The Impact of Wavelet Coefficient Correlations on Fractionally Differenced Process Estimation

Peter F. Craigmile Donald B. Percival Peter Guttorp

NRCSE

Technical Report Series

NRCSE-TRS No. 049

June 9, 2000

The **NRCSE** was established in 1996 through a cooperative agreement with the United States Environmental Protection Agency which provides the Center's primary funding.

The Impact of Wavelet Coefficient Correlations on Fractionally Differenced Process Estimation

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

Last updated June 9, 2000.

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

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

[3]MathSoft Inc., 1700 Westlake Avenue North, Suite 500, Seattle. WA 98109–3044.

Abstract

The discrete wavelet transform (DWT) approximately decorrelates a fractionally differenced (FD) process, allowing for simple maximum likelihood estimation of the FD process parameters using the wavelet coefficients. In previous work we have established limit theorems for the parameters based on a model where scales are uncorrelated and two simple models for within-scale correlation, namely, white noise and a first order autoregressive (AR) process. Here we assess the adequacy of these simple models for handling between- and within-scale correlations. We compare the performance of these simple models for estimating the FD process parameters against procedures that use longer wavelet filters (to reduce between-scale correlations) and use AR models of higher order (to more precisely model within-scale correlations).

1 Introduction

Time series collected in areas such as atmospheric sciences, geosciences and hydrology often exhibit a long range dependence, i.e., slowly decaying auto-correlations or, equivalently, a spectral density function (SDF) that is proportional to $|f|^{\alpha}$ at low frequencies for some $\alpha < 0$. A convenient model for such series is a fractionally differenced (FD) process (5; 6). Specifically, for $d \in [-\frac{1}{2}, \frac{1}{2})$ and $\sigma_\epsilon^2 > 0$, $\{X_t\}_{t\in\mathcal{Z}}$ is an FD(d, σ_ϵ^2) process if its SDF is given by

$$S_X(f) \quad = \quad \sigma_\epsilon^2 |2\sin(\pi f)|^{-2d} \quad f \in [-\tfrac{1}{2}, \tfrac{1}{2}]. \tag{1}$$

d is known as the *difference parameter* and σ_ϵ^2 is the *innovation variance*. When $d = 0$, $\{X_t\}$ is a white noise (i.e. uncorrelated) process. Extending this model by letting $d \geq \frac{1}{2}$ in equation (1), we obtain a class of non-stationary FD processes that are stationary if we difference $\lfloor d + \frac{1}{2} \rfloor$ times.

Given a time series that is a realisation of a portion $\{X_t\}_{t=0}^{N-1}$ of a stationary FD process, McCoy and Walden (9) extended earlier work by Wornell (13) to obtain effective approximate maximum likelihood (ML) estimators of the FD parameters d and σ_ϵ^2. The basis of their scheme was to formulate the likelihood function in terms of the discrete wavelet transform (DWT) of $\{X_t\}$ by making use of the assumption that the DWT of a Gaussian FD process yields approximately independent deviates. In previous work (10; 3), we extended the McCoy and Walden estimator to handle both stationary and non-stationary FD processes observed in the presence of a trend; i.e., the observed time series is taken to be a realisation of

$$Y_t \quad = \quad T_t + X_t \qquad t = 0, \ldots, N-1. \tag{2}$$

Here $\{T_t\}$ is a deterministic polynomial trend of order K, and $\{X_t\}$ is a realisation of a Gaussian FD(d, σ_ϵ^2) process. As with the McCoy and Walden scheme, the key assumption behind this extension is the independence of certain wavelet coefficients across and between scales. In this paper we re-examine this assumption. After a review of background material in Section 2, we argue in Section 3 that the correlation between scales can be made arbitrarily small by increasing the length of the wavelet filter. This increase in filter length, however, does not help reduce the correlation within scales, so we consider in Section 4 modelling this correlation using autoregressive (AR) models whose coefficients are scale-dependent but are solely determined by d. We conclude that a first order AR model is adequate for modelling the correlation structure within scales.

2 Definitions and Background on Wavelet Coefficients

For an even integer L, let $\{h_l\}_{l=0}^{L-1}$ denote a Daubechies (4) wavelet filter. By definition this filter has squared gain function

$$\mathcal{H}_{1,L}(f) \equiv 2 \sin^L(\pi f) \textstyle\sum_{l=0}^{L/2-1} \binom{L/2-1+l}{l} \cos^{2l}(\pi f). \tag{3}$$

Associated with the wavelet filter we define the scaling filter by $g_l \equiv (-1)^{l+1} h_{L-1-l}$ (with a squared gain function of $\mathcal{G}_{1,L}(f) = \mathcal{H}_{1,L}(\frac{1}{2} - f)$). Assume for convenience that $N = 2^J$ for some integer J, and let $N_j \equiv N2^{-j}$. The level j wavelet coefficients can be computed using the level j wavelet filter $\{h_{j,l}\}_{l=0}^{L_j-1}$:

$$W_{j,k} = \sum_{l=0}^{L_j-1} h_{j,l} X_{2^j(k+1)-l-l \bmod N_{j-1}} \qquad j = 1, \ldots, J, \quad k = 0, \ldots, N_j - 1$$

where $L_j \equiv (2^j - 1)(L - 1) + 1$ and $\{h_{j,l}\}$ has squared gain function

$$\mathcal{H}_{j,L}(f) \quad \equiv \quad \mathcal{H}_{1,L}(2^{j-1} f) \prod_{k=0}^{j-2} \mathcal{G}_{1,L}(2^k f). \tag{4}$$

These coefficients are associated with changes in averages on scale $\tau_j \equiv 2^{j-1}$ and with times spaced $\lambda_j \equiv 2^j$ units apart. In practice we use the pyramid algorithm (Mallat (8)) to calculate these DWT coefficients efficiently (see Percival and Walden (11)). Since the DWT handles filtering operations periodically, the first $B_j \equiv \lceil (L-2)(1-2^{-j}) \rceil$ wavelet coefficients are explicitly affected by the circularity assumption. We call these coefficients the *boundary dependent (BD) coefficients.* We call the remaining $M_j \equiv N_j - B_j$ which are unaffected by boundaries the *boundary independent (BI) coefficients.* Let $M \equiv \sum_{j=1}^{J} M_j$.

In Craigmile *et al.* (3) we noted that the BI wavelet coefficients are unaffected by the polynomial trend if $K \leq \frac{L}{2}$, and thus we can estimate the parameters of $\{X_t\}$ via Gaussian likelihood using these coefficients (if $\frac{L}{2} \geq \lfloor d - \frac{1}{2} \rfloor$). We further assumed that the BI wavelet coefficients are uncorrelated between scales, and either a white noise or AR(1) model was a good fit for these coefficients on each level. We now investigate this further.

3 Between-Scale Decorrelation

Let $(W_w)_{j,k}$ denote the BI wavelet coefficients ($j = 1, \ldots, J,\ k = 0, \ldots, M_j$). From chapter 9 of Percival and Walden (11) we have that

$$\begin{aligned}\mathrm{Cov}((W_w)_{j,k}, (W_w)_{j',k'}) \quad &= \quad 2^{1-2d}\sigma_\epsilon^2 \int_0^{\frac{1}{2}} \cos(2\pi f(2^j(k+1) - 2^{j'}(k'+1))) \\ &\qquad \times H_{j,L}(f)\, H^*_{j',L}(f)\, \sin^{-2d}(\pi f)\, df,\end{aligned}$$

where $H_{j,L}(f)$ is the Fourier transform of $\{h_{j,l}\}$ and denotes * the complex conjugation operator. An extension of Theorem 3.2 in Craigmile *et al.* (3) shows that this integral is finite for $d < \frac{L+1}{2}$. $\mathcal{H}_{j,L}(f)$ corresponds to an approximate band-pass filter with pass-band $[2^{-(j+1)}, 2^{-j}]$ (see, e.g., Daubechies

(4)). This approximate filter has a squared gain function given by $\mathcal{H}_{1,bp}(f) \equiv 2^j 1_{[2^{-(j+1)},2^{-j}]}(f)$. Lai (7) defines the following squared gain function

$$\mathcal{H}_{1,I}(f) \equiv \begin{cases} 0, & f \in [0, \frac{1}{4}); \\ 1, & f = \frac{1}{4}; \\ 2, & f \in (\frac{1}{4}, \frac{1}{2}], \end{cases}$$

and shows that $\mathcal{H}_{1,L}(f) \to \mathcal{H}_{1,I}(f)$ as $L \to \infty$ for all $f \in [0, \frac{1}{2}]$. Thus if we define $\mathcal{H}_{j,I}(f) \equiv \mathcal{H}_{1,I}(2^{j-1}f) \prod_{k=0}^{j-2} \mathcal{H}_{1,I}(\frac{1}{2} - 2^k f)$, we have $\mathcal{H}_{j,L}(f) \to \mathcal{H}_{j,I}(f)$ as $L \to \infty$ for all $f \in [0, \frac{1}{2}]$ and $j \geq 1$. $\mathcal{H}_{j,I}(f)$ differs from $\mathcal{H}_{1,bp}(f)$ on a countable set of points and thus an integral involving either of these two squared gain functions will be the same. By the spectral representation theorem we can therefore see that the BI wavelet coefficients at different scales are asymptotically uncorrelated for large L, since the pass-bands of these squared gain functions do not intersect. (see Craigmile (2) for additional details).

Lai (7) also proves that convergence of $\mathcal{H}_{1,L}(f)$ is monotone in the following sense. For all even L, $\mathcal{H}_{1,L}(\frac{1}{4}) = 1$,

$$\begin{aligned} \mathcal{H}_{1,L}(f) \geq \mathcal{H}_{1,L+2}(f) \geq \mathcal{H}_{1,I}(f), \quad f \in [0, \tfrac{1}{4}); \\ \mathcal{H}_{1,L}(f) \leq \mathcal{H}_{1,L+2}(f) \leq \mathcal{H}_{1,I}(f), \quad f \in (\tfrac{1}{4}, \tfrac{1}{2}]. \end{aligned}$$

For $j > 1$ this translates into

$$\mathcal{H}_{j,L}(f) \geq \mathcal{H}_{j,L+2}(f) \geq \mathcal{H}_{j,I}(f), \quad f \in [2^{-j}, \tfrac{1}{2}]$$

meaning that the side lobe behaviour of $\mathcal{H}_{j,L}(f)$ reduces with increasing L. Also the decorrelation between higher and lower scales is rapid with increasing L because there is less intersection of the squared gain functions. Figure 1 illustrates this decay for a number of wavelet filter lengths and $j = 1, \ldots, 4$.

See Tewfik and Kim (12) for a related discussion on the correlation structure of a DWT of fractional Brownian motion.

4 An AR(p) wavelet model

We now consider the within-scale dependence. On level j we can write the lag τ auto-covariance as

$$\sigma_\epsilon^2 \sigma_{j,\tau}(d) \equiv \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{i2\pi f\tau} S_j(f)\, df, \tag{5}$$

where we define the SDF of the level j BI wavelet coefficients by

$$S_j(f) \equiv \sigma_\epsilon^2\, 2^{-j-2d} \sum_{k=0}^{2^j-1} \mathcal{H}_{j,L}(2^{-j}(f+k))\ \sin^{-2d}(2^{-j}(f+k)). \tag{6}$$

In Craigmile *et al.* (3) we showed that, if we assume that the BI wavelet coefficients are either a portion of a white noise or AR(1) process, then the estimate of d is asymptotically normal, and the estimate of σ_ϵ^2 follows a scaled chi-squared distribution. Both estimates are consistent. Since we can approximate any continuous SDF by an AR(p) SDF for large enough p (Anderson (1)), a better approximation is given by supposing that the BI wavelet coefficients $\{(W_w)_{j,k} : k = 0, \ldots, M_j - 1\}$ are a portion of an AR(p) process, i.e.,

$$(W_w)_{j,k} = \sum_{r=1}^{p} \phi_{r,p}(j,d)(W_w)_{j,k-r} + Z_{j,k} \qquad (k = p, \ldots, M_j - 1) \tag{7}$$

where $\{Z_{j,k} \sim$ i.i.d. $N(0, \eta(j,d)\sigma_\epsilon^2) : j = 1 \ldots J,\ k = 0, \ldots, M_j - 1\}$. Figure 2 illustrates this for an FD(0.25,1) process analysed using an wavelet filter with $L = 8$. The top left panel shows the SDF of the process along with the approximate band-passes that correspond to the first five wavelet levels. The top middle panel shows the actual SDF of the BI wavelet coefficients (equation 6) and the right-hand panel is the SDF if we assume that that the BI coefficients are uncorrelated per each wavelet level. If we assume the AR(p) model as above for $p = 1, 2, 5$ we have an SDF given in the lower panels of figure 2. The SDF looks better for higher p, but not by too much (the AR(1) approximation is very good as it stands). We now employ the Levinson-Durbin (LD) recursions. Let $\phi_{1,1}(j,d) \equiv \sigma_1(j,d)/\sigma_0(j,d)$, $P_0(j,d) \equiv \sigma_0(j,d)$ and $P_1(j,d) \equiv \sigma_0(j,d)(1 - \phi_{1,1}^2(j,d))$. Then for $s > 1$

$$\begin{aligned}
\phi_{s,s}(j,d) &= P_{s-1}^{-1}(j,d)\left(\sigma_s(j,d) - \sum_{r=1}^{s-1} \phi_{r,s-1}(j,d)\, \sigma_{s-r}(j,d)\right); \\
\phi_{r,s}(j,d) &= \phi_{r,s-1}(j,d) - \phi_{s,s}(j,d)\, \phi_{s-r,s-1}(j,d) \qquad (r = 1, \ldots, s-1); \\
P_s(j,d) &= \sigma_0(j,d) - \sum_{r=1}^{s} \phi_{r,s}(j,d)\, \sigma_r(j,d) = P_{s-1}(j,d)(1 - \phi_{s,s}^2(j,d)).
\end{aligned}$$

The Yule-Walker equations show that $P_p(j,d) = \eta(j,d)$. Letting $M_{jp} \equiv M_j - p$, the likelihood for one wavelet level ($j = 1 \ldots J$) is

$$\begin{aligned} l(j,d,\sigma_\epsilon^2) \;\; &\equiv \;\; -\tfrac{M_j}{2}\left[\log(2\pi\sigma_\epsilon^2)\right] - \tfrac{1}{2}[M_{jp}\log(P_p(j,d)) + \textstyle\sum_{k=0}^{p-1}\log(P_k(j,d))] \\ & \quad -\tfrac{1}{2\sigma_\epsilon^2}\left[\textstyle\sum_{k=0}^{p-1}\frac{(\vec{Z}_{j,k}^{(k)}(d))^2}{P_k(j,d)} + \textstyle\sum_{k=p}^{M_j-1}\frac{Z_{j,k}^2}{P_p(j,d)}\right] \end{aligned}$$

where

$$\vec{Z}_{j,k}^{(k)}(d) \;\; \equiv \;\; \begin{cases} (W_w)_{j,0}, & k=0; \\ (W_w)_{j,k} - \sum_{r=1}^{k}\phi_{r,k}(j,d)\,(W_w)_{j,k-r}, & k = 1\ldots p-1, \end{cases}$$

and hence assuming that wavelet levels are uncorrelated

$$l_N(d,\sigma_\epsilon^2) \equiv \textstyle\sum_{j=1}^{J} l(j,d,\sigma_\epsilon^2). \tag{8}$$

Maximising with respect to σ_ϵ^2 yields the ML estimate

$$\widehat{\sigma}_{\epsilon,N,p}^2(d) \;\; = \;\; \tfrac{1}{M}\textstyle\sum_{j=1}^{J}\left[\textstyle\sum_{k=0}^{p-1}\frac{(\vec{Z}_{j,k}^{(k)}(d))^2}{P_k(j,d)} + \textstyle\sum_{k=p}^{M_j-1}\frac{Z_{j,k}^2}{P_p(j,d)}\right]. \tag{9}$$

The profile likelihood with respect to d is

$$\begin{aligned} l_N(d,\widehat{\sigma}_{\epsilon,N,p}^2(d)) \;\; &\equiv \;\; -\tfrac{M}{2}\left[\log(2\pi\widehat{\sigma}_{\epsilon,N,p}^2(d)) + 1\right] \\ & \quad -\tfrac{1}{2}\textstyle\sum_{j=1}^{J}\left[M_{jp}\log(P_p(j,d)) + \textstyle\sum_{k=0}^{p-1}\log(P_k(j,d))\right]. \end{aligned} \tag{10}$$

We maximise this expression to obtain $\hat{d}_{N,p}$. Now let $\hat{\theta}^T \equiv (\hat{d}_{N,p}, \widehat{\sigma}_{\epsilon,N,p}^2(d))$ denote the vector of estimates. We can extend the results of Craigmile *et al.* (3) as follows (see (2) for a proof).

Theorem 4.1 *For a differentiable function $g(\cdot)$, let $\Delta_1(g(x)) \equiv [\frac{\partial}{\partial y}g(y)|_{y=x}]/g(x)$. Suppose that equation (7) holds. For $d < \frac{L+1}{2}$, as $N \to \infty$*

(a) $(\hat{\theta} - \theta) \to_p 0$;

(b) $\sqrt{N}(\hat{\theta} - \theta) \to_d N(0, \Gamma^{-1}(\theta))$;

(c) $\sqrt{N}(\hat{d}_{N,p} - d) \to_d N(0, \sigma_{d,p}^2)$,

where

$$2\Gamma(\theta) \;\; \equiv \;\; \begin{bmatrix} \sum_j \Delta_1^2(P_p(j,d))\,2^{-j} & \sigma_\epsilon^{-2}\sum_j \Delta_1(P_p(j,d))\,2^{-j} \\ \sigma_\epsilon^{-2}\sum_j \Delta_1(P_p(j,d))\,2^{-j} & \sigma_\epsilon^{-4} \end{bmatrix},$$

and $\sigma^2_{d,p} \equiv 2[\sum_j \Delta_1^2(P_p(j,d))\, 2^{-j} - (\sum_j \Delta_1(P_p(j,d))\, 2^{-j})^2]^{-1}$.

For the same range of d *and any* N, $\hat{\sigma}^2_{\epsilon,N,p}(d) =_d M^{-1}\sigma^2_\epsilon \chi^2_M$.

Table 1 shows $\sigma^2_{d,p}$ for various values of filter length L ($L = \infty$ refers to using the ideal wavelet filter $\mathcal{H}_{j,I}(f)$) and difference parameter d under different AR(p) wavelet models ($p = 0$ refers to the white noise wavelet model of Craigmile *et al.* (3)). We analyse to J=6. In general, keeping L fixed, the asymptotic variance decreases with increasing d (especially for shorter values of L). It also decreases with increasing L for stationary $d < \frac{1}{2}$, but increases with L for non-stationary $d \geq \frac{1}{2}$. The limit variance only changes slightly for $L = 2$ as we increase p. For $L > 2$ there is little change in the asymptotic variance with $p \geq 1$. In fact Monte Carlo studies to estimate d for various samples sizes, filter lengths and values of the difference parameter showed that an AR(1) model in this case was sufficient. An AR(p) ($p > 1$) gave no improvement to estimation.

5 Conclusions

In this paper we have further examined the estimation of the parameters of a trend contaminated FD process using the DWT. We have demonstrated that as we increase the filter length we can decorrelate between wavelet scales and decrease side-lobe behaviour. By extending the white noise and AR(1) wavelet models to the AR(p) ($p > 1$) case, we do not improve the estimation of d from that of the AR(1) model. Clearly these results give an attractive framework in which to model other short and long dependent processes.

Acknowledgments

The authors gratefully acknowledge support for this research from an STTR grant from AFOSR (MathSoft, Inc., and the University of Washington), an NSF grant (MathSoft, Inc., and the University of Washington) and an EPA grant (National Research Center for Statistics and the Environment).

References

[1] T.W. Anderson, *The Statistical Analysis of Time Series (Classics Edition)*, Wiley-Interscience, (1994).

[2] P.F. Craigmile, *Wavelet-Based Parameter Estimation for Trend Contaminated Long Memory Processes*, Ph.D. Thesis, University of Washington, (2000).

[3] P.F. Craigmile, D.B. Percival and P. Guttorp, *Wavelet-Based Parameter Estimation for Trend Contaminated Fractionally Differenced Processes*, Technical Report, National Research Center for Statistics and the Environment, University of Washington, (2000).

[4] I. Daubechies, *Ten Lectures on Wavelets*, CBMS-NSF Series in Applied Mathematics, 61, SIAM, Philadelphia, (1992).

[5] C.W.J. Granger, and R. Joyeux, *An Introduction to Long-memory Time Series Models and Fractional Differencing*, Journal of Time Series Analysis, **1** (1980) 15–29.

[6] J. R. M. Hosking, *Fractional Differencing*, Biometrika, **68(1)** (1981) 165–176.

[7] M-J Lai, *On the digital filter associated with Daubechies' wavelets*, IEEE Trans. on Signal Processing, **43** (1995) 2203–2205.

[8] S. Mallat, *A theory for multiresolution signal decomposition: The wavelet representation*, IEEE Trans. on Pattern Analysis and Machine Intelligence, **11** (1989) 674–693.

[9] E. McCoy and A. Walden, *Wavelet Analysis and Synthesis of Stationary Long-Memory Processes*, J. Computational and Graphical Statistics, **5(1)** (1996) 26–56.

[10] D.B. Percival, and A.G. Bruce, *Wavelet-Based Approximate maximum likelihood estimates for Trend-Contaminated Fractional Difference Processes*, MathSoft Research Report, **67** (1998).

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

[12] A.H. Tewfik and M. Kim *Correlation structure of the DWT of fractional Brownian motion*, IEEE Trans. Information Theory, **38(2)** (1992) 904–909.

[13] G. W. Wornell, *Signal Processing with Fractals: A Wavelet-Based Approach*, Prentice Hall, Upper Saddle River, New Jersey, (1995).

		d				
L	p	0	0.25	0.50	0.75	1.00
2	0	1.260	1.036	0.896	0.781	0.664
2	1	1.260	1.052	0.906	0.767	0.664
2	2	1.260	1.055	0.908	0.762	0.664
2	5	1.260	1.056	0.912	0.754	0.664
4	0	1.060	0.982	0.921	0.867	0.816
4	1	1.060	1.005	0.960	0.909	0.839
4	2	1.060	1.004	0.963	0.916	0.836
4	5	1.060	1.004	0.963	0.917	0.829
8	0	0.991	0.956	0.923	0.893	0.864
8	1	0.991	0.977	0.965	0.952	0.934
8	2	0.991	0.975	0.963	0.953	0.937
8	5	0.991	0.975	0.964	0.953	0.937
∞	0	0.944	0.931	0.919	0.906	0.894
∞	1	0.944	0.944	0.945	0.944	0.944
∞	2	0.944	0.944	0.944	0.944	0.944
∞	5	0.944	0.944	0.945	0.945	0.945

Table 1: Calculation of $\sigma^2_{d,p}$ for various filter lengths, L, ($L = \infty$ refers to using the ideal wavelet filter $\mathcal{H}_{j,I}(f)$), AR(p) wavelet model ($p = 0$ refers to the white noise model of Craigmile *et al.* (3)) and difference parameter d.

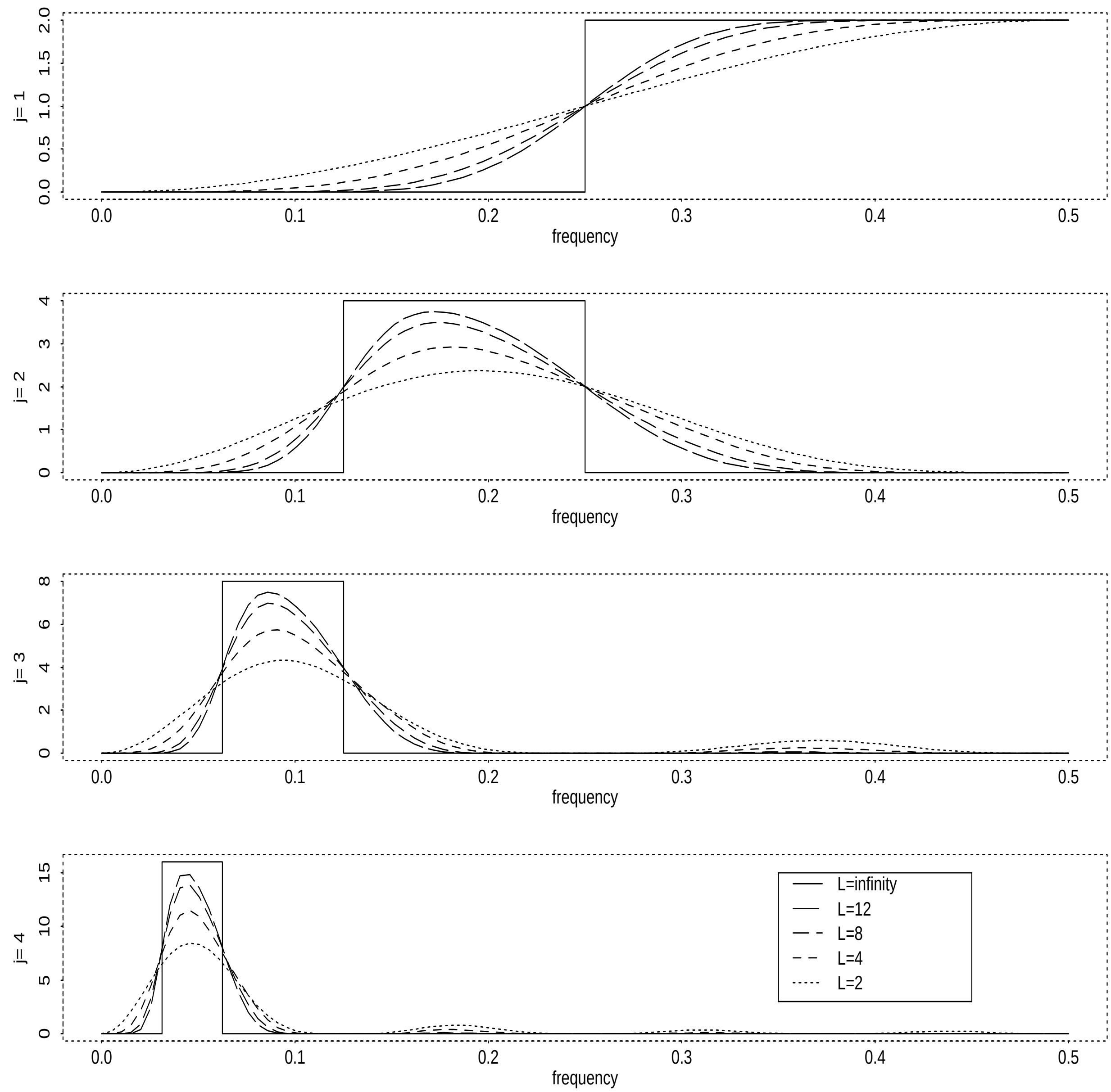

Figure 1: Squared gain functions for various filter lengths, L, and $j = 1, \ldots, 4$ ($L = \infty$ denotes the ideal wavelet filter). For example, with $j = 3$ the side-lobes for $f > \frac{1}{4}$ decrease with increasing L.

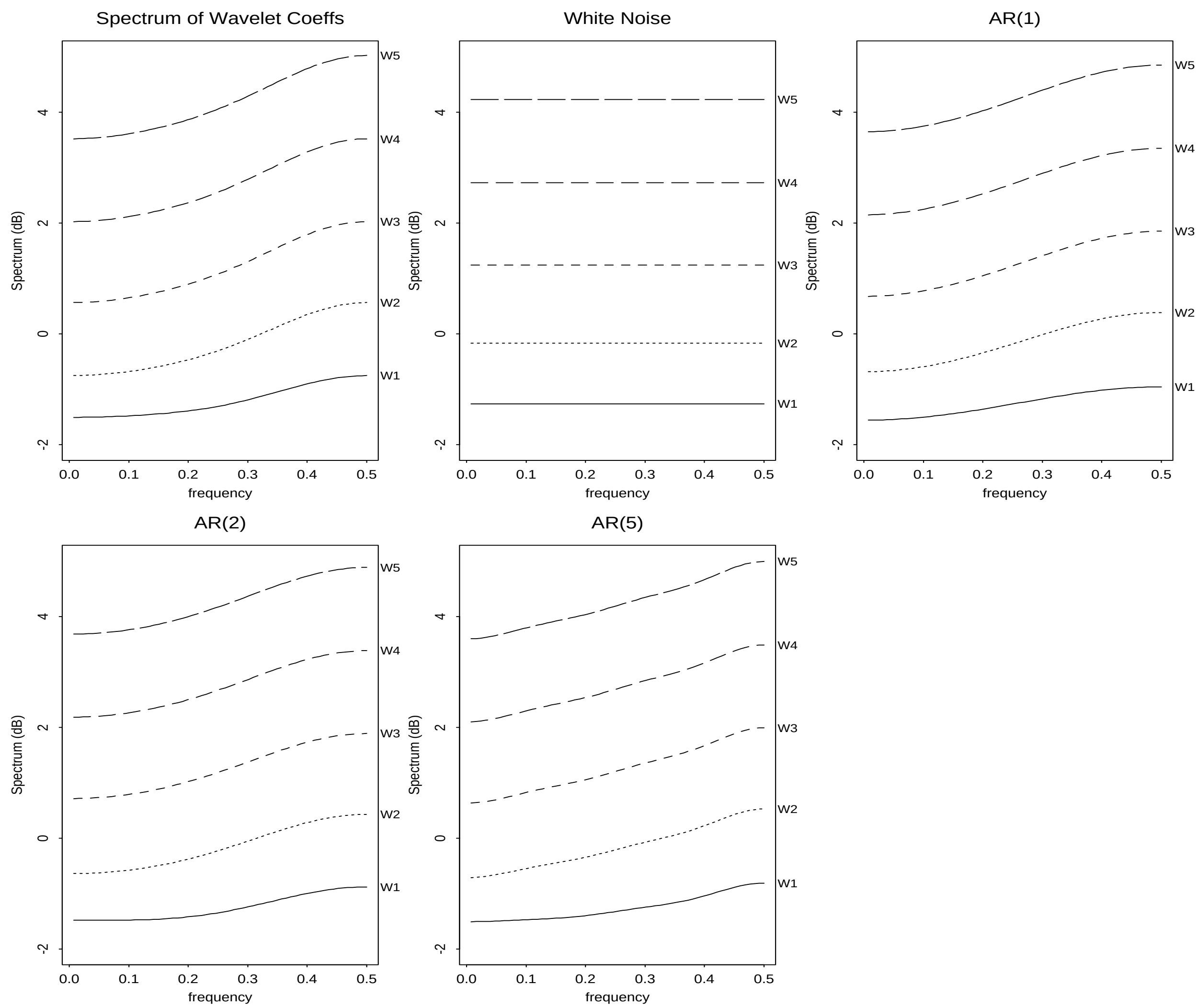

Figure 2: Going from left to right, top to bottom, plots show the SDF of a FD(0.25,1) process (dotted vertical lines indicate the approximate bandpasses for the first five wavelet levels), the SDF of the BI wavelet coefficients with $L = 8$, and the SDF assumed in the white noise and AR(p) models for $p = 1, 2, 5$.