Technical results for the paper, "Wavelet-Based Parameter Estimation for Trend Contaminated Fractionally Differenced Processes"

Peter F. Craigmile[1], **Peter Guttorp**[2] **and Donald B. Percival**[3].

[1] Department of Statistics, 1958 Neil Avenue, The Ohio State University, Columbus, OH 43210.

[2] Department of Statistics, Box 354322, University of Washington, Seattle, WA 98195–4322.

[3] Applied Physics Laboratory, Box 355640, University of Washington, Seattle, WA 98195–5640.

Proof of 4.1 See Theorem 5.1 of Craigmile and Percival (2003).

Proof of 4.2 For $d < (L+1)/2$, the wavelet coefficients have mean zero by the differencing properties of the Daubechies wavelet filter. When $\{X_t\}$ is stationary ($d < 1/2$) the theorem follows from Exercise [348b] of Percival and Walden (2000) (a solution to this exercise can be found in Craigmile (2000)). To establish the result for $d \geq 1/2$ we need only show that equation (6) is finite when $\tau = 0$ for all j since $S_j(\cdot)$ is then the spectrum for a stationary process. For brevity define $C_L(l) = \binom{L/2+l-1}{l}$. Then

$$\sigma^2 \, s_{j,0}(d) \;=\; 2\sigma^2 \int_0^{1/2} \mathcal{H}_{j,L}(f)\,(2\sin(\pi f))^{-2d}\, df.$$

For $j = 1$, we have (using equation (1))

$$s_{1,0}(d) \;=\; \frac{2^{2-2d}}{\pi} \sum_{l=0}^{L/2-1} C_L(l) \int_0^{\frac{\pi}{2}} \cos^{2l}(w) \sin^{L-2d}(w)\, dw,$$

by the substitution $w = \pi f$. From standard results the final integral exists for $L - 2d > -1$, i.e., $d < (L+1)/2$. Since L is even, standard trigonometry yields for $j > 1$

$$s_{j,0}(d) \;=\; \frac{2^{j+(j-1)L+1-2d}}{\pi} \sum_{l_0} .. \sum_{l_{j-1}} C_L(l_0)\ldots, C_L(l_{j-1}) \int_0^{\frac{\pi}{2}} \sin^{L-2d}(w) \prod_{k=0}^{j-2} \cos^{2L}(2^k w)$$
$$\times \cos^{2l_{j-1}}(2^{j-1}w) \prod_{k=0}^{j-2} \sin^{2l_k}(2^k w)\, dw,$$

which exists for $d < (L+1)/2$.

Lemma 1 *Under the white noise model, $R_j =_d s_{j,0}(d_0)\ \sigma_0^2\ \chi^2_{M_j}$ for each j. It then follows that $[R_j/(M_j s_{j,0}(d_0)) - 1] \to_p 0$, for each j, as $M \to \infty$.*

Proof The result follows directly from the white noise model, the weak law of large numbers, and the fact that $M_j \to \infty$ as $M \to \infty$.

We now define some extra notation. For any twice differentiable function, g, define the operator

$$\Delta_2(g(x)) = \frac{d}{dx}\Delta_1(g(x)) = \frac{\frac{d^2}{dx^2}g(x)}{g(x)} - \left(\frac{\frac{d}{dx}g(x)}{g(x)}\right)^2.$$

We also note that the second derivative of $s_{j,\tau}(d)$ with respect to d is

$$s''_{j,\tau}(d) \quad = \tfrac{d^2}{dd^2}s_{j,\tau}(d) \quad = 8\int_0^{1/2} [\log\sin(\pi f)]^2\, \mathcal{H}_{j,L}(f)\cos(2^{j+1}\pi f\tau)(2\sin(\pi f))^{-2d}\, df.$$

Lemma 2 *For the white noise model, the first two derivatives of the log-likelihood satisfy,*

$$-2\,\dot{l}_M(\boldsymbol{\theta}) \quad = \quad -2\begin{bmatrix} \frac{\partial}{\partial d}\, l_M(\boldsymbol{\theta}) \\ \frac{\partial}{\partial\sigma^2}\, l_M(\boldsymbol{\theta}) \end{bmatrix} \quad = \quad \begin{bmatrix} -\sum_{j=1}^{J}\left[\frac{R_j}{s_{j,0}(d)\sigma^2} - M_j\right]\Delta_1(s_{j,0}(d)) \\ -\frac{M(\widehat{\sigma}^2_M(d) - 1)}{\sigma^2} \end{bmatrix},$$

and

$$-2\,\ddot{l}_M(\boldsymbol{\theta}) \quad = \quad -2\begin{bmatrix} \frac{\partial^2}{\partial d^2}\, l_M(\boldsymbol{\theta}) & \frac{\partial^2}{\partial d\partial\sigma^2}\, l_M(\boldsymbol{\theta}) \\ \frac{\partial^2}{\partial d\partial\sigma^2}\, l_M(\boldsymbol{\theta}) & \frac{\partial^2}{\partial\sigma^4}\, l_M(\boldsymbol{\theta}) \end{bmatrix},$$

where

$$-2\,[\ddot{l}_M(\boldsymbol{\theta})]_{1,1} \quad = \quad \sum_{j=1}^{J} M_j\ \Delta_2(s_{j,0}(d)) - \sum_{j=1}^{J}\frac{R_j}{s_{j,0}(d)\sigma^2}\ \left(\Delta_2(s_{j,0}(d)) - \Delta_1^2(s_{j,0}(d))\right),$$

$$-2\,[\ddot{l}_M(\boldsymbol{\theta})]_{1,2} \quad = \quad \sum_{j=1}^{J}\frac{R_j}{s_{j,0}(d)\sigma^4}\ \Delta_1(s_{j,0}(d)),\ \textit{and}$$

$$-2\,[\ddot{l}_M(\boldsymbol{\theta})]_{2,2} \quad = \quad -\frac{M}{\sigma^4} + 2\sum_{j=1}^{J}\frac{R_j}{s_{j,0}(d)\sigma^6}.$$

Proof Using the notation of equation (13) and taking derivatives of equation (9) with respect to d yields

$$
\begin{aligned}
-2\,\frac{\partial}{\partial d}\,l_M(d,\sigma^2|(W_{nb})_{j,k}) &= \sum_{j=1}^{J} M_j\,\Delta_1(s_{j,0}(d)) - \sum_{j=1}^{J}\frac{R_j}{s_{j,0}(d)\sigma^2}\,\Delta_1(s_{j,0}(d)).\\
&= -\sum_{j=1}^{J}\left[\frac{R_j}{s_{j,0}(d)\sigma^2} - M_j\right]\Delta_1(s_{j,0}(d)),
\end{aligned}
$$

and with respect to σ^2,

$$
-2\,\frac{\partial}{\partial \sigma^2}\,l_M(d,\sigma^2|(W_{nb})_{j,k}) \;=\; \frac{M}{\sigma^2} - \sum_{j=1}^{J}\frac{R_j}{s_{j,0}(d)\sigma^4} \;=\; -\frac{M(\widehat{\sigma}^2_M(d)-1)}{\sigma^2}.
$$

The second derivatives follow in a similar way.

Lemma 3 *Under the white noise model.*

(a) $-2M^{-1/2}\dot{l}_M(\boldsymbol{\theta}_0) \rightarrow_d N_2(0, 4\boldsymbol{\Sigma}_0(\boldsymbol{\theta}_0))$ *as* $M \rightarrow \infty$;

(b) $-2M^{-1}\ddot{l}_M(\boldsymbol{\theta}_0) \rightarrow_{as} 4\boldsymbol{\Sigma}_0(\boldsymbol{\theta}_0)$ *as* $M \rightarrow \infty$.

Proof (a) Result follows by using the Cramér–Wold theorem and examining the characteristic function of $-2M^{-1/2}\dot{l}_M(\boldsymbol{\theta}_0)$. In the limit as $M \rightarrow \infty$ we can use lemma 1 and the uniqueness of characteristic functions to establish the result.

(b) Follows directly from lemma 2 and lemma 1.

Lemma 4 $f_L(d) \equiv s_{2,0}(d)/s_{1,0}(d)$ *is a strictly increasing function of* d *for* $d < (L+1)/2$.

Proof For $L = 2$ it can be shown that (Craigmile 2000, p. 45)

$$
f_2(d) \;=\; \frac{6}{(2-d)(3-d)},
$$

which is a strictly increasing function for $d < \frac{3}{2}$. Since $f'_L(d)$ is a continuous function of d for all L, one can validate graphically that $f'_L(d) > 0$ for a particular $L > 2$ and $d < (L+1)/2$. Craigmile (2000, Figure 3.9, p. 46) demonstrates this for $L = 4, 6, 8, \ldots, 20$.

Proof of 7.1 (a) To prove consistency, we follow the proof of Lehmann (1998), Theorem 3.7. For $r > 0$, let $B_r = \{\boldsymbol{\theta} \in \Theta_L : ||\boldsymbol{\theta} - \boldsymbol{\theta}_0|| = r\}$. We want to show that

$$P(l_M(\boldsymbol{\theta}) < l_M(\boldsymbol{\theta}_0) \text{ for all } \boldsymbol{\theta} \in B_r) \to 1, \tag{1}$$

as $M \to \infty$, or equivalently

$$P(2M^{-1}(l_M(\boldsymbol{\theta}) - l_M(\boldsymbol{\theta}_0)) < 0 \text{ for all } \boldsymbol{\theta} \in B_r) \to 1, \tag{2}$$

as $M \to \infty$. This implies that the likelihood equations have a local maximum inside B_r. Since the equations are satisfied at a local maximum, for any $r > 0$ with probability converging to one, the likelihood equations have a solution within B_r. Utilizing equation (9)

$$\begin{aligned} 2M^{-1}\left[l_M(\boldsymbol{\theta}) - l_M(\boldsymbol{\theta}_0)\right] &= -\log(2\pi\sigma^2) - \sum_{j=1}^{J}(M_j/M)\log(s_{j,0}(d)) - \sum_{j=1}^{J}\frac{R_j}{Ms_{j,0}(d)\sigma^2} \\ &\quad + \log(2\pi\sigma_0^2) + \sum_{j=1}^{J}(M_j/M)\log(s_{j,0}(d_0)) + \sum_{j=1}^{J}\frac{R_j}{Ms_{j,0}(d_0)\sigma_0^2} \\ &= \sum_{j=1}^{J}\frac{M_j}{M}\left[\log\left(\frac{s_{j,0}(d)\sigma^2}{s_{j,0}(d_0)\sigma_0^2}\right) + \left(1 - \frac{s_{j,0}(d)\sigma^2}{s_{j,0}(d_0)\sigma_0^2}\right)\right] \\ &\quad + \sum_{j=1}^{J}\frac{M_j}{M}\left(\frac{R_j}{M_j s_{j,0}(d_0)} - 1\right)\left(1 - \frac{s_{j,0}(d)\sigma^2}{s_{j,0}(d_0)\sigma_0^2}\right). \end{aligned}$$

Remembering that $m_j = \lim_{M\to\infty}(M_j/M)$ is a constant and using lemma 1, we see that the second term goes in probability to zero as $M \to \infty$. We will establish the result if we show that

$$\frac{M_j}{M}\left[\log\left(\frac{s_{j,0}(d)\sigma^2}{s_{j,0}(d_0)\sigma_0^2}\right) + \left(1 - \frac{s_{j,0}(d)\sigma^2}{s_{j,0}(d_0)\sigma_0^2}\right)\right]$$

is nonpositive for all $M >$ (some)M_0 and for each j, and is negative for at least one j. Now let $f(x) \equiv \log(x) + 1 - x$. Considering the derivatives of f we see that the function has maximum value at $x = 1$ with value $f(x) = 0$. Thus the quantity above is always nonpositive and negative if at least one of values, $(s_{j,0}(d)\sigma^2)/(s_{j,0}(d_0)\sigma_0^2)$, is not equal to one. As we assume that $\boldsymbol{\theta} \in B_r$, we have that $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$. This leads to three possible cases:

1. $d_0 = d$ and $\sigma_0^2 \neq \sigma^2$. In that case $(s_{j,0}(d)\sigma^2)/(s_{j,0}(d_0)\sigma_0^2) = \sigma^2/\sigma_0^2 \neq 1$.

2. $d_0 \neq d$ and $\sigma_0^2 = \sigma^2$. In that case $(s_{j,0}(d)\sigma^2)/(s_{j,0}(d_0)\sigma_0^2) = s_{j,0}(d)/s_{j,0}(d_0) \neq 1$ by definition of $s_{j,0}(d)$ (it is an increasing function of d for fixed j).

3. $\sigma_0^2 \neq \sigma^2$ and $d_0 \neq d$. By Lemma 4 it follows that if $(s_{1,0}(d)\sigma^2)/(s_{1,0}(d_0)\sigma_0^2) = 1$, then $(s_{2,0}(d)\sigma^2)/(s_{2,0}(d_0)\sigma_0^2) \neq 1$.

Thus with probability one the likelihood evaluated at $\boldsymbol{\theta} \in B_r$ is smaller than that at $\boldsymbol{\theta}_0$. If we let $r \to 0$ we obtain the consistency result by always taking the root of the likelihood equations closest to $\boldsymbol{\theta}_0$.

(b) A Taylor series expansion for $\dot{l}_M(\widehat{\boldsymbol{\theta}}_M)$ about $\boldsymbol{\theta}_0$ is given by

$$\dot{l}_M(\widehat{\boldsymbol{\theta}}_M) = \dot{l}_M(\boldsymbol{\theta}_0) + \ddot{l}_M(\boldsymbol{\theta}^*)(\widehat{\boldsymbol{\theta}}_M - \boldsymbol{\theta}_0)$$

where $\boldsymbol{\theta}^*$ lies between $\boldsymbol{\theta}_0$ and $\widehat{\boldsymbol{\theta}}_M$. Since $\dot{l}_M(\widehat{\boldsymbol{\theta}}_M) = \mathbf{0}$

$$\dot{l}_M(\boldsymbol{\theta}_0) = [-\ddot{l}_M(\boldsymbol{\theta}^*)](\widehat{\boldsymbol{\theta}}_M - \boldsymbol{\theta}_0).$$

The asymptotic normality of $\widehat{\boldsymbol{\theta}}_M$ follows by dividing this equation by $\sqrt{M}$, and noting that

(1) $M^{-1/2}\dot{l}_M(\boldsymbol{\theta}_0) \to_d N_2(0, \Sigma_0(\boldsymbol{\theta}_0))$ (by Lemma 3);

(2) $-M^{-1}\ddot{l}_M(\boldsymbol{\theta}_0) \to_p \Sigma_0(\boldsymbol{\theta}_0))$ (by Lemma 3);

(3) $M^{-1}\left[\ddot{l}_M(\boldsymbol{\theta}_0) - \ddot{l}_M(\boldsymbol{\theta}^*)\right] \to_p 0$ (by part (a), lemma 3 and continuity of the second derivative).

Invertibility of $\Sigma_0(\boldsymbol{\theta}_0)$, consistency of the ML estimate and Slutsky's theorem yields the required result.

(c) The result follows from the Cramér–Wold theorem using the vector $\boldsymbol{a} = (1, 0)$. In that case $\psi_0^2(d_0) = (1,0)^T \Sigma_0^{-1}(\boldsymbol{\theta}_0)(1,0)$.

(d) This result follows immediately from lemma 1, noting that different wavelet levels are independent under the white noise model.

Proof of 7.2 The proof of this result is similar in style to the proof of Theorem 7.1 – see Craigmile (2000) for further details.

References

Craigmile, P. F. (2000). *Wavelet Based Estimation for Trend Contaminated Long Memory Processes.* Ph. D. thesis, University of Washington, Department of Statistics.

Craigmile, P. F. and D. B. Percival (2003). Asymptotic decorrelation of between-scale wavelet coefficients. Technical Report 711, Department of Statistics, The Ohio State University.

Lehmann, E. L. (1998). *Theory of Point Estimation* (Second ed.). New York: Springer-Verlag.

Percival, D. and A. Walden (2000). *Wavelet Methods for Time Series Analysis.* Cambridge: Cambridge University Press.