Adaptive wavelet decompositions of stationary time series *†‡

Gustavo Didier
Tulane University

Vladas Pipiras§
University of North Carolina

June 9, 2009

Abstract

A general and flexible framework for the wavelet-based decompositions of stationary time series in discrete time, called Adaptive Wavelet Decompositions (AWDs), is introduced. It is shown that several particular AWDs can be constructed with the aim of providing decomposition (approximation and detail) coefficients that exhibit certain nice statistical properties, where the latter can be chosen based on a range of theoretical or applied considerations.

AWDs make use of a Fast Wavelet Transform-like algorithm whose filters - in contrast with their counterparts in Orthogonal Wavelet Decompositions (OWDs) - may depend on the scale. As with OWDs, this algorithm has good properties such as computational efficiency and invariance to polynomial trends.

A property whose pursuit plays a central role in this work is the decorrelation of the detail coefficients. For many time series models (e.g., FARIMA$(0, \delta, 0)$), the AWD filters can be defined so that the resulting AWD detail coefficients are all (exactly) decorrelated. The corresponding AWDs, called Exact AWDs (EAWDs), are particularly useful in simulation of Gaussian stationary time series if the associated filters have a fast decay. The proposed simulation methods generalize and improve upon existing wavelet-based ones.

AWDs for which the detail coefficients are not exactly decorrelated, but still more decorrelated than those of OWDs, are referred to as approximate AWDs (AAWDs). They can be obtained by truncating EAWD filters, or by adopting some of the existing approaches to modeling the dependence structure of the OWD detail coefficients (e.g., Craigmile et al. (2005)). AAWDs naturally lead to new wavelet-based Maximum Likelihood estimators. The performance of these estimators is investigated through simulations and from some theoretical standpoints. The focus in estimation is also on Gaussian stationary series, though the method is expected to work for non-Gaussian stationary series as well.

1 Introduction

Orthogonal Wavelet Decompositions (OWDs, in short) are useful in several areas of Time Series Analysis. For example, they are used to analyze and to synthesize (Gaussian) long memory time series (Abry, Flandrin, Taqqu and Veitch (2003), Pipiras (2005), Moulines, Roueff and Taqqu (2008)) and in connection to unit roots (Fan and Gençay (2006)). See also a nice monograph on the subject by Percival and Walden (2000), and many more references therein. Other applications, but in continuous time, concern locally stationary time series (Mallat, Papanicolaou and Zhang (1998), Nason, von Sachs and Kroisandt (2000)), multifractal processes (Ossiander and Waymire (2000), Resnick, Samorodnitsky, Gilbert and Willinger (2003), Jaffard, Lashermes and Abry (2006)).

*The second author was supported in part by the NSF grant DMS-0505628.

†*AMS Subject classification.* Primary 60G10, 42C40; secondary 62M10, 62F10.

‡*Keywords and phrases*: stationary time series, wavelets, scaling and wavelet filters, zero moments, adaptive wavelet decompositions, long memory, simulation, maximum likelihood estimation.

§The corresponding author. E-mail: *pipiras@email.unc.edu*.

An appealing property of OWDs in Time Series Analysis is the decorrelation property of detail (wavelet) coefficients. This fact has by now become an integral part of the "folklore" and has been formalized to some degree. The most studied cases are probably those of fractional Brownian motion and long memory time series. Flandrin (1992), Tewfik and Kim (1992), Dijkerman and Mazumdar (1994) studied dependence structure of detail coefficients of fractional Brownian motion. These authors found that the decay of correlations between detail coefficients is controlled by the number of zero moments of the underlying wavelet. Similar findings were also reported for long memory time series. In particular, within scales, the dependence structure of detail coefficients was found to be weak and well-modeled as AR(1) or AR(2) series (see, for example, Craigmile, Guttorp and Percival (2005)). More recently, Craigmile and Percival (2005) showed that detail coefficients across scales become decorrelated asymptotically, as the length of the wavelet filter increases. See also McCoy and Walden (1996), Craigmile (2005), Craigmile et al. (2005), and Vannuci and Corradi (1999), Delbeke (1998) for other studies concerning dependence structure of detail coefficients.

The decorrelation properties described above have been exploited in at least two applications of OWDs:

- simulation, and
- maximum likelihood estimation (MLE)

of typically Gaussian stationary time series. For example, OWD-based simulation can be found in Craigmile (2005), and OWD-based MLE is the subject of Craigmile et al. (2005), Section 9 in Percival and Walden (2000), Jensen (1999). The interest in these OWD-based methods is described in the above references, and often refers to computational efficiency, ease of implementation and invariance to polynomial trends. Two basic simplifying frameworks are used in these studies: either detail coefficients are assumed to be uncorrelated within and across scales, or decorrelation is assumed across scales but dependence is modeled within scale by AR series. Similar simplifying assumptions are also made in other OWD applications, for example, Veitch and Abry (1999). Two exceptions are Moulines, Roueff and Taqqu (2007, 2008), where theoretical results are established without simplifying assumptions.

Simplifications are made because dealing with the exact dependence structure of the detail coefficients is generally quite difficult. More specifically, OWD decomposes a time series $X^0 = \{X^0_n\}_{n\in\mathbb{Z}}$ into approximation coefficients $X^j = \{X^j_n\}_{n\in\mathbb{Z}}$ and detail coefficients $\xi^j = \{\xi^j_n\}_{n\in\mathbb{Z}}$, $j \geq 1$, according to the formulae (the so-called Fast Wavelet Transform or FWT)

$$X^j = \downarrow_2 (\overline{u} * X^{j-1}), \quad \xi^j = \downarrow_2 (\overline{v} * X^{j-1}), \quad j \geq 1, \tag{1.1}$$

where $*$, $\overline{x}$ and $\downarrow_2$ denote, respectively, convolution, time reversion of x and down-sampling by the factor of 2 (see Section 2 for more details) and u, v are suitable sequences, called here OWD (scaling and wavelet) filters. If X^0 is a stationary time series, detail coefficients ξ^j at a fixed scale j, for example, are also stationary and have the spectral density

$$f_{\xi^j}(w) = \frac{1}{2^j} \sum_{n=0}^{2^j-1} f_{X^0}\left(\frac{w+2\pi n}{2^j}\right) \left|\widehat{V}^j\left(\frac{w+2\pi n}{2^j}\right)\right|^2, \tag{1.2}$$

where f_{X^0} is the spectral density of the series X^0, $\widehat{V}^j(w) = \widehat{v}(2^{j-1}w)\prod_{k=1}^{j-1}\widehat{u}(2^{k-1}w)$ and the "hats" indicate the Fourier transform throughout (see Section 3.1 for more details). The orthonormality of OWD ensures that, for any $j \geq 1$,

$$\frac{1}{2^j} \sum_{n=0}^{2^j-1} \left|\widehat{V}^j\left(\frac{w+2\pi n}{2^j}\right)\right|^2 \equiv 1. \tag{1.3}$$

In particular, if $f_{X^0}(w) \equiv \text{const}$ (i.e. X^0 is a white noise), then $f_{\xi^j}(w) \equiv \text{const}$ as well (i.e. ξ^j is a white noise as well). Otherwise, i.e., if $f_{X^0}(w) \not\equiv \text{const}$, the detail coefficients ξ^j have a nontrivial dependence structure.

In this work, we introduce several new wavelet-based decompositions of stationary time series in discrete time. More precisely, we consider:

- Adaptive Wavelet Decompositions (AWDs),
- Exacts AWDs (EAWDs), and
- Approximate AWDs (AAWDs).

In brief, AWDs are wavelet-based decompositions where scaling and wavelet filters in (1.1) can depend on a scale j. These filters can be chosen, in particular, so that the detail coefficients in (1.1) with AWD filters are exactly decorrelated. The resulting decompositions are called EAWDs. When the detail coefficients are not exactly decorrelated, but still more decorrelated than those obtained from OWDs, we refer to the resulting AWDs as AAWDs. To attain decorrelation, EAWD and AAWD filters need to involve the correlation structure of the considered time series. Since this structure is unknown in some applications (such as MLE), these decompositions will not be decorrelating when applied to time series with "misspecified" correlation structure. The more general framework of AWDs also allows one to analyze such situations. We provide next a more detailed description of these new decompositions, followed by a discussion of their applications to simulation and MLE.

AWDs are defined in the same way as OWDs except that OWD filters u, v are now replaced by AWD filters U_d^j, V_d^j which can depend on a scale j and are defined through

$$U_d^j = \overline{c}^j * u, \quad V_d^j = \overline{d}^j * v, \tag{1.4}$$

for some "external" sequences c^j, d^j. The approximation coefficients X^j and the detail coefficients ξ^j of AWD are defined as in (1.1) but by using filters U_d^j, V_d^j instead of u, v. In the definition of AWDs, the sequences c^j, d^j are essentially arbitrary (apart from all quantities being well-defined) and hence AWDs have a general and flexible structure. (In particular, taking $\widehat{c}^j(w) \equiv 1$ and $\widehat{d}^j(w) \equiv 1$ leads to OWDs.) For AWDs, detail coefficients at a fixed scale j, for example, now have the spectral density

$$f_{\xi^j}(w) = \frac{1}{2^j} \sum_{n=0}^{2^j-1} f_{X^0}\left(\frac{w+2\pi n}{2^j}\right) \left|\widehat{D}^j\left(\frac{w+2\pi n}{2^j}\right)\right|^2 \left|\widehat{V}^j\left(\frac{w+2\pi n}{2^j}\right)\right|^2, \tag{1.5}$$

where $\widehat{D}^j(w) = \widehat{d}^j(2^{j-1}w) \prod_{k=1}^{j-1} \widehat{c}^k(2^{k-1}w)$ (cf. (1.2)). AWDs also inherit other nice properties of OWDs such as invariance to polynomial trends.

EAWDs are AWDs whose detail coefficients are exactly decorrelated (within and across scales). It should not be surprising that EAWDs are possible for suitable choices of "external" sequences c^j, d^j. In view of (1.5) and (1.3), for example, detail coefficients of AWD are decorrelated at a fixed scale j as long as

$$f_{X^0}(w)|\widehat{D}^j(w)|^2 = f_{X^0}(w)|\widehat{d}^j(2^{j-1}w)|^2 \prod_{k=1}^{j-1} |\widehat{c}^k(2^{k-1}w)|^2 \equiv 1, \quad j \geq 1. \tag{1.6}$$

A number of sequences c^j, d^j satisfy (1.6). For example, one can take $|\widehat{c}^1(w)|^2 = |\widehat{d}^1(w)|^2 = f_{X^0}(w)^{-1}$, $|\widehat{c}^j(w)|^2 \equiv |\widehat{d}^j(w)|^2 \equiv 1$, $j \geq 2$. But this is obviously not the only possibility. Among various possible EAWDs, we are particularly interested in those whose filters U_d^j, V_d^j are finite or exhibit a *fast decay* to zero. (Why this is of interest is explained below in this section.) We will show that it is possible to construct EAWDs with finite or fast decaying filters for a wide range of time series models, and that this possibility seems particularly suitable in at least two situations of interest, namely,

- long memory, and
- near unit roots.

It is quite intriguing that these are exactly the two situations where OWDs were found particularly useful (see the discussion in the beginning of this section). As will be seen, the number of zero moments of the underlying OWD filters plays here a fundamental role. Moreover, in these decompositions, approximation coefficients at various analysis scales typically turn out to be dependent. It also goes unsaid that in order to attain decorrelation, EAWDs filters U_d^j, V_d^j have to depend on the correlation structure of the series X^0 itself.

Finally, AAWDs refer to, somewhat loosely, AWDs whose filters U_d^j, V_d^j are of finite practical length and whose detail coefficients are expected to be more decorrelated than those obtained from OWDs. Two types of AAWDs are considered in this work:

- EAWD-based AAWDs, and
- AR-based AAWDs.

EAWD-based AAWDs are obtained by truncating EAWD filters U_d^j, V_d^j when the latter are infinite (and have a fast decay). AR-based AAWDs formalize earlier approaches (e.g. Craigmile et al. (2005)) that model OWD detail coefficients using AR models. In this work, for the sake of illustration and understanding, we provide various EAWDs and AAWDs of FARIMA$(0, \delta, 0)$, AR(p) and MA(q) models.

Because of the exact decorrelation property, EAWDs are particularly suitable to generate Gaussian stationary series. The scheme (1.1) with AWD filters (1.4) is called FWT at decomposition. There is also such FWT at reconstruction and that is the scheme which is used for simulation. Having associated EAWD filters with a fast decay allows one to take them of small length in practice. The resulting method is most natural for wavelet-based simulation and improves upon earlier wavelet-based simulation approaches. For FARIMA$(0, \delta, 0)$ series, for example, it is also faster than the popular simulation method based on Circulant Matrix Embedding (Dietrich and Newsam (1997)).

Application to MLE carries quite a different flavor from that to simulation. First, MLE uses FWT at decomposition, not FWT at reconstruction. Second, whereas a particular model for the series to be generated is fixed in simulations, a whole class of models is considered in MLE and the exact model (within the considered class of models) of an analyzed time series is not known in advance. In order to explain basic ideas behind this application, suppose the class of models of interest is that of FARIMA$(0, \delta, 0)$ series. We can construct EAWDs for the latter with fast decaying filters. Moreover, these filters can be truncated in practice to obtain EAWD-based AAWD. Let $\xi^{j,\delta}$ be detail coefficients of an analyzed time series (FARIMA$(0, \delta_0, 0)$ series with a true parameter δ_0) obtained from such AAWD when parameter δ is used. The estimator of the true parameter δ_0, for example, is then defined as

$$\widehat{\delta} = \operatorname*{argmin}_{\delta} \sum_{j=1}^{J} \frac{2^{-j}}{n_j} \sum_{n=1}^{n_j} (\xi_n^{j,\delta})^2, \tag{1.7}$$

where n_j is the number of detail coefficients available at scale j (that is, not affected by the boundary at scale j).

The estimation based on (1.7) works, in fact, quite well. We shall explain reasons for this in detail below (see Section 7). Moreover, we will argue that (1.7) can be viewed as an approximate (Gaussian) ML estimator, and hence that it has the variance of ML estimators in the asymptotic sense. In finite samples, however, at least for sample sizes smaller than $2^{11} = 2,048$, we find that the estimator (1.7) has a slightly larger Mean Squared Error than that of the Whittle approximate ML estimator or, for that matter, the OWD estimator. This effect for smaller sample sizes will be explained by making an argument in the Fourier domain. Despite its slight inferiority in smaller samples, the proposed

estimator has an interesting structure, arises naturally in the AWD framework, and sheds light on various relations between Fourier and wavelet methods. The focus in estimation is also on Gaussian stationary series, though the method is expected to work for non-Gaussian stationary series as well.

In its approach, this study is also closest to our parallel work on AWDs in continuous time in Didier and Pipiras (2008). See also Meyer et al. (1999) and a series of papers by Unser and Blu (2007), Blu and Unser (2007) and references therein. Despite some similarities, however, the focus and content of this work are very different from those in Didier and Pipiras (2008).

The rest of the paper is organized as follows. In Section 2, we gather some basic notions and facts on time series and wavelets that will be used throughout the paper. In Section 3, we introduce and examine AWDs of stationary time series, and their particular variants EAWDs and AAWDs. Examples are considered in Sections 4 and 5. Applications of AWDs to simulation and MLE can be found in Sections 6 and 7. Proofs have been moved to Section 8, and all tables and figures can be found at the end of the paper.

2 Preliminaries on time series and wavelets

We focus throughout on stationary time series $X = \{X_n\}_{n\in\mathbb{Z}}$ in discrete time. Stationarity refers to 2^{nd} order (wide-sense) stationarity, that is, the case when, for any $h \in \mathbb{Z}$,

$$EX_{k+h}X_h = EX_kX_0 =: r(k), \quad k \in \mathbb{Z}, \tag{2.1}$$

where r is the autocovariance function. We suppose, in addition, that a time series X is *Gaussian*. (In this case, decorrelation is equivalent to independence.) This assumption is not restrictive. Since the law of a Gaussian time series is determined by second moments, our arguments can be based only on considerations of the second moments. After removing Gaussianity, the same arguments then apply to 2^{nd} order stationary time series. In our applications, however, we consider only Gaussian time series.

We will also work only with linear time series

$$X_n = \sum_{k=-\infty}^{\infty} a_k\epsilon_{n-k} = (a * \epsilon)_n, \quad n \in \mathbb{Z}, \tag{2.2}$$

where $a = \{a_k\} \in l^2(\mathbb{Z})$ and $*$ denotes the usual convolution. In the Gaussian case, $\epsilon = \{\epsilon_n\}$ are independent, $\mathcal{N}(0,1)$ random variables. We will refer to such ϵ as a *Gaussian white noise* (sequence). One of the main tools we will use is the spectral representation of X in (2.2) (see e.g. Brockwell and Davis (1991)):

$$X_n = \int_0^{2\pi} e^{inw}dW(w) = \int_0^{2\pi} e^{inw}\widehat{a}(w)dZ(w), \quad n \in \mathbb{Z}, \tag{2.3}$$

where $W(w)$, $w \in (0, 2\pi)$, is a Gaussian, orthogonal (independent) increment, complex-valued process such that $EdW(w)\overline{dW(w')} = |\widehat{a}(w)|^2 dw 1_{\{w=w'\}}/2\pi$, $Z(w)$, $w \in (0, 2\pi)$, is a Gaussian, orthogonal (independent) increment process such that $EdZ(w)\overline{dZ(w')} = dw 1_{\{w=w'\}}/2\pi$, and

$$\widehat{a}(w) = \sum_{k=-\infty}^{\infty} a_k e^{-ikw}, \quad w \in (0, 2\pi), \tag{2.4}$$

is the discrete Fourier transform of a sequence $a \in l^2(\mathbb{Z})$. The function $|\widehat{a}(w)|^2/2\pi$ is known as a spectral density of X. Observe also that $r = a * \overline{a}$, $\widehat{r}(w) = |\widehat{a}(w)|^2$, where $\{\overline{x}_k\} = \{x_{-k}\}$ stands for time reversal of a sequence $\{x_k\}$.

In regard to wavelets, since we work in discrete time, we will use the so-called OWD scaling and wavelet filters, also called Conjugate Mirror Filters, associated with an orthogonal Multiresolution

Analysis (MRA). See, for example, Mallat (1998). OWD filters are made up of a low pass filter $u = \{u_n\}$ and a high pass filter $v = \{v_n\}$ satisfying a number of properties. In particular, for any $w \in \mathbb{R}$,

$$|\widehat{u}(w)|^2 + |\widehat{u}(w+\pi)|^2 = 2, \tag{2.5}$$

$$\widehat{v}(w) = e^{-iw}\overline{\widehat{u}(w+\pi)} \tag{2.6}$$

and hence

$$|\widehat{v}(w)|^2 + |\widehat{v}(w+\pi)|^2 = 2, \tag{2.7}$$

$$\widehat{u}(w)\overline{\widehat{v}(w)} + \widehat{u}(w+\pi)\overline{\widehat{v}(w+\pi)} = 0. \tag{2.8}$$

Popular OWD filters are those of Daubechies with N zero moments, $N \geq 1$. For fixed N, both low and high pass Daubechies filters are of finite length $2N$. Moreover, for any OWD filters of finite length and N zero moments, it is also known (e.g. Mallat (1998), p. 241) that

$$\widehat{u}(w) = (1+e^{-iw})^N \widehat{u}_{0,N}(w), \quad \widehat{v}(w) = (1-e^{-iw})^N \widehat{v}_{0,N}(w), \tag{2.9}$$

with $u_{0,N}, v_{0,N}$ of finite length as well.

OWD filters u and v appear in the (orthogonal) Fast Wavelet Transform (FWT) of a deterministic sequence $x = \{x_n\}$. FWT relates decomposition coefficients across scales. Setting $a_0 = x$, at the decomposition step, one defines the approximation and detail coefficients as

$$a_j = \downarrow_2 (\overline{u} * a_{j-1}), \quad d_j = \downarrow_2 (\overline{v} * a_{j-1}), \quad j = 1, 2, \ldots, \tag{2.10}$$

where $(\downarrow_2 x)_k = x_{2k}$ is the downsampling (decimation) by factor 2 operator. At the reconstruction step, one has

$$a_j = u* \uparrow_2 a_{j+1} + v* \uparrow_2 d_{j+1}, \quad j = 0, 1, \ldots, \tag{2.11}$$

where $(\uparrow_2 x)_k = x_{k/2} 1_{\{\text{even } k\}} + 0 1_{\{\text{odd } k\}}$ is the upsampling by factor 2 operator. One can easily verify that

$$\widehat{(\downarrow_2 x)}(w) = \frac{1}{2}\Big(\widehat{x}\Big(\frac{w}{2}\Big) + \widehat{x}\Big(\frac{w}{2}+\pi\Big)\Big), \quad \widehat{(\uparrow_2 x)}(w) = \widehat{x}(2w). \tag{2.12}$$

3 Adaptive Wavelet Decompositions

Definition and basic properties of Adaptive Wavelet Decompositions can be found in Section 3.1. Sections 3.2 and 3.3 introduce what we call Exact and Approximate Adaptive Wavelet Decompositions.

3.1 Definition and basic properties of AWDs

Let $X^0 = a^0 * \epsilon^0$ be a Gaussian, stationary time series with $a^0 \in l^2(\mathbb{Z})$ and a Gaussian white noise ϵ^0. For $j \geq 1$, let also

$$c^j, d^j \in l^2(\mathbb{Z}) \tag{3.1}$$

and

$$\widehat{U}_d^j(w) = \overline{\widehat{c^j}(w)}\widehat{u}(w), \quad \widehat{V}_d^j(w) = \overline{\widehat{d^j}(w)}\widehat{v}(w), \tag{3.2}$$

where u, v are OWD filters. Similarly, for $j \geq 0$, set

$$\widehat{U}_r^j(w) = \widehat{c}^{j+1}(w)^{-1}\widehat{u}(w), \quad \widehat{V}_r^j(w) = \widehat{d}^{j+1}(w)^{-1}\widehat{v}(w). \tag{3.3}$$

For $j \geq 1$, define

$$X^j = \downarrow_2 (\overline{U}_d^j * X^{j-1}), \quad \xi^j = \downarrow_2 (\overline{V}_d^j * X^{j-1}). \tag{3.4}$$

By analogy to the OWD relation (2.11), one expects that, for $j \geq 1$,

$$X^{j-1} = U_r^{j-1} * \uparrow_2 X^j + V_r^{j-1} * \uparrow_2 \xi^j, \tag{3.5}$$

when the following "reconstruction" identity holds

$$\frac{\widehat{c}^j(w+\pi)}{\widehat{c}^j(w)} = \frac{\widehat{d}^j(w+\pi)}{\widehat{d}^j(w)}. \tag{3.6}$$

Definition 3.1 The decomposition of a stationary time series $X = X^0$ into the collection of series X^j, ξ^j, $j \geq 1$, in (3.4) will be called *Adaptive Wavelet Decomposition* (*AWD*, in short) of a stationary time series X. We will refer to X^j as *approximation coefficients* (approximations, in short), to ξ^j as *detail coefficients* (details, in short) and to $U_d^j, V_d^j, U_r^j, V_r^j$ as AWD (decomposition and reconstruction) *filters*.

The next theorem provides simple technical conditions for AWD of a stationary time series to be well-defined through the FWT relation (3.4) at decomposition, and for the resulting AWD coefficients to satisfy the FWT relation (3.5) at reconstruction.

Theorem 3.1 *The following assertions hold:*

(i) (Decomposition step) If $\widehat{U}_d^j, \widehat{V}_d^j \in L^2(0, 2\pi)$ *and the corresponding filters*

$$U_d^j, V_d^j \in l^1(\mathbb{Z}), \quad j \geq 1, \tag{3.7}$$

then the sequences X^j, ξ^j *are well-defined series through (3.4). Moreover, the collection* $\{X^J, \xi^j, 1 \leq j \leq J\}$ *consists of Gaussian stationary series, and is jointly Gaussian.*

(ii) (Reconstruction step) If the "reconstruction" identity (3.6) holds and, in addition, $\widehat{U}_r^j, \widehat{V}_r^j \in L^2(0, 2\pi)$ *and the corresponding filters*

$$U_r^j, V_r^j \in l^1(\mathbb{Z}), \quad j \geq 0, \tag{3.8}$$

then the relation (3.5) holds.

Remark 3.1 Another interpretation of Theorem 3.1 is to say that $(U_d^j, V_d^j, U_r^{j-1}, V_r^{j-1})$ forms a perfect reconstruction filter bank (see, for example, Mallat (1991), p. 259). Indeed, by Theorem 7.8 in Mallat (1998), this is so if and only if

$$\overline{\widehat{U}_d^j(w)}\widehat{U}_r^{j-1}(w+\pi) + \overline{\widehat{V}_d^j(w)}\widehat{V}_r^{j-1}(w+\pi) = 0,$$

$$\overline{\widehat{U}_d^j(w)}\widehat{U}_r^{j-1}(w) + \overline{\widehat{V}_d^j(w)}\widehat{V}_r^{j-1}(w) = 2.$$

The left-hand side of the first relation is

$$\frac{\widehat{c}^j(w)}{\widehat{c}^j(w+\pi)}\overline{\widehat{u}(w)}\widehat{u}(w+\pi) + \frac{\widehat{d}^j(w)}{\widehat{d}^j(w+\pi)}\overline{\widehat{v}(w)}\widehat{v}(w+\pi)$$

$$= \frac{\widehat{c}^j(w)}{\widehat{c}^j(w+\pi)}\left(\overline{\widehat{u}(w)}\widehat{u}(w+\pi) + \overline{\widehat{v}(w)}\widehat{v}(w+\pi)\right) = 0,$$

using the "reconstruction" identity (3.6) and the perfect reconstruction property of OWD filters. The second relation can be proved similarly. Note also that Theorem 3.1 is not a consequence of perfect reconstruction because the filtering scheme involves (random) time series.

The term "detail coefficients" in Definition 3.1, in particular, is motivated by the fact that the filter v in (3.2) and (3.3) is a high pass filter. An important related property of any AWD is that AWD detail coefficients ξ^j are invariant to polynomial trends up to the order of the number of zero moments of the underlying orthogonal MRA. An analogous fact is well-known for OWDs. (In discrete time, this follows immediately from Theorem 7.4, (iv), in Mallat (1998).) We show that it holds here as well (see Section 8 for a proof).

Theorem 3.2 *Suppose that the underlying orthogonal MRA has N zero moments with factorization (2.9). Let $p_n = p(n)$, $n \in \mathbb{Z}$, where a polynomial p is of degree $D < N$. Consider AWD with decomposition filters U_d^j, V_d^j such that $|U_{d,n}^j|, |V_{d,n}^j| \leq C_j|n|^{-D-2}$, where C_j is a constant. Then, for any $j \geq 1$,*

$$\xi^j(p) = 0, \tag{3.9}$$

where $\xi^j(p)$ are details in AWD when applied to the polynomial p.

By its definition, AWDs are OWDs when

$$\widehat{c}^j(w) \equiv 1, \quad \widehat{d}^j(w) \equiv 1. \tag{3.10}$$

In the context of this work, however, we will be interested in those AWDs whose detail coefficients are more decorrelated than those obtained through OWDs. In this regard, it is useful to have a result characterizing covariance structures of X^j and ξ^j, and this is established next. The corresponding result for OWDs is well-known (see, for example, Percival and Walden (2000), p. 348). We shall use the following notation, which stands in analogy to the corresponding notation for OWDs. For $j \geq 1$, let

$$\widehat{C}^j(w) = \prod_{k=1}^{j} \widehat{c}^k(2^{k-1}w), \quad \widehat{D}^j(w) = \widehat{d}^j(2^{j-1}w) \prod_{k=1}^{j-1} \widehat{c}^k(2^{k-1}w) \tag{3.11}$$

and

$$\widehat{U}^j(w) = \prod_{k=1}^{j} \widehat{u}(2^{k-1}w), \quad \widehat{V}^j(w) = \widehat{v}(2^{j-1}w) \prod_{k=1}^{j-1} \widehat{u}(2^{k-1}w), \tag{3.12}$$

where the last products in (3.11) and (3.12) are interpreted as 1 for $j = 1$. (The difference in notation (3.12) and (3.2) is the subindex d.) Let also

$$\widehat{F}^j(w) = \widehat{C}^j(w)\overline{\widehat{U}^j(w)}, \quad \widehat{G}^j(w) = \widehat{D}^j(w)\overline{\widehat{V}^j(w)}. \tag{3.13}$$

Theorem 3.3 *Under the assumptions in part (i) of Theorem 3.1, the collection $\{X^J, \xi^j, 1 \leq j \leq J\}$ has the following covariance structure: for $1 \leq j, j' \leq J$, $n, n' \in \mathbb{Z}$, $k \in \mathbb{Z}$,*

$$EX_n^J X_{n+k}^J = \int_0^{2\pi} e^{i2^J kw} |\widehat{a}^0(w)|^2 |\widehat{F}^J(w)|^2 \frac{dw}{2\pi}, \tag{3.14}$$

$$EX_n^J \xi_{n'}^{j'} = \int_0^{2\pi} e^{i(2^J n - 2^{j'} n')w} |\widehat{a}^0(w)|^2 \widehat{F}^J(w) \overline{\widehat{G}^{j'}(w)} \frac{dw}{2\pi}, \tag{3.15}$$

$$E\xi_n^j \xi_{n'}^{j'} = \int_0^{2\pi} e^{i(2^j n - 2^{j'} n')w} |\widehat{a}^0(w)|^2 \widehat{G}^j(w) \overline{\widehat{G}^{j'}(w)} \frac{dw}{2\pi}. \tag{3.16}$$

In particular,

$$E\xi_n^j \xi_{n+k}^j = \int_0^{2\pi} e^{i2^j kw} |\widehat{a}^0(w)|^2 |\widehat{G}^j(w)|^2 \frac{dw}{2\pi}. \tag{3.17}$$

Observe that, since

$$\int_0^{2\pi} e^{i2^j kw} |\widehat{a}^0(w)|^2 |\widehat{G}^j(w)|^2 \frac{dw}{2\pi} = \int_0^{2\pi} e^{ikw} f_{\xi^j}(w) dw \tag{3.18}$$

with

$$f_{\xi^j}(w) = \frac{1}{2\pi 2^j} \sum_{n=0}^{2^j-1} \left| \widehat{a}^0 \left(\frac{w+2\pi n}{2^j} \right) \right|^2 \left| \widehat{G}^j \left(\frac{w+2\pi n}{2^j} \right) \right|^2, \tag{3.19}$$

relation (3.17) shows that ξ^j is a (Gaussian) stationary series with the spectral density f_{ξ^j} in (3.19). Similarly, X^J is also a (Gaussian) stationary series with the spectral density

$$f_{X^J}(w) = \frac{1}{2\pi 2^J} \sum_{n=0}^{2^J-1} \left| \widehat{a}^0 \left(\frac{w+2\pi n}{2^j} \right) \right|^2 \left| \widehat{F}^J \left(\frac{w+2\pi n}{2^J} \right) \right|^2, \tag{3.20}$$

AWD is, in general, not an orthogonal transformation. Under suitable assumptions, however, there is a simple formula for energy transformation under AWD (that is, a formula for the sum of the second moments of AWD detail coefficients across scales – see (3.23) below). These assumptions are satisfied, in particular, by Exact AWDs introduced below, and play a key role in estimation (Section 7).

Theorem 3.4 *Under the assumptions in (i) of Theorem 3.1, suppose that the following "energy transformation" condition holds: for $j \geq 1$ and some function $\widehat{b}(w)$,*

$$|\widehat{D}^j(w)|^2 = |\widehat{d}^j(2^{j-1}w)|^2 \prod_{k=1}^{j-1} |\widehat{c}^k(2^{k-1}w)|^2 =: \frac{1}{|\widehat{b}(w)|^2}. \tag{3.21}$$

Then,

$$\sum_{j=1}^{\infty} 2^{-j} E(\xi_n^j)^2 = \frac{1}{2\pi} \int_0^{2\pi} \frac{|\widehat{a}^0(w)|^2}{|\widehat{b}(w)|^2} \left\{ \sum_{j=1}^{\infty} 2^{-j} |\widehat{V}^j(w)|^2 \right\} dw \tag{3.22}$$

$$= \frac{1}{2\pi} \int_0^{2\pi} \frac{|\widehat{a}^0(w)|^2}{|\widehat{b}(w)|^2} dw \tag{3.23}$$

and, for any $J \geq 1$,

$$\sum_{j=1}^{J} 2^{-j} E(\xi_n^j)^2 = \frac{1}{2\pi} \int_0^{2\pi} \frac{|\widehat{a}^0(w)|^2}{|\widehat{b}(w)|^2} \left\{ \sum_{j=1}^{J} 2^{-j} |\widehat{V}^j(w)|^2 \right\} dw. \tag{3.24}$$

Remark 3.2 The factor 2^{-j} in (3.22) and (3.24) is for proper normalization. It would not be needed if $= 2$ is replaced by $= 1$ in (2.5), which is another common convention in the wavelet literature.

3.2 Exact AWDs

A particular case of AWDs is when detail coefficients ξ^j in (3.4) are exactly decorrelated (within and across scales). We shall refer to this case as that of *Exact Adaptive Wavelet Decompositions* (*EAWDs*, in short). Decorrelation can be achieved for a wide range of filters c^j, d^j in (3.1). In this work, we consider the following particular situation, which turns out to be quite interesting for a range of time series models.

Consider sequences $a^j \in l^2(\mathbb{Z})$, $j \geq 1$, and set

$$\widehat{c}^j(w) = \frac{\widehat{a}^j(2w)}{\widehat{a}^{j-1}(w)}, \quad \widehat{d}^j(w) = \frac{1}{\widehat{a}^{j-1}(w)}. \tag{3.25}$$

Note that, because of the argument $2w$ of $\widehat{a}^j$ and the periodicity with period 2π of $\widehat{a}^j$, the filters c^j, d^j in (3.25) satisfy the "reconstruction" identity (3.6). AWDs with the choice (3.25) have properties stated in the following result.

Theorem 3.5 *Suppose the filters $U_d^j, V_d^j, U_r^j, V_r^j$ defined through (3.2), (3.3) and (3.25) satisfy the conditions of Theorem 3.1. Consider approximation and detail coefficients X^j and ξ^j defined by (3.4). Then, for $J \geq 1$,*

$$X^J = a^J * \epsilon^J \tag{3.26}$$

with a Gaussian white noise ϵ^J, and ξ^j, $j \geq 1$, are independent, Gaussian white noise sequences. Moreover, ϵ^J (and hence X^J) and ξ^j, $1 \leq j \leq J$, are independent.

Though (3.25) is a particular case of EAWDs, it is still quite flexible in the choice of moving average filters a^j, and hence the corresponding time series X^j. In fact, EAWDs can be defined for many different choices of a^j's according to the properties one wishes to obtain. The latter can be suggested by an application at hand or other considerations, for example,

(a) X^J and ξ^j, $1 \leq j \leq J$, consisting of uncorrelated (independent) variables,

(b) $U_d^j, V_d^j, U_r^j, V_r^j$ having a fast decay, or

(c) X^j being a natural approximation to X^0 at scale 2^j.

The property (a) is important in Signal Processing, as it is typically associated with optimality in coding. For this work, we were motivated by (b). In regard to (c), one natural approximation of a series X^0 at scale 2^j is

$$X^j = \{X^0_{2^j k}\}_{k \in \mathbb{Z}}. \tag{3.27}$$

In particular, taking $\widehat{a}^0(w) = \widehat{a}(w)$ as in the spectral representation (2.3) of X^0, the spectral representation of X^j is defined through the function

$$\widehat{a}^j(w) = \frac{1}{2}\Big(\widehat{a}^{j-1}\Big(\frac{w}{2}\Big) + \widehat{a}^{j-1}\Big(\frac{w}{2} + \pi\Big)\Big). \tag{3.28}$$

See Remark 5.1 in Section 5 below for further discussion on (c).

Remark 3.3 Note that $\widehat{c}^j$ and $\widehat{d}^j$ in (3.25) satisfy the "energy transformation" condition (3.21) with $\widehat{b}(w) = \widehat{a}^0(w)$. This fact plays a central role in estimation based on EAWDs (see Section 7).

3.3 Approximate AWDs

Approximate Adaptive Wavelet Decompositions (*AAWDs*, in short) will refer to, somewhat loosely, AWDs which are not EAWDs, whose filters $U_d^j, V_d^j, U_r^j, V_r^j$ are of finite practical length and whose detail coefficients are expected to be more decorrelated than those obtained from OWDs. We shall focus on two types of AAWDs:

- EAWD-based AAWDs, and
- AR-based AAWDs.

EAWD-based AAWDs are defined by truncating EAWD filters when the latter are infinite and have a fast decay. We postpone a further discussion of EAWD-based AAWDs to examples found in Sections 4 and 5.

AR-based AAWDs are motivated by AR modeling of OWD detail coefficients found in, for example, Craigmile et al. (2005). The basic idea here is the following. As mentioned in Section 1, with many time series models of interest, OWD detail coefficients are correlated mostly at the first few lags and within scales. With AR-based AAWDs, the filters c^j, d^j in (3.1) are chosen such that the corresponding detail coefficients are essentially the residuals in AR(1) models of OWD coefficients. Hence, these new AAWD detail coefficients are expected to be more decorrelated than the original OWD detail coefficients. This approach is detailed in the next example.

Example 3.1 (*AR-based AAWDs*) Take

$$\widehat{c^j}(w) \equiv 1, \quad \widehat{d^j}(w) = \widehat{b^j}(2w) = \sigma_j(1 - b_j e^{-i2w}) \tag{3.29}$$

in (3.2) and (3.3) of AWD. AR refers here to the autoregressive form of the filters b^j in (3.29). Note that, in this case, X^j are the approximation coefficients obtained through OWD of X^0, and $\xi^j = b^j * W^j$ or

$$\xi_n^j = \sigma_j(W_n^j - b_j W_{n-1}^j), \tag{3.30}$$

where W^j are the detail coefficients obtained through OWD of X^0.

Since in AAWD we want the detail coefficients ξ^j to be more decorrelated than the coefficients W^j, the parameters b_j, σ_j in (3.29) have to be chosen based on the correlation structure of the underlying time series X^0. One way this can be done is the following (adapting the approach used in Craigmile et al. (2005)). The relation (3.30) can be rewritten as

$$W_n^j = b_j W_{n-1}^j + \frac{1}{\sigma_j}\xi_n^j. \tag{3.31}$$

The covariance structure of W_n^j is given by

$$r_{W^j}(k) = EW_n^j W_{n+k}^j = \int_0^{2\pi} e^{ikw} f_{W^j}(w)dw, \tag{3.32}$$

where, from (3.19),

$$f_{W^j}(w) = \frac{1}{2\pi 2^j}\sum_{n=0}^{2^j-1}\left|\widehat{a}^0\left(\frac{w+2\pi n}{2^j}\right)\right|^2\left|\widehat{V}^j\left(\frac{w+2\pi n}{2^j}\right)\right|^2 \tag{3.33}$$

is the spectral density of W^j. Since we expect ξ^j to be approximately white noise, the coefficients b_j, σ_j in (3.31) could be chosen based on the correlation structure of W^j, namely,

$$r_{W^j}(0) = \frac{1}{\sigma_j^2(1-b_j^2)}, \quad r_{W^j}(1) = \frac{b_j}{\sigma_j^2(1-b_j^2)} \tag{3.34}$$

or

$$b_j = \frac{r_{W^j}(1)}{r_{W^j}(0)}, \quad \frac{1}{\sigma_j^2} = r_{W^j}(0)(1-b_j^2) \tag{3.35}$$

(cf. (11) in Craigmile et al. (2005)).

Remark 3.4 Unlike EAWDs, which involve the exact correlation structure of the series X^0, AR-based AAWDs in Example 3.1 could be considered more model-free. In particular, they can be implemented in practice by replacing $r_{W^j}(0), r_{W^j}(1)$ in (3.35) by their sample counterparts or, equivalently, by fitting AR(1) model to detail coefficients obtained through OWD. Let us also add that the approach of Example 3.1 could be extended to AR(p) models with higher order p by following the corresponding AR modeling of OWD coefficients as in Craigmile (2000).

4 Examples of AWDs: FARIMA$(0, \delta, 0)$ series

We construct here particular AWDs for (Gaussian) FARIMA(0,δ,0) time series with $\delta \in (-1/2, 1/2)$ ($\delta \neq 0$). These are series $X = a * \epsilon$ with a (Gaussian) white noise ϵ and

$$\widehat{a}(w) = \sigma(1 - e^{-iw})^{-\delta}, \tag{4.1}$$

where $\sigma > 0$ (see, for example, Brockwell and Davis (1991), p. 520, or Beran (1994)). The case $\delta \in (0, 1/2)$ corresponds to the so-called long memory, and is generally considered more difficult to deal with. Exact AWDs are studied in Section 4.1 and a finer analysis of the resulting fractional filters can be found in Section 4.2. Approximate AWDs are considered in Section 4.3.

4.1 Exact AWDs

We introduce here exact AWDs of the form (3.25) where we would like the associated filters $U_d^j, V_d^j, U_r^j, V_r^j$ to decay to zero fast. Consider EAWDs with the choice of

$$\widehat{a}^j(w) = \widehat{a}(w), \quad j \geq 1, \tag{4.2}$$

in (3.25), where $\widehat{a}(w)$ is given by (4.1) and we assume for simplicity that $\sigma = 1$. Observe that the filters entering into (3.2) and (3.3) are

$$\widehat{c}(w) \equiv \widehat{c}^j(w) = \frac{\widehat{a}(2w)}{\widehat{a}(w)} = \frac{(1 - e^{-i2w})^{-\delta}}{(1 - e^{-iw})^{-\delta}} = (1 + e^{-iw})^{-\delta}, \tag{4.3}$$

$$\widehat{d}(w) \equiv \widehat{d}^j(w) = \frac{1}{\widehat{a}(w)} = (1 - e^{-iw})^{\delta} \tag{4.4}$$

and

$$\widehat{c}(w)^{-1} = (1 + e^{-iw})^{\delta}, \quad \widehat{d}(w)^{-1} = (1 - e^{-iw})^{-\delta}. \tag{4.5}$$

They have the general forms (with $s = \pm\delta$)

$$(1 + e^{-iw})^s = \sum_{k=0}^{\infty} f_k^{(s)} e^{-iwk}, \quad (1 - e^{-iw})^s = \sum_{k=0}^{\infty} g_k^{(s)} e^{-iwk} \tag{4.6}$$

and, in fact, decay extremely slowly when $s \in (-1/2, 1/2) \setminus \{0\}$: by using Stirling's formula, one can show that, as $k \to \infty$,

$$f_k^{(s)} \sim (-1)^k \frac{k^{-s-1}}{\Gamma(-s)}, \quad g_k^{(s)} \sim \frac{k^{-s-1}}{\Gamma(-s)}. \tag{4.7}$$

(For example, when $s \in (-1/2, 0)$, these filters are not even absolutely summable.)

It is therefore quite surprising that, in fact, the resulting filters $U_d^j \equiv U_d, V_d^j \equiv V_d, U_r^j \equiv U_r, V_r^j \equiv V_r$ may decay to 0 very rapidly. As mentioned in Section 1, this results from the number of zero moments of the underlying orthogonal MRA. Letting N denote the number of zero moments and using (2.9), observe that

$$\widehat{U}_d(w) = \overline{(1 + e^{-iw})^{-\delta+N}} \widehat{u}_{0,N}(w), \quad \widehat{V}_d(w) = \overline{(1 - e^{-iw})^{\delta+N}} \widehat{v}_{0,N}(w), \tag{4.8}$$

$$\widehat{U}_r(w) = (1 + e^{-iw})^{\delta+N} \widehat{u}_{0,N}(w), \quad \widehat{V}_r(w) = (1 - e^{-iw})^{-\delta+N} \widehat{v}_{0,N}(w). \tag{4.9}$$

For example, by (4.7),

$$(1 + e^{-iw})^{\delta+N} = \sum_{k=0}^{\infty} f_k^{(\delta+N)} e^{-iwk} \quad \text{with} \quad f_k^{(\delta+N)} \sim (-1)^k \frac{k^{-\delta-N-1}}{\Gamma(-\delta - N)}, \tag{4.10}$$

as $k \to \infty$. Comparing (4.10) and (4.7) with $s = \delta$, we see that these filters now decay rapidly when N is large.

The latter observation by itself does not show that the resulting filters U_d, V_d, U_r, V_r in (4.8) and (4.9) decay faster as N increases because $u_{0,N}$ and $v_{0,N}$ also grow in size (not length). We address this and other issues in more detail in the next section.

4.2 Finer analysis of fractional filters

We examine here in greater detail the fractional filters U_d, V_d, U_r, V_r in (4.8) and (4.9). In Table 1, we provide lengths of filters U_r, V_r truncated at a priori specified cutoff levels ϵ and the underlying Daubechies OWD filters for various choices of zero moments N. We choose $\delta = 0.4$. More precisely, the length of the truncated filter U_r is computed as follows. Observe that

$$|U_{r,n}| \le \sum_{k=0}^{N-1} |u_{0,N,k}||f_{n-k}^{(\delta+N)}|, \tag{4.11}$$

where $U_{r,n}$ denotes the nth element of U_r and similarly for $u_{0,N,n}$ representing $u_{0,N}$. (We also used the fact that $u_{0,N}$ is of length N.) The right-hand side of (4.11) decreases monotonically for $n \ge N-1$ and the length of a truncated filter U_r is chosen as the smallest $n = n_0 + 1$ ($n_0 \ge N-1$) for which the right-hand side of (4.11) is smaller than ϵ.

Remark 4.1 In fact, determining the filter length based on the right-hand side of (4.11) is essentially equivalent to finding the smallest n_0+1 ($n_0 \ge N-1$) such that $|U_{r,n_0}| \le \epsilon$. This is so because the sign of $f_n^{(\delta+N)}$ eventually oscillates, and the sign of $u_{0,N,n}$ is also oscillating (see Table 6.2 in Daubechies (1992), p. 196). The same could be said about the filter V_r: the sign of $g_n^{(-\delta+N)}$ is eventually the same, and the sign of $v_{0,N,n}$ is always the same.

Remark 4.2 The faster decay in (4.10) has also the following simple explanation that is useful more generally. According to (4.6)–(4.7), the elements $f_k^{(s)}$ of $(1+e^{-iw})^s$ decay as

$$f_k^{(s)} \sim (-1)^k \frac{k^{-s-1}}{\Gamma(-s)}.$$

Application of the filter $(1+e^{-iw})^N$ to $(1+e^{-iw})^s$ corresponds to taking sums in blocks of size N. Since $f_k^{(s)}$ oscillates and decays, the sums will become smaller. A similar explanation with difference instead of sums applies to the elements $g_k^{(s)}$ of $(1-e^{-iw})^s$.

Remark 4.3 Because of the explanation provided in Remark 4.2, we expect that analogous fractional filters have fast decay for long memory time series other than FARIMA series given by (4.1), for example, for the series characterized by $\widehat{a}(w) = w^{-\delta}$. A note of caution here is that our MLE procedure in Section 7 below assumes, in particular, that the filter $a = \{a_n\}_{n\in\mathbb{Z}}$ is causal (that is, $a_n = 0$ for $n \le 0$). This is not the case for the filter $\widehat{a}(w) = w^{-\delta}$ mentioned above.

Table 1 shows, in particular, that increasing N substantially reduces the lengths of truncated fractional filters at small levels of cutoff ϵ. This is relevant in simulations discussed in Section 6: from a practical perspective, say even for cutoff $\epsilon = 10^{-13}$, the related reconstruction filters U_r, V_r can be chosen of a fairly short length. Note also that Table 1 provides the lengths of the truncated filters U_d and V_d which are exactly those of V_r and U_r, respectively. The latter fact is easy to understand. For example, by (7.58) in Mallat (1998), p. 236, one has (2.6) and hence

$$\widehat{v}_{0,N}(w) = (-1)^N e^{i(N-1)w} \overline{\widehat{u}_{0,N}(w+\pi)}. \tag{4.12}$$

Plugging this, for example, into V_d in (4.8) yields

$$\widehat{V}_d(w) = (-1)^N e^{i(N-1)w}\overline{(1-e^{-iw})^{\delta+N}\widehat{u}_{0,N}(w+\pi)}$$

$$= (-1)^N e^{i(N-1)w}\overline{(1+e^{-i(w+\pi)})^{\delta+N}\widehat{u}_{0,N}(w+\pi)} = (-1)^N e^{i(N-1)w}\overline{\widehat{U}_r(w+\pi)}. \tag{4.13}$$

Entries in Table 1 for larger cutoff levels ϵ suggest that only a small number of elements of filters U_d, V_d, U_r, V_r take on larger values. This is indeed what happens. Figure 1, for example, depicts the first 30 elements of filters U_r for $\delta = 0.4$ and $N = 2, 4, 6$ and 10. Table 2 also refines the information given in Table 1, for larger ϵ, by considering several values of δ.

It can be seen from Table 2 that, for $N \geq 4, \epsilon = 10^{-3}$ and $N \geq 6, \epsilon = 10^{-4}$, one may essentially take fractional filters of length

$$2N + 3, \tag{4.14}$$

with the second term 3 becoming 2 or 1 as N increases. Moreover, this choice works across the whole range of δ. This observation also suggests one form of filters in AAWDs considered in the next section.

The Daubechies OWD filters u and v used in Tables 1, 2 and Figure 1 have the so-called extremal phase and can be thought of as being the most asymmetric ones. Other analogous filters are possible having different phase properties, for example, the least asymmetric Daubechies filters (Daubechies (1992), Percival and Walden (2000)). It is interesting to discuss here briefly how these various choices affect the fractional filters (and, in particular, their phase). In Figures 2 and 3, we depict fractional filters obtained by considering all such possible OWD filters u and v, when $\delta = 0.4$ and $N = 4$. (The filters V_r are plotted as starting with index $n = 0$.) As can be seen from these figures, various choices of u and v lead to visually different fractional filters. In particular, the choice appearing in subplots 3 seems to yield least asymmetric fractional filters for $N = 4$. A higher degree of symmetry might be of interest in some applications (such as denoising) and in the visualization of transformation coefficients, though not in the applications appearing in Sections 6 and 7.

4.3 Approximate AWDs

The discussion in the beginning of Section 3.3 and around (4.14) suggests the construction of AAWDs based on truncated fractional filters. First, motivated by (4.14), one can set

$$U_d^{a_1}, V_d^{a_1}, U_r^{a_1}, V_r^{a_1} \tag{4.15}$$

as the filters U_d, V_d, U_r, V_r truncated to the length $2N + 3$ (the superscript a_1 stands for the first approximation). Note, however, that $V_d^{a_1}$ in particular would not exactly eliminate polynomial trends. This can be corrected by considering the filters

$$U_d^{a_2}, V_d^{a_2}, U_r^{a_2}, V_r^{a_2} \tag{4.16}$$

defined as follows. For example, the filter $V_r^{a_2}$ is defined through

$$\widehat{V}_r^{a_2}(w) = (1-e^{-iw})_t^{-\delta}\widehat{v}(w), \tag{4.17}$$

where

$$(1-e^{-iw})_t^{-\delta} = \sum_{k=0}^{K} g_k^{(-\delta)} e^{-iwk} \tag{4.18}$$

is the truncation of $(1-e^{-iw})^{-\delta}$ to K terms. Choosing $K = 2N + 3$ ensures that the first $2N + 3$ elements of $V_r^{a_2}$ are the same as those of V_r. The length of all filters (4.16) becomes

$$2N + (2N + 3) - 1 = 4N + 2. \tag{4.19}$$

For example, for $N = 4$ and 5, this equals 18 and 22, compared to 11 and 13 according to (4.14).

Another possibility of AAWDs is detailed in Example 3.1. Implementing this involves computing $r_{W^j}(0)$ and $r_{W^j}(1)$ in (3.32) and (3.33). Two methods for calculations are suggested in Craigmile et al. (2005): the exact method based on numerical integration of (3.32), and the ideal bandpass approximation method. The latter consists of using the ideal bandpass approximations of Daubechies filters u and v,

$$|\widehat{u}(w)| \approx |\widehat{u}_I(w)| := \begin{cases} \sqrt{2}, & |w| \in [0, \pi/2), \\ 0, & |w| \in [\pi/2, \pi), \end{cases} \quad |\widehat{v}(w)| \approx |\widehat{v}_I(w)| := \begin{cases} 0, & |w| \in [0, \pi/2), \\ \sqrt{2}, & |w| \in [\pi/2, \pi) \end{cases} \tag{4.20}$$

(which can be replaced by convergence as the number of zero moments $N \to \infty$), and hence

$$|\widehat{V}^j(w)| \approx \begin{cases} 0, & \text{otherwise}, \\ 2^{j/2}, & |w| \in [\pi/2^j, \pi/2^{j-1}), \end{cases} \tag{4.21}$$

for V^j appearing in (3.12). Substituting this into (3.33), the covariance (3.32) is approximated as

$$r_{W^j}(k) \approx \frac{2^j}{\pi} \int_{\pi/2^j}^{\pi/2^{j-1}} \cos(2^j w k) |\widehat{a}^0(w)|^2 dw. \tag{4.22}$$

Several finer numerical schemes for (4.22) are discussed in Craigmile (2000), pp. 47-48. Using the approximation (4.22), however, seems the only practical way of dealing with r_{W^j} when j is larger.

To give an idea of the resulting dependence structure of detail coefficients, we briefly report here on their within-scales spectral densities. Figure 4 depicts deviations from $(2\pi)^{-1}$ (that is, the spectral density of white noise with variance 1) of spectral densities f_{ξ^j} of detail coefficients at scales $j = 2, 3, 4$ and 5. The detail coefficients are computed from AAWD based on filters (4.15) or AAWD based on AR approximation. The chosen model is FARIMA$(0, \delta, 0)$ with $\delta = 0.4$ and the number of zero moments used is $N = 4$. As seen from the figure (in particular, its vertical scale with the small maximum value), both decompositions yield almost decorrelated coefficients, with that based on (4.15) achieving a more precise decorrelation. The fact that AAWD performs here well for AR approximation can also be seen by examining spectral densities of OWD detail coefficients and their approximations. Figure 5 shows such densities in theory and their approximations based on white noise or AR models. Consistently with Figure 4, the AR-based approximations provide a good fit to theoretical densities.

5 Examples of AWDs: AR and MA series

We study here AWDs of (Gaussian) AR or MA series. The focus will be on AR series throughout. (As will be seen below, the discussion on MA series can be essentially reduced to that for AR series with the roles of decomposition and reconstruction filters reversed.) These are series $X = a * \epsilon$ with a (Gaussian) white noise ϵ and

$$\widehat{a}(w) = \sigma(1 - \phi_1 e^{-iw} - \ldots - \phi_p e^{-ipw})^{-1} = \sigma(1 - \gamma_1 e^{-iw})^{-1} \ldots (1 - \gamma_p e^{-iw})^{-1}, \tag{5.1}$$

where $\sigma > 0$, the roots γ_i^{-1}, $i = 1, \ldots, p$, are outside the unit circle, and p denotes the order of series, that is, AR(p) series. The simplest case is that of AR(1) series with

$$\widehat{a}(w) = \sigma(1 - \gamma_1 e^{-iw})^{-1}, \tag{5.2}$$

where $-1 < \gamma_1 < 1$ ($\gamma_1 \neq 0$). The case of $\gamma_1 = \pm 1$, not considered here, corresponds to unit roots, and the case of γ_1 close to ± 1 ($-1 < \gamma_1 < 1$) is referred to as near unit roots. Exact and approximate AWDs of AR series are considered in Sections 5.1 and 5.2.

5.1 Exact AWDs

5.1.1 AR series

We consider here exact AWDs of the form (3.25). If the decomposition of X is of interest (as, for example, in maximum likelihood estimation: see Section 7), we can consider EAWD of the form (3.25) with

$$\widehat{a}^j(w) \equiv 1, \quad j \geq 1. \tag{5.3}$$

Then,

$$\widehat{U}_d^1(w) = \overline{\widehat{a}(w)}^{-1}\widehat{u}(w) = \sigma^{-1}(1 - \phi_1 e^{iw} - \ldots - \phi_p e^{ipw})\widehat{u}(w), \quad \widehat{V}_d^1(w) = \overline{\widehat{a}(w)}^{-1}\widehat{v}(w), \tag{5.4}$$

where $\widehat{a}(w)$ is given by (5.1), and

$$\widehat{U}_d^j(w) = \widehat{u}(w), \quad \widehat{V}_d^j(w) = \widehat{v}(w), \quad j \geq 2. \tag{5.5}$$

Hence, the corresponding filters U_d^j, V_d^j are of short and finite length (supposing that so are u and v). Note also that, in this case, all approximations X^j and details ξ^j are (Gaussian) white noise sequences.

With the choice (5.3),

$$\widehat{U}_r^0(w) = \widehat{a}(w)\widehat{u}(w), \quad \widehat{V}_r^0(w) = \widehat{a}(w)\widehat{v}(w) \tag{5.6}$$

and

$$\widehat{U}_r^j(w) = \widehat{u}(w), \quad \widehat{V}_r^j(w) = \widehat{v}(w), \quad j \geq 1. \tag{5.7}$$

When γ_i, $i = 1, \ldots, p$, in (5.1) are close to 0, the elements of $(1 - \gamma_i e^{-iw})^{-1} = \sum_{k=0}^{\infty} \gamma_i^k e^{-iwk}$ decay to zero rapidly and hence the filters U_r^0, V_r^0 can be taken of short length in practice. When γ_i is close to ± 1, however, the decay of γ_i^k is much slower, resulting in longer filters U_r^0, V_r^0. Zero moments are not helpful for U_r^0 when $0 < \gamma_i < 1$, and for V_r^0 when $-1 < \gamma_i < 0$. For example, this occurs for V_r^0 because, in simple terms, the elements of $(1 - \gamma_1 e^{-iw})^{-1} = \sum_{k=0}^{\infty} (-1)^k |\gamma_1|^k e^{-iwk}$ oscillate and the difference operator $(1 - e^{-iw})^N$ does not make them decrease to 0 faster (see Remark 4.2 above). Hence, in these cases, if the reconstruction is of interest, the choice (5.6) may not be the best one in practice.

When $0 < \gamma_i < 1$, the decay of U_r^0 can be improved by considering a different EAWD. Consider EAWD with the filter

$$\widehat{a}^j(w) = \sigma(1 - \gamma_1^{2^j} e^{-iw})^{-1} \ldots (1 - \gamma_p^{2^j} e^{-iw})^{-1}, \tag{5.8}$$

so that

$$\begin{aligned} \widehat{U}_r^j(w) &= (1 + \gamma_1^{2^j} e^{-iw}) \ldots (1 + \gamma_p^{2^j} e^{-iw})\widehat{u}(w), \\ \widehat{V}_r^j(w) &= \sigma(1 - \gamma_1^{2^j} e^{-iw})^{-1} \ldots (1 - \gamma_p^{2^j} e^{-iw})^{-1}\widehat{v}(w), \quad j \geq 0. \end{aligned} \tag{5.9}$$

In this case, U_r^j are of finite length. Moreover, the filters V_r^j can also be taken of finite and short length in practice, irrespective of the value of $0 < \gamma_i < 1$. This is a consequence of two complementary facts: first, when $\gamma_i^{2^j}$ is close to 0, the filter $(1 - \gamma_i^{2^j} e^{-iw})^{-1}$ already decays rapidly and, second, when $\gamma_i^{2^j}$ is close to 1, the larger number of zero moments make the filter V_r^j decay faster.

We illustrate the above for $p = 1$ (AR(1) series) in the following way. Focus on the filter

$$\widehat{V}_r^0(w) = \frac{\widehat{v}(w)}{1 - \gamma_1 e^{-iw}}, \tag{5.10}$$

where we took $\sigma = 1$ (note that V_r^j is V_r^0 by replacing γ_1 by $\gamma_1^{2^j}$). Let v be the Daubechies OWD wavelet filter with N zero moments, so that its length is $2N$. The filter V_r^0 is obtained by convolving

the sequence $(1, \gamma_1, \gamma_1^2, \ldots)$ with the filter v. Note that the $(2N + j)$th nonzero element of the convolution is

$$\gamma_1^j c := \gamma_1^j (1, \gamma_1, \ldots, \gamma_1^{2N-1}) v', \quad j \geq 0, \tag{5.11}$$

and decays as a geometric sequence. In Table 3, we provide the absolute values of the $(2N)$th nonzero element of the filter V_r^0 for various choices of the parameter γ_1 and the number of zero moments N. In parentheses, we provide the value of γ_1^{2N} for comparison. Note from the table that the filter V_r^0 is short from a practical perspective not only for smaller γ_1 but also when γ_1 is closer to 1. In particular, for $N \geq 6$, the length

$$2N + 3 \tag{5.12}$$

of the truncated filters ensures that all remaining elements are smaller than 10^{-3} (this uses the relation (5.11)). For illustration sake, in Figure 6, we provide plots of V_r^0 for several values of γ_1.

Note also that, with the choice (5.8) for EAWD, the approximations X^j become AR(p) time series with the roots $\gamma_i^{-2^j}$ of the corresponding AR polynomial. The decomposition filters associated with (5.8) are

$$\widehat{U}_d^j(w) = \frac{\widehat{u}(w)}{(1 + \gamma_1^{2^j} e^{iw}) \ldots (1 + \gamma_p^{2^j} e^{iw})}, \quad \widehat{V}_d^j(w) = \sigma^{-1}(1 - \gamma_1^{2^j} e^{iw}) \ldots (1 - \gamma_p^{2^j} e^{iw}) \widehat{v}(w). \tag{5.13}$$

When γ_i are close to 1, the filters U_d^j can also be seen to decay faster as the number of zero moments increases.

When $-1 < \gamma_1 < 0$ and especially when γ_1 is close to -1, the EAWD with (5.8) is not helpful because the decay of V_r^0 is not affected by the increasing number of zero moments (see Remark 4.2). In this case, the EAWD with (5.3) is probably the best that one can do.

5.1.2 Discussion on MA series

We studied above the case of AR time series. Suppose now that X is an MA time series and focus on MA(1) time series, that is, $X = a * \epsilon$ with

$$\widehat{a}(w) = 1 + \theta_1 e^{-iw}, \tag{5.14}$$

where $-1 < \theta_1 < 1$ ($\theta_1 \neq 0$). Since $\widehat{a}(w)$ in (5.14) is reciprocal to the filter in (5.2), our discussion in Section 5.1.1 also covers the case of MA(1) time series. For example, reconstruction filters are natural to define by taking EAWD with the choice (5.3), that is,

$$\widehat{a}^j(w) \equiv 1, \quad j \geq 1.$$

The resulting filters are

$$\widehat{U}_r^0(w) = (1 + \theta_1 e^{-iw}) \widehat{u}(w), \quad \widehat{V}_r^0(w) = (1 + \theta_1 e^{-iw}) \widehat{v}(w)$$

and $\widehat{U}_r^j(w) = \widehat{u}(w), \widehat{V}_r^j(w) = \widehat{v}(w)$, $j \geq 1$. In analogy to (5.8), the decomposition filters can naturally be defined for

$$\widehat{a}^j(w) = 1 - (-\theta_1)^{2^j} e^{-iw}, \quad j \geq 1. \tag{5.15}$$

Remark 5.1 Part (c) at the end of Section 3.2 suggests natural approximation (3.27), characterized by (3.28), as one possible choice for EAWD. In fact, such natural approximations underly some of the EAWDs for MA and AR series introduced above. Observe that, if X^0 is an MA(1) time series with $\widehat{a}(w)$ as in (5.14), then the functions (3.28) characterizing the natural approximations X^j in (3.27) become

$$\widehat{a}^j(w) \equiv 1. \tag{5.16}$$

If X^0 is an AR(1) time series with $\widehat{a}(w)$ in (5.2) (with $\sigma = 1$), then the functions (3.28) are

$$\widehat{a}^j(w) = (1 - \gamma_1^{2^j} e^{-iw})^{-1}. \tag{5.17}$$

Observe that (5.16) and (5.17) are exactly what was proposed above for EAWDs at reconstruction for MA(1) and AR(1) time series.

5.2 Approximate AWDs

As in Section 4.3 for fractional filters, we examine two possibilities of AAWDs for AR series. We focus on AR(1) time series, and consider approximation to EAWD given by (5.8). (EAWD based on (5.3) has finite associated decomposition filters and needs no approximation.) The corresponding decomposition filters are found in (5.13), and we shall approximate them to the length $2N + 3$ in (5.12). Alternatively, AAWD can be defined as in Example 3.1.

Figure 7 shows deviations from $(2\pi)^{-1}$ of spectral densities f_{ξ^j} of detail coefficients at scales $j = 2, 3, 4$ and 5, for the two AAWDs discussed above. The chosen model is AR(1) series with $\gamma_1 = 0.5$ and the number of zero moments is $N = 4$. The plots in the figure indicate that both methods do well in decorrelating detail coefficients. In Figure 8, we also provide spectral densities of OWD detail coefficients and their approximations based on white noise and AR models.

6 Applications of AWDs: simulation

We apply here AWDs to the simulation of Gaussian stationary time series. In Section 6.1, general AWD-based simulation methods are described. The performance of the simulation methods is investigated in Section 6.2.

6.1 Simulation

We first describe our methods and then examine them from various angles. Suppose that a Gaussian stationary time series X^0 of length $T = 2^J$ is desired. It can be simulated through AWD by the steps described next. We distinguish between EAWD-based and AAWD-based simulation methods for clarity, though the two methods could also be viewed under one framework.

EAWD-based simulation:

1. For $j = 0, 1, \ldots, J-1$, determine the largest length L_J of the reconstruction filters U_r^j, V_r^j in EAWD truncated at a chosen cutoff level $\epsilon > 0$. Let $\widetilde{U}_r^j, \widetilde{V}_r^j$, $j = 0, 1, \ldots, J-1$, be the reconstruction filters U_r^j, V_r^j truncated so that each one has length L_J.

2. Use some simulation method to generate an initial time series vector X^J of length $L_J + 1$. (The choice of this series is further discussed below.)

3. Apply the reconstruction scheme (3.5) recursively J times with the truncated reconstruction filters $\widetilde{U}_r^j, \widetilde{V}_r^j$ and taking into account the border effect to obtain the time series X^0 of length 2^J.

AAWD-based simulation:

1. For $j = 0, 1, \ldots, J-1$, determine the largest length L_J of the finite reconstruction filters U_r^j, V_r^j in AAWD.

2. Perform Step 2 in EAWD-based simulation above.

3. Perform Step 3 with filters U_r^j, V_r^j in EAWD-based simulation above.

Several observations regarding these methods and steps are in order. Consider first EAWD-based simulation method. By a cutoff level $\epsilon > 0$ in Step 1, we mean an extremely small value, such as 10^{-15} or smaller, which from a practical perspective effectively means that an exact reconstruction is being performed. For example, for FARIMA$(0, \delta, 0)$ series, we know from Section 4.1 that fractional filters truncated at even such small cutoff levels are fairly short if the number of zero moments is large. The second step refers to the fact that application of the reconstruction scheme (3.5) requires some initial approximation X^j. We take $j = J$ because X^J can be taken of the smallest possible length $L_J + 1$ in order to apply the simulation scheme (3.5). According to EAWD, the exact X^J is given by

$$X^J = a^J * \epsilon^J, \tag{6.1}$$

where a^J and the series X^J depend on a specific EAWD. Ignoring truncation of the EAWD reconstruction filters (which is insignificant from a practical perspective), the choice (6.1) leads to an exact reconstruction of the series X^0. The series X^J in (6.1) can be simulated by the popular Circulant Matrix Embedding (CME) method (Dietrich and Newsam (1997)) or, since L_J is often small, by the Durbin-Levinson algorithm (Brockwell and Davis (1991)). Another simpler possibility is just to take, for example,

$$X^J = 0, \tag{6.2}$$

especially for large J. The idea here is that approximations X^J can be thought of as representing frequencies $[0, 2\pi/2^{J-1})$ in the spectrum of X^0. Hence, as J increases, the choice of X^J should be less and less relevant. The choices (6.1) and (6.2) will be compared through simulations in Section 6.2 below. Regarding the third step, observe that by applying the scheme (3.5) with $\widetilde{U}_r^{J-1}, \widetilde{V}_r^{J-1}$ to X^J of length $L_J + 1$, we obtain $2(L_J + 1) - L_J = L_J + 2$ number of observations of the time series X^{J-1} which are unaffected by the border. Here, $2(L_J + 1)$ is the number of observations after the operation $\uparrow_2$ and $(-L_J)$ takes into account the border effect. By repeating this argument, the number of observations of the resulting time series X^0 which are unaffected by the border is $L_J + 2^J > 2^J$.

The AAWD-based simulation is performed in a similar way. Consider, for example, FARIMA$(0, \delta, 0)$ series for which several AAWDs are discussed in Section 4.3. In EAWD-based AAWD, the fractional filters U_r^j, V_r^j in Step 1 are taken of length $2N+3$. The initial series X^J could be taken as in (6.1) or (6.2). For AR-based AAWD, the exact choice of X^J consists of approximation coefficients obtained from OWD at scale 2^J. One possibility is also to use AR approximations for the series X^J. Another possibility is to set the series X^J equal to 0 again.

It is important to discuss here statistical properties of the series X^0 simulated according to the above methods. Because of the linear structure and Gaussian inputs, the series X^0 is always Gaussian. For the EAWD-based simulation with the initial series in (6.1), from a practical perspective the series X^0 is stationary and has the desired correlation structure. Strictly speaking, however, the series X^0 based on other suggested methods (that is, the EAWD-based method with zero initialization (6.2) and all AAWD-based methods) is not exactly stationary. This could easily be seen with $X^J = 0$ in (6.2). With this choice, note that

$$X^{J-1} = V_r^{J-1} * \uparrow_2 \xi^J,$$

which cannot be expected to be stationary because $\uparrow_2 \xi^J$ is not stationary. This is also the issue with previous wavelet-based simulations such as in, for example, Craigmile (2005) or Section 9.2 in Percival and Walden (2000).

6.2 Examining proposed simulation methods

We shall briefly evaluate here various simulation methods proposed in Section 6.1. Several approaches are possible (see, for example, Bardet et al. (2003)) and we somewhat follow part of the analysis done

in Craigmile (2005). We focus on simulation of FARIMA$(0, \delta, 0)$ time series models and take $\delta = 0.4$. We shall compare various simulation methods by studying how they perform in the estimation of long memory parameter δ. For estimation based on a series $X_0, \ldots, X_{T-1}$ of length T, we use an approximate Whittle ML estimator based on minimizing (see, for example, Beran (1994))

$$\log\left(\frac{1}{M}\sum_{k=1}^{M}\frac{I(\lambda_k)}{|1-e^{-i\lambda_k}|^{-2\delta}}\right) + \frac{1}{M}\sum_{k=1}^{M}\log(|1-e^{-i\lambda_k}|^{-2\delta}), \tag{6.3}$$

where $\lambda_k = 2\pi k/T$ are the Fourier frequencies, $I(\lambda) = (2\pi T)^{-1}|\sum_{t=1}^{T} e^{it\lambda}X_t|^2$ is the periodogram, and M is a threshold. We consider $M = [(T-1)/2]$, corresponding to the full approximate likelihood function, and also the popular choice of $M = T^{4/5}/4$, which corresponds to a local likelihood.

Tables 4 and 5 present differences in bias in the estimation of δ when simulation was performed based on CME method and simulation based on one of the methods described in Section 6.1. Several sample sizes T and initializations X_J are considered. The number of zero moments of the underlying MRA is $N = 10$ and fractional filters of the EAWD-based method were truncated at $\epsilon = 10^{-12}$. The tables, in fact, present 95% bootstrap confidence intervals and their centers with entries in bold. Note from these bold entries that EAWD-based simulation outperforms the other methods, and is one of the methods that includes 0 in the 95% confidence interval in the largest number of considered cases. Note also that the performance of all methods with initialization $X^J = 0$ is quite satisfactory, which provides evidence in favor of its use (as discussed in Section 6.1).

7 Applications of AWDs: MLE

We examine an application of AWDs to approximate MLE. In Section 7.1, our MLE method is described. In Sections 7.2 and 7.3, we examine its performance with FARIMA$(0, \delta, 0)$ and AR(1) models.

7.1 Estimation method

AWDs naturally lead to approximate MLE. MLE based on OWDs and AR(1)-based AAWDs was considered, in particular, by Craigmile et al. (2005). We shall therefore focus on MLE using EAWDs, when the associated filters are finite (as in Section 7.3 for AR(1) series), and EAWD-based AAWDs, when the corresponding filters have to be truncated (as in Section 7.2 for FARIMA$(0, \delta, 0)$ series). To explain the estimation method, we shall take a slightly less conventional approach which we find more revealing and better suited for our purposes.

Suppose one is considering a class of time series models indexed by unknown parameters $\widetilde{\theta}$, and having spectral densities $\widetilde{f}_{\widetilde{\theta}}(w)$. As is the case with many models (including those considered here), suppose also that, with a series $a_\theta = \{a_{\theta,n}\}_{n\geq 0} \in l^2(\mathbb{Z})$,

$$\widetilde{f}_{\widetilde{\theta}}(w) = \frac{\sigma^2}{2\pi}|\widehat{a}_\theta(w)|^2 =: \sigma^2 f_\theta(w), \tag{7.1}$$

where $\widetilde{\theta} = (\sigma^2, \theta)$ and

$$a_{\theta,0} = 1. \tag{7.2}$$

Let $\widetilde{\theta}_0 = (\sigma_0^2, \theta_0)$ be the true parameters of a time series to be estimated.

For later reference, suppose for the moment that the whole true time series X^0 is available. Then, under mild assumptions, one expects that

$$\theta_0 = \operatorname*{argmin}_\theta \int_0^{2\pi} \frac{|\widehat{a}_{\theta_0}(w)|^2}{|\widehat{a}_\theta(w)|^2} dw = \operatorname*{argmin}_\theta E(a_\theta^{-1} * X^0)^2, \tag{7.3}$$

where a_θ^{-1} denotes the inverse filter of a_θ, or even stronger,

$$a_{\theta_0} = \underset{a}{\operatorname{argmin}} \int_0^{2\pi} \frac{|\widehat{a}_{\theta_0}(w)|^2}{|\widehat{a}(w)|^2} dw = \underset{a}{\operatorname{argmin}}\, E(a^{-1} * X^0)^2, \tag{7.4}$$

where $a = (a_0, a_1, \ldots)$ is such that $a_0 = 1$. The basic idea behind (7.3)–(7.4) is that, for $b = \{b_n\}_{n\geq 0}$ with $b_0 = 1$, we obviously have

$$\underset{b}{\operatorname{argmin}} \int_0^{2\pi} |\widehat{b}(w)|^2 dw = \underset{b}{\operatorname{argmin}} \{1 + |b_1|^2 + |b_2|^2 + \ldots\} = (1, 0, 0, \ldots).$$

Relations (7.3)–(7.4) say that the true parameter value is (and hence could be estimated in practice as) the one for which the corresponding filter best decorrelates the series.

With EAWDs, the same argument as above is applicable due to Theorem 3.4. More precisely, consider EAWD based on (3.25), and such that

$$\widehat{a}^0(w) = \widehat{a}_\theta(w) \quad (\text{or } a^0 = a_\theta). \tag{7.5}$$

Denote by $\xi_n^{j,\theta}$ the detail coefficients in EAWD based on (3.25) and (7.5). By Theorem 3.4 and (7.3)–(7.4), we have that

$$\theta_0 = \underset{\theta}{\operatorname{argmin}} \sum_{j=1}^{\infty} 2^{-j} E(\xi_n^{j,\theta})^2. \tag{7.6}$$

Relation (7.6) suggests the following estimation procedure. Let $\xi_n^{j,\theta}$, $n = 1, \ldots, n_j$, $j = 1, \ldots, J$, be all available detail coefficients in AWD, based on data and *not* affected by the boundary. In practice, θ_0 could be estimated as

$$\widehat{\theta} = \underset{\theta}{\operatorname{argmin}} \sum_{j=1}^{J} \frac{2^{-j}}{n_j} \sum_{n=1}^{n_j} (\xi_n^{j,\theta})^2 =: \underset{\theta}{\operatorname{argmin}}\, \widehat{l}(\theta, J). \tag{7.7}$$

Since

$$n_j \approx \frac{T}{2^j}, \tag{7.8}$$

where T is the length of a given series, the estimator in (7.7) can also be thought of as

$$\widehat{\theta} \approx \underset{\theta}{\operatorname{argmin}} \sum_{j=1}^{J} \sum_{n=1}^{n_j} (\xi_n^{j,\theta})^2 =: \widehat{\theta}_1. \tag{7.9}$$

For later reference, it is useful to introduce

$$l(\theta, J) = \sum_{j=1}^{J} 2^{-j} E(\xi_n^{j,\theta})^2 = \frac{1}{2\pi} \int_0^{2\pi} \frac{|\widehat{a}_{\theta_0}(w)|^2}{|\widehat{a}_\theta(w)|^2} \left\{ \sum_{j=1}^{J} 2^{-j} |\widehat{V}^j(w)|^2 \right\} dw, \tag{7.10}$$

where the last equality follows from (3.24). (Strictly speaking, the second equality in (7.10) is true only when estimation is based on EAWD with finite filters, and is an approximation when these filters are truncated and used in AAWD.) Note also that it is natural to estimate the variance σ_0^2 through

$$\widehat{\sigma}^2 = \sum_{j=1}^{J} \frac{2^{-j}}{n_j} \sum_{n=1}^{n_j} (\xi_n^{j,\widehat{\theta}})^2 \tag{7.11}$$

(see Remark 7.1 below).

Remark 7.1 The estimator (7.7) can obviously be thought of as an approximate (Gaussian) ML estimator. Given observations $X^0 = (X_0, X_1, \ldots, X_{T-1})$, the *exact* negative (Gaussian) log-likelihood is (up to additive and multiplicative constants)

$$\log |\widetilde{\Sigma}_{\widetilde{\theta}}| + X^0 \widetilde{\Sigma}_{\widetilde{\theta}}^{-1} X^{0\prime}, \tag{7.12}$$

where $\widetilde{\Sigma}_{\widetilde{\theta}}$ is the covariance matrix of the model with unknown parameters $\widetilde{\theta}$, and $|\cdot|$ denotes the determinant. Under (7.1) and (7.2), one common simplification (approximation of Grenander and Szegö (1958)) of the log-determinant in (7.12) is

$$\log |\widetilde{\Sigma}_{\widetilde{\theta}}| \approx \frac{T}{2\pi} \int_0^{2\pi} \log\{2\pi f_{\widetilde{\theta}}(w)\} dw = T \log \sigma^2, \tag{7.13}$$

where the last equality follows from the Kolmogorov formula. See, for example, Fox and Taqqu (1986), Dahlhaus (1989), Chan and Palma (2006) for similar simplifications in the MLE context. The second term in (7.12) can be approximated using AWD. Using the above notation $\xi_n^{j,\theta}$, let $Y_\theta = (\xi_1^{1,\theta}, \ldots, \xi_{n_J}^{J,\theta})$ be the vector of all detail coefficients. The second term in (7.12) can be approximated by $(\sigma^2)^{-1} Y_\theta Y_\theta'$. The approximate log-likelihood function then becomes

$$T \log \sigma^2 + (\sigma^2)^{-1} Y_\theta Y_\theta' \tag{7.14}$$

and the corresponding estimators are

$$\widehat{\theta}_1 = \operatorname*{argmin}_\theta Y_\theta Y_\theta', \quad \widehat{\sigma}_1^2 = \frac{1}{\text{length}(Y_{\widehat{\theta}_1})} Y_{\widehat{\theta}_1} Y_{\widehat{\theta}_1}'. \tag{7.15}$$

Note that the estimator $\widehat{\theta}_1$ of θ is that given in (7.9). Modulo a possible truncation of filters in EAWD, all of the above estimators are asymptotically equivalent to exact ML estimators.

Remark 7.2 As mentioned in Section 1, the estimation method introduced in (7.5) is of interest for several reasons: it is an approximate MLE, is computationally fast, takes more into account the correlation structure of the series, and ignores polynomial trends. Moreover, though not emphasized in this work, the methodology can be extended to nonstationary time series.

In the remaining two sections, we examine the proposed method on two classes of models.

7.2 Estimation in the class of FARIMA$(0, \delta, 0)$ models

We examine here the performance of our estimation method in the class of FARIMA$(0, \delta, 0)$ models (Section 4). We focus on estimation of the parameter δ only.

First, it is interesting to examine $l(\delta, J)$ in (7.10) and $\widehat{l}(\delta, J)$ in (7.7). In Figure 9, we plot $l(\delta, J)$, for $J = 1, \ldots, 9$, on the intervals $\delta \in [0, 1/2]$ and $[-1/2, 1/2]$ when the true parameter value is $\delta_0 = 0.2$. With $J = 8$ or 9, the minimum of $l(\delta, J)$ is close to the true value $\delta_0 = 0.2$. (For example, with $J = 9$, the minimum is at $\delta = 0.19$ at the resolution 0.01 for δ.) For smaller values of J, the minimum is farther from the true value, and is in the interval $[-1/2, 1/2]$ till about $J = 4$. Numerical calculations in Figure 10 show that this difference (bias) is constant across δ_0 and hence could be taken into account in estimation. This invariance to δ_0 can also be easily seen in theory. Note that,

by using (7.10),

$$\begin{aligned}
\underset{\delta}{\operatorname{argmin}}\, l(\delta, J) &= \underset{\delta}{\operatorname{argmin}} \int_0^{2\pi} \frac{|\widehat{a}_{\delta_0}(w)|^2}{|\widehat{a}_{\delta}(w)|^2} \left\{ \sum_{j=1}^{J} 2^{-j} |\widehat{V}^j(w)|^2 \right\} dw \\
&= \underset{\delta}{\operatorname{argmin}} \int_0^{2\pi} |1 - e^{-iw}|^{2(\delta-\delta_0)} \left\{ \sum_{j=1}^{J} 2^{-j} |\widehat{V}^j(w)|^2 \right\} dw \\
&= \delta_0 + \underset{x}{\operatorname{argmin}} \int_0^{2\pi} |1 - e^{-iw}|^{2x} \left\{ \sum_{j=1}^{J} 2^{-j} |\widehat{V}^j(w)|^2 \right\} dw, \qquad (7.16)
\end{aligned}$$

where the last term does not depend on δ_0 (this is the bias which is computed numerically and plotted in Figure 10). In Figure 11, we present plots of $\widehat{l}(\delta, J)$ for 20 realizations of FARIMA$(0, 0.2, 0)$ series. We consider several values of the sample size $T = 2^8, 2^9, 2^{10}$ and 2^{11} with the corresponding (largest available) $J = 4, 5, 6$ and 7, and plot the corresponding true function $l(\delta, J)$ (in thicker line). EAWD-based AAWD is used based on the filters in (4.15). The number of zero moments used is $N = 5$ throughout.

In Table 6, we examine the performance of the proposed estimator (EAWD-based AAWD in the table) and compare it to those of several other estimators. Other estimators considered are the Whittle approximate ML estimator (based on minimizing (6.3) with $M = [(T-1)/2]$), the estimator based on OWD and the estimator based on AR(1)-based AAWD (as discussed in Craigmile et al. (2005)). The results are reported for FARIMA$(0, 0.4, 0)$ series of length $T = 2^8, 2^9, 2^{10}$ and 2^{11}. For wavelet-based methods, Daubechies filters with $N = 3$ and 6 vanishing moments are considered. The first line entries in the table cells are bias and standard deviations of the considered estimators (all based on 1,000 replications), with bias given by BS$\times 10^{-3}$ and standard deviation given by SD$\times 10^{-2}$. The entries on the second line under BS for EAWD-based AAWD estimator are bias corrections according to (7.16). (The other lines under "Whittle" are explained below.)

The results of the table show that the proposed estimator is outperformed by all the other estimators for the considered sample sizes, in terms of their respective standard deviations (the same is not exactly true for the bias). This fact might appear surprising in view of Remark 7.1, where it is argued that our estimator should be viewed as approximate ML estimator. Several explanations might seem plausible for the loss of efficiency in smaller samples, detailed next.

First, by comparing (7.7) and (6.3), this might be attributed to the presence of the additional, second term

$$\frac{1}{[(T-1)/2]} \sum_{k=1}^{[(T-1)/2]} \log(|1 - e^{-i\lambda_k}|^{-2\delta}) \qquad (7.17)$$

in (6.3). (The first term without the logarithm in (6.3) can be thought of as (7.7) in view of (7.10) and the ideal bandpass approximation (4.21).) For estimation here, however, the term (7.17) has little effect, and this should be expected (and is well known) in view of the approximation (7.13) and the discussion in Remark 7.1. The second line under SD for "Whittle" in Table 6 presents estimation results ignoring the term (7.17) and, indeed, these are not that different from those when the term (7.17) is taken into account. Second, in view of (7.10) and the ideal bandpass approximation (4.21), the EAWD-based estimator can be thought of as the Whittle-type estimator based on minimizing

$$\frac{1}{[(T-1)/2] - m + 1} \sum_{k=m}^{[(T-1)/2]} \frac{I(\lambda_k)}{|1 - e^{-i\lambda_k}|^{-2\delta}} \qquad (7.18)$$

for some $m \geq 1$. The third line under SD for "Whittle" in Table 6 presents estimation results based on minimizing (7.18) with the same $m = 30$ for all considered sample sizes T. Note an accord of

the values with the corresponding ones for EAWD-based AAWD, confirming the above heuristics. While the effect of fixed m is negligible in the asymptotic sense, the results reported here show that this still plays a significant role for smaller sample sizes.

7.3 Estimation in the class of AR(1) models

In this section, we present analogous plots and simulation results for estimation in AR(1) models. Figures 12 (left plot) and 13 are based on simulated AR(1) series with the true parameter $\gamma_{1,0} = 0.5$. The bias in Figure 12 (right plot) is not constant in $\gamma_{1,0}$, in contrast to Figure 10. This is immediately evident in theory as the argument in (7.16) cannot be repeated for an AR(1) filter.

The estimation results are reported in Table 7 and analogous observations can be made here as in Section 7.2. A small difference is that the SD based on EAWD is much closer to those based on other wavelet methods (and already smaller than that based on OWD when $T = 2^{11}$).

8 Proofs of the main results

Proof of Theorem 3.1: The condition (3.7) ensures that the time series in (3.4) and (3.5) are well-defined (Theorem 4.10.1 and Remark 1 in Brockwell and Davis (1991), pp. 154-155). The other parts of statement (i) are trivial. To show that (3.5) holds, consider for simplicity the case $j = 1$. (The general case can be proved in an analogous fashion.) Suppose that, as in (2.3),

$$X_n^0 = \int_0^{2\pi} e^{inw} \widehat{a}^0(w) dZ(w)$$

is the spectral representation of X^0. We first establish spectral representations of X^1 and ξ^1.

By Theorem 4.10.1 in Brockwell and Davis (1991), we obtain that

$$X_n^1 = \Big(\downarrow_2 (\overline{U}_d^1 * X^0) \Big)_n = \int_0^{2\pi} e^{i2nw} \widehat{c}^1(w) \widehat{a}^0(w) \overline{\widehat{u}(w)} dZ(w) = \int_0^{2\pi} e^{inw} dZ_1(w)$$

with

$$dZ_1(w) = \widehat{c}^1\Big(\frac{w}{2}\Big)\widehat{a}^0\Big(\frac{w}{2}\Big)\overline{\widehat{u}\Big(\frac{w}{2}\Big)} dZ\Big(\frac{w}{2}\Big) + \widehat{c}^1\Big(\frac{w}{2}+\pi\Big)\widehat{a}^0\Big(\frac{w}{2}+\pi\Big)\overline{\widehat{u}\Big(\frac{w}{2}+\pi\Big)} dZ\Big(\frac{w}{2}+\pi\Big).$$

Similarly,

$$\xi_n^1 = \Big(\downarrow_2 (\overline{V}_d^1 * X^0) \Big)_n = \int_0^{2\pi} e^{inw} dZ_2(w)$$

with

$$dZ_2(w) = \widehat{d}^1\Big(\frac{w}{2}\Big)\widehat{a}^0\Big(\frac{w}{2}\Big)\overline{\widehat{v}\Big(\frac{w}{2}\Big)} dZ\Big(\frac{w}{2}\Big) + \widehat{d}^1\Big(\frac{w}{2}+\pi\Big)\widehat{a}^0\Big(\frac{w}{2}+\pi\Big)\overline{\widehat{v}\Big(\frac{w}{2}+\pi\Big)} dZ\Big(\frac{w}{2}+\pi\Big).$$

We shall next establish (3.5) only at even times $n = 2s$. (The case $n = 2s+1$ can be proved in a similar way.) Using the spectral representation of X^1 above, we obtain that

$$(U_r^0 * \uparrow_2 X^1)_n = (\downarrow_2 U_r^0 * X^1)_s = \frac{1}{2}\int_0^{2\pi} e^{isw}\Big(\widehat{U}_r^0\Big(\frac{w}{2}\Big) + \widehat{U}_r^0\Big(\frac{w}{2}+\pi\Big)\Big) dZ_1(w)$$

$$= \frac{1}{2}\int_0^{2\pi} e^{inw} \widehat{U}_r^0(w) dZ_1(2w).$$

Similarly,

$$(V_r^0 * \uparrow_2 \xi^1)_n = \frac{1}{2}\int_0^{2\pi} e^{inw} \widehat{V}_r^0(w) dZ_1(2w).$$

Hence,

$$(U_r^0 * \uparrow_2 X^1)_n + (V_r^0 * \uparrow_2 \xi^1)_n$$
$$= \int_0^{2\pi} e^{inw}\left(\frac{1}{2}\widehat{U}_r^0(w)dZ_1(2w) + \frac{1}{2}\widehat{V}_r^0(w)dZ_2(2w)\right) = \int_0^{2\pi} e^{inw}\widehat{a}^0(w)dZ(w) = X_n^0,$$

since

$$\frac{1}{2}\widehat{U}_r^0(w)dZ_1(2w) + \frac{1}{2}\widehat{V}_r^0(w)dZ_2(2w)$$
$$= \frac{1}{2}\widehat{c}^1(w)^{-1}\widehat{u}(w)\left(\widehat{c}^1(w)\widehat{a}^0(w)\overline{\widehat{u}(w)}dZ(w) + \widehat{c}^1(w+\pi)\widehat{a}^0(w+\pi)\overline{\widehat{u}(w+\pi)}dZ(w+\pi)\right)$$
$$+\frac{1}{2}\widehat{d}^1(w)^{-1}\widehat{v}(w)\left(\widehat{d}^1(w)\widehat{a}^0(w)\overline{\widehat{v}(w)}dZ(w) + \widehat{d}^1(w+\pi)\widehat{a}^0(w+\pi)\overline{\widehat{v}(w+\pi)}dZ(w+\pi)\right)$$
$$= \frac{1}{2}\Big(\frac{\widehat{c}^1(w+\pi)}{\widehat{c}^1(w)}\widehat{u}(w)\overline{\widehat{u}(w+\pi)} + \frac{\widehat{d}^1(w+\pi)}{\widehat{d}^1(w)}\widehat{v}(w)\overline{\widehat{v}(w+\pi)}\Big)\widehat{a}^0(w+\pi)dZ(w+\pi)$$
$$+\frac{1}{2}\Big(|\widehat{u}(w)|^2 + |\widehat{v}(w)|^2\Big)\widehat{a}^0(w)dZ(w) = \widehat{a}^0(w)dZ(w). \qquad \square$$

Proof of Theorem 3.2: We will establish first that approximations $X^j = X^j(p)$ and details $\xi^j = \xi^j(p)$ are well-defined. In fact, we will show that

$$|X_n^j| \le C(1+|n|)^D, \tag{8.1}$$

where a constant C may depend on j. This bound is trivial for $j = 0$ since $X^0 = p$ is a polynomial of degree D. Suppose that (8.1) holds for $j-1$ and consider it for j. Then,

$$|X_n^j| \le \sum_k |U_{d,k}^j X_{n-k}^{j-1}| \le C_1 \sum_k (1+|k|)^{-D-2}(1+|n-k|)^D$$
$$\le C_2 \sum_k (1+|k|)^{-D-2}(1+|n|^D+|k|^D) \le C_3(1+|n|)^D,$$

where the constants C_1, C_2 and C_3 may depend on j. Using (8.1) and the assumed bound for $V_{d,n}^j$, the argument above also shows that ξ^j is well-defined.

To prove (3.9), we will first establish the formula

$$\widehat{X}^j(w) = \frac{1}{2^j}\sum_{n=0}^{2^j-1}\left\{\prod_{k=1}^{j}\overline{\widehat{U}_d^k\Big(\frac{w}{2^{j+1-k}} + b_{n,k}\Big)}\right\}\widehat{p}\Big(\frac{w}{2^j} + \frac{n\pi}{2^{j-1}}\Big), \tag{8.2}$$

where $b_{n,k} \in [0, 2\pi)$. Since p is not in $l^2(\mathbb{Z})$, the use of $\widehat{p}$ has to be clarified. Here and below, equations in the "spectral domain" should be interpreted through the "time domain" where, in particular, all products of Fourier transforms should be regarded as convolutions. The relation (8.2) is trivial for $j = 1$. Assume it holds for $j-1$ and consider it for j. Then, by (2.12),

$$\widehat{\downarrow_2 (\overline{U}_d^j * X^{j-1})}(w) = \frac{1}{2}\left(\overline{\widehat{U}_d^j\Big(\frac{w}{2}\Big)}\widehat{X}^{j-1}\Big(\frac{w}{2}\Big) + \overline{\widehat{U}_d^j\Big(\frac{w}{2}+\pi\Big)}\widehat{X}^{j-1}\Big(\frac{w}{2}+\pi\Big)\right)$$

$$= \frac{1}{2^j}\sum_{n=0}^{2^{j-1}-1}\prod_{k=1}^{j}\left(\overline{\widehat{U}_d^k\Big(\frac{w}{2^{j+1-k}} + b_{n,k}\Big)}\widehat{p}\Big(\frac{w}{2^j} + \frac{n\pi}{2^{j-2}}\Big) + \overline{\widehat{U}_d^k\Big(\frac{w}{2^{j+1-k}} + b'_{n,k}\Big)}\widehat{p}\Big(\frac{w}{2^j} + \frac{\pi}{2^{j-1}} + \frac{n\pi}{2^{j-2}}\Big)\right)$$

$$= \frac{1}{2^j}\sum_{n=0}^{2^{j-1}-1}\prod_{k=1}^{j}\left(\overline{\widehat{U}_d^k\Big(\frac{w}{2^{j+1-k}} + b_{n,k}\Big)}\widehat{p}\Big(\frac{w}{2^j} + \frac{2n\pi}{2^{j-1}}\Big) + \overline{\widehat{U}_d^k\Big(\frac{w}{2^{j+1-k}} + b'_{n,k}\Big)}\widehat{p}\Big(\frac{w}{2^j} + \frac{(2n+1)\pi}{2^{j-1}}\Big)\right)$$

$$= \frac{1}{2^j}\sum_{n=0}^{2^j-1}\prod_{k=1}^{j}\overline{\widehat{U}_d^k\Big(\frac{w}{2^{j+1-k}}+c_{n,k}\Big)}\widehat{p}\Big(\frac{w}{2^j}+\frac{n\pi}{2^{j-1}}\Big).$$

Since

$$\widehat{\xi}^j(w) = \frac{1}{2}\left(\overline{\widehat{V}_d^j\Big(\frac{w}{2}\Big)}\widehat{X}^{j-1}\Big(\frac{w}{2}\Big) + \overline{\widehat{V}_d^j\Big(\frac{w}{2}+\pi\Big)}\widehat{X}^{j-1}\Big(\frac{w}{2}+\pi\Big)\right)$$

and

$$\widehat{V}_d^j(w) = \overline{\widehat{d}^j(w)}\widehat{v}(w),$$

it suffices to prove that, for $n = 0, 1, ..., 2^{j-1}-1$,

$$\overline{\widehat{v}\Big(\frac{w}{2}\Big)}\widehat{p}\Big(\frac{w}{2^j}+\frac{2n\pi}{2^{j-1}}\Big) = 0 \tag{8.3}$$

and

$$\overline{\widehat{v}\Big(\frac{w}{2}+\pi\Big)}\widehat{p}\Big(\frac{w}{2^j}+\frac{(2n+1)\pi}{2^{j-1}}\Big) = 0. \tag{8.4}$$

Observe that, by using (8.2), the relation (8.3) follows from

$$\overline{\widehat{v}\Big(2^{j-1}\Big(\frac{w}{2^j}+\frac{2n\pi}{2^{j-1}}\Big)\Big)}\widehat{p}\Big(\frac{w}{2^j}+\frac{2n\pi}{2^{j-1}}\Big) = \overline{\widehat{v}(2^{j-1}w')}\widehat{p}(w')$$

$$= \overline{\widehat{v}_{0,N}(2^{j-1}w')}\prod_{k=2}^{j}(1+e^{i2^{j-k}w'})^N(1-e^{iw'})^N\widehat{p}(w') = 0,$$

since $(1-e^{-iw'})^N\widehat{p}(w') = 0$. A similar argument applies to (8.4). □

Proof of Theorem 3.3: It is convenient to continue using the spectral domain arguments as in the proofs of Theorems 3.1 and 3.2. For $X^0 = a^0 * \epsilon^0$, for example, we write $\widehat{X}^0 = \widehat{a}^0\widehat{\epsilon}^0$ and interpret $\widehat{\epsilon}^0$ as $dZ(w)$ with the latter as in (2.3), and $X_n^0 = \int_0^{2\pi} e^{inw}\widehat{X}^0(w)dw = \int_0^{2\pi} e^{inw}\widehat{a}^0(w)\widehat{\epsilon}^0(dw)$ in accord with (2.3). With this convention, one can see that

$$\widehat{X}^j(w) = \frac{1}{2^j}\sum_{k=0}^{2^j-1}\widehat{a}^0\Big(\frac{w}{2^j}+\frac{2\pi k}{2^j}\Big)\widehat{F}^j\Big(\frac{w}{2^j}+\frac{2\pi k}{2^j}\Big)\widehat{\epsilon}\Big(\frac{w}{2^j}+\frac{2\pi k}{2^j}\Big),$$

$$\widehat{\xi}^j(w) = \frac{1}{2^j}\sum_{k=0}^{2^j-1}\widehat{a}^0\Big(\frac{w}{2^j}+\frac{2\pi k}{2^j}\Big)\widehat{G}^j\Big(\frac{w}{2^j}+\frac{2\pi k}{2^j}\Big)\widehat{\epsilon}\Big(\frac{w}{2^j}+\frac{2\pi k}{2^j}\Big).$$

The proofs for (3.14)–(3.16) are similar and we only examine (3.16). This relation follows from the following argument. Note that, for $j' \geq j$, $n, n' \in \mathbb{Z}$,

$$E\xi_n^j\xi_{n'}^{j'} = E\int_0^{2\pi}\int_0^{2\pi} e^{inw-in'w'}\sum_{k=0}^{2^j-1}\widehat{a}^0\Big(\frac{w}{2^j}+\frac{2\pi k}{2^j}\Big)\widehat{G}^j\Big(\frac{w}{2^j}+\frac{2\pi k}{2^j}\Big)\widehat{\epsilon}\Big(\frac{w}{2^j}+\frac{2\pi k}{2^j}\Big)\cdot$$

$$\cdot\sum_{k'=0}^{2^{j'}-1}\overline{\widehat{a}^0\Big(\frac{w'}{2^{j'}}+\frac{2\pi k'}{2^{j'}}\Big)}\ \overline{\widehat{G}^{j'}\Big(\frac{w'}{2^{j'}}+\frac{2\pi k'}{2^{j'}}\Big)}\ \overline{\widehat{\epsilon}\Big(\frac{w'}{2^{j'}}+\frac{2\pi k'}{2^{j'}}\Big)}d\frac{w}{2^j}d\frac{w'}{2^{j'}}$$

$$= E\int_0^{2\pi/2^j}\int_0^{2\pi/2^{j'}} e^{i2^jnx-i2^{j'}n'y}\sum_{k=0}^{2^j-1}\widehat{a}^0\Big(x+\frac{2\pi k}{2^j}\Big)\widehat{G}^j\Big(x+\frac{2\pi k}{2^j}\Big)\widehat{\epsilon}\Big(x+\frac{2\pi k}{2^j}\Big)\cdot$$

$$\cdot \sum_{k'=0}^{2^{j'}-1} \overline{\widehat{a}^0\Big(y+\frac{2\pi k'}{2^{j'}}\Big)}\ \overline{\widehat{G}^{j'}\Big(y+\frac{2\pi k'}{2^{j'}}\Big)}\ \overline{\widehat{\epsilon}\Big(y+\frac{2\pi k'}{2^{j'}}\Big)}dxdy$$

Since

$$E\int_0^{2\pi/2^j}\int_0^{2\pi/2^{j'}} e^{i2^j nx - i2^{j'}n'y}\widehat{a}^0\Big(x+\frac{2\pi k}{2^j}\Big)\overline{\widehat{a}^0\Big(y+\frac{2\pi k'}{2^{j'}}\Big)}\cdot$$

$$\cdot\ \widehat{G}^j\Big(x+\frac{2\pi k}{2^j}\Big)\overline{\widehat{G}^{j'}\Big(y+\frac{2\pi k'}{2^{j'}}\Big)}\ \widehat{\epsilon}\Big(x+\frac{2\pi k}{2^j}\Big)\overline{\widehat{\epsilon}\Big(y+\frac{2\pi k'}{2^{j'}}\Big)}dxdy$$

$$=\int_{2\pi(k'-2^{j'-j}k)/2^{j'}}^{2\pi(k'+1-2^{j'-j}k)/2^{j'}} e^{i2^j nx - i2^{j'}n'x}\widehat{a}^0\Big(x+\frac{2\pi k}{2^j}\Big)\overline{\widehat{a}^0\Big(x+\frac{2\pi k}{2^j}\Big)}\widehat{G}^j\Big(x+\frac{2\pi k}{2^j}\Big)\overline{\widehat{G}^{j'}\Big(x+\frac{2\pi k}{2^j}\Big)}\frac{dx}{2\pi}$$

for $k' = 2^{j'-j}k, \ldots, 2^{j'-j}(k+1)-1$, and 0 otherwise, or

$$\int_{2\pi k'/2^{j'}}^{2\pi(k'+1)/2^{j'}} e^{i(n2^j - n'2^{j'})w}|\widehat{a}^0(w)|^2\widehat{G}^j(w)\overline{\widehat{G}^{j'}(w)}\frac{dw}{2\pi}$$

for $k' = 2^{j'-j}k, \ldots, 2^{j'-j}(k+1)-1$, and 0 otherwise, we obtain that

$$E\xi_n^j\xi_{n'}^{j'} = \int_0^{2\pi} e^{i(n'2^{j'} - n2^j)w}|\widehat{a}^0(w)|^2\widehat{G}^j(w)\overline{\widehat{G}^{j'}(w)}\frac{dw}{2\pi}. \qquad \square$$

Proof of Theorem 3.4: Note from (3.17)–(3.19) that

$$E(\xi_n^j)^2 = \int_0^{2\pi} f_{\xi^j}(w)dw = \frac{1}{2\pi}\int_0^{2\pi}|\widehat{a}^0(w)|^2|\widehat{G}^j(w)|^2dw.$$

The relation (3.22) then follows immediately from the definition (3.13) of $\widehat{G}^j$, and the relation (3.23) is a consequence of using underlying OWD. $\square$

References

Abry, P., Flandrin, P., Taqqu, M. S. & Veitch, D. (2003), Self-similarity and long-range dependence through the wavelet lens, *in* 'Theory and Applications of Long-Range Dependence', Birkhäuser Boston, Boston, MA, pp. 527–556.

Bardet, J.-M., Lang, G., Oppenheim, G., Philippe, A. & Taqqu, M. S. (2003), Generators of long-range dependence processes: a survey, *in* P. Doukhan, G. Oppenheim & M. S. Taqqu, eds, 'Theory and Applications of Long-Range Dependence', Birkhäuser.

Beran, J. (1994), *Statistics for Long-Memory Processes*, Chapman & Hall, New York.

Blu, T. & Unser, M. (2007), 'Self-Similarity: Part II Optimal Estimation of Fractal Processes', *IEEE Transactions on Signal Processing* **55**(4), 1364–1378.

Brockwell, P. J. & Davis, R. A. (1991), *Time Series: Theory and Methods*, 2nd edn, Springer-Verlag, New York.

Chan, N. & Palma, W. (2006), Estimation of long-memory time series models: A survey of different likelihood-based methods, *in* T. Fomby & D. Terrell, eds, 'Econometric Analysis of Financial and Economic Time Series/Part B', Vol. 20 of *Advances in Econometrics*, Elsevier, pp. 89–121.

Craigmile, P. F. (2000), Wavelet-based estimation for trend contaminated long memory processes, PhD thesis, University of Washington, Seattle, Washington.

Craigmile, P. F. (2005), 'Approximate wavelet-based simulation of long memory processes', *J. Stat. Comput. Simul.* **75**(5), 363–379.

Craigmile, P. F. & Percival, D. B. (2005), 'Asymptotic decorrelation of between-scale wavelet coefficients', *IEEE Transactions on Information Theory* **51**(3), 1039–1048.

Craigmile, P. F., Guttorp, P. & Percival, D. B. (2005), 'Wavelet-based parameter estimation for polynomial contaminated fractionally differenced processes', *IEEE Transactions on Signal Processing* **53**(8, part 2), 3151–3161.

Dahlhaus, R. (1989), 'Efficient parameter estimation for self similar processes', *The Annals of Statistics* **17**(4), 1749–1766.

Daubechies, I. (1992), *Ten Lectures on Wavelets*, SIAM Philadelphia. CBMS-NSF series, Volume 61.

Delbeke, L. (1998), Wavelet based estimators for the Hurst parameter of a self-similar process, PhD thesis, KU Leuven, Belgium.

Didier, G. & Pipiras, V. (2008), 'Gaussian stationary processes: adaptive wavelet decompositions, discrete approximations, and their convergence', *The Journal of Fourier Analysis and Applications* **14**(2), 203–234.

Dietrich, C. R. & Newsam, G. N. (1997), 'Fast and exact simulation of stationary Gaussian processes through circulant embedding of the covariance matrix', *SIAM Journal on Scientific Computing* **18**(4), 1088–1107.

Dijkerman, R. W. & Mazumdar, R. R. (1994), 'On the correlation structure of the wavelet coefficients of fractional Brownian motion', *IEEE Transactions on Information Theory* **40**(5), 1609–1612.

Fan, Y. & Gençay, R. (2006), Unit root and cointegration tests with wavelets, Preprint.

Flandrin, P. (1992), 'Wavelet analysis and synthesis of Fractional Brownian motion', *IEEE Transactions on Information Theory* **38**, 910–917.

Fox, R. & Taqqu, M. S. (1986), 'Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series', *The Annals of Statistics* **14**, 517–532.

Grenander, U. & Szego, G. (1958), *Toeplitz Forms and their Applications*, Chelsea, New York.

Jaffard, S., Lashermes, B. & Abry, P. (2006), Wavelet leaders in multifractal analysis, *in* T. Qian, M. I. Vai & X. Yuesheng, eds, 'Wavelet Analysis and Applications', Birkhäuser Verlag, Basel, Switzerland, pp. 219–264.

Jensen, M. J. (1999), 'An approximate wavelet MLE of short and long memory parameters', *Studies in Nonlinear Dynamics and Econometrics* **3**, 239–253.

Mallat, S. (1998), *A Wavelet Tour of Signal Processing*, Academic Press Inc., San Diego, CA.

Mallat, S., Papanicolaou, G. & Zhang, Z. (1998), 'Adaptive covariance estimation of locally stationary processes', *The Annals of Statistics* **26**(1), 1–47.

McCoy, E. J. & Walden, A. T. (1996), 'Wavelet analysis and synthesis of stationary long-memory processes', *Journal of Computational and Graphical Statistics* **5**(1), 26–56.

Meyer, Y., Sellan, F. & Taqqu, M. S. (1999), 'Wavelets, generalized white noise and fractional integration: the synthesis of fractional Brownian motion', *The Journal of Fourier Analysis and Applications* **5**(5), 465–494.

Moulines, E., Roueff, F. & Taqqu, M. S. (2007), 'On the spectral density of the wavelet coefficients of long-memory time series with application to the log-regression estimation of the memory parameter', *Journal of Time Series Analysis* **28**(2), 155–187.

Moulines, E., Roueff, F. & Taqqu, M. S. (2008), 'A wavelet Whittle estimator of the memory parameter of a nonstationary Gaussian time series', *The Annals of Statistics* **36**(4), 1925–1956.

Nason, G. P., von Sachs, R. & Kroisandt, G. (2000), 'Wavelet processes and adaptive estimation of the evolutionary wavelet spectrum', *Journal of the Royal Statistical Society. Series B. Statistical Methodology* **62**(2), 271–292.

Ossiander, M. & Waymire, E. C. (2000), 'Statistical estimation for multiplicative cascades', *The Annals of Statistics* **28**(6), 1533–1560.

Percival, D. B. & Walden, A. T. (2000), *Wavelet Methods for Time Series Analysis*, Vol. 4 of *Cambridge Series in Statistical and Probabilistic Mathematics*, Cambridge University Press, Cambridge.

Pipiras, V. (2005), 'Wavelet-based simulation of fractional Brownian motion revisited', *Applied and Computational Harmonic Analysis* **19**(1), 49–60.

Resnick, S., Samorodnitsky, G., Gilbert, A. & Willinger, W. (2003), 'Wavelet analysis of conservative cascades', *Bernoulli* **9**(1), 97–135.

Tewfik, A. H. & Kim, M. (1992), 'Correlation structure of the discrete wavelet coefficients of fractional Brownian motions', *IEEE Transactions on Information Theory* **IT-38**(2), 904–909.

Unser, M. & Blu, T. (2007), 'Self-Similarity: Part I Splines and Operators', *IEEE Transactions on Signal Processing* **55**(4), 1352–1363.

Vannucci, M. & Corradi, F. (1999), 'Covariance structure of wavelet coefficients: theory and models in a Bayesian perspective', *Journal of the Royal Statistical Society. Series B. Statistical Methodology* **61**(4), 971–986.

Veitch, D. & Abry, P. (1999), 'A wavelet-based joint estimator of the parameters of long-range dependence', *IEEE Transactions on Information Theory* **45**(3), 878–897.

Gustavo Didier
Mathematics Department
Tulane University
6823 St. Charles Avenue
New Orleans, LA 70118, USA
gdidier@tulane.edu

Vladas Pipiras
Dept. of Statistics and Operations Research
UNC at Chapel Hill
CB#3260, Smith Bldg.
Chapel Hill, NC 27599, USA
pipiras@email.unc.edu

Lengths of truncated filters					
Filters	Cutoff ϵ	$N = 2$	$N = 4$	$N = 6$	$N = 10$
U_r, V_d	10^{-3}	8	9	12	19
	10^{-5}	27	16	16	20
	10^{-9}	375	70	39	30
	10^{-13}	5,597	370	124	56
V_r, U_d	10^{-3}	11	10	13	20
	10^{-5}	52	19	18	23
	10^{-9}	1,739	111	49	35
	10^{-13}	60,007	793	175	65

Table 1: Lengths of truncated fractional filters at cutoff level ϵ with $\delta = 0.4$ and the Daubechies MRA with N zero moments.

Lengths of truncated filters											
		$N = 2$		$N = 4$		$N = 6$		$N = 8$		$N = 10$	
Filters	$\delta\backslash\epsilon$	10^{-3}	10^{-4}	10^{-3}	10^{-4}	10^{-3}	10^{-4}	10^{-3}	10^{-4}	10^{-3}	10^{-4}
U_r, V_d	-0.4	10	22	9	12	11	13	15	16	18	19
	-0.2	8	17	9	11	12	13	15	16	18	19
	0.2	8	12	9	11	12	13	15	16	18	19
	.4	8	15	9	12	12	13	15	16	19	19
V_r, U_d	-0.4	9	15	11	13	14	15	17	18	21	22
	-0.2	8	14	10	13	13	15	17	18	21	21
	0.2	9	17	10	13	13	15	16	18	20	21
	.4	11	23	10	14	13	15	16	18	20	21

Table 2: Lengths of truncated fractional filters at cutoff level ϵ, parameter δ and the Daubechies MRA with N zero moments.

Size of the $(2N)$th nonzero element					
γ_1	$N=2$	$N=4$	$N=6$	$N=8$	$N=10$
0.1	-0.4016 (0.0001)	-0.1652 (1×10^{-8})	-0.0692 (1×10^{-12})	-0.0293 (1×10^{-16})	-0.0124 (1×10^{-20})
0.3	-0.2556 (0.0081)	-0.0718 (6.56×10^{-5})	-0.0207 (5.31×10^{-7})	-0.006 (4.3×10^{-9})	-0.00179 (3.48×10^{-11})
0.5	-0.1369 (0.0625)	-0.0221 (0.0039)	-0.0037 (2.4×10^{-4})	-6.32×10^{-4} (1.52×10^{-5})	-1.089×10^{-4} (9.536×10^{-7})
0.7	-0.0516 (0.2401)	-0.00336 (0.0576)	-2.3012×10^{-4} (0.0138)	-1.606×10^{-5} (0.0033)	-1.1353×10^{-6} (7.979×10^{-4})
0.9	-0.0059 (0.6561)	-4.85×10^{-5} (0.4304)	-4.155×10^{-7} (0.2824)	-3.649×10^{-9} (0.1853)	-3.251×10^{-11} (0.12157)
0.999	-6.1×10^{-7} (0.996)	-5.22×10^{-13} (0.992)	3.33×10^{-16} (0.988)	1.55×10^{-15} (0.9841)	-5.51×10^{-16} (0.98018)

Table 3: The $(2N)$th nonzero element of the filter V_r^0 for various choices of γ_1 and the Daubechies MRA with N zero moments. The parentheses contain the value γ_1^{2N} for comparison.

Differences in bias from CME method; 95% CI and its center								
			AAWD based on					
method	EAWD		EAWD		AR		WN	
$T\backslash X_J$	exact	0	exact	0	AR	0	WN	0
256 (2^8)	−.00393 **−.00039** .00313	.00273 **.00594** .00916	−.00689 **−.00358** −.00028	−.00049 **−.00265** .00580	−.00450 **−.00120** .00208	.00012 **.00336** .00661	−.00108 **.00219** .00547	.01589 **.01937** .02291
1024 (2^{10})	−.00115 **−.00051** .00218	.00123 **.00271** .00418	−.00243 **−.00071** .00099	.00212 **.00376** .00539	0.000096 **.00174** .00340	.00174 **.00335** .00496	.00175 **.00345** .00515	.00963 **.01123** .01282
65,536 (2^{16})	−.00077 **.000019** .00081	.000009 **.00072** .000144	−.00087 **−.000012** .00063	−.00059 **.00020** .00101	.00077 **.00095** .00113	.00095 **.00113** .00132	−.00036 **.00043** .00122	.00045 **.00123** .00201

Table 4: Differences in bias from CME method when $M=[(T-1)/2]$: 95% bootstrap confidence intervals with center in bold.

Differences in bias from CME method; 95% CI and its center								
			AAWD based on					
method	EAWD		EAWD		AR		WN	
$T\backslash X_J$	exact	0	exact	0	AR	0	WN	0
	−.01615	.00654	−.01240	.01128	−.01268	−.00072	−.03408	−.00326
256	**−.00919**	**.01535**	**−.00499**	**.019**	**−.00523**	**.00353**	**−.02679**	**.00415**
(2^8)	−.00223	.02417	.00241	.02671	.00220	.00779	−.01951	.01156
	−.00380	.00754	−.00474	.00692	−0.00546	−.00319	−.01334	.00575
1024	**.00041**	**.01165**	**−.00071**	**.01094**	**−.00182**	**.00095**	**−.00906**	**.00967**
(2^{10})	.00464	.01577	.00331	.01497	.00181	.00510	−.00479	.01359
	−.00015	−.00014	−.00024	−.000047	.00077	−.00018	−.00108	.01589
65,536	**.000030**	**.000033**	**−.000057**	**.00015**	**.00095**	**.00052**	**−.00219**	**−.01937**
(2^{16})	.00021	.00021	.00012	.00035	.00113	.00123	.00547	.02291

Table 5: Differences in bias from CME method when $M = T^{4/5}/4$:
95% bootstrap confidence intervals with its center (in bold).

	Whittle			OWD		AR-based AAWD		EAWD-based AAWD	
T	BS	SD	N	BS	SD	BS	SD	BS	SD
2^8	-0.1181	5.3301	3	-24.0570	7.5179	-11.5524	7.3680	-149.6794	13.5040
(256)		5.2870						-49.6794	
		8.6105	6	-19.9726	9.6167	-14.0797	9.0117	-300.6189	20.4007
								-70.6189	
2^9	2.4903	3.7302	3	-13.3664	4.7617	-8.7847	4.9632	-67.4455	6.8943
(512)		3.9384						-17.4456	
		8.2993	6	-6.6289	5.8080	-2.6785	5.5099	-123.3027	8.4662
								-13.3027	
2^{10}	2.0819	2.5361	3	-10.4865	3.0138	-5.5848	2.9768	-32.9087	3.9097
(1024)		2.5690						-2.9088	
		3.9825	6	-6.2808	3.5593	-4.0755	3.5223	-57.3903	4.6499
								-7.3904	
2^{11}	0.3009	1.8405	3	-7.7201	2.0635	-5.8618	1.9474	-17.6581	2.4173
(2048)		1.7666						-7.6581	
		2.4087	6	-5.5773	2.3556	-1.3739	2.1987	-31.0907	2.6786
								-21.0907	

Table 6: Estimation results for FARIMA$(0, 0.4, 0)$ series with estimator bias given by BS$\times 10^{-3}$, and standard deviation by SD$\times 10^{-2}$. The entries on the second line under BS for EAWD-based AAWD are bias corrected values as discussed in Section 7.2. The entries on the second and third lines under SD for Whittle are explained at the end of Section 7.2.

	Whittle			OWD		AR-based AAWD		EAWD	
T	BS	SD	N	BS	SD	BS	SD	BS	SD
2^8 (256)	-6.0450	5.4433 5.3440 6.2196	3	-32.6106	6.9459	-22.8381	6.6773	-59.3497	7.5639
			6	-17.6067	8.9911	-9.5455	8.7830	-121.1312	9.2544
2^9 (512)	-2.6256	3.7908 3.8608 4.4643	3	-26.0404	4.4280	-17.8881	4.2618	-26.6541	4.6981
			6	-12.1983	5.0739	-5.9867	4.9876	-52.5886	5.5276
2^{10} (1024)	-1.0599	2.7298 2.6334 2.9856	3	-23.4273	3.0862	-16.0067	2.9562	-11.9731	3.0987
			6	-10.8367	3.2882	-5.2459	3.2112	-25.2154	3.4521
2^{11} (2048)	-0.8880	1.9464 1.8699 2.0217	3	-23.1767	2.1439	-15.9476	2.0701	-6.3215	2.0855
			6	-11.0269	2.1998	-5.8309	2.1211	-12.6888	2.2509

Table 7: Estimation results for AR(1), $\gamma_1 = 0.5$, series with estimator bias given by BS$\times 10^{-3}$, and standard deviation by SD$\times 10^{-2}$. The entries on the second and third lines under SD for Whittle are obtained as for Table 6: see the end of Section 7.2 for an explanation.

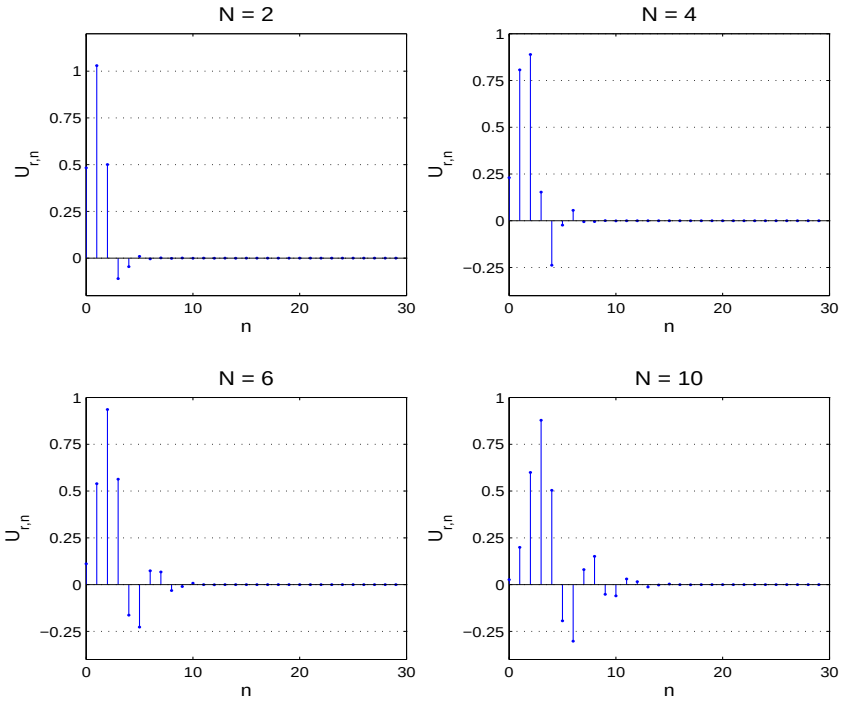

Figure 1: Fractional filters U_r for $\delta = 0.4$ and various N.

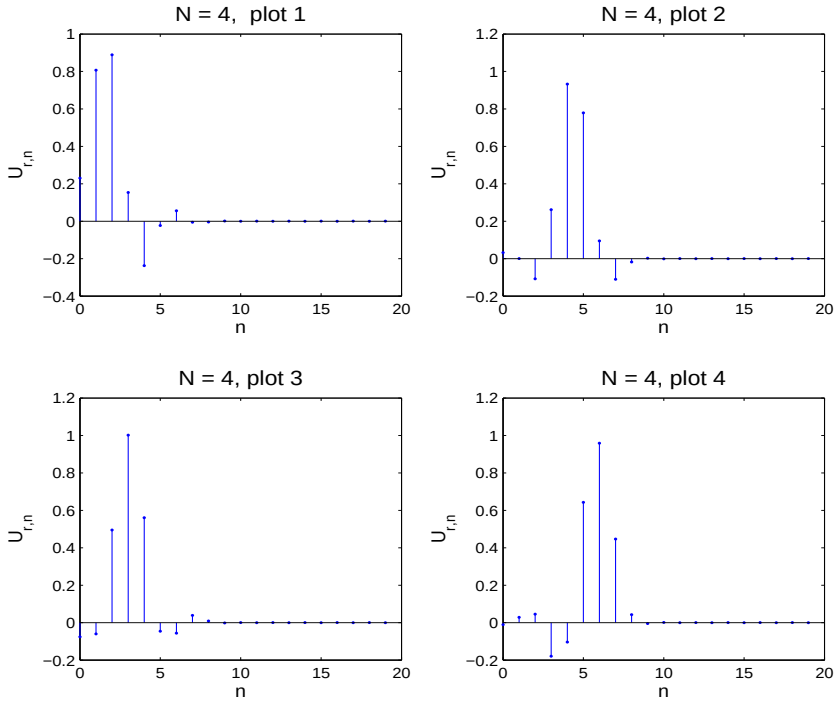

Figure 2: Fractional filters U_r for $\delta = 0.4$ and $N = 4$ for all possible underlying Daubechies OWD filters.

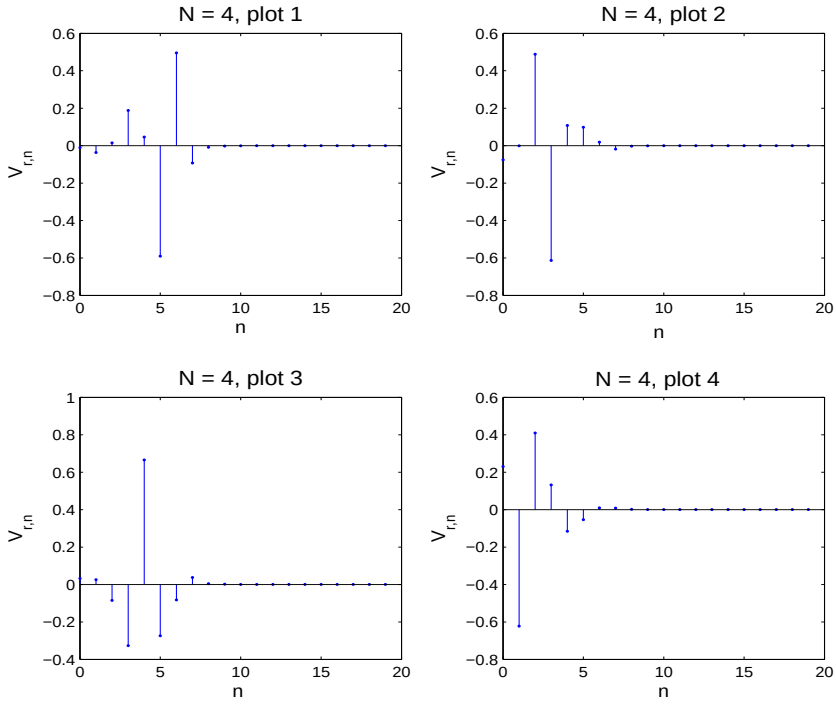

Figure 3: Fractional filters V_r for $\delta = 0.4$ and $N = 4$ for all possible underlying Daubechies OWD filters.

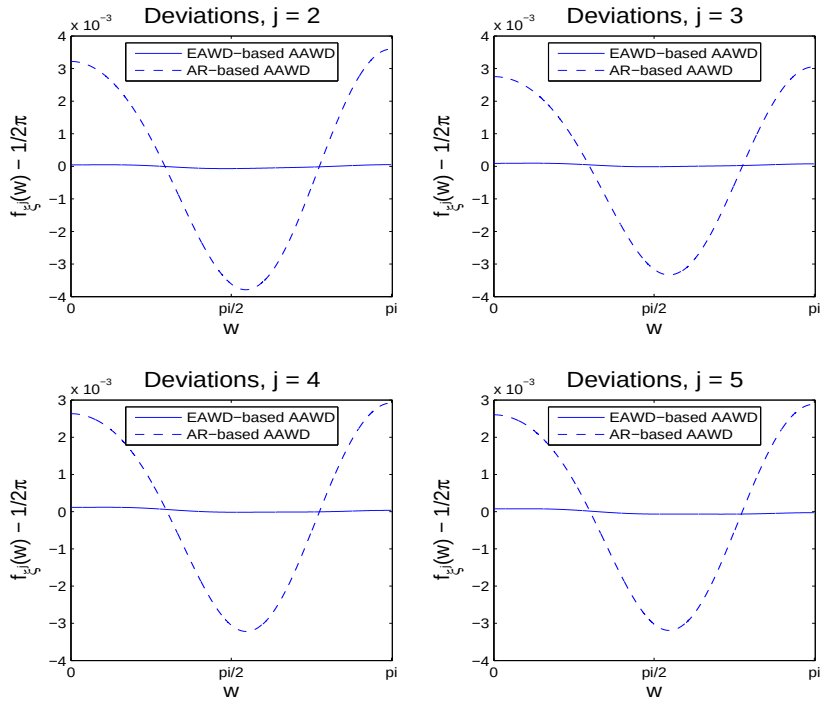

Figure 4: Deviations of spectral densities $f_{\xi j}$ from $(2\pi)^{-1}$ in FARIMA$(0, 0.4, 0)$ model.

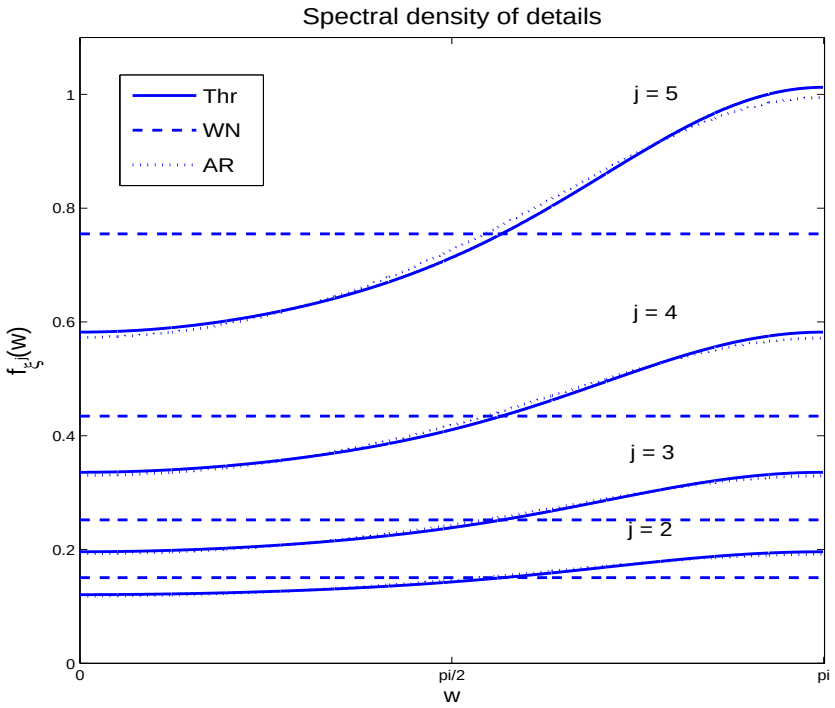

Figure 5: Spectral densities of OWD detail coefficients and their approximations using WN and AR models in FARIMA$(0, 0.4, 0)$ model.

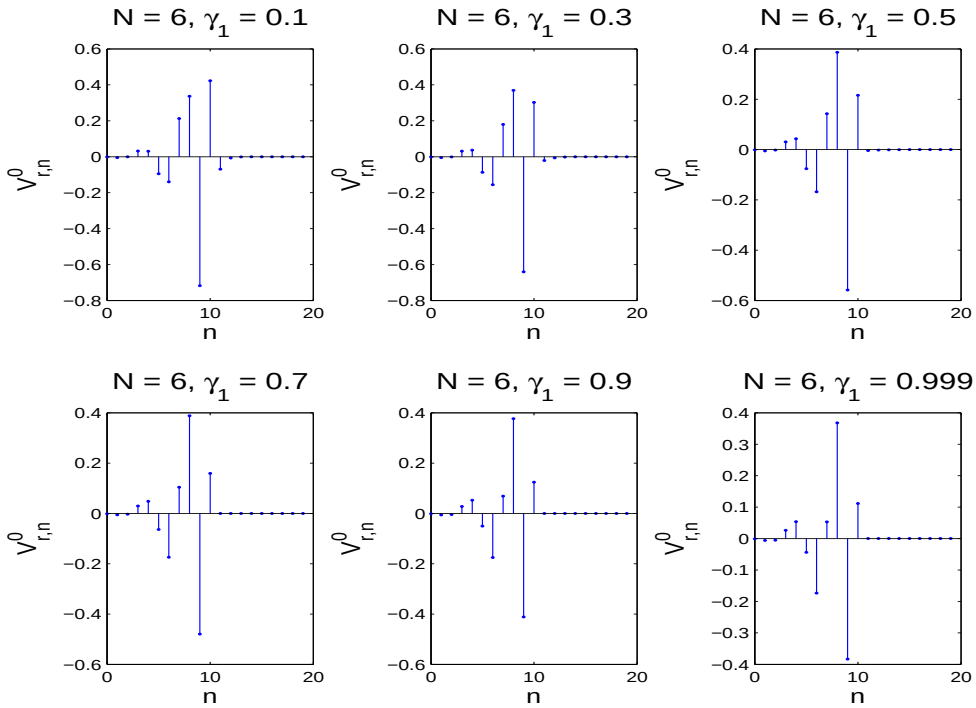

Figure 6: Filters V_r^0 for various choices of γ_1 and the Daubechies MRA with $N = 6$ zero moments.

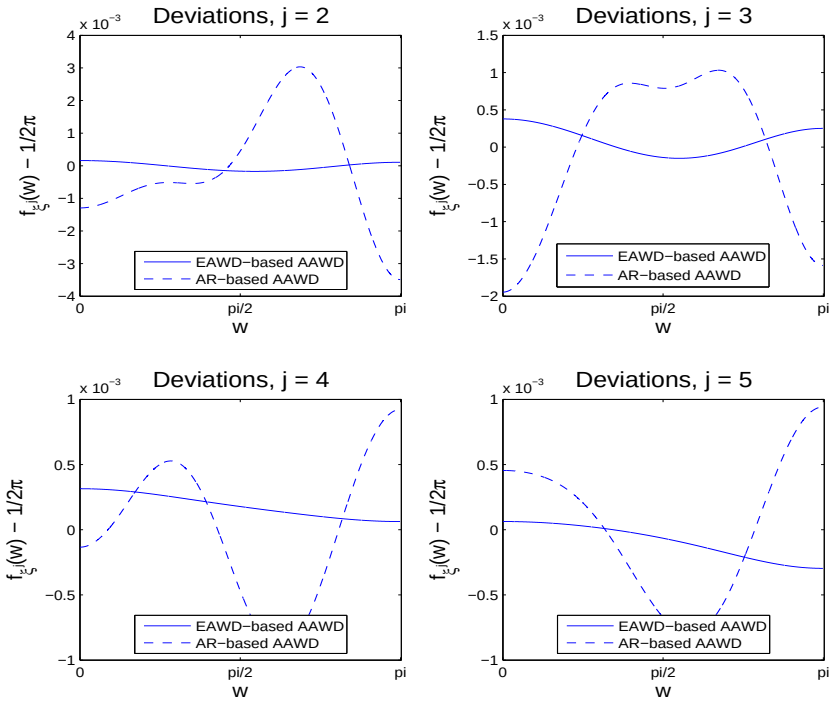

Figure 7: Deviations of spectral densities f_{ξ^j} from $(2\pi)^{-1}$ in AR(1) model with parameter $\gamma_1 = 0.5$.

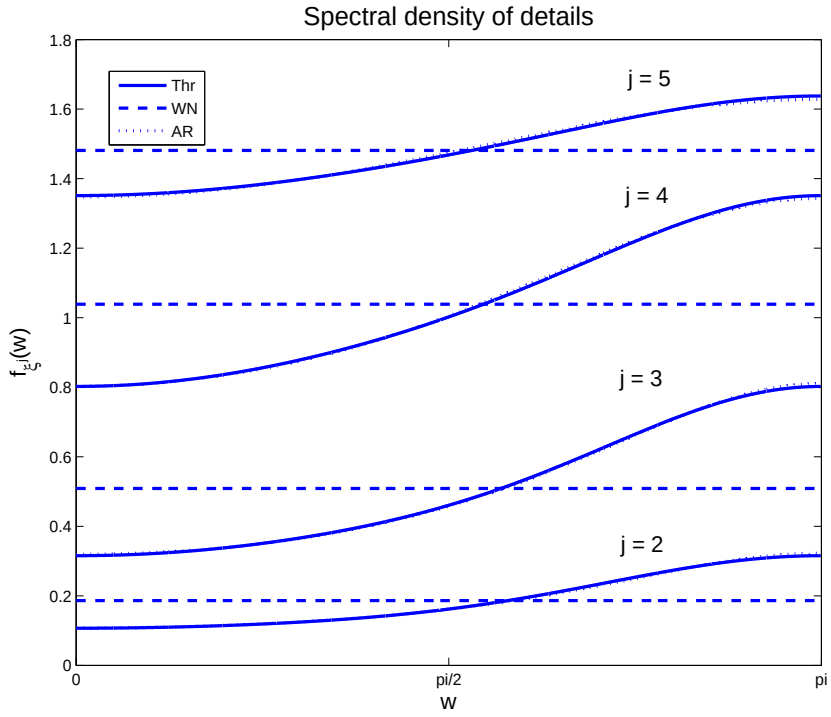

Figure 8: Spectral densities of OWD detail coefficients and their approximations using WN and AR models in AR(1) model with parameter $\gamma_1 = 0.5$.

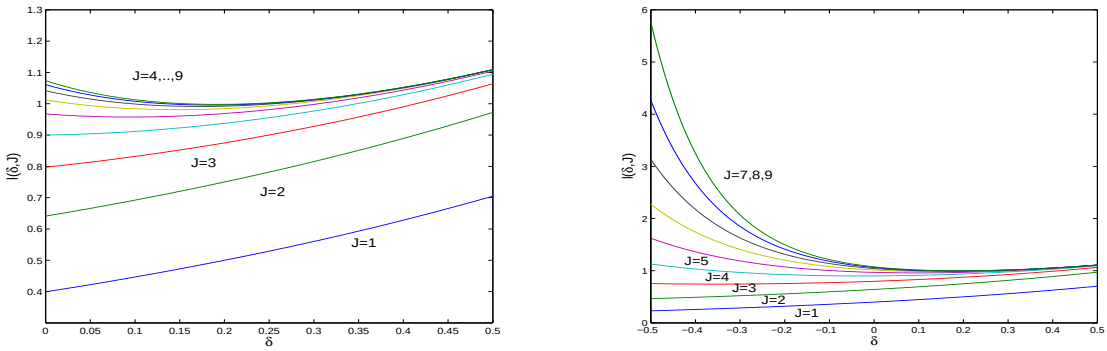

Figure 9: Function $l(\delta, J)$ for FARIMA$(0, 0.2, 0)$ model.

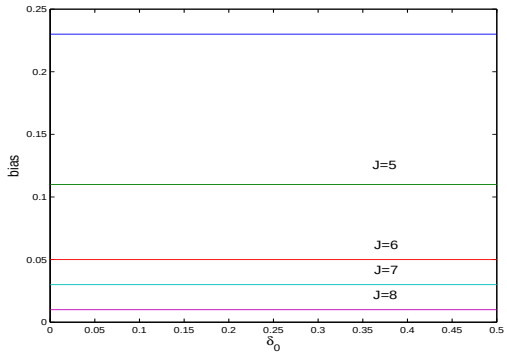

Figure 10: Bias for FARIMA$(0, \delta_0, 0)$ model.

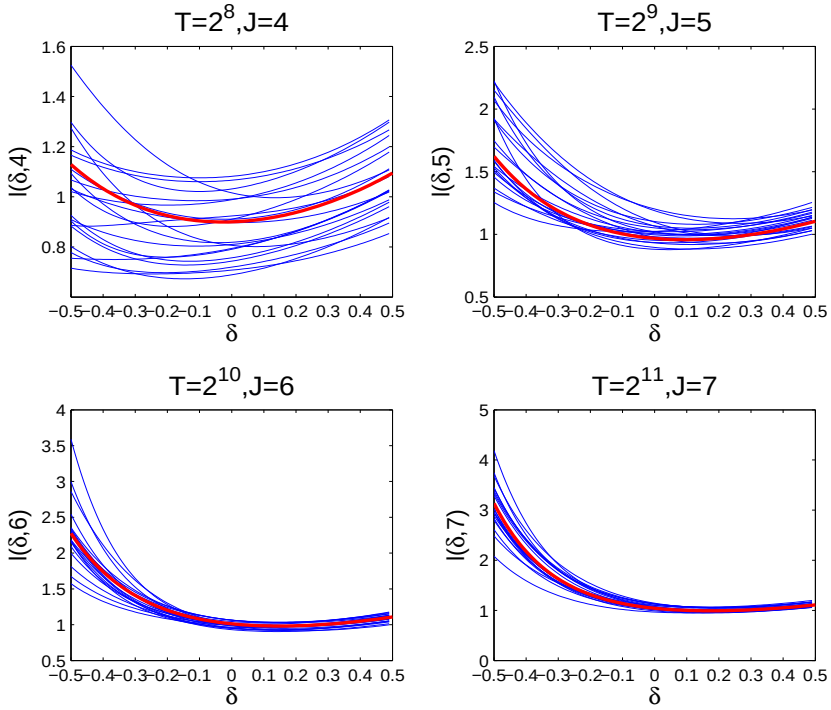

Figure 11: Functions $\widehat{l}(\delta, J)$ for several realizations of FARIMA$(0, 0.2, 0)$ series.

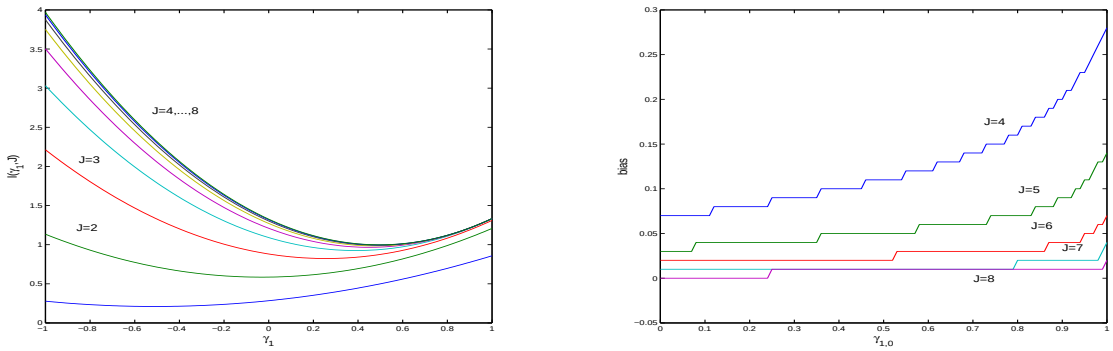

Figure 12: Function $l(\gamma_1, J)$ for AR(1) model with true parameter $\gamma_{1,0} = 0.5$ (left plot). Bias for AR(1) model with true parameter $\gamma_{1,0}$.

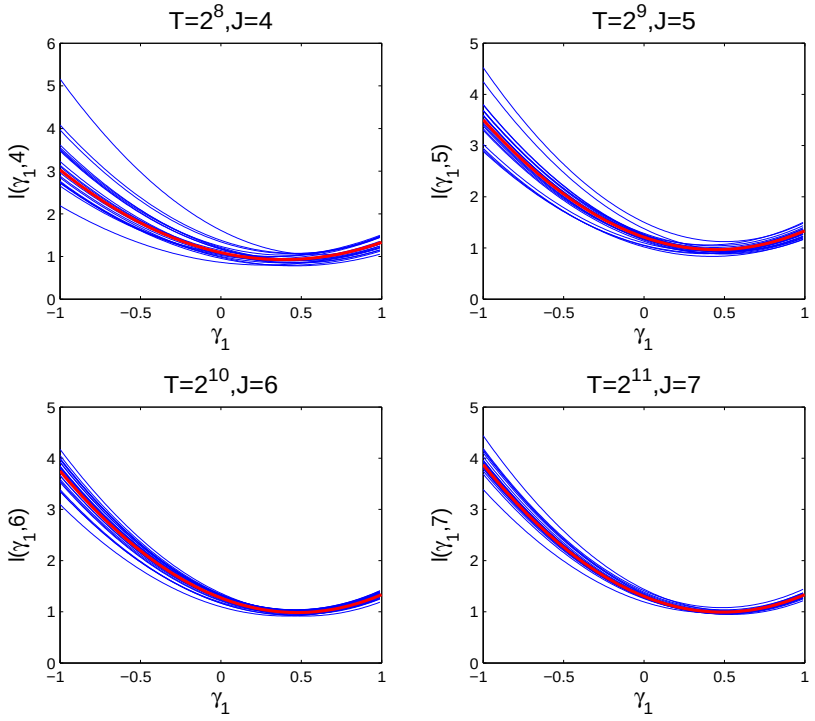

Figure 13: Functions $\widehat{l}(\gamma_1, J)$ for several realizations of AR(1) series with true parameter $\gamma_{1,0} = 0.5$.