An intuitive guide to wavelets for economists

Patrick M. Crowley*†

February 2005

Abstract

Wavelet analysis, although used extensively in disciplines such as signal processing, engineering, medical sciences, physics and astronomy, has not fully entered the economics literature. In this discussion paper, wavelet analysis is introduced in an intuitive manner, and the existing economics and finance literature that utilises wavelets is explored and explained. Examples of exploratory wavelet analysis using industrial production data are given. Finally, potential and possible future applications for wavelet analysis in economics are discussed and explored.

Keywords: Business cycles, economic growth, statistical methodology, multiresolution analysis, wavelets.

JEL Classification: C19, C65, C87, E32

*Visiting Research Scholar, Bank of Finland, Helsinki, Finland and College of Business, Texas A&M University, Corpus Christi, TX, USA.

†Contact details: Research department, Bank of Finland, P.O. Box 101, 00101 Helsinki, Finland. Email: patrick.crowley@bof.fi

1 Introduction

The wavelet literature has rapidly expanded over the past 15 years, with over 1600 articles and papers now published using this methodology in a wide variety of disciplines. Applications using wavelets in disciplines other than economics are extensive, with many papers published in areas such as acoustics, astronomy, engineering, medicine, forensics, and physics. Economics (and to a lesser degree, finance) is conspicuous in its absence from this list, largely because for some reason the potential for using wavelets in economic applications has been overlooked. Although some enterprising economists have attempted to use wavelet analysis, given the discipline's fixation on traditional time-series methods, these papers have not been widely cited and have in fact been largely ignored. The main aim of this discussion paper is to shed new light on wavelet analysis, to highlight the work that has already been done using this approach, and to suggest future areas where wavelet analysis might be able to make a contribution to our discipline. To maximise accessibility to this material, the discussion paper is pitched at a less technical level than most other introductions to wavelets, although a fairly complete list of references is provided for those who might wish to obtain a more technical introduction.

So, what are wavelets? Wavelets are, by definition, small waves. That is, they begin at a finite point in time and die out at a later finite point in time. As such they must, whatever their shape, have a defined number of oscillations and last through a certain period of time or space. Clearly these small wavelike functions are ideally suited to locally approximating variables in time or space as they have the ability to be manipulated by being either "stretched" or "squeezed" so as to mimic the series under investigation.

Wavelets possess many desirable properties, some of which are useful in economics and finance, but many of which are not. In this paper the focus is placed on the ability of wavelet analysis to deal with both stationary and non-stationary data, their localization in time, and on their ability to decompose and analyse the fluctuation in a variable (or in signal processing, what is called a "signal").

In economics, while providing a review of the possible future contributions of wavelets to the economics discipline, Ramsey (2000) explored four ways in which wavelets might be used to enhance the empirical toolkit of our profession. These are as follows:

- exploratory analysis - time scale versus frequency: in economics and finance an examination of data to assess the presence and ebb and flow of frequency components is potentially valuable.

- density estimation and local inhomogeneity: wavelets estimators are superior to kernel estimators whenever local inhomogeneity is present (e.g. modelling impact of minimum wage legislation, tax legistlation, rigidities, innovation)

- time scale decomposition: recognition that relationships between economic variables can possibly be found at the disaggregate (scale) level rather than at an aggregate level.

- forecasting: disaggregate forecasting, establishing global versus local aspects of series, therefore whether forecasting is really possible.

As shall soon be apparent, in this paper the focus is on the first and third areas of empirical analysis. The second and fourth areas have certainly gained popularity and are still largely unexplored. This paper constitutes an extension to Schleicher (2002) and Ramsey (2002), given that many of the advances in wavelet analysis have only recently been made widely available on commercially available statistical software packages. For those more mathematically inclined, they should refer to the mathematics literature where Debauchies (1992) and Walnut (2002) are good starting points. Other perhaps less intuitive but nonetheless technically sound introductions to wavelets can be found in Walker (1999), Percival and Walden (2000) and Addison (2002). There are three other entry points to this literature that are specifically aimed at economists - the excellent book by Gençay, Selçuk, and Whicher (2001), an article by Ramsey (2002) which provides a nice rationale for wavelets in economics, and the discussion paper by Schleicher (2002).

2 What are wavelets?

Wavelet analysis has various points of similarity and contrast with Fourier analysis. The Fourier transform is based on the usage of the sum of sine and cosine functions at various wavelengths to represent a given function. One of the drawbacks of Fourier transforms though is that they assume that the frequency content of the function is stationary along the time axis. Imagine a minimalist symphony (say John Adams or Steve Reich) - the analogue here would be each instrument playing a note, with a specific loudness: $a_0 + a_1 \cos t + a_2 \cos 2t + ...$ - to represent this signal one would only need the list of amplitudes $(a_0, a_1, a_2, ...)$ and the corresponding frequencies $(0, 1, 2, ...)$. In this sense Fourier

analysis involves the projection of a signal onto an orthonormal[1] set of components which are trigonometric in the case of the Fourier approach. The Fourier transform makes particular sense when projecting over the range $(0, 2\pi)$, as Fourier series have infinite energy (they do not die out) and finite power (they cannot change over time).

Windowed Fourier analysis extends basic Fourier analysis by transforming short segments of the signal (or "symphony") separately. In other words there are breaks where we just repeat the exercise above. In a visual world we see an edge - in an economics world we see regime shifts. Once again these are just separate sets of orthonormal components - one set for each window.

Wavelets are localized in both time and scale. They thus provide a convenient and efficient way of representing complex variables or signals, as wavelets can cut data up into different frequency components. They are especially useful where a variable or signal lasts only for a finite time, or shows markedly different behavior in different time periods. Using the symphonic analogy, wavelets can be thought of representing the symphony by transformations of a basic wavelet, $w(t)$. So at $t = 0$, if cellos played the same tune twice as fast as the double basses, then the cello would be playing $c_1 w(2t)$ while the double basses play $b_1 w(t)$, presumably with $b_1 \geq c_1$. At $t = 1$, the next bass plays $b_2 w(t-1)$, and the next cello plays $c_2 w(2t-1)$ starting at $t = 0.5$. Notice that we need twice as many cellos to complete the symphony as double basses, as long as each cello plays the phrase once. If violas played twice as fast as cellos and violins twice as fast as violas, then obviously this would be 8 times as fast as double basses, and if there were such an instrument as a hyper-violin, then it would play 16 times faster than a double bass. In general the n-violins play "translations" of $w(2^n t)$.

In wavelet analysis, we only need to store the list of amplitudes, as the translations automatically double the frequency. With Fourier analysis a single disturbance affects all frequencies for the entire length of the series, so that although one can try and mimic a signal (or the symphony) with a complex combination of waves, the signal is still assumed to be "homogeneous over time". In contrast, wavelets have finite energy and only last for a short period of time. It is in this sense that wavelets are not homogeous over time and have "compact support". In wavelet analysis to approximate series that continue over a long period, functions with compact support are strung together over the period required, each indexed by location. In other words at each point in time, several wavelets can be

[1] An orthonormal tranform is one which preserves the energy of the series and is not affected by shifts in the data.

analysing the same variable or signal. The difference between Fourier analysis and wavelet anaysis is shown in diagrammatic form in figure 1.

2.1 Elementary wavelets

Wavelets have genders: there are father wavelets ϕ and mother wavelets ψ. The father wavelet integrates to 1 and the mother wavelet integrates to 0:

$$\int \phi(t)dt = 1 \tag{1}$$

$$\int \psi(t)dt = 0 \tag{2}$$

The father wavelet (or scaling function) essentially represents the smooth, trend (low-frequency) part of the signal, whereas the mother wavelets represent the detailed (high frequency) parts by scale by noting the amount of stretching of the wavelet or "dilation". In diagrammatic terms, father and mother wavelets can be illustrated for the Debechies wavelet, as shown in figure 2.

Wavelets also come in various shapes, some are discrete (as in the Haar wavelet - the first wavelet to be proposed many decades ago, which is a square wavelet with compact support), some are symmetric (such as the Mexican hat wavelet, symmlets and coiflets), and some are asymmetric (such as daublets)[2]. Four are illustrated in figure 3.

All the wavelets given above are mother wavelets. The upper left hand wavelet is a Haar wavelet, and this is a discrete symmetric wavelet. The upper right hand box contains a daublet, which is asymmetric, the lower left hand box contains a symmlet and the lower right hand box a coiflet. The latter two wavelets are nearly symmetric.

The dilation or scaling property of wavelets is particulary important in exploratory analysis of time series. Consider a double sequence of functions:

$$\psi(t) = \frac{1}{\sqrt{s}}\psi\left(\frac{t-u}{s}\right) \tag{3}$$

where s is a sequence of scales. The term $\frac{1}{\sqrt{s}}$ ensures that the norm of $\psi(.)$ is equal to one. The function $\psi(.)$ is then centered at u with scale s. In wavelet language, we would say that the energy of $\psi(.)$ is concentrated in a neighbourhood of u with size proportional to s. As s increases the length of support in terms of t increases So, for example, when

[2]Other wavelets also exist - notably Morlets, DOGS, Pauls, and bi-orthogonal wavelet functions.

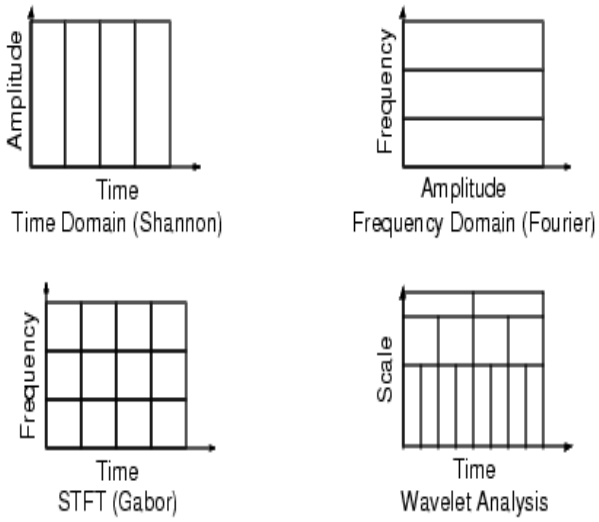


Figure 1: Fourier vs Wavelet analysis

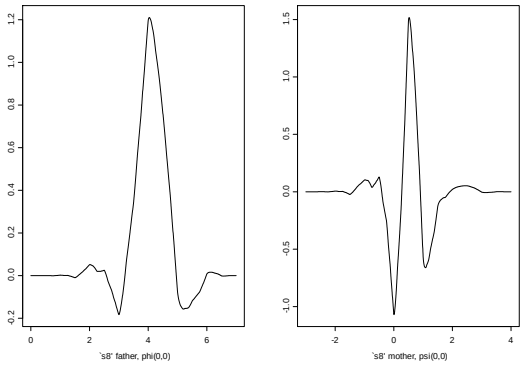


Figure 2: Mother and Father Wavelets

$u = 0$, the support of $\psi(.)$ for $s = 1$ is $[d, -d]$. As s is increased, the support widens to $[sd, -sd]$.[3] Dilation is particularly useful in the time domain, as the choice of scale indicates the "packets" used to represent any given variable or signal. A broad support wavelet yields information on variable or signal variations on a large scale, whereas a small support wavelet yields information on signal variations on a small scale. The important point here is that as projections are orthogonal, wavelets at a given scale are not affected by features of a signal at scales that require narrower support[4]. Lastly, using the language introduced above, if a wavelet is shifted on the time line, this is referred to as translation or shift of u. An example of the dilation and translation property of wavelets is shown in figure 4.

The left hand box contains symmlet of dilation 8, scale 1, shifted 2 to the right, while the right hand box contains the same symmlet with scale 2 and no translation.

2.2 Multiresolution decomposition

The main feature of wavelet analysis is that it enables the researcher to separate out a variable or signal into its constituent multiresolution components. In order to retain tractability (- many wavelets have an extremely complicated functional form), assume we are dealing with symmlets, then the father and mother pair can be given respectively by the pair of functions:

$$\phi_{j,k} = 2^{-\frac{j}{2}} \phi(\frac{t - 2^j k}{2^j}) \tag{4}$$

$$\psi_{j,k} = 2^{-\frac{j}{2}} \psi(\frac{t - 2^j k}{2^j}) \tag{5}$$

where j indexes the scale, and k indexes the translation. It is not hard to show that any series $x(t)$ can be built up as a sequence of projections onto father and mother wavelets indexed by both j, the scale, and k, the number of translations of the wavelet for any given scale, where k is often assumed to be dyadic. As shown in Bruce and Gao (1996), if the wavelet coefficients are approximately given by the integrals:

[3]The rescaling characteristic of wavelets in the time domain is thus equivalent to the rescaling of frequencies in Fourier analysis.

[4]It is only true in the simplest of cases, however, that large scale wavelets are associated with low frequencies. Consider for example, a given scale, say 2^3, or 8 months, and that the signal has components at that scale. It could still be true that the signal could contain projections that oscillate with any period greater than 8 months as well.

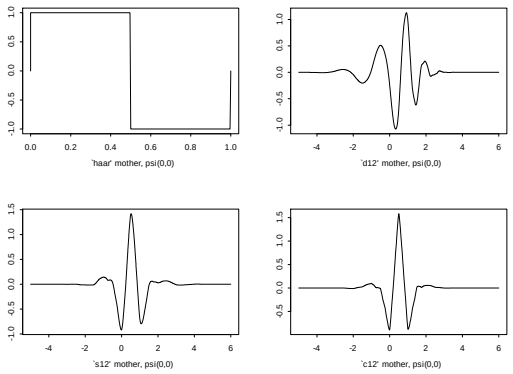


Figure 3: Families of wavelets

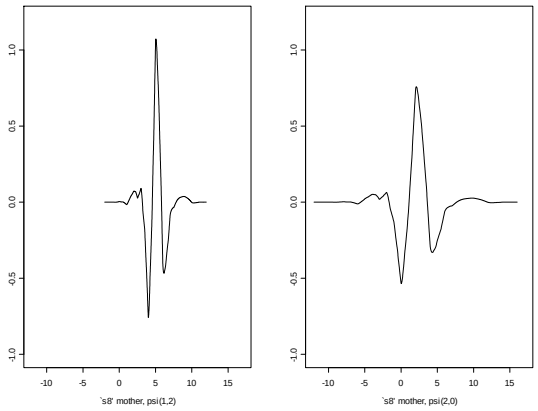


Figure 4: Scaled and translated Symmlet "s8" wavelets

$$s_{J,k} \approx \int x(t)\phi_{J,k}(t)dt \tag{6}$$

$$d_{j,k} \approx \int x(t)\psi_{j,k}(t)dt \tag{7}$$

$j = 1, 2, ...J$ such that J is the maximum scale sustainable with the data to hand. A multiresolution representation of the signal $x(t)$ is now be given by:

$$x(t) = \sum_k s_{J,k}\phi_{J,k}(t) + \sum_k d_{J,k}\psi_{J,k}(t) + \sum_k d_{J-1,k}\psi_{J-1,k}(t) + ... + \sum_k d_{1,k}\psi_{1,k}(t) \tag{8}$$

where the basis functions $\phi_{J,k}(t)$ and $\psi_{J,k}(t)$ are assumed to be orthogonal, that is:

$$\begin{array}{c} \int \phi_{J,k}(t)\phi_{J,k'}(t) = \delta_{k,k'} \\ \int \psi_{J,k}(t)\phi_{J,k'}(t) = 0 \\ \int \psi_{J,k}(t)\psi_{J',k'}(t) = \delta_{k,k'}\delta_{j,j'} \end{array} \tag{9}$$

where $\delta_{i,j} = 1$ if $i = j$ and $\delta_{i,j} = 1$ if $i \neq j$. Note that when the number of observations is dyadic, the number of coefficients of each type is given by:

- at the finest scale 2^1 :there are $\frac{n}{2}$coefficients labelled $d_{1,k}$.
- at the next scale 2^2 :there are $\frac{n}{2^2}$coefficients labelled $d_{2,k}$.
- at the coarsest scale 2^J:there are $\frac{n}{2^J}$coefficients $d_{J,k}$ and $S_{J,k}$

In wavelet language, each of these coefficients is called an "atom" and the coefficients for each scale are termed a "crystal"[5]. The multiresolution decomposition (MRD) of the variable or signal $x(t)$ is then given by:

$$\{S_J, D_J, D_{J-1}, ...D_1\} \tag{10}$$

For ease of exposition, the informal description above assumes a continuous variable or signal, which in signal processing is usually the case, but in economics although variables we use for analysis represent continuous "real time signals", they are invariably sampled at pre-ordained points in time. The continuous version of the wavelet transform (known as the CWT) assumes an underlying continuous signal, whereas a discrete wavelet transform (DWT) assumes a variable or signal consisting of observations sampled at evenly-spaced

[5]Hence the atoms make up the crystal for each scale of the wavelet resolution.

points in time. Apart from a later section in the paper, only the DWT (or variations on the DWT) will be used from this point onwards.

The interpretation of the MRD using the DWT is of interest in terms of understanding the frequency at which activity in the time series occurs. For example with a monthly or daily time series table 1 shows the interpretation of the different scale crystals:

Scale crystals	Annual frequency resolution	Monthly frequency resolution	Daily frequency resolution
d1	1-2	1-2	1-2
d2	2-4	2-4	2-4
d3	4-8	4-8	4-8
d4	8-16	8-16=8m-1yr4m	8-16
d5	16-32	16-32=1yr4m-2yr8m	4-8=4d-1wk 3d
d6	32-64	64-128=6yrs-10yr8m	32-64=6wks 2d-12wks 4d
d7	64-128	128-256=10yr8m-21yr4m	64-128=12wks 4d-25wks 3d
d8	128-256	etc	etc

Table 1: Frequency interpretation of MRD scale levels

Note four things from table 1:

i) the number of observations dictates the number of scale crystals that can be produced - only j scales can be used given that the number of observations, $N \geq 2^j$. Using the example given in this paper of industrial production, we have 1024 monthly observations, so the maximum number of scales is in theory $j = 9$. If there are only 255 observations, then no more then 7 crystals can be produced, but as the highest scale (lowest frequency crystal) can only just be properly resolved, it is usually recommended that only 6 crystals be produced. With 512 observations, although the d7 crystal can be produced, the trend crystal (or "wavelet smooth"), denoted s7 will yield further fluctuations from trend for all periodicities above a 256 month period.

ii) the choice of wavelet used in the analysis also figures into the number of scale crystals that can be produced. Say you have a choice between an "s4" and an "s8" wavelet for the MRD analysis. An "s4" wavelet means that the symmetric wavelet starts with a width of 4 observations for its support - this corresponds to the wavelet used to obtain the d1 crystals.[6] Using an s4 wavelet with annual or quarterly data will

[6]Care should be taken not to confuse what the letter "d" means here. The d4 crystal refers to the detail crystal at the 4th scale, where as a d4 wavelet refers to a Debauchies wavelet of length 4.

still capture the correct periodicities, but will enable the researcher to decompose to higher order scales. Clearly the higher the frequency of data, the more likely the researcher is to use a longer supported wavelet though, as very short wavelets are unlikely to yield any additional information.

iii) wavelet MRD analysis assumes that data is sampled at equally spaced intervals. The frequency resolution interpretation is more difficult with daily data, as daily (or hourly or even more frequently sampled) data is not evenly sampled. Note that with yearly data the resolution limit on a d3 crystal is half the time period on the minimum frequency picked up by a d4 crystal. Clearly this would not be the case when using daily data though.

iv) existing stylized facts need to be taken into account when applying an MRD to economic data. For example as economists know that business usually last for a decade at the most, it does not make sense to decompose a series beyond this level - so with annual data it wouldn't make sense to use anything more than the d4 scale crystals - doing so would cause "redundancy". So for example, with monthly economic data, if business cycles and their sub-cycles are to be identified, then it is desirable to have at least 512 observations.

Obviously the first and fourth point noted above pose major constraints for MRD with economic data, as virtually no annual economic time series contain more than 100 observations, and very few quarterly data series contain more than 200 observations. In finance, when using high frequency data, the MRD yields more information on activity at many different scale levels in the data, perhaps explaining the more frequent usage of wavelet analysis in this domain. With most economic and finance data of a reasonable time span, choice of wavelet type doesn't make a significant difference to the MRD (- perhaps with the exception of the Haar wavelet).

2.3 Examples

2.3.1 Continuous variable: a noisy doppler

A doppler signal[7] is defined by:

$$x(t) = \sqrt{t(1-t)}\sin(\frac{2.1\pi}{t+0.05}) \tag{11}$$

[7] The "waning" of a signal heard when for example passing a police siren.

Hence a doppler is sinusoid with a changing amplitude and decreasing frequency. If we now generate a doppler signal with 1024 datapoints and add white noise to the doppler signal the MRD plot using a symmlet is obtained in figure 5.

The first line gives the signal, then the plots labelled D1-D4 plot the crystals (or detail signals) associated with the scale of resolution, while the final plot gives the coarseset approximation of the signal. Clearly the MRD separates out the noise in the D1-D4 plots, with most of the action appearing within the first 200 observations of the plot. Beyond this, crystals D1 and D2 contain mostly white noise, while D3 and D4 contain hardly any of the signal and S4 contains the majority.

2.3.2 Discretely sampled variable: Canadian industrial production

Canadian industrial production data is available on a monthly basis from 1919 to 2004 by splicing together three Statistics Canada series. The data was used in annual percent change format, giving 1014 datapoints. Because of the advantages of using dyadic series in wavelet analysis, the series was padded with extra data (- the August value was continued through to the end of the series) so as to render a series of 1024 datapoints. The data is shown in appendix A, together with data for US industrial production and Finnish industrial production. It can be immediately seen that the index was extremely volatile during the inter-war years, and then also during the second world war, but stabilised in the late 1940s.

Spectral analysis reveals that there appears to be more than one cycle active in this series, and using a spectral analysis of autocovariances this is confirmed for both the Canadian industrial production series and the US industrial production series, as shown in figure **??** . Inspection of the figures shows that i) there appears to be activity at 5 different frequencies in the series, and that ii) clearly the US industrial production series also exhibits the same frequency patterns.

When wavelet MRD analysis is used for the Canadian series, 7 scales are used, which then encompasses business cycles with frequencies of up to nearly 11 years. The results are illustrated in figure 8 which indicates that small scale frequency changes mostly took place at the beginning of the series in the 1920s and 1930s but that other cycles have been active in the data since.the 1920s and 30s.

In figure 8, as wavelets for each scale are convolved with the data, crystal values are given at increasingly large intervals - these are the "spikes" that appear in the multiresolution decomposition in the stack plot, where the crystals at each scale level are plotted in ascending order. The crystals d1, d2 contain mostly noise, d3 appears to have contained

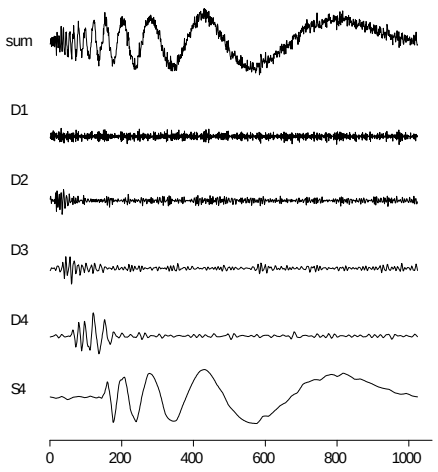


Figure 5: MRD of a noisy doppler signal

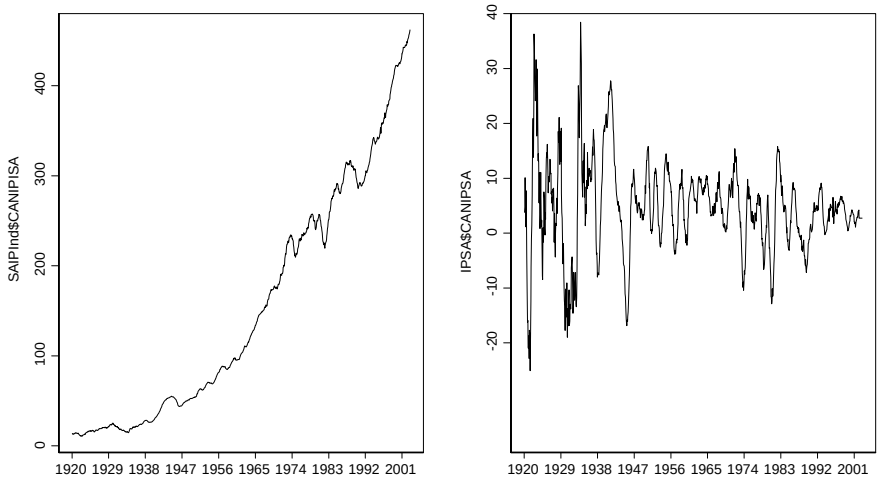


Figure 6: Canadian Industrial Production (sa): 1919-2004

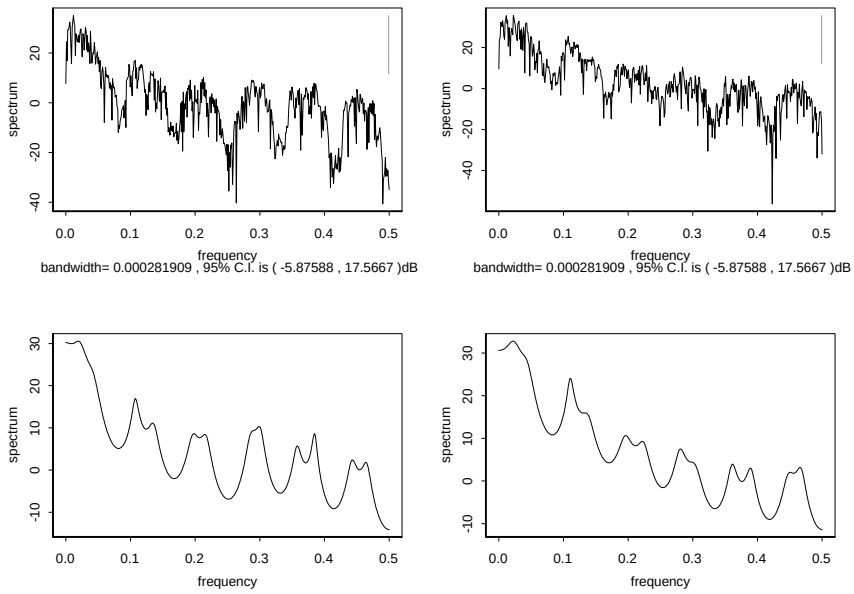


Figure 7: Spectra for Canadian and US industrial production

some explanatory power during the 20s and 30s, but now most variability is to be found in crystals d4 to d7 and s7. These cyclical interpretation of these crystals corroborates the spectra obtained in figure 7 as 4 separate cycles are identified (crystals d4, d5, d6 and d7), with crystals d1 to d3 containing mostly noise. Crystal s7 is interpreted as a trend or drift variable, and does not capture any distinct cycles in the data.

There are various ways of looking at the resulting DWT in graphical terms as shown in figure 9.

First, a in the upper left plot, a time-frequency graph shows activity in each scale through time. The y-axis is the inverse of scale - so large-scale (dilated) wavelets are given by flat wide boxes, and small-scale (very compact) wavelets are given by tall thin boxes. Each box should have the same area. This illustrates the localization in time of both large-scale and small-scale wavelets. Figure 10 shows this idea schematically.

The colour of the boxes reflects the size of the crystal coefficients - in this schema green, black and brown represent increasingly large coefficient sizes. Second, in the upper right hand panel, a dot chart shows the percentage of energy by crystal for scale j, E_j, which is given by:

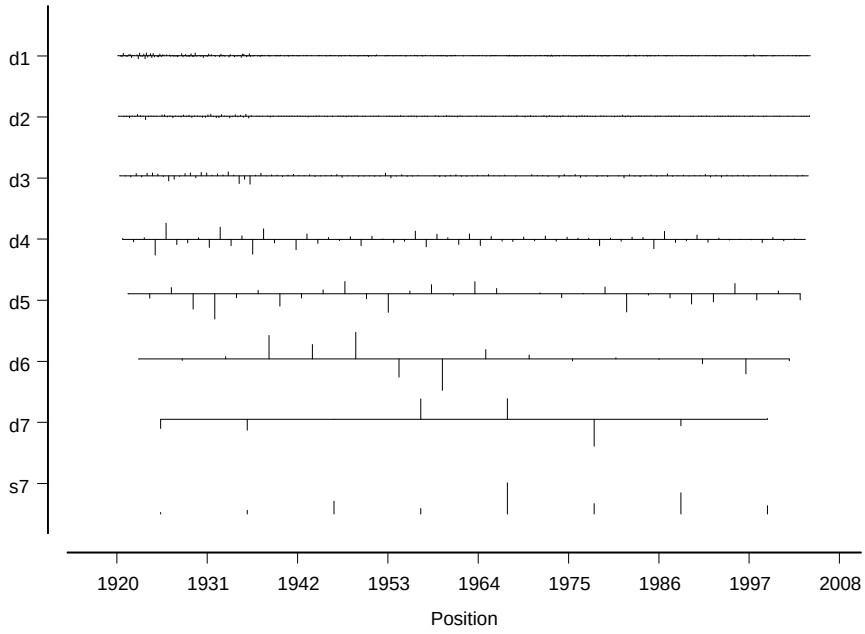


Figure 8: DWT for Canadian industrial production

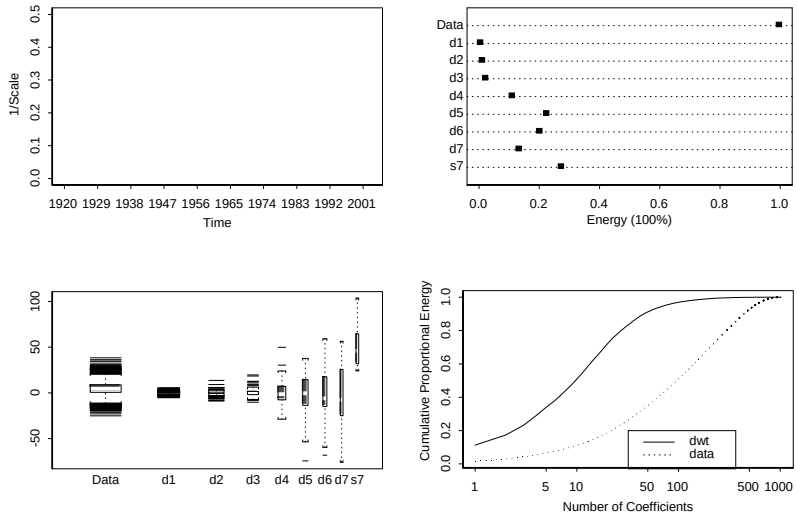


Figure 9: Summary of MRD for Canadian industrial production

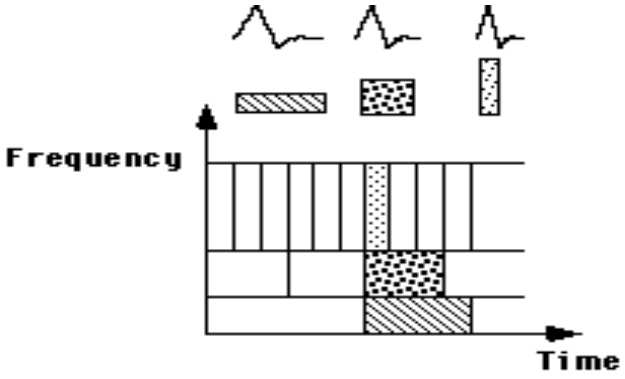


Figure 10: The time-frequency plot

$$E_j^d = \frac{1}{E} \sum_{k=1}^{\frac{n}{2^j}} d_{j,k}^2 \tag{12}$$

where d refers to the detail crystals, but the same can be done for the smoothness crystals. Orthogonal wavelets they are energy (variance) preserving, so that:

$$E = E_j^s + \sum_{i=1}^{j} E_j^d \tag{13}$$

As already noted, crystals d4, d5, d6 d7 and s7 contain most of the series energy. For the bottom left hand panel, a box plot should be self explanatory, and here the width of the boxes represent the number of data points or coefficients. Lastly the bottom right hand figure shows an energy plot which provides the cumulative energy function for the data and the DWT. Clearly the DWT contains the salient information about the series much better than the original untransformed data, which helps to explain why wavelets are so popular as a means of data compression.

Extracting information about individual crystals is also possible and easily implemented. For example in this instance, we have confirmed interest in crystals d4, d5 and d6. For example a panel plot for crystal d4 is illustrated in figure **??**. The upper left hand plot gives the d4 coefficients together with the reconstructed compoenent signal D4. The upper right hand plot is an ACF for the d4 coefficients with 95% confidence intervals. The bottom left hand plot is a quantile-quantile plot of the d4 coefficients versus those for a normal distribution. Clearly the distribution is fat-tailed at one end, which perhaps emphasises the role that the d4 crystal played in representing the increased volatility in the series in the earlier part of the series.

2.4 Multiresolution analysis (MRA)

The sequence of partial sums of crystals:

$$S_{j-1}(t) = S_J(t) + D_J(t) + D_{J-1}(t) + ... + D_j(t) \tag{14}$$

provides a multiresolution approximation (MRA) to the variable. This works by building up the variable from the highest numbered (coarsest) scale downwards. An MRA therefore could be viewed as a filtered version of the series which retains the most important parts of the series, but de-noises the series to a greater or lesser degree. Indeed, because wavelet analysis essentially filters certain information at different scales, many of those involved in the development of wavelets label them filters (of limited bandwidth)[8]. To show how the inverse discrete wavelet transform (IDWT) can approximate the series by acting as a band-pass filter, a smooth MRA representation of the data using the 4-6 scale crystals is calculated as follows:

$$S_4 = S_6 + D_4 + D_5 + D_6 \tag{15}$$

Clearly one could also reconstruct the signal again by adding lower scale crystals to equation 15. One interesting application for wavelet analysis, given that we have determined which crystals are most relevant for describing the Canadian industrial production, would be to invert the wavelet transform so as to reconstruct the series using an MRA. If this is done, figure 12 shows what is obtained. S7 represents both the s7 and d7 crystals combined, and then when d6 is superimposed, S6 is obtained etc.

Putting the wavelet analysis of the Canadian industrial production series together, it is now quite apparent (from figures 8, **??** and 12) that crystals d7, d6, d5, and d4 largely show the movement of the series. Adding d3 (to get S3 in figure 12), adds very little to explaining any movement in the series (except perhaps towards the beginning), and if shorter term fluctuations are desired, then clearly s2 captures some more of the noise but adds very little to the analysis. The real value added here is the recognition from figures **??** and 12 that there appear to be 5 different sources of variation in the series - one longer term, two medium term, one shorter term, and lastly very short-term "noise" variations, the latter appearing not to be cyclical in nature.

[8] In this parallel literature the mother wavelet is usually called a "wavelet filter" and the father wavelet is called a "scale filter", and a DWT can be thought of as an "octave band" filter bank (see Bruce, Hong-Ye, and Ragozin (1995))

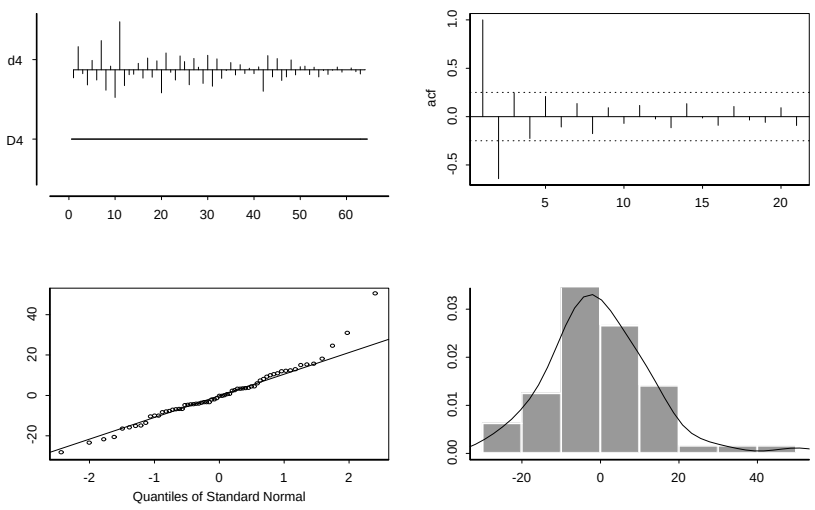


Figure 11: Canadian industrial production: d4 coefficient vector

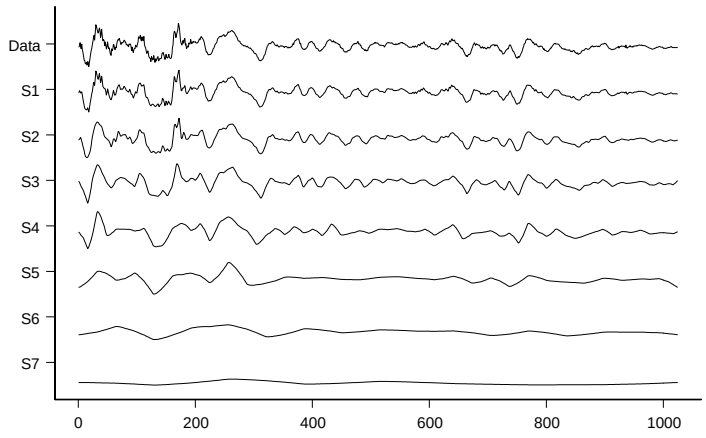


Figure 12: MRA for Canadian industrial production

3 How does a DWT work?

The principle behind the notion of the wavelet transform is deceptively simple, and originated in the pioneering work of Mallat (1989) in signal processing. The core of this approach is the usage of a "pyramid algorithm" which uses 2 filters at each stage (or scale) of analysis.figure 14 represents the pyramid algorithm approach for MRD, where L represents a low-band filter and H represents a high-band filter.

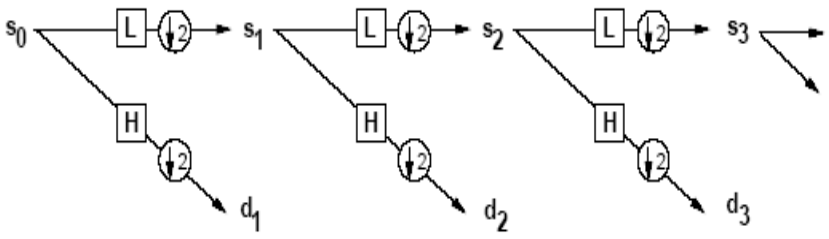


Figure 13: Mallat's DWT Pyramid Algorithm

If the input to the algorithm is $X = (x_1, x_2, ...x_n)$ and then define a filter to be $F = (f_1, f_2, ...f_m)$,then the convolution of the filter and variable is given by[9]:

$$y_t = \sum_{i=1}^{m} f_i x_t \tag{16}$$

If we use two filters, one low-pass (- a father wavelet) and one high-pass (mother wavelet) filter, then this will produce two series. Now drop every other data point in terms of the output from these two filters to get:

$$y_t = \sum_{i=1}^{m} f_i x_{2t+i-2} \tag{17}$$

The output will be s_1 for the low-pass filter and d_1 for the high-pass filter. These details coefficients are kept, and s_1 is now put through a further high-pass and low-pass filter, etc, to finally produce:

$$d_j = W_{H,\downarrow}(s_{j-1}) \tag{18}$$

[9] An interesting applet showing convolution is located at http://www.jhu.edu/~signals/discreteconv/ on the web.

a set of j detail coefficients and:

$$s_j = W_{L,\downarrow}(s_{j-1}) \tag{19}$$

a set of level j smooth coefficients. The choice of filter obviously aligns with the choice of wavelet here. This is the output given by the MRD of a variable. To construct an MRA an inverse DWT needs to be performed. In algorithm terms, this is shown as figure 14.Here the crystals are taken and convolved with a synthesis filter, and at the same time

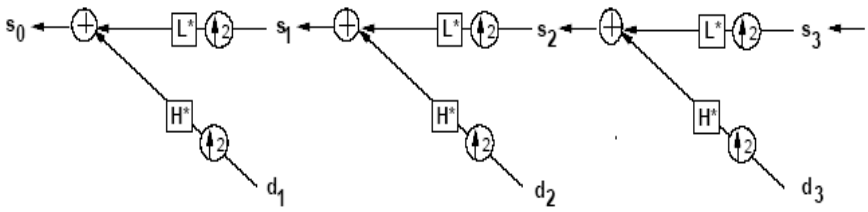


Figure 14: The IDWT (MRA) algorithm

"upsampled" by inserting zeros between every other value of the filter input. The smoothed coefficients for scale $j-1$ are obtained as:

$$s_{j-1} = W_{L,\uparrow}(s_j) + W_{H,\uparrow}(d_j) \tag{20}$$

where $W_{F,\uparrow}$ is an upsampling convolution operator for filter F.

4 Wavelet packet transforms

One important extension to wavelet analysis was introduced by Ronald Coifman and others Coifman, Meyer, Quake, and Wickerhauser (1990). Wavelet packets are a generalisation of wavelets, as they take a wavelet of a specific scale and add oscillations. Following the notation used above, in mathematical terms they can be represented as functions $W_{j,b,k}$where j corresponds to the scale/resolution level, k corresponds to the shift and b indicates the number of oscillations[10]. A discrete wavelet packet table is shown schematically in figure 15, unfortunately using slightly different notation than used in previous sections The first line of the figure gives the original data or variable. The data is first filtered (convolved) with

[10]Only for the Haar wavelet does b represent the number of zero crossings - for other wavelets the number of zero crossings is usually larger than b.

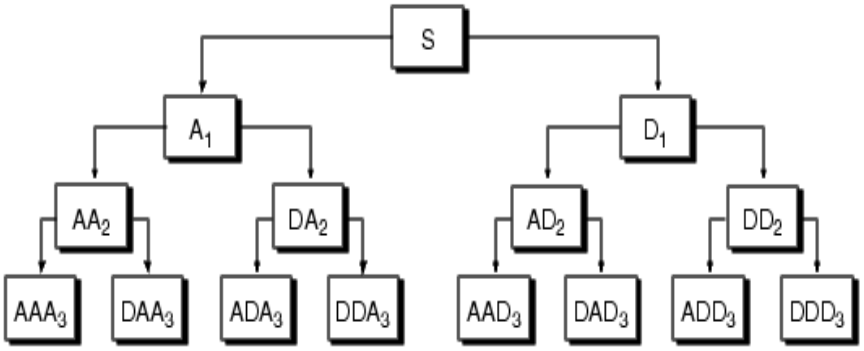


Figure 15: Wavelet packet tree

a high level filter to get $A_1$and a low level filter to get D_1 The wavelet packet transform then departs with the DWT by continue to apply high level and low level filters to these crystals, with the result that wavelets with oscillations are introduced. Mathematically these wavelet packet basis functions can be written as:

$$W_{j,b,k} = 2^{-j/2} W_b(2^{-j}t - k) \tag{21}$$

and the theoretical wavelet packet coefficients can be given as approximately:

$$w_{j,b,k} \approx \int W_{j,b,k}(t) f(t) dt \tag{22}$$

Put another way, a wavelet packet crystal $\mathbf{w}_{j,b}$ can be written as a vector:

$$\mathbf{w}_{j,b} = (w_{j,b,1}, w_{j,b,2}, ..., w_{j,b,\frac{n}{2^j}})' \tag{23}$$

where $\mathbf{w}_{j,b}$ is the result of selecting n linearly independent rows from a matrix $\mathbf{W}$, such that:

$$\mathbf{w} = \mathbf{W}\mathbf{x} \tag{24}$$

where $\mathbf{x}$ is the original signal/series.

Wavelet packet functions are illustrated in figure 16.

The first step in doing wavelet packet analysis is to use a wavelet packet table. Suppose the series of interest has n observations, where n is a muliple of 2^j(i.e. dyadic), then the wavelet packet table will have $J + 1$ resolution levels where J is the maximum resolution

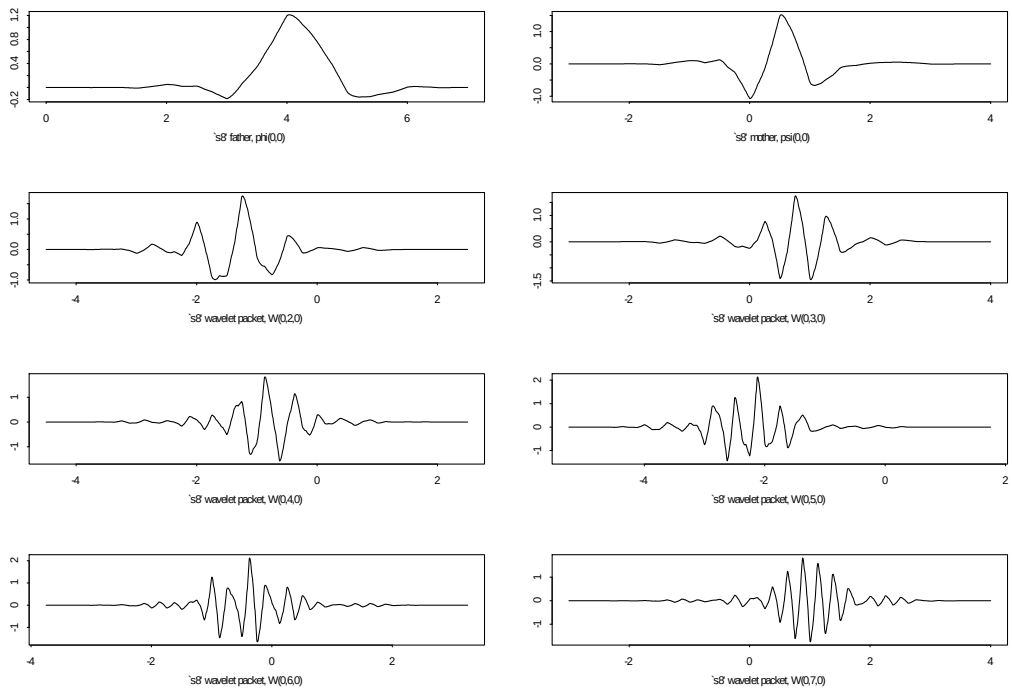


Figure 16: Debauchies wavelet packet functions

level. If the $(J+1)$ resolution levels are stacked in order a table of $n*(J+1)$ coefficients is obtained. At level J the table has n coefficients divided into 2^jcoefficient blocs (crystals). In other words, each row represents a certain scale, and as you read across the wavelet packet table you see wavelet filters with increasing large numbers of oscillations at that scale. To illustrate, a generated linear chirp signal of 1024 datapoints is analysed in a wavelet packet table in figure 17 and then the Canadian industrial production series is also analysed in figure 18

Figure 17 shows the crystals for each wavelet basis in a box bordered by dots. The series is well represented by certain wavelets, and notably those wavelets that have a lower number of osceillatoins, although the higher the scale level, the more larger oscillation wavelets appear to have crystal values significantly different from zero. Take level 2 for example: here there are 4 crystals, $w_{2,0}, w_{2,1}, w_{2,2}, w_{2,3}$- as the chirp oscillates more rapidly so the coefficients in $w_{2,0}$seem to wane, only for the coefficients in $w_{2,1}$ to provide a better fit. Clearly $w_{2,2}$,and $w_{2,3}$ have little explanatory power, although towards the end of the series the values of these wavelet coefficients appear to start to respond.

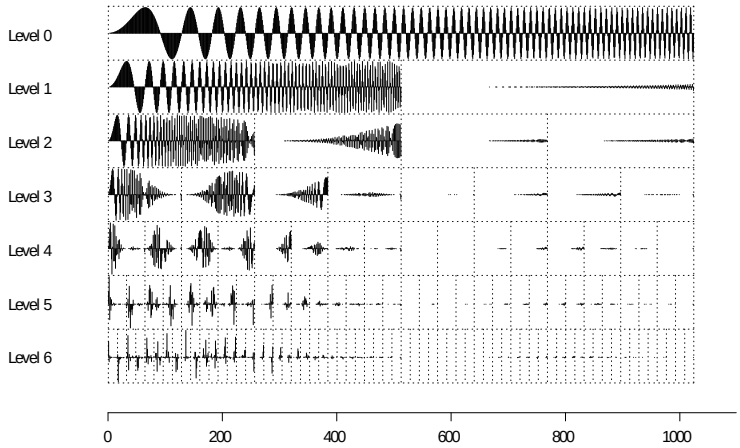


Figure 17: Linear chirp wavelet packet table

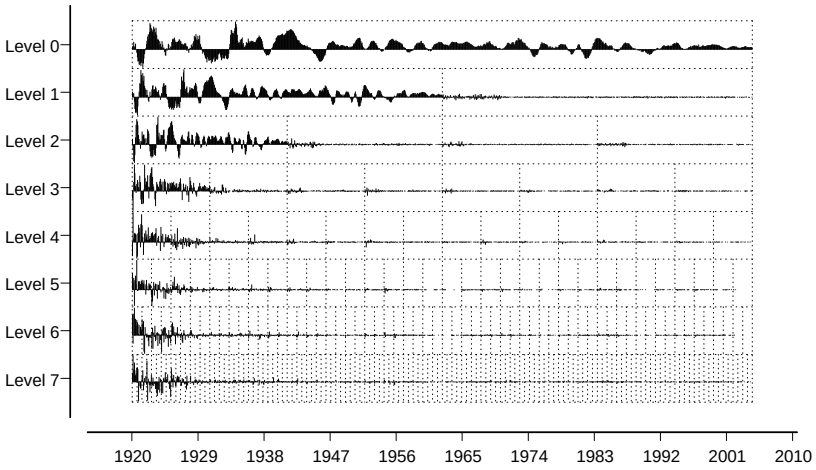


Figure 18: Wavelet packet table for Canadian industrial production

When turning to Canadian industrial production data, figure 18 tends to suggest that the zero oscillation wavelets characterise the series fairly well, so the improvement over a DWT by using a packet transform might be significant, as the DWT uses single oscillation wavelets at all scales. One of the more interesting uses of wavelet packet transforms is to characterise series in terms of a particular set of crystals of a certain scale. So for example, using an inverse wavelet packet transform with say level 4 crystals, a reconstruction of a series can be made. This is done in figure 19 for the Canadian industrial production series.

Figure 20 presents the time-frequency plot for the level 4 packet transform for the Canadian industrial production series, and compares it with the original time-frequency plot for the DWT.

First note that in the left hand plot in the panel, the boxes are exactly the same size - that is because all the crystals are level 4, but representing different numbers of oscillations. In figure 19 there are 16 crystals, and these are stacked up in the left hand panel of figure 20 - once again, wavelets with large numbers of oscillations tend to have significant coefficients only for the first part of the series.

5 Optimal transforms

Coifman and Wickerhauser (1992) developed a "best basis" algorithm for selecting the most suitable bases for signal representation using a wavelet packet table. The best basis algorithm finds the wavelet packet transforms W that minimises a cost function E:

$$E(W) = \sum_{j,b \in I} E(w_{j,b}) \tag{25}$$

where I is the set of index pairs $(j, b)$of the crystals in the transform W. Typically, the entropy cost function is used, in which case the cost function is of the form:

$$E_{j,b}^{entropy} = \sum_{k} \left[\frac{w_{j,b,k}}{\|\mathbf{w}_{0,0}\|_2} \right]^2 \log \left\{ \left[\frac{w_{j,b,k}}{\|\mathbf{w}_{0,0}\|_2} \right]^2 \right\} \tag{26}$$

where $w_{j,b,k}$ represents the crystal coefficients and $\|\|_2$ is the L_2 norm of the matrix $\mathbf{w}$ defined above in equation (equation 23). Other cost functions are also used in wavelet analysis, such as the threshold function (- which takes the coefficients with values greater than a certain threshold), Stein's unbiased risk estimate (SURE) function and the L_p.norm cost function. The entropy cost function essentially applies higher costs to packets with higher energy, thus favouring large crystal coefficients in packets with lower levels of energy.

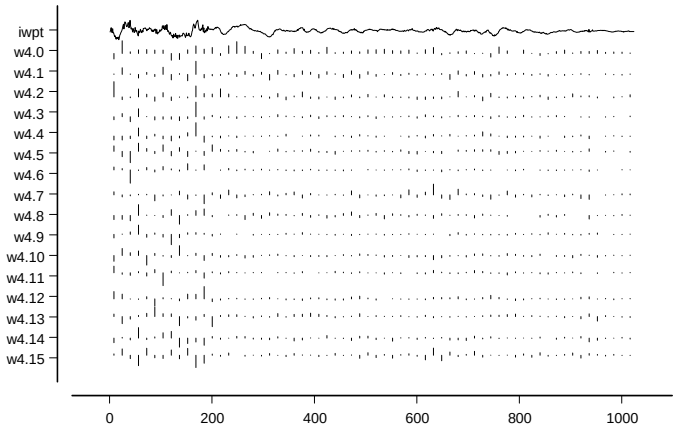


Figure 19: Level 4 packet transform for Canadian industrial production

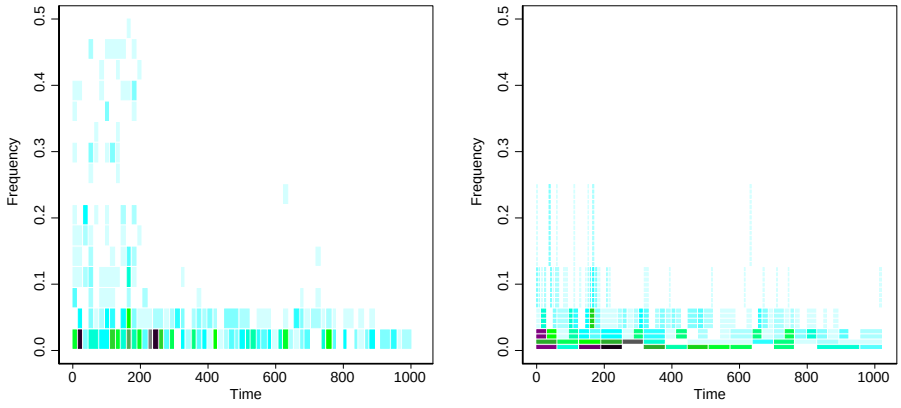


Figure 20: Time-freq plots for Canadian industrial production a) level 4 DWPT and b) DWT

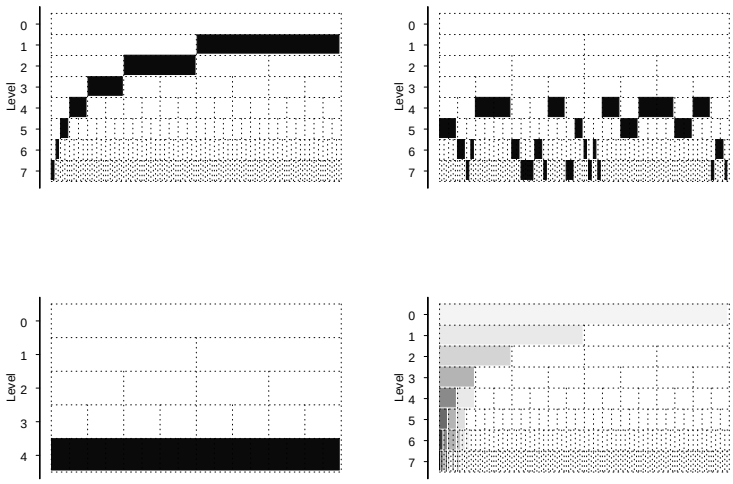


Figure 21: Wavelet packet tables for Canadian industrial production: a) DWT; b) best basis using entropy cost function; c) scale level 4 basis; d) entropy cost function basis

Once the cost function is applied to the series in question, a best basis plot is obtained. Of course, as the algorithm for choosing the best representation of the series uses orthogonal wavelet transforms, so the the choice of best wavelets must not overlap at each scale. This is shown in the wavelet packet table as a series of shaded boxes such that every column in the table is covered by one crystal so the series can be reconstructed, while at the same time no column has more than one crystal. To compare the best basis with a DWT and an arbitrary choice of crystal to represent the series (d4 here), figure 21 plots the best basis cost function for the Canadian series (lower right hand plot) with the wavelet packet representations for three other cases (top left hand plot: WPT; top right: best basis; bottom left: scale 4 crystals).

The best basis transform for Canadian industrial production clearly shows the best basis packets emanating from the crystals at scales 4 to 7, with scale 4 actually being more strongly represented than might have been suggested by the DWT. Once the best basis transform has been found.the original series can be plotted with an indication of which crystals best represent the variable/signal over various periods of time. .The coefficients in figure are ordered roughly by oscillation and scale.

Another way to view packet wavelet transforms is to display them in the form of a tree

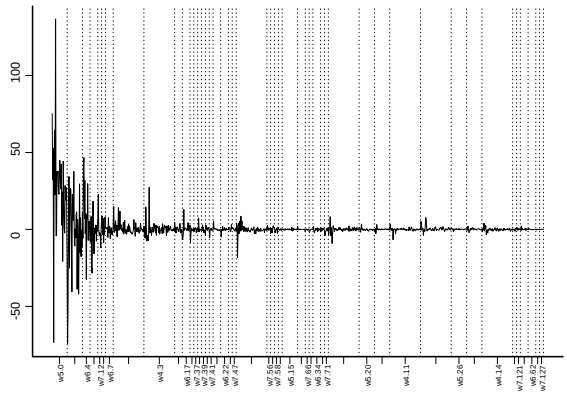


Figure 22: Time representation of Canadian industrial production using best basis packet selection.

plot - the tree plots for the best basis, the level 4 basis and the DWT are shown in figure 23. The length of the arc denotes the cost savings by using the alternative transforms (the two left hand plots) compared with the DWT (right hand plot). There are clearly considerable cost savings by using the best basis as compared with the DWT in this instance, as the higher scales crystals are not particularly relevant for the series in question here.

6 Other useful wavelets and wavelet transforms

6.1 Bi-orthogonal wavelets

These are wavelets that are not orthogonal.but are symmetric, and were introduced by Cohen, Daubechies, and Feauveau (1992) Biorthogonal wavelets are characterised by wavelets of four types: $\phi, \psi, \widetilde{\phi}$ and $\widetilde{\psi}$, where ϕ and ψ are the usual father and mother wavelets respectively, but biorthogonal wavelets necessitate the introduction of two new types of wavelets, $\widetilde{\phi}$ and $\widetilde{\psi}$, which are called the "dual" wavelets. These dual wavelets have different lengths of support so that while the original mother and father wavelets analyze the variable in question the dual wavelets act to synthesize the wavelets through time. The wavelet approximation can be expressed as a variation on the DWT, but using the dual wavelet functions, as in equation 27 :

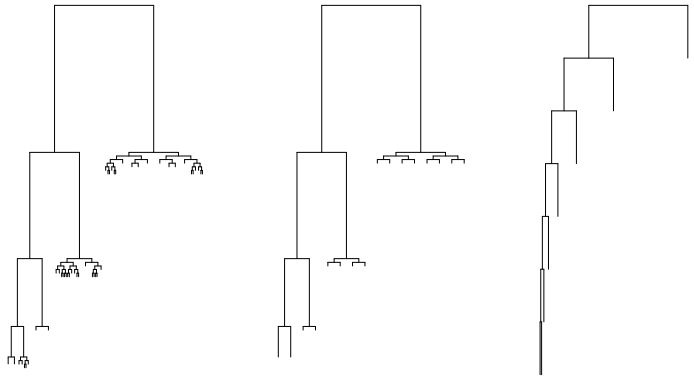


Figure 23: Tree plots for Canadian industrial production: a) best basis; b) level 4 crystal; c) DWT

$$f(t) = \sum_k s_{J,k}\widetilde{\phi}_{J,k}(t) + \sum_k d_{J,k}\widetilde{\psi}_{J,k}(t) + \sum_k d_{J-1,k}\widetilde{\phi}_{J-1,k}(t) + ... + \sum_k d_{1,k}\widetilde{\psi}_{1,k}(t) \qquad (27)$$

Biorthogonal wavelets satisfy the following conditions:

$$\begin{array}{c} \int \phi_{J,k}(t)\widetilde{\phi}_{J,k'}(t) = \delta_{k,k'} \\ \int \psi_{J,k}(t)\widetilde{\phi}_{J,k'}(t) = 0 \\ \int \psi_{J,k}(t)\widetilde{\psi}_{J',k'}(t) = \delta_{k,k'}\delta_{j,j'} \end{array} \qquad (28)$$

so that is, there is orthogonality between the father and mother dual wavelets (and vice-versa), but not between the mother and dual or father and dual. Note from equation 27 that if either of the duals are set equal to the usual father and mother wavelet, then we revert back to a DWT (as given by equation **??**. There are two popular forms of bi-orthogonal wavelets, namely B-spline and S-spline wavelets - with the former based on a simple polynomial spline function and the latter designed specifically to try and mimic an orthogonal wavelet. Examples of biorthogonal wavelets are given in figure 24. Bi-orthogonal wavelets are used in storing information in JPEG file formats.

6.2 Maximal-overlap DWT (MODWT)

Although extremely popular due to its intuitive approach, the classic DWT suffers from two drawbacks: dyadic length requirements and the fact that the DWT is non-shift invariant.

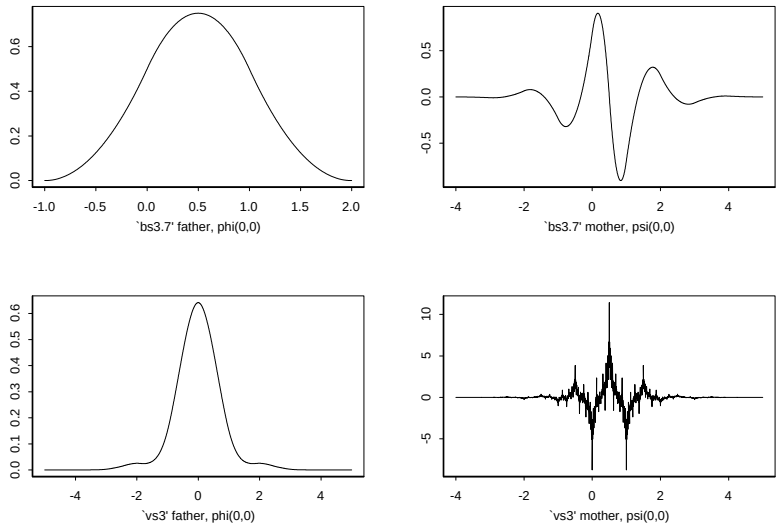


Figure 24: Examples of biorthogonal wavelets

In order to address these two drawbacks, the maximal-overlap DWT (MODWT)[11] gives up the orthogonality property of the DWT to gain other features, such as given in Percival and Mofjeld (1997) as:

- the ability to handle any sample size regardless of whether dyadic or not;
- increased resolution at coarser scales as the MODWT oversamples the data;
- translation-invariance - in other words the MODWT crystal coefficients do not change if the time series is shifted in a "circular" fashion; and
- the MODWT produces a more asymptotically efficient wavelet variance estimator than the DWT.

Both Gençay, Selçuk, and Whicher (2001) and Percival and Walden (2000) give a description of the matrix algebra involved in the MODWT, but for our purposes the MODWT can be described simply by referring back to figure **??**. In contrast to the DWT the

[11] As Percival and Walden (2000) note, the MODWT is also commonly referred to by various names in the wavelet literature. Equivalent labels for this transform are non-decimated DWT, time-invariant DWT, undecimated DWT, translation-invariant DWT and stationary DWT. The term "maximal overlap" comes from its relationship with the literature on the Allan variance (the variation of time-keeping by atomic clocks) - see Greenhall (1991).

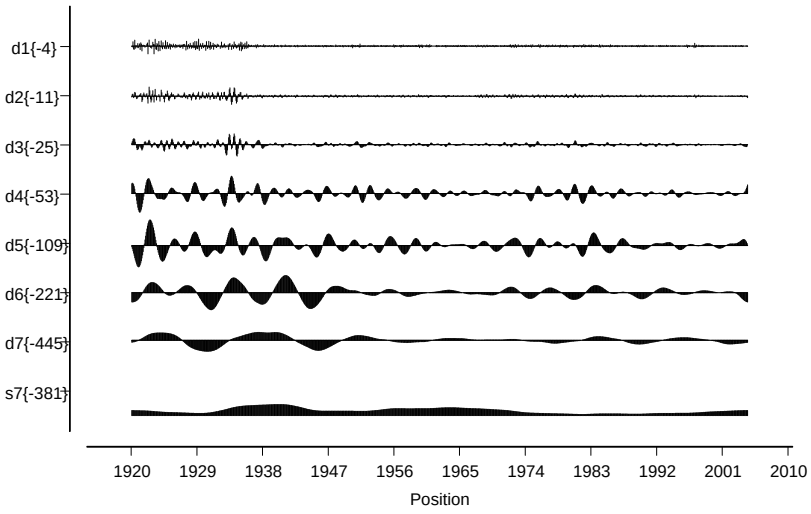


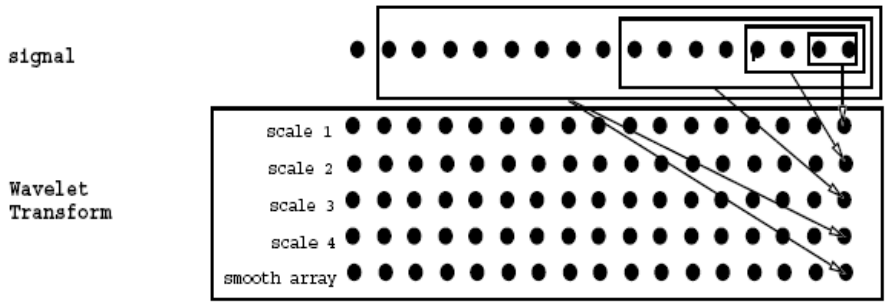
Figure 25: MODWT for Canadian industrial production

MODWT simply skips the downsampling after filtering the data, and everything else described in the section on MRDs using DWTs above follows through, including the energy (variance) preserving property and the ability to reconstruct the data using MRA with an inverse MODWT. A simple derivation of the MODWT following both Gençay, Selçuk, and Whicher (2001) and Percival and Walden (2000) can be found in appendix C.

Figure 25 shows the MODWT for Canadian industrial production. Clearly the resolution dramatically increases for the coarser scales, and now the intermediate cycles are more clearly apparent in the data.

One of the problems with the DWT is that the calculations of crystals occurs at roughly half the length of the wavelet basis (length) into the series at any given scale. Thus in figure 8 crystal coefficients start further and further along the time axis as the scale level increases. As the MODWT is shift invariant, the MRD will not change with a circular shift in the time series, so that each scale crystal can be appropriately shifted so that the coefficients approximately line up with the original data (known as "zero phase" in the signalling literature). This is done by shifting the scales to the left by increasingly large amounts as the scale order increases, and the y=axis of figure **??** shows.

Although the MODWT has a number of highly desirable properties, the transform leads to a large amount of "redundancy", as even though the transform is energy preserving, the



distribution of energy is clearly inferior to the original data. Similar analysis can be done with MODWTs as with DWTs and appendix C shows an example of a MODWPT, which as expected yields almost exactly the same results as obtained for its discrete counterpart.

6.3 "à trous" WT

For smaller time series, the scale crystals obtained may contain few data points, another simpler approach sometimes used in wavelet analysis which pre-dates the MODWT is called the *à trous* non-decimated wavelet approach. The algorithm over-samples at coarser scales so as to provide better spatial resolution but it does this by calculating the detail coefficients as simple differences between the smooth coefficients at different levels:

$$d_{j,k} = s_{j-1,k} - s_{j,k} \tag{29}$$

While this is computationally simpler, the MODWT possesses other properties which make it superior in most cases to the *à trous*, such as the shift invariance property referred to above. Figure 6.3 shows the construction of the *à trous* WT in diagrammatic terms.

6.4 Matching pursuit decompositions

This is another form of non-orthogonal wavelet decomposition, but using so-called waveform "dictionaries". The original idea here was to represent a series using linear combinations of a small number of wave-like functions (waveforms) selected from a large and redundant collection of functions. The original work on this algorithm started with Mallat and Zhang (1993) who used either wavlet packet tables, cosine packet tables, or gabor function tables to generate non-orthogonal "dictionaries" of waveforms to fit to any given series. The algorithm essentially takes each part of the series and attempts to fit a waveform to that

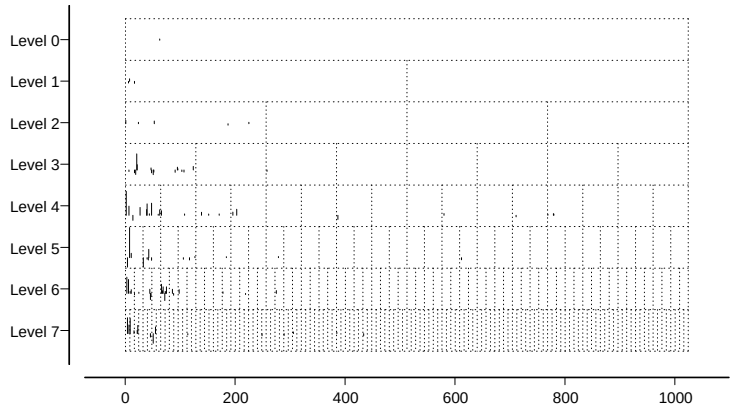


Figure 26: Matching pursuit decomposition for Canadian industrial production

part (- many wave shapes are used). The approximation to the actual series can be expressed as:

$$f(t) = \sum_{n=1}^{N} \alpha_n g_{\gamma_n}(t) + R_N(t) \tag{30}$$

where $g_{\gamma_n}(t)$ are a list of waveforms coming from a dictionary, α_n is the matching pursuit coefficient, and $R_N(t)$ is the residual which is still unexplained once the best atoms have been located. The best fitting waveforms from the dictionary are then preserved in a list, and are called "atoms". Figure 26 shows the best 100 atom coefficients for the Canadian industrial production series.

Most of the atoms are once again in the 4th to 7th crystals, and indeed if split down in terms of energy, 42% of the total energy is located in the scale level 4 atoms. It is noteworthy that the matching pursuit actually gives the scale 3 atoms 18% of the energy and scale 5 and 6 around 15% each, which is a somewhat different result that we obtained when using the DWT, as with the DWT the level 3 crystal had little explanatory power but the level 7 had significant explanatory power. Figure 27 compares the top 15 crystals from the matching pursuit algorithm with the top 15 from the DWT.

Clearly the matching pursuit algorithm does a better job of matching waveforms to the series than the DWT. It uses a much richer collection of waveforms than the DWT, which is always limited to one particular wavelet form.

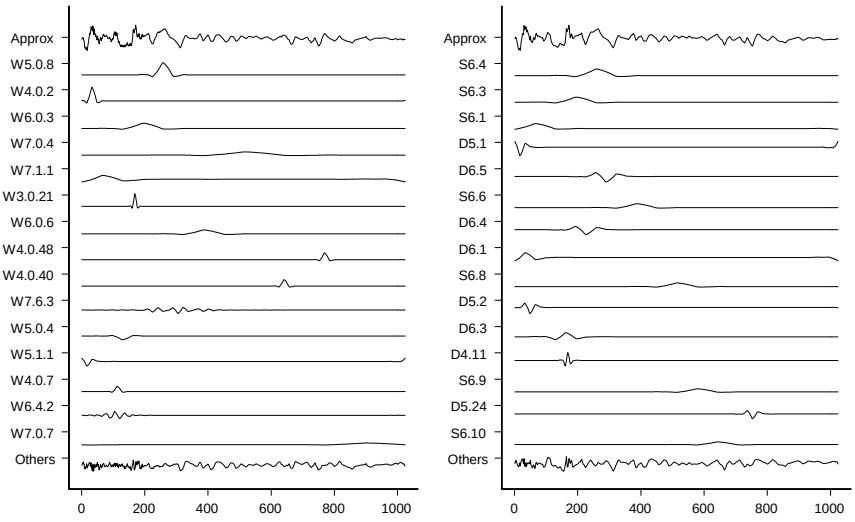


Figure 27: Top 15 atoms for Canadian industrial production: a) matching pursuit and b) DWT

6.5 Wavelet shrinkage

Donoho and Johnstone (1995) first introduced the idea of wavelet shrinkage in order to denoise a time series. The basic idea here is to shrink wavelet crystal coefficients, either proportionately, or selectively, so as to remove certain features of the time series. The initial version of the waveshrink algorithm was able to de-noise signals by shrinking the detail coefficients in the lower-order scales, and then applying the inverse DWT to recover a de-noised version of the series. This idea was then extended to different types of shrinkage function, notably so-called "soft" and "hard" shrinkage (reviewed in Bruce and Gao (1995)) . In mathematical terms, following the notation of Bruce, Hong-Ye, and Ragozin (1995) these different types of shrinkage can be defined by application of a shrinkage function to the crystal coefficients such that:

$$\widetilde{\mathbf{d}}_j = \delta_{j\sigma}(x)\mathbf{d}_j \tag{31}$$

where $\mathbf{d}_j$ is a vector of scale j crystal coefficients and $\delta_{j\sigma}(x)$ is defined as a shrinkage function which has as parameters the variance of the noise at level j, σ_j, and the threshold defined within the shrinkage function. The threshold argument can be defined in three different ways:

$$\delta_j^S(x) = \begin{cases} 0 & if \quad |x| \leq \widetilde{c} \\ sign(x)(|x| - \widetilde{c}) & if \quad |x| > \widetilde{c} \end{cases} \tag{32}$$

$$\delta_j^H(x) = \begin{cases} 0 & if \quad |x| \leq \widehat{c} \\ x & if \quad |x| > \widehat{c} \end{cases} \tag{33}$$

$$\delta_j^{SS}(x) = \begin{cases} 0 & if \quad |x| \leq c_L \\ sign(x)\frac{c_U(|x|-c_L)}{c_U-c_L} & if \quad c_L > |x| \leq c_U \\ x & if \quad |x| > c_U \end{cases} \tag{34}$$

where $\delta^S(x)$ is a generic soft shrinkage function, $\delta^H(x)$ is a generic hard shrinkage function and $\delta^{SS}(x)$ is a generic semi-soft shrink function for coefficients of any scale crystal. $\delta^H(x)$ essentially sets all crystal coefficients to zero below a specified coefficient value $\widetilde{c}$, $\delta^S(x)$ just lowers the crystal coefficients above a defined scale by the same amount, and the semi-soft version has the hard function property above scale c_U and below scale c_L but in between these two scales adopts a linear combination of the two approaches. There are a variety of ways of choosing $\widetilde{c}$, $\widehat{c}$ or $c_L$and c_U, some based on statistical theory and others rather more subjective. In some cases, such as the universal threshold case (setting $\widehat{c} = \sqrt{2\log n}$), the threshold is constant at all scale levels, and in other cases such as the "adapt" threshold, the threshold changes according to scale level.

7 Extensions

7.1 Wavelet variance, covariance and correlation analysis

7.1.1 Basic analysis

Given that wavelet analysis can decompose a series into sets of crystals at various scales, it is not such a big leap to then take each scale crystal and use it as a basis for decomposing the variance of a given series into variances at different scales. Here we follow a very simplified version according to Constantine and Percival (2003) which is originally based on Whitcher, Guttorp, and Percival (2000b) (with full-blown mathematical background provided in Whitcher, Guttorp, and Percival (1999)). Other more technical sources for this material are Percival and Walden (2000) and Gençay, Selçuk, and Whicher (2001).

Let x_t be a (stationary or non-stationary) stochastic process, then the time-varying wavelet variance is given by:

$$\sigma_{x,t}^2(\lambda_j) = \frac{1}{2\lambda_j} V(w_{j,t}) \tag{35}$$

where λ_j represents the jth scale level, and $w_{j,t}$ is the jth scale level crystal. The main complication here comes from making the wavelet variance time independent, the calculation of the variance for different scale levels (because of boundary problems) and accounting for when decimation occurs, as with the DWT. For ease of exposition, assume we are dealing with an MODWT, and assume that a finite, time-independent wavelet variance exists, then we can write equation 35 as:

$$\widetilde{\sigma}^2_{x,t}(\lambda_j) = \frac{1}{M_j} \sum_{t=L_j-1}^{N-1} d^2_{j,t} \tag{36}$$

where M_j is the number of crystal coefficients left once the boundary coefficients have been discarded. These boundary coefficients are obtained by combining the beginning and end of the series to obtain the full set of MODWT coefficients, but if these are included in the calculation of the variance this would imply biasedness. If L_j is the width of the wavelet (filter) used at scale j, then we can calculate M_j as $(N - L_j + 1)$[12].

Calculation of confidence intervals is a little more tricky. Here the approach is to first assume that $d_j \sim iid(0, \widetilde{\sigma}^2_j)$ with a Gaussian distribution, so that the sum of squares of d_j is distributed as $\kappa\chi^2_\eta$, and then to approximate what the distribution would look like if the d_j are correlated with each other (- as they are likely to be). This is done by approximating η so that the random variable $(\sigma^2_{x,t}\chi^2_\eta)/\eta$ has roughly the same variance as $\widetilde{\sigma}^2_{x,t}$ - hence η is not an actual degrees of freedom parameter, but rather is known as an "equivalent degrees of freedom" or EDOF. There are three ways of estimating the EDOF in the literature, and these can be summarised as i) based on large sample theory, ii) based on a priori knowledge of the spectral density function and iii) based on a band-pass approximation. Gençay, Selçuk, and Whicher (2001) show that if d_j is not Gaussian distributed then by maintaining this assumption this can lead to narrower confidence intervals than should be the case.

As an example of calculating wavelet variances for the three industrial production series, figure 28 shows the variances and confidence intervals by scale using a band-pass approximation to EDOF. The change in wavelet variance by scale is quite similar for Canada and the US, but Finland's wavelet variance appears not to show a similar pattern, even when the US series is adjusted so as to coincide with the Finnish IP period.

Once the wavelet variance has been derived, the covariance between two economic series can be derived, as shown by Whitcher, Guttorp, and Percival (2000b) (with mathematical

[12] $L_j = [(2^j - 1)(L - 1) + 1]$ as an L tap filter will clearly have a larger base at larger scales.

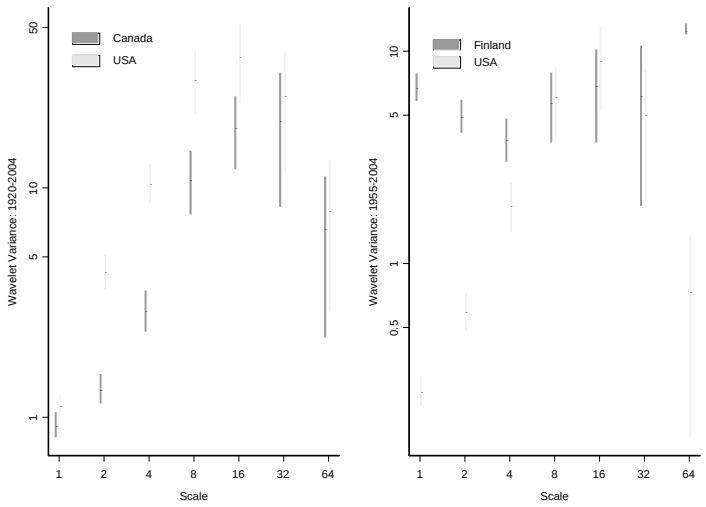


Figure 28: Wavelet variance by scale for Canada and Finland industrial production vs the US.

background provided in Whitcher, Guttorp, and Percival (1999)). The covariance of the series can be decomposed by scale, and thus different "phases" between the series can be detected. Figure 29 shows how the covariance between Canadian and US industrial production series breaks down by scale and then figure 30 shows cross covariances.between the Canadian and Finnish series and the US equivalent.

Once covariance by scale has been obtained, the wavelet variances and covariances can be used together to obtain scale correlations between series. Once again, confidence intervals can be derived for the correlation coefficients by scale (these are derived in Whitcher, Guttorp, and Percival (2000b)). The correlation between the Canadian and Finnish industrial production series and their US counterpart are estimated and plotted in figure 31 by scale, and then the cross-correlations for a 25 lag span.are plotted in figure 32.

Several interesting points emerge from figures 31 and 32:

- First, even for time series that are not particularly highly correlated such as that of Finland and the US, the cyclical nature of the co-correlation at every scale implies that there is co-movement of business cycles at a variety of different levels;

- Second, for Canada the co-correlation appears to increase as scale increases - this would be perhaps be expected for series that are highly correlated (or co-integrated)

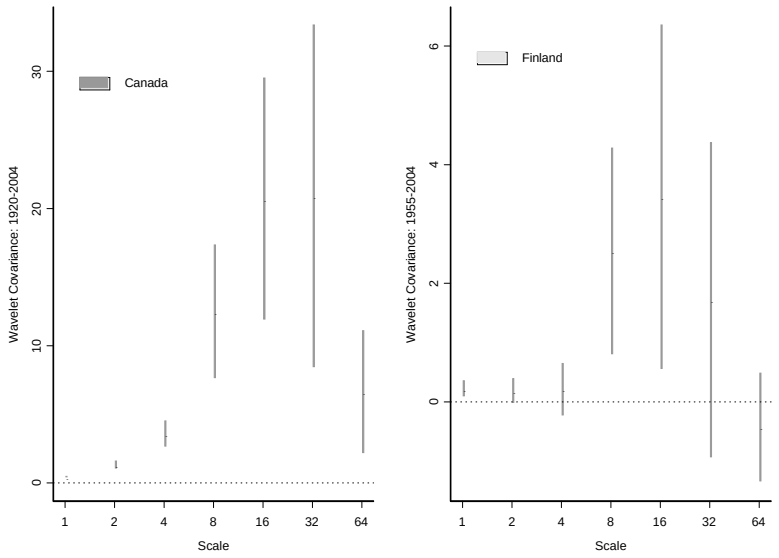


Figure 29: Wavelet covariance by scale for Canadian and Finnish industrial production vs the US.

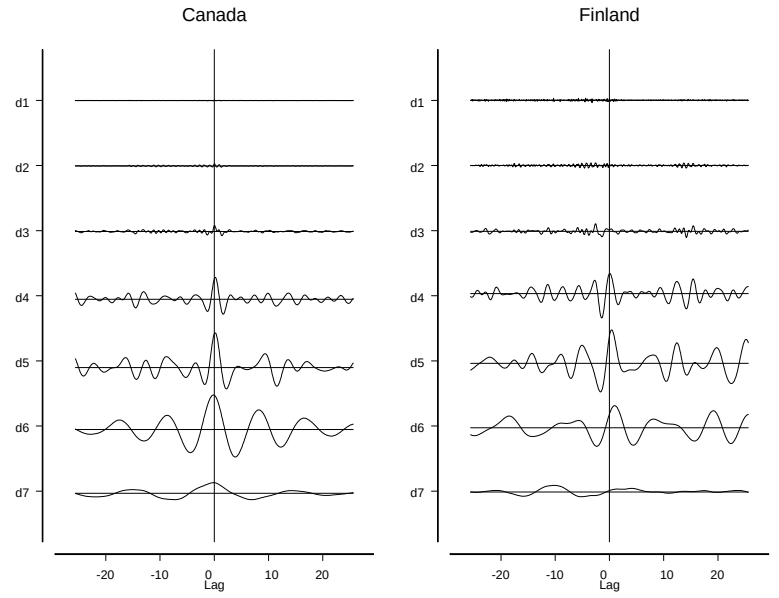


Figure 30: Cross-covariances for Canadian and Finnish industrial production series vs the US.

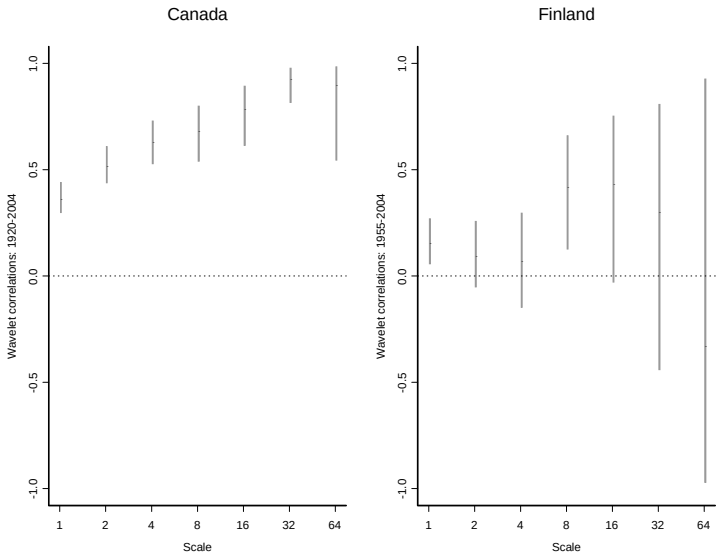


Figure 31: Correlation by scale for Canadian and Finnish industrial production vs the US.

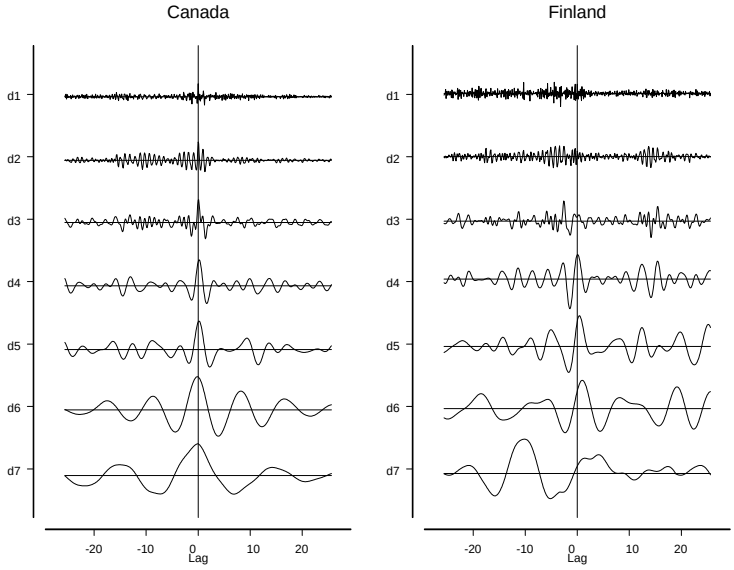


Figure 32: Cross-correlation for Canadian and Finnish industrial production vs the US.

over time; whereas for Finland, beyond scale 4, the co-correlation doesn't appear to follow any pattern by scale;

- Thirdly, as we are comparing crystals for increasingly wide wavelet support, we would expect the positive correlations to persist much longer as lag length is increased - this is found to be the case here;

- Lastly, at the same lag length the co-correlation can be quite different according to scale (- for example for Canada at a lag length of 12, scales 1-4 show negative co-correlation, scale 5 shows zero co-correlation and scales 6 and 7 exhibit positive correlation).

7.1.2 Testing for homogeneity

Whitcher, Byers, Guttorp, and Percival (1998) developed a framework for applying a test for homogeneity of variance on a scale-by-scale basis to long-memory processes. A good summary of the procedure is located in Gençay, Selçuk, and Whicher (2001). The test relies on the usual econometric assumption that the crystals of coefficients, $w_{j,t}$ for scale j at time t have zero mean and variance $\sigma_t^2(\lambda_j)$. This allows us to formulate a null hypothesis of:

$$H_0 : \sigma^2_{L_j}(\lambda_j) = \sigma^2_{L_j+1}(\lambda_j) = ... = \sigma^2_{N-1}(\lambda_j) \tag{37}$$

against an alternative hypothesis of:

$$H_A : \sigma^2_{L_j}(\lambda_j) = ... = \sigma^2_{k}(\lambda_j) \neq \sigma^2_{k+1}(\lambda_j) = ... = \sigma^2_{N-1}(\lambda_j) \tag{38}$$

where k is an unknown change point and L_j represents the scale once the number of boundary coefficients have been discarded.. The assumption is that the energy throughout the series builds up linearly over time, so that for any crystal, if this is not the case, then the alternative hypothesis is true. The test statistic used to test this is the D statistic, which has previously been used by Inclan and Tiao (1994) for the purpose of detecting a change in variance in time series. Define P_k as:

$$P_k = \frac{\sum_{j=1}^{k} w_j^2}{\sum_{j=1}^{N} w_j^2} \tag{39}$$

then define the D statistic as $D = \max(D^+, D^-)$ where $D^+ = (\widehat{L} - P_k)$ and $D^- = (P_k - \widehat{L})$ where $\widehat{L}$ is the cumulative fraction of a given crystal coefficient to the total number of

coefficients in a given crystal. Percentage points for the distribution of D can be obtained through Monte Carlo simulations, if the number of coefficients is less than 128 and is obtained through large sample properties. Table 2 shows the results of applying this test to the industrial production series at a 5% level of significance.

Scale	Canada IP	US IP	Finnish IP
1	F	F	F
2	F	F	F
3	F	F	F
4	F	F	T
5	T	F	T
6	F	F	T
7	T	T	NA

Table 2: Testing for Homogeneity

Gençay, Selçuk, and Whicher (2001) provides two examples of tests for homogeneity of variance - one for an IBM stock price and another for multiple changes in variance using methodology extending this framework which was developed by Whitcher, Guttorp, and Percival (2000a).

7.2 Wavelet analysis of fractionally differenced processes

Granger and Joyeux (1980) first developed the notion of a fractionally differenced time series and in turn, Jensen (1999) and Jensen (2000) developed the wavelet approach to estimating the parameters for this type of process. A fractionally differenced process generalises the notion of an ARIMA model by allowing the order of integration to assume a non-integer value, giving what some econometricians refer to as an ARFIMA process. In the frequency domain, this means that a fractionally differenced process has a spectral density function that varies as a power law over certain ranges of frequencies - implying that a fractionally differenced process will have a linear spectrum when plotted with log scales. A fractionally differenced process (FDP) can be defined as difference stationary if:

$$(1-B)^d x_t = \sum_{k=0}^{d} \binom{d}{k} (-1)^k x_{t-k} = \epsilon_t \tag{40}$$

where ϵ_t is a white-noise process, d is a real number and $\binom{d}{k} = \frac{d!}{k!(d-k)!} = \frac{\Gamma(d+1)}{\Gamma(k+1)\Gamma(d-k+1)}$, Γ being Euler's gamma function. Wavelets can be used to estimate the fractional difference parameter, d, as the variance of an ARFIMA type process can be expressed in the form:

$$\sigma_x^2(\lambda_j) \approx \sigma_\epsilon^2 \widetilde{c}(\widetilde{d})[\lambda_j]^{2\widetilde{d}-1} \tag{41}$$

where $\widetilde{c}$ is a power function of $\widetilde{d}$ and $\widetilde{d} = (d - 0.5)$. Equation 41 suggests that to estimate $\widetilde{d}$ a least squares regression could be done to the log of an estimate of the wavelet variance obtained from an MODWT. This represents the first way of estimating $\widetilde{d}$, but there is a problem as here the MODWT yields correlated crystals, so the estimate of d would likely be biased upwards.

Another way of estimating the parameters of the FDP is to go back to the DWT, as the scale crystals are uncorrelated, and use the likelihood function for the interior coefficients (- in other words after discarding the boundary coefficients). Let $\mathbf{d}_I$ be a vector of length M containing all the DWT wavelet coefficients for all j scales (for which there are M_j elements for each scale), then the liklihood function for σ_ϵ^2 and $\widetilde{d}$ will be:

$$\mathcal{L}(\widetilde{d}, \sigma_\epsilon^2 \mid \mathbf{d}_I) = \frac{\exp\{(-\mathbf{d}_I' \Sigma_{\mathbf{d}_I}^{-1} \mathbf{d}_I)/2\}}{(2\pi)^{M/2} \left|\Sigma_{\mathbf{d}_I}\right|^{1/2}} \tag{42}$$

where $\Sigma_{\mathbf{d}_I}$ is the variance covariance matrix. Using the fact that the wavelet coefficients at different scales are uncorrelated, we can use an approximation to the variance $C_j(\widetilde{d}, \sigma_\epsilon^2)$ so that equation 42 can be rewritten as:

$$\mathcal{L}(\widetilde{d}, \sigma_\epsilon^2 \mid \mathbf{d}_I) = \prod_{j=1}^{J} \prod_{t=0}^{M_j - 1} \frac{\exp\{-d_{j,t}^2 / 2C_j(\widetilde{d}, \sigma_\epsilon^2)\}}{(2\pi C_j(\widetilde{d}, \sigma_\epsilon^2))^{1/2}} \tag{43}$$

where $d_{j,t}$ is once again a crystal coefficient. Equivalently we can estimate a reduced log likelihood, such as:

$$\ell(\widetilde{d} \mid \mathbf{d}_I) = M \ln(\widetilde{\sigma}_\epsilon^2(\widetilde{d})) + \sum_{j=1}^{J} M_j \ln(\widehat{C}_j(\widetilde{d})) \tag{44}$$

where $\widetilde{\sigma}_\epsilon^2(\widetilde{d}) = \frac{1}{M} \sum_{j=1}^{J} \frac{1}{\widehat{C}_j(\widetilde{d})} \sum_{t=0}^{M_j - 1}$ and $\widehat{C}_j(\widetilde{d}) = C_j(\widetilde{d}, \sigma_\epsilon^2)/\sigma_\epsilon^2$. Minimizing equation 44 yields a maximum likelihood estimate for $\widetilde{d}$, $\widehat{\widetilde{d}}_{MLE}$, which can then be substituted into the definition of $\widetilde{\sigma}_\epsilon^2(\widetilde{d})$ to obtain an estimate for the variance. Percival and Walden (2000) devotes a whole chapter to the issue of FDPs and provides examples of MLEs for non-stationary processes, and Constantine and Percival (2003) provides routines for chopping the series up into blocks to obtain "instantaneous" estimators for the FDP parameters.

To illustrate wavelets applied to a financial time series, take the British pound daily series against the US dollar (details given in appendix D), and after taking log differences

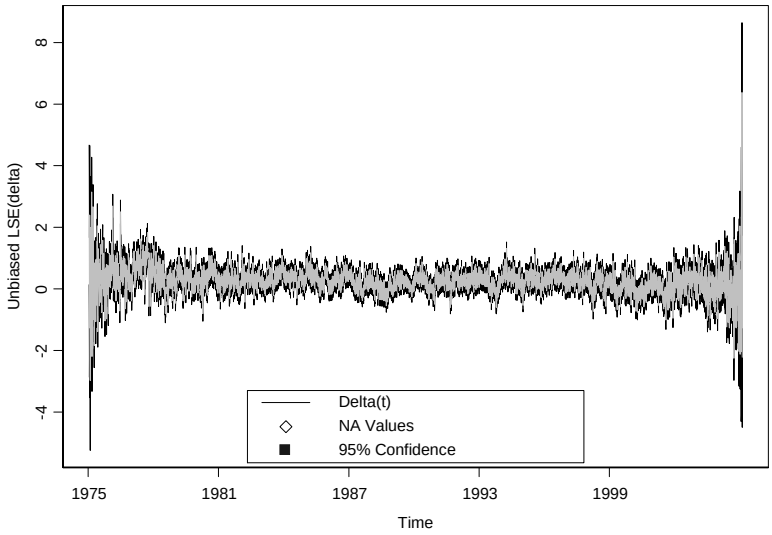


Figure 33: Instantaneous difference parameter estimates using LS for abs(GBP/USD)

of the series, take the absolute value of the log differences and use an instantaneous parameter estimation of the difference coefficient $\widetilde{d}$. This series will likely have long memory as volatility appears in clusters, so that the difference coefficient is unlikely to be an integer.

Figure 33 plots the instantaneous estimate of the difference coefficient together with a 95% confidence interval band. The mean of the parameter estimates is 0.236, and the mean of the variance estimates for the 10 scale levels estimated is 0.1002.

7.3 Dual Tree WTs using wavelet pairs

This is relatively recent research and uses a so-called dual tree wavelet transform (DTWT or complex wavelet transform), which was developed by Kingsbury (see Kingsbury (1998) and Kingsbury (2000)) with particular applications in image compression and reconstruction in mind. The DTWT differs from a traditional DWT in two distinct ways:

i) after the first application of a high filter (as in figure 13), then the data are not decimated and are handed to two different banks of filters - the even observations are passed to one set of filters that then convolve with the data in the usual DWT process (with decimation), but the other filter bank takes the odd numbered observations and convolves that set of data with the filters coefficients reversed.

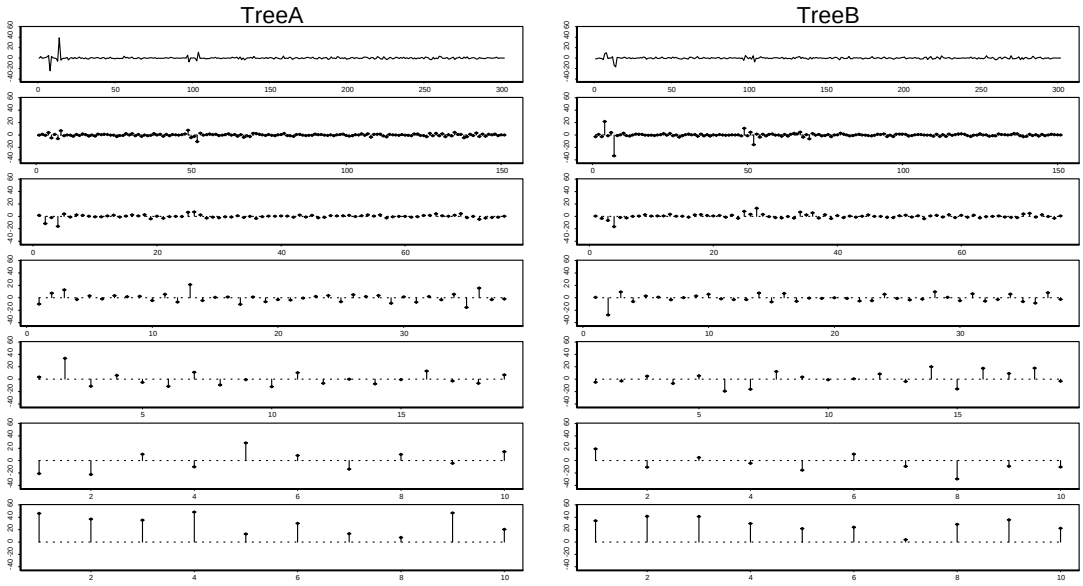


Figure 34: Dual Tree WT for Finnish Industrial Production

ii) the pair of filters have the character of "real" and "imaginary" parts of an overall complex wavelet transform. If the mother wavelets are the reverse of one another, they are called Hilbert wavelet pairs.

The big advantages of the DTWT are that they are shift invariant, without the "redundancy" associated with the MODWT. This is particularly useful for data and image compression applications. Figure 34 shows the DTWT for the Finnish industrial production series. The two branches of the DTWT yield quite different results, as might be expected given the somewhat erratic movements at the beginning of the series.

Kingsbury (2000) defines 4 types of biorthogonal filters for the first stage of the DTWT and then another 4 types of so called "Q shift" filters for the subsequent stages of the process. Two of these filters, one from each set, are shown in appendix B. Craigmile and Whitcher (2004) further extend the DTWT to a maximal overlap version (MODHWT) in order to produce wavelet based analogues for spectral analysis. This enables the frequency content between two variables to be analysed simultaneously in both the time and frequency domain.

8 Frequency domain analysis

Wavelet analysis is neither strictly in the time domain nor the frequency domain: it straddles both. It is therefore quite natural that wavelets applications have also been developed in the frequency domain using spectral analysis. Perhaps the best introduction into the theoretical side of this literature can be found in Lau and Weng (1995) and Chiann and Morettin (1998), while Torrence and Compo (1998) probably provides the most illuminating examples of applications to time series from meteorology and the atmospheric sciences.

Spectral analysis is perhaps the most commonly known frequency domain tool used by economists (see Collard (1999), Camba Mendez and Kapetanios (2001), Valle e Azevedo (2002), Kim and In (2003), Süssmuth (2002) and Hughes Hallett and Richter (2004) for some examples), and therefore needs no extended introduction here. In brief though, a representation of a covariance stationary process in terms of its frequency components can be made using Cramer's representation, as follows:

$$x_t = \mu + \int_{-\pi}^{\pi} e^{i\omega t} z(\omega) d\omega \tag{45}$$

where $i = \sqrt{-1}$, μ is the mean of the process, ω is measured in radians and $z(\omega)d\omega$ represents complex orthogonal increment processes with variance $f_x(\omega)$, where it can be shown that:

$$f_x(\omega) = \frac{1}{2\pi}\left(\gamma(0) + 2\sum_{\tau=1}^{\infty}\gamma(\tau)\cos(\omega\tau)\right) \tag{46}$$

where $\gamma(\tau)$ is the autocorrelation function. $f_x(\omega)$ is also known as the specturm of a series as it defines a series of orthogonal periodic functions which essentially represent a decomposition of the variance of the series into an infinite sum of waves of different frequencies. A given large value for equation, say at $f_x(\omega_i)$ implies that frequency ω_i is a particularly important component of the series.

To intuitively derive the wavelet spectrum, recall using the definition of a mother wavelet convolution from equation 5, that the convolution of the wavelet with the series renders an expression such as that given in equation 16, and associated crystals given by equation 18, then a continuous wavelet transform (CWT) for a frequency index k will analogously yield:

$$W_k(s) = \sum_{k=0}^{N} \widehat{x}_t \widehat{\psi}^*(s\omega_k) e^{i\omega_k t \partial t} \tag{47}$$

where $\widehat{x}_k$ is the discrete Fourier transform of x_t:

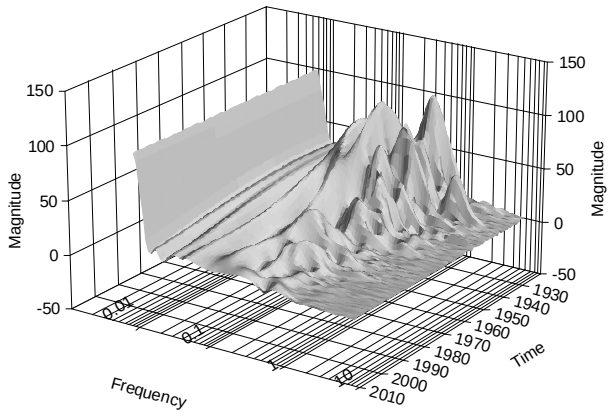


Figure 35: CWT 3D magnitude spectrum for US industrial production

$$\widehat{x}_k = \frac{1}{(N+1)} \sum_{k=0}^{N} x_t \exp\left\{\frac{-2\pi i k t}{N+1}\right\} \tag{48}$$

so $\widehat{x}_k$,represents the Fourier coefficients, $\widehat{\Omega}^*$ is the Fourier transform of a normalized scaling function defined as:

$$\Omega^* = \left\{\frac{t}{s}\right\}^{0.5} \Omega\left[\frac{t-\mu}{s}\right] \tag{49}$$

so that in the frequency domain the CWT $W_k(s)$ is essentially the convolution of the Fourier coefficients and the Fourier transform of Ω^*, $\widehat{\Omega}^*$. The angular frequency for ω_k in equation 47 is given by:

$$\omega_k = \begin{cases} \frac{2\pi k}{(N+1)} & : \quad k \leq \frac{N+1}{2} \\ -\frac{2\pi k}{(N+1)} & : \quad k > \frac{N+1}{2} \end{cases} \tag{50}$$

The CWT, $W_k(s)$, can now be split into a real and complex part, $\Re\left\{W_k(s)\right\}$, and $\Im\left\{W_k(s)\right\}$, or amplitude/magnitude, $|W_k(s)|$ and phase, $\tan^{-1}\left[\frac{\Re\{W_k(s)\}}{\Im\{W_k(s)\}}\right]$.

To show some examples of wavelet spectra, first figure 35 shows a 3-dimensional plot for the US industrial production series.using a log frequency scale and altering the reversing the time sequencing so as to show the evolution of the spectrum to best effect.

This CWT spectral plot displays similar characteristics to that of the time-frequency plots of figure 20, in that the high frequency content is particularly marked in the earlier

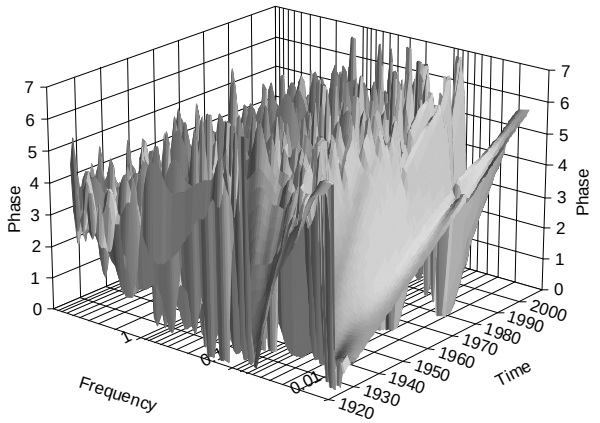


Figure 36: CWT 3D phase spectrum for US industrial production

part of the 19th century, with mostly low frequency content evident throughout the entire time period. As might be expected, the ridge lines in the 3D magnitude plot for the series suggest that there are roughly 5 distinct frequencies active in the series, as was already confirmed with the use of a DWT. Figure 36.now plots the phase for the US industrial production series - here the phasing pattern is very unclear..

It is probably a better strategy to project both 3D plots onto the magnitude or phase axis, respectively, so as to view the plots in terms of their contours. This is done in figure 37 first for the scale-averaged (or integrated) power spectrum, $|W_k(s)|^2$, which, as expected, shows that the volatile short-term fluctuations in the early part of the series tend to dominate and then in figure 38 for the phase contour plot. With reference to figure 42, figures 39 gives CWT a magnitude contour plot (against log of frequency) using a Paul (complex) wavelet with initial width of 6 months. This is an asymmetric wavelet, and so the frequency is only roughly related to the scale of the wavelets, hence the approximation that s^{-1} is equivalent to the maximum of $\widehat{\Omega}$ is not analytically correct[13]. The contour plots point to several other features from the frequency domain, notably:

i) three areas of notable high frequency activity, two of which coincide with the rebounds in production in the early 1920s and mid 1930s;

[13]The relationship between scale and frequency depends on the form of the wavelet. For more on this see the discussion in Torrence and Compo (1998) (p67).

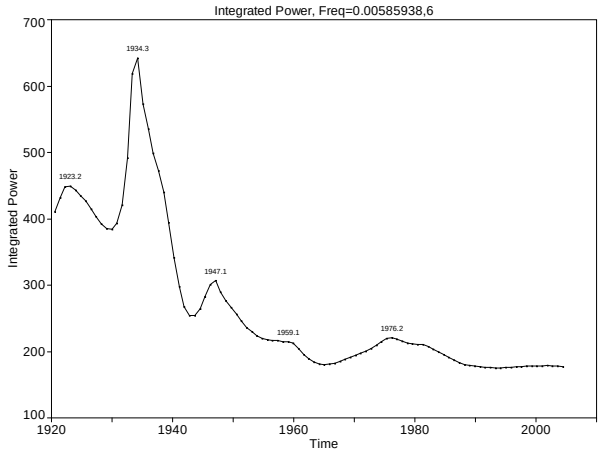


Figure 37: CWT scale-averaged wavelet power spectrum for US industrial production

ii) the constant presence of a relatively low frequency pattern throughout the period of the data;

iii) the disappearance of almost all high frequency activity since 1945 and the dying out of a long medium frequency pattern over the period 1945-1980.

iv) a range of frequencies where phases interact quite frequently (the "fingers" in the phase contour plot), sometimes at a large range of frequencies (the longer fingers) and sometimes over a smaller range (the smaller thinner fingers).

Torrence and Compo (1998) rightly observes that one of the major drawbacks of spectral analysis has been that significance testing was difficult to implement in the frequency domain. But by assuming a "background" spectra such as white noise, or an AR(1) process, a null hypothesis is formed by subtracting a Monte Carlo simulated background spectrum from the time series wavelet spectrum to yield a "differenced" spectrum. By assuming a gaussian distribution confidence intervals can then be constructed as any point on the differenced spectrum should be distributed as $0.5P_k\chi_2^2$ where P_k is the Fourier power spectrum. Here two different background spectra are assumed for the US industrial production series. First, a random walk is assumed for the series, and then an AR(1) series, which then allows for confidence intervals to be constructed for differences between the two different spectra.

The largest (red) shaded areas for figure 40 shows the parts of the US industrial produc-

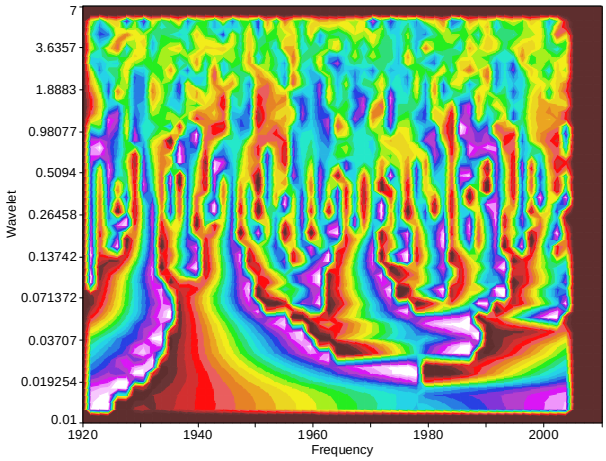


Figure 38: CWT phase spectrum contour plot for US industrial production

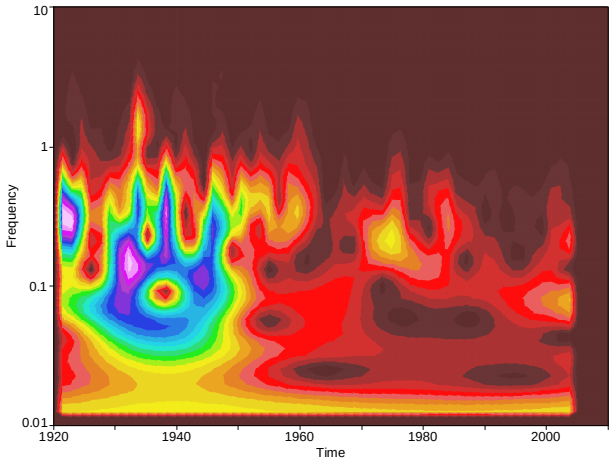


Figure 39: CWT magnitude contour plot using US industrial production

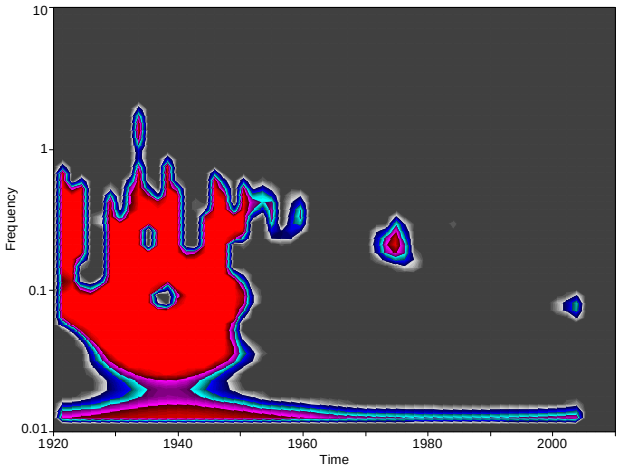


Figure 40: CWT magnitude contour plot assuming white noise background spectrum

tion spectrum that are outside the 99% confidence interval[14] for a white noise spectrum. Similarly, 41 shows the parts of the spectrum that are outside the 99% confidence interval for a AR(1) process with coefficient 0.95[15].

Clearly, as this series could easily be characterised as an AR(1) process, only in the earlier parts of the period under study does the spectrum show a significant departure from an AR(1) process, and this is only for the higher frequencies. When testing against white noise, however, the departures are significant in the higher frequencies during the early part of the series, but continue to be significant in the lower frequency bands throughout the series.

Torrence and Compo (1998) also go on to use a similar approach to analyze co-spectrum for use in multivariate time series analysis.

9 Economics and wavelet analysis

As with many applied statistical techniques, wavelet analysis undoubtedly has unfulfilled potential in economics. Several pioneering economic researchers should be recognized as having seen this potential[16], and then devoting considerable amounts of time and effort to understanding and using wavelets in their work. As many of the advances in this area

[14] The software used here actually uses critical limits, such that a 95% critical limit means that in only 1 in 20 similar size random data sets would the largest CWT spectral peak attain this height.

[15] The estimated AR(1) coefficient attained for this series was 0.96.

[16] James Ramsay of New York University is particularly notable in this regard.

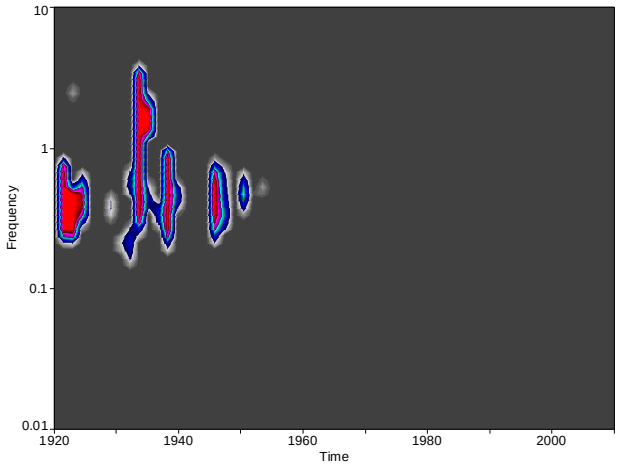


Figure 41: CWT magnitude contour plot assuming AR(1) process with 0.95 coefficient for background spectrum

are not being made in economics, but rather in signal and image processing, engineering, astronomy and meteorology, it is perhaps understandable that the uptake by the economics profession has not been swift. Like spectral analysis, wavelet analysis will never dominate mainstream time-series econometrics, but it should be seen as another useful mode of analysis in the applied economist's toolkit.

In this final section a brief list of practical issues is first presented and then a description of the rapidly expanding literature on wavelets in economics is provided.

9.1 Practical issues

These points are largely taken from Percival and Walden (2000) and Torrence and Compo (1998) where, in the latter case the approach is related specifically to atmospheric and oceanic research data.

1. Sufficient data is required given the observed nature of the time series. This implies that in economics, if for example business cycles are to be analysed, quite long time series are necessary if all scales are to be properly resolved. Many economists believe that wavelet analysis requires high frequency data, as the technique has been applied much more frequently in finance.- this is not the case, as table 1 details.

2. Choice of wavelet function - issues arise such as whether to use orthogonal or non-orthogonal, complex or real, and the choice of width and type of wavelet function

is also an issue. For analysis of economic time series, clearly the Haar wavelet is not really appropriate, given the discontinuous nature of its waveform, but beyond this, the choice is largely one for the researcher, and is likely limited by the software being used. Width of wavelet function is also very data dependent, but as most economic time series are not high frequency, large numbers of oscillations are usually not relevant.

3. Dyadic series - some wavelet applications require dyadic time series. This is obviously not attractive for usage with economic time series as it implies discarding data or padding series with extra data if the series is not dyadic. In the examples used in this paper, the Canadian and US time series were padded to 1024 datapoints, but the Finnish industrial production series was left with 604 data points. Clearly the MODWT and the DTWT show most promise here, as they do not require dyadic series.

4. Boundary or cone of influence effects. With the DWT these are usually labelled boundary effects, but with the CWT they are called cone of influence effects. There are many different types of boundary treatment rules, such as i) a periodic rule - assumption that the series is periodic, that is for any $x_i = x_{i+nk}$, where k is an integer and $i = 1, 2, 3, .., n$; ii) a reflection rule - assumption that x_n is reflected at the boundaries and then periodically extended; iii) a zero padding rule - at each step of the filtering process the series is padded at the beginning and end with zeros; iv) a polynomial extension - padding the beginning and end of the series by repeating the beginning and end series values before and after the series respectively; and v) an interval rule - using special "interval" wavelets at the boundaries, whose filter coefficients are zero outside the range of the data.

Obviously software is also an issue, as many of these procedures in wavelet analysis are not easy to implement unless the researcher is an experienced programmer. Appendix E provides a list of software resources for wavelet analysis.

9.2 The wavelet literature in economics

There are relatively few contributions to economics that use wavelets in the existing literature, although in recent months this list has expanded considerably, which signals that wavelet analysis is at last being recognised as a legitimate methodological approach in economics. The contributions located to date are described in chronological order below:

a. Ramsey and Zhang (1995) analyses high-frequency foreign exchange rates using waveform dictionaries and a matching pursuit algorithm and determines that there is structure only at the lowest frequencies. At the highest frequencies "chirp" like bursts of energy are distributed across frequencies throughout the year.

b. Ramsey and Lampart (1998b) takes seasonally unadjusted monthly money supply data and nominal personal income for the US from 1960-1998, and then use Granger causality tests for each of the sets of crystals. They find that the relationship between money and income varies according to scale, with money supply Granger causing income at higher scale levels, and income Granger causing money supply at low scale levels.

c. Ramsey and Lampart (1998a) does the same thing as in a. except for durable and non-durable expenditures and income.

d. Conway and Frame (2000) analyses different measures of the output gap in New Zealand using time dependent spectra with wavelet analysis.

e. Tkacz (2000) uses the Jensen wavelet estimator Jensen (2000) to estimate the fractional order of integration for Canadian and US interest rates. He finds that US nominal interest rates likely follow long-memory processes, and that Canadian rates also exhibit strong persistence, giving even a larger order of integration than for the US.

f. Renaud, Starck, and Murtagh (2003) propose a new forecasting method based on wavelets whereby MRD crystals are used with a simple AR process.

g. Atkins and Sun (2003) uses Jensen's estimator of the long-memory parameter coupled with Fisher regressions in the wavelet domain to show that in the short run, there is no relationship between interest rates and inflation for shorter time scales but is statistically significant at longer time scales.

h. Kim and In (2003) investigate whether US financial variables have predictive power over US industrial production data over various frequency domains and time scales using spectral, wavelet analysis and Granger causality tests. They find that the relationship between US financial and real variables runs from financial variables to real at short time scales and from real to financial at longer time scales.

i. Capobianco (2004) uses matching pursuit algorithms with waveform dictionaries are used to scale decompose intra-day stock return dynamics. Wavelet analysis helps to identify hidden intraday periodicities both at the 1 minute and 5 minute timescale.

j. Crivellini, Gallegati, Gallegati, and Palestrini (2004) scale decompose G6 (G7 minus Canada) industrial production data and then look at the characteristics of the different scale cycles and do a rolling correlation analysis of these scale components between countries[17]. Their results indicated an increase in correlation of business cycles at all scales in the 1970s, a decrease in the 1980s at the medium and short run scales, and in the 1990s relatively high long term correlations. The EU countries also tended to have high medium term correlations in the 1990s, with the UK an exception to this rule.

k. Dalkir (2004) scale decomposes personal income, sum and divisia monetary aggregates for the US and uses Granger causality tests to review the Ramsey and Lampart (1998b) results. The findings are that in the majority of cases money Granger causes personal income, although this relationship does reverse, usually during a shift in monetary policy regime.

l. Fernandez (2004) looks at returns spillover in stockmarkets at different time scales using wavelet analysis. She finds evidence of price spillovers from North America to Latin America, emerging Asian and Far East markets, and Pacific markets, and also evidence of spillovers from European and Latin America to North American markets. In the latter part of the paper she also controls for conditional heteroskadasticity and serial correlation by using an asymmetric power GARCH model.

m. Lee (2004) uses wavelets to analyse the relationship between the South Korean and US stockmarkets. Using MRA at different scales, it is determined that there is strong evidence of price and volatility spillovers from developed country to developing country stockmarkets.

n. Neumann and Greiber (2004) use scale decompositions of M3 money growth and eurozone inflation to compare MRDs with other filters and also to look at the relationship between M3 and inflation for the eurozone on a scale by scale basis. They find that short- to medium-term fluctuations of money growth with cycles of up to about 8 years were not signficant causes of affecting trend inflation.

o. Vuorenmaa (2004) analyses Nokia share volatility using wavelet MRA analysis.and finds that wavelet variance and covariance analysis reveals a considerable amount about

[17] One of the problems with the analysis here, is that the monthly industrial production data is only analysed to the 6th scale level, which implies that longer business cycles are not properly resolved.

stock market activity at intra-day levels. He then applies a local scaling law and long memory in volatility, and finds that time-varying long-memory is supported over a medium term period (months).

p. Gençay and Whitcher (2005) use wavelets with high-frequency financial data and a hidden Markov tree model to establish a "new" stylized fact about volatility - that low volatility at a long time horizon is most likely followed by low volatility at shorter time horizons - the reverse for high volatility doesn't seem to be the case. The authors label this phenomenon *asymmetric vertical dependence*.

10 Conclusions

Perhaps because economics does not usually deal with long data series with "natural" periodicity, traditional spectral analysis never had a large amount to offer the economics discipline. Time series analysis is probably much better suited to analysis of shorter time series with variable "non-natural" periodicity, coupled with its emphasis on causation and underlying processes. Wavelet analysis, however, differs from spectral analysis in that it straddles both the time and frequency domains, thereby allowing identification of both time period and scale. In this sense, it likely has a lot more to offer economists.

Wavelet analysis has the potential to offer much to empirical economic research. The potential is particularly apparent in two areas of economics: that of business cycle analysis, which naturally lends itself to analysis of periodicities, and where filtering has been a particularly controversial issue; and in any part of macro or monetary economics where theoretical long-run and short-run relationships can be distinguished. It's ability to separate out the dynamics in a time series over a variety of different time horizons, the number of which being directly related to the type of time series and the number of observations available, can reveal interesting insights into cycles at different time scales. It's ability to work with non-stationary data is particularly advantageous, as most econometric methodology assumes stationarity, which may or may not be apparent in economic data (either locally or globally).

In particular, wavelet analysis now offers economic researchers the availability of the MODWT, so that time series can be decomposed into identical numbers of scale-decomposed equivalent series, making statistical testing and time series analysis possible for individual scale crystals. The real challenge will be to integrate the wavelet approach with the traditional time-series toolkit of the econometrician, as perhaps here the potential con-

tribution is greatest, and here there is the greatest chance of impressing upon the wavelet research.community that this methodology has something to offer outside of its most widely-applied research areas.

Appendices

A Industrial production indices

In the paper, three industrial production indices are used. These are as follows:

- Canadian industrial production - this is available in a monthly format from January 1919, although it is necessary to construct a full data series from three different Statistics Canada (CANSIM) series, and then splice together. The series was seasonally adjusted by the Bank of Finland using the Stamp programme. The resulting series gave 1014 percentage year-over-year observations, just shy of a dyadic series of 1024 points. The series was therefore padded by using the August 2004 value for the percent change of the index over the remaining 10 months.

- US industrial production - this is available from the Bureau of Economic Analysis (BEA) of the US Department of Commerce. The data is available on the web, and was again seasonally adjusted by the Bank of Finland. Again, only 1015 monthly data points were available obtained when transformed to a percentage year-over-year series, so these were padded using the last observation for September 2004.

- Finnish industrial production - the source was Statistics Finland, seasonally unadjusted. The series was seasonally adjusted by the Bank of Finland using the Stamp programme. As this was a non-dyadic series, it is only used in the later sections of the paper..

The seasonally adjusted data are shown in figure 42 with right hand side plots representing the industrial production index and the right hand side plots representing the year-on-year change in the three indices.

The Finnish industrial production series starts in 1954 and so has a shorter time-span than the other two series Figure 43 shows a spectral decomposition for the Finnish series.

Note that the raw periodogram appears to be quite similar to the periodograms for the Canadian and US series, which suggests that monthly industrial production series tend to embed 5 frequencies.

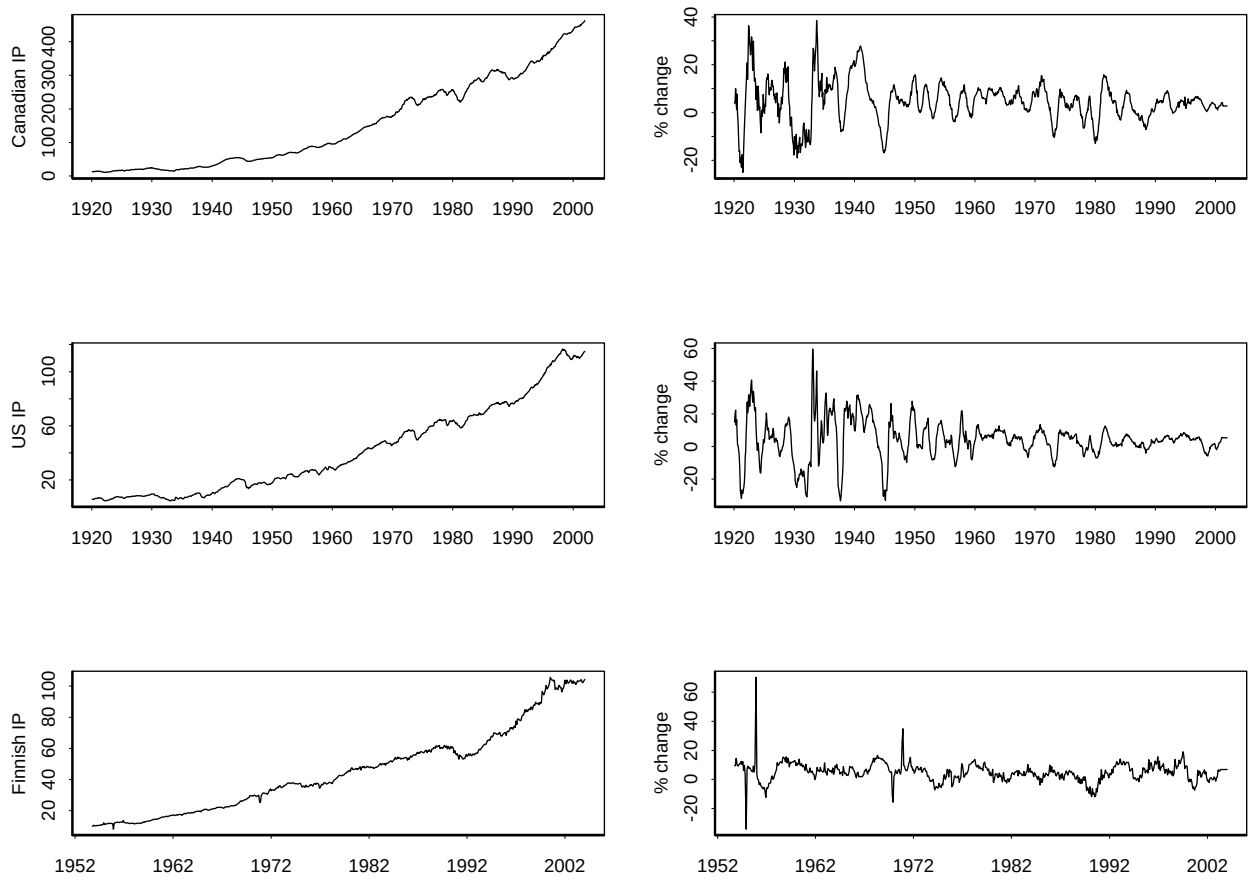


Figure 42: Canadian, US and Finnish Industrial Production Series

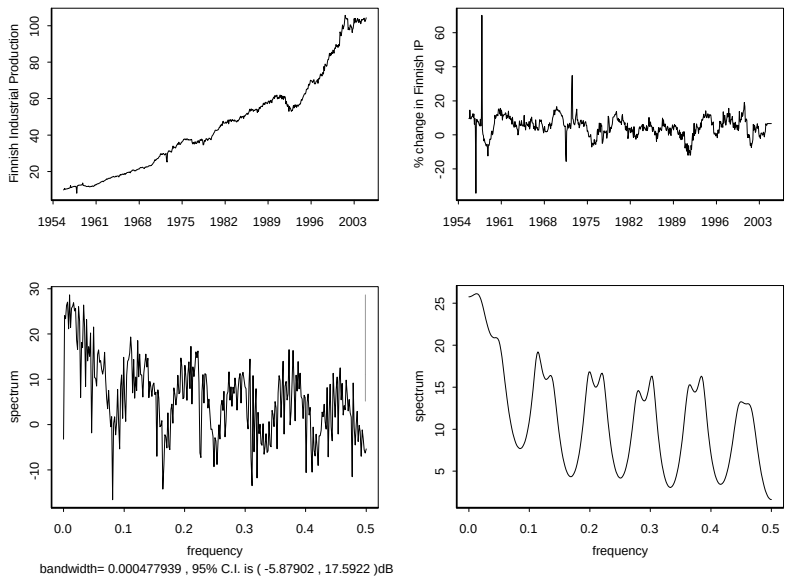


Figure 43: Finnish industrial production: a) series; b) annual % change; c)raw periodogram; d) autocorrelation spectra.

B Wavelets as filters

For most purposes in practical applications wavelets only need to be defined over a specific span of points in time. In place of actual wavelets sequences of values representing the wavelet are used - these are commonly called wavelet filters. The number of values in the series is called the length of the wavelet filter and the number of values is known as the "tap" of the filter. Given a wavelet filter $\{h_l\}$ where $l = 0, , , (L-1)$, the set of wavelet coefficients h_l have to satisfy:

$$\begin{array}{c} \sum_{l=0}^{L-1} h_l = 0 \\ \sum_{l=0}^{L-1} h_l^2 = 1 \\ \sum_{l=0}^{L-1} h_l h_{l+2k} = 0 \end{array} \tag{51}$$

- in other words, the coefficients must sum to zero have unit energy and be orthogonal to even shifts in the filter. The last two conditions in equation are called the orthonormality conditions. This defines the father wavelet in filter terms. To obtain a mother wavelet filter, which is known in the filter literature as a "scaling wavelet", the wavelet filter is converted using a quadrature mirror filter so that the coefficients of the mother wavelet filter are given by:

$$g_l = (-1)^{l+1} h_{L-1-l} \tag{52}$$

The mother wavelet filter also satisfies the orthonormality condition but does not have unit energy:

$$\begin{array}{c} \sum_{l=0}^{L-1} g_l = \sqrt{2} \\ \sum_{l=0}^{L-1} g_l^2 = 1 \\ \sum_{l=0}^{L-1} g_l g_{l+2k} = 0 \end{array} \tag{53}$$

The shape of the wavelet filter is known in this literature as the "impulse response". There are several different types of wavelet filters, which all correspond to a continuous wavelet counterparts discussed in the main part of the text, but are referred to using filter terminology. Examples of these are as follows:

1. Extremal phase filters - here energy increases rapidly - these are asymmetric filters, which are also known as minimum phase filters or "daublets", as they roughly correspond to Daubechies wavelets.

2. Least asymmetric filters - filters in this group are smoother and are nearly symmetric - the magnitudes of their discrete fourier transforms are the same, and they correspond to Symmlet wavelets.

3. Best localized filters - these filters are also nearly symmetric and they also have the same gain functions.

4. Coiflet filters - these filters approximate linear phase filters, and have different gain functions from the other three types above - they clearly correspond to a Coiflet wavelet.

A DWT simply corresponds to repeated application of a chosen filter, which is often referred to as a "filter bank". Obviously for any scale level J, an amplitude reduced and dilated filter can be derived (using the pyramid algorithm) for the original data to be convolved with, giving one filter for each level of the transform. In this sense the filter bank acts as a bandpass filter, and the frequency response of any combined filter can be obtained. For example, figure 44 shows the frequency response for a Coiflet 6 length filter at scale levels of $J = 1, 2, ..5$.

For the DTWT, there are four types of scale filters and four types of wavelet filters to select from. In figure nearsyma and Qshiftc filters are shown. The first row of figure

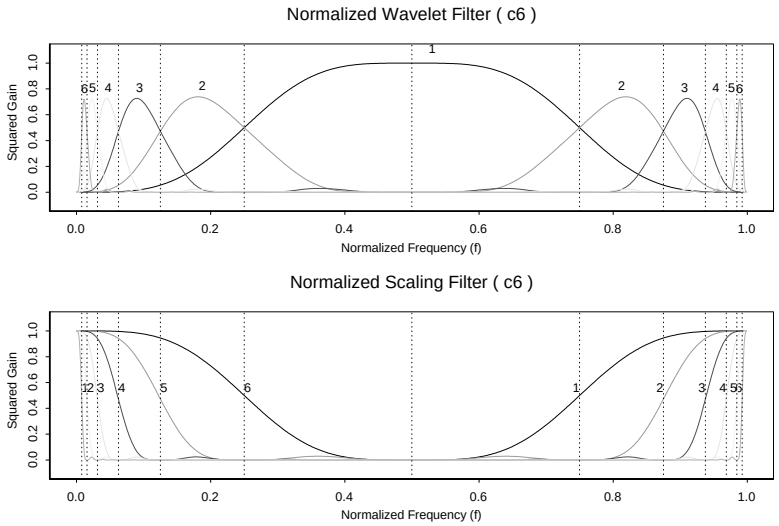


Figure 44: Squared gain function for Coiflet c6 filter

shows the filters that are applied to all of the data, then the second row shows the filters that are applied to one tree and the third row shows the filters that are applied to the second tree. Clearly the frequency response for the Qshift scaling filter will be oriented towards the higher frequency parts of the series and the nearsyma wavelet filters will be oriented towards the lower frequency fluctuations of the series.

C MODWT

Here we use matrix notation. Let $\mathbf{x}$ be a vector of $\mathbf{N}$ observations. The vector of MODWT coefficients is given by:

$$\widetilde{\mathbf{w}} = \widetilde{\mathbf{W}}\mathbf{x} \tag{54}$$

where $\widetilde{\mathbf{W}}$ is the (J+1)NxN matrix defining the MODWT, and so $\widetilde{\mathbf{w}}$ is a (J+1)N vector of wavelet coefficients which result from the transform. Similarly to a DWT, the MODWT matrix $\widetilde{\mathbf{W}}$ can be rewritten as:

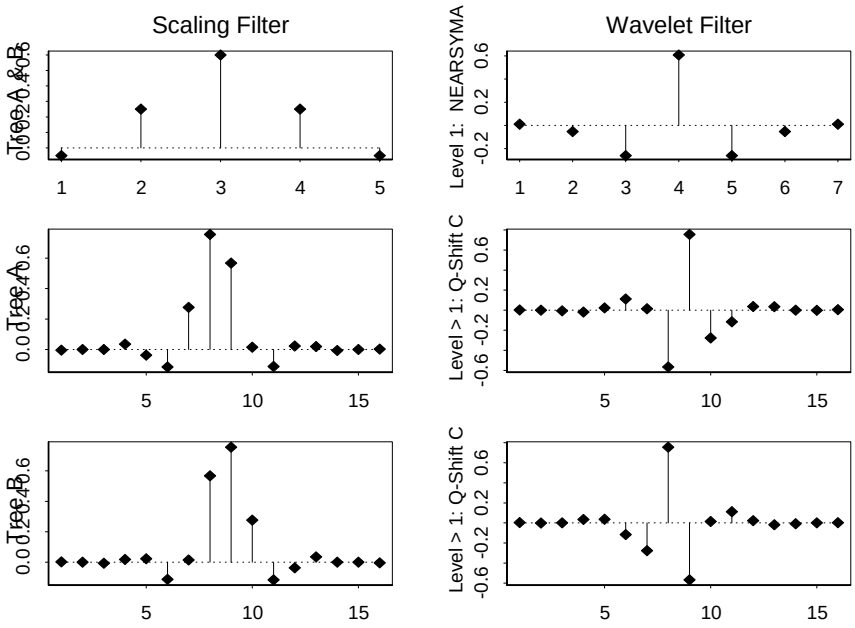


Figure 45: DTWT filters. Top row - level 1 filters; middle row - tree A filters, bottom row - tree B filters

$$\widetilde{\mathbf{W}} = \begin{bmatrix} \widetilde{\mathbf{W}}_1 \\ \widetilde{\mathbf{W}}_2 \\ ... \\ \widetilde{\mathbf{W}}_J \\ \widetilde{\mathbf{V}}_J \end{bmatrix} \tag{55}$$

where $\widetilde{\mathbf{W}}_1$ is an NxN matrix representing the filter components at each scale level. But also each scale level in the matrix can be rewritten as:

$$\widetilde{\mathbf{W}}_1 = \begin{bmatrix} \widetilde{\mathbf{h}}_1^1 & \widetilde{\mathbf{h}}_1^2 & \widetilde{\mathbf{h}}_1^3 & ... & \widetilde{\mathbf{h}}_1^{N-2} & \widetilde{\mathbf{h}}_1^{N-1} & \widetilde{\mathbf{h}}_1 \end{bmatrix}^T \tag{56}$$

for scale 1, where the vector of filter coefficients $\widetilde{\mathbf{h}}_1^k$ represent the rescaled jth scale filter coefficients:

$$\widetilde{\mathbf{h}}_j = \frac{\mathbf{h}_j}{2^j} \tag{57}$$

where $\mathbf{h}_j$ is the vector of DWT filter coefficients shifted k intergers to the right.

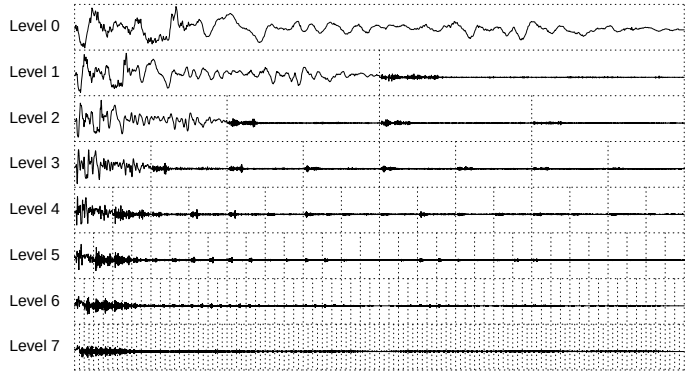


Figure 46: MODWPT for Canadian industrial production

In other words, the MODWT essentially takes a DWT and shifts the the wavelet filters one to the right to interweave with any given level of the equivalent DWT submatrix $\widetilde{\mathbf{W}}$. In practice a pyramid algorithm is used, but without the downsampling inherent in a DWT.

Figure 46 shows the MODWT packet table for Canadian industrial production:

D British pound exchange rate

This daily series was taken from the Bank of England's exchange rate database. The series runs from 1975 to 26 October 2004, and contains 7540 datapoints. The original series and the absolute value of the log differenced series are plotted below in figure 47

E Software sources

Software for wavelets is now quite widely available. Both *SPLUS* and *Matlab* have modules for extensive wavelet analysis that can be separately purchased from the publishers themselves, and a separate wavelet module is available as freeware for use with the open source *R* statistical software. Apart from this, software is available for specialist wavelet applications, for example *Autosignal* implements the CWTs spectral analysis described in this paper, and freeware is also available to implement this with *Matlab.* An additive

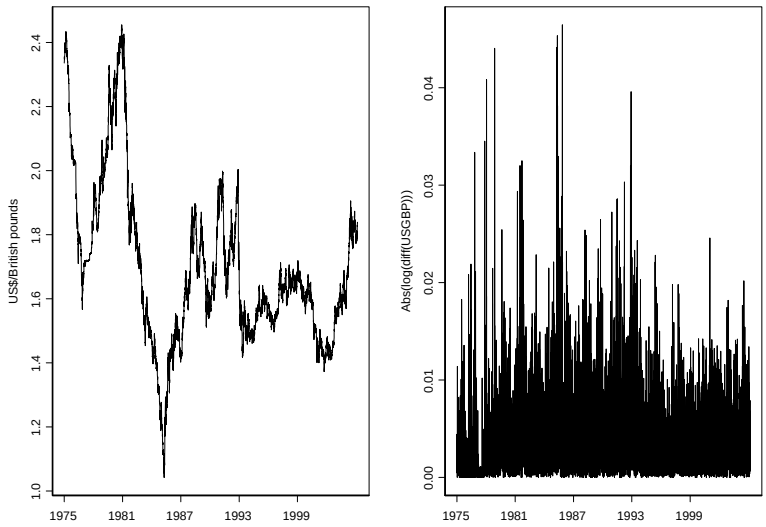


Figure 47: USD/GBP exchange rate: a) daily exchange rate; b) abs(diff(log(USD/GBP)))

package called *TSM* is available for *GAUSS* as well, although this appears to only do a limited degree of wavelet analysis.

References

Addison, P. (2002). *The Illustrated Wavelet Transform Handbook*. Bristol, UK: Institute of Physics.

Atkins, F. and Z. Sun (2003). Using wavelets to uncover the fisher effect. Technical Report DP 2003-09, Department of Economics, University of Calgary, Canada.

Bruce, A. and H.-Y. Gao (1995). Waveshrink: Shrinkage functions and thresholds. In A. Laine and M. Unser (Eds.), *Wavelet Applications in Signal and Image Processing III*, Volume 2569, pp. 270–283. SPIE.

Bruce, A. and H.-Y. Gao (1996). *Applied Wavelet Analysis with S-PLUS*. New York, NY, USA: Springer-Verlag.

Bruce, A., G. Hong-Ye, and D. Ragozin (1995). S+WAVELETS: Algorithms and technical details. Technical report, StatSci Division of MathSoft Inc.

Camba Mendez, G. and G. Kapetanios (2001). Spectral based methods to identify common trends and common cycles. Technical Report Working paper 62, European Central Bank, Frankfurt, Germany.

Capobianco, E. (2004). Multiscale analysis of stock index return volatility. *Computational Economics 23*(3), 219–237.

Chiann, C. and P. Morettin (1998). A wavelet analysis for time series. *Nonparametric Statistics 10*, 1–46.

Cohen, A., I. Daubechies, and J. Feauveau (1992). Biorthogonal bases of compactly supported wavelets. *Communications in Pure Applied Mathematics 45*, 485–560.

Coifman, R., Y. Meyer, S. Quake, and V. Wickerhauser (1990). Signal processing and compression with wavelet packets. Technical report, Yale University.

Coifman, R. and V. Wickerhauser (1992). Entropy-based algorithms for best basis selection. *IEEE Transactions on Information Theory 38*(2), 713–718.

Collard, F. (1999). Spectral and persistence properties of cyclical growth. *Journal of Economic Dynamics and Control 23*, 463–488.

Constantine, W. and D. Percival (2003, October). S+Wavelets 2.0.

Conway, P. and D. Frame (2000, June). A spectral analysis of new zealand output gaps using fourier and wavelet techniques. DP2000/06.

Craigmile, P. and B. Whitcher (2004). Multivariate spectral analysis using hilbert wavelet pairs. *International Journal of Wavelets, Multiresolution and Information Processing 2*(4), 567–587.

Crivellini, M., M. Gallegati, M. Gallegati, and A. Palestrini (2004). Industrial output fluctuations in developed countries: A time-scale decomposition analysis. Technical report, European Commission, Brussels, Belgium. Working Papers and Studies: Papers from the 4th Eurostat and DGFin Colloquium "Modern Tools for Business Cycle Analysis".

Dalkir, M. (2004). A new approach to causality in the frequency domain. *Economics Bulletin 3*(4), 1–14.

Debauchies, I. (1992). *Ten Lectures on Wavelets*. Montpelier, VT, USA: Capital City Press.

Donoho, D. and I. Johnstone (1995). Adapting to unknown smoothness by wavelet shrinkage. *Journal of the American Statistical Association 90*, 1200–1224.

Fernandez, V. (2004). Time-scale decomposition of price transmission in international markets. Technical Report Working paper 189, Center for Applied Economics, University of Chile, Santiago, Chile.

Gençay, R., F. Selçuk, and B. Whicher (2001). *An Introduction to Wavelets and Other Filtering Methods in Finance and Economics*. San Diego, CA, USA: Academic Press.

Gençay, R., S. F. and B. Whitcher (2005). Multiscale systematic risk. *Journal of International Money and Finance*, to appear.

Granger, C. and R. Joyeux (1980). An introduction to long-memory time series models and fractional differencing. *Journal of Time Series Analysis 1*, 15–29.

Greenhall, C. (1991). Recipes for degrees of freedom of frequency stability estimators. *IEEE Transactions on Instrumentation and Measurement 40*, 994–9.

Hughes Hallett, A. and C. Richter (2004). Spectral analysis as a tool for financial policy: An analysis of the short-end of the british term structure. *Computational Economics 23*, 271–288.

Inclan, C. and G. Tiao (1994). Use of cumulative sums of squares for retrospective detection of changes of variance. *Journal of the American Statistical Association 89*, 913–923.

Jensen, M. (1999). Using wavelets to obtain and consistent ordinary least squares estimator of the long-memory parameter. *Journal of Forecasting 18*, 17–32.

Jensen, M. (2000). An alternative maximum likelihood estimator of long-memory processes using compactly supported wavelets. *Journal of Economic Dynamics and Control 24*, 361–387.

Kim, S. and F. In (2003). The relationship between financial variables and real economic activity: Evidence from spectral and wavelet analyses. *Studies in Nonlinear Dynamics and Econometrics 7*(4).

Kingsbury, N. (1998, September). The dual-tree complex wavelet transform: A new efficient tool for image restoration and enhancement. In *Proceedings EUSIPCO 98*, Rhodes, Greece.

Kingsbury, N. (2000, September). A dual-tree complex wavelet transform with improved orthogonality and symmetry properties. In *Proceedings of the IEEE Conference on Image Processing*, Vancouver, Canada.

Lau, K.-M. and H.-Y. Weng (1995). Climate singal detection using wavelet transform: How to make a time series sing. *Bulletin of the American Meteorological Society 76*, 2391–2402.

Lee, H. (2004). International transmission of stock market movements: A wavelet analysis. *Applied Economics Letters 11*, 197–201.

Mallat, S. (1989). A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence 11*(7), 674–693.

Mallat, S. and Z. Zhang (1993). Matching pursuits with time frequency dictionaries. *IEEE Transactions on Signal Processing 41*(12).

Neumann, M. and C. Greiber (2004). Inflation and core money growth in the euro area. Technical Report Discussion paper 36/2004, Deutsche Bundesbank, Frankfurt, Germany.

Percival, D. and H. Mofjeld (1997). Analysis of subtidal coastal sea level flusctuations using wavelets. *Journal of the American Statistical Association 92*, 868–80.

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

Ramsey, J. (2000). The contribution of wavelets to the analysis of economic and financial data. Volume Wavelets: the key to intermittent information, pp. 221–236. OUP.

Ramsey, J. (2002). Wavelets in economics and finance: Past and future. *Studies in Nonlinear Dynamics Econometrics 6*(3), 1–27.

Ramsey, J. and C. Lampart (1998a). The decomposition of economic relationships by time scale using wavelets: Expenditure and income. *Studies in Nonlinear Dynamics and Econometrics 3*(1), 23–41.

Ramsey, J. and C. Lampart (1998b). Decomposition of economic relationships by timescale using wavelets. *Macroeconomic Dynamics 2*, 49–71.

Ramsey, J. and Z. Zhang (1995, January). The analysis of foreign exchange data using waveform dictionaries. Technical Report 95-03, C.V. Starr Center for Applied Economics, Department of Economics, New York University, NYC, USA.

Renaud, O., J.-L. Starck, and F. Murtagh (2003). Prediction based on multiscale decomposition. *International Journal of Wavelets, Multiresolution and Information Processing 1*(2), 217–232.

Schleicher, C. (2002). An introduction to wavelets for economists. Technical Report Working Paper 2002-3, Bank of Canada.

Süssmuth, B. (2002, January). National and supranational business cycles (1960-2000): A multivariate description of central g7 and euro15 NIPA aggregates. Technical Report WP 658(5), CESifo.

Tkacz, G. (2000). Estimating the fractional order of integration of interest rates using wavelet OLS estimator. Technical report, Bank of Canada. Working Paper 2000-5.

Torrence, C. and G. Compo (1998). A practical guide to wavelet analysis. *Bulletin of the American Meteorological Society 79*(1), 61–78.

Valle e Azevedo, J. (2002, April). Business cycles: Cyclical comovement within the european union in the period 1960-1999. a frequency domain approach. Technical Report WP 5-02, Banco do Portugal.

Vuorenmaa, T. (2004, July). A multiresolution analysis of stock market volatility using wavelet methodology. Draft Bank of Finland discussion paper.

Walker, J. (1999). *A Primer on Wavelets and their Scientific Applications*. Boca Raton, FL, USA: CRC Press.

Walnut, D. (2002). *An Introduction to Wavelet Analysis*. Boston, MA, USA: Birkhäuser Boston.

Whitcher, B., S. Byers, P. Guttorp, and D. Percival (1998). Testing for homogeneity of variance in time series: Long memory, wavelets and the nile river. Technical Report Technical Report 9, National Research Center for Statistics and the Environment.

Whitcher, B., P. Guttorp, and D. Percival (1999). Mathematical background for wavelet estimators of cross-covariance and cross-correlation. Technical Report TSE No. 038, NRCSE.

Whitcher, B., P. Guttorp, and D. Percival (2000a). Multiscale detection and location of multiple variance changes in the presence of long memory. *Journal of Statistical Computation and Simulation 68*, 65–88.

Whitcher, B., P. Guttorp, and D. Percival (2000b). Wavelet analysis of covariance with application to atmospheric time series. *Journal of Geophysical Research 105*(D11), 14,941–14,962.