[H\left(k_1,\ldots,k_4\right)\right]=0$, we have by Lemma S1.2 with $\eta=\delta/2$,

$$\begin{aligned} \sum_{\substack{1\le k_1<\ldots<k_4\le T\\ k_2-k_1=d_1}} \left|E\left[H\left(k_1,\ldots,k_4\right)\right]\right| &\le 4T^{-2}M_T^{2/(2+\delta)}\sum_{k_1=1}^{T-3}\sum_{k_2=k_1+\max_{j\ge3}\{k_j-k_{j-1}\}}^{T-2}\sum_{k_3=k_2+1}^{T-1}\sum_{k_4=k_3+1}^{T}\left[\alpha\left(k_2-k_1\right)\right]^{\delta/(2+\delta)} \\ &\quad\times\left(\max_s r_{1s}\right)^2 \\ &\le 4T^{-2}M_T^{2/(2+\delta)}\sum_{k_1=1}^{T-3}\sum_{k_2=k_1+1}^{T-2}\left(k_2-k_1\right)^2\left[\alpha\left(k_2-k_1\right)\right]^{\delta/(2+\delta)}\left(\max_s r_{1s}\right)^2 \\ &\le 4M_T^{2/(2+\delta)}\sum_{j=1}^{T} j\alpha\left(j\right)^{\delta/(2+\delta)}\left(\max_s r_{1s}\right)^2 = O(T^{-2}). \end{aligned}$$

Similarly, $\sum_{\substack{1\le k_1<\ldots<k_4\le T\\ k_4-k_3=d_1}}\left|E\left[H\left(k_1,\ldots,k_4\right)\right]\right| = O(T^{-3})$ by using $E_{k_4}\left[H\left(k_1,\ldots,k_4\right)\right]=0$. If $k_3-k_2=d_1$, then either $k_2-k_1=d_2$ or $k_4-k_3=d_2$. Then we can use the fact that $E_{k_1}\left[H\left(k_1,\ldots,k_4\right)\right]$

$= 0$ in the first case and $E_{k_4}\left[H\left(k_1,\ldots,k_4\right)\right] = 0$ in the second case to obtain

$$
\begin{aligned}
\sum_{\substack{1\le k_1<\ldots<k_4\le T\\ k_3-k_2=d_1}} \left|E\left[H\left(k_1,\ldots,k_4\right)\right]\right| &\le 4T^{-2}M_T^{2/(2+\delta)} \sum_{\substack{1\le k_1<\ldots<k_4\le T\\ k_3-k_2=d_1,k_2-k_1=d_2}} \left[\alpha\left(k_2-k_1\right)\right]^{\delta/(2+\delta)}\left(\max_s r_{1s}\right)^2\\
&+4T^{-2}M_T^{2/(2+\delta)} \sum_{\substack{1\le k_1<\ldots<k_4\le T\\ k_3-k_2=d_1,k_4-k_3=d_2}} \left[\alpha\left(k_4-k_3\right)\right]^{\delta/(2+\delta)}\left(\max_s r_{1s}\right)^2\\
&\le 8T^{-1}M_T^{2/(2+\delta)}\sum_{k_2=1}^{T-1}\sum_{k_3=k_2+1}^{T}\left(k_2-k_1\right)\left[\alpha\left(k_2-k_1\right)\right]^{\delta/(2+\delta)}\left(\max_s r_{1s}\right)^2\\
&\le 8M_T^{2/(2+\delta)}\sum_{j=1}^{T} j\alpha\left(j\right)^{\delta/(2+\delta)}\left(\max_s r_{1s}\right)^2 = O(T^{-2}).
\end{aligned}
$$

It follows that $I_{3T} = O(T^{-2})$. Consequently, we have shown that $E\left[(\hat{D}_{2,2})^2\right] = O(T^{-2})$. Then $\hat{D}_{2,2} = O_p(T^{-1})$ by the Chebyshev inequality. Similarly, $\hat{D}_{2,3} = O_p(T^{-1})$. It follows that $\hat{D}_2 = O_p(T^{-1})$. Analogously, $\hat{D}_3 = O_p(T^{-1})$ and $\hat{D}_4 = O_p(T^{-1})$. In sum, we have shown that

$$\hat{D} = \hat{D}_1 + O_p(T^{-1}),$$

where $\hat{D}_1 = \frac{1}{T}\sum_{s,t=1}^{T}\breve{h}_{1st}\breve{h}_{2st}$.

Step 2. We prove that $\hat{D}_1 \xrightarrow{d} \sum_{j=0}^{\infty}\lambda_j\mathcal{Z}_j^2$.

Let $\xi_{tT} \equiv (Y_t', t/T)'$. The dependence of ξ_{tT} on T will complicate the asymptotic analysis. But $\{t/T\}_{t=1}^{T}$ behaves like the T realizations of a uniform random variable on the interval $[0,1]$. This motivates us to introduce $\xi_t \equiv (Y_t', V_t)'$, where $\{V_t\}$ are i.i.d. $U[0,1]$ and are independent of the process $\{Y_t\}$. Let $\xi = (y', \tau)'$, $\tilde{\xi} = (\tilde{y}', \tilde{\tau})'$, and $\varphi(\xi,\tilde{\xi}) = \breve{h}_1\left(y,\tilde{y}\right)\breve{h}_2\left(\tau,\tilde{\tau}\right)$, where

$$
\begin{aligned}
\breve{h}_1\left(y,\tilde{y}\right) &= h_1\left(y,\tilde{y}\right) - E\left[h_1\left(y,Y_t\right)\right] - E\left[h_1\left(Y_t,\tilde{y}\right)\right] + E_sE_t\left[h_1\left(Y_t,Y_s\right)\right], \text{ and}\\
\breve{h}_2\left(\tau,\tilde{\tau}\right) &= h_2\left(\tau,\tilde{\tau}\right) - E\left[h_2\left(\tau,V_t\right)\right] - E\left[h_2\left(V_t,\tilde{\tau}\right)\right] + E_sE_t\left[h_2\left(V_t,V_s\right)\right].
\end{aligned}
$$

By construction, both $\breve{h}_1$ and $\breve{h}_2$ are canonical in the sense that they are $E[\breve{h}_1\left(y,Y_t\right)] = E[\breve{h}_1\left(Y_t,y\right)] = 0$ for all y and $E[\breve{h}_2\left(\tau,V_t\right)] = E[\breve{h}_2\left(V_t,\tau\right)] = 0$ for all τ. Note that $\varphi\left(\xi_{tT},\xi_{sT}\right) = \breve{h}_{1st}\breve{h}_{2st}$ and $\hat{D}_1 = \frac{1}{T}\sum_{s,t=1}^{T}\breve{h}_{1st}\breve{h}_{2st} = \frac{1}{T}\sum_{s,t=1}^{T}\varphi\left(\xi_{tT},\xi_{sT}\right)$. When $W\left(u,v\right) = W_1\left(u\right)W_2\left(v\right)$, we have

$$\breve{h}_1\left(y,\tilde{y}\right) = \int\left[e^{\mathbf{i}u'y} - E(e^{\mathbf{i}u'Y_t})\right]\left[e^{-\mathbf{i}u'\tilde{y}} - E(e^{-\mathbf{i}u'Y_t})\right]\mathrm{d}W_1(u), \text{ and}$$

$$\breve{h}_2(\tau, \tilde{\tau}) \quad = \quad \int \left[e^{\mathbf{i}v2\pi\tau} - E(e^{\mathbf{i}v2\pi V_t}) \right] \left[e^{-\mathbf{i}v2\pi\tilde{\tau}} - E(e^{-\mathbf{i}v2\pi V_t}) \right] \mathrm{d}W_2(v).$$

With the above expressions, we can readily show that $\varphi(\xi, \tilde{\xi})$ is a positive semidefinite (p.s.d.) kernel, i.e.,

$$\int \int \varphi(\xi, \tilde{\xi}) g(\xi) g(\tilde{\xi}) \mathrm{d}\xi \mathrm{d}\tilde{\xi} \geq 0$$

for any $g \in L_2(\Xi, F)$ where $L_2(\Xi, F)$ denotes the Hilbert space defined on the support Ξ of ξ_t and F denotes the cumulative distribution function (CDF) of ξ_t. Then by the Mercer theorem (e.g., Vapnik (1998, p.423) or Sun (2005)), $\varphi\left(\xi, \tilde{\xi}\right)$ exhibits the following spectral decomposition

$$\varphi\left(\xi, \tilde{\xi}\right) = \sum_{j=0}^{\infty} \lambda_j \phi_j(\xi) \phi_j\left(\tilde{\xi}\right), \tag{S1.2}$$

where $\left\{\phi_j(\cdot)\right\}_{j=0}^{\infty}$ denotes the orthonormal eigenfunctions with $E\left[\phi_j(\xi_t)\right] = 0$, and $E[\phi_j(\xi_t)\phi_k(\xi_t)]$ $= \delta_{j,k}$ where $\delta_{j,k} = \mathbf{1}(j = k)$, and $\{\lambda_j\}_{j=1}^{\infty}$ the corresponding nonnegative eigenvalues of the integral equation

$$\int_0^1 \int_{\mathbb{R}^{dm}} \varphi(\xi, \tilde{\xi}) \phi_j(\tilde{y}, \tilde{\tau}) \mathrm{d}F_Y(\tilde{y}) \mathrm{d}\tilde{\tau} \ = \lambda_j \phi_j(\xi), \tag{S1.3}$$

where F_Y denotes the marginal CDF of Y_t. Here, without loss of generality, we can restrict λ_j's to be strictly positive in the above expansion as they will not contribute to the summation otherwise. Indeed, for any $\lambda_j > 0$, we have by (S1.3),

$$\phi_j(\xi_t) = \frac{1}{\lambda_j} \int_0^1 \int_{\mathbb{R}^{dm}} \varphi\left(\xi_t, \tilde{\xi}\right) \phi_j(\tilde{y}, \tilde{\tau}) \mathrm{d}F_Y(\tilde{y}) \mathrm{d}\tilde{\tau} = \frac{1}{\lambda_j} E_s\left[\varphi(\xi_t, \xi_s) \phi_j(\xi_s)\right].$$

Taking expectations on both sides and applying Fubini theorem, we obtain

$$E\left[\phi_j(\xi_t)\right] = \frac{1}{\lambda_j} E_t E_s\left[\varphi(\xi_t, \xi_s) \phi_j(\xi_s)\right] = \frac{1}{\lambda_j} E_s\left[E_t\left[\varphi(\xi_t, \xi_s)\right] \phi_j(\xi_s)\right] = 0,$$

where the last equality holds by the fact that $E_t\left[\varphi(\xi_t, \xi_s)\right] = 0$. By the fact that $\varphi(\xi, \tilde{\xi}) = \breve{h}_1(y, \tilde{y}) \breve{h}_2(\tau, \tilde{\tau})$ and both $\breve{h}_1$ and $\breve{h}_2$ are canonical, we have

$$E\left[\varphi\left((y, V_t), \tilde{\xi}\right)\right] = E\left[\breve{h}_1(y, \tilde{y}) \breve{h}_2(V_t, \tilde{\tau})\right] = 0 \text{ for any nonrandom } y \text{ and } \tilde{\xi},$$

and

$$E\left[\varphi\left((Y_t, \tau), \tilde{\xi}\right)\right] = E\left[\breve{h}_1(Y_t, \tilde{y}) \breve{h}_2(\tau, \tilde{\tau})\right] = 0 \text{ for any nonrandom } \tau \text{ and } \tilde{\xi}.$$

It follows that

$$E\left[\phi_j(y,V_t)\right]=\frac{1}{\lambda_j}E_tE_s\left[\varphi\left((y,V_t),\xi_s\right)\phi_j(\xi_s)\right]=\frac{1}{\lambda_j}E_s\left[E_t\left[\varphi\left((y,V_t),\xi_s\right)\right]\phi_j(\xi_s)\right]=0 \text{ for any } y,$$

and

$$E\left[\phi_j(Y_t,\tau)\right]=\frac{1}{\lambda_j}E_tE_s\left[\varphi\left((Y_t,\tau),\xi_s\right)\phi_j(\xi_s)\right]=\frac{1}{\lambda_j}E_s\left[E_t\left[\varphi\left((Y_t,\tau),\xi_s\right)\right]\phi_j(\xi_s)\right]=0 \text{ for any } \tau.$$

In particular $E\left[\phi_j\left(\xi_{tT}\right)\right]=E[\phi_j(Y_t,\frac{t}{T})]=0$ for all j and t. In addition, the series in (S1.2) converges absolutely and the eigenvalues are summable. Difficulty lies in the fact that we observe ξ_{tT} but not ξ_t and that $E\left[\phi_j\left(\xi_{tT}\right)\right]\neq 0$. In fact, ξ_{tT} is not even stationary despite the stationarity of ξ_t. Then we have

$$\begin{aligned}\hat{D}_1 &= \frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T}\sum_{j=0}^{N}\lambda_j\phi_j\left(\xi_{tT}\right)\phi_{jT}\left(\xi_{sT}\right)+\frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T}\sum_{j=N+1}^{\infty}\lambda_j\phi_j\left(\xi_{tT}\right)\phi_j\left(\xi_{sT}\right)\\ &\equiv \hat{D}_{1,1}^N+\hat{D}_{1,2}^N,\end{aligned}$$

where $\hat{D}_{1,1}^N$ is the leading term and $\hat{D}_{1,2}$ is the remainder term for a well chosen integer N.

First, we show that $\hat{D}_{1,2}^N=o_p(1)$. Noting that $\hat{D}_{1,2}^N=\frac{1}{T}\sum_{j=N+1}^{\infty}\lambda_j\left[\sum_{t=1}^{T}\phi_j\left(\xi_{tT}\right)\right]^2\geq 0$, we can readily apply Lemma S1.2

$$E\left|\hat{D}_{1,2,1}^N\right|=\sum_{j=N+1}^{\infty}\lambda_j\frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T}E\left[\phi_j\left(\xi_{tT}\right)\phi_j\left(\xi_{sT}\right)\right]\lesssim\sum_{j=N+1}^{\infty}\lambda_j=o(1) \text{ as } N\to\infty.$$

Then $\hat{D}_{1,2}^N=o_p(1)$ as $N\to\infty$.

Next, by the fact that $E\left[\phi_j\left(\xi_{tT}\right)\right]=0$ for each j and t and the summability of $\{\lambda_j\}$, we can readily follow the proof of Theorem 1 in Borisov and Volodko (2008) (see also the proof of Theorem 1 in Lee (1990, Ch.3.2.2)) and show that

$$\hat{D}_{1,1}^N=\sum_{j=0}^{N}\lambda_j\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\phi_j\left(\xi_{tT}\right)\right]^2\xrightarrow{d}\sum_{j=1}^{\infty}\lambda_j\mathcal{Z}_j^2,$$

where $\{\mathcal{Z}_j\}_{j=1}^{\infty}$ is a centered Gaussian sequence with the covariances

$$E\left[\mathcal{Z}_j\mathcal{Z}_k\right]=\lim_{T\to\infty}\frac{1}{T}\sum_{s,t=1}^{T}E\left[\phi_j\left(\xi_{sT}\right)\phi_k\left(\xi_{tT}\right)\right], \qquad \text{(S1.4)}$$

where the last summation and limit are well defined by the Davydov inequality for strong mixing processes. To simplify the expression in (S1.4), let

$$\begin{aligned} V_{jk,T} &= \frac{1}{T}\sum_{s,t=1}^{T} E\left[\phi_j\left(\xi_{sT}\right)\phi_k\left(\xi_{tT}\right)\right] \\ &= \frac{1}{T}\sum_{t=1}^{T} E\left[\phi_j\left(\xi_{tT}\right)\phi_k\left(\xi_{tT}\right)\right] + \frac{1}{T}\sum_{1\leq s<t\leq T} E\left[\phi_j\left(\xi_{sT}\right)\phi_k\left(\xi_{tT}\right)\right] + \frac{1}{T}\sum_{1\leq t<s\leq T} E\left[\phi_j\left(\xi_{sT}\right)\phi_k\left(\xi_{tT}\right)\right] \\ &\equiv V_{jk,T,1} + V_{jk,T,2} + V_{jk,T,3}. \end{aligned}$$

For $V_{jk,T,1}$, we have

$$\begin{aligned} V_{jk,T,1} &= \frac{1}{T}\sum_{t=1}^{T} E\left[\phi_j\left(Y_t, \frac{t}{T}\right)\phi_k\left(Y_t, \frac{t}{T}\right)\right] \\ &= \int \phi_j\left(y,\tau\right)\phi_k\left(y,\tau\right)\mathrm{d}F_Y(y)\mathrm{d}\tau + o\left(1\right) = E\left[\phi_j\left(\xi_t\right)\phi_k\left(\xi_t\right)\right] + o\left(1\right). \end{aligned}$$

For $V_{jk,T,2}$, we have

$$\begin{aligned} V_{jk,T,2} &= \sum_{l=1}^{T-1}\frac{1}{T}\sum_{t=l+1}^{T} E\left[\phi_j\left(Y_{t-l}, \frac{t-l}{T}\right)\phi_k\left(Y_t, \frac{t}{T}\right)\right] \\ &= \sum_{l=1}^{T-1}\frac{T-l}{T}\frac{1}{T-l}\sum_{t=l+1}^{T} E\left[\phi_j\left(Y_{1-l}, \frac{t}{T}-\frac{l}{T}\right)\phi_k\left(Y_1, \frac{t}{T}\right)\right] \\ &= \sum_{l=1}^{T-1}\left(1-\frac{l}{T}\right)\int_0^1 E\left[\phi_j\left(Y_{1-l}, \tau-\frac{l}{T}\right)\phi_k\left(Y_1,\tau\right)\right]\mathrm{d}\tau + o\left(1\right) \\ &\to \sum_{l=1}^{\infty}\int_0^1 E\left[\phi_j\left(Y_{1-l},\tau\right)\phi_k\left(Y_1,\tau\right)\right]\mathrm{d}\tau = \sum_{l=1}^{\infty} E\left[\phi_j\left(Y_{t-l},V_t\right)\phi_k\left(Y_t,V_t\right)\right], \end{aligned}$$

where the second equality follows from the strict stationarity of $\{Y_t\}$, the third holds by the Riemann summation approximation of a definite integral, and the convergence holds by the dominated convergence theorem. By the same token,

$$V_{jk,T,3} \to \sum_{l=1}^{\infty}\int_0^1 E\left[\phi_j\left(Y_1,\tau\right)\phi_k\left(Y_{1-l},\tau\right)\right]\mathrm{d}\tau = \sum_{l=1}^{\infty} E\left[\phi_j\left(Y_t,V_t\right)\phi_k\left(Y_{t-l},V_t\right)\right].$$

It follows that

$$E\left[\mathcal{Z}_j\mathcal{Z}_k\right]=E\left[\phi_j\left(\xi_t\right)\phi_k\left(\xi_t\right)\right]+\sum_{l=1}^{\infty}\left\{E\left[\phi_j\left(Y_{t-l},V_t\right)\phi_k\left(Y_t,V_t\right)\right]+E\left[\phi_j\left(Y_t,V_t\right)\phi_k\left(Y_{t-l},V_t\right)\right]\right\}. \tag{S1.5}$$

In sum, we have shown that $\hat{D}_1\overset{d}{\to}\sum_{j=0}^{\infty}\lambda_j\mathcal{Z}_j^2$, where $\{\mathcal{Z}_j\}_{j=0}^{\infty}$ is a sequence of zero-mean Gaussian variables with covariance defined in (S1.5). ■

Remark. When $\{Y_t\}$ are i.i.d., it is easy to see that $E\left[\mathcal{Z}_j\mathcal{Z}_k\right]=E\left[\phi_j\left(\xi_t\right)\phi_k\left(\xi_t\right)\right]=\delta_{j,k}$. In particular, $E(\mathcal{Z}_j^2)=1$ and $\{\mathcal{Z}_j\}_{j=1}^{\infty}$ are i.i.d. $N\left(0,1\right)$. Note that the major difference between the main distributional results in Theorems S1.1 and 3.1 lies in the fact that one is represented in terms of discrete Gaussian process while the other in terms of continuous Gaussian process. Despite such different representations, the two limiting distributions are equivalent when the conditions in both theorems are satisfied. Here we only show that the two limiting distributions share the same first moment. It is easy to show that under Assumption S.2,

$$\lim_{T\to\infty}\frac{1}{T}\sum_{t,s=1}^{T}E\left[\breve{h}_1\left(Y_t,Y_s\right)\right]\breve{h}_{2st}=\int\Gamma_1\left(u,u\right)W_1\left(u\right)\mathrm{d}u\int\Gamma_2\left(v,v\right)W_2\left(v\right)\mathrm{d}v.$$

Then

$$\begin{aligned}E\left[\int\left|S\left(u,v\right)\right|^2W\left(u,v\right)\mathrm{d}u\mathrm{d}v\right] &= \int E\left[S\left(u,v\right)S\left(u,v\right)^*\right]W_1\left(u\right)W_2\left(v\right)\mathrm{d}u\mathrm{d}v\\ &= \int\Gamma_1\left(u,u\right)W_1\left(u\right)\mathrm{d}u\int\Gamma_2\left(v,v\right)W_2\left(v\right)\mathrm{d}v\\ &= \lim_{T\to\infty}\frac{1}{T}\sum_{t,s=1}^{T}E\left[\breve{h}_1\left(Y_t,Y_s\right)\breve{h}_{2st}\right]\\ &= \lim_{T\to\infty}\frac{1}{T}\sum_{t,s=1}^{T}E\left[\varphi\left(\xi_{tT},\xi_{sT}\right)\right]\\ &= \lim_{T\to\infty}\sum_{j=0}^{\infty}\lambda_j\frac{1}{T}\sum_{t,s=1}^{T}E\left[\phi_j\left(\xi_{tT}\right)\phi_j\left(\xi_{sT}\right)\right]\\ &= \sum_{j=0}^{\infty}\lambda_jE(\mathcal{Z}_j^2).\end{aligned}$$

The above calculations show that the two limiting distributions in S1.1 and 3.1 match in the first moment. By more tedious arguments, one can show that they also match in the second moments.

S2 Discussion on the Asymptotic Pivotality

An anonymous referee asks whether it is possible to standardize the statistic $\hat{D}$ to obtain an asymptotically pivotal test in our framework. Unfortunately, we are unable to provide a positive solution. To appreciate the technical challenge, we notice that Proposition 3.1 implies that under $\mathbb{H}_0$, $\sqrt{T}\hat{A}(u,v)$ converges to a complex-valued normal distribution for each fixed pair (u,v) with a well-defined long-run variance $\mathcal{K}_0(w,w)$ where $w=(u',v)'$. Then it is tempting to construct the following standardized version of $\hat{A}(u,v)$:

$$\widehat{SA}(u,v)=\frac{\sqrt{T}\hat{A}(u,v)}{\sqrt{\hat{\Gamma}_1(u,u)\hat{\Gamma}_2(v,v)}},$$

where $\hat{\Gamma}_2(v,v)=1-\left|\int_0^1 e^{\mathbf{i}v2\pi\tau}\mathrm{d}\tau\right|^2$ and $\hat{\Gamma}_1(u,u)$ is a consistent HAC estimator of the long-run variance $\Gamma_1(u,u)$. By the continuous mapping theorem, it is standard to show that $|\widehat{SA}(u,v)|^2\xrightarrow{d}\chi^2(1)$, where $\chi^2(1)$ denotes the chi-squared distribution with 1 degree of freedom. But $\widehat{SA}(u,v)$ alone cannot serve as a consistent test statistic for our null hypothesis as it only checks the spectrum at a single pair (u,v) . To obtain a consistent test, one may consider the following integrated version of $\widehat{SA}(u,v)$:

$$\widehat{SD}^{(1)}=\int_{\mathbb{R}^{dm+1}}\left|\widehat{SA}(u,v)\right|^2 W(u,v)\mathrm{d}u\mathrm{d}v.$$

Even if $|\widehat{SA}(u,v)|^2\xrightarrow{d}\chi^2(1)$ for each fixed pair (u,v), $\widehat{SD}^{(1)}$ does not follow an asymptotically pivotal distribution under the null since $\widehat{SA}(u,v)$'s are dependent across various pairs of (u,v)'s, and we do not know how to take into account the dependence structure among them to deliver a test statistic that is asymptotically pivotal under the null.

Alternatively, it is tempting to consider the following standardization:

$$\widehat{SD}^{(2)}=(\hat{D}-\hat{\mathcal{B}})/\sqrt{\hat{\mathcal{V}}},$$

where $\hat{\mathcal{B}}$ and $\hat{\mathcal{V}}$ are consistent estimators of $E(\hat{D})$ and $\mathrm{var}(\hat{D})$. As demonstrated in Section S1, the limiting distribution of $\hat{D}$ can be written as a weighted sum of chi-squared distributions with a countable number of mixture components and complex dependence structures among the mixture components that depend on certain long-run variance. The estimation of $E(\hat{D})$ and $\mathrm{var}(\hat{D})$ would inevitably require the estimation of parameters in the limiting distribution that include the eigenvalues and the long-run variance. Either one turns out to be extremely challenging if possible at all. As a result, we are unable to obtain a version of our test statistic that is asymptotically pivotal under the null.

S3 Some Additional Simulation and Application Results

In this section, we report some additional simulation and application results. First, we follow the suggestion of an anonymous referee and study the sensitivity of our test to the choice of tuning parameter b_T used in the DWB for both simulated and real data. Second, we compare our test with that of Francq and Zakoïan (2012) to test for strict stationarity of GARCH(1,1) processes. Then, we consider the pth-order stationarity tests for both simulated and real data.

S3.1 Sensitivity of the test statistics

In this subsection, we use simulations to show the impact of the tuning parameter b_T on our test based on the dependent wild bootstrap (DWB). Compared to several existing tests, our test does not need the choice of any smooth parameter in constructing the test statistic for the original data. But we need to choose a tuning parameter b_T that plays a similar role to the block length in the moving block bootstrap (MBB) to implement the bootstrap version of our test.[1] In addition, we also need to choose a weighting function as in Hong et al. (2017). However, as long as it satisfies Assumption A.2, our test is consistent against various types of nonstationarity, and the simulations in the paper suggest that our test is not sensitive to the choice of weighting function.

Below we show that our test can perform well using various choices of b_T. In the main text, we use a data-driven method, i.e., the MV approach, to determine b_T. Now, we also consider the rule-of-thumb (RoT) approach proposed by Rho and Shao (2019) and examine the finite sample performance of our test under various choices of b_T. Rho and Shao (2019) proposed a RoT choice for $b_T : b_T \equiv l_r \equiv \lfloor 6(T/100)^{1/4} \rfloor$, where $\lfloor \cdot \rfloor$ denotes the integer part of $\cdot$. Shao (2010) conjectured that nonparametric plug-in methods (e.g., Bülmann and Künsch, 1999; Paparoditis and Politis,

[1] Shao (2010) calls the tuning parameter b_T as "bandwidth". To avoid confusion with the bandwidth in nonparametric kernel estimation, we refer to b_T as "block length".

2001, 2002; Politis and White, 2004; Lahiri et al., 2007) can be extended to the DWB. So, we rerun the simulation of DGPs.S1–S5 and P1–P6 using the following choices of b_T :

- the rule-of-thumb block length: $b_T = l_r = \lfloor 6(T/100)^{1/4} \rfloor$;
- the over- and under-estimated RoT block length: $b_T = l_r \pm b$ for $b = 1$ and 2;
- the data-dependent block length via Politis and White's (2004, PW04) procedure: $b_T = l_{\text{PW04}}$;
- a combined approach that chooses the smaller one of l_r and l_{PW04}: $b_T = l_{\text{Combined}} = \min(l_r, l_{\text{PW04}})$.

Table S1 reports the finite sample size performance of our test using DWB with the aforementioned choices of block length b_T. Even though our test is a bit undersized under l_r for four out of five DGPs when the sample size is small, the rejection rates approach the corresponding nominal levels quickly as the sample size grows. As expected, the under-rejection in small samples is alleviated with $b_T = l_r - 1$ and $l_r - 2$ but is worsened with $b_T = l_r + 1$ and $l_r + 2$. The empirical rejection rates are all reasonable when the sample size is large enough. This shows the insensitivity of our results to the choice of b_T. Furthermore, the plug-in method by Politis and White (2004) and the combined approach both perform well in finite samples under all DGPs. Even though a formal justification is not provided, our simulation results support the conjecture made by Shao (2010).

Table S2 provides the finite sample power performance of our test under various choices of b_T. It shows that our test is powerful for all the considered DGPs and such performance is quite robust to the choices of b_T. We do observe that the rejection can be affected by b_T when $T = 100$. Nevertheless, the empirical rejection rates are all close to 1 when $T = 300$ and 500. We note that an exception occurs under l_{PW04}. Our test does not exhibit good finite sample power for DGP.P1 and P4–P6 when l_{PW04} is used. That is because Politis and White's (2004) approach tends to generate a large b_T under these DGPs. Notice that the adopted DWB mimics the serial dependence in the data using an Ornstein-Uhlenbeck process. When b_T is large, $\eta_t^\star$ in Section 3.4 behaves like a near-unit root process. That will inflate the value of the bootstrap test statistics and result in under-rejection for an overly large value of b_T. Such a problem can be alleviated if

Table S1: Size of DFT tests with different DWB block lengths under DGPs.S1–S5

		l_r-2		l_r-1		l_r		l_r+1		l_r+2		l_{PW04}		l_{Combined}	
	T	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
S1	100	0.044	0.134	0.034	0.100	0.010	0.084	0.006	0.080	0.010	0.072	0.024	0.106	0.040	0.110
	300	0.052	0.128	0.056	0.138	0.036	0.110	0.030	0.106	0.020	0.108	0.038	0.108	0.042	0.108
	500	0.060	0.112	0.052	0.126	0.048	0.120	0.034	0.094	0.046	0.124	0.042	0.108	0.046	0.114
S2	100	0.040	0.162	0.010	0.100	0.002	0.076	0.000	0.044	0.002	0.056	0.044	0.132	0.052	0.150
	300	0.048	0.136	0.060	0.112	0.024	0.090	0.032	0.098	0.030	0.100	0.044	0.110	0.072	0.122
	500	0.076	0.166	0.044	0.124	0.054	0.128	0.038	0.108	0.022	0.086	0.046	0.116	0.054	0.140
S3	100	0.050	0.142	0.018	0.094	0.006	0.062	0.008	0.070	0.002	0.038	0.026	0.082	0.042	0.144
	300	0.070	0.126	0.032	0.116	0.040	0.112	0.028	0.102	0.030	0.104	0.028	0.108	0.046	0.110
	500	0.028	0.112	0.068	0.130	0.042	0.094	0.036	0.100	0.034	0.092	0.040	0.106	0.054	0.098
S4	100	0.052	0.168	0.044	0.138	0.020	0.122	0.016	0.088	0.006	0.082	0.016	0.102	0.024	0.112
	300	0.070	0.140	0.052	0.164	0.046	0.132	0.034	0.110	0.042	0.122	0.030	0.096	0.058	0.128
	500	0.074	0.134	0.076	0.164	0.044	0.102	0.042	0.104	0.044	0.104	0.040	0.120	0.054	0.130
S5	100	0.048	0.136	0.010	0.072	0.008	0.080	0.000	0.044	0.002	0.036	0.018	0.080	0.016	0.092
	300	0.044	0.118	0.048	0.106	0.036	0.108	0.024	0.084	0.022	0.092	0.040	0.130	0.032	0.108
	500	0.078	0.150	0.058	0.122	0.038	0.118	0.062	0.122	0.048	0.098	0.036	0.114	0.044	0.096

Notes: (i) l_r denotes using the rule-of-thumb block length $l_r = \lfloor 6(T/100)^{1/4} \rfloor$; (ii) $l_r \pm b, b = 1, 2$ denote using block lengths $l_r \pm b$, respectively; (iii) l_{PW04} denotes using the block length selection method of Politis and White (2004); (iv) l_{Combined} denotes choosing the smaller value between l_r and l_{PW04}; and (v) the number of bootstrap samples is 1000.

one adopts the combined approach that picks a smaller value between l_r and l_{PW04}. Since l_r only depends on the sample size, it will offset the effect of an excessively large value of b_T selected by PW04 when the serial dependence is strong in the data.

Table S2: Power of DFT tests with different DWB block lengths under DGPs.P1–P6

		l_r-2		l_r-1		l_r		l_r+1		l_r+2		l_{PW04}		l_{Combined}	
DGP	T	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
P1	100	0.788	0.904	0.670	0.868	0.526	0.812	0.420	0.716	0.256	0.690	0.000	0.072	0.534	0.796
	300	0.938	0.956	0.936	0.974	0.884	0.942	0.876	0.946	0.834	0.932	0.010	0.430	0.888	0.948
	500	0.964	0.980	0.970	0.984	0.960	0.978	0.928	0.976	0.920	0.970	0.078	0.568	0.958	0.990
P2	100	0.530	0.776	0.448	0.746	0.320	0.704	0.264	0.686	0.122	0.642	0.682	0.820	0.692	0.838
	300	0.998	1.000	0.994	0.996	1.000	1.000	1.000	1.000	0.998	1.000	0.996	0.998	1.000	1.000
	500	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
P3	100	0.566	0.774	0.458	0.730	0.316	0.696	0.252	0.676	0.122	0.562	0.660	0.790	0.668	0.802
	300	0.998	1.000	0.992	0.998	0.998	1.000	0.994	1.000	0.988	0.998	1.000	1.000	0.998	1.000
	500	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
P4	100	0.770	0.906	0.694	0.886	0.608	0.868	0.474	0.820	0.284	0.768	0.080	0.478	0.588	0.866
	300	0.998	0.998	0.996	1.000	0.990	1.000	0.994	1.000	0.986	0.998	0.158	0.926	0.994	1.000
	500	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.720	1.000	1.000	1.000
P5	100	0.972	1.000	0.910	0.994	0.746	0.990	0.480	0.962	0.248	0.926	0.028	0.278	0.706	0.992
	300	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.070	0.986	1.000	1.000
	500	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.504	1.000	1.000	1.000
P6	100	0.962	0.990	0.920	0.980	0.852	0.988	0.668	0.948	0.440	0.912	0.000	0.056	0.834	0.972
	300	1.000	1.000	0.996	1.000	0.996	0.998	0.990	0.996	0.966	0.994	0.002	0.406	0.992	1.000
	500	1.000	1.000	1.000	1.000	0.998	0.998	0.998	1.000	1.000	1.000	0.006	0.572	1.000	1.000

Notes: (i) l_r denotes using the rule-of-thumb block length $l_r = \lfloor 6(T/100)^{1/4} \rfloor$; (ii) $l_r \pm b, b = 1, 2$ denote using block lengths $l_r \pm b$, respectively; (iii) l_{PW04} denotes using the block length selection method of Politis and White (2004); (iv) l_{Combined} denotes choosing the smaller value between l_r and l_{PW04}; and (v) the number of bootstrap samples is 1000.

Table S3 provides the empirical results for our test with different choices of b_T. Compared with the results in the main text, we can see that our test is quite robust to the choices of b_T, even though we may draw different conclusions for different choices of b_T in some cases. As the simulations suggest, in case of conflicting conclusions, we recommend the use of l_{Combined}.

Table S3: Stationarity tests for exchange rate returns with different block lengths

	$l_r - 2$	$l_r - 1$	l_r	$l_r + 1$	$l_r + 2$	l_{PW04}	l_{Combined}
Nominal-univariate							
GBP	0.022	0.030	0.032	0.049	0.037	0.006	0.003
CAD	0.000	0.000	0.000	0.000	0.000	0.000	0.000
JAY	0.051	0.074	0.066	0.078	0.112	0.006	0.008
EUR	0.086	0.083	0.100	0.110	0.118	0.031	0.018
Real-univariate							
GBP	0.013	0.012	0.030	0.018	0.016	0.000	0.001
CAD	0.000	0.000	0.000	0.000	0.000	0.000	0.000
JAY	0.059	0.071	0.065	0.083	0.069	0.014	0.011
EUR	0.100	0.105	0.115	0.120	0.120	0.028	0.028
Nominal-bivariate							
(GBP, CAD)	0.000	0.000	0.000	0.002	0.000	0.000	0.000
(GBP, JAY)	0.008	0.018	0.012	0.025	0.026	0.001	0.000
(GBP, EUR)	0.192	0.221	0.249	0.236	0.247	0.123	0.123
(CAD, JAY)	0.000	0.001	0.000	0.002	0.004	0.000	0.000
(CAD, EUR)	0.068	0.070	0.101	0.109	0.113	0.022	0.028
(JAY, EUR)	0.066	0.088	0.109	0.112	0.150	0.025	0.028
Real-bivariate							
(GBP,CAD)	0.000	0.000	0.000	0.002	0.000	0.000	0.000
(GBP, JAY)	0.006	0.009	0.011	0.014	0.016	0.001	0.001
(GBP, EUR)	0.262	0.265	0.300	0.271	0.290	0.153	0.154
(CAD, JAY)	0.001	0.000	0.001	0.001	0.003	0.000	0.000
(CAD, EUR)	0.077	0.077	0.089	0.090	0.119	0.037	0.034
(JAY, EUR)	0.075	0.093	0.126	0.135	0.139	0.044	0.029

Notes: (i) numbers in main entries are the bootstrap p-values; (ii) l_r denotes using the rule-of-thumb block length $l_r = \lfloor 6(T/100)^{1/4} \rfloor$; (iii) $l_r \pm b, b = 1, 2$ denote using block lengths $l_r \pm b$, respectively; (iv) l_{PW04} denotes using the block length selection method of Politis and White (2004); (v) l_{Combined} denotes choosing the smaller value between l_r and l_{PW04}; and (vi) the number of bootstrap samples is 1000.

S3.2 A Comparison of the DFT test with Francq and Zakoïan's (2012) test for the strict stationarity of GARCH(1,1) processes

As suggested by one referee, we further investigated how the proposed test performs in GARCH models, in comparison with Francq and Zakoïan's (2012) strict stationarity test which is developed specifically for GARCH$(1,1)$. Specifically, we consider the following four GARCH(1,1) processes.

$$
\begin{aligned}
&\text{DGP.S6}: \quad Y_t = \sqrt{h_t}\varepsilon_t,\ h_t = 0.01 + 0.01Y_{t-1}^2 + 0.92h_{t-1};\\
&\text{DGP.S7}: \quad Y_t = \sqrt{h_t}\varepsilon_t,\ h_t = 0.001 + 0.8Y_{t-1}^2 + 0.2h_{t-1};\\
&\text{DGP.P7}: \quad Y_t = \sqrt{h_t}\varepsilon_t,\ h_t = \begin{cases} 0.01 + 0.01Y_{t-1}^2 + 0.92h_{t-1}, & t \leq 0.3T \\ 0.001 + 0.8Y_{t-1}^2 + 0.15h_{t-1}, & t > 0.3T \end{cases};\\
&\text{DGP.P8}: \quad Y_t = \sqrt{h_t}\varepsilon_t,\ h_t = 0.001 + 0.28Y_{t-1}^2 + 0.8h_{t-1};
\end{aligned}
$$

where $\{Y_t\}_{t=1}^T$ is the process to be tested, and $\{\varepsilon_t\} \sim t(7)$ is an i.i.d. error term, which is adopted in Francq and Zakoïan (2012). Obviously, DGP.S6 is a stationary GARCH(1,1) process, where the coefficients of Y_{t-1}^2 and h_{t-1} (denoted as α and β, respectively) satisfy the weak stationarity restriction $\alpha + \beta = 0.93 < 1$. DGP.S7 is an integrated GARCH(1,1) (IGARCH) process with $\alpha + \beta = 1$. As illustrated by Nelson (1990), DGP.S7 satisfies the strict stationarity condition but not the weak stationarity condition. The coefficients considered in DGPs.S6–S7 are common in empirical studies as $\alpha + \beta \simeq 1$. As for the nonstationary cases, DGP.P7 is a nonstationary GARCH(1,1) process with an abrupt break in the volatility dynamics. DGP.P8 is a nonstationary GARCH(1,1) process investigated by Francq and Zakoïan (2012) with the corresponding coefficients satisfying $\alpha + \beta > 1$. (See e.g., Table IV of their paper.)

Table S4 reports the rejection rates for both our test and that of Francq and Zakoïan (2012). As expected, our test shows reasonable empirical rejection rates around the nominal significance levels under DGPs.S6–S7. Even though the test of Francq and Zakoïan (2012) shows under-rejection for DGPs.S6–S7, it is consistent with their theory. Define the top Lyapunov exponent associated with a GARCH(1,1) model: $Y_t = \sqrt{h_t}\varepsilon_t,\ h_t = \omega + \alpha Y_{t-1}^2 + \beta h_{t-1}$ as

$$\gamma \ \equiv \ E\left[\log\left(\alpha\varepsilon_t^2 + \beta\right)\right].$$

Nelson (1990) shows that the necessary and sufficient condition for the strict stationarity of the

GARCH(1,1) process is $\gamma < 0$. Francq and Zakoïan (2012) consider a test for strict stationarity by testing the null hypothesis $H_0: \ \gamma < 0$. But the empirical rejection rate of their test only converges to the nominal significance level when the top Lyapunov exponent equals 0 (see Corollary 3.3 in their paper). This explains the under-size issue of their test since $\gamma < 0$ in DGPs.S6 and S7.

As for the power performance, we find that Francq and Zakoïan's (2012) test does not have power against DGP.P7 when the nonstationarity results from an abrupt structural break instead of explosive volatility dynamics. In contrast, our tests have good power and the empirical rejection rate increases as the sample size grows. This shows a power improvement of our test over Francq and Zakoïan (2012). Intuitively, their test is designed to detect whether the top Lyapunov exponent γ exceeds 0. DGP.P7 depicts two regimes of the volatility dynamics which are both non-explosive with $\gamma < 0$. This explains why Francq and Zakoïan's (2012) test does not have power. In addition, our tests also have good power against the nonstationary explosive GARCH process under DGP.P8 in which $\gamma > 0$. In contrast, Francq and Zakoïan's (2012) test is less powerful and even exhibits non-increasing power against DGP.P8 as the sample size increases. We note that the empirical rejection rate of Francq and Zakoïan's (2012) test increases as the sample size increases from 500 to 4000 as indicated by their Table IV. The empirical rejection rates under 5% and 10% significance levels are 69.2% and 80.9% when $T = 2000$, and 91.7% and 97% when $T = 4000$, respectively. When conducting our simulation study, we adopted their code directly from their online supplement at
https://www.econometricsociety.org/content/supplement-strict-stationarity-testing-and-estimation-explosive-and-stationary-garch-models.

S3.3 Tests for the second-order stationarity via the DFT

In this subsection, we report some simulation results and empirical evidence for testing the second-order stationarity via the DFT. We use DGPs.S1–S6 and P1–P7 studied in Sections 5 and S3.2. Note that the second-order moment does not exist under DGPs.S7 and P8, the moment condition of the second-order stationarity test is not satisfied. So the results for DGPs.S7 and P8 are omitted. Furthermore, we note that testing second-order stationarity should contain testing constancy for both the first and second moment of a time series. To concisely illustrate the idea of our test, we only report the results for testing constancy of the second moment.

Table S4: Rejection rates of strict stationarity tests for DGPs.S6–S7 and DGPs.P7–P8

		$\hat{D}^N$		$\hat{D}^L$		$\hat{F}$	
DGP	T	5%	10%	5%	10%	5%	10%
S6	100	0.030	0.096	0.023	0.109	0.002	0.002
	300	0.036	0.099	0.044	0.109	0.000	0.000
	500	0.050	0.114	0.056	0.132	0.000	0.000
S7	100	0.048	0.129	0.060	0.161	0.001	0.001
	300	0.036	0.139	0.046	0.183	0.000	0.000
	500	0.040	0.128	0.040	0.124	0.000	0.000
P7	100	0.108	0.295	0.231	0.583	0.001	0.003
	300	0.643	0.787	0.819	0.916	0.000	0.000
	500	0.760	0.856	0.888	0.968	0.000	0.000
P8	100	0.848	0.927	0.884	0.956	0.335	0.382
	300	0.994	0.998	0.993	1.000	0.279	0.370
	500	0.942	0.999	0.934	1.000	0.312	0.436

Notes: (i) $\hat{D}^N$ and $\hat{D}^L$ denote DFT tests for strict stationarity with normal and Laplace weighting functions, respectively; (ii) $\hat{F}$ denotes strict stationarity tests of Francq and Zakoïan (2012); (iii) for each test, the number of repetitions is 1000; and (iv) for our test, the number of bootstrap samples is 500.

For our test, we implement $\hat{D}^{(2)}$ defined in Section 4 when the weighting function $\tilde{W}(v)$ is chosen to be the normal and Laplace density functions, respectively. To implement the DWB, we consider the following three choices of b_T : (1) $b_T = l_{\text{MV}}$ with l_{MV} selected via the Rho and Shao's (2019) MV approach, (2) $b_T = l_r$ with l_r being the Rho and Shao's (2019) rule-of-thumb block length, and (3) $b_T = l_{\text{Combined}} = \min(l_r, l_{\text{PW04}})$ with l_{PW04} selected via Politis and White's (2004) procedure. In addition, we also implement Hong et al.'s (2017) test for the second-order stationarity and choose their bandwidth and block length parameters as they suggested.

Tables S5 reports the size performance of our test for DGPs.S1–S6 in comparison with Hong et al.'s (2017) test for testing the second-order stationarity. As Table S5 suggests, our tests can be undersized for some of the DGPs when the sample size T is small, but the size generally improves as the sample size increases. Exception occurs for DGP.S6 where moderate oversize distortion is observed when T is small. In contrast, Hong et al.'s (2017) test has superb size control for all DGPs but DGP.S6 where severe size distortion exists even when $T = 500$.

Table S6 reports the finite sample power and size performance of our test for DGPs.P1–P7 in

Table S5: Finite sample size of the second-order stationarity tests under DGPs.S1–S6

DGP		T	$\hat{D}^{(2),N}_{l_{\mathrm{MV}}}$ 5%	10%	$\hat{D}^{(2),N}_{l_r}$ 5%	10%	$\hat{D}^{(2),N}_{l_{\mathrm{Combined}}}$ 5%	10%	$\hat{D}^{(2),L}_{l_{\mathrm{MV}}}$ 5%	10%	$\hat{D}^{(2),L}_{l_r}$ 5%	10%	$\hat{D}^{(2),L}_{l_{\mathrm{Combined}}}$ 5%	10%	$\hat{H}^{(2)}$ 5%	10%
S1	$Y_{t,1}$	100	0.022	0.087	0.023	0.103	0.032	0.110	0.012	0.059	0.000	0.035	0.023	0.088	0.073	0.123
		300	0.036	0.105	0.036	0.096	0.058	0.116	0.021	0.070	0.026	0.092	0.042	0.116	0.062	0.106
		500	0.039	0.093	0.032	0.082	0.046	0.112	0.036	0.094	0.026	0.076	0.032	0.094	0.050	0.100
S2	$Y_{t,1}$	100	0.010	0.060	0.016	0.078	0.009	0.051	0.005	0.030	0.000	0.018	0.006	0.037	0.078	0.134
		300	0.006	0.035	0.008	0.046	0.008	0.048	0.005	0.024	0.004	0.028	0.004	0.030	0.060	0.120
		500	0.005	0.032	0.010	0.044	0.008	0.044	0.004	0.024	0.006	0.038	0.006	0.038	0.052	0.118
S3	$Y_{t,1}$	100	0.016	0.071	0.024	0.082	0.032	0.093	0.010	0.051	0.002	0.026	0.022	0.076	0.076	0.143
		300	0.019	0.065	0.020	0.066	0.032	0.088	0.009	0.052	0.020	0.050	0.030	0.074	0.062	0.104
		500	0.022	0.068	0.020	0.076	0.038	0.086	0.023	0.063	0.020	0.064	0.036	0.082	0.050	0.094
	$Y_{t,2}$	100	0.017	0.086	0.019	0.095	0.028	0.097	0.008	0.051	0.002	0.028	0.020	0.090	0.059	0.109
		300	0.040	0.089	0.032	0.086	0.050	0.118	0.016	0.075	0.024	0.074	0.042	0.102	0.056	0.102
		500	0.036	0.074	0.034	0.086	0.038	0.104	0.035	0.084	0.032	0.070	0.036	0.088	0.042	0.092
S4	$Y_{t,1}$	100	0.025	0.091	0.033	0.111	0.027	0.097	0.014	0.063	0.002	0.038	0.020	0.080	0.056	0.106
		300	0.035	0.093	0.036	0.102	0.046	0.112	0.020	0.052	0.032	0.086	0.048	0.106	0.048	0.106
		500	0.034	0.096	0.036	0.100	0.038	0.106	0.032	0.098	0.032	0.078	0.038	0.090	0.040	0.106
	$Y_{t,2}$	100	0.017	0.070	0.022	0.082	0.027	0.085	0.010	0.058	0.001	0.029	0.021	0.067	0.079	0.153
		300	0.023	0.073	0.026	0.082	0.036	0.098	0.010	0.050	0.020	0.074	0.028	0.090	0.072	0.142
		500	0.020	0.098	0.022	0.096	0.044	0.116	0.023	0.099	0.020	0.078	0.030	0.102	0.066	0.148
	$Y_{t,3}$	100	0.022	0.072	0.028	0.079	0.038	0.091	0.015	0.050	0.003	0.033	0.027	0.077	0.062	0.118
		300	0.022	0.082	0.016	0.072	0.052	0.114	0.027	0.063	0.010	0.062	0.046	0.102	0.060	0.112
		500	0.018	0.078	0.024	0.086	0.050	0.122	0.028	0.078	0.016	0.082	0.050	0.116	0.064	0.142
S5	$Y_{t,1}$	100	0.012	0.069	0.013	0.081	0.023	0.090	0.007	0.048	0.000	0.020	0.019	0.078	0.046	0.104
		300	0.032	0.090	0.026	0.076	0.040	0.124	0.019	0.059	0.020	0.054	0.038	0.104	0.052	0.106
		500	0.034	0.078	0.038	0.080	0.046	0.100	0.032	0.076	0.032	0.078	0.046	0.096	0.044	0.096
	$Y_{t,2}$	100	0.018	0.067	0.016	0.071	0.022	0.086	0.007	0.045	0.004	0.028	0.018	0.070	0.055	0.118
		300	0.032	0.084	0.030	0.088	0.046	0.114	0.023	0.071	0.020	0.080	0.042	0.102	0.056	0.112
		500	0.028	0.078	0.026	0.070	0.036	0.104	0.027	0.069	0.024	0.060	0.038	0.098	0.056	0.114
	$Y_{t,3}$	100	0.019	0.075	0.025	0.078	0.036	0.096	0.009	0.054	0.003	0.029	0.030	0.079	0.061	0.116
		300	0.023	0.084	0.030	0.090	0.044	0.124	0.013	0.049	0.020	0.074	0.040	0.116	0.048	0.106
		500	0.026	0.084	0.026	0.088	0.048	0.102	0.021	0.087	0.022	0.076	0.044	0.098	0.044	0.088
	$Y_{t,4}$	100	0.011	0.064	0.017	0.068	0.028	0.083	0.008	0.048	0.002	0.024	0.018	0.071	0.048	0.092
		300	0.027	0.081	0.028	0.068	0.038	0.100	0.019	0.067	0.018	0.060	0.030	0.094	0.054	0.092
		500	0.040	0.084	0.042	0.088	0.050	0.106	0.040	0.089	0.040	0.082	0.056	0.108	0.056	0.112
S6	$Y_{t,1}$	100	0.060	0.185	0.025	0.121	0.093	0.214	0.046	0.150	0.009	0.082	0.079	0.183	0.226	0.341
		300	0.071	0.190	0.065	0.178	0.051	0.167	0.061	0.165	0.046	0.164	0.038	0.146	0.218	0.342
		500	0.056	0.160	0.060	0.150	0.054	0.152	0.040	0.136	0.048	0.148	0.038	0.140	0.198	0.302

Notes: (i) $\hat{D}^{(2),N}_{(\cdot)}$ and $\hat{D}^{(2),L}_{(\cdot)}$ denote DFT tests for the second-order stationarity with normal and Laplace weighting functions, respectively; (ii) among the subscripts of $\hat{D}^{(2),\cdot}_{(\cdot)}$, l_{MV} denotes using minimum volatility method to select DWB block length, l_r denotes using the rule-of-thumb block length $l_r = \lfloor 6(T/100)^{1/4} \rfloor$, and l_{Combined} denotes choosing the smaller block length selected by l_r and l_{PW04}; (iii) $\hat{H}^{(2)}$ denotes the second-order stationarity test of Hong et al. (2017); and (iv) $Y_{t,i}, i = 1, \dots, d$ denotes the ith entry of vector Y_t.

comparison with Hong et al.'s (2017) test for testing the second-order stationarity. Please note that the null hypothesis of second-order stationarity is violated in DGPs.P1, P2, P4, and $\{Y_{t,2}\}$ in DGP.P5, and all three series in DGP.P6, whereas it holds for DGP.P3 and $\{Y_{t,1}\}$ in DGP.P5. Table S6 indicates that like the test of Hong et al. (2017), our test has reasonable power against the alternatives in DGP.P1, P2, and $\{Y_{t,2}\}$ in DGP.P5, all three series in DGP.P6, and DGP.P7, but has low power against the locally stationary alternative in DGP.P4. For this latter DGP, both the first and second moments of Y_t are time-varying but they are of the order t/T, which is $o(1)$ for a large number of time series observations. For DGP.P3 and $\{Y_{t,1}\}$ in DGP.P5 where the null hypothesis of second-order stationarity holds, both our test and that of Hong et al. (2017) have reasonable size and our test outperforms that of Hong et al. (2017) under DGP.P3.

We also apply the pth-order stationarity tests to the exchange rate returns data. In the case of rejection of the null of strict stationarity, it is interesting to know the source of nonstationarity. One way is to check whether it is caused by the nonstationarity in the first or second moments. Since we reject strict stationarity for the GBP and CAD series, we apply the proposed pth-order stationarity test to these two series. Specifically, we conduct the first- and second-order tests (i.e., $p = 1$ and 2) to both the nominal and real GBP/USD and CAD/USD returns with the normal and Laplace weighting functions, respectively. We set the number of bootstrap replications to be $B = 1000$. Table S7 reports the bootstrap p-values. It shows that we cannot reject the null hypothesis when $p = 1$, but find strong evidence of the time-varying second-order moment under $p = 2$. This indicates that one source of rejection for the strict stationarity test of GBP and CAD is the time-varying second-order moment. The results also coincide with the conclusions drawn by Malik (2003) and Rapach and Strauss (2008), which document structural breaks in the variance of exchange rate returns.

Table S6: Finite sample power/size of the second-order stationarity tests under DGPs.P1–P7

DGP		T	$\hat{D}^{(2),N}_{l_{\mathrm{MV}}}$ 5%	10%	$\hat{D}^{(2),N}_{l_r}$ 5%	10%	$\hat{D}^{(2),N}_{l_{\mathrm{Combined}}}$ 5%	10%	$\hat{D}^{(2),L}_{l_{\mathrm{MV}}}$ 5%	10%	$\hat{D}^{(2),L}_{l_r}$ 5%	10%	$\hat{D}^{(2),L}_{l_{\mathrm{Combined}}}$ 5%	10%	$\hat{H}^{(2)}$ 5%	10%
P1	$Y_{t,1}$	100	0.195	0.469	0.053	0.266	0.053	0.266	0.164	0.472	0.031	0.233	0.031	0.234	0.250	0.379
		300	0.443	0.675	0.491	0.720	0.491	0.720	0.453	0.719	0.496	0.764	0.496	0.764	0.590	0.682
		500	0.508	0.746	0.712	0.844	0.712	0.844	0.513	0.785	0.754	0.892	0.754	0.892	0.750	0.822
P2	$Y_{t,1}$	100	0.750	0.936	0.556	0.902	0.858	0.960	0.662	0.919	0.441	0.868	0.830	0.952	0.924	0.970
		300	1.000	1.000	0.999	1.000	0.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		500	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
P3	$Y_{t,1}$	100	0.069	0.160	0.038	0.111	0.089	0.154	0.053	0.133	0.030	0.093	0.082	0.147	0.133	0.203
		300	0.052	0.113	0.065	0.117	0.084	0.127	0.047	0.098	0.056	0.108	0.074	0.122	0.108	0.166
		500	0.074	0.130	0.054	0.116	0.054	0.122	0.063	0.131	0.060	0.106	0.062	0.114	0.092	0.146
P4	$Y_{t,1}$	100	0.008	0.042	0.000	0.010	0.003	0.042	0.005	0.033	0.000	0.005	0.003	0.038	0.019	0.038
		300	0.013	0.058	0.017	0.071	0.021	0.088	0.010	0.061	0.014	0.086	0.022	0.102	0.040	0.110
		500	0.024	0.088	0.032	0.118	0.040	0.138	0.024	0.081	0.040	0.162	0.054	0.184	0.120	0.200
P5	$Y_{t,1}$	100	0.017	0.078	0.005	0.054	0.038	0.108	0.011	0.059	0.001	0.034	0.026	0.095	0.054	0.100
		300	0.021	0.088	0.033	0.106	0.058	0.126	0.015	0.075	0.026	0.088	0.048	0.120	0.044	0.102
		500	0.034	0.078	0.039	0.102	0.039	0.102	0.029	0.080	0.022	0.078	0.030	0.100	0.046	0.094
	$Y_{t,2}$	100	0.007	0.043	0.001	0.013	0.032	0.084	0.007	0.040	0.002	0.013	0.034	0.104	0.036	0.067
		300	0.018	0.139	0.022	0.185	0.143	0.340	0.034	0.261	0.042	0.322	0.216	0.488	0.164	0.376
		500	0.058	0.368	0.186	0.642	0.294	0.694	0.075	0.399	0.398	0.864	0.484	0.890	0.680	0.898
P6	$Y_{t,1}$	100	0.162	0.415	0.041	0.217	0.042	0.225	0.141	0.410	0.016	0.173	0.017	0.182	0.247	0.401
		300	0.501	0.707	0.545	0.748	0.545	0.748	0.483	0.737	0.560	0.788	0.560	0.788	0.614	0.700
		500	0.614	0.772	0.736	0.854	0.736	0.854	0.626	0.813	0.760	0.884	0.760	0.884	0.782	0.846
	$Y_{t,2}$	100	0.248	0.524	0.052	0.269	0.052	0.271	0.214	0.526	0.026	0.239	0.026	0.240	0.278	0.443
		300	0.530	0.736	0.567	0.766	0.567	0.766	0.532	0.767	0.586	0.820	0.586	0.820	0.638	0.718
		500	0.610	0.794	0.746	0.860	0.746	0.860	0.636	0.810	0.756	0.902	0.756	0.902	0.790	0.858
	$Y_{t,3}$	100	0.218	0.491	0.057	0.255	0.057	0.255	0.176	0.494	0.028	0.222	0.028	0.222	0.275	0.416
		300	0.464	0.701	0.521	0.725	0.521	0.725	0.464	0.744	0.544	0.770	0.544	0.770	0.606	0.688
		500	0.538	0.756	0.724	0.858	0.724	0.858	0.546	0.767	0.746	0.896	0.746	0.896	0.726	0.822
P7	$Y_{t,1}$	100	0.357	0.637	0.154	0.530	0.226	0.577	0.289	0.587	0.098	0.460	0.181	0.528	0.620	0.686
		300	0.594	0.690	0.608	0.672	0.602	0.672	0.574	0.673	0.616	0.674	0.605	0.675	0.708	0.770
		500	0.684	0.752	0.644	0.692	0.636	0.688	0.668	0.744	0.628	0.684	0.628	0.684	0.708	0.744

Notes: (i) $\hat{D}^{(2),N}_{(\cdot)}$ and $\hat{D}^{(2),L}_{(\cdot)}$ denote DFT tests for the second-order stationarity with normal and Laplace weighting functions, respectively; (ii) among the subscripts of $\hat{D}^{(2),\cdot}_{(\cdot)}$, l_{MV} denotes using minimum volatility method to select DWB block length, l_r denotes using the rule-of-thumb block length $l_r = \lfloor 6(T/100)^{1/4} \rfloor$, and l_{Combined} denotes choosing the smaller block length selected by l_r and l_{PW04}; (iii) $\hat{H}^{(2)}$ denotes the second-order stationarity test of Hong et al. (2017); and (iv) $Y_{t,i}, i = 1, \ldots, d$ denotes the ith entry of vector Y_t.

Table S7: The pth-order stationarity tests for GBP/USD and CAD/USD returns

	$\hat{D}^{(p),N}$		$\hat{D}^{(p),L}$	
	$p=1$	$p=2$	$p=1$	$p=2$
Nominal GBP/USD	0.832	0.028	0.872	0.026
Nominal CAD/USD	0.472	0.004	0.399	0.005
Real GBP/USD	0.865	0.010	0.882	0.007
Real CAD/USD	0.753	0.003	0.530	0.011

Notes: (i) the main entries are the bootstrap p-values of the tests; (ii) $\hat{D}^{(p),N}$ and $\hat{D}^{(p),L}$ denote the pth-order stationarity test with the normal and Laplace weighting functions, respectively; (iii) $p=1$ and $p=2$ denote the first-order and the second-order stationarity tests, respectively; and (iv) the number of bootstrap samples is 1000.

Additional References

Borisov, I.S. and N.V. Volodko (2008) Orthogonal series and limit theorems for canonical U- and V-statistics of stationary connected observations. *Siberian Advances in Mathematics* 18, 242–257.

Bühlmann, P. and H.R. Künsch (1999) Block length selection in the bootstrap for time series. *Computational Statistics & Data Analysis* 31, 295–310.

Francq, C. and J.-M. Zakoïan (2012) Strict stationarity testing and estimation of explosive and stationary generalized autoregressive conditional heteroscedasticity models. *Econometrica* 80, 821–861.

Hong, Y., X. Wang and S. Wang (2017) Testing strict stationarity with applications to macroeconomic time series. *International Economic Review* 58, 1227–1277.

Lahiri, S.N., K. Furukawa and Y.-D. Lee (2007) A nonparametric plug-in rule for selecting optimal block lengths for block bootstrap methods. *Statistical Methodology* 4, 292–321.

Lee, A.J. (1990) *U-Statistics: Theory and Practice.* CRC Press, New York.

Malik F. (2003) Sudden changes in variance and volatility persistence in foreign exchange markets. *Journal of Multinational Financial Management* 13, 217–230.

Nelson, D.B. (1990) Stationarity and persistence in the GARCH(1,1) model. *Econometric Theory* 6, 318–334.

Paparoditis, E. and D.N. Politis (2001) Tapered block bootstrap. *Biometrika* 88, 1105–1119.

Paparoditis, E. and D.N. Politis (2002) The tapered block bootstrap for general statistics from stationary sequences. *The Econometrics Journal* 5, 131–148.

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

Rapach D.E. and J.K. Strauss. (2008) Structural breaks and GARCH models of exchange rate volatility. *Journal of Applied Econometrics* 23, 65–90.

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

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

Sun, H. (2005) Mercer theorem for RKHS on noncompact sets. *Journal of Complexity* 21, 337–349.

Sun, S. and C.-Y. Chiang (1997) Limiting behavior of the perturbed empirical distribution functions evaluated at U-statistics for strongly mixing sequences of random variables. *Journal of Applied Mathematics and Stochastic Analysis* 10, 3–20.

Vapnik, V.N. (1998) *Statistical Learning Theory.* John Wiley & Sons, New Jersey.