Department of Economics

Wavelet Analysis and Denoising: New Tools for Economists

Iolanda Lo Cascio

Working Paper No. 600 May 2007 ISSN 1473-0278

WAVELET ANALYSIS AND DENOISING: NEW TOOLS FOR ECONOMISTS

Iolanda LO CASCIO

Department of Economics, Queen Mary College,
University of London,
Mile End Road, London E1 4NS, UK

Abstract

This paper surveys the techniques of wavelets analysis and the associated methods of denoising. The Discrete Wavelet Transform and its undecimated version, the Maximum Overlapping Discrete Wavelet Transform, are described.

The methods of wavelets analysis can be used show how the frequency content of the data varies with time. This allow us to pinpoint in time such events as major structural breaks.

The sparse nature of the wavelets representation also facilitates the process of noise reduction by nonlinear *wavelet shrinkage,* which can be used to reveal the underlying trends in economic data.

An application of these techniques to the UK real GDP (1873–2001) is described. The purpose of the analysis is to reveal the true structure of the data—including its local irregularities and abrupt changes—and the results are surprising.

KEY WORDS: Wavelets, Denoising, Structural Breaks, Trend Estimation.
JEL: C22, C14, C53.

1 Introduction: From Fourier to Wavelet Analysis

Econometric methodology is dominated by time-invariant linear parametric models and, as a consequence, it is rooted in the time-domain. However,

the frequency-domain approach sometimes provides a more insightful representation of econometric data by decomposing it into sinusoidal components of various frequencies, which have intensities that vary across the frequency spectrum. However neither mode of analysis is sufficiently flexible to cater for truly evolving phenomena, which may be subject to gradual drifts or to occasional abrupt changes.

This inadequacy of Classical Fourier analysis in the face of evolutionary phenomena is a consequence of the assumption that the intensities are approximately constant through time. The analysis has a global nature; and, whereas they are localised in frequency, the sines and cosines functions are not localised in time but extend over the entire real line. This feature makes Fourier methods ineffective in analysing signals containing local irregularities, such as spikes or discontinuities.

Another drawback, when transforming to the frequency domain, is the loss of the time information which would indicate the incidence of events and their duration. Such time information is encoded in the phase spectrum from which it cannot be easily deciphered. The amplitude spectrum or, equivalently, the power spectrum, which is the primary object of a Fourier analysis, contains no time information. It tells what a signal contains at each frequency, but it does not indicate when these frequency components were emitted. For example, a particular frequency might occur continuously or over a very short time interval. However, a classical Fourier analysis cannot distinguish between these two cases.

The short-time Fourier transform (Gabor, 1946), also known as windowed Fourier transform, was the first attempt to achieve a resolution in both time and frequency. The idea was to study the frequency content of a signal, segment by segment, by multiplying it by a windowing function, usually a unit box, which is zero outside a finite interval and whose size is fixed throughout the sample.

In such a way, a part of the data is isolated which can then be subjected to a Fourier transform. When one segment of the signal has been analysed, the window is slid along the signal so as to analyse further segments, until the entire signal has been covered. The technique has two main drawbacks.

First, fixing the size of the window involves accepting compromises. If one is interested in the high-frequency components, then a small window is appropriate; but then is impossible to get information about the low frequency components of the signal. Conversely, a large window allows one to analyse only the coarse features of the data. Also, analysing the data with various window widths, in order to capture the features existing at different frequencies, will generate a considerable amount of data compared with what is entailed in a simple Fourier transform of the data sequence.

To achieve additional flexibility, methods are called for that can combine both temporal and frequency information. An answer is provided by *wavelets*.

Often, they have finite supports, which means that they exist only within a given interval of time. If they have infinite supports, then their effects tend to disappear as the distance on either side from a central time point increases.

The term *wavelet* has been introduced by Grossman and Morlet (1984) to describe a square integrable function whose translation and dilation form a basis of $L^2(\mathcal{R})$. All wavelet basis functions are *self-similar*, and they differ from one another only by translations along the time axis and changes of scale.

Wavelet analysis can be considered as a windowing technique with variable-sized regions. Instead of combining sines and cosines at different frequencies which extend throughout time from minus to plus infinity, wavelet analysis decomposes a signal into shifted (translated) and scaled (dilated or compressed) versions of a *mother* wavelet. In this way, discontinuities in signals, can be located in time by means of very short basis functions, whereas accurate discrimination amongst low frequencies can be achieved via highly dilated low-frequency basis functions.

Throughout the 20th century, scientists in many areas of research have been trying to overcome the limitations of the classical Fourier Analysis. They have described, in their own languages, their particular versions of what has amounted to a wavelets analysis. Under different names, the *pyramid algorithm* (Mallat, 1987, 1989) used in image processing, the *subband coding* (Vetterli and Kovacevic, 1995) of signal processing, the *quadrature mirror filters* of digital speech processing (Vaidyanathan, 1990;1993) have bee addressing the same problem.

The manner in which wavelet analysis is usually presented is, therefore, a consequence of the field in which it has originated. There have been two main sources: the functional analysis, which gives strong emphasis to continuous wavelet analysis and the theory of digital signal processing which approaches the subject as an application of the techniques of multi-rate filtering.

A common belief has been that, if the aim of the analysis is the perfect reconstruction of a signal, then a *Continuous Wavelet Transform* (CWT) must have been adopted. The signal is analysed at all possible resolutions and the wavelets are displaced by all values, and not only by integer translations. Because the parameters governing the dilation and the translation are real numbers, the outcome is an infinite number of wavelet coefficients which retain a mass of redundant information.

In fact, in the CWT the wavelets overlap each other, so that most of the information encoded by one wavelet is also encoded by its neighbours.

The signal is said to be over-sampled. That redundancy can be avoided by employing orthogonal transforms of the data that allow for perfect reconstruction (Hubbard, 1998). The use of an orthogonal basis implies the use of a *Discrete Wavelet Transform* whereas a non-orthogonal function can be used with either the Discrete or the Continuous Wavelet Transform.

In the process of selecting a wavelet, one is faced with two kinds of choices, The first concerns the system of representation which may be discrete, continuous or semi-continuous [1]. The second choice concerns the properties of the wavelets themselves. These include the number of vanishing moments and the degree of regularity, which will be explained later [2].

A Continuous Wavelet function centred at τ and with scale s can be expressed as

$$\psi_{s,\tau} = \frac{1}{\sqrt{|s|}}\psi(\frac{t-\tau}{s}) \tag{1}$$

where $\tau, s \in \mathbf{R}$ and $s > 0$. The continuous wavelet transform can then be thought of as a measuring the correlation between the signal and the wavelet function. In mathematical terms, it performs the following convolution:

$$\mathcal{CWT}(s,\tau) = \frac{1}{\sqrt{|s|}}\int_{-\infty}^{\infty} f(t)\psi(\frac{t-\tau}{s})dt \tag{2}$$

The transform is a function of scale and translation and it will be large valued if the signal and the wavelet match in shape at a specific scale and location. If not, the transform will be small.

The so-called *admissibility condition* must be satisfied by the wavelet in order to reconstruct the signal from its transform. This is

$$C_\psi = \int_{-\infty}^{\infty} \frac{|\psi(\omega)|^2}{\omega}d\omega < \infty \tag{3}$$

where $\psi(\omega)$ is the Fourier transform of $\psi(t)$. That condition implies that $\psi(0) = 0$—$\psi(\omega)$ goes to zero as ω goes to zero—and that

$$\int_{-\infty}^{\infty} \psi(t)dt = 0 \tag{4}$$

[1] The semi-continuous transform is a translation-invariant transform in which the scale parameter, but not the translation parameter, is discreteised and sampled. It is a highly redundant transform for which a fast algorithm, known as *algorithm a trou* (algorithm with holes) has been developed by Holsschneider at al., 1989.

[2] The first of these two choices is affected by the *Shannon-Whittaker Sampling Theorem*. According to that theorem, any continuous and band limited function with frequency content in the interval $[0, \pi]$ can be reconstituted from its sampled ordinates by using a basis of *sinc* functions at unit displacements.

In other words, the mother wavelet is an oscillatory function of limited duration and with average value equal to zero. It is a well localised function both in time and frequency and it decays rapidly in both. If a wavelet satisfies the above condition, then, the transform is left-invertible and the signal can be reconstructed (synthesised) from its wavelet coefficients in the following way

$$f(t) = \frac{1}{C_\psi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathcal{W}(s,\tau)\psi_{\tau,s}(t) \frac{ds d\tau}{s^2} \tag{5}$$

The *CWT* is well suited to detecting singularities or sudden localised changes. Singularities are mathematically expressed in terms of the so called *Lipschitz* exponent, α, which measures the degree of regularity of a signal. In the resulting time-scale decomposition, the lines that link the maxima of the moduli of the coefficients at the different scales converge to the singularities of the signal. By proper manipulation, it is possible to estimate the degree of regularity. Jumps, for example, can then be modelled as points of local Lipschitz regularity with a very small exponent α.

In the case of band-limited functions, it is possible to think of the Discrete Wavelet Transform as a critical sampling of the Continuous Time Wavelet Transform. This is a consequence of the Shannon-Whittaker *sampling theorem* (Shannon and Weaver, 1964). More often, the DWT is derived independently from the CWT through an appropriate choice of the filter coefficients. For the dyadic case, which dominates the wavelet literature, $s = 2^j$ and $\tau = k \cdot 2^j$.

Most of the accounts of wavelet analysis start with the Haar wavelet. The Haar wavelet is an excellent tool for teaching purposes but less useful for most applications because of its discontinuity. On the other hand, it is well suited to analysing and synthesising time series with sharp jumps or steps. Its compact support makes it highly localised in time but dispersed in the frequency domain. It turns out to be the only orthogonal wavelet that has symmetric analysis and synthesis filters.

The popularity of compact support wavelets, like the ones belonging to the Daubechies' family (1992), is mainly due to their relation to the dyadic multiresolution analysis which dominates wavelet research.[3] The frequency bands of a multiresolution analysis are obtained by dividing the frequency range into successively smaller intervals in a descent from the high frequencies to the low frequencies, each band having half the width of its predecessor.

[3]Among the infinite support wavelets are, on the other hand, *Gaussian* wavelets, the *Mexican Hat*, *Morlet* and *Meyer*' wavelet. Gaussian wavelets are obtained from the derivatives of the Gaussian function. Mexican Hat is similar to the Gaussian $G(2)$. Among these wavelets, only Meyer wavelet has a scaling function.

In the process, the dispersion of the wavelets that constitute the bases of the frequency subspaces is doubled, and the number of bases wavelets is halved.

The main disadvantage of the conventional dyadic wavelet analysis is the restriction on the sample size N, which has to be a power of 2, and the restriction on the location of the frequency and time bands.

This sort of rigidly structured analysis is not always appropriate for statistical data analysis, for the structure to be investigated might not fall neatly into dyadic time and frequency bands. These difficulties can be overcame by pursuing a *Mixed-Radix Wave-Packet Analysis*—see Pollock and Lo Cascio, (2006)—which can be applied to sample sizes with arbitrary factorisations and can be seen as an extension of the dyadic wave-packet analysis (Wickerhauser, 1994).

Some flexibility can be achieved if one is prepared to accept an incomplete DWT by identifying a scale beyond which a wavelet analysis is no longer of interest. (The stage at which the dyadic decomposition must cease is when no further divisions of the sample size by higher powers of two are available.) A partial DWT of level J_0 allows us to relax the restriction that $N = 2^J$ for some J and to replace it with the condition that N be an integer multiple of 2^{J_0}.

Another transform variant is the *Maximum Overlapping Discrete Wavelet Transform (MODWT)* (known under many different names, i.e., *Stationary Wavelet Transform* (Nason and Silverman, 1995), *Translation Invariant DWT* (Coifman and Donoho, 1995), *Time Invariant DWT* (Pesquet et al. 1996), *Non-Decimated DWT* (Bruce and Gao, 1996)).

This represents an attempt to generate a transform which is not sensitive to the choice of the starting point for the data series. In order to eliminate that sensitivity, one does not subsample the filtered output at each stage of the pyramid algorithm. As a consequence, the number of coefficients generated at any stage of the decomposition are in number equal to the sample size, T, instead that equal to $T/2^j$.

However, even though the MODWT is an energy-preserving transform (the variance of the signal is perfectly captured by the variance of the coefficients, but, contrary to the DWT case not by the variance of the components), the MODWT components are not orthogonal.

An important feature of the MODWT is that, besides handling any sample size, the detail and smooth coefficients of the multiresolution analysis are associated with linear phase filters. The consequence is that it is possible to align the features of the original time series with those of the multiresolution analysis. Hence the redundancy of the MODWT turns out to be a useful which we exploit it in the context of denoising schemes.

Translation invariance can also be achieved by means of a semi-continuous

transform, in which only the scale parameter (but not the translation parameter) is discreteised and sampled. For these highly redundant cases a fast algorithm is available, known as the *algorithm a trous* and developed by Holschneider, Kronland-Martinet, Morlet and Tchamitchian (1989).

In many applications in signal (and image) processing, the data points Y_i, are assumed to represent noisy observations on a smooth underlying function $g(t)$ defined, for convenience, as a function of $t \in [0, 1]$, with the unit interval as its domain. Thus

$$Y_i = g(t_i) + \varepsilon_i \quad i = 1, \ldots, n \tag{6}$$

where $n = 2^J$, $t_i = i/n$ and $\epsilon \sim i.i.d \quad \mathcal{N}(0, \sigma^2)$, and the question is how to strip away the noise ϵ. Wavelet theory has provided statisticians with new and powerful techniques for *nonparametric* function estimation.

The fact that the signal energy is often mostly concentrated in only few big coefficients (i.e., *sparsity* of the DWT) together with the natural ability of wavelets to adapt to local variations is the basis for noise reduction by nonlinear *wavelet shrinkage* and *wavelet thresholding* estimators (Donoho and Johnstone (1994) and Donoho *et al.*, (1995)).

A review of these techniques is given in section 3. Among their advantages is the fact that they can be applied without prior knowledge of the family $\{g_\theta; \theta \in \Theta\}$— Θ being an infinite dimensional parameter set—to which g may belong.

A linear approach to wavelet thresholding is also available and an exposition is given by Antoniadis (1996); its main limitation is that, it is not well suited to handling spatially or temporally inhomogeneous functions with a low degree of regularity.

The structure of the paper is as follows. In section 2, a lengthy and detailed account of wavelet multiresolution analysis is given, together with concepts and theorems that relate to denoising. Section 3 describes wavelet denoising procedures which are then applied in Section 4 to the wavelet transform of the UK log real annual GDP—available from 1873 to 2001—for the purposes of detecting structural breaks and of estimating the trend.

2 Dyadic Multiresolution Analysis

Whatever the field in which it has originated, wavelet analysis is dominated by the so called *Dyadic Multiresolution Analysis*. The concept of *multiresolution analysis* was introduced by Mallat (1987,1989) and Meyer (1992) and it provides a natural framework in which to understand the orthonormal and compactly supported wavelet bases.

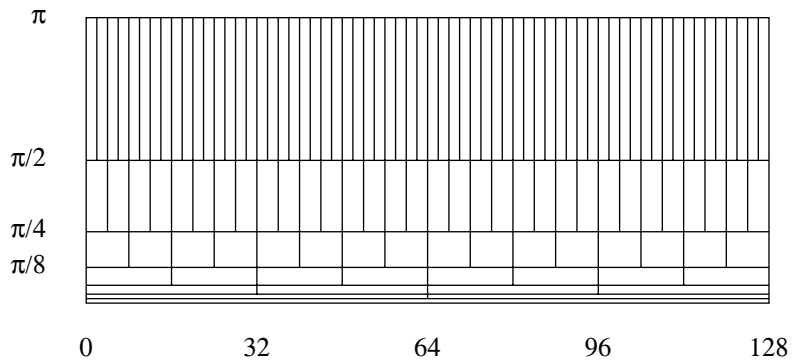

Figure 1: The partitioning of the time-frequency plane according to a dyadic multiresolution analysis of a data sequence of $T = 128 = 2^7$ points.

The purpose of a mutliresolution analysis is to decompose a data sequence into a set of wavelets associated with the cells of a grid that partitions the time-frequency plane[4]. The partitioning is known as *Mosaic Diagram* and it is illustrated by Figure 1, which is for a sample size of $T = 128 = 2^7$. In the conventional dyadic multiresolution analysis, the sample size $T = 2^n$ is restricted to be a power of two and the scale resolution is $s = 2^j$. Here, the integer j also refers to the iterative stage of an algorithm of wavelet analysis.

In Figure 1, the height of a cell corresponds to the bandwidth in the frequency domain whereas its width represents the temporal duration. The highest observable frequency in the sampled data is the Nyquist frequency of π radians per observation interval, which is the upper limit of the range of frequencies detectable via regular sampling.

Centred on each cell there is a wavelet. Each band contains a succession of wavelets that have a common frequency range and an equal temporal duration but different amplitudes. These wavelets form a set of orthogonal functions and, in case of finite samples, there are as many wavelets as there are data points. From the mosaic, it is clear that high frequency wavelets are tightly packed in time and that they cover a wide frequency range.

As we descend the frequency scale, the wavelets become increasingly dispersed in time but their resolution in frequency is enhanced. At the lower end of the frequency spectrum, wavelets are widely dispersed in time but they subsist within narrow frequency bands. In descending the frequency range, at each scale, the frequency band is halved and the temporal dispersion is

[4]As pointed out by Priestley (1995), it is more appropriate to think in terms of level of resolution rather than in terms of frequency, given that the term *frequency* refers only to sines, cosines and exponential of an imaginary argument.

doubled.

However, wavelets are never completely confined to a cell. In all but the case of the Haar wavelets, they break the bounds that demarcate the nominal time intervals and, in all but the case of the Shannon wavelets, they break the bounds of the nominal frequency intervals. Therefore, in general, we cannot say that a succession of basis wavelets spans a nominal frequency interval, and we shall say, instead that it corresponds to the interval.

Conventional dyadic multiresolution analysis considers a succession of frequency intervals in the form of $(\pi/2^j, \pi/2^{j-1}); j = 1, 2, \ldots, n$, whose bandwidths are halved repeatedly in descending from high frequencies to low frequencies. By the jth round, there will be j wavelet bands and one accompanying scaling function band. The set $\{\psi_j(t-2^j k); t, k \in I\}$ containing $T/2^j$ mutually orthogonal wavelets, separated from each other by 2^j points, that correspond to the nominal frequency range $[\pi/2^j, \pi/2^{j-1})$, will be complemented by the set $\{\phi_j(t-2^j k); t, k \in I\}$ of scaling functions that correspond to the lower nominal frequency range $[0, \pi/2^j)$. These wavelets and scaling functions form bases for the corresponding details and approximation spaces, V_j and W_j respectively.

More formally, a sequence $\{V_j\}_{j\in\mathcal{Z}}$ of closed subspaces of $L^2(\mathcal{R})$ is a multiresolution approximation if and only if the following properties under (7)–(12) are satisfied:

$$\forall j, k \in \mathcal{Z}, \quad f(t) \in V_j \Leftrightarrow f(t-2^j k) \in V_j \tag{7}$$

This condition declares that, if the function $f(t)$ belongs to the space V_j, such that it can be expressed in terms of the basis of that space, then the same must be true of its translated version $f(t-2^j k)$.

$$\forall j \in \mathcal{Z}, \quad \mathrm{V}_{j+1} \subset V_j \tag{8}$$

$$\forall j \in \mathcal{Z}, \quad f(t) \in V_j \Leftrightarrow f\left(\frac{t}{2}\right) \in V_{j+1} \tag{9}$$

The two conditions above assert that every subspace V_j contains a subspace V_{j+1} comprising the functions that have twice the dilations of those of V_j.

$$\bigcap_{j\in\mathcal{Z}} V_j = \{0\} \tag{10}$$

$$\bar{\bigcup_{j\in\mathcal{Z}}} V_j = L^2(\mathcal{R}) \tag{11}$$

The condition (10) and (11) indicate that the direct sum of the subspaces constitutes a space of square-integrable functions. Finally, there must be a space that contains all levels of resolution up to finest; and so it must be declared that there exists $\phi \in V_0$ such that

$$\{\phi_{(0)}(t-k); t, k \in \mathcal{Z}\} \quad \text{is an orthonormal basis in } V_0 \tag{12}$$

Here V_0, corresponds to the nominal frequency range $[0, \pi)$ in the Mosaic diagram.

The basis for the subspace V_j is a set of orthonormal translated functions; each of these function sets is a fixed dilation of the scaling function, $\{\phi_{j,k}; j, k \in \mathcal{Z}\}$. Let W_j be the orthogonal complement of V_j in V_{j-1}. Then we have

$$V_{j-1} = V_j \oplus W_j \tag{13}$$

and

$$W_j \perp W_{j'} \quad \text{if} \quad j \neq j' \tag{14}$$

The wavelet basis $\{\psi_{j,k}(t) = 2^{-j/2}\psi(2^{-j}t - k); j, k \in \mathcal{Z}\}$ forms an orthonormal basis in W_j. W_j is the space containing the *detail* information necessary to go from an approximation with resolution 2^{j-1} to a coarser approximation with resolution 2^j. Thus, if a dyadic decomposition is pursued from the resolution level j down to the level $n > j$, there will be the following direct sum of subspaces:

$$V_j = V_n \oplus W_n \oplus W_{n-1} \oplus \cdots \oplus W_{j+1} \tag{15}$$

The outcome is that, whenever a collection of closed subset satisfies the MRA definition, there exists an orthonormal wavelet basis $\{\psi_{j,k}; j, k \in \mathcal{Z}\}$ of $L^2(\mathcal{R})$, $\psi_{j,k}(t) = 2^{-j/2}\psi(2^{-j}t - k)$ such that any function $f(t) \in L^2(\mathcal{R})$ can be represented as a sequence of projections onto father (scaling) and mother (wavelets) functions, namely:

$$f(t) = \sum_k s_{J,k}\phi_{J,k}(t) + \sum_j \sum_k d_{j,k}\psi_{j,k} \tag{16}$$

Here, J is the furthest level of decomposition and $s_{j,k}$ and $d_{j,k}$ are the following projections of $f(t)$ on the bases $\phi_{j,k}$ and $\psi_{j,k}$ respectively:

$$\begin{aligned} s_{j,k} &= \int f(t)\phi_{j,k}(t)dt \\ d_{j,k} &= \int f(t)\psi_{j,k}(t)dt \end{aligned} \tag{17}$$

The signal can then be written as a set of orthogonal components at resolutions 1 to J:

$$f(t) = S_J + D_J + D_{J-1} + \ldots + D_1 \tag{18}$$

A direct consequence of the definition of multiresolution is that, given that the original scaling function ϕ and the wavelet function ψ are both members of V_0, they also belong to V_{-1}. As $\{\sqrt{2}\phi(2t-k)\}$ is an orthonormal basis for V_{-1}, it follows that both ϕ and ψ can be expressed as a linear combination of $\{\sqrt{2}\phi(2t-k)\}$. In other words:

$$\phi(t) = \sqrt{2}\sum_k h_k \phi(2t-k) \tag{19}$$

$$\psi(t) = \sqrt{2}\sum_k g_k \phi(2t-k) \tag{20}$$

These two equations are known as *dilation* and *wavelet* equation respectively, but also as *refinement* or *two-scale* equations, and they are used to move from the continuous wavelet transform to the discrete wavelet transform. They express a wavelet in one frequency band and a scaling function in the band below as a linear combination of the more densely packed and less dispersed scaling functions. In the process, they generate the *wavelet coefficients* $\{g_k\}$ and the *scaling-function coefficients* $\{h_k\}$ that are the basis of the discrete transform.

In the case of compactly supported wavelets, ψ assumes non-zero values only on a bounded interval; the two coefficient sequences $\{h_k\}$ and $\{g_k\}$ contain the same finite number L of non-zero terms.

A somewhat curious fact is that, it is possible to analyze a signal into its components at different scales without ever referring to wavelet and scaling functions. All we need are *filters*. In the two equations above, $\{h_k\}$ and $\{g_k\}$ behave as lowpass and highpass filters respectively and they are such that:

$$g_k = (-1)^{k+1} h_{L-k-1} \quad k = 0, \ldots, L-1 \tag{21}$$

where L is the length of the filter. A set of conditions must be satisfied by the coefficients before they can represent an orthonormal wavelet. A wavelet filter must sum to zero, have unit energy and be orthogonal to its even shift; in other words:

$$\sum_{k=0}^{K-1} g_k = 0 \tag{22}$$

$$\sum_{k=0}^{K-1} g_k^2 = 1 \tag{23}$$

$$\sum_{k=0}^{K-1} g_k g_{k+2l} = \sum_{k=-\infty}^{\infty} g_k g_{k+2l} = 0 \tag{24}$$

for all $l \neq 0$, where the integer $2l$ is the index of the even displacement of the wavelet.

The two filters $\{h_k\}$ and $\{g_k\}$ are also called *Quadrature Mirror filters* (Smith and Barnwell, 1986; Croisiere, Esteban and Galand, 1976).They play the main role in a two-channel filter bank, where, in the analysis channel, a signal of length $T = 2^n$ is passed in parallel through lowpass and highpass filters. The output is downsampled by 2, so as to guarantee the non redundancy in the information contained in the filtered output. The downsampled output from the first highpass filter is the set of wavelet coefficients at the highest time resolution. The downsampled output from the first lowpass filter is, instead, the input for the next lowpass/highpass filtering and decimation process. The iteration continues throughout the n levels of a wavelet decomposition.

The synthesis channel aims at reconstructing the original signal from the coefficients produced in the analysis channel. Prior to reassembling them, and starting from the final average and the wavelet coefficients at the highest level of decomposition, zeros are interpolated between the elements of the component signals to replace the discarded elements. These are then replaced by estimates of the missing values obtained by passing the upsampled sequences through separate highpass and lowpass filters. Then the two signals are added together to recreate the original signal.

An important condition, which must be fulfilled if perfect reconstruction is to be to achieved, is that wavelets that are displaced relative to each other by multiples of 2 sampling intervals should be mutually orthogonal, in which case they form an orthonormal basis of the corresponding function space.

The orthogonality of the wavelets will be reflected in the autocorrelation function $\rho(\tau) = \{\rho_\tau; \tau = 0, \pm 1, \pm 2, \ldots\}$ of the sequence of wavelets coefficients. There should be $\rho_0 = 1$; and $\rho_\tau = 0$; for $\tau \in \{\pm 2, \pm 4, \ldots\}$. Therefore, the subsampled sequence $\rho(2\tau) = \{\rho_{2\tau}; \tau = 0, \pm 1, \pm 2, \ldots\}$ should have characteristics akin to those of the autocovariance function of a discrete white-noise process.

In particular, the Fourier transform of the subsampled sequence should be a constant function defined over the interval $[-\pi, \pi]$. If the Fourier transform of the sequence $\rho(\tau)$ is denoted by $\mathcal{G}(\omega)$, then the Fourier transform of the

subsequence will be

$$\mathcal{G}(\omega) + \mathcal{G}(\omega + \pi) = 2 \quad \text{for all} \quad \omega \tag{25}$$

This is a consequence of the process of wrapping in the frequency domain, or of "aliasing", that accompanies the process of subsampling in the time domain.

For a proof of this result, consider the z-transform of the autocovariance function. Taking account of the fact that the autocovariances with an even index are zero-valued, this becomes

$$\mathcal{G}(z) = \{1 + \rho_1(z + z^{-1}) + \rho_2(z^3 + z^{-3}) + \cdots\} \tag{26}$$

Now let

$$z = e^{-i\omega} \quad \text{and} \quad s = e^{-i(\omega+\pi)} = e^{-i\pi}e^{-i\omega} = -e^{-i\omega} \tag{27}$$

Then we have

$$s + s^{-1} = -(z + z^{-1}), \quad s^3 + s^{-3} = (z^3 + z^{-3}) \quad \text{etc.} \tag{28}$$

and it is clear that

$$\mathcal{G}(z) + \mathcal{G}(s) = \mathcal{G}(\omega) + \mathcal{G}(\omega + \pi) = 2. \tag{29}$$

Due to the QMF relation, that follows:

$$\mathcal{H}(\omega) = \mathcal{G}(\pi - \omega) \tag{25}$$

where $\mathcal{H}(\omega)$ is the squared gain of the lowpass filter. As a consequence, for all ω, it is possible to write:

$$\mathcal{H}(\omega) + \mathcal{H}(\omega + \pi) = 2 \quad \text{or} \quad \mathcal{H}(\omega) + \mathcal{G}(\omega) = 2 \tag{26}$$

The Multiresolution Analysis is implemented via the *pyramid algorithm*, which has been introduced in the context of wavelets by Mallat (1989). The iterative nature of the algorithm is fully encapsulated in the frequency domain version of the dilation equation:

$$\hat{\phi}(\omega) = H\left(\frac{\omega}{2}\right)\hat{\phi}\left(\frac{\omega}{2}\right) \tag{27}$$

which results from taking the Fourier transform of both sides of the dilation equation (19) after replacing $2t - k$ with u. Here $H(\omega) = \frac{1}{\sqrt{2}}\sum_{k=0}^{L-1} h_k e^{i\omega k}$

is a 2π periodic function and $\phi(\omega) = \frac{1}{\sqrt{2}} \int_{-\infty}^{\infty} \phi(t) e^{i\omega t} dt$. From the wavelet equation (20) we have likewise:

$$\hat{\psi}(\omega) = G\left(\frac{\omega}{2}\right)\hat{\phi}\left(\frac{\omega}{2}\right) \tag{28}$$

where $G(\omega) = 2^{-1/2} \sum_{k=0}^{L-1} g_k e^{i\omega k} = e^{-i\omega} \bar{H}_0(\omega + \pi)$.

The second step of iteration produces

$$\hat{\phi}(\omega) = H\left(\frac{\omega}{2}\right)\left[H\left(\frac{\omega}{4}\right)\hat{\phi}\left(\frac{\omega}{4}\right)\right] \tag{29}$$

After the j iteration,[5] we have:

$$\hat{\phi}(\omega) = H\left(\frac{\omega}{2}\right)H\left(\frac{\omega}{4}\right)\ldots H\left(\frac{\omega}{2^j}\right)\hat{\phi}\left(\frac{\omega}{2^j}\right) = \prod_{j=1}^{\infty} H\left(\frac{\omega}{2^j}\right)\hat{\phi}\left(\frac{\omega}{2^j}\right) \tag{30}$$

As $j \to \infty$ the final factor goes to 1, therefore, it is possible to write

$$\hat{\phi}(\omega) = \prod_{j=1}^{\infty} H\left(\frac{\omega}{2^j}\right) \tag{31}$$

If the factors approach 1 as j goes to infinity, then, the infinite product converges. In particular, the natural requirement on $H(\omega)$, for the decay of $\hat{\phi}(\omega)$, is that $H(\pi) = 0$ $H(0) = 1$. That can easily be seen from the fact that, given the periodicity of $H(\omega)$, we have $H(2\pi) = H(4\pi) = \ldots = 1$. As a consequence, the values of, for example, $\hat{\phi}(2\omega)$, $\hat{\phi}(4\omega)$, $\hat{\phi}(8\omega)$ are all equal. If $H(\pi) = 0$, then, these values will all equal zero—(i.e. $\hat{\phi}(2\pi) = H(\pi)\hat{\phi}(\pi)$).

Notwithstanding the limitation of using finite response impulse filters (their are not band limited, therefore, they admit substantial leakage across the frequency range), the pyramid algorithm reaches its greatest efficiency when it is used in conjunction with wavelets, such as the Daubechies', which are compactly supported in the time-domain. In particular, it owes its efficiency to the manner in which each bandpass can be constructed from elementary component filters (Pollock and Lo Cascio, 2006) and is faster ($\mathcal{O}(N)$) than the Fast Fourier Transform algorithm $\mathcal{O}(N \log_2 N)$. It can be recognised that, once the filter $H = \{h_k\}_{k=0}^{L-1}$ has been chosen, it completely determines ϕ and ψ. Indeed, there focus of the DWT is upon the computation of wavelet amplitude coefficients rather than upon the wavelets themselves which may never need to be represented explicitly.

[5] j is also the degree of the polynomial that can be reconstructed.

The ability of the discrete wavelet transform to encode a signal with only a few coefficients different from zero, which is its so-called sparsity property, and the tendency for most of the fine-scale wavelet coefficients to be very small, depends on the number of vanishing moments, the degree of regularity and the length of the support. These features are strictly related to each other.

$$\mathcal{N}_m \equiv \int_{-\infty}^{\infty} t^m \psi(t) dt = 0 \quad 0 \leq m < r \tag{32}$$

or if its scaling function can generate polynomials of degree up to r. The vanishing order corresponds to the number of zeros, r, at $\omega = \pi$ for the frequency response function of the lowpass filter and the number of zeros at $\omega = 0$ for the frequency response function of the highpass filter[6]. A wavelet with many vanishing moments is a rapidly oscillating function which gives small coefficients when it is used to analyze low frequencies. As a consequence, it produces a multiresolution analysis characterized by a smoother transition from one approximation to the next one. That is the reason why Daubechies (1988, 1992) decided to incorporate the vanishing moments and the regularity requirements into her filter design.

The continuity of ψ and of $r-1$ of its derivatives is a pre-requisite for the existence of r vanishing moments. Given that regularity is concerned with differentiability, vanishing moments and regularity are related to each other. A signal is *regular* if it can be approximated by a polynomial. On the other hand, non-smooth basis functions (like the *Coiflets*) introduce artificial discontinuity which creates spurious artifacts in the reconstructed signal. For that reason, one must be cautious in choosing the appropriate basis keeping well in mind the trade-off between regularity and localisation (in terms of support, the smoother is a function the less local it is.). Daubechies proved that a wavelet with r vanishing moments must have a support at least of $2r-1$. As a consequence, an increase in the number of vanishing moment increases the burden of computation but, more importantly, makes more difficult to detect local features in the signal (i.e., to locate *singularities*)[7].

[6]The vanishing moment property implies that the Fourier transform of $\psi(.)$ and of certain number of derivatives are zero at zero frequency. That can be seen by considering the m-th order derivative $\psi^m(\omega) = (-i\omega)^m \int_{-\infty}^{+\infty} t^m \psi(t) e^{-i\omega t} dt$. If $\mathcal{N}_m = 0$, then, $\psi^m(0) = 0$.

[7]A more accurate reconstruction of the signal can be done by using biorthogonal filters; good accounts are given in Chui (1997), Burrus *et al.* (1998) amongst others. In contrast to the orthogonal case where the analysis and synthesis wavelets are the same (filters are time reversal of each others) and bear the same characteristics in terms of regularity and

Singularities, abrupt changes and frequency transients, in time series analysis, are often more interesting than smooth and regular behavior. The wavelet transform is an excellent tool for its ability to characterize the local regularity of signals. Singularities and edges are detected with accuracy through an appropriate choice of the basis function.

Local irregularities are reflected within a wavelet analysis by the presence of high coefficients associated with fine scales, or high frequency, wavelets. Mallat, amongst others, has made a formal analysis of the phenomenon which makes reference to the concept of *Lipischtz* regularity. It is also common to look at the asymptotic decay of its Fourier transform, $\hat{f}(\omega)$. In particular, a function f is bounded and uniformly *Lipschitz-*α over R if[8]

$$\int_{-\infty}^{\infty} |\hat{f}(\omega)|(1+|\omega|^{\alpha})d\omega < \infty \tag{33}$$

This is still a global regularity condition, given the nature of the Fourier transform. In fact, it does not tell us if the function is locally more regular at a particular point. A regular function can be approximated by a polynomial; then, for any $\alpha > 0$, a function is *pointwise Lipschitz* α at v, if and only if there exists a constant K and a polynomial of degree m, $m < \alpha \leq m+1$ such that

$$|f(t) - P_v(t)| \leq K|t-v|^{\alpha} \quad \forall \mathrm{t} \in R \tag{34}$$

[9] The Lipschitz regularity of f at v or over $[a,b]$ is the sup of α such that f is Lipschitz-α. Moreover, a function is singular at v if it is not Lipschitz-1 at v (A function that is continuously differentiable at a point is Lipschitz-1 at this point.).

An important theorem by Holschneider and Tchamitchian (1990), in some cases attributed to Jaffard (1989), relates the pointwise regularity of a signal to the decay of its wavelet transform' modulus. It states that, if $f \in L^2(R)$ is a *Lipschitz-*$\alpha \leq m$ at v, then, there exists A such that

vanishing moments, in the biorthogonal case more flexibility can be reached by assigning the two offsetting properties to two different functions. In particular, the vanishing order refers to the analysis wavelet; the synthesis wavelet needs instead to be regular so as to minimize reconstruction artifacts.

[8] A proof can be found in Mallat ()

[9] For the case of sharp cusp detection (Wang, 1995, 1999), $0 \leq \alpha < 1$. The case of $\alpha = 0$ corresponds to a jump at v. If $\alpha \geq 1$, $f(t)$ is continuous and differentiable. If $0 < \alpha < 1$ $f(t)$ is continuous but not differentiable. For $-1 < \alpha \leq 0$ $f(t)$ is discontinuous and not differentiable. If $\alpha \leq -1$ $f(t)$ is no longer locally integrable.

$$\forall(\tau,s)\in R\times R^{+},|\,W(\tau,s)|\leq As^{\alpha+\frac{1}{2}}\left(1+\left|\tfrac{\tau-v}{s}\right|^{\alpha}\right) \tag{35}$$

Conversely, if $\alpha < m$ is not an integer and there exists A and $\alpha^{'} < \alpha$ such that

$$\forall(\tau,s)\in R\times R^{+},|\,W(\tau,s)|\leq As^{\alpha+\frac{1}{2}}\left(1+\left|\tfrac{\tau-v}{s}\right|^{\alpha^{'}}\right) \tag{36}$$

then, f is *Lipschitz-α* at v.

In other words, if f is differentiable at v, its wavelet transform has order $s^{3/2}$ as s tends to zero; if f has an α-cusp at v, the maximum of its transform over a neighborhood of τ of size proportional to the scale s converges to zero at a rate no faster than $s^{\alpha+1/2}$ as s tends to zero.

If the uniform Lipschitz regularity is positive, then, the previous two equations (after taking logs) imply that the amplitude of the wavelet transform modulus maxima should decrease when the scale decreases. In particular, an important theorem by Mallat and Hwang (1992) proves that there cannot be a singularity in a signal without a local maximum of its wavelet transform at the finer scale.

It should be clear from that discussion that the Holder regularity features do not hold in general for the discrete wavelet transform (Abry, 1994). This is due to the lack of translation invariance of the transform; in other words the transform of a shifted version of the signal does not corresponds to the shifted version of the transform of the original signal. As a consequence, the change points are usually located by mean of a continuous transform evaluated at discretised and dyadic scales, overcoming, in this way, also the lack of fast computational algorithm through the use of a modified version of the *algorithm a trous* (Abry, 1994).

The same sort of transform is used by Xu *et al.* (1994); in order to detect singularities and edges, they provide a spatially selective noise filtration technique which makes use of the direct spatial correlation of the nonorthogonal wavelet transform at several adjacent scales ($Corr_l(m,n) = \prod_{i=0}^{l-1} W(m+i,n)$ where $n = 1,\ldots N$, l is the number of scales involved in the direct multiplication, $m < M - l + 1$ and M is the total number of scales.). The use of a nonorthogonal transform is motivated by the fact that, it produces a transformed signal which is correlated across scales. A high correlation indicates the presence of a significant feature at the position that should be passed through the filter, whereas a moderate correlation leads to the cancelation.

In general, edges can be identified as part of the signal showing peaks across many scales. Then, the signal at small scales, where most of the noise is usually confined, is passed if it is around an identified edge, otherwise it is eliminated. That technique is very convenient for the optimal trade-off between noise reduction and edges preservation. Now let us attempt to relate the concept of Lipschitz regularity to the number of vanishing moments that characterize the Daubechies wavelet we shall employ in the subsequent analysis.

The Daubechies'compactly supported wavelets constitute a family of mother wavelets with varying smoothness, ψ_r, where r is the number of vanishing moments; $\psi_r \in \mathcal{C}^{\mu r}(R)$ with $\mu \approx 0.2$ and $\mathcal{C}^{\alpha}(R)$ is the space of α times continuous differentiable functions on R.

The Daubechies wavelets have no explicit expression except for the Haar case, which corresponds to a DB(1)[10]. However the squared modulus of the transfer function of the filter coefficients h_k is explicit and quite simple. The transfer function has a compact support and is r continuously differentiable.

The key to achieve that, while maintaining the orthogonality condition $\mathcal{H}(\omega) + \mathcal{H}(\omega + \pi) = 2$, is given by

$$H(\omega) = \left(\frac{1 + e^{i\omega}}{2}\right)^r Q(\omega) \tag{37}$$

where $Q(\omega)$ is such that $|Q(\omega)|^2 = \mathcal{Q}(\omega)$ and $\mathcal{Q}(\omega.)$ is a polynomial in $\sin^2(\omega/2)$. Implicit in that equation is the fact that a (unitary) scaling filter is said to be r-regular if its z-transform has r zeros at $z = e^{i\pi}$, namely if

$$H(z) = \left(\frac{1 + z^{-1}}{2}\right)^r Q(z) \tag{38}$$

where $H(z) = \sum_k h_k z^{-k}$ is the z-transform of the scaling coefficients h_k and $Q(z)$ has no poles or zeros at $z = e^{i\pi}$

If the filter is r-regular, then, its length is $K = 2r$ and $H(z)$ is a polynomial of degree $K - 1$; moreover, given that there are r zeroes at $z = -1$ the polynomial $Q(z)$ is of degree $K - 1 - r$.

The squared gain of the scaling filter is given by

$$\mathcal{H}(\omega) = 2\cos^{2r}(\omega/2) \sum_{k=0}^{r-1} \binom{r - 1 + k}{k} \sin^{2k}(\omega/2) \tag{39}$$

whereas the one for the wavelet is

[10] DB(K) corresponds to a filter with (K) vanishing moments and of length $2K$.

$$\mathcal{G}(\omega) = 2\sin^{2r}(\omega/2)\sum_{k=0}^{r-1}\binom{r-1+k}{k}\cos^{2k}(\omega/2) \tag{40}$$

The same squared gain function can generate several real-valued wavelet filters with the same number of coefficients when subject to spectral factorization, namely, when one takes the square root of $\mathcal{H}(\omega)$.

The squared gain associated with the transfer function of the scaling filter can be written as

$$\mathcal{H}(\omega) = |H(\omega)|^2 = \left|\frac{e^{i\omega}+1}{2}\right|^{2r}|Q(\omega)|^2 = \cos^{2r}(\omega/2)\mathcal{Q}(\omega) \tag{41}$$

$\mathcal{Q}(\omega)$ can be written as a polynomial in $\cos(\omega)$ (because $\{h_k\}$ is a real-valued sequence, and $\mathcal{H}(\omega)$ is a polynomial in $\cos(\omega/2)$. Given that $\cos(\omega) = 1 - 2\sin^2(\omega/2)$, it is possible to write $\mathcal{H}(\omega)$ as a polynomial in $\sin^2(\omega/2)$:

$$\mathcal{H}(\omega) = |\cos^{2r}(\omega/2)|\mathcal{P}(sin^2(\omega/2)) \tag{42}$$

By letting $y = \sin^2(\omega/2) = 1 - \cos^2(\omega/2)$ in $|\cos^2(\omega/2)|^r\mathcal{P}(\sin^2(\omega/2))$ we get $(1-y)^r\mathcal{P}(y)$ and the orthogonality condition $\mathcal{H}(\omega)+\mathcal{H}(\omega+\pi) = 2$ takes the following form[11]

$$(1-y)^r\mathcal{P}(y) + y^r\mathcal{P}(1-y) = 2 \tag{43}$$

which should hold for all $y \in [0,1]$. Daubechies proved that

$$\mathcal{P}(y) = 2[\mathcal{P}_r(y) + y^r\mathcal{R}(\frac{1}{2}-y)] \tag{44}$$

where

$$\mathcal{P}_r(y) \equiv \sum_{k=0}^{r-1}\binom{r-1-k}{k}y^k \tag{45}$$

and $\mathcal{R}(.)$ is an odd polynomial chosen so that $\mathcal{P}(y) \geq 0$ for $y \in [0,1]$. By choosing $\mathcal{R}(.) = 0$, she derives the class of filters having the squared gain functions given in (50) and (51). Derived under certain regularity conditions, once applied to data sequences, the Daubechies filters produce coefficients

[11] $$\cos^{2r}(\omega/2)\mathcal{P}(\sin^2(\omega/2)) + \sin^{2r}(\omega/2)\mathcal{P}(\cos^2(\omega/2)) = 2$$

that can be easily interpreted in terms of changes in adjacent weighted averages. They can be interpreted in terms of a cascade of filters made up by $K/2 = r$ difference (yielding the overall differencing operation) along with a lowpass filter (yielding the weighted average). Two different criteria imposed led to the *extremal phase* filters, conventionally known as $DB(r)$, and to the *Least Asymmetric* filters, known as $LA(.)$ or *Symmlets*.

3 Wavelet Denoising

The sparsity property, also called *coherence* of the wavelet transform, which is its ability to encode a signal in few big coefficients, is the basis for noise reduction techniques by thresholding the wavelet coefficients . These techniques outperform classical linear filtering.

Only if the signal and the noise components reside in separate frequency bands can such linear filtering techniques succeed in separating them perfectly. If the preponderant of the signal relies in the low frequency range and if the noise extends uniformly over the range or if it is biased towards the high frequency then the relevant filtering techniques are liable to be described as smoothing. These procedures often blur edges in signals and images.

The distinction between *smoothing* and *denoising* has been pointed out by Barclay *et al.* (1997) and becomes more apparent in a transform domain (i.e, Fourier or Wavelet). A white noise, not only has an *incoherent* structure, *i.e.* there is no way to compactly represent it, but it carries no useful information. It manifests itself in small amplitude coefficients which are widely dispersed across the frequency domain; the signal, instead, is well localised to parts of the frequency range.

Donoho and Johnstone (1994, 1995) have developed a simple way to suppress the noisy part of a signal, X, and to recover f in the context of a non-parametric regression

$$X(t) = f(t) + \sigma e(t) \tag{46}$$

where, in the simplest case, $e(t)$ is a Gaussian white noise with mean zero and variance, σ^2, equal to one. The method does involve *shrinkage* in the wavelet domain –an overall reduction in the size of the wavelet coefficients which will reduce coefficients of negligible value to zero– and is quite effective when applied to functions with a sparse wavelet representation. In such cases, given that the noise affects all the coefficients regardless of scale, shrinking them towards zero has the effect of suppressing the noise while preserving the essential features of the signal.

The denoising scheme consists of three steps. In the first step, a discrete wavelet transform is applied to the observed data sequence to produce

$$\eta_{j,k} = d_{j,k} + \epsilon_{j,k} \tag{47}$$

where $j = 1, \ldots, J_0$ and $t = 0, \ldots, N_j - 1$. Rather than a full transform, a partial at level J_0 is usually recommended. The detail coefficients are the superposition of the transform coefficients of $f(t)$ and of $\epsilon(t)$. Moreover the transform of the Gaussian white noise is also Gaussian white noise.

In the second step, the coefficients above a *primary resolution* are subjected to thesholding rules, which may reduce their values or set them to zero if they are below a certain level (see Johnstone and Silverman, 1997). Below the primary resolution level no thresholding is performed because it is presumed the signal is must predominate over the noise in the region of low frequencies.

There are two main rules of thresholding In the case *soft* thresholding, the coefficients below the threshold are regarded as noise and are set to zero; and the ones above are shrunk towards zero according to the following rule:

$$\eta(d_{j,k}) = \begin{cases} 0, & \text{if } |d_{j,k}| \leq \lambda \\ sign(d_{j,k})(|d_{j,k}| - \lambda) & \text{if } |d_{j,k}| > \lambda \end{cases} \tag{48}$$

In the case of *hard* thresholding, the coefficients are not shrunk. Instead they are set them equal to zero only if their absolute values are below the threshold:

$$\eta(d_{j,k}) = \begin{cases} 0, & \text{if } |d_{j,k}| \leq \lambda \\ d_{j,k} & \text{if } |d_{j,k}| > \lambda \end{cases} \tag{49}$$

The rule that is mostly used is the soft one. Even though the hard rule is better able to preserve the peaks, it also produces greater spurious oscillations close discontinuities. For this reason it is not especially favoured However, the problem can be mitigated through the use of a time-invariant transform

In the last step, the shrunken coefficients are transformed back in the signal domain, generally time, to give a sparser representation of the signal. However, we need to clarified how to choose the threshold. There are different ways corresponding to different trade-offs between optimal denoising and oversmoothing of the signal details.

Bruce and Gao (1996) and Marron *et al.* (1998) showed that the application of a particular threshold λ according to the *hard* rule results in larger variance in the function to be estimated. On the other hand, the same

threshold value with *soft* thresholding shifts the estimated coefficients by an amount of λ even when the empirical detail coefficients stands way above the noise level. This creates avoidable biases when the true coefficients are large.

Another issue is the instability of the hard threshold that is attributable to its discontinuity. It is sensitive to small changes in the data. As a remedy, Bruce and Gao (1997) proposed the *firm threshold* rule, known as a *keep* or *shrink* or *kill* rule. Firm threshold makes a transition from the soft to the hard threshold; that is to say, coefficients below the threshold level are nullified. Those at an intermediate level are reduced to an extent that diminishes as the values increases. When their value reach an upper bound they are preserved. However, the firm threshold is seldom applied because, by requiring two threshold values, it makes the estimation procedure more computationally expensive.

All of the above procedures entail thresholds and discontinuities. Attempts to remove such discontinuities have resulted in various formulations which have been described as *garroting* procedures (Gao, 1998).

The *universal* or *fixed-form* threshold, $\lambda = \hat{\sigma}\sqrt{2 * \log(n)}$, (Donoho and Johnstone, 1994), where $\hat{\sigma}$ is an estimate of the noise standard deviation, is a conservative choice. The motive for this, as explained by Donoho and Johnstone, is the fact that, for a standard normal sequence, $\{z_i; i = 1, \ldots, n\}$, the probability of exceeding that threshold tends to zero as n increases, i.e.

$$P\Big\{ \max |z_i| > \sqrt{2 * \log(n)} \Big\} \to 0 \quad \text{as } n \to \infty \tag{50}$$

whether or not they are independent. The expected number of $|z_i|$ exceeding $\sqrt{2 * \log(n)}$ tends to zero, and so is the probability as $n \to \infty$. Equivalently, the maximum of n i.i.d. standard Gaussian variables is smaller than $\sqrt{2 \log(n)}$ with probability equal to one as n increases.

Donoho and Johnstone (1994) proved the asymptotical optimality of the universal threshold, in *Mean Square Error* (MSE) sense, for mimicking the MSE of an ideal estimator designed via an *oracle* that provides crucial information regarding the functional form of the signal that is not conveyed by the data.

The estimate $\hat{\sigma}$ of the noise standard deviation can be obtained by applying a median absolute deviation (MAD) estimator to the $N/2$ wavelet coefficients at the first level of decomposition, incorporating a scale factor equal to 0.6745 [12]. It can be presumed that these high frequency coefficients

[12]The scale factor 0.6745 is the upper quartile of the standard normal distribution; given that the DWT gives rise to a sparse representation of a signal, only upper few $\eta_{j,k}$ will contain signal as well noise.

belong almost entirely to the noise component and contain virtually nothing of the signal .

The Median Absolute Standard Deviation (MAD) estimate of the standard deviation is a robust estimator against large deviations and reflects the noise variance rather then the signal variance.

The universal threshold is asymptotically optimal for many classes of functions and needs no *a priori* assumption concerning the class that f may belong to. For that reason is said to produce an asymptotically adaptive estimator. However, the use of the universal method comes at the price of a higher threshold level compared to other methods. As a consequence, it gives rise to a smoother representation of the signal which can be too far from the original. Better performance is obtained with smaller thresholds[13].

To address this problem, Donoho and Johnstone (1995) proposed that the threshold should depend upon the level of the wavelet decomposition. They have adopted the so-called *SURE* (Stein's Unbiased Risk Estimator) estimator due to Stein (1981), which leads to an unbiased estimate for the mean squared error of soft thresholding for each possible threshold choice λ. The level dependent thresholds are derived by considering the different resolution levels j as independent multivariate normal estimation problems and, then, by minimizing the estimate with respect to λ over the range $[0, \sigma\sqrt{(2\log n)}]$. The risk estimator is derived as follows. Suppose $Y \sim \mathcal{N}(d, \sigma^2)$ and let $\hat{d}$ be a k-dimensional vector of the form $\hat{d} = Y + g(Y)$. Stein showed that it is possible to estimate the loss $\|\hat{d} - d\|^2$ unbiasedly also in the case of nonlinear biased estimators. The risk is defined as:

$$\begin{aligned} \mathcal{R}(\hat{d}, d) = & E(\hat{d} - d)^2 = E(Y + g(Y) - d)^2 \\ & = E(Y - d)^2 + 2E(g(Y))(Y - d) + E(g(Y))^2 \end{aligned} \tag{51}$$

If $g(Y)$ is differentiable, $E[g^{'}(Y)] < \infty$, then

$$Eg(Y)(Y - d) = \sigma^2 Eg^{'}(Y) \tag{52}$$

and the equation above become

$$\begin{aligned} E(\hat{d} - d)^2 = & E(Y - d)^2 + E[g(Y)^2 + 2\sigma^2 g^{'}(Y)] \\ & = \sigma^2 + E[g(Y)^2 + 2\sigma^2 g^{'}(Y)] \end{aligned} \tag{53}$$

[13]However, when used with translation invariant transforms, the lower threshold produces a very large number of noise spikes, apparently much larger than in the non-invariant case.

If the soft-threshold estimator is used, we get

$$g(Y) = \begin{cases} \lambda, & \text{if } Y < -\lambda \\ -Y & \text{if } |Y| \leq \lambda \\ -\lambda & \text{if } Y > \lambda \end{cases} \tag{54}$$

Then

$$g^{'}(Y) = \begin{cases} 0 & \text{if } Y < \lambda \\ -1 & \text{if } |Y| \leq \lambda \\ 0 & \text{if } Y > \lambda \end{cases} \tag{55}$$

$$g(Y)^2 = \min(Y^2, \lambda^2) \tag{56}$$

As a consequence, an unbiased estimate of the risk $E\|\hat{d} - d\|^2$ can be written as

$$SURE(\lambda, Y) = k\sigma^2 - 2\sigma^2 \sum_{i=1}^{k} \{|Y| \leq \lambda\} + \sum_{i=1}^{k} \min(Y^2, \lambda^2) \tag{57}$$

from which the threshold is derived according to

$$\lambda = argmin_{0 \leq \lambda \leq \sigma\sqrt{2logk}} SURE(\lambda, Y) \tag{58}$$

As outlined by Donoho and Johnstone (1995) that minimization can be accomplished in $O(klogk)$ time. However, if most of the wavelet coefficients are zeros, then, the universal threshold is to be preferred. In other words, if the signal-to noise ratio is very small the SURE estimate is very noisy. So if such a situation is detected, the fixed form threshold is used. In order to detect that situation, Donoho and Johnstone proposed a measure of the sparsity of the wavelet representation and an hybrid method called *HeurSure* which, based on that measure, determines at each resolution level a threshold which is either the universal or the SURE one.

Another method proposed by Donoho and Johnstone (1998) is the *minimax thresholding*. The minimax principle is widely used in statistics to design estimators. Some prior information is often available concerning the vector of parameters, θ, and used to design some estimators the performance of which is then compared. This principle is also known as the *worst case* analysis: given some error measures, i.e. the mean squared error, compute the maximum expected error over the restricted parameter space and then determine

the estimator that minimizes the maximum risk[14].The *minimax risk* serves as a benchmark or ideal risk against which to measure the performance of alternative possible estimators. In our nonparametric framework, we are endowed with n noisy samples of a function $f(t_i); i = 1, \ldots, n$ which has to be estimated on the basis of n available observations on $x(t_i)$. We want the estimate $\hat{f}$ to have a small mean square error or *Risk* $\mathcal{R}_n(\hat{f}, f) = E\|\hat{f} - f\|^2$. As the universal threshold, the minimax threshold is of a fixed form and is chosen so as to yield minimax performance for the mean square error against an ideal procedure. The performances of the shrinkage estimators are evaluated in terms of a benchmark that can be thought as obtained through an *oracle* that tells which of the details, $d_{j,k}$, is close to zero. Two oracles have been used by Donoho and Johnstone (1994): the *Diagonal Linear Projection* which tells when to *keep or kill* the coefficient and the *Diagonal Linear Shrinker* which tells how much to shrink each empirical wavelet coefficient. In the diagonal projection case, the oracle sets $\hat{\theta}_i = \delta_i d_i$ with $\delta_i = 1$ or $\delta_i = 0$, depending if it keep or kill the wavelet coefficient; then the risk for the ideal choice of the sequence $\{\delta_i\}$ is

$$\mathcal{R}(DLP, \theta) = \min_{\{\delta_i\}} \Big\{ \sum_i E(\delta_i d_i - \theta_i)^2 \Big\} = \sum_i \min\{\theta_i^2, E(d_i - \theta_i)^2\} = \sum_i (\theta_i^2 \wedge \sigma^2) \tag{59}$$

Ideally, the oracle must tell us which of the $\hat{\theta}_i$ exceeds the noise standard deviation σ; in fact, $\delta_i = I[\theta_i^2 \geq E(x_i - \theta_i)^2] = I[\|\theta_i\| \geq \sigma]$. The empirical wavelet coefficient is kept if its contribution to the energy of the function is

[14]The *minimax principle* can be seen as the results of adapting some *game theory* tools for use in statistics (Johnstone, 1999). There are two players facing a statistical decision problem. Player I, *the scientist*, faced with the problem of recovering θ_k (basis coefficients) from $y_k = \theta_k + \sigma_n z_k; k = 1, \ldots, n$, is allowed to choose any estimator $\hat{\theta}(y)$, linear, threshold or of more complicated type. Player II, *the nature*, may choose $\theta \in \mathcal{R}^n$ at random with a probability, π, associated, subject only to a sparsity constraint, which is usually measured in terms of l_p norms, $\|\theta_p\|_p = (\sum_1^n |\theta_k|^p)^{1/p}$ for $p < 2$. The pay-off is calculated as the expected mean squared error of $\hat{\theta}(Y)$ when θ is chosen according to π and Y satisfies $Y = \theta + \sigma_n z$ for $z \sim N_n(0, I)$. As a consequence, the pay-off averages over both θ and Y and is such that

$$r(\hat{\theta}, \pi) = E_\pi E_{Y|\theta} \|\hat{\theta}(Y) - \theta\|_2^2$$

Obviously, the scientist aims at minimizing the pay-off and nature tries to maximize it. Therefore, it is possible to apply the *minimax theorem* of von Newmann to get the *minimax risk*

$$R_n = \inf_\theta \sup_\pi r(\hat{\theta}, \pi) = \sup_\pi \inf_{\hat{\theta}} r(\hat{\theta}, \pi)$$

The first part of the equality is the *minimax estimator for player I*, while a prior distribution π attaining the right end supremum is an optimal strategy for player II and is called *the least favourable* (Johnstone, 1999).

bigger than the variance of the noise, otherwise is discarded. This ideal situation will never happen but the risk attained $\mathcal{R}_{ORACLE}(DLP, \theta) = \sum_i(\theta_i^2 \wedge \sigma^2)$ which is the MSE of $\hat{\theta}$, is used as a benchmark [15]. More precisely, the benchmark includes also an additional term, σ^2. Donoho and Johnstone (1994) showed that the performance of the universal threshold comes surprisingly close to the performance of the ideal estimator built via an oracle and derived an upper bound of the risk of the minimax estimator in terms of the unattainable ideal risk for the entire vector of wavelet coefficients; in the case of soft thresholding, it is

$$E\|\hat{\theta} - \theta\|^2 \leq (1 + 2\log n)\left\{\sigma^2 + \sum_i(\theta_i^2 \wedge \sigma^2)\right\} \quad \text{for all } \theta \in \mathcal{R}^n \tag{60}$$

In other words, for all possible θ, the estimator $\hat{\theta}$ can mimic, within a factor $(1 + \log n)$ the performance of the ideal risk plus the additional parameter, $\overline{\sigma}^2$; moreover, if the soft threshold is used, the estimate is with high probability at least as smooth as the original function.

In the case of colored noise, instead, the upper bound is

$$E\|\hat{\theta} - \theta\|^2 \leq (1 + 2\log n)\left\{\bar{\sigma}^2 + \sum_i(\theta_i^2 \wedge \sigma^2)\right\} \quad \text{for all } \theta \in \mathcal{R}^n \tag{61}$$

Other approaches are available for the choice of the threshold value. We give only a brief outline of some of them and provide the main references for them.

The *Cross-Validation* approach is a technique which has been widely used in statistics for the purpose of choosing a smoothing parameter in a nonparametric regression (Green and Silverman, 1994). For wavelet thresholding, it is applied in the *iid* Gaussian noise framework. The approach (Nason, 1994, 1995, 1996; Weyrich and Warhola (1995)) focus on minimizing the prediction error generated by two competing approximations of the unknown function. Two methods are available, the *Twofold Cross-Validation* and the *Leave-One-Out*. [16]

[15] In the diagonal linear shrinker case, instead, the risk is $\mathcal{R}_{ORACLE}(DLS, \theta) = \sum_i(d_i^2/(d_i^2 + 1))$

[16] The former consists of splitting the data series into odd and even indexed observations. Given a threshold, the odd-indexed sample is used to estimate a signal which is, then, interpolated and compared, using squared differences, with the even-indexed downsamples. The threshold which minimizes the sum of the squared differences is chosen. The latter, instead, removes from the sample, $X_0, X_1, \ldots, X_{N-1}$, one observation, let's say X_t, and

The Bayesian approach to wavelet shrinking and denoising is based on a *prior* imposed on the wavelet coefficients of the function that has to be recovered. That prior intends to capture the sparseness of the transform and assigns to each of the wavelet coefficients a scale mixture of two distributions, one corresponding to *negligible* coefficients and the other to *significant* coefficients. It is common practice to use a scale mixture of two normal distributions or a mixture of a normal and a point mass at zero[17]. For details see, amongst others, Abramovich *et al.* (1998), Chipman *et al.* (1997); Crouse *et al.* (1998).

Despite the near-optimality properties in comparison to other methods, classical wavelet denoising suffers from a serious drawback which sees it as responsible of creating artifacts in the recovered function (Coifman and Donoho, 1995). In particular, in the neighborhood of a discontinuity in the noisy sampled series, the denoised signal can exhibit pseudo-Gibbs phenomena. These arise as a consequence of using only a subset of coefficients in the reconstruction of the function; however, they behave much better than those in the Fourier domain, i.e. they are better localized and of smaller oscillations. The presence of these oscillations is related to the lack of shift-invariance of the discrete wavelet transform which does not allow alignment between the signal features and those of the wavelet basis chosen. Shift invariance can be achieved by means of overcomplete or redundant transforms according to different methods. In the attempt to realign signal and basis features, so as to improve the denoising performance, one can shift the signal by an optimal amount, then denoise it and unshift it. However, if a signal has more than one discontinuity, a shift that is optimal for one of the features might turn out to be not appropriate for the others. To overcome that, Coifman and Donoho (1995) proposed the idea of *Cycle Spinning* which consists in denoising all possible shifts of a signal and, then, averaging them (in sequence: circular shift, DWT, thresholding, IDWT, unshift, average). The algorithm has complexity $\mathcal{O}(n^2)$; however, Beylkin (1992) and Shensa *et al.* (1992) have proved that it can be implemented efficiently in terms of an

considers two subsamples, one made up by the observations before X_t and the other by the observations after X_t. For a given threshold, an estimate of the unknown function is produced and also an estimate of the removed value which is, then compared to the true value of the expelled point. This is repeated for each t and the threshold is chosen so as to minimize the sum of the squared differences between the estimated and the true removed point. That procedure is faced with the dimensionality problem. By removing one point, we end up with $2^M - 1$ observations and, unless one pads with zeros, a Discrete Wavelet Transform cannot be used. A MODWT can be applied instead.

[17] A point mass in statistics is a discontinuous segment in a probability distribution; in physics, is an idealization of a body whose dimensions are very small compared to the other distances that are relevant to that problem.

overcomplete transform such as the MODWT. Removing the downsampling step at each stage of the pyramidal algorithm is equivalent to compute only $n \log n$ different coefficients among those corresponding to all shifts of the input signal. That procedure is computationally heavier than the classical DWT - which is $\mathcal{O}(n)$ - but the complexity is still far lower than that of the averaging-shift method.

The MODWT gives rise to components that are not mutually orthogonal but, in terms of its wavelet coefficients, is still an energy preserving transform. As the DWT, it is well suited to analyze the scale dependence in an ANOVA framework. Moreover it is possible to build an unbiased estimator of the sample variance as the sum of the variances at each level of decomposition[18].

4 Empirical Application

As an illustration of the power of the shrinkage techniques, we consider the problem of identifying structural breaks in the process that has generated the UK real GDP from 1873 to 2001. These data have often been quoted in the debate concerning the appropriate means of reducing trended sequences to stationarity. The issue is whether this should be achieved by subtracting an interpolated trend from the data or by applying a difference operator to them (*i.e*, Duck, 1992; Mills, 1994; Mills and Crafts, 1996).

It has been claimed (by Rappoport and Reichlin, 1989, amongst others,) that the tendency to opt for the difference stationary model can be attributed to the difficulty in finding an adequate functional representation the trend.

In fact, the period in question, of slightly more than a century and a quarter, has witnessed many policy regime shifts, two world wars and two oil crises, as well as major legislative and technological changes. To accommodate such upheavals, one might be justified in using a broken or a segmented function. However, this is not the only available recourse; *i.e.*, given their ability to decompose functions into localised oscillations, wavelets are extremely good at capturing sudden changes in noisy sequences.

Disjunctions or singularities in the data are usually detected, as we have discussed in details, by identifying the values of the abscissae where the

[18]For a sample size N and a wavelet filter of length L, the wavelet variance $\sigma_x^2(\lambda_j)$ can be estimated by

$$\sigma_x^2(\lambda_j) = \frac{1}{\tilde{N}_j} \sum_{t=L_j-1}^{N-1} \tilde{w}_{j,t}^2$$

where $\tilde{w}_{j,t}$ are the jth level MODWT coefficients, $L_j = (2^j - 1)(L-1) + 1$ is the length of the scale j wavelet filter, $\tilde{N}_j = N - L_j + 1$ and $L_j = 2^{j-1}$.

modulus maxima of the continuous wavelet transform converge at fine scales (Mallat and Hwang, 1992; Wang, 1995;1999). In the presence of singularities, the asymptotic decay is much slower than the case of Holder-regular functions, as we have already described. However, the DWT cannot be characterized in terms of Holder regularity (Abry, 1994) due to to the lack of translation invariance, which is a crucial property for pattern recognition applications, such as the detection of structural changes.

A redundant transform, such as the MODWT, which gives rise to components that are correlated across the scales,can be used in detecting singularities.

The MODWT comprises an analysis transform that generates the wavelet coefficients and a synthesis transform that forms the components of the data decomposition. Since the analysis transform entails a backward looking causal filter, the coefficients of the MODWT embody a phase lag which throws them out of alignment with the events in the data to which they correspond (this difficulty also affects the DWT). In the synthesis stage the MODWT coefficients are processed by an anticausal filter which eliminates the phase lag. The effect is to produce a set of sequences at all levels that are precisely aligned with the events in the data. The avoidance of downsampling also enhances the manner in which the MODWT components reflects the events in the data. The phase lag of the analysis stage can be minimized by choosing wavelet filters that are nearly symmetric and by applying an appropriate phase adjustment. However, the choice of the wavelet filter is less critical in the case of the MODWT than in the case of the DWT.

There is also the question of what happens when the wavelet filter approaches the end of a finite data sequence. The difficulties increase with the length of the support of the filter.

A filter with a wider support is usually more closely confined to its nominal passband band within the frequency domain. Therefore, it tends to produce a smoother reconstruction by reducing the artefacts in a MRA that are due to the shape of the filter. However, a wide filter also produces a larger number of boundary coefficients, especially at higher scales. In a MODWT the support of the filter increases in the descent from high frequency to low frequency.

A shorter filter, on the other hand, corresponds to a less regular wavelet with a shorter support; and it is better suited to detecting sudden changes and jumps. However, the corresponding wavelet filters at different scales entail more leakage from frequencies outside the passband compared to longer filters. On the basis of these considerations, we have chosen to perform a

MODWT by using a *symmlet*(2)[19]

The *Least Asymmetric* family is characterised by approximate linear phase and near symmetry about the filter midpoint.

It is usual, in the case of MODWT, to handle the end effects by making a periodic extension of the data. Of course, this is equivalent to applying the wavelet filters to a data sequence that has been wrapped around the circle.A problem with the periodic extension can occur when there is a large discontinuity between the end of one replication of the sample end the beginning of the next. To reduce this problem the data should be suitably detrended. An alternative to periodic extension is an extension by symmetric reflection in which successive segments of the extension contain the data of their predecessors in reverse order.

A failure adequately to detrend the data may result in radical disjunction in passing from the end of one replication of the sample to the beginning of the next. This can cause severe border distortions, albeit that, in a wavelets analysis, they are less severe than they would be in a conventional Fourier analysis. For these reasons, the series that is subject of the analysis has been reduced to stationary by taking first differences. The good news, however, is that the boundary condition has little impact on detecting jumps and sharp cusps away from the boundary points which are at the beginning and the end of the sample (Percival and Walden, 2000)[20].

Departing from the middle of the sample in either direction there is a decreasing variability in the annual UK GDP growth (Figure 2) that is even more manifest from its multiresolution decomposition performed up till level $j = 5$ (Figure 3).

The components at each scale of resolution are synthesized from the wavelet and scaling coefficients. In the attempt to reveal the true structure of the data, i.e. to locate the dates of sudden changes, jumps and disjunctions in general, these wavelet coefficients have been subject to soft shrinkage once the threshold parameter, λ, has been determined.

We have chosen the universal threshold method to derive the threshold $\lambda = \hat{\sigma}\sqrt{2\log n}$. Here $\hat{\sigma} = \sigma_{MAD}$ is the MODWT-based MAD (*Median Absolute Deviation*) estimator of the standard deviation and is calculated from the elements of the smallest scale vector of wavelet coefficients. These coefficients are noise dominated with the exception of the largest values and are

[19] $N = 2$ denotes the number of vanishing moments. accordingly the length of the filter is $2N = 4$

[20] The MODWT that are affected by circularity are $\tilde{W}_{j,t}$ and $\tilde{V}_{j,t}$ for $t = 0, \ldots, \min\{L_j - 2, N - 1\}$, where $L_j \equiv (2^j - 1)(L - 1) + 1$. Since the size of these vectors of coefficients is always N and since L_j increases with N, the proportion of boundary coefficients increases with J and reaches its maximum, unity, when $L_j - 1 \geq N$.

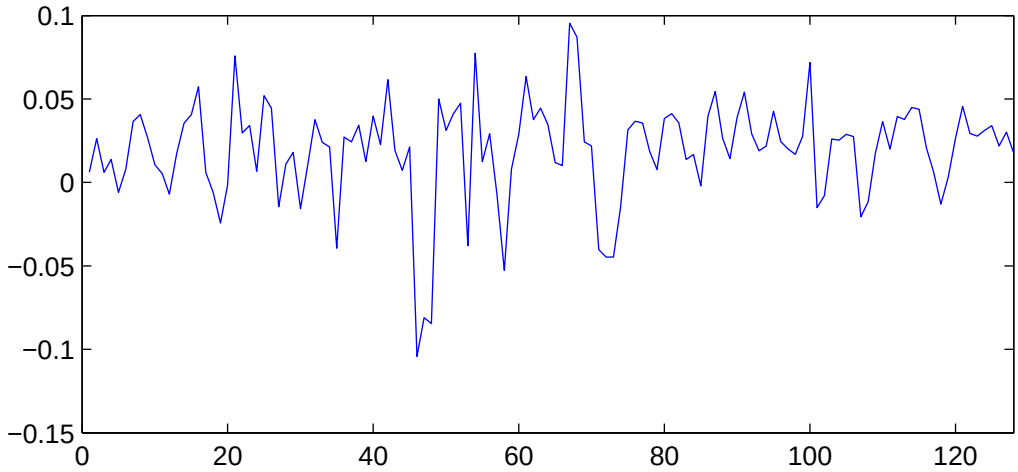

Figure 2: Logarithms of UK GDP annual growth.

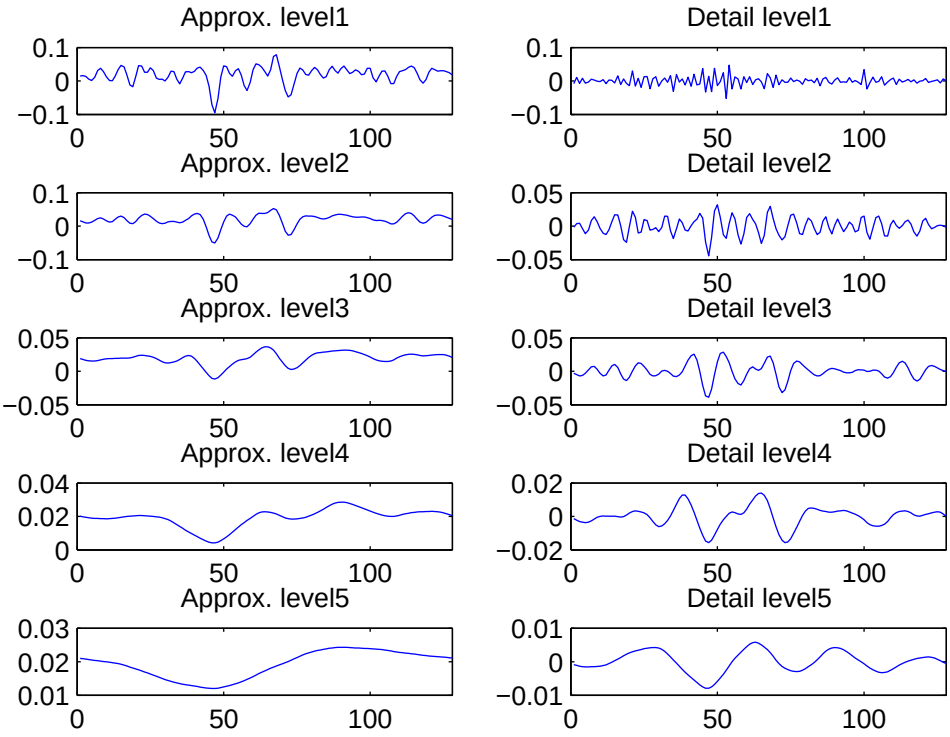

Figure 3: MODWT multiresolution of logarithms of the UK GDP annual growth using the LA(2) wavelet.

used to derive:

$$\hat{\sigma}_{MAD} \equiv \frac{2^{1/2} median\{|\tilde{W}_{1,0}|, |\tilde{W}_{1,1}|, \ldots, |\tilde{W}_{1,N-1}|\}}{0.6745} \tag{62}$$

The resulting threshold is equal to 0.051. All the coefficients below it are set to zero. The others have been shrunk according to the formula (49).

Figure 4 shows the denoised version of UK GDP. Whereas Figure 2 gives an obscure picture of the events occurring in the sample, the sharp decline in the growth series, revealed by Figure 4, is a clear indication of a structural break occurring in 1918. A major shock in the economy commonly results in a dramatic decline in the levels of output, followed by a recovery. It is easy to recognise that the post-break growth rate is higher than the pre-break one. To a lesser extent, this applies also to the second break corresponding to the second world war. The effects of that war, however, are not easily discernible in Figure 2.

Manifestly the period that is bounded by the 2 world wars was much more volatile than the preceding and succeeding periods. It is also surprising to note that in terms of the growth rate the effects of the crash of 1929 are smaller than the effects of the cessation of hostilities after both World Wars. The break is an event that has a structure that extends across the scales of resolution. It is well located temporally and stands out across the frequency range. Cycles of length between 4 and 8 years are detectable from the MRA, corresponding to a substantial concentration of details energy at level 3. Cycles shorter than 16 years are also important (level $j = 4$) but to a lesser extent.

To conclude the analysis, we have used the denoised annual growth to recover the trend that has been, then, plotted with the UK GDP in levels.

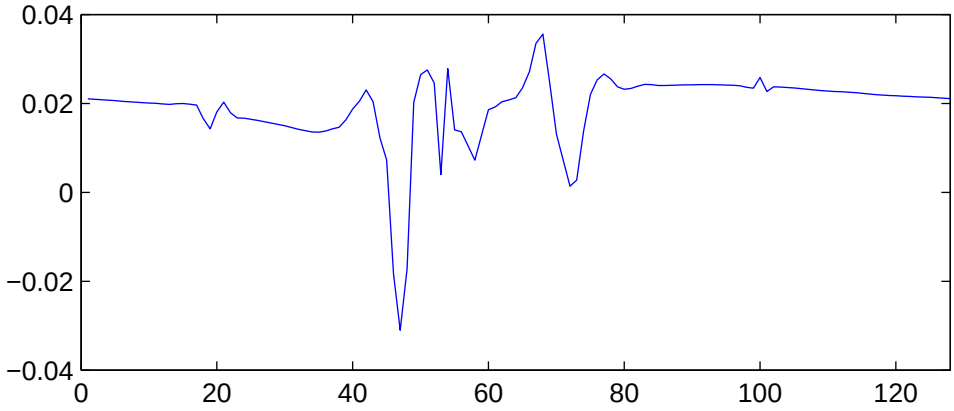

Figure 4: Denoised version of the logs of the UK GDP annual growth.

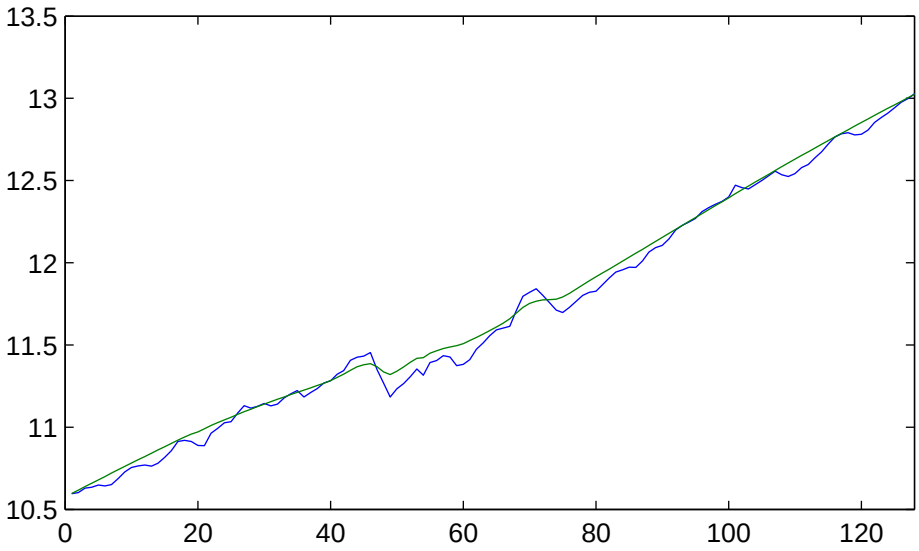

Figure 5: UK GDP and its trend.

References

Arfken, G., (1985), *Mathematical Methods for Physicists,* Third Edition, Academic Press, San Diego, California.

Abramovich F., Sapatinas T. and B.W. Silverman, (1998), *Wavelet thresholding via a Bayesian Approach,* Journal of the Royal Statistic Society, Series B, 60, pp.725-49.

Abry P., (1994), *Transformes en Ondelettes - Analyses Multirsolutions et Signaux de Pression en Turbulence,* These de Doctorat. Universit Claude Bernard, Lyon, France.

Antoniadis, A., (1996), *Smoothing noisy data with tapered coiflets series,* Scandinavian Journal of Statistics, vol. 23, pp. 313-330.

Barclay V.J., Bonner R.F. and I.P. Hamilton, (1997) *Application of Wavelet Transform to Experimental Spectra: Smoothing, Denoising, and Data Set Compression,* Analytical Chemistry, 69, pp.78-90.

Beylkin G., (1992), *On the Representation of Operators in Bases of Compactly Supported Wavelets,* SIAM Journal of Numerical Analysis, vol.29, n.6, pp.1716-1740.

Bruce, A. G. and H. Y.Gao, (1996), *Understanding WaveShrink: variance and bias estimation,* Biometrika, 83, pp. 727-745.

Burrus C.S., Gopinath R. A. and H. Guo, (1996), *Introduction to Wavelets and Wavelet Transform: a Primer,* Prentice Hall.

Chipman, H.A., Kolaczyk, E.D and R.E. McCulloch, (1997), *Adaptive Bayesian Wavelet Shrinkage,* Journal of the American Statistical Association, vol.92, pp.1413-21.

Chui, C.K., (1997), *Wavelets: A Mathematical Tool for Signal Analysis,* SIAM Monographs on Mathematical Modeling and Computation.

Coifman, R.R. and D.L Donoho, (1995), *Translation-Invariant Denoising,* in Wavelets and Statistics (Lecture Notes in Statistics, Vol.103), edited by Antoniadis A. and G. Oppenheim. New York: Springer-Verlag, 125-50.

Croisiere A., Esteban D. and C. Galand, (1976), *Perfect Channel Splitting by Use of Interpolation/Decimation/Tree Decomposition Techniques,* in International conference on Information Sciences and systems, pp.443-446, Patras, Greece, August 1976.

Crouse, M.S., Novak, R.D. and R.G. Baraniuk, (1998), *Wavelet-Based Statistical Signal Processing Using Hidden Markov Models,* IEEE Transactions on Signal Processing, vol. 46, pp. 886-902.

Daubechies, I., (1988), *Orthonormal Bases of Compactly Supported Wavelets,* Communications in Pure and Applied Mathematics, vol. 41, pp.909-996.

Daubechies, I., (1992), *Ten Lectures on Wavelets,* Society for Industrial and Applied Mathematics, Philadelphia.

Donoho D.L. and I.M. Johnstone, (1994), *Ideal Spatiation Adaptation Via Wavelet Shrinkage,* Biometrika, 81, pp.425-455.

Donoho D.L. and I.M Johnstone, (1995), *Adapting to Unknown Smoothness via Wavelet Shrinkage,* Journal of the American Statistical Association, Vol 90, n.432, pp.1200-1224.

Donoho D.L. and I.M Johnstone ,(1998), *Minimax Estimation Via Wavelet Shrinkage,* The Annals of Statistics, 1998, vol.26 n.3, pp.879-921.

Duck N.W.,(1992), *UK Evidence on Breaking Trend Functions,* Oxford Economic Papers, 44, pp.426-439.

Gabor, D., (1946), *Theory of Communication,* Journal of the Institute for Electrical Engineers, 93, pp.429-457.

Gao, H.-Y. and A.G. Bruce, (1997), *WaveShrink with firm shrinkage,* Statistica Sinica, vol. 7, pp.855-874.

Gao, H.Y., (1998), *Wavelet shrinkage denoising using the non-negative garrote,* Journal of Computational and Graphical Statistics, vol. 7, pp. 469-488.

Green, P.J. and B.W. Silverman, (1994), *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach,* London: Chapman Hall.

Grossmann, A and J. Morlet, (1984), *Decomposition of Hardy functions into Square Integrable Wavelets of Constant Shape,* SIAM Journal on Mathematical Analysis, 15, 723-36.

Holschneider, M, Kronland-Martinet, R., Morlet J. and P. Tchamitchian, (1989), *A Real-Time algorithm fo Signal Analysis with the Help of the wavelet Transform,* in Wavelets, Time-Frequency Methods and Phase Space, pp. 289-297, Sprinter-Verlag, Berlin .

Holschneider, M and P. Tchamitchian, (1990), *Regularit locale de la function 'nondifferentiable' de Riemann,* pp.102-124 in Lemari P.G. (1990) Les ondelettes en 1989 , Lecture Notes in Mathematics no. 1438, Springer-Verlag, Berlin.

Hubbard, B., (1998), *The World According to Wavelets. The Story of a Mathematical Technique in the Making,* A.K. Peters. Natick, Massachusetts.

Jaffard, S., (1989), *Exposants de Holder en des points donns et coefficients d'ondelettes,* C.R. Acad. Sci. Paris, 308, Serie I, pp.79-81.

Jaffard S. and Y. Meyer, (1996), *Wavelet Methods for Pointwise Regularity and Local Oscillations of Functions,* Memoirs of AMS, 123 (587).

Johnstone, I.M, (1999), *Wavelet and Theory of Non-Parametric Function Estimation,* Philosophical Transactions of the Royal Society of London, SERIES A, 357, pp. 2475-2493.

Johnstone I.M and B.W. Silverman, (1997), *Wavelet Threshold Estimators for Data with Correlated Noise,* Journal of the Royal Statistical Society. Series B (Methodological), vol. 59, n.2, pp.319-351.

Mallat, S.G., (1987), *A Compact Multiresolution Representation: The Wavelet Model,* in Proceedings of the IEEE Computer Society Workshop on Computer Vision, pp. 2-7, IEEE Computer Society Press, Washington, D.C., 1987.

Mallat S.G., (1989),*A Theory for Multiresolution signal Decomposition: the Wavelet Representation,* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.11, pp.647-693.

Mallat S.G. and W.L.Hwang, (1992), *Singularity Detection and Processing with Wavelets,* IEEE Transactions on Information Theory, vol.38,n.2.

Mallat S.G., (1999), *A Wavelet Tour of Signal Processing,* Second Edition, Academic Press.

Marron, J.S., Adak, S., Johnstone, I.M., Neumann, M.H. and P. Patil, (1998), *Exact Risk Analysis of Wavelet Regression,* Journal of Computational and Graphical Statistics, Vol.7, n.3, pp.278-309.

Meyer, Y., (1992), *Wavelets and Operators,* Cambridge University, Cambridge, 1992.

Mills, T. C.,(1994), *Segmented Trends and the stochastic Properties of UK Output,* Applied Economic Letters, 1, pp.132-133.

Mills, T. C.and N.F.R. Crafts,(1996), *Modelling Trends in Economic History,* The Statistician, 45, pp.153-159.

Nason, G.P., (1994), *Wavelet Regression by Cross-Validation,* Technical report 447, Department of Statistics, Stanford University, USA.

Nason, G.P., (1995), *Wavelet function Estimation using Cross-Validation,* In Wavelets and Statistics, Antoniadis, A. and G. Oppenheim (Eds), Lecture Notesin statistics, 103, pp.261-280, New York:Springer-Verlag.

Nason, G.P., (1996), *Wavelet Shrinkage Using Cross-Validation,* Journal of the Royal statistic Society, Series B, 58, 463-479.

Nason, G.P. and B. Silverman, (1995), *The Stationary Wavelet Transform and Some Statistical Applications,* In Wavelets and Statistics, vol. 103 edited by A.Antoniadis and G. Oppenheim. New York: Springer-Verlag, 281-99.

Nelson, C.R. and C.I. Plosser, (1982), *Trends and Random Walks in Macroeconomic Time Series,* Journal of Monetary Economics, 10,pp.139-162.

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

Perron, P., (1989),*The Great Crash, the Oil Price Shock, and the Unit Root Hypothesys,* Econometrica, 57, pp.1361-1401.

Perron, P., (1990), *Evidence on Breaking Trends and Non-stationary Time Series*Princeton University, Mimeo.

Pesquet, J.-C., Krim, H. and Carfantan H., (1996), *Time-Invariant Orthonormal Wavelet Representations,* IEEE Transactions on signal Processing, 44, pp. 1964-70.

Pollock D.S.G. and I. Lo Cascio, (2006), *Non-Dyadic Wavelet Analysis,* in Optimisation, Econometric and Financial Analysis, ed. Kontoghiorghes E.J. and C. Gatu.

Priestley M.B., (1995), *Spectral Analysis and Time Series,* Academic Press.

Rappoport P. and L. Reichlin, (1989), *Segmented Trends and Non-Stationary Time Series,* The Economic Journal, 99 (Conference 1989) pp. 168-177.

Shannon, C.E., and W. Weaver, (1964), *The Mathematical theory of Communications,* The University of Illinois Press, Urbana, Illinois.

Shensa, M.J., (1992), *The Discrete Wavelet Transform: Wedding the Trous and Mallat Algorithms,* IEEE Transactions on Signal Processing, vol. 41, pp.3584-3590.

Smith, M.J. and T. P. Barnwell, (1986), *Exact Reconstruction Techniques for tree-Structured Subband Coders,* IEEE Transaction on Acoustic Speach and Signal Processing, vol. 34, n.3, pp.434-441.

Stein, C., (1981), *Estimation of the Mean of a Multivariate Normal Distribution,* Annals of Statistics, 9 (6), pp.1135-1151.

Vaidyanathan, P.P., (1990), *Multirate Digital Filters, Filter Banks, Polyphase Networks, and Applications: A Tutorial,* Proceedings of the IEEE, vol. 78, n.1, pp. 56-93.

Vaidyanathan (1993),

Vetterli, M. and J. Kovacevic, (1995), *Wavelet and Subband Coding,* Prentice Hall.

Wang, Y., (1995), *Jump and Sharp Cusp Detection by Wavelets,* Biometrika, 82, pp.385-397.

Wang, Y., (1999), *Change-Points Via Wavelets for Indirect Data,* Statistica Sinica, 9, pp.103-117.

Weyrich, N. and G.T. Warhola, (1995), *Wavelet shrinkage and generalized cross-validation for de-noising with applications to speech,* In Wavelets and Multilevel Approximation, Chui, C.K. Schumaker, L.L. (Eds.), Approximation Theory VIII 2, pp. 407-414, Singapore: World Scientific

Wickerhauser, V.M., (1994), *Adapted Wavelet Analysis from Theory to Software,* A.K. Peters, Natick, Massachussets.

Woodward, P.M., (1953), *Probability and Information Theory with Applications to Radar,* McGraw-Hill Book Company, New York.

Xu Y., Weaver J.B., Healy, Jr., and Jian Lu, (1994), *Wavelet Transform Domain Filters: A Spatially Selective Noise Filtration Technique,* IEEE Transactions on Image Processing, vol.3, n.6.

This working paper has been produced by
the Department of Economics at
Queen Mary, University of London

Department of Economics
Queen Mary, University of London
Mile End Road
London E1 4NS
Tel: +44 (0)20 7882 5096
Fax: +44 (0)20 8983 3580
Web: www.econ.qmul.ac.uk/papers/wp.htm