WAVELET TRANSFORM FOR REGRESSION ESTIMATION OF NON-STATIONARY LONG MEMORY TIME SERIES

By Jin Lee[1]

We consider log periodogram regression estimation of memory parameter d for non-stationary long memory time series. Non-stationary long memory processes obtain co-variance stationarity through wavelet transforms, then the spectral representation at zero frequency is well defined. Thus, we can make use of the statistical inferences developed in the stationary long memory context for non-stationary long memory time series.

We propose wavelet-based log periodogram regression estimator, which does not depend on data differencing or data tapering to obtain the consistency and asymptotic normality for non-stationary long memory time series. Asymptotic property of wavelet-based estimator is derived, in line with Robinson (1995), Hurbich, Deo and Brodsky (1998) and Andrews and Guggenberger (2003). The convergence rate of mean squared error is the same as in the stationary case. Simulation studies show that wavelet-based estimator works reasonably well without using data differencing or data tapering.

KEYWORDS: Non-stationary long memory time series, log periodogram regression, wavelet transform.

1. INTRODUCTION

WE CONSIDER LOG PERIODOGRAM (LP) regression estimation of memory parameter d for non-stationary long memory time series. While statistical inferences of LP estimation have been well developed in the stationary case (Robinson (1995), Hurvich, Deo and Brodsky (1998; HDB), for example), their direct extension to the non-stationary case raises some problems. In the LP context, it is known that LP estimator is not consistent for $d > 1$ (Hurbich and Ray (1995)). One can use data differencing or data tapering to achieve the consistency and asymptotic normality for non-stationary range of d. Such methods, as mentioned in Shimotsu and Phillips (2002), expose some disadvantages. Data differencing requires a prior knowledge of the order of integration for the data. Data tapering for LP estimator proposed by Hurbich and Ray (1995) and Velasco (1999) reduces the bias, but inflates the variance. Thus, one can obtain the consistency of LP estimator at the cost of efficiency loss. One way to overcome such problems is using local whittle estimator by Shimotsu and Phillips (2002).

In this paper, we propose a new way of estimating non-stationary memory parameter in the LP context. We consider wavelet transform and develop a regression estimator, which does not depend on data differencing or data tapering. Jensen (1997) also considers wavelet transform of long memory time series, but in the stationary case, and studies variance regression to estimate d, proposed by Masry (1993).

It is known that wavelet transform for non-stationary long memory time series obtains covariance stationarity (Flandrin (1992), Masry (1993), and Kato and Masry (1999), etc.). While non-stationary long memory processes do not admit valid spectral representation, wavelet transform of such process has well defined spectral representation. Since LP regression estimation sources on the spectral density at zero frequency proposed by Geweke and Porter-Hudak (1983; GPH), wavelet methods motivate one to make use of the statistical inferences developed in the stationary context for non-stationary long memory time series.

We derive the asymptotic bias and the variance, in line with Robinson (1995), HDB and Andrews and Guggenberger (2003; AG). Thus, we show the consistency of wavelet-based LP regression estimator. We can also obtain the convergence rate of the mean squared error, which is the same as in the stationary case.

2. THE MODEL

We consider discrete time series $\{x_t,\ t = 1, 2, \cdots, n\}$, observed from fractional Brownian motion (fBm) with the memory parameter d for $d \in (0.5, 1.5)$. It is non-stationary Gaussian process with stationary increment of mean-zero and variance $\sigma^2 < \infty$, with covariance function

$$(1) \quad E(X_t X_s) = \frac{\sigma^2}{2}(t^{2d-1} + s^{2d-1} - |t-s|^{2d-1}).$$

If $d = 1$, then $\{X_t\}$ returns to well known ordinary Brownian motion with variance equal to $\sigma^2 t$.

There has been a lot of literature which study wavelet transform for the fBm. (Flandrin (1992), Masry (1993), and Kato and Masry (1999) to name a few). One interesting finding in their works is that the wavelet transform for fBM is covariance stationary. We briefly summarize the known result. Define the wavelet transform for X_t

$$(2) \quad W_j(q) = 2^{j/2} \int_{-\infty}^{\infty} X_t \psi(2^j t - q) dt,$$

where j and q are integers. The function ψ is a wavelet, which is a well localized function. Below, we explicitly introduce the properties of the wavelet functions.

If we consider the covariance between the wavelet transforms

$$(3) \quad \begin{aligned} Cov(\tau) &= E(W_j(q) W_j(q+\tau)) \\ &= -\frac{\sigma^2}{2} 2^{-j} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |2^{-j}(u-v) - 2^{-j}\tau|^{2d-1} \psi(u)\psi(v) du dv, \end{aligned}$$

then the covariance only depends on the distance τ, since the first two terms which involve non-stationarity in (1) equal to zero due to the mean zero property of the wavelet function. Then, wavelet transform becomes covariance stationary process on the wavelet domain. Further, it is noted that the covariance of wavelet transform (3) has the same expression in the stationary case, since $E(X_t X_s)$ behaves as $|t-s|^{2d-1}$ for $d \in (-0.5, 0.5)$ (Jensen (1999)).

Now, we formally introduce the wavelet function.

$$\begin{aligned} \text{ASSUMPTION 1(a)} \quad &: \quad \psi : \mathbb{R} \to \mathbb{R} \textit{ forms an orthonormal basis for } L_2(\mathbb{R}) \textit{ such that} \\ \textstyle\int_{-\infty}^{\infty} \psi(x) dx &= 0\ ,\ \textstyle\int_{-\infty}^{\infty} \psi(x)\psi(x-k) dx = 0 \textit{ for all } k \in \mathbb{Z}, k \neq 0, \\ \textstyle\int_{-\infty}^{\infty} |\psi(x)| dx &< \infty, \textit{ and } \textstyle\int_{-\infty}^{\infty} \psi^2(x) dx = 1. \\ \text{(b)} \quad &: \quad \psi \textit{ has the } v \textit{ vanishing moments such that} \\ \textstyle\int_{-\infty}^{\infty} x^r \psi(x) dx &= 0 \textit{ for } r = 0, 1, 2, \cdots, v-1. \end{aligned}$$

The assumption 1(a) describes the wavelet function. See Hernandez and Weiss (1996) and Daubechies (1992) for details. By Assumption 1(a), the doubly infinite system $\{\psi_{j\tau}(\cdot)\}$ is an orthonormal basis for $L^2(\mathbb{R})$, where $\psi_{jq}(x) = 2^{j/2}\psi(2^j x - q)$, for $j, \tau \in \mathbb{Z}$.

The integer j is called a scale parameter, and the integer q a translation parameter. Intuitively, j localizes analysis in frequency and q localizes analysis in time. The simultaneous time-frequency localization of information is the key feature of wavelet analysis.

Next, assumption 2(b) is implied when the function ψ is in $C^v(\mathbb{R})$ where $C^v(\mathbb{R})$ is the set of all functions f defined on the real line $\mathbb{R}$ such that all the derivatives up to order v exist, and $f^{(v)}$ is continuous on $\mathbb{R}$. Fast decaying wavelet functions satisfy this condition. Examples include a family of compactly supported wavelets constructed by Daubechies (1992), which is popular in the wavelet literature. A class of orthonormal wavelets has the support of $[0, 2v-1]$, where v is positive integer, and enjoy the property that the first v moments equal to zero. For example, Haar wavelet is defined as $\psi(x) = 1$ for $x \in (0, 0.5]$, and -1 for $x \in (0.5, 1]$, and has $v = 1$. We consider Daubechies' family of wavelets for the analysis in our paper. Another example is a class of spline wavelets. For example, the first order spline wavelet, often called Franklin wavelet, has $v = 2$, and spline wavelet of order 2 has $v = 3$. In general, the spline wavelets of order n has the property that $\int_{-\infty}^{\infty} x^r \psi(x) dx = 0 \; for \; r = 0, 1, 2, \cdots n.$

We now consider the spectral representation of long memory time series for LP estimation. The spectral density in the stationary case of $d \in (-0.5, 0.5)$ is often modelled as

$$(4) \quad f(\lambda) = |\lambda|^{-2d} f^*(\lambda)$$

where $f^*(\lambda)$ is assumed to be even and positive function which is continuous on $[-\pi, \pi]$ with $0 < f^*(0) < \infty$. The asymptotic bias of LP estimator basically depends on how to assume the behavior of f^* around zero frequency. In HDB, it is assumed that the first derivative of $f^*(0)$ is zero, and the second and third derivatives of f^* are bounded in a neighborhood of zero frequency. In AG, it is assumed that f^* is smooth of order s at zero frequency for some $s \geq 1$.

The spectral density of the wavelet transform around zero frequency is introduced as (for example, see Kato and Masry(1999))

$$(5) \quad f_j(\lambda) = C_j |\lambda|^{-2d} |\hat{\psi}(\lambda)|^2 \text{ as } \lambda \to 0$$

where $C_j = c_j/2\pi < \infty$ is a constant term, which depends on integer-valued scale j, and $\hat{\psi}(\cdot)$ is the Fourier transform of wavelet ψ, that is to say, $\hat{\psi}(\lambda) = (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{\infty} \psi(x) e^{-i\lambda x} dx$.

One difference between two spectral representations of (4) and (5) is that the function $|\hat{\psi}(\lambda)|^2$ in (5) do not have the same property as $f^*(\lambda)$ in (4). The function $|\hat{\psi}(\lambda)|^2$ is known to be even, positive and continuous on $[-\pi, \pi]$, but it is not bounded away from zero at zero frequency, say, $|\hat{\psi}(0)| = 0$. The mean zero property, $\int_{-\infty}^{\infty} \psi(x) dx = 0$ is fundamental in wavelet analysis, which implies $\hat{\psi}(0) = 0$. It is necessary that $\hat{\psi}(\lambda)$ is bounded away from zero around $\lambda = 0$ to obtain the explicit form of asymptotic bias, we express the spectral density of the wavelet transform in an alternative form.

In doing so, we use the property of vanishing moment of the wavelet function. The vanishing moment condition in assumption 1(b) is equivalently expressed in the frequency domain that the first v spectral derivatives around zero frequency are zero, say, $\frac{d^r}{d\lambda^r} \hat{\psi}(\lambda)\Big|_{\lambda=0} = 0$ for $r = 0, 1, 2, \cdots, v-1$. Here, we notice the relation, $\frac{d^r}{d\lambda^r} \hat{\psi}(\lambda)\Big|_{\lambda=0} = (-i)^r \int_{\mathbb{R}} x^r \psi(x) dx$. The term, $|\hat{\psi}(\lambda)|^2$ has the first $2v$ zero spectral derivatives at the neighborhood of $\lambda = 0$. Also, odd numbered derivatives are equal to zero, since $|\hat{\psi}(\lambda)|^2$ is even function.

We have the following spectral density around zero frequency,

$$
\begin{aligned}
f_j(\lambda) &= C_j^* |\lambda|^{-2d} \{ g(\lambda) |\lambda|^{2v} + O(|\lambda|^{2v+2}) \} \text{ as } \lambda \to 0 \\
&= C_j^* |\lambda|^{-2(d-v)} g(\lambda) + O(|\lambda|^{-2(d-v)+2}) \text{ as } \lambda \to 0
\end{aligned}
\tag{6}
$$

where $g(\lambda) = \frac{d^{2v}}{d\lambda^{2v}} |\hat{\psi}(\lambda)|^2$, and $C_j^* = \frac{2^{-2jv}}{(2v)!} C_j$.

The expression of transformed spectral density satisfies the Assumption 1 in Robinson (1995). The $O(\cdot)$ term in (6) becomes negligible by imposing suitable conditions on the rate of growth m in $\lambda_k = 2\pi k/n$, where $k = 1, 2, \cdots, m$. The condition on m is stated as follows, which is standard in the long memory literature.

ASSUMPTION 2 : $m = m(n) \to \infty$, *and* $m/n \to 0$ *as* $n \to \infty$.

The positive integer m is restricted to increase at slower rate than n. This is identical to assumption 1 in AG.

3. WAVELET-BASED ESTIMATOR

We show the consistency of wavelet-based LP estimator for $d \in (0.5, 1.5)$. First, we discretize the wavelet transform in (2) by re-indexing the time t as $t = (2v-1)i/n$, where

$i = 1, 2, \cdots, n$, given the known value of v. Then, the support of the wavelet ψ, $[0, 2v-1]$ is entirely covered. Define the discrete wavelet transform for $\{X_t\}$

$$(7) \quad w_j(q) = 2^{j/2} \sum_t x_t \psi(2^j t - q),$$

where $t = (2v-1)(\frac{i}{n})$, with $i = 1, 2, \cdots, n$.

It is natural to define a periodogram for wavelet transform, since it is covariance stationary process. Define a periodogram at fixed scale j,

$$(8) \quad I_k^{(j)} \equiv I_k = \frac{1}{2\pi n} \sum_{q=0}^{(2^j-1)(2v-1)} |w_j(q) \exp(i\lambda_k q)|^2,$$

where $\lambda_k = 2\pi k/n$, and $k = 1, 2, \cdots, m$. The scale j is suppressed as it is fixed during the analysis, which helps simplify the notation. The maximum number of q equals to $(2^j - 1)(2v - 1)$, which depends on the effective support of the wavelet. If Haar wavelet is used, then q runs from 0 to $2^j - 1$.

Write the LP regression as

$$(9) \quad \log I_k = \alpha + \beta X_k + \log(g(\lambda_k)/g(0)) + \varepsilon_k,$$

where $\alpha = (\log(g(0)) + \log C_j^* - \eta)$, $\beta = (d - v)$, $X_k = -2\log(\lambda_k)$, and $\varepsilon_k = \log(I_k/f_k) + \eta$, for $k = 1, 2, \cdots m$.

The term, $\log(g(\lambda_k)/g(0))$ is dominating term for the asymptotic bias. To get the explicit form of the asymptotic bias, we have Taylor expansion for $\log(g(\lambda_k)/g(0))$ at $\lambda = 0$,

$$\log \frac{g(\lambda_k)}{g(0)} = \frac{1}{2} \frac{g''(0)}{g(0)} \lambda_k^2 + O(\lambda_k^4)$$

where we have used the fact that $|\hat{\psi}(\cdot)|^2$ is even function, then odd numbered spectral derivatives of $g(\cdot)$ equal to zero at $\lambda = 0$.

We show the consistency of LP estimator for d, by examining the asymptotic bias, and variance.

THEOREM 1 : *Suppose Assumptions 1 and 2 hold. Then,*

$$(a)\ E\widehat{d} - d = -\frac{8\pi^2}{9} \frac{g''(0)}{g(0)} \frac{m^2}{n^2} (1 + o(1)) + O(m^4/n^4) + O(\frac{\log^3 m}{m}).$$

$$(b)\ Var(\widehat{d}) = \frac{\pi^2}{24m} + o(\frac{1}{m}).$$

Theorem 1 shows that wavelet-based estimator $\widehat{d}$ is consistent for $d \in (0.5, 1.5)$ in the L_2 sense. The variance takes the same form as in the stationary case. Unlike the case of data tapering, we do not have efficiency loss to achieve the consistency. The proof is basically adapted from AG and HDB, as well as Robinson(1995).

Further, we obtain the form of MSE

$$(10)\quad MSE(\widehat{d}) = \left(\frac{8\pi^2}{9}\frac{g''(0)}{g(0)}\right)^2 \frac{m^4}{n^4}(1+o(1)) + O\left(\frac{m^3 \log^3 m}{n^4}\right) + \frac{\pi^2}{24m}(1+o(1)).$$

Given the form of MSE, we obtain the optimal m^*

$$(11)\quad m^* = 0.2661 \cdot \left(\frac{g(0)}{g''(0)}\right)^{2/5} n^{4/5}.$$

In the case of stationary GPH estimator, the optimal m^* is not available, since it depends on unknown spectral density. In our case, both $g(0)$ and $g''(0)$ depend on the known wavelet function, thus the optimal m^* is available if the functional form of the wavelet function is given. For example, Haar wavelet has $|\hat{\psi}(\lambda)| = (\lambda/4) \cdot (\frac{\sin(\lambda/4)}{(\lambda/4)})^2$, which generates the optimal number of frequencies,

$$(12)\quad m^* = 0.207 \cdot n^{4/5}.$$

It follows from (11) and (12) that we have $MSE(\widehat{d}) = O(n^{-4/5})$. This is the same convergence rate as that of GPH estimator in the stationary case, developed by HDB. Faster rate of convergence of MSE can be only obtained by including additional regressors of frequencies in the LP regression as in AG. In such case, $MSE = O(n^{-(4+4r)/(5+4r)})$, given that $s \geq 2 + 2r$, with s is the order of smoothness of f^* around zero frequency, and r the number of additional regressors.

Given the optimal rate of m above, we construct the asymptotic normality of wavelet-based estimator as similarly done in HDB and AG in stationary case.

COROLLARY 1 : *Suppose Assumption 1 hold, and* $m = o(n^{4/5})$, *then*

$$m^{1/2}(\widehat{d} - d) \rightarrow_d N(0, \frac{\pi^2}{24}) \text{ as } n \rightarrow \infty.$$

The proof, briefly stated in the Appendix, follows from Robinson (1995), HDB, and AG.

4. SIMULATION STUDIES

We investigate the finite sample performance of the wavelet-based LP regression estimator. We consider ARFIMA$(1, d, 0)$ process with autoregressive parameter $\phi = 0, 0.5$, and 0.8, as well as $d =$ 0.6, 0.8, 1.0, 1.2 and 1.4. The $I(d)$ process $\{X_t\}_{t=1}^{n}$ is generated through $X_t = \sum_{k=0}^{t-1} \frac{(d)_k}{k!} u_{t-k}$, and $(d)_k = d(d+1)\cdots(d+k-1)$, where $u_t \sim i.i.d.N(0,1)$. (Shimotsu and Phillips (2002)). One thousand iterations for sample size of $n = 512$ are conducted.

For wavelet-based estimator, we use Haar wavelet, which has $v = 1$ in (9). The optimal number of the frequencies m is used according to (12). The integer-valued j for the transformed periodogram is chosen as large as possible so that the periodogram is well behaved. We use $j = 8$ for $n = 512$ in our simulation, which generates the transformed series, $\{w_j(q),\ q = 0, 1, \cdots, 2^8\}$. Neighboring values of j give similar results. We label wavelet-based estimator as W.

For comparison, we include two LP regression estimators in our simulations. One is tapered GPH estimator using cosine taper (Velasco (1999) and Hurvich and Ray (1995)). Data tapering using cosine bell is applied to raw data, where periodogram ordinates at λ_k for $k = l+1, l+2, \cdots, m$ with $l = 1$ are used in the regression as in Hurvich and Ray (1995). We label this estimator as DT. The other estimator is the GPH estimator using data differencing. We estimate $d^* = d - 1$ for the first differenced data, then obtain the estimate of d. There is another approach to combine data differencing and data tapering proposed by Hurvich and Chen (2000). Since both tapered and untapered estimator using data differencing show pretty similar results for $d \in (0.5, 1.5)$ in their simulations, we only include untapered GPH estimator through data differencing. We label this as DF. For DT and DF, we set $m = [0.25n^{4/5}]$, which is, as an optimal rate, used in the simulations in Hurvich and Chen (2000).

Table I reports the bias and mean squared error (MSE) of LP regression estimators for different values of d with $\phi = 0, 0.5$ and 0.8, when $n = 512$. We assess the performances according to the values of ϕ. For $\phi = 0$, W performs between DT and DF. As for the bias, W is more positively biased than DF, and less positively biased than DT in most cases. The MSE of W is rather larger than that of DF, which is the smallest for all values of d, and smaller than that of DT. Thus, when autoregression parameter is zero, we see that DF performs the best, followed by W, and by DT.

Next, we increase the value of ϕ to 0.5. All the estimators become more positively biased than in the case of $\phi = 0$. Now, W obtains the smallest bias for all values of d.

The bias of W is almost half as much as that of DF, and again DT has the largest bias. In terms of MSE, DF still has the smallest MSE due to smallest variance, followed by W, then by DT. The MSE of W is slightly larger than that of DF, but the difference becomes smaller than in the case of $\phi = 0$.

Lastly, we increase the value of ϕ to 0.8. It is again pronounced that the bias of W is less affected by the magnitude of autoregression than that of DF and of DT. For all values of d, W obtains the smallest bias. The bias of W is between half and nearly a third of that of DF. Due to this significant bias reduction, W now achieves the smallest MSE. Thus, when autoregression parameter is strong and positive, W performs the best.

In sum, wavelet-based estimator works reasonably well without data tapering or data differencing. In particular, it is attractive when autoregression parameter is strongly positive.

Department of Economics, National University of Singapore, 10 Kent Ridge Crescent, Singapore 119260; ecsleej@nus.edu.sg

APPENDIX

PROOF OF THEOREM 1: Let I, X, R, and ε denote $m \times 1$ column vectors whose k-th elements are $\log I_k$, X_k, $\log(g(\lambda_k)/g(0))$, and ε_k, respectively. As in AG, we write the regression equation in matrix form as $\log I = (\log g(0) + \log C^* - \eta)1_m + X\beta + R + \varepsilon$. By letting $Z = X - 1_m\overline{X}$ with $\overline{X} = (X'1_m)/m$, we write

$$(13) \quad \log I = ((\log g(0) + \log C^* - \eta + \overline{X}d)1_m + Z\beta + R + \varepsilon$$

The bias term can be written as $E\widehat{d} - d = (Z'Z)^{-1}Z'(R + \varepsilon)$.

The proof consists of the three parts: (a) $Z'Z$, (b) $Z'R$, and (c) $Z'E(\varepsilon)$. First, note that $X_k = -2\log\lambda_k$, then from HDB (page22), we have $Z'Z = 4m(1 + o(1))$. Next, we write

$$(14) \quad Z'R = \frac{1}{2}\frac{g''(0)}{g(0)}Z'\lambda_k^2 + \sum_{k=1}^{m}(X_k - \overline{X})O(\lambda_k^4).$$

The first term in (14) is written as

$$\begin{aligned} \frac{1}{2}\frac{g''(0)}{g(0)}Z'\lambda_k^2 &= \frac{1}{2}\frac{g''(0)}{g(0)}Z'(\frac{k}{m})^2(\frac{2\pi m}{n})^2 \\ &= -\frac{2}{9}\frac{g''(0)}{g(0)}(\frac{2\pi m}{n})^2 m(1 + o(1)) \\ &= -\frac{8\pi^2}{9}\frac{g''(0)}{g(0)}\frac{m^3}{n^2}(1 + o(1)), \end{aligned}$$

where the first line follows from $\lambda_k = 2\pi k/n$, and the second line from $Z'(k/m)^2 = -[4/9]m(1 + o(1))$ by Lemma 2(c) in AG. The order of magnitude for the second term in (14) follows from AG or HDB (page 38) that $\sum_{k=1}^{m}(X_k - \overline{X})O(\lambda_k^4) = O(m^5/n^4)$.

Lastly, with the assumption of Gaussianity, we directly apply the Lemma 8 in HDB or Lemma 2(f) in AG. Then, we have $Z'E(\varepsilon) = O(\log^3 m)$. The proof of the variance term comes directly from HDB (proof of Theorem 1), then we omit it. This completes the proof.

PROOF OF COROLLARY 1: We verify the Theorem 2 in Robinson (1995), which is essential to show the asymptotic normality. The proof of asymptotic normality directly follows from that of Theorem 2 in HDB or of Theorem 2 in AG, which are based on

Robinson (1995). Write discrete Fourier transform of transformed series $\{w_j(q)\}$ for fixed j, and its normalized version as

(15) $u(\lambda_k) = (2\pi n)^{-1/2} \sum_{q=0}^{(2^j-1)(2v-1)} w_j(q) \exp(i\lambda_k q)$, and $v(\lambda_k) = u(\lambda_k)/f^{1/2}$.

The normalization is made by using $f^{1/2}$ rather than $C_j \lambda^{-2d} |\hat{\psi}(\lambda)|^2$.

$$
\begin{aligned}
& E\{u(\lambda_k)\overline{u}(\lambda_k)\} \\
= & (2\pi n)^{-1} \sum_{q=0}^{(2^j-1)(2v-1)} \sum_{r=0}^{(2^j-1)(2v-1)} E\{w_j(q) w_j(r)\} \exp\{i(q-r)\lambda_k\} \\
= & \int_{-\pi}^{\pi} f(\lambda)(2\pi n)^{-1} \sum_{q=0}^{(2^j-1)(2v-1)} \sum_{r=0}^{(2^j-1)(2v-1)} \exp\{-i(q-r)\lambda\} \exp\{i(q-r)\lambda_k\} d\lambda \\
= & \int_{-\pi}^{\pi} f(\lambda) K(\lambda_k - \lambda) d\lambda.
\end{aligned}
$$

where $K(\lambda) = (2\pi n)^{-1} \sum_{q=0}^{(2^j-1)(2v-1)} \sum_{r=0}^{(2^j-1)(2v-1)} \exp\{i(q-r)\lambda\}$. Then, we obtain the same expression as that of (4.1) in Robinson (1995), and the right hand side of (4.2) does not appear due to normalization. Thus, the proof of Theorem 2 in Robinson (1995) is applied to have

$$E\{v(\lambda_k)\overline{v}(\lambda_k)\} = 1 + O(\frac{\log k}{k}).$$

By similar reasoning, we also obtain

$$
\begin{aligned}
E\{u(\lambda_k)u(\lambda_k)\} &= \int_{-\pi}^{\pi} f(\lambda) D(\lambda_k - \lambda)(\lambda + \lambda_k) d\lambda, \\
E\{u(\lambda_k)\overline{u}(\lambda_s)\} &= \int_{-\pi}^{\pi} f(\lambda) D(\lambda_k - \lambda) D(\lambda - \lambda_s) d\lambda, \\
E\{u(\lambda_k)u(\lambda_s)\} &= \int_{-\pi}^{\pi} f(\lambda) D(\lambda_k - \lambda) D(\lambda + \lambda_s) d\lambda,
\end{aligned}
$$

where $D(\lambda) = (2\pi n)^{-1} \sum_{q=0}^{(2^j-1)(2v-1)} \exp(iq\lambda)$. Then, again by the proof of Robinson (1995), we verify that $E\{u(\lambda_k)u(\lambda_k)\} = O(k/\log k)$, $E\{u(\lambda_k)\overline{u}(\lambda_s)\} = O(k/\log s)$, and $E\{u(\lambda_k)u(\lambda_s)\} = O(k/\log s)$.

Given the above results, the proof of Theorem 2 in HDB or of Theorem 2 in AG follows.

REFERENCES

ANDREWS, D.W.K., AND P. GUGGENBERGER (2003): "A Bias-reduced Log Periodogram Regression Estimator for the Long-memory Parameter," *Econometrica*, forthcoming.

DAUBECHIES, I. (1992): *Ten Lectures on Wavelets*, CBS-NSF Regional Conferences in Applied Mathematics 61, Society for Industrial and Applied Mathematics.

FLANDRIN, P. (1992): "Wavelet Analysis and Synthesis of Fractional Brownian Motion," *IEEE Transactions on Information Theory*, 38, 910-917.

GEWEKE, J., AND S. PORTER-HUDAK (1983): " The Estimation and Application of Long Memory Time Series Models," *Journal of Time Series Analysis*, 4, 221-237.

HERNANDEZ, E., AND G. WEISS (1996): *A First Course on Wavelets*, CRC Press.

HURBICH, C.M., AND W. CHEN (2000): "An Efficient Taper for Potentially Overdifferenced Long-memory Time Series," *Journal of Time Series Analysis*, 21, 155-180.

HURBICH, C.M., R. DEO, AND J. BRODSKY (1998): "The Mean Squared Error of Geweke and Porter-Hudak's Estimator of the Memory Parameter of a Long-memory Time Series," *Journal of Time Series Analysis*, 19, 19-46.

HURBICH, C.M., AND B. RAY (1995): "Estimation of the Memory Parameter for Nonstationary or Noninvertible Fractionally Integrated Processes," *Journal of Time Series Analysis*, 16, 17-41.

JENSEN, M. (1997): "Using Wavelets to Obtain a Consistent Ordinary Least Squares Estimator of the Long-memory Parameter," *Journal of Forecasting*, 18, 17-32.

KATO, T., AND E. MASRY (1999): "On the Spectral Density of the Wavelet Transform of Fractional Brownian Motion," *Journal of Time Series Analysis*, 20, 559-563.

LEE, J., AND Y. HONG (2001): "Testing for Serial Correlation of Unknown Form using Wavelet Methods," *Econometric Theory*, 17, 386-423.

MASRY, E. (1993): "The Wavelet Transform of Stochastic Processes with Stationary Increments and Its Application to Fractional Brownian Motion," *IEEE Transactions on Information Theory*, 39, 260-264.

PHILLIPS, P.C.B. (1999): "Unit Root Log Periodogram Regression," discussion paper, Cowles Foundation, Yale University.

ROBINSON, P.M. (1995): "Log Periodogram Regression of Time Series with Long Range Dependence ," *Annals of Statistics*, 23, 1048-1072.

SHIMOTSU, K., AND P.C.B. PHILLIPS (2002): "Exact Local Whittle Estimation of Fractional Integration," discussion paper, Cowles Foundation, Yale University.

TANAKA, K. (1999): "The Nonstationary Fractional Unit Root," *Econometric Theory*, 15, 549-582.

TEWFIK, A.H., AND M. KIM (1992): "Correlation Structure of the Discrete Wavelet Coefficients of Fractional Brownian Motion," *IEEE Transactions on Information Theory*, 38, 904-909.

VELASCO, C. (1999): "Non-stationary Log Periodogram Regression," *Journal of Econometrics*, 91, 325-371.

FOOTNOTE

The author is grateful to Yongmiao Hong, Cheol Park, and the participants at the New Zealand Econometrics Study Group Meeting in 2003.

TABLE I

BIAS AND MSE OF LP ESTIMATORS: $n = 512$.

		Bias			MSE		
$\phi = 0$	d	W	DT	DF	W	DT	DF
	0.6	0.0200	0.0079	0.0104	0.0285	0.0349	0.0171
	0.8	0.0085	0.0156	0.0024	0.0289	0.0345	0.0169
	1.0	0.0079	0.0256	-0.0011	0.0291	0.0347	0.0166
	1.2	0.0116	0.0401	-0.0020	0.0297	0.0355	0.0163
	1.4	0.0218	0.0597	0.0027	0.0306	0.0374	0.0166
$\phi = 0.5$	0.6	0.0326	0.0671	0.0558	0.0285	0.0392	0.0206
	0.8	0.0211	0.0747	0.0504	0.0292	0.0398	0.0193
	1.0	0.0183	0.0847	0.0473	0.0297	0.0412	0.0189
	1.2	0.0213	0.0991	0.0467	0.0302	0.0437	0.0185
	1.4	0.0317	0.1184	0.0514	0.0312	0.0480	0.0191
$\phi = 0.8$	0.6	0.1000	0.3207	0.2662	0.0380	0.1374	0.0875
	0.8	0.0908	0.3285	0.2621	0.0376	0.1419	0.0852
	1.0	0.0883	0.3384	0.2594	0.0380	0.1484	0.0837
	1.2	0.0919	0.3525	0.2586	0.0386	0.1582	0.0832
	1.4	0.1013	0.3721	0.2615	0.0405	0.1723	0.0848