《人文及社會科學集刊》
第十三卷第四期 (90/12), pp. 401-416

Alternative Proof for the Consistency of the KPSS Tests Against Fractional Alternatives

Wen-Jen Tsay
The Institute of Economics
Academia Sinica, Taiwan

ABSTRACT

This paper investigates the asymptotic properties of the semiparametric long run variance estimator when we demean or detrend a stationary $I(d)$ process. The analytic results are used to show that the KPSS (Kwiatkowski, *et al.*, 1992) test of the $I(0)$ null is consistent against $I(d)$ alternatives when the growth rate of the bandwidth parameter l is $o(T^{1/4})$.

Key Words: Semiparametric long run variance estimator; long memory process; fractionally integrated process.

1. Introduction

There is much research concerning the semiparametric estimation of the long run variance $\Omega=\sum_{j=-\infty}^{\infty}\gamma_j$ of a weakly dependent process ϵ_t, where γ_j is the autocovariance function of ϵ_t at lag j. For example, to account for the autocorrelation in ϵ_t, Newey and West (1987), Andrews (1991), and Hansen (1992) considered the long run variance estimator, $\hat{\Omega}_l$, to be defined as

Received: December 15, 2000; Accepted: June 19, 2001

$$\hat{\Omega}_l = T^{-1}\sum_{t=1}^{T} e_t^2 + 2T^{-1}\sum_{j=1}^{l} k\left(\frac{j}{l+1}\right)\sum_{t=j+1}^{T} e_t e_{t-j}, \tag{1}$$

where e_t can be the raw data (ϵ_t) itself, or the residuals from the regression model with stochastic or deterministic trends; the kernel weights $k(.)$ are assumed to satisfy some specific conditions (for more details, please refer to Andrews, 1991, p.821); and l is the bandwidth parameter and depends upon the sample size T. The major concern of the current literature is to find the conditions under which we can prove that $\hat{\Omega}_l \xrightarrow{p} \Omega$. However, the asymptotic properties of $\hat{\Omega}_l$ under various choices of kernel function and bandwidth parameter are also of interest to researchers.

We have recently witnessed fast-growing studies on the $I(d)$ process, that is, the integrated process of order d where d is a fractional number. The main feature of the $I(d)$ process is that its autocovariance function declines at a slower hyperbolic rate (instead of the geometric rate found in the conventional ARMA processes). Many test statistics have been reevaluated when the data generating process (DGP) is an $I(d)$ process. For example, Diebold and Rudebusch (1991) showed that the *simple* Dickey-Fuller unit root tests have low power against $I(d)$ alternatives. Moreover, Lo (1991) used the modified rescaled range (MRS) statistic to test the $I(0)$ null of the asset return against $I(d)$ alternatives, where $d \in (-0.5, 0) \cup (0, 0.5)$.

However, Lo (1991) only demonstrated the exact proof for cases $d > 0$ when $k(.)$ is the Bartlett kernel. The proof for cases $d < 0$ was omitted and left to the readers. Nevertheless, it is questionable to apply Lo's arguments to cases $d < 0$, because the bandwidth parameter $l = o(T)$ chosen by Lo (1991) is not accurate.

Another example is when Lee and Schmidt (1996) analyzed the power of the KPSS tests against $I(d)$ alternatives by calculating the asymptotic properties of $\hat{\Omega}_l$ as ϵ_t is an $I(d)$ process. However, instead of investigating $\hat{\Omega}_l$ itself, Lee and Schmidt (1996) utilized the Bartlett kernel to study the population counterpart of $\hat{\Omega}_l$, i.e., Ω_l, which is defined as

$$\Omega_l = \gamma_0 + 2\sum_{j=1}^{l}\left(1 - \frac{j}{l+1}\right)\gamma_j.$$

Therefore, the proof derived by Lee and Schmidt (1996) is not complete, because they did not show that the difference between $\hat{\Omega}_l$ and Ω_l diminishes asymptotically.

The results mentioned previously show that the asymptotic properties of $\hat{\Omega}_l$ are crucial to determine the consistency of the KPSS and MRS tests against $I(d)$ alternatives. We thus investigate the asymptotic properties of $\hat{\Omega}_l$ when the DGP is the $I(d)$ process, and use the analytic results to prove the consistency of the KPSS tests against $I(d)$ alternatives. Moreover, to extend the coverage of our analysis, we impose very mild restrictions on the kernel function. Our results reveal that the bandwidth parameter must be $o(T^{1/4})$ to ensure the consistency of the KPSS test against $I(d)$ alternatives. Nevertheless, Andrews (1991) showed that the Bartlett kernel cannot result in an efficient long run variance estimator for the weakly dependent process if l is not $O(T^{1/3})$. To obtain an efficient long run variance estimator under the $I(0)$ null and sustain the consistency of the KPSS tests against $I(d)$ alternatives, we cannot use the Bartlett kernel. In fact, the joint use of the Parzen kernel and $l=O(T^{1/5})$ is what we need. This explains why we extend the findings in Lo (1991) and Lee and Schmidt (1996) where they only considered the Bartlett kernel.

2. The Model and The Main Results

KPSS (1992) used their test as a test for trend stationarity. That is, they tested the hypothesis that deviations of a series X_t from a deterministic trend is short memory. Let e_t be the residuals from a regression of X_t on an intercept and a time (t), and let S_t be the partial sum of e_t: $S_t=\sum_{j=1}^{t} e_j$, $t=1,\ldots,T$. The KPSS test is expressed as

$$\hat{\eta}_\tau = T^{-2}\sum_{t=1}^{T} S_t^2/\hat{\Omega}_l.$$

Based on the construction of the KPSS test, we note that the term $\hat{\Omega}_l$ is the key ingredient of the KPSS statistic. Since the paper's objective is to consider the power of the KPSS tests against $I(d)$ alternatives when the data is generated as $X_t=\mu+\beta t+\epsilon_t$, and μ and β are arbitrary constants, we first investigate the asymptotic properties of $\hat{\Omega}_l$ when we demean or detrend a stationary $I(d)$ process. We then use the asymptotic properties of $\hat{\Omega}_l-\Omega_l$ and Ω_l to establish the consistency of the KPSS test against $I(d)$ alternatives. Moreover, we impose very few restrictions on the kernel function to extend the applicability of our results in empirical applications. Throughout this paper, we only require the kernels to satisfy the following conditions.

Assumption 1. (i) *For all* $x \in (0, 1]$, $k(x) \leq 1$; $k(x)$ *is continuous for almost all* $x \in (0, 1]$; $\int_R k(x)dx < \infty$. (ii) $|k(x) - k(y)| < C|x - y|$ *for some* $C > 0$. (iii) $k(x)$ *is a monotonic decreasing function, and there is only a finite number of points in* $(0, 1]$ *such that* $k(x) - k(y) = 0$, *for some* $x < y$.

Conditions (i) and (ii) of Assumption 1 are standard and have been used in Anderson (1971) and Andrews (1991). Condition (iii) of Assumption 1 is not as restrictive as it looks, because it covers some kernels, including the Bartlett and Parzen kernels.

Before presenting our theoretical results, let us first review some basic properties of the $I(d)$ process. A process ϵ_t is said to be an autoregressive fractionally integrated moving average process of order p, d, q, denoted as ARFIMA (p, d, q) or $I(d)$, if it is defined as

$$\phi(L)(1-L)^d \epsilon_t = \theta(L) a_t, \tag{2}$$

where L is the usual lag operator; $\phi(L)$ is a pth degree polynomial; d is the differencing parameter which can be a fractional number; $\theta(L)$ is a qth degree polynomial; the zeroes of $\phi(L)$ and $\theta(L)$ lie outside the unit circle; $\phi(L)$ and $\theta(L)$ have no common zeroes; and the innovation sequences a_t is a white noise with zero mean and variance σ_a^2. The fractional differencing operator $(1-L)^d$ has the following binomial expansion: $(1-L)^d = \sum_{j=0}^{\infty} \psi_j L^j$, where $\psi_j = \Gamma(j-d)/\Gamma(j+1)\Gamma(-d)$, and $\Gamma(\cdot)$ is the gamma function. The fractional white noise process is defined as

$$(1-L)^d \epsilon_t = a_t,$$

which is the simplest case of the ARFIMA model. This process was first introduced by Granger (1980, 1981), Granger and Joyeux (1980), and Hosking (1981). They showed that ϵ_t is stationary when $d < 0.5$ and is invertible when $d > -0.5$. Please refer to Baillie (1996) for more details.

Given the preceding conditions on the stationary ARFIMA (p, d, q) process, Hosking (1996, Theorem 8) showed that the exact order of magnitude of $Var(\sum_{t=1}^{T} \epsilon_t)$ is equal to $O(T^{1+2d})$. This asymptotic result is crucial to the derivation of our theoretical results and was established by Hosking (1996) with no more than a second moment condition on a_t.

Lemma 1 presents the asymptotic properties of $\hat{\Omega}_l - \Omega_l$ and Ω_l, which are the cornerstone of our analysis.

Lemma 1. *Given that* $E(a_t^4) < \infty$, *and* $k(.)$ *satisfies the conditions in Assumption 1, then as* $l \to \infty$, $T \to \infty$, $l/T \to 0$, *we have the following results:*

1. $\hat{\Omega}_l - \Omega_l = O_p(lT^{2d-1})$, *when* $d \in (0.25, 0.5)$.
2. $\hat{\Omega}_l - \Omega_l = O_p(lT^{-0.5}(\log T)^{0.5})$, *when* $d \in (-0.5, 0.25]$.
3. The exact order of magnitude of Ω_l *is equal to* $O(l^{2d})$.

The proofs of Lemma 1 and the following Theorem 1 are in the Appendix. Lemma 1 indicates that we cannot replace $\hat{\Omega}_l$ with Ω_l unless we set some restrictions on the growth rate of l. For example, $l = o(T^{1-2d})$ must be set when $d \in (0.25, 0.5)$; otherwise, the difference between $\hat{\Omega}_l$ and Ω_l will not converge in probability to zero. Furthermore, if the exact order of magnitude of l is $O(T^{1/5})$, which was recommended by Andrews (1991) when the Parzen kernel is used, then we cannot simply replace $\hat{\Omega}_l$ with Ω_l when $d > 0.4$. In addition to the above restriction, we note that the order of magnitude of Ω_l must exceed that of $\hat{\Omega}_l - \Omega_l$, or we cannot replace $\hat{\Omega}_l$ with Ω_l either.

The results in Lemma 1 hold for any kernel function which satisfies the conditions in Assumption 1. Therefore, Lemma 1 greatly extends the finding in Lee and Schmidt (1996) where they only considered the Bartlett window. Moreover, following equation (6) of Lee and Schmidt (1996, p.289), we note that $\Omega_l / l^{2d} = O(1)$ and is bounded away from zero.

Given the preceding results, we realize that the usual "consistency" properties of the long run variance estimator for the weakly dependent process cannot be trivially extended to that of the stationary $I(d)$ process. Moreover, it is quite odd mentioning the word "consistency" for the semiparametric long run variance estimator of the $I(d)$ process. The reason is that the exact order of magnitude of Ω_l is $O(l^{2d})$. This implies that Ω_l either diverges without bound or converges to zero when the DGP is the $I(d)$ process with $d \in (-0.5, 0) \cup (0, 0.5)$. However, with the help of Lemma 1 and the restriction that $l = o(T^{1/4})$, the consistency of the KPSS tests against $I(d)$ alternatives can still be established.

Theorem 1. *Given that* $E(a_t^4) < \infty$, $k(.)$ *satisfies the conditions in Assumption 1, and* $l = o(T^{1/4})$, *then as* $T \to \infty$, *the KPSS test of the* $I(0)$ *null is consistent against* $I(d)$ *alternatives with* $d \in (-0.5, 0) \cup (0, 0.5)$, *no matter whether the null hypothesis is level stationarity or trend stationarity.*

Theorem 1 reveals the conditions under which we can establish the consistency of the KPSS test of the $I(0)$ null against $I(d)$ alternatives. In fact, the condition that $l = o(T^{1/4})$ is not as restrictive as it looks, because $l = O(T^{1/5})$ is popularly recommended in the calculation of $\hat{\Omega}_l$ when the DGP is a weakly dependent process. For example, the Parzen kernel can be employed when the exact order of magnitude of l is $O(T^{1/5})$. In particular, when the DGP is the $I(0)$ process and the exact growth rate of l is $O(T^{1/5})$, $\hat{\Omega}_l$ attains its efficiency level when the Parzen, or Quadratic Spectral, or Tukey-Hanning kernels are used. However, Quadratic Spectral and Tukey-Hanning kernels do not satisfy the conditions in Assumption 1.

On the other hand, when $l=o(T^{1/4})$ is imposed, the Bartlett kernel should not be used according to the results in Andrews (1991). His paper showed that the Bartlett window cannot result in an efficient long run variance estimator for the weakly dependent process when the exact order of magnitude of l is not $O(T^{1/3})$. Therefore, in order to obtain an efficient long run variance estimator under the $I(0)$ null and sustain the consistency of the KPSS tests against $I(d)$ alternatives, we should not employ the Bartlett kernel. This also explains why we extend the findings in Lo (1991) and Lee and Schmidt (1996) where they only considered the Bartlett kernel.

3. Conclusion

This paper investigates the asymptotic properties of the semiparametric long run variance estimator when we demean or detrend a stationary $I(d)$ process. The analytic results are used to show that the KPSS test of the $I(0)$ null is consistent against $I(d)$ alternatives when the growth rate of the bandwidth parameter l is $o(T^{1/4})$. Moreover, the consistency of the modified Durbin-Watson (MDW) test of $I(1)$ null against fractional alternatives proposed by Tsay (1998) can also be proved. Therefore, our analysis provides theoretical support underpinning the practical use of the KPSS tests and other related test statistics against fractional alternatives.

Appendix

Two $I(d)$ processes are used in the following proofs. The first one is the stationary $I(d)$ process ϵ_t, defined in equation (2), with differencing parameters d such that $-0.5<d<0.5$. We then define one nonstationary $I(1+d)$ process by integrating ϵ_t:

$$y_t=y_{t-1}+\epsilon_t.$$

We also assume, without loss of generality, that the initial values of processes ϵ_0 and y_0 are both zero. Hence, y_t can be viewed as the partial sum of ϵ_t;

i.e., $y_t=\sum_{i=1}^{t}\epsilon_i$. Obviously, the order of integration of the nonstationary process y_t lies between 0.5 and 1.5.

A.1. Proof of Lemma 1:

We divide our analysis into two parts. The first part considers the cases where the DGP is $X_t=\mu+\epsilon_t$. The second part deals with the cases where the DGP is $X_t=\mu+\beta t+\epsilon_t$.

If $X_t=\mu+\epsilon_t$, then we obtain

$$e_t=\epsilon_t-\bar{\epsilon}=\epsilon_t-T^{-1}\sum_{t=1}^{T}\epsilon_t,$$

and $\bar{\epsilon}=O_p(T^{d-0.5})$. From equation (1), we have

$$\hat{\Omega}_l=T^{-1}\sum_{t=1}^{T}(\epsilon_t-\bar{\epsilon})^2+2T^{-1}\sum_{j=1}^{l}k\left(\frac{j}{l+1}\right)\sum_{t=j+1}^{T}(\epsilon_t-\bar{\epsilon})(\epsilon_{t-j}-\bar{\epsilon}).$$

Moreover, if we define

$$\tilde{\Omega}_l=T^{-1}\sum_{t=1}^{T}\epsilon_t^2+2T^{-1}\sum_{j=1}^{l}k\left(\frac{j}{l+1}\right)\sum_{t=j+1}^{T}\epsilon_t \qquad \epsilon_{t-j}=c_0+2\sum_{j=1}^{l}k\left(\frac{j}{l+1}\right)c_j,$$

then we can divide $\hat{\Omega}_l-\Omega_l$ into two parts:

$$\hat{\Omega}_l-\Omega_l=(\hat{\Omega}_l-\tilde{\Omega}_l)+(\tilde{\Omega}_l-\Omega_l).$$

To understand the asymptotic property of $\hat{\Omega}_l-\Omega_l$, we first discuss the part $\hat{\Omega}_l-\tilde{\Omega}_l$.

Given the preceding definition for $\hat{\Omega}_l$ and $\tilde{\Omega}_l$, we have

$$\begin{aligned}\hat{\Omega}_l-\tilde{\Omega}_l&=-2(\bar{\epsilon})\left\{T^{-1}\sum_{t=1}^{T}\epsilon_t+T^{-1}\sum_{j=1}^{l}k\left(\frac{j}{l+1}\right)\sum_{t=j+1}^{T}(\epsilon_t+\epsilon_{t-j})\right\}\\&\quad+(\bar{\epsilon})^2\left\{T^{-1}\sum_{t=1}^{T}1+2T^{-1}\sum_{j=1}^{l}k\left(\frac{j}{l+1}\right)\sum_{t=j+1}^{T}1\right\}\\&\le 2|\bar{\epsilon}|\left|T^{-1}\sum_{t=1}^{T}\epsilon_t+T^{-1}\sum_{j=1}^{l}k\left(\frac{j}{l+1}\right)\sum_{t=j+1}^{T}(\epsilon_t+\epsilon_{t-j})\right|\\&\quad+|\bar{\epsilon}|^2\left|T^{-1}\sum_{t=1}^{T}1+2T^{-1}\sum_{j=1}^{l}k\left(\frac{j}{l+1}\right)\sum_{t=j+1}^{T}1\right|\\&\le 2|\bar{\epsilon}|\left\{\left|T^{-1}\sum_{t=1}^{T}\epsilon_t\right|+\left|T^{-1}\sum_{j=1}^{l}k\left(\frac{j}{l+1}\right)\sum_{t=j+1}^{T}\epsilon_t\right|+\left|T^{-1}\sum_{j=1}^{l}k\left(\frac{j}{l+1}\right)\sum_{t=j+1}^{T}\epsilon_{t-j}\right|\right\}\\&\quad+|\bar{\epsilon}|^2\left|1+2T^{-1}\sum_{j=1}^{l}\sum_{t=j+1}^{T}1\right|\\&\equiv 2|\bar{\epsilon}|\left(\left|T^{-1}\sum_{t=1}^{T}\epsilon_t\right|+|M|+|M'|\right)+|\bar{\epsilon}|^2\,|1+2N|.\end{aligned}$$

However, we note that

$$N=T^{-1}\sum_{j=1}^{l}\sum_{t=j+1}^{T}1=T^{-1}\left\{\sum_{j=1}^{l-1}j+l+l(T-l)\right\}=T^{-1}\left(lT-\frac{l^2+l}{2}\right)=O(l).$$

Moreover, we observe that

$$\begin{aligned}M&=T^{-1}\sum_{j=1}^{l}k\left(\frac{j}{l+1}\right)\sum_{t=j+1}^{T}\epsilon_t=T^{-1}\sum_{j=1}^{l}k\left(\frac{j}{l+1}\right)(y_T-y_j)\\&=T^{-1}\sum_{j=1}^{l}k\left(\frac{j}{l+1}\right)y_T-T^{-1}\sum_{j=1}^{l}k\left(\frac{j}{l+1}\right)y_j\\&=M_1+M_2.\end{aligned}$$

For M_1, we note that

$$\begin{aligned}l^{-1}\sum_{j=1}^{l}k\left(\frac{j}{l+1}\right)y_T&=y_Tl^{-1}\sum_{j=1}^{l}k\left(\frac{j}{l+1}\right)\\&\le y_Tl^{-1}\sum_{j=1}^{l}\left|k\left(\frac{j}{l+1}\right)\right|\\&=y_T\int_0^{\infty}|k(x)|dx=O_p(T^{0.5+d}),\end{aligned}$$

and we prove that $M_1=O_p(lT^{d-0.5})$. To calculate the asymptotic behavior of M_2, we apply Abel's transformation:

$$\sum_{i=1}^{n}u_iv_i=\sum_{i=1}^{n-1}U_i(v_i-v_{i+1})+U_nv_n,$$

where $U_k=u_1+u_2+\ldots+u_k$ for $k=1, 2, \ldots, n$. Furthermore, we note that

$$M_2=T^{-1}\sum_{j=1}^{l}k\left(\frac{j}{l+1}\right)y_j.$$

If we define

$$U_j=y_1+y_2+\ldots+y_j \quad \text{and} \quad v_j=k\left(\frac{j}{l+1}\right),$$

then we have

$$\sum_{j=1}^{l}k\left(\frac{j}{l+1}\right)y_j=\sum_{j=1}^{l-1}U_j\left\{k\left(\frac{j}{l+1}\right)-k\left(\frac{j+1}{l+1}\right)\right\}+U_lk\left(\frac{l}{l+1}\right)=E_1+E_2.$$

We observe that

$$E_2=U_lk\left(\frac{l}{l+1}\right),$$

and

$$k\left(\frac{l}{l+1}\right)-0=k\left(\frac{l}{l+1}\right)-k\left(\frac{l+1}{l+1}\right)\leq\frac{C}{l+1}=O(l^{-1}).$$

Furthermore,

$$U_l=\sum_{j=1}^{l}y_j=O_p(l^{1.5+d}).$$

Combining the preceding results, we prove that $E_2=O_p(l^{1.5+d})$.

We also note that

$$E_1=\sum_{j=1}^{l-1}U_j\left\{k\left(\frac{j}{l+1}\right)-k\left(\frac{j+1}{l+1}\right)\right\}\leq\frac{C}{l+1}\sum_{j=1}^{l-1}U_j,$$

where C is a positive constant. Without loss of generality, we assume $C>1$ throughout this paper. We also observe that

$$\begin{aligned}\sum_{j=1}^{l-1}U_j&=\sum_{j=1}^{l-1}\sum_{k=1}^{j}y_k=\sum_{j=1}^{l-1}(1-j)y_j=\sum_{j=1}^{l-1}ly_j-\sum_{j=1}^{l-1}jy_j\\&=O_p(l^{2.5+d}).\end{aligned}$$

Thus, we prove that $E_1=O_p(l^{1.5+d})$ and the order of magnitude of E_1 dominates that of E_2. This implies that

$$M_2=O_p(T^{-1}l^{1.5+d}),$$

and we prove that $M=O_p(lT^{d-0.5})$.

We can similarly show that

$$M'=O_p(lT^{d-0.5}).$$

Given the preceding results, we have

$$\begin{aligned}\hat{\Omega}_l-\tilde{\Omega}_l&=O_p(T^{d-0.5})O_p(lT^{d-0.5})+O_p(T^{2d-1})O(l)\\&=O_p(lT^{2d-1}).\end{aligned}$$

Let us discuss the remaining part $\tilde{\Omega}_l-\Omega_l$. We note that $\tilde{\Omega}_l-\Omega_l=\tilde{\Omega}_l-E(\tilde{\Omega}_l)+E(\tilde{\Omega}_l)-\Omega_l$. First, we have

$$\begin{aligned}E(\tilde{\Omega}_l)-\Omega_l&=E\left\{c_0+2\sum_{j=1}^{l}k\left(\frac{j}{l+1}\right)c_j\right\}-\left\{\gamma_0+2\sum_{j=1}^{l}k\left(\frac{j}{l+1}\right)\gamma_j\right\}\\&=\left\{\gamma_0+2\sum_{j=1}^{l}k\left(\frac{j}{l+1}\right)\frac{T-j}{T}\gamma_j\right\}-\left\{\gamma_0+2\sum_{j=1}^{l}k\left(\frac{j}{l+1}\right)\gamma_j\right\}\end{aligned}$$

$$
\begin{aligned}
&= -2T^{-1}\sum_{j=1}^{l} k\left(\frac{j}{l+1}\right) j\gamma_j \\
&\leq 2T^{-1}\sum_{j=1}^{l} |j\gamma_j| \\
&= \begin{cases} T^{-1}O(l^{2d+1}) & \text{if} \quad d>0, \\ T^{-1}O(l(\log l)) & \text{if} \quad d=0, \\ T^{-1}O(l^{2d+1}) & \text{if} \quad d<0, \end{cases}
\end{aligned}
$$

since γ_j decays geometrically as $d=0$. At least we can say that $\gamma_j = O(j^{-1})$ as $j \to \infty$ and $d=0$.

Second, we note that

$$
\begin{aligned}
\tilde{\Omega}_l - E(\tilde{\Omega}_l) &= \left\{c_0 + 2\sum_{j=1}^{l} k\left(\frac{j}{l+1}\right) c_j\right\} - E\left\{c_0 + 2\sum_{j=1}^{l} k\left(\frac{j}{l+1}\right) c_j\right\} \\
&= \{c_0 - E(c_0)\} + 2\sum_{j=1}^{l} k\left(\frac{j}{l+1}\right)\{c_j - E(c_j)\}.
\end{aligned}
$$

However, under the restriction that $E(a_t^4) < \infty$, the results of Hosking (Theorem 3, 1996) and Chung (Corollary 1, 1996) showed that

$$
Var(c_j) = \begin{cases} O(T^{4d-2}) & \text{if} \quad d>0.25, \\ O(T^{-1}(\log T)) & \text{if} \quad d=0.25. \end{cases}
$$

Moreover, the preceding results do not depend on j. Thus, we have

$$
c_j - E(c_j) = \begin{cases} O_p(T^{2d-1}) & \text{if} \quad d>0.25, \\ O_p(T^{-0.5}(\log T)^{0.5}) & \text{if} \quad d=0.25. \end{cases}
$$

For cases $d<0.25$, $Var(c_j)$ can be uniformly bounded by $O(T^{-1}(\log T))$, and we have

$$
c_j - E(c_j) = \begin{cases} O_p(T^{2d-1}) & \text{if} \quad d>0.25, \\ O_p(T^{-0.5}(\log T)^{0.5}) & \text{if} \quad d\leq 0.25. \end{cases}
$$

Given the preceding results, if $d>0.25$, then

$$
\begin{aligned}
l^{-1}T^{1-2d}\sum_{j=1}^{l} k\left(\frac{j}{l+1}\right)\{c_j - E(c_j)\} &\leq l^{-1}T^{1-2d}\sum_{j=1}^{l}\left|k\left(\frac{j}{l+1}\right)\right| |c_j - E(c_j)| \\
&= l^{-1}\sum_{j=1}^{l}\left|k\left(\frac{j}{l+1}\right)\right| T^{1-2d}|c_j - E(c_j)| \\
&= O_p(1).
\end{aligned}
$$

For $d\leq 0.25$, the reasoning is similar and is omitted. Overall,

$$\sum_{j=1}^{l} k\left(\frac{j}{l+1}\right)\{c_j - E(c_j)\} = \begin{cases} O_p(lT^{2d-1}) & \text{if } d > 0.25, \\ O_p(lT^{-0.5}(\log T)^{0.5}) & \text{if } d \leq 0.25. \end{cases}$$

Thus, we have

$$\begin{aligned} \hat{\Omega}_l - \Omega_l &\leq |\hat{\Omega}_l - \tilde{\Omega}_l| + |\tilde{\Omega}_l - \Omega_l| \\ &\leq |\hat{\Omega}_l - \tilde{\Omega}_l| + |\tilde{\Omega}_l - E(\tilde{\Omega}_l)| + |E(\tilde{\Omega}_l) - \Omega_l| \\ &= \begin{cases} O_p(lT^{2d-1}) & \text{if } d > 0.25, \\ O_p(lT^{-0.5}(\log T)^{0.5}) & \text{if } d \leq 0.25. \end{cases} \end{aligned}$$

and items 1 and 2 of Lemma 1 are proved.

To prove item 3 of Lemma 1, we define

$$U_j = \gamma_2 + \gamma_2 + \ldots + \gamma_j \quad \text{and} \quad v_j = k\left(\frac{j}{l+1}\right).$$

We then have

$$\sum_{j=1}^{l} k\left(\frac{j}{l+1}\right)\gamma_j = \sum_{j=1}^{l-1} U_j\left\{k\left(\frac{j}{l+1}\right) - k\left(\frac{j+1}{l+1}\right)\right\} + \left(\sum_{j=1}^{l}\gamma_j\right)k\left(\frac{l}{l+1}\right) = A' + B'.$$

Therefore, we have

$$\Omega_l = \gamma_0 + 2A' + 2B'.$$

For the term A', we note that

$$A' = \sum_{j=1}^{l-1} U_j\left\{k\left(\frac{j}{l+1}\right) - k\left(\frac{j+1}{l+1}\right)\right\} \leq \frac{C}{l+1}\sum_{j=1}^{l-1} U_j,$$

and

$$\sum_{j=1}^{l-1} U_j = \sum_{j=1}^{l-1}\sum_{k=1}^{j} \gamma_k = \sum_{j=1}^{l-1}(l-j)\gamma_j.$$

For the term B', we note that

$$\sum_{j=1}^{l} \gamma_j = \begin{cases} O(l^{2d}) & \text{if } d > 0, \\ O(1) & \text{if } d = 0, \\ O(1) & \text{if } d < 0, \end{cases}$$

and we prove that

$$B'=\begin{cases} O(l^{2d-1}) & \text{if} \quad d>0, \\ O(l^{-1}) & \text{if} \quad d=0, \\ O(l^{-1}) & \text{if} \quad d<0. \end{cases}$$

Thus, we have

$$\begin{aligned}\Omega_l=\gamma_0+2\sum_{j=1}^{l}k\left(\frac{j}{l+1}\right)\gamma_j+2B' &\leq \gamma_0+2\frac{C}{l+1}\sum_{j=1}^{l-1}U_j+2B' \\ &\leq C\gamma_0+2\frac{C}{l+1}\sum_{j=1}^{l-1}(1-j)\gamma_j+2B' \\ &=(l+1)\frac{C}{l+1}\gamma_0+2\frac{C}{l+1}\sum_{j=1}^{l-1}(1-j)\gamma_j+2B' \\ &=\frac{C}{l+1}\gamma_0+\frac{Cl}{l+1}\gamma_0+2\frac{C}{l+1}\sum_{j=1}^{l-1}(1-j)\gamma_j+2B' \\ &=\frac{C}{l+1}\gamma_0+\frac{C}{l+1}\left\{l\gamma_0+2\sum_{j=1}^{l-1}(1-j)\gamma_j\right\}+2B' \\ &=\frac{C}{l+1}\gamma_0+\frac{C}{l+1}Var\left(\sum_{t=1}^{l}\epsilon_t\right)+2B' \\ &=O(l^{-1})+O(l^{-1})O(l^{1+2d})+2B'=O(l^{2d}).\end{aligned}$$

Item 3 of Lemma 1 is proved. Note that $O(l^{2d})$ is the exact order of magnitude of Ω_l, because the exact order of magnitude of $Var(\sum_{t=1}^{l}\epsilon_t)$ is $O(l^{1+2d})$. The choice of constant C will not change the asymptotic properties of Ω_l.

For the second part, we note that $X_t=\mu+\beta t+\epsilon_t$, and we have

$$e_t=\epsilon_t-\bar{\epsilon}-(\hat{\beta}-\beta)(t-\bar{t}),$$

where $\hat{\beta}-\beta=O_p(T^{d-1.5})$. We only calculate $\hat{\Omega}_l-\tilde{\Omega}_l$, because the asymptotic properties of $\tilde{\Omega}_l-\Omega_l$ will not change when the DGP is $X_t=\mu+\beta+\epsilon_t$. Following the previous analysis, we obtain

$$\begin{aligned}\hat{\Omega}_l=&\ T^{-1}\sum_{t=1}^{T}\{\epsilon_t-\bar{\epsilon}-(\hat{\beta}-\beta)(t-\bar{t})\}^2 \\ &+2T^{-1}\sum_{j=1}^{l}k\left(\frac{j}{l+1}\right)\sum_{t=j+1}^{T}\{\epsilon_t-\bar{\epsilon}-(\hat{\beta}-\beta)(t-\bar{t})\}\{\epsilon_{t-j}-\bar{\epsilon}-(\hat{\beta}-\beta) \\ &(t-j-\bar{t})\},\end{aligned}$$

and

$$\hat{\Omega}_l-\tilde{\Omega}_l\leq 2|\bar{\epsilon}|\left[\left|T^{-1}\sum_{t=1}^{T}\epsilon_t\right|+\left|T^{-1}\sum_{j=1}^{l}k\left(\frac{j}{l+1}\right)\sum_{t=j+1}^{T}\epsilon_t\right|+\left|T^{-1}\sum_{j=1}^{l}k\left(\frac{j}{l+1}\right)\sum_{t=j+1}^{T}\epsilon_{t-j}\right|\right]$$

$$+|\bar{\epsilon}|^2\left|T^{-1}\sum_{t=1}^{T}1+2T^{-1}\sum_{t=1}^{l}k\left(\frac{j}{l+1}\right)\sum_{t=j+1}^{T}1\right|$$
$$+2|\hat{\beta}-\beta|\left[\left|T^{-1}\sum_{t=1}^{T}(t-\bar{t})\epsilon_t\right|+\left|T^{-1}\sum_{j=1}^{l}k\left(\frac{j}{l+1}\right)\sum_{t=j+1}^{T}(t-\bar{t})\epsilon_{t-j}\right|\right]$$
$$+2|\hat{\beta}-\beta|\left|T^{-1}\sum_{j=1}^{l}k\left(\frac{j}{l+1}\right)\sum_{t=j+1}^{T}(t-j-\bar{t})\epsilon_t\right|$$
$$+2|\hat{\beta}-\beta|\,|\bar{\epsilon}|\left[\left|T^{-1}\sum_{t=1}^{T}(t-\bar{t})\right|+\left|T^{-1}\sum_{j=1}^{l}k\left(\frac{j}{l+1}\right)\sum_{t=j+1}^{T}(t-\bar{t})\right|\right]$$
$$+2|\hat{\beta}-\beta|\,|\bar{\epsilon}|\left|T^{-1}\sum_{j=1}^{l}k\left(\frac{j}{l+1}\right)\sum_{t=j+1}^{T}(t-j-\bar{t})\right|$$
$$+|\hat{\beta}-\beta|^2\left[\left|T^{-1}\sum_{t=1}^{T}(t-\bar{t})^2\right|+2\left|T^{-1}\sum_{j=1}^{l}k\left(\frac{j}{l+1}\right)\sum_{t=j+1}^{T}(t-\bar{t})(t-j-\bar{t})\right|\right]$$

After some calculations, we prove that the order of magnitude of each term on the right-hand side of the inequality sign are bounded by $O_p(lT^{2d-1})$. The remaining proof is similar to the previous analysis. The details are omitted.

A.2. Proof of Theorem 1:

For cases $0\leq d\leq 0.25$, if we set $l=o(T^{0.5})$, then $O_p(lT^{-0.5}(\log T)^{0.5})=o_p(1)$. Moreover, the order of magnitude of Ω_l dominates that of $O_p(lT^{-0.5}(\log T)^{0.5})$. Thus, $\hat{\Omega}_l$ can be replaced by Ω_l. The proof of Lee and Schmidt (1996) is still correct and the details are omitted.

For cases $d>0.25$, Lemma 1 implies that

$$\hat{\Omega}_l-O(l^{2d})=O_p(lT^{2d-1}).$$

We also note that $\sum_{t=1}^{T}S_t^2=O_p(T^{2+2d})$. From Lemma 1, we note that $\hat{\Omega}_l/l^{2d}-\Omega_l/l^{2d}=O_p(l^{1-2d}T^{2d-1})=o_p(1)$. Moreover, Ω_l/l^{2d} is bounded away from zero. This implies that $\hat{\Omega}_l/l^{2d}$ is also bounded away from zero. Thus, we have $\hat{\Omega}_l/l^{2d}\xrightarrow{p}\Omega_l/l^{2d}=O(1)$. We then have

$$\hat{\eta}_\tau=T^{2d}\frac{T^{-2-2d}\sum_{t=1}^{T}S_t^2}{\hat{\Omega}_l}=\frac{T^{2d}}{l^{2d}}\frac{T^{-2-2d}\sum_{t=1}^{T}S_t^2}{l^{-2d}\hat{\Omega}_l}=\frac{T^{2d}}{l^{2d}}\frac{T^{-2-2d}\sum_{t=1}^{T}S_t^2}{l^{-2d}\Omega_l}=\left(\frac{T}{l}\right)^{2d}O_p(1).$$

This implies that $\hat{\eta}_\tau\xrightarrow{p}\infty$ as $d>0.25$.

For cases $-0.5<d<0$, we note that $\Omega_l=O(l^{2d})$. This implies that Ω_l also converges in probability to zero under this circumstance. Thus, we have to find the conditions under which the order of magnitude of Ω_l dominates that of $O_p(lT^{-0.5}(\log T)^{0.5})$. Otherwise, we cannot simply replace $\hat{\Omega}_l$ with Ω_l.

In other words, we need to specify the restriction on l, such that $l^{2d}/(lT^{-0.5}(\log T)^{0.5})=l^{2d-1}T^{0.5}(\log T)^{-0.5}\rightarrow\infty$ as l and T both approach infinity. It is not difficult to figure out that $l=o(T^{1/4})$ is the restriction that we need. Given $l=o(T^{1/4})$, we can then follow the preceding arguments in the proof of the case $0\leq d\leq 0.25$. That is, we can replace $\hat{\Omega}_l$ with Ω_l and follow the proof of Lee and Schmidt (1996). Combining the results of the preceding three cases, we prove the consistency of the KPSS test against $I(d)$ alternatives.

Acknowledgements

This research is supported by NSC-89-2415-H-001-036. We are very grateful for two referees for their valuable suggestions.

References

Anderson, T.W.

1971 *The Statistical Analysis of Time Series*, New York: Wiley.

Andrews, D.W.K.

1991 "Heteroskedasticity and autocorrelation consistent matrix estimation," *Econometrica*, 59, 817-858.

Baillie, R.T.

1996 "Long memory processes and fractional integration in econometrics," *Journal of Econometrics*, 73, 5-59.

Chung, C.-F.

1996 "Sample variances, sample covariances, and linear regression of stationary multivariate long memory processes," Working Paper, Michigan State University.

Diebold, F.X. and G.D. Rudebusch

1991 "On the power of Dickey-Fuller tests against fractional alternatives," *Economics Letters*, 35, 155-160.

Granger, C.W.J.

1980 "Long memory relationships and the aggregation of dynamic models," *Journal of Econometrics*, 14, 227-238.

1981 "Some properties of time series data and their use in econometric model specification," *Journal of Econometrics*, 16, 121-130.

Granger, C.W.J. and R. Joyeux

1980 "An introduction to long-memory time series models and fractionally differencing," *Journal of Time Series Analysis*, 1, 15-29.

Hansen, B.E.

1992 "Consistent covariance matrix estimation for dependent heterogeneous processes," *Econometrica*, 60, 967-972.

Hosking, J.R.M.
1981 "Fractional differencing," *Biometrika*, 68, 165-176.
1996 "Asymptotic distributions of the sample mean, autocovariances, and autocorrelations of long-memory time series," *Journal of Econometrics*, 73, 261-284.
Kwiatkowski, D., Phillips, P.C.B., Schmidt, P. and Y. Shin
1992 "Testing the null hypothesis 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, D. and P. Schmidt
1996 "On the power of the KPSS test of stationarity against fractionally-integrated alternatives," *Journal of Econometrics*, 73, 285-302.
Lo, A.W.
1991 "Long-term memory in stock market prices," *Econometrica*, 59, 1279-1313.
Newey, W.K. and K.D. West
1987 "A simple positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix," *Econometrica*, 55, 703-708.
Tsay, W.J.
1998 "On the power of Durbin-Watson statistic against fractionally integrated processes," *Econometric Reviews*, 17, 361-386.

再探KPSS檢定統計量在對立假設為部分差分時間數列之一致性

蔡文禎
中央研究院經濟研究所副研究員

摘　要

本文探討半參數化長期變異數估計法（semiparametric long run variance estimator）應用於平穩部分差分時間數列（stationary fractionally integrated process）的極限性質，以進一步得出 KPSS (Kwiatkowski, *et al.*, 1992) 檢定統計量在對立假設為平穩部分差分時間數列達成一致性所需之條件，其中之一指出 bandwidth parameter 隨樣本數增加的速度必須低於樣本數的四分之一次方。

關鍵詞：部分差分時間數列、KPSS 統計量