

Wavelet Transform for Estimating the Memory Parameter in Long Memory Stochastic Volatility Model

JIN LEE

Department of Economics, National University of Singapore, Singapore

E-mail: ecsleej@nus.edu.sg

Abstract: We consider semiparametric estimation of memory parameter in long memory stochastic volatility models. It is known that log periodogram regression estimator by Geweke and Porter-Hudak (1983) results in significant negative bias due to the existence of the spectrum of non-Gaussian noise process. Through wavelet transform of the squared process, we effectively remove the noise spectrum around zero frequency, and approximate the spectral density of squared process to that of long memory process only. We propose wavelet-based regression and local Whittle estimators. Simulation studies show that wavelet-based estimation is an effective way in reducing the bias. We present empirical applications to foreign exchange rate returns.

KEY WORDS: Spectral density of wavelet transforms, Gaussian semiparametric estimation, Long range dependence.

We consider semiparametric estimation of memory parameter of the latent process in long memory stochastic volatility (LMSV) models. In LMSV models, the spectral density of the nonlinear processes such as squared or log squared process is the sum of the spectral density of Gaussian long memory process and that of non-Gaussian noise process. Given the spectral representation of the squared processes, log periodogram (LP) estimator of Geweke and Porter-Hudak (1983; GPH) violates the Gaussian or Martingale assumption. As a result, GPH estimator suffers from significant negative bias mainly due to the existence of the spectrum of non-Gaussian noise. (Breidt et al (1998) and Deo and Hurvich (2001)).

In this paper, we introduce wavelet transform of the squared process to effectively remove the noise spectrum around zero frequency. In doing so, it is expected to attain improved rate of convergence of the mean squared error including bias reduction for memory estimation. We apply wavelet method to both LP and local Whittle (LW) estimation. Our simulation studies clearly confirms such theoretical conjecture, compared with GPH estimator. Since spectral density of squared process is approximated to that of the long memory process via wavelet transform, statistical inferences of Robinson (1995) are readily applicable. It is also noted that unlike GPH, the conditions on the growth rate of the fundamental frequency do not depend on unknown memory parameter for the consistency and for the asymptotic normality.

In Section 1, we study the LMSV models and introduce spectral representation of wavelet transforms. In sections 2 and 3, wavelet-based LP and LW estimators are proposed. Finite sample performances are presented in section 4, followed by an empirical example in section 5.

1. THE MODEL

We consider a LMSV model for discrete time series $\{X_t, t = 1, 2, \dots, n\}$

$$x_t = \sigma \exp(z_t/2)e_t \tag{1}$$

where $\{z_t\}$ is a latent Gaussian long memory process with the memory parameter $d \in (0, 0.5)$, which is independent of mean zero i.i.d. process $\{e_t\}$. We assume that the spectral behavior of z_t at zero frequency, which is standard in the long memory context.

Assumption A1: $f_z(\lambda) = \lambda^{-2d}g(\lambda)$ as $\lambda \rightarrow 0$, for $d \in (0, 0.5)$

where $g(\lambda)$ is an even function on $[-\pi, \pi]$, and $0 < g(0) < \infty$.

The log squared process as a volatility measure is written by

$$y_t = \log(x_t^2) = \eta + z_t + u_t \quad (2)$$

where $\eta = \log \sigma^2 + E(\log e_t^2)$ and $u_t = \log e_t^2 - E(\log e_t^2)$. Here, $\{u_t\}$ is mean zero i.i.d. with variance σ_u^2 . Autocovariances $R(j)$ of $\{y_t\}$ is identical to that of $\{z_t\}$ for $j \neq 0$.

Given A1, the spectral density of y_t is the sum of the spectral density of Gaussian long memory process and that of non-Gaussian noise process,

$$f_y(\lambda) = \lambda^{-2d}g(\lambda) + \frac{\sigma_u^2}{2\pi} = \lambda^{-2d}(g(\lambda) + \frac{\sigma_u^2}{2\pi}\lambda^{2d}), \quad \text{as } \lambda \rightarrow 0. \quad (3)$$

It is known that given the above spectral representation, GPH estimator violates the Gaussian or Martingale assumption which the asymptotic theory is built upon in the long memory context (Bollerslev and Wright (2000)). In particular, due to the existence of constant spectrum of $\{u_t\}$, the dominant term of the bias of GPH estimator behaves at the order of λ^{2d} . Then, GPH estimator suffers from significant negative bias, which is clearly pointed out in Deo and Hurvich (2001). It is noted that bias-reduced method of Andrews and Guggenberger (2003) is not directly applicable to LMSV model since the term $g(\lambda) + \frac{\sigma_u^2}{2\pi}\lambda^{2d}$ in (3) is not even function and the first and higher order biases can not be eliminated by including additional regressors to the power of even-numbered frequencies.

Below we make use of wavelet transform of the squared process and remove the noise spectrum around zero frequency. Define the wavelet transform for y_t

$$w_{jq} = 2^{j/2} \sum_t y_t \psi(2^j t - q), \quad (4)$$

where t is suitably re-indexed so that the support of the wavelet is fully covered. For example, if the support of ψ is $[0, 1]$, then we let $t = i/n$, for $i = 1, 2, \dots, n$. The integer valued j and q are scale and translation parameter, respectively, where $j = 0, 1, \dots, J$, $q = 0, 1, \dots, 2^j - 1$. The finest (maximum) scale is often set to $n = 2^J$. It can be seen that the transformed series w_{jq} is simply a linear combination of y_t over a varying interval which is determined by j and q . The function ψ is a wavelet, which is a well localized function.

We explicitly introduce the properties of the wavelet functions.

Assumption A2:

(a) $\psi : \mathbb{R} \rightarrow \mathbb{R}$ such that $\int_{-\infty}^{\infty} \psi(x)dx = 0$, $\int_{-\infty}^{\infty} |\psi(x)|dx < \infty$,

and $\int_{-\infty}^{\infty} (1+x^2)\psi(x)dx < \infty$.

(b) $|\hat{\psi}(\lambda)| = \lambda^v b(\lambda)$, with $b(t\lambda)/b(\lambda) = 1$ for all t , as $\lambda \rightarrow 0$,

with v integer, $0 < b(0) < \infty$,

where $\hat{\psi}(\lambda)$ is Fourier transform of ψ , $\hat{\psi}(\lambda) = (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{\infty} \psi(x)e^{-i\lambda x}dx$.

The assumption 2(a) describes the wavelet function. By Assumption 2(a), the spectral density function of w_{jq} is well defined (Kato and Masry(1999)). It is not necessary in our analysis that ψ forms an orthonormal basis for L_2 , though it is often the case in the wavelet literature. Next, assumption 2(b) models the spectral behavior of $\hat{\psi}(\lambda)$ around $\lambda = 0$. Integer-valued v is the number of vanishing moment of ψ in the sense that $\int_{-\infty}^{\infty} x^r \psi(x)dx = 0$ for $r = 0, 1, \dots, v-1$. The v vanishing moment is equivalent to saying that the first v spectral derivatives are zero at zero frequency, $\frac{d^r}{d\lambda^r} \hat{\psi}(\lambda) = 0$ at $\lambda = 0$ for $r = 0, 1, \dots, v-1$, from the relation, $\left. \frac{d^r}{d\lambda^r} \hat{\psi}(\lambda) \right|_{\lambda=0} = (-i)^r \int_{\mathbb{R}} x^r \psi(x)dx$. This assumption is satisfied if ψ has a compact support and belongs to $C^v(\mathbb{R})$, where $C^v(\mathbb{R})$ is the class of the functions f on the real line \mathbb{R} such that all the derivatives up to the order v exist, and the v -th derivative $f^{(v)}$ is continuous on \mathbb{R} . For example, Haar wavelet, defined as

$$\psi(x) = \begin{cases} 1 & 0 \leq x \leq 0.5 \\ -1 & 0.5 < x \leq 1, \end{cases} \quad (5)$$

satisfies A2 with $v = 1$. Further, $|\hat{\psi}(\lambda)| = (\lambda/4)[\sin^2(\lambda/4)/(\lambda/4)^2]$, where we have $b(0) = 1/4$. Another example includes a class of spline wavelets. The first order spline wavelet, often called Franklin wavelet, has $v = 2$. In general, the spline wavelets of order n has $n - 1$ vanishing moment (Hernandez and Weiss (1996)). Also, $b(\lambda)$ is assumed to be a slowly varying function at zero frequency. We use Haar wavelet for the analysis and for the simulation in our paper.

Write the wavelet transform of y_t using Haar wavelets,

$$w_{jq} = \alpha_{jq} + \beta_{jq} \quad (6)$$

where $\alpha_{jq} = 2^{j/2} \sum_{t=1/n}^1 z_t \psi(2^j t - q)$ and $\beta_{jq} = 2^{j/2} \sum_{t=1/n}^1 u_t \psi(2^j t - q)$.

We show that the spectral density of wavelet transform of $\{w_{jq}\}$ is approximated to that of wavelet transform of long memory, $\{\alpha_{jq}\}$ at zero frequency. Thus, transformed noise process does not contribute the spectral behavior of the squared process.

First, we analyze the spectral density of β_{jq} at zero frequency. Let $R_\beta(m) = E\beta_{jq}\beta_{jq+m}$ be the autocovariances of the transformed series β_{jq} at scale j , and $f_\beta^{(j)}$ be the spectral density at scale j . The wavelet transform β_{jq} is a linear combination of i.i.d. noise process u_t . Using Haar wavelet in (5), it is simply the difference of local sums of u_t over $t \in [2^{-j}q, 2^{-j}(q+0.5)]$ and over $t \in (2^{-j}q, 2^{-j}(q+0.5)]$. Moreover, β_{jq} becomes 1-dependent process. Thus, it behaves as MA(p) process of i.i.d. series, where the order p is determined by j . As j increases, the width of the interval decreases, and β_{Jq} becomes MA(1) process, that is to say, $\beta_{Jq} = 2^{J/2}(U_{2^{-J}q} - U_{2^{-J}(q+1)})$. Then, it can be seen that when $j = J$,

$$f_\beta^{(J)}(0) = \frac{1}{2\pi} \sum_{m=-\infty}^{\infty} R_\beta(m) = 0, \quad (7)$$

where $R_\beta(0) = 2\sigma_U^2$, $R_\beta(1) = -\sigma_U^2$, and $R_\beta(m) = 0$ for $|m| > 1$.

Next, we let $f_\alpha^{(j)}(\lambda)$ be the spectral density function of α_{jq} at scale j . Given A1, we directly obtain $f_\alpha^{(j)}(\lambda)$ as follows. We write autocovariances of α_{jq} at scale j as

$$\begin{aligned} E\alpha_{jq}\alpha_{j\tau} &= 2^j \sum_t \sum_s EZ_t Z_s \psi(2^j t - q) \psi(2^j s - \tau) \\ &= 2^{-j} \int_{-\pi}^{\pi} f_Z(\lambda) |\hat{\psi}(2^{-j}\lambda)|^2 e^{i2^{-j}(q-\tau)\lambda} d\lambda, \end{aligned}$$

where the second line follows from discrete Fourier transform of ψ and the change of variable. It is inferred that when $j = J$,

$$f_\alpha^{(J)}(\lambda) = 2^{-J} f_Z(\lambda) |\hat{\psi}(2^{-J}\lambda)|^2, \quad \lambda \in [-\pi, \pi] \quad (8)$$

Since the scale parameter is restricted to $j = J$, we suppress J as $f_w^{(J)}(\lambda) = f_w(\lambda)$. Combining (7) and (8), we have Gaussian-approximate spectral representation of the Haar wavelet transform around zero frequency

$$f_w(\lambda) = C_J \lambda^{-2(d-1)} g(\lambda) h(\lambda) \quad \text{as } \lambda \rightarrow 0, \text{ for } d \in (0, 0.5) \quad (9)$$

where $C = 2^{-3J}$ and $h(\lambda) = b^2(\lambda)$.

The functions $g(\lambda)$ and $h(\lambda)$ arise from short-run dependence in z_t and from wavelet transform, respectively. It is noted that both are even, continuous on $[-\pi, \pi]$, and bounded away from zero at zero frequency. Spectral representation (9) provides a basis for semi-parametric estimation of d . Note that the term $d-1$ appears from the vanishing moment condition, which is indeed de-correlation property of the Haar wavelet function (Tewfik and Kim (1992)). Below we use wavelet transformation to construct LP and LW estimators.

2. WAVELET-BASED LOG PERIODOGRAM ESTIMATOR

Now we construct wavelet-based LP estimator. We define a periodogram for wavelet transform at scale J ,

$$I_k^{(J)} \equiv I_k = \frac{1}{2\pi n} \sum_{q=0}^{2^J-1} |w_{Jq} \exp(i\lambda_k q)|^2, \quad k = 1, 2, \dots, m \quad (10)$$

where $\lambda_k = 2\pi k/n$. The periodogram can be simply computed by using the relation, $I_k = A_k^2 + B_k^2$, where $A_k = (2\pi n)^{-1/2} \sum_{q=0}^{2^J-1} w_{Jq} \cos(\lambda_k q)$ and $B_k = (2\pi n)^{-1/2} \sum_{q=0}^{2^J-1} w_{Jq} \sin(\lambda_k q)$.

As standard in the long memory literature, suitable conditions on the rate of growth m are imposed for the frequencies $\lambda_k = 2\pi k/n$, where $k = 1, 2, \dots, m$.

Assumption 3 : $m = m(n) \rightarrow \infty$, and $m/n \rightarrow 0$ as $n \rightarrow \infty$.

The positive integer m is restricted to increase at slower rate than n .

Under the spectral representation in (9) with $s(\lambda) = g(\lambda)h(\lambda)$, we write the LP regression as

$$\log I_k = \alpha + \beta X_k + \log(s(\lambda_k)/s(0)) + \varepsilon_k, \quad k = 1, 2, \dots, m \quad (11)$$

where $\alpha = (\log C_J + \log(s(0)))$, $\beta = (d - 1)$, $X_k = -2 \log(\lambda_k)$, and $\varepsilon_k = \log(I_k/f_k)$.

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

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

We obtain the asymptotic bias, and variance.

Theorem 1. Suppose Assumptions 1- 3 hold. Then,

$$\begin{aligned} (a) \ E\hat{d} - d &= -\frac{2\pi^2}{9} \frac{s''(0)}{s(0)} \frac{m^2}{n^2} (1 + o(1)) + O(m^4/n^4) + O\left(\frac{\log^3 m}{m}\right). \\ (b) \ Var(\hat{d}) &= \frac{\pi^2}{24m} + o\left(\frac{1}{m}\right). \end{aligned}$$

Theorem 1 shows that wavelet-based estimator \hat{d} is consistent for $d \in (0, 0.5)$ in the L_2 sense. The variance takes the same form as in the stationary Gaussian case. The proof is basically adapted from Hurvich, Deo and Brodsky (1998; HDB) and Andrews

and Guggenberger (2003;AG), as well as Robinson(1995). Further, we obtain the form of MSE

$$MSE(\widehat{d}) = \left[\frac{2\pi^2 s''(0)}{9 s(0)} \right]^2 \frac{m^4}{n^4} (1 + o(1)) + O\left(\frac{m^3 \log^3 m}{n^4}\right) + \frac{\pi^2}{24m} (1 + o(1)). \quad (12)$$

Given the expression of MSE, we directly obtain the optimal m^*

$$m^* = \left[0.4634 \cdot \left(\frac{s(0)}{s''(0)} \right)^{2/5} n^{4/5} \right] \quad (13)$$

where $[z]$ denotes the closest integer to z . Both $s(0)$ and $s''(0)$ are unknown, though the function h depends the known wavelet function. Thus, only the rate of the optimal m^* is available. It follows that we have $MSE(\widehat{d}) = O(n^{-4/5})$. This is the same convergence rate as that of GPH estimator for the mean long memory process, which is developed by HDB. Further, the rate of convergence does not depend on unknown parameter d , since the noise process does not contribute the bias and variance through wavelet transform.

Given the optimal rate of m above, the asymptotic normality can be applied to the wavelet-based estimator.

Corollary 1. Suppose Assumption 1-3 hold, and $m = o(n^{4/5})$, then

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

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

3. WAVELET-BASED LOCAL WHITTLE ESTIMATOR

In this section, we apply wavelet method to construct local Whittle estimator. Arteche (2004) considers LW estimator for the signal plus noise models including LMSV models. The LW estimator is constructed from the squared process Y_t directly, thus, as in the case of GPH estimator, it is not able to decrease the bias of the memory parameter estimates and the rate of convergence still depends on unknown parameter d . On the other hand, spectral density of squared process is approximated to that of only long memory process via wavelet transforms, where transformed noise spectrum becomes zero at zero frequency.

Given the spectral density function around zero frequency, we consider the following minimization problem

$$Q(G, d) = \frac{1}{m} \sum_{k=1}^m \left(\log G \lambda_k^{-2(d-1)} + \frac{\lambda_k^{2(d-1)}}{G} I_k \right).$$

In particular, the local Whittle estimator \widehat{d}^{lw} is obtained as

$$\widehat{d}^{lw} = \arg \min_{d \in (0, 0.5)} \left(\log \widehat{G}(d) - 2(d-1) \sum_{k=1}^m \log \lambda_k \right) \quad (14)$$

where $\widehat{G}(d) = \frac{1}{m} \sum_{k=1}^m \lambda_k^{2(d-1)} I_k$ with I_k in (10).

For consistency and asymptotic normality of \widehat{d}^{lw} , the proof directly follows from Robinson (1995b) by imposing suitable assumptions on the spectral density functions of the wavelet transforms. We provide a set of assumptions for consistency and for normality in the Appendix. First, we have consistency of wavelet-based LW estimator.

Theorem 2. Suppose A2, and A4-A8 in the Appendix hold. Then,

$$\widehat{d}^{lw} \rightarrow_p d \quad \text{as } n \rightarrow \infty.$$

The proof follows from Robinson (1995b, Theorem 1) if we write true spectral density of wavelet transform as $f_w = G_0 \lambda^{-2(d_0-1)}$ as $\lambda \rightarrow 0$. Next, we have asymptotic normality of \widehat{d}^{lw} .

Corollary 2. Suppose A2, A4, and A9-A11 in the Appendix hold. Then,

$$m^{1/2}(\widehat{d}^{lw} - d) \rightarrow_d N\left(0, \frac{1}{4}\right) \quad \text{as } n \rightarrow \infty.$$

If we define $R(d) = \log \widehat{G}(d) - 2(d-1) \sum_{k=1}^m \log \lambda_k$, then the Theorem 2 of Robinson (1995b) is readily applied by writing

$$0 = \frac{dR(\widehat{d})}{d\widehat{d}} = \frac{dR(d_0)}{dd} + \frac{d^2R(\widetilde{d})}{dd^2}(\widehat{d} - d_0),$$

where $|\widetilde{d} - d_0| \leq |\widehat{d} - d_0|$. Normality of $m^{1/2}(\widehat{d} - d_0)$ comes from the fact that $2m^{1/2}dR(d_0)/dd \rightarrow N(0, 1)$ together with $d^2R(\widetilde{d})/dd^2 \rightarrow 4$. See also Arteche (2004).

Wavelet transform enables one to obtain LW estimator out of the spectral density of long memory process only. Thus, we have the condition on m , which is identical to Robinson (1995b, A4')

$$\frac{1}{m} + \frac{m^{1+2\gamma}(\log m)^2}{n^{2\gamma}} \rightarrow 0, \quad \text{as } n \rightarrow \infty.$$

It follows that the optimal number of frequency $m^* = O(n^{2\gamma/(1+2\gamma)})$ and the convergence rate of MSE to zero is $O(n^{-2\gamma/(1+2\gamma)})$, where $\gamma \in (0, 2]$ such that $f_w(\lambda) = G_0 \lambda^{-2(d_0-1)}(1 + O(\lambda^\gamma))$, as $\lambda \rightarrow 0$. On the other hand, As in Arteche (2004), the condition on m depends

on unknown parameter d , that is to say, $1/m + m^{1+4d}(\log m)^2/n^{4d} \rightarrow 0$, as $n \rightarrow \infty$. This generates the convergence rate of MSE to zero behaves as $O(n^{-4d/1+4d})$. Thus, the convergence rate of MSE of wavelet-based LW estimator is faster than that of LW estimator in Arteche (2004) if $\gamma > 2d$. This condition is satisfied by parametric long memory models. For example, in the case of fractional ARMA model with $\gamma = 2$, wavelet transform attains faster convergence rate of $n^{-4/5}$, which comes from bias reduction, as in the case of LP estimation. We verify the theoretical conjecture by simulation studies below.

4. SIMULATION STUDIES

We compare the finite sample performance of the wavelet-based estimators and GPH estimator. In data generating process (1), we let $\sigma = 1$, and consider ARFIMA(1, d , 0) process for $\{z_t\}$,

$$(1 - \phi)(1 - L)^d z_t = \varepsilon_t \quad (15)$$

where ϕ is the autoregressive parameter, and ε_t is i.i.d. with variance σ_ε^2 . The $I(d)$ process $\{z_t\}_{t=1}^n$ is generated through $z_t = \sum_{k=0}^{t-1} \frac{(d)_k}{k!} \eta_{t-k}$, and $(d)_k = d(d+1) \cdots (d+k-1)$, where $\eta_t \sim i.i.d.N(0, 1)$. Sample size is set to $n = 1024$. The value of σ_ε^2 is set to 0.37 as in Deo and Hurvich (2001) and Breidt et al (1998). We only consider the combination of $(d, \phi) = (0.2, 0)$, $(0.2, 0.3)$ and $(0.3, 0.6)$. Other combinations show qualitatively similar results. Regarding the number of frequencies m , we include four values of m as $m_1 = [n^{0.5}]$, $m_2 = [n^{0.55}]$, $m_3 = [n^{0.6}]$, and $m_4 = [n^{0.65}]$, where $[x]$ denotes the integer part of x . Similar choice of m is found in Breidt and et al (1998). For each value of m , one thousand iterations are conducted. We report the bias and mean squared error (MSE) over different values of m .

For wavelet-based estimator, we use Haar wavelet. The integer-valued scale j is set to the finest scale J for the transformed periodogram. For $n = 1024 = 2^{10}$, we set $J = 10$, which generates the transformed series, $\{w_j(q), q = 0, 1, \dots, 2^{10} - 1\}$. For GPH estimator, we do not truncate the low frequency components, which is known to perform better than the truncated version of GPH estimator.(Deo and Hurvich (2001)). We label wavelet-based LP and LW estimators as \hat{d}_W and \hat{d}_W^w . The counterparts of GPH estimators are denoted as \hat{d}_G and \hat{d}_G^w , respectively.

The Table I presents the bias and MSE of wavelet-based LP estimator and GPH estimator. For $(d, \phi) = (0.2, 0)$, \hat{d}_G shows significant negative bias and the magnitude of

the bias nearly remains unchanged for all values of m . On the other hand, \widehat{d}_W significantly reduces the magnitude of the bias. Its bias slowly changes from positive to negative direction as m grows. In terms of MSE, \widehat{d}_G has smaller MSE than \widehat{d}_W for small values of m . As m become large, \widehat{d}_W delivers smaller MSE. This makes sense since the optimal rate of convergence of MSE of \widehat{d}_W grows at the rate of $n^{0.8}$, which is larger than that of GPH. In the cases of $(d, \phi) = (0.2, 0.3)$ and $(0.3, 0.6)$, which allow some short-run dependence, basic pattern of the bias and MSE remains unaffected. It is again pronounced that \widehat{d}_W attains significantly smaller bias than \widehat{d}_G in all cases.

Next, we compare \widehat{d}_W^{lw} and \widehat{d}_G^{lw} . We only present the results in the case of $m = [n^{0.65}]$, as the qualitative pattern of LW estimation is the same as that of LP estimation. In all combination of d and ϕ , \widehat{d}_W^{lw} greatly reduces the bias. Further, given the same values of m , \widehat{d}_W^{lw} shows smaller MSE than \widehat{d}_W , which confirms that LW estimator is more efficient than LP estimator. The simulations results for other values of m are available upon request.

In sum, our proposed wavelet-based estimators are effective in reducing the bias for memory parameter estimation in LMSV models.

5. EMPIRICAL EXAMPLE

The proposed estimator is applied to a set of exchange rate data. We consider the daily spot exchange rates of Yen, Euro and pound against US Dollar, denoted as Y/US, US/EU, and US/P, respectively. Data is collected from the website of federal reserve bank of St. Louis, and it ranges from 3, January, 2000 to 16, April, 2004, which has 1078 observations. The return is simply defined as logarithmic difference between the daily rates. We conduct wavelet-based local Whittle estimation for the memory parameters in squared return and log squared return. Two choices of $m = [n^{0.65}]$ and $[n^{0.8}]$ are included, where we can see the effects of values of m on the estimates.

Table III report the estimates and the test statistic for the hypothesis of $d = 0$. The test statistic is constructed as $m^{1/2}(\widehat{d}_W^{lw} - d)/(0.25)^{1/2}$. The results show strong evidence of long memory in the volatility dependencies in foreign exchange returns. This finding is consistent with earlier empirical studies, for example, of Bollerslev and Wright (2000) which analyzes high frequency exchange return. Also, we find that log squared returns show larger value of the memory estimates than squared returns. As m increases, the values of the estimates rather decrease. Nonetheless, the null hypothesis of $d = 0$ is clearly rejected in all cases.

6. CONCLUSIONS

We propose wavelet transform for semiparametric estimation of memory parameter in long memory stochastic volatility models. Though widely used in the empirical study, Geweke and Porter-Hudak (1983)'s log periodogram regression estimator results in significant negative bias due to the spectrum of the noise process. One can remove the noise spectrum around zero frequency through suitable wavelet transforms of the squared process. We develop wavelet-based regression and local Whittle estimators. Simulation studies show that wavelet method significantly reduces the bias of the memory parameter estimates, compared with GPH estimator.

REFERENCES

- Anderson, T.G., and Bollerslev, T. (1997), "Heterogeneous Information Arrivals and Return Volatility Dynamics: Uncovering the Long-Run in High Frequency Returns," *Journal of Finance*, 52, 975-1005.
- Andrews, D.W.K., and Guggenberger, P. (2003), "A Bias-reduced Log Periodogram Regression Estimator for the Long-Memory Parameter," *Econometrica*, 71, 675-712.
- Arteche, J. (2004), "Gaussian Semiparametric Estimation in Long Memory in Stochastic Volatility and Signal Plus Noise Models," *Journal of Econometrics*, 119, 131-154.
- Bollerslev, T., and Wright, J. (2000), "Semiparametric Estimation of Long-Memory Volatility Dependencies: The Role of High-Frequency Data," *Journal of Econometrics*, 98, 81-106.
- Breidt, F., Crato, N., and de Lima, P. (1998), "The Detection and Estimation of Long Memory in Stochastic Volatility," *Journal of Econometrics*, 83, 325-348.
- Deo, R. and Hurvich, C.M. (2001), "On the Log Periodogram Regression Estimator of the Memory Parameter in Long Memory Stochastic Volatility Models," *Econometric Theory*, 17, 686-710.
- Geweke, J., and Porter-Hudak, S. (1983), "The Estimation and Application of Long Memory Time Series Models," *Journal of Time Series Analysis*, 4, 221-237.
- Hernandez, E., and Weiss, G. (1996), *A First Course on Wavelets*. CRC Press.
- Hurvich, C.M., Deo, R. and Brodsky, J. (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.
- Kato, T., and Masry, E. (1999), "On the Spectral Density of the Wavelet Transform of Fractional Brownian Motion," *Journal of Time Series Analysis*, 20, 559-563.
- Lee, J. (2004), "Wavelet Transform for Regression Estimation of Non-Stationary Fractional Time Series," Manuscript, Department of Economics, National University of Singapore.
- Lobato, I. (1999), "A Semiparametric Two-Step Estimator in a Multivariate Long Memory Model," *Journal of Econometrics*, 90, 129-153.
- 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.
- Moulin, P. (1994), "Wavelet Thresholding Techniques for Power Spectrum Estimation," *IEEE Transactions on Signal Processing*, 42, 3126-3136.

- Robinson, P.M. (1995a), “Log Periodogram Regression of Time series with Long Range Dependence,” *Annals of Statistics*, 23, 1048-1072.
- (1995b), “Gaussian Semiparametric Estimation of Long-Range Dependence,” *Annals of Statistics*, 23, 1630-1661.
- (2001), “The Memory of Stochastic Volatility Models” *Journal of Econometrics*, 101, 195-218.
- Shimotsu, K., and Phillips, P.C.B. (2004), “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 Kim, M. (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.

APPENDIX

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

$$\log I = (\log C_J + \log s(0) + \bar{X}d)1_m + Z\beta + R + \varepsilon \quad (\text{A1})$$

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

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

$$Z'R = \frac{1}{2} \frac{s''(0)}{s(0)} Z'\lambda_k^2 + \sum_{k=1}^m (X_k - \bar{X})O(\lambda_k^4). \quad (\text{A2})$$

The first term in (A2) is written as

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

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 (A2) follows from AG or HDB (page 38) that $\sum_{k=1}^m (X_k - \bar{X})O(\lambda_k^4) = O(m^5/n^4)$. Lastly, under the 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: 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). Below we only verify the Theorem 2 in Robinson (1995), which is essential to show the asymptotic normality.

Write discrete Fourier transform of transformed series $\{w_{Jq}\}$ for fixed J , and its normalized version as

$$u(\lambda_k) = (2\pi n)^{-1/2} \sum_{q=0}^{2^J} w_{Jq} \exp(i\lambda_k q), \text{ and } v(\lambda_k) = u(\lambda_k)/f_J^{1/2}. \quad (\text{A3})$$

where f_J is the spectral density function of $\{w_{Jq}\}$, and the normalization is made by using $f_J^{1/2}$ for $v(\lambda_k)$. It follows that

$$\begin{aligned} & E\{u(\lambda_k)\bar{u}(\lambda_k)\} \\ &= (2\pi n)^{-1} \sum_{q=0}^{2^J-1} \sum_{r=0}^{2^J-1} E(w_{Jq}w_{Jr}) \exp\{i(q-r)\lambda_k\} \\ &= \int_{-\pi}^{\pi} f_J(\lambda) (2\pi n)^{-1} \sum_{q=0}^{2^J-1} \sum_{r=0}^{2^J-1} \exp\{-i(q-r)\lambda\} \exp\{i(q-r)\lambda_k\} d\lambda \\ &= \int_{-\pi}^{\pi} f_J(\lambda) K(\lambda_k - \lambda) d\lambda. \end{aligned}$$

where $K(\lambda) = (2\pi n)^{-1} \sum_{q=0}^{2^J-1} \sum_{r=0}^{2^J-1} \exp\{i(q-r)\lambda\}$. Then, we obtain the same expression as that of (4.1) in Robinson (1995). Thus, the proof of Theorem 2 in Robinson (1995) is applied to have

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

By similar reasoning, we also obtain

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

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

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

Assumptions for Theorem 2:

A4: For $\gamma \in (0, 2]$,

$$f_w(\lambda) = G_0 \lambda^{-2(d_0-1)} (1 + O(\lambda^\gamma)), \quad \text{as } \lambda \rightarrow 0$$

where $G_0 \in (0, \infty)$.

A5: In a neighborhood of $(0, \delta)$ of the zero frequency, $f_w(\lambda)$ is differentiable, and

$$\left| \frac{d}{d\lambda} \log f_w(\lambda) \right| = O(\lambda^{-1}), \quad \text{as } \lambda \rightarrow 0.$$

Both A4 and A5 governs the behavior of spectral density function of transforms $\{w_{Jq}\}$.

A6:

$$\alpha_{Jq} = \sum_{j=0}^{\infty} \theta_j \epsilon_{t-j}, \sum_{j=0}^{\infty} \theta_j^2 < \infty,$$

where $E(\epsilon_t | F_{t-1}) = 0$, and $E(\epsilon_t^2 | F_{t-1}) = 1$.

Wavelet transform α_{Jq} is simply a linear combination of Z_t . Then, Assumption A6 is implied by assuming that the long memory process Z_t is linear in square integrable martingale difference sequence in Robinson (1995b, A3).

A7:

$$\frac{1}{m} + \frac{m}{n} \rightarrow 0.$$

A8: α_{Jq} and β_{Jr} are independent for all q and r .

This is satisfied if the signal Z_t is independent of the noise e_s for all t and s .

Assumptions for Corollary 2:

A9: In a neighborhood of $(0, \delta)$ of the zero frequency, $\theta(\lambda) = \sum_{j=0}^{\infty} \theta_j e^{ij\lambda}$ is differentiable and $\frac{d}{d\lambda} \theta(\lambda) = O\left(\frac{|\theta(\lambda)|}{\lambda}\right)$, as $\lambda \rightarrow 0$, where γ_j is in A6 above.

A10: In addition to A6, $E(\epsilon_t^3 | F_{t-1}) = \mu_3$, $E(\epsilon_t^4 | F_{t-1}) = \mu_4$.

A11:

$$\frac{1}{m} + \frac{m^{1+2\gamma} (\log m)^2}{n^{2\gamma}} \rightarrow 0, \quad \text{as } n \rightarrow 0.$$

TABLE I

Bias and MSE of Regression Estimators: $n = 1024$.

$(d, \phi) = (0.2, 0)$	Bias		MSE	
	\hat{d}_G	\hat{d}_W	\hat{d}_G	\hat{d}_W
m_1	-0.1577	0.0407	0.0365	0.0377
m_2	-0.1582	0.0232	0.0337	0.0292
m_3	-0.1612	-0.0207	0.0324	0.0180
m_4	-0.1695	-0.0489	0.0332	0.0177
$(d, \phi) = (0.2, 0.3)$				
m_1	-0.1313	0.0471	0.0289	0.0350
m_2	-0.1309	0.0285	0.0257	0.0270
m_3	-0.1350	-0.0131	0.0246	0.0163
m_4	-0.1382	-0.0349	0.0235	0.0151
$(d, \phi) = (0.3, 0.6)$				
m_1	-0.0663	0.0185	0.0162	0.0211
m_2	-0.0670	0.0085	0.0134	0.0156
m_3	-0.0703	-0.0140	0.0113	0.0104
m_4	-0.0721	-0.0177	0.0098	0.0082

TABLE II

Bias and MSE of Local Whittle Estimators:
 $n = 1024, m = \lceil n^{0.65} \rceil$.

(d, ϕ)	Bias		MSE	
	\hat{d}_G^w	\hat{d}_W^w	\hat{d}_G^w	\hat{d}_W^w
(0.2, 0)	-0.1584	-0.0232	0.0269	0.0140
(0.2, 0.3)	-0.1383	-0.0145	0.0216	0.0125
(0.3, 0.6)	-0.0774	-0.0179	0.0097	0.0068

TABLE III

Estimates of memory parameter for volatilities in foreign exchange rate return:

	squared return			log squared return		
	Y/US	US/EU	US/P	Y/US	US/EU	US/P
$m = [n^{0.65}]$.	0.31 (5.97)	0.32 (6.17)	0.57 (10.99)	0.64 (12.34)	0.65 (12.53)	0.63 (12.15)
$m = [n^{0.8}]$.	0.19 (6.19)	0.2 (6.52)	0.34 (11.09)	0.42 (13.69)	0.42 (13.69)	0.44 (14.35)

NOTE: The value in the paranthesis is the Z-value for testing $H_0 : d = 0$.