JETEM

Journal of Economic Theory and Econometrics

The Journal of the Korean Econometrics Society

(계 량 경 제 학 보)

March 2010

Volume 21. Issue 1. pp.1-22

Long-run Variance Estimation for Linear Processes Under Possible Degeneracy*

Jin Lee†

Abstract We analyze the asymptotic behavior of the long-run variance estimator for linear processes under degeneracy, where the spectral density function near the origin equals to zero. Given degeneracy which typically arises from over-differencing, standard results in the literature of heteroskedasticity and autocorrelation consistent (HAC) estimation are invalid. We provide asymptotic distribution of the long-run variance estimator from long term trends in linear processes. Further, we propose a test statistic to testing degeneracy, which achieves asymptotic normality. Our test is directly applied to testing for trend stationarity. Under the null of trend stationarity, the spectrum near the origin for the differenced process becomes zero. On the other hand, under the alternative of difference stationarity, the spectrum becomes strictly positive at the zero frequency. It is found that, depending on the signal-to-ratio, our test has significant power advantages over the KPSS test. Thus, the proposed test becomes an useful complement to the KPSS test

Keywords Long-run variance; linear process; spectral density estimator; degeneracy, trend stationarity

JEL Classification C12, C14, C22

* The author is grateful to Yongmiao Hong and two anonymous referees.

† 2Associate Professor, Department of Economics, Ewha Womans University, Seoul, Korea, 120-750, Email address: leejin@ewha.ac.kr.

Received June 26, 2008, Revised December 27, 2008, Accepted July 29, 2009

1. Introduction

The magnitude of the long-run variance (LRV), which is the spectral density function near the origin draws a lot of attention in time series context. The power spectrum at the zero frequency represents long term trends in the data. In the macroeconometrics context, long-term forecasts or long-run responses to certain shocks are concentrated on the zero frequency in terms of spectral representation of the underlying processes. Also, in the regression context, it is well known that heteroskedasticity and autocorrelation consistent (HAC) covariance matrix estimator is simply proportional to the spectral density estimator near the origin. The HAC estimators are widely used to estimate long-run variances of the data in contexts such as unit root testing, cointegration estimation and so on. In doing so, it is often assumed that the spectrum near the origin is strictly positive. Then, associated theories of the LRV estimator have been well developed in the literature (Hannan (1970), Andrews (1991) and Newey and West (1994), to name a few). On the other hand, relevant theories when the true spectrum is zero are seldom known.

In our work, we concentrate on a degenerate situation, where the power spectrum at the zero frequency equals to zero. This typically arises from over-differencing the data. For instance, if the true process is trend stationary, but mistakenly first-differenced, then it generates nil spectrum near the origin. Similar arguments can be applied to both stationary and nonstationary processes, which are differenced more than the true order of the magnitude in the data. Simply put, over-differencing excessively removes the long-run information in the data. If the true spectrum is zero, existing theories in the HAC context do not work, particularly due to the degenerate asymptotic variance. The asymptotic variance of the LRV estimator equals to zero with the well-known rate of T/M, where T is the sample size and M is the bandwidth. Thus, in order to derive a valid limiting distribution, it is necessary to deal with higher-order expansions of the asymptotic variance of the estimator. We find that the asymptotic variance converges to a non-degenerating quantity with the rate as big as $T \times M^3$.

We specifically deal with degeneracy in linear processes. A weakly dependent linear process can be effectively decomposed into long term trend parts and short rum components through Beveridge-Nelson techniques. We derive asymptotic normality of the estimator based on long-term trends and discuss the relationships between the variances of HAC estimator and of long term trends. Given the result, we propose a test statistic to testing trend stationarity. Under the null of trend stationarity, the spectrum near the origin for the first-differenced process becomes zero. On the other hand, under the alternative of difference stationarity, the spectrum becomes strictly positive at the zero frequency. It is found from simulation studies that our test has power advantages over the KPSS test, depending on the signal-to-ratio. Thus, the proposed test can serve as an useful complement to the KPSS test.

2. Linear Process and Degenerate Spectral Density

We consider a scalar time series x_t, $t = 1, 2, \cdots, T$, which follows a linear process with weakly dependent structures as studied in Phillips and Solo (1992),

$$x_t - \mu = a(L)e_t, = \sum_{k=0}^{\infty} a_k e_{t-k}, \tag{1}$$

where e_t is martingale difference sequence (mds) with zero mean and, $Ee_t^4 = \mu_4 < \infty$, and $\sum_{k=1}^{\infty} k a_k^2 < \infty$.

In order to obtain spectral representation for this linear process, we introduce the spectral density function of a scalar series process x_t,

$$f(\lambda) = (2\pi)^{-1} \sum_{j=-\infty}^{\infty} R(j) e^{-i\lambda j}, \qquad -\pi \leq \lambda \leq \pi, \tag{2}$$

where $R(j) = E(x_t - \mu)(x_{t-|j|} - \mu)$, and $\mu = Ex_t$.

The linear relationship specified in (1) can lead to the following relationship in frequency domain (Priestley (1981, p.669)),

$$f(\lambda) = |A(\lambda)|^2 f_e(\lambda), \qquad -\pi \leq \lambda \leq \pi, \tag{3}$$

where $f_e(\lambda)$ are the spectral density functions of e_t, and $A(\lambda)$ is the Fourier transforms of the one-sided infinite sequence $\{a_k\}_{k=0}^{\infty}$, given as

$$A(\lambda) = (2\pi)^{-1/2} \sum_{k=0}^{\infty} a_k e^{-ik\lambda}, \tag{4}$$

with $a_k = 0$ for $k < 0$, and $|A(\lambda)|^2 = A(\lambda)A^*(\lambda)$. The quantity in (4) is also called a transfer function. Useful discussions are given in Priestley (1981, sec.9.2)).

We particularly interested in long-run information, which is reflected in the spectrum at the zero frequency,

$$f(0) = |A(0)|^2 f_e(0). \tag{5}$$

Then, we restrict our attention to the case when $f(0)=0$. Since $f_e(0)>0$, degeneracy arises from zero long-run impulse responses,

$$f(0)=0 \text{ is equivalent to } A(0)=0. \tag{6}$$

Under this degeneracy, we study the asymptotic behavior of the estimator of $f(0)$. Introduce a conventional kernel-based spectral density estimator at the zero frequency,

$$\widehat{f}(0)=(2\pi)^{-1}\textstyle\sum_{j=1-T}^{T-1}k(j/M)\widehat{R}(j), \tag{7}$$

where

$$\widehat{R}(j)=\frac{1}{T}\textstyle\sum_{t=|j|+1}^{T}(x_t-\overline{x})(x_{t-|j|}-\overline{x}),$$

where $\overline{x}=T^{-1}\sum_{t=1}^{T}x_t$, $k(x)$ is a kernel function and M is the bandwidth parameter.

Our main task is to derive the asymptotic distribution of the LRV estimator $\widehat{f}(0)$ out of linear processes. Following Phillips and Solo (1992), we first write

$$\begin{aligned} x_t x_{t-j} &= a(L)e_t a(L)e_{t-j} \\ &= \sum_{k=0}^{\infty} a_k a_{k-j} e_{t-k}^2 + \sum_{k=0}^{\infty}\sum_{r=j-k,\neq 0}^{\infty} a_k a_{k+r-j} e_{t-k} e_{t-k-r} \\ &= C_j(L)e_t^2 + \sum_{r=-\infty,\neq 0}^{\infty} C_{j-r}(L) e_t e_{t-r}, \end{aligned} \tag{8}$$

where $C_j(L)=\sum_{k=0}^{\infty}a_k a_{k-j}L^k$. Using the technique of Beveridge-Nelson decomposition, we put

$$C_j(L)=C_j(1)-(1-L)\widetilde{C}_j(L), \tag{9}$$

where $\widetilde{C}_j(L)=\sum_{k=0}^{\infty}\sum_{s=k+1}^{\infty}a_s a_{s+j}L^k=\sum_{k=0}^{\infty}\widetilde{C}_{jk}L^k$. Then, we have

$$\begin{aligned} x_t x_{t-j} &= C_j(1)e_t^2 + \sum_{r=-\infty,\neq 0}^{\infty}[C_{j-r}(1)e_t e_{t-r}] \\ &\quad -(1-L)\widetilde{C}_j(L)e_t^2-(1-L)\sum_{r=-\infty,\neq 0}^{\infty}[\widetilde{C}_{j-r}(L)e_t e_{t-r}]. \end{aligned} \tag{10}$$

The first and second terms contribute the long-term trends, whereas the remaining terms are related with short-run components. Denote the LRV estimator based on long-term trends as $\widehat{f}_L(0)$, and it can be written as

$$T^{1/2}\widehat{f}_L(0)=T^{-1/2}\sum_{t=1}^{T}(2\pi)^{-1}\sum_{j=1-T}^{T-1}k(j/M)\left\{\sum_{r=-\infty}^{\infty}C_{j-r}(1)[e_t e_{t-r}-Ee_t e_{t-r}]\right\}. \tag{11}$$

Then, by Brown's (1971) martingale limit theorem, we can obtain desired asymptotic normality. Below, we impose some technical conditions to derive the normality for $T^{1/2}\widehat{f}_L(0)$.

Assumption 1: (a) $k(x) : \mathbf{R} \to [-1, 1]$ is symmetric and continuous at zero with $k(0) = 1$, where $\mathbf{R}$ is the set of real numbers. The Fourier transform of $k(x)$ is defined as

$$K(\lambda) = (2\pi)^{-1} \int_{-\infty}^{\infty} k(x)e^{-i\lambda x}dx,$$

for all $\lambda \in [-\pi, \pi]$.

(b) $\int_{-\pi}^{\pi} \lambda^r K(\lambda)d\lambda = 0$, for $r = 1, 2, ..q - 1$, and $\neq 0$, for $r = q$.

Assumption 1 is a regularity condition for kernel functions, which is quite standard in the nonparametric literature. For more discussion, see Hannan (1971), Andrews (1991). The function $K(\lambda)$ is called spectral window generator, with the property that $\int_{-\pi}^{\pi} K(\lambda)d\lambda = 1$, which is equivalent to $k(0) = 1$. The moment condition on $K(\cdot)$ can be equivalently understood as derivative of its inverse Fourier transforms, $k(\cdot)$ evaluated at zero, that is to say, $d^r k(x)/dx^r|_{x=0} = 0$ for $r = 1, 2, ..q - 1$, and $\neq 0$, for $r = q$, where $k(x) = \int_{-\pi}^{\pi} K(\lambda)e^{ix\lambda}d\lambda$.

The smoothness of the kernel function is characterized by

$$k_q = \lim_{x\to 0} \frac{1 - k(x)}{|x|^q}, \qquad \text{for } q \in [0, \infty). \tag{12}$$

The positive value q needs not be an integer. Well-known quadratic kernels such as Parzen and quadratic spectral kernel satisfy the above conditions with $q = 2$. In our analysis, however, the value of q needs to be larger than 2, which is closely related with higher-order Taylor expansions of the estimators. In particular, we consider the following fourth-order kernel with $q = 4$,

$$K(\lambda) = \frac{15}{32\pi}[7(\frac{\lambda}{\pi})^4 - 10(\frac{\lambda}{\pi})^2 + 3], \tag{13}$$

see Gasser *et al* (1985). The fourth order kernel generates $\int_{-\pi}^{\pi} \lambda^4 K(\lambda)d\lambda = -4.6385$, which equals to $d^4k(x)/dx^4|_{x=0}$. Then, by Taylor expansions of $k(x)$ around $x = 0$, where $k(0) = 1$, $d^r k(x)/dx^r|_{x=0} = 0$ for $r = 1, 2, 3$, we obtain the smoothness index for the kernel in (6), $k_4 = (-1/24) \cdot (-4.6385) = 0.1932$. This value appears in the asymptotic bias of the kernel-based spectral density estimator. We use the above kernel for our simulations. Other higher-order kernels with $q > 4$ can be considered (e.g. Velasco and Robinson (2001)).

Next, we state some smoothness conditions for $f(\lambda)$ and for $A(\lambda)$ near the origin.

Assumption 2: $\sum_{j=-\infty}^{\infty} ||R(j)|| < \infty$, and $\sum_{j=-\infty}^{\infty} |j|^q R(j) < \infty$,for $q \in [0, \infty)$.

The smoothness for $f(\cdot)$ at the zero frequency is given by the q-th order generalized spectral derivative,

$$f^{(q)}(0) = (2\pi)^{-1} \sum_{j=-\infty}^{\infty} |j|^q R(j), \qquad \text{for } q \in [0, \infty), \tag{14}$$

where q may not be integer-valued. The generalized spectral derivative $f^{(q)}(0)$ is not necessarily equal to q-th derivative in the usual sense given by $f_{(q)}(0) = d^q f(\lambda)/d\lambda^q|_{\lambda=0}$. If q is even-numbered, then

$$f^{(q)}(0) = (-1)^{q/2} f_{(q)}(0). \tag{15}$$

For example, when $q = 2$, then $f^{(2)}(0) = -f_{(2)}(0)$. If $q = 4$, then $f^{(4)}(0) = f_{(4)}(0)$. The larger values of q, the smoother the spectral density function near the origin. Thus, when the true $f(0) = 0$, it is necessary to obtain higher-order expansions of the spectral density estimator near zero, which leads to a large value of q.

Assumption 3: Given the transfer function, $A(\lambda) = (2\pi)^{-1/2} \sum_{k=0}^{\infty} a_k e^{-ik\lambda}$, and $B(\lambda) = |A(\lambda)|^2 = (2\pi)^{-1} \sum_{k=0}^{\infty} \sum_{r=0}^{\infty} a_k a_r e^{-i(k-r)\lambda}$, the q-th derivatives of $B(\lambda)$ near the origin are given as $B_{(q)}(0) = d^q B(\lambda)/d\lambda^q|_{\lambda=0} < \infty$.

The smoothness condition on linear filters is analogous to the smoothness on the spectral density near the origin. Degeneracy entails higher order expansions for the spectral density and for the transfer function near the origin. Thus, it is necessary to introduce spectral derivatives given in Assumptions 2 and 3.

As a preliminary step, we study the mean squared properties of the LRV estimator when $f(0) = 0$. The asymptotic variance is particularly essential in dealing with the degenerate situation in our work.

Proposition 1: (Lee (2008)) Suppose Assumptions 1-2 hold, and $M = C \times T^{\alpha}$, where $0 < \alpha < \min[q/4, 1]$, for $q \geq 2$ and $0 < C < \infty$. Then, under $f(0) = 0$,

(a) $\lim_{N\to\infty}(T \times M^3) \times Var(\widehat{f}(0)) = V \equiv \pi f_{(2)}^2(0)\int_{-\pi}^{\pi} u^4 K^2(u)du$.
(b) If $M^q/T \to 0$, then $\lim_{T\to\infty} M^q E\widehat{f}(0) = -(2\pi)^{-1}k_q \sum_{j=-\infty}^{\infty} |j|^q R(j)$.

The part (a) comes from Lee (2008). The convergence rate of the asymptotic variance provides different results from the case of $f(0) > 0$. When $f(0) > 0$, conventional HAC estimator converges at the rate of T/M, where T is the sample size and M is the bandwidth. On the other hand, the rate of convergence of the variance of $\widehat{f}$ under the null of $f(0) = 0$ becomes much faster, at the rate of $T \cdot M^3$. Faster rate of convergence intuitively makes sense when the true spectrum is zero. This result is due to higher order expansions of the spectrum at the zero frequency, whereas such expansions are not considered when $f(0) > 0$ in the context of traditional spectral density estimations. Velasco and Robinson (2002) proves useful Edgeworth expansions of the HAC estimator, and though they assume $f(0) > 0$, provide the result that $Var(\widehat{f}(0)) = O(M^{1-r} \times T^{-1})$, for $r = 0, 1, 2, ...$ It is noted that conventional HAC estimators correspond to $r = 0$, but our results comes from the case of $r = 4$, to obtain valid convergences under the zero spectrum. The proof of part (a) is given in Lee (2008), whereas the part (b) is not new, following from Hannan (1970), for instance.

Note that for Proposition 1, we assume that x_t is a stationary Gaussian process. Gaussian assumptions facilitate derivation of asymptotic variance of spectral density estimators since fourth-order cumulants are simply removed. Without Gaussianity, however, one needs to estimate general dependence of fourth order cumulants, causing a great deal of difficulty (Velasco and Robinson (2001)).

3. Main Results

We present the asymptotic normality of the LRV estimator.

Theorem 1: Suppose Assumptions 1-3 hold, and $M = C \times T^{\alpha}$, and $1/(2q-3) < \alpha < 1/4$, and $0 < C < \infty$. If $f(0) = 0$, then

$$Z_T = \sqrt{T \times M^3}\widehat{f}(0) \to N(0, V),$$

where

$$V = \pi\sigma^4 B_{(2)}^2(0)\int_{-\pi}^{\pi} u^4 K^2(u)du = \pi f_{(2)}^2(0)\int_{-\pi}^{\pi} u^4 K^2(u)du.$$

The above asymptotic results, which are our main results, show new convergence rate and the asymptotic form of the variance for the LRV estimator. Under non-degeneracy, it can be inferred that

$$\sigma^2[|A(\lambda)|^2] = f(\lambda), \tag{16}$$

where $|A(\lambda)|^2 = A(\lambda)A^*(\lambda)$. The zero spectrum $f(0) = 0$ corresponds to $A(0) = 0$. Such degeneracy becomes intuitively quite clear. Since the quantity $A(0)$ can be regarded as long-run impulse response functions, zero long-run variance implied by $f(0) = 0$ is indeed associated with zero long-run impulse responses under linear models.

Under degeneracy, the non-zero second spectral derivative $f_2(0)$ determines the variance. Further, we can infer the associated quantity from linear processes. Let $B(\lambda) = |A(\lambda)|^2$, and we have

$$\sigma^4 B^2_{(2)}(0) = f^2_{(2)}(0). \tag{17}$$

where $B_{(2)}(0)$ is the second derivative of the squared transfer function $B(\lambda)$ at zero. The above relationship is based on the fact that the asymptotic variance V from linear processes should be equal to that from HAC estimator. The proof is given in the Appendix.

Besides, technical conditions on the bandwidth are derived through establishing asymptotic normality. In light of the Proposition 1, the bias terms asymptotically vanish if $(T \times M^3)^{1/2} M^{-q} \to 0$, as $N \to \infty$. While the condition on α effectively restricts the growth rate of the bandwidth. For example, if $q = 4$, then $1/5 < \alpha < 1/4$, and if $q = 6$, then $1/9 < \alpha < 1/5$. We note that the q needs be larger than that required in Proposition 1, which implies that stronger condition on smoothness should be necessary to construct the asymptotic normality.

Also, the asymptotic variance V involves unknown quantity, which can be estimated by $\widehat{f}_{(2)} = \pi^{-1} \sum_{j=-s}^{s} j^2 \widehat{R}(j)$, where s is the lag truncation. Alternatively, kernel weights can be employed in practice. Then, we can use $\widehat{V} = \pi \widehat{f}^2_{(2)}(0) \int_{-\pi}^{\pi} u^4 K^2(u) du$. For the fourth-order kernel, the value of $\int_{-\pi}^{\pi} u^4 K^2(u) du = 0.6582$.

Given the result of Theorem 1, we obtain a t-type test statistic. Under the null of degeneracy,

$$S_T = \widehat{V}^{-1/2} \sqrt{T \times M^3} \widehat{f}(0) \to N(0, 1). \tag{18}$$

Our test is applied to testing for trend stationarity. Under the null of trend stationarity, the spectrum near the origin for the first-differenced process becomes zero. On the other hand, under the alternative of difference stationarity, the spectrum becomes strictly positive at the zero frequency. In this regard, we can discuss the consistency of the test statistic. In particular, we compare powers of S_T and of tests by Kwiatkowski, D., Phillips, P.C.B., Schmidt, P., Shin, Y.(1992; KPSS hereafter) under fixed alternatives, $f(0) > 0$. We note that the KPSS test requires upper-bounded bandwidth to obtain consistency under fixed alternatives. The test diverges at the rates of $O(T^{2/3})$ for Bartlett kernel, and $O(T^{4/5})$ for QS kernel under iid innovations. On the other hand, our test diverges the rate of $O((T^{1/2}M^{3/2}))$. Thus, it is seen that the S_T diverges faster than Bartlett kernel-based KPSS test if $\alpha > 1/9$, and QS kernel-based KPSS if $\alpha > 1/5$. Such power advantages can be seen in simulation studies in the next section. Also, as a reference, Lee (2008) considers local power analysis, while local powers of the KPSS test are studies in Kurozumi (2002).

4. Simulation Studies

In this section, we investigate finite sample performance of the test statistic S_T, in comparison with the KPSS test. The simulation results of size and size-adjusted powers of the tests are reported in Table 1 through Table 4.

The data generating process that we consider is given as

$$y_t = \alpha + \beta t + z_t + u_t, \qquad z_t = z_{t-1} + v_t, \tag{19}$$

where $u_t = \rho u_{t-1} + e_t$, e_t and v_t are i.i.d. normal (0,1). We set $\alpha = 0$, and $\beta = 0.3$ without loss of generality. The autoregressive coefficient ρ takes values of 0, 0.2, 0.5 and 0.8. The innovations $(e_t, v_t)'$ are allowed to be contemporaneously correlated by setting $Ee_tv_t = 0.3$. Two sample sizes of $T = 200$, and 500 are included, and 10,000 iterations are performed.

For the S_T test, the fourth-order kernel-based spectral density estimator is used. The bandwidths for $\widehat{f}(0)$ takes the form, $M = C \cdot N^{\alpha}$, where we let α equal to 0.22 and constant C range from 1 to 3. Since no optimal choice for C is available in our case of degeneracy, we can employ sieve bootstrap method developed by Park (2006), which is particularly appealing when underlying process u_t is weakly dependent. Steps to obtain bootstrap critical values are as follows. First, given the null of $\sigma_v = 0$, we regress y_t on the intercept and linear trend to obtain residuals $\{\widehat{u}_t\}_{t=1}^{T}$. Second, we fit the model, $\widehat{u}_t = \widehat{\rho}\widehat{u}_{t-1} + \widehat{\varepsilon}_t$, and get the bootstrapped series $\{\varepsilon_t^*\}$ from demeaned residuals $\{\widehat{\varepsilon}_t - T^{-1}\sum_{t=1}^{T}\widehat{\varepsilon}_t\}_{t=1}^{T}$. Third, we generate bootstrapped $\{u_t^*\}_{t=1}^{T}$ from $u_t^* = \widehat{\rho}u_{t-1}^* + \varepsilon_t^*$, then we finally obtain bootstrap sample $y_t^* = \alpha + \beta t + u_t^*$. As seen in the simulations below, this method works reasonably well. It is also useful since size performances become less sensitive to the choice of the bandwidths than the case of using asymptotic critical values. See Park (2006) for theoretical results for the sieve bootstrap method. We note that bandwidth selections in case of degeneracy would be an open question.

Besides, it is crucial to choose the lag selection s to estimate the $f_{(2)}(0)$ and $f_4(0)$. We use a following simple procedure to select s, which performs reasonably well. Notice that the first differenced process follows MA(1), $x_t = \beta + u_t - u_{t-1}$. Denote the theoretical correlation of x_t as $cor(j) = R(j)/R(0)$. For example, when u_t is iid, then $cor(1) = -0.5$, and $cor(j) = 0$ for $j > 1$. On the other hand, if u_t is dependent, then $cor(j)$ becomes nonzero for some $j > 1$. When u_t follows AR(1) process, $u_t = \rho u_{t-1} + e_t$, then $cor(1) = -(1+2\rho+\rho^2)/(2+2\rho+2\rho^2)$, $cor(2) = \rho/(2+2\rho+2\rho^2)$, and $cor(j) = 0$ for $j > 2$, where the sum, $\sum_{j=1}^{2} cor(j) = -0.5$. Then, we select the lag selection $s = \widetilde{s}$ such that $\sum_{j=1}^{\widetilde{s}} \widehat{cor}(j)$ reaches to -0.5, where $\widehat{cor}(j)$ is a sample correlation. It is then expected that the integer-valued $\widetilde{s}$ is small if the data behaves closely to i.i.d., and becomes larger when the dependence becomes stronger. Our selection rule is given as $s = \min(\widetilde{s}, M)$, which does not affect the asymptotic results. Other methods can be explored in alternative ways. Note that conventional lag selection rules in the HAC context assuming positive spectrum near origin (e.g., Andrews (1991) and Newey and West (1994)) work poorly in our experiment.

As the KPSS tests involve HAC estimators, we use Bartlett and QS kernel-based estimators. In order to choose the bandwidth which makes the test consistent under the alternatives, we follow the procedures in Kurozumi (2002), which modifies Andrews' parametric plug-in methods. For example, the bandwidth is chosen as

$$M^* = \min[1.1447(\frac{4\widehat{a}^2 N}{(1+\widehat{a})^2(1-\widehat{a})^2})^{1/3}, 1.1447(\frac{4b_j^2 N}{(1+b_j)^2(1-b_j)^2})^{1/3}], \quad (20)$$

for Bartlett kernel, and

$$M^* = \min[1.3221(\frac{4\widehat{a}^2 N}{(1-\widehat{a})^4})^{1/5}, 1.3221(\frac{4b_j^2 N}{(1-b_j)^4})^{1/5}], \quad (21)$$

for QS kernel, where $\widehat{a}$ is autoregressive coefficient estimate in AR(1) approximating model (See Andrews(1991)), and b_j is a pre-specified constant. Such upper bound b_j makes the

KPSS test consistent (e.g., Choi and Ahn (1999)). We include three cases of $b_j = 0.7$, 0.8, and 0.9, for $j = 1, 2, 3$. As we will see below, different values of upper bounds critically affect the power performance of the KPSS test.

Table 1 and Table 2 present size performances of tests. The size of S_T is presented with a selective range of bandwidth, i.e., we only include constants $C = 1, 2$ and 3. Then, we can see how the size varies with this range of different bandwidths. Results are summarized as follows. First, when $T = 200$, our test equipped with sieve bootstrap method show reasonable size performance. As expected, results are insensitive to different values of bandwidth. Though not reported here, when we use asymptotic critical values, size of S_T sensitively depends on the choice of the bandwidths as well as lag selections. Thus, bootstrap method is appealing under dependence as in our experiment. Second, when $T = 500$, sizes become more distorted than the case of $T = 200$ in some cases. For instance, when $\rho = 0.2$, S_T underrejects a little for $C = 2$ and 3, whereas when $\rho = 0.8$, S_T rather overrejects for all values of C. Except for these cases, S_T show reasonable sizes in general.

The size of KPSS is pretty reasonable and become more distorted as dependence measure ρ increases. Further, larger values of upper bound tend to lead to better sizes. The power, however, can be negatively affected by large values of pre-set bound b_j for the KPSS test, as seen below. As KPSS test is not combined with any bootstrap method, we note that it is not fair to directly compare the size performance of the two tests. One can expect size improvements through bootstrapping KPSS tests, but it is not our main interests here. We instead concentrate on the power comparison between these two tests.

Next, we investigate the size-adjusted powers for the tests. We only report powers under i.i.d. errors and under $\rho = 0.5$ in Table 3 and 4, respectively. For the powers, we gradually change the signal to noise ratio $\lambda = \sigma_v^2/\sigma_u^2$ from 0.1 up to 100. In table 3, the upper panel includes powers when $T = 200$, and the lower panel when $T = 500$. First, S_T clearly dominates KPSS for values of $\lambda \geq 0.5$. As λ grows, S_T monotonically reaches to unit power regardless of the choice of bandwidths. Changes in powers of S_T due to different values of bandwidths becomes negligible when $\lambda \geq 1$. Second, for very small departures out of the null, say, $\lambda = 0.1$ or 0.2, KPSS are more powerful in some cases than S_T, but their powers abruptly retreat and maintain to a certain level, which is strictly lower than unity as λ continues to increase. The power of KPSS also crucial depends on pre-set values of the upper bounds. It is clearly pronounced that larger the values of the bound b, lower the powers. These unfavorable features of power properties of the KPSS arise mainly from pre-set upper bounds in the bandwidths (See Kurozumi (2002)). Third, as the sample size increases to $T = 500$, powers of the tests qualitatively remain unchanged. Our test quickly climbs up to unity as long as $\lambda \geq 0.5$, where it dominates KPSS regardless of the bandwidth choice.

The Table 4 reports powers of the tests when the errors follow AR(1) with $\rho = 0.5$. Dependence certainly lowers powers of both S_T and KPSS. Other than that, powers of S_T and KPSS show qualitatively same patterns as those in Table 3. We can still obtain power advantages of S_T over the KPSS test. Our test have lower powers for very small λ than KPSS, but rapidly dominate the KPSS tests as λ continues to grow. In sum, our proposed test can be a useful complement to KPSS test or its variation.

5. Conclusion

We provide the asymptotic results for the long-run variance estimator when true power spectrum equals to zero, where conventional results of HAC estimation can not be applied. Formally, we investigate degeneracy issue in linear processes. While long-run variance or long-term trend is associated with the spectrum power evaluated at the zero frequency in spectral analysis, our analysis is based on higher-order expansions of the long-run variance. Asymptotic mean squared errors and normality based on long-term trends is explicitly derived. It provides convergence rates of long-run variance estimator, which are clearly different from those in HAC estimators. Further, we propose a t-test for degeneracy, which converges to standard normal distribution under the null hypothesis of degeneracy near the origin. Evidence is found that the tests deliver better power performances than prevailing KPSS test statistic. In particular, when the signal-to-noise ratio is very small, KPSS is more powerful, whereas when the ratio grows, the proposed test dominates the KPSS. Thus, our test can serve as a complement to the KPSS test. Further research direction may include degeneracy issues in multivariate cases and in panel regression models.

References

Andrews, D.W.K., 1991, Heteroskedasticity and autocorrelation consistent covariance matrix estimation, Econometrica, 59, 817-858.

Brockwell, P. and R. Davis, 1991, Time Series: Theory and Methods, second edition: Springer-Verlag.

Brown, B., 1971, Martingale central limit theorems, Annals of Mathematical Statistics, 42, 59-66.

Choi, I. and B.C. Ahn, 1999, Testing the null of stationarity for multiple time series, Journal of Econometrics, 88, 41-77.

Gasser, T., H-G. Müller and V. Mammitzsch, 1985, Kernels for nonparametric curve estimation, Journal of Royal Statistical Society B, 47, 238-252.

Hannan, E.J., 1970, Multiple Time Series: New York, Wiley.

Kurozumi, E., 2002, Testing for stationarity with a break, Journal of Econometrics, 108, 63-99.

Kwiatkowski, D., P.C.B. Phillips, P. Schmidt and Y. Shin, 1992, Testing the null of stationarity against the alternative of a unit root: how sure are we that economic time series have a unit root? Journal of econometrics, 54, 159-178.

Lee, J., 2008, Testing for trend stationarity using spectral density estimators, manuscript, Ewha Womans University.

Newey, W. and K. West, 1994, Automatic lag selection in covariance matrix estimation, Review of Economic Studies, 61, 631-653.

Phillips, P.C.B. and V. Solo, 1992, Asymptotics for linear processes, Annals of Statistics, 20, 971-1001.

Phillips, P.C.B., Y. Sun, and S. Jin, 2006, Spectral density estimation and robust hypothesis testing using steep origin kernels without truncation, International Economic Review, 47, 837-894.

Priestley, M. B., 1981, Spectral Analysis and Time Series: New York: Academic Press.

Robinson, P.M., 1991, Automatic frequency domain inference on semiparametric and nonparametric models, Econometrica, 59, 755-786.

Taniguchi, M. and Y. Kakizawa, 2000, Asymptotic theory of statistical inference for time series: New York, Springer-Verlag.

Velasco, C., and P.M. Robinson, 2001, Edgeworth expansions for spectral density estimates and studentized sample mean, Econometric Theory, 17, 497-539.

Appendix: Proofs

Proof of Theorem 1: We derive the asymptotic normality of the LRV estimator from linear process $\{x_t\}_{t=1}^T$. For linear processes, we employ useful results in Phillips and Solo (1992, eq.(28); PS hereafter). Put

$$\begin{aligned} x_t x_{t-j} &= a(L)e_t a(L)e_{t-j} \\ &= \sum_{k=0}^{\infty} a_k a_{k-j} e_{t-k}^2 + \sum_{k=0}^{\infty} \sum_{r=j-k,\neq 0}^{\infty} a_k a_{k+r-j} e_{t-k} e_{t-k-r} \\ &= C_j(L)e_t^2 + \sum_{r=-\infty,\neq 0}^{\infty} C_{j-r}(L) e_t e_{t-r}, \end{aligned} \tag{A1}$$

where $C_j(L) = \sum_{k=0}^{\infty} a_k a_{k-j} L^k$.

Using the technique of Beveridge-Nelson decomposition,

$$C_j(L) = C_j(1) - (1-L)\widetilde{C}_j(L), \tag{A2}$$

where $\widetilde{C}_j(L) = \sum_{k=0}^{\infty} \sum_{s=k+1}^{\infty} a_s a_{s+j} L^k = \sum_{k=0}^{\infty} \widetilde{C}_{jk} L^k$. Then,

$$\begin{aligned} x_t x_{t-j} &= C_j(1)e_t^2 + \sum_{r=-\infty,\neq 0}^{\infty} [C_{j-r}(1) e_t e_{t-r}] \\ &\quad -(1-L)\widetilde{C}_j(L)e_t^2 - (1-L) \sum_{r=-\infty,\neq 0}^{\infty} [\widetilde{C}_{j-r}(L) e_t e_{t-r}]. \end{aligned}$$

Put $\widehat{f}(0) = (2\pi T)^{-1} \sum_{t=1}^T x_t^2 + (\pi T)^{-1} \sum_{j=1}^{T-1} k(j/M) \sum_{t=1+j}^T x_t x_{t-j}$, then

$$\begin{aligned} & T^{1/2}\widehat{f}(0) \\ =\; & (2\pi)^{-1} T^{-1/2} \sum_{t=1}^T x_t^2 + \pi^{-1} T^{-1/2} \sum_{j=1}^{T-1} k(j/M) \sum_{t=1}^T x_t x_{t-j} \\ =\; & A_{1T} + A_{2T}, \end{aligned} \tag{A3}$$

where $x_j = 0$ for $j \leq 0$.

We then decompose A_{1T} and A_{2T} to extract the martingale sequences, as in PS. Let $e_j = 0$, for $j \leq 0$, then

$$A_{1T} = (2\pi)^{-1} T^{-1/2} \sum_{t=1}^T x_t^2 = U_{1T} - D_{1T} - D_{2T}, \tag{A4}$$

where

$$\begin{aligned}
U_{1T} &= (2\pi)^{-1}T^{-1/2}\sum_{t=1}^{T}\left[C_0(1)e_t^2 + \sum_{r=-\infty,\neq 0}^{\infty} C_{-r}(1)e_te_{t-r})\right] \\
D_{1T} &= (2\pi)^{-1}T^{-1/2}\widetilde{C}_0(L)e_T^2 \\
D_{2T} &= (2\pi)^{-1}T^{-1/2}\sum_{r=-\infty,\neq 0}^{\infty}[\widetilde{C}_{-r}(L)]e_Te_{T-r},
\end{aligned}$$

where the expressions for D_{1T} and D_{2T} arise when summed over t from 1 to T. Also for A_{2T},

$$A_{2T} = \pi^{-1}\sum_{j=1}^{T-1}k(j/M)T^{-1/2}\sum_{t=1}^{T}x_tx_{t-j} = U_{2T} - D_{3T} - D_{4T}, \tag{A5}$$

where

$$\begin{aligned}
U_{2T} &= \pi^{-1}\sum_{j=1}^{T-1}k(j/M)T^{-1/2}\sum_{t=1}^{T}\left[C_j(1)e_t^2 + \sum_{r=-\infty,\neq 0}^{\infty} C_{j-r}(1)e_te_{t-r}\right], \\
D_{3T} &= \pi^{-1}\sum_{j=1}^{T-1}k(j/M)T^{-1/2}\widetilde{C}_j(L)e_T^2, \\
D_{4T} &= \pi^{-1}\sum_{j=1}^{T-1}k(j/M)T^{-1/2}\sum_{r=-\infty,\neq 0}^{\infty}\widetilde{C}_{j-r}(L)e_Te_{T-r}.
\end{aligned}$$

It can be easily shown that $D_{jN} \to^p 0$, for $j = 1, 2, 3, 4$. Note that $D_{1T} = D_{3T}$ for $j = 0$, and

$$\begin{aligned}
E[\widetilde{C}_j(L)e_T^2]^2 &= E[\sum_{k=0}^{\infty}\widetilde{C}_{jk}L^ke_T^2]^2 \\
&= \sum_{k=0}^{\infty}\sum_{k'=0}^{\infty}\widetilde{C}_{jk}\widetilde{C}_{jk'}E(e_{T-k}^2e_{T-k'}^2) \\
&= \sum_{k=0}^{\infty}\widetilde{C}_{jk}^2\mu_4 + \sum_{k=0}^{\infty}\sum_{k'=0,\neq k}^{\infty}\widetilde{C}_{jk}\widetilde{C}_{jk'}\sigma^2,
\end{aligned}$$

where $\widetilde{C}_{jk} = \sum_{s=k+1}^{\infty}a_sb_{s+j}$, $\mu_4 = Ee_t^4$ and $\sigma^2 = Ee_t^2$. By lemma 3.6 in PS and Cauchy-Schwarz inequality, $E[\widetilde{C}_j(L)e_T^2]^2 = O(1)$. It follows that

$$E(D_{3T}^2) = \pi^{-2}T^{-1}\sum_{j=1}^{T-1}\sum_{j'=1}^{T-1}k(j/M)k(j'/M)E[\widetilde{C}_j(L)e_T^2]^2 = O(M/T), \tag{A6}$$

where $M^{-1}\sum_{j=1}^{T-1}\sum_{j'=1}^{T-1}k(j/M)k(j'/M) = O(1)$. Similarly, $D_{2T} = D_{4T}$ for $j = 0$, and by similar reasonings in lemma 5.9 in PS to have $ED_{4T}^2 = O(M/T)$. It follows that $E(D_{jT}^2) = o(1)$, for $j = 1, 2, 3, 4$, given that $M/T \to 0$.

Now consider demeaned process, $U_{1T} + U_{2T} - EU_{1T} - EU_{2T}$. Write

$$[U_{1T} + U_{2T} - EU_{1T} - EU_{2T}] = T^{-1/2} \sum_{t=1}^{T} v_t, \tag{A7}$$

where

$$v_t = (2\pi)^{-1} \sum_{j=1-T}^{T-1} k(j/M) \left\{ \sum_{r=-\infty}^{\infty} C_{j-r}(1)[e_t e_{t-r} - Ee_t e_{t-r}] \right\}.$$

Note that v_t is a martingale difference sequence with

$$\lambda = T^{-1} \sum_{t=1}^{T} Ev_t^2 \to \Omega, \tag{A8}$$

Thus, we can employ Brown's (1971) martingale limit theorem to obtain

$$T^{-1/2} \sum_{t=1}^{T} v_t \to^d N(0, \Omega). \tag{A9}$$

The Lindberg conditions are (a) $\lambda^{-1}T^{-1}\sum_{t=1}^{T} E[v_t^2 1(v_t^2 > \epsilon T\lambda)] \to 0$, for all $\epsilon > 0$, and (b) $\lambda^{-1}T^{-1}\sum_{t=1}^{T} E[v_t^2|F_{t-1}] \to^p 1$. The condition (a) comes from dominated convergence theorem (equation (50) in PS), and the condition (b) is obvious since $E\{E[v_t^2|F_{t-1}]\} = Ev_t^2 = \Omega$ (See Theorem 3.8 in PS).

Given the martingale difference sequence v_t, we obtain

$$\begin{aligned} & Ev_t^2 \\ = \; & (2\pi)^{-2} \sum_{j=1-T}^{T-1} k(j/M) \sum_{j'=1-T}^{T-1} k(j'/M) \sum_{r=-\infty}^{\infty} \sum_{s=-\infty}^{\infty} C_{j-r}(1) C_{j'-s} \times \\ & E[e_t e_{t-r} - Ee_t e_{t-r}] \times [e_t e_{t-s} - Ee_t e_{t-s}] \\ = \; & (2\pi)^{-2}\sigma^4 \sum_{j=1-T}^{T-1} \sum_{j'=1-T}^{T-1} k(j/M)k(j'/M) \sum_{r=0}^{\infty} [C_{j-r}(1)C_{j'-r} + C_{j+r}C_{j'+r}] \\ = \; & \xi_1 + \xi_2, \end{aligned} \tag{A10}$$

where the second line follows from the property of $\{e_t\}$ sequence and from the fact that $E[e_t e_{t-r} - Ee_t e_{t-r}]^2 = 2\sigma^4$ if $r = 0$, and $= \sigma^4$ if $r \neq 0$ under Gaussian assumption.

Given $C_j(1) = \sum_{k=0}^{\infty} a_k a_{k-j}$, and the Fourier transforms of $\{a_k\}$ is given by

$$A(\lambda) = (2\pi)^{-1/2} \sum_{k=0}^{\infty} a_k e^{-ik\lambda}, \quad -\pi \leq \lambda \leq \pi, \tag{A11}$$

where $a_k = 0$ for $k < 0$. Then, we can make use of the following relations,

$$\begin{aligned} \sum_{r=-\infty}^{\infty} C_{j-r}(1) &= \sum_{r=-\infty}^{\infty}\sum_{k=0}^{\infty} a_k a_{k-j+r} \qquad \text{(A12)} \\ &= 2\pi \left[(2\pi)^{-1/2}\sum_{k=0}^{\infty} a_k\right]\left[(2\pi)^{-1/2}\sum_{r=-\infty}^{\infty} a_{k-j+r}e^{-i(k-j+r)\lambda}\right] e^{i(k-j+r)\lambda} \\ &= 2\pi e^{i(r-j)\lambda}|A(\lambda)|^2, \end{aligned}$$

where $|A(\lambda)|^2 = A(\lambda)A^*(\lambda)$. Then, rearrange the terms in (A10) to obtain

$$\begin{aligned} &\xi_1 \qquad \text{(A13)} \\ =\ & \frac{\sigma^4}{(2\pi)^2}\sum_{j=1-T}^{T-1} k(j/M)\sum_{r=0}^{\infty} C_{j-r}(1)\left[\int_{-\pi}^{\pi} MK(M\lambda)\sum_{j'=1-T}^{T-1} e^{ij'\lambda}\sum_{k=0}^{\infty} a_k a_{k-(j'-r)}d\lambda\right] \\ =\ & \frac{\sigma^4}{(2\pi)^2}\sum_{j=1-T}^{T-1} k(j/M)\sum_{r=0}^{\infty} C_{j-r}(1)\left[2\pi\int_{-\pi}^{\pi} MK(M\lambda)|A(\lambda)|^2 e^{ir\lambda}d\lambda\right] \\ =\ & \frac{\sigma^4}{(2\pi)^2}2\pi\int_{-\pi}^{\pi} MK(M\lambda)|A(\lambda)|^2\sum_{j=1-T}^{T-1} k(j/M)\sum_{r=0}^{\infty}\sum_{k=0}^{\infty} a_k a_{k-j+r}e^{ir\lambda}d\lambda \\ =\ & \frac{\sigma^4}{(2\pi)^2}(2\pi)^3\int_{-\pi}^{\pi} MK(M\lambda)|A(\lambda)|^2\left[(2\pi)^{-1}\sum_{j=1-T}^{T-1} k(j/M)e^{ij\lambda}\right]\times \\ & \left[(2\pi)^{-1/2}\sum_{k=0}^{\infty} a_k e^{-ik\lambda}\right]\left[(2\pi)^{-1/2}\sum_{r=0}^{\infty} a_{k-j+r}e^{i(k-j+r)\lambda}\right]d\lambda \\ =\ & 2\pi\sigma^4\int_{-\pi}^{\pi} B(\lambda)^2 M^2 K^2(M\lambda)d\lambda, \end{aligned}$$

where real-valued $B(\lambda) = |A(\lambda)|^2$. We similarly obtain the identical expression for ξ_2.

Below, we obtain the expression for the variance of the HAC estimator, which is written as

$$Var[\widehat{f}(0)] = (2\pi)^{-2}\sum_{j=1-T}^{T-1}\sum_{j'=1-T}^{T-1} k(j/M)k(j'/M)Cov[\widehat{R}(j),\widehat{R}(j'), \qquad \text{(A14)}$$

where $Cov[\widehat{R}(j),\widehat{R}(j')$ denotes the covariance between $\widehat{R}(j)$ and $\widehat{R}(j')$. See Hannan (1970, p.313), Priestley (1981, p.326).

It can be decomposed as

$$T * Var[\widehat{f}(0)] = (V_{1T} + V_{2T})(1 + o(1)) \qquad \text{(A15)}$$

where

$$V_{1T} = \frac{1}{(2\pi)^2 T} \sum_{j=1-T}^{T-1} \sum_{j'=1-T}^{T-1} k(j/M)k(j'/M) \sum_{h=1-T}^{T-1} R(h)R(h+j'-j),$$

$$V_{2T} = \frac{1}{(2\pi)^2 T} \sum_{j=1-T}^{T-1} \sum_{j'=1-T}^{T-1} k(j/M)k(j'/M) \sum_{h=1-T}^{T-1} R(h+j')R(h-j).$$

It follows that

$$V_{1T} + V_{2T} = 2\pi \int_{-\pi}^{\pi} f^2(\lambda) M^2 K^2(M\lambda) d\lambda + o(1). \quad \text{(A16)}$$

The asymptotic variances in (A13) and (A16) should be equal, that is to say,

$$\Omega = 4\pi\sigma^4 \int_{-\pi}^{\pi} B^2(\lambda) M^2 K^2(M\lambda) d\lambda = 4\pi \int_{-\pi}^{\pi} f^2(\lambda) M^2 K^2(M\lambda) d\lambda. \quad \text{(A17)}$$

Now, we proceed the arguments to the case of degeneracy near origin, $f(0) = 0$. Through Taylor expansions for the variance in the right-hand side in (A17),

$$\begin{aligned}
& 4\pi \int_{-\pi}^{\pi} f^2(\lambda) M^2 K^2(M\lambda) d\lambda \qquad \text{(A18)} \\
= \; & 4\pi M \int_{-\pi}^{\pi} f^2(u/M) K^2(u) du \\
= \; & 4\pi M \Big\{ f^2(0) \int_{-\pi}^{\pi} K^2(u) du + 2f(0) f_{(1)}(0) \int_{-\pi}^{\pi} (u/M) K^2(u) du + \\
& [f_{(1)}(0)^2 + f(0) f_{(2)}(0)] \int_{-\pi}^{\pi} (u/M)^2 K^2(u) du + \\
& \frac{1}{6} [6 f_{(1)}(0) f_{(2)}(0) + 2 f(0) f_{(3)}(0)] \int_{-\pi}^{\pi} (u/M)^3 K^2(u) du + \\
& \frac{1}{24} [6 f_{(2)}^2(0) + 8 f_{(1)}(0) f_{(3)}(0) + 2 f(0) f_{(4)}(0)] \int_{-\pi}^{\pi} (u/M)^4 K^2(u) du \Big\} + O(M^{-5}) \\
= \; & \pi M^{-3} f_{(2)}^2(0) \int_{-\pi}^{\pi} u^4 K^2(u) du + O(M^{-5}),
\end{aligned}$$

where $f_{(j)}(0)$ is $j-th$ spectral derivative at zero frequency.

Then, we have

$$\begin{aligned}
(T \times M^3) \times Var[\widehat{f}(0)] &= \pi f_{(2)}^2(0) \int_{-\pi}^{\pi} u^4 K^2(u) du + o(1) \qquad \text{(A19)} \\
&= V + o(1),
\end{aligned}$$

where V is as in Theorem 1. Thus, we obtain

$$\sigma^4 B_{(2)}^2(0) = f_{(2)}^2(0), \text{ or } \sigma^2 |B_{(2)}(0)| = |f_{(2)}(0)|. \quad \text{(A20)}$$

This completes the proof of Theorem 1.

Table 1. Rejection Rates in Percentages of the Test Statistics under the Null: Sample Size $T = 200$ and AR(1) Model for errors

$\rho =$		0 (iid)		0.2		0.5		0.8	
		10%	5%	10%	5%	10%	5%	10%	5%
S_T	C								
	1	10.33	4.63	9.59	5.19	12.18	7.65	12.81	7.74
	2	11.28	4.30	10.39	4.90	10.50	5.08	12.18	6.90
	3	10.59	3.53	9.31	4.25	11.92	5.49	14.39	9.03
$K_{BT}(b_1)$		10.19	4.93	14.24	7.09	16.06	8.39	24.29	13.05
$K_{QS}(b_1)$		10.07	4.94	12.55	6.28	13.24	6.78	18.13	8.97
$K_{BT}(b_2)$		10.19	4.93	14.24	7.09	16.06	8.39	20.68	9.99
$K_{QS}(b_2)$		10.07	4.94	12.55	6.28	13.24	6.78	14.78	6.33
$K_{BT}(b_3)$		10.19	4.93	14.24	7.09	16.06	8.39	20.12	9.55
$K_{QS}(b_3)$		10.07	4.94	12.55	6.28	13.24	6.78	14.52	5.96

(1) DGP: $y_t = \beta_0 + \beta_1 t + z_t + u_t$, and $z_t = z_{t-1} + v_t$, where v_t is i.i.d. normal$(0, \sigma_v^2)$, $u_t = \rho u_{t-1} + e_t$, and e_t is i.i.d. Normal(0,1). The innovations are correlated as $E e_t v_t = 0.3$. We set $\beta_0 = 0$, $\beta_1 = 0.3$. The value $\sigma_v^2 = 0$ for the size.

(2) S_T : Spectrum-based test with the fourth-order kernel-based spectral density estimator. The bandwidths $M = C \cdot T^{\alpha}$, with $\alpha = 0.22$ and C takes values from 1 to 3. Bootstrapped critical values are used through sieve methods by Park (2006).

(3) $K_{BT}(b_j)$, $K_{QS}(b_j)$: KPSS tests with HAC estimators using Bartlett kernel and QS kernel, where the upper bounds for the bandwidth are given as $b_1 = 0.7$, $b_2 = 0.8$ and $b_3 = 0.9$, (See Kurozumi (2002), Andrews(1991)).

Table 2. Rejection Rates in Percentages of the Test Statistics under the Null: Sample Size $T = 500$ and AR(1) Model for errors

$\rho =$		0 (iid)		0.2		0.5		0.8	
		10%	5%	10%	5%	10%	5%	10%	5%
S_T	C								
	1	16.82	10.64	10.40	6.20	18.49	11.20	24.93	14.44
	2	12.85	6.72	5.97	3.12	10.36	5.07	16.15	8.06
	3	10.64	5.92	6.47	2.27	8.67	3.80	21.21	14.12
$K_{BT}(b_1)$		10.49	5.70	13.37	7.30	14.86	7.84	20.19	11.43
$K_{QS}(b_1)$		10.48	5.73	12.31	6.70	12.71	6.67	17.03	9.07
$K_{BT}(b_2)$		10.49	5.70	13.37	7.30	14.86	7.84	17.55	8.65
$K_{QS}(b_2)$		10.48	5.73	12.31	6.70	12.71	6.67	13.82	6.53
$K_{BT}(b_3)$		10.49	5.70	13.37	7.30	14.86	7.84	17.17	8.49
$K_{QS}(b_3)$		10.48	5.73	12.31	6.70	12.71	6.67	13.58	6.43

(1) See the explanatory notes in Table 1.

Table 3. Size-adjusted Powers in Percentages of the Test Statistics under the Alternatives At the 5% Level: Sample Size $T = 200$ and 500 under i.i.d. Normal Errors

$T = 200 : \lambda =$		0.1	0.2	0.5	1	2	5	10	100
S_T ($M = CT^{0.22}$)									
C	1	53.73	78.78	95.77	99.26	99.88	99.99	100	100
	2	74.20	82.75	89.10	94.45	97.35	98.55	98.82	99.08
	3	76.50	78.00	69.08	73.77	81.17	86.06	87.50	88.84
KPSS	$K_{BT}(b_1)$	87.99	80.83	73.90	73.07	73.07	73.26	73.28	73.32
	$K_{QS}(b_1)$	86.13	76.04	66.08	65.21	65.43	65.62	65.66	65.56
	$K_{BT}(b_2)$	87.91	78.68	62.89	58.62	57.88	57.80	57.86	57.88
	$K_{QS}(b_2)$	85.83	71.98	50.62	45.59	45.07	44.88	44.84	44.71
	$K_{BT}(b_3)$	87.91	78.05	55.17	42.52	37.49	35.72	35.37	35.15
	$K_{QS}(b_3)$	85.80	70.62	37.03	20.68	14.87	12.88	12.59	12.58
$T = 500$									
S_T ($M = CT^{0.22}$)									
C	1	86.08	95.60	99.87	99.99	100	100	100	100
	2	95.70	94.89	98.27	99.68	99.91	99.95	99.96	99.97
	3	96.65	84.84	89.32	95.37	97.69	98.62	98.91	99.06
KPSS	$K_{BT}(b_1)$	93.15	90.00	89.62	89.68	89.66	89.68	89.70	89.65
	$K_{QS}(b_1)$	93.94	90.08	89.51	89.56	89.56	89.56	89.53	89.58
	$K_{BT}(b_2)$	92.26	84.72	81.30	81.29	81.30	81.33	81.39	81.39
	$K_{QS}(b_2)$	92.78	83.34	78.72	78.59	78.60	78.70	78.74	78.74
	$K_{BT}(b_3)$	92.06	80.29	63.96	60.53	60.26	60.31	60.30	60.28
	$K_{QS}(b_3)$	92.31	75.98	53.74	50.01	49.59	49.56	49.59	49.58

(1) Empirical critical values for each test statistic are obtained from simulations in Table 1 and 2.

(2) The value $\lambda = \sigma_v^2/\sigma_u^2 > 0$ for the powers.

(3) See the explanatory notes in Table 1.

Table 4. Size-adjusted Powers in Percentages of the Test Statistics under the Alternatives At the 5% level: Sample Size $T = 200$ and 500 under AR(1) errors with $\rho = 0.5$

$T = 200 : \lambda =$		0.1	0.2	0.5	1	2	5	10	100
S_T $(M = CT^{0.22})$									
$C =$	1	11.66	19.11	38.82	60.15	77.94	89.41	92.98	96.45
	2	36.59	58.64	82.77	92.05	96.00	97.83	98.44	98.76
	3	41.67	57.46	76.75	85.71	90.32	92.58	93.45	94.53
KPSS	$K_{BT}(b_1)$	55.66	58.96	61.45	62.47	63.04	63.29	63.50	63.52
	$K_{QS}(b_1)$	52.68	54.84	57.46	58.45	58.85	59.17	59.28	59.24
	$K_{BT}(b_2)$	48.75	45.61	44.74	45.14	45.51	45.68	45.74	45.59
	$K_{QS}(b_2)$	43.84	38.92	37.52	37.79	37.97	38.16	38.26	38.30
	$K_{BT}(b_3)$	46.58	36.13	24.71	21.28	20.14	19.83	19.71	19.69
	$K_{QS}(b_3)$	39.17	22.35	7.92	4.23	3.26	3.01	2.96	2.96
$T = 500$									
S_T $(M = CT^{0.22})$									
C	1	24.95	49.23	86.10	97.27	99.54	99.96	99.99	99.99
	2	71.76	91.50	99.08	99.84	99.94	99.98	99.98	99.99
	3	76.90	89.50	97.59	99.02	99.57	99.72	99.78	99.84
KPSS	$K_{BT}(b_1)$	84.71	85.82	86.32	86.67	86.85	86.96	86.99	87.06
	$K_{QS}(b_1)$	85.80	86.72	87.46	87.70	87.92	88.09	88.10	88.16
	$K_{BT}(b_2)$	76.26	76.15	76.74	76.97	77.12	77.10	77.05	77.05
	$K_{QS}(b_2)$	75.70	75.26	75.84	76.08	76.13	76.17	76.18	76.15
	$K_{BT}(b_3)$	66.62	56.11	52.67	52.65	52.77	52.96	53.00	53.05
	$K_{QS}(b_3)$	62.77	49.55	45.34	45.10	45.12	45.14	45.15	45.16

(1) Empirical critical values for each test statistic are obtained from simulations in Table 1 and 2.

(2) DGP: The value $\lambda = \sigma_v^2/\sigma_u^2$, where $\sigma_u^2 = 1/(1-\rho^2)$.

(3) See the explanatory notes in Table 3.