Wavelets in Economics and Finance: Past and Future

James B. Ramsey[*]

[*]New York University,

Wavelets in Economics and Finance: Past and Future

James B. Ramsey

Abstract

In this paper I review what insights we have gained about economic and financial relationships from the use of wavelets and speculate on what further insights we may gain in the future. Wavelets are treated as a 'lens' that enables the researcher to explore relationships that previously were unobservable.

1.   Introduction

The development of any new statistical procedure potentially yields one or more of four types of gain on the existing set of statistical tools. A new procedure could provide estimators for novel situations, improve the efficiency of estimation or reduce bias, enhance the robustness to modeling errors, or provide new insights into the properties of the cognate discipline. While each of these goals is worthy in its own right and provides useful gains to the profession, the last goal is for me the most important potential gain of all. While not denying the benefits of enhanced estimation efficiency and bias reduction, or the improvement in robustness, the ability to apply a new "lens" to inspect the relationships in economics or finance provides great promise for the development of the discipline. This is particularly applicable in economics and finance where the potential for experimentation is limited and the measured variables are complex aggregations of disparate components.

Wavelets provide a unique decomposition of time series observations that enable one to deconstruct the data in ways that are potentially revealing. The situation is similar to, but distinct from, the insights gained from analyzing data using Fourier series. In so far as Fourier series are applicable to data sets under examination, the application of Fourier analysis yields many interesting insights into dynamic relationships. Even the failure of Fourier analysis in the context of many data sets provides insights into the underlying dynamical relationships., For example in the GDP indices of production, we can discover that there are no simple Fourier components except in the context of non-durable goods. Using advanced techniques, Fourier analysis can be used to clarify and confirm the nature of the non-stationarity that has long been suspected. See for example, Ramsey and Thomson (1999), who demonstrate non-stationarity for a variety of macro variables. They do so by using a more efficient estimation procedure than the standard FFT. The procedure capitalizes on the properties of oblate spheroid wave functions that enable one to estimate efficiently the time rate of change in the spectrum.

The balance of this paper is in three sections. The first reviews schematically the structure of wavelets and atomistic decomposition using waveform dictionaries. The second and largest section discusses the insights obtained from applying wavelets and waveform dictionaries to economic and financial data. This section will examine the crucial distinction between noise smoothing and de-noising, the critical role of time scale in economics, the analysis of non-stationary and complex functions, the discovery of time delays in economic relationships that are functions of the state space, the discovery of Fourier frequencies that wax and wane over time, and lastly, insight is gained on the long standing complaint that economists can fit data well, but forecasts are routinely poor. The last section provides a brief

summary and speculates about some areas in which wavelet analysis may generate further insights into the analysis of economic and financial data.

An earlier review of the contribution of wavelets to the analysis of economic and financial data is in Ramsey (1999) that contains a different perspective and further details.

2.   Schematic Review of Wavelets and Waveform Dictionaries

In this paper I will only give a brief schematic review of wavelets and atomistic decomposition by waveform dictionaries. There are many excellent expositions in the literature, for example, Chui (1992), Gencay et al. (2002), Carmona et al (1998), Hardle et al. (1998), Percival and Walden (2000), and Strang and Nguyen (1996) to which the interested reader is recommended. In particular, Gencay et al. (2002) is useful for economists as it contains a number of economic and financial applications. Both Percival and Walden (2000) and Strang and Nguyen (1996) are very well written books that stress the filter development of wavelets. Brillinger (1994), Brillinger (1996) and Ramsey (1999a) develop some of the distribution theory for wavelet analysis. Mallat and Zhang have developed the use of projection pursuit methods in the atomistic decomposition of waveform dictionaries, see Mallat (1989) and Mallat and Zhang (1993).

There are some points of comparison between wavelets and Fourier series, but there are also important points of difference; the latter are more important to keep in mind than the former. The points of comparison are that both wavelets and Fourier representations are exactly that, representations, and both are obtained by projecting the signal onto a basis space.

The differences stem from the differences in the bases used by the two procedures. In the Fourier integral representation, one is projecting on to an expansion in terms of trigonometric functions assuming stationarity over the entire history of the signal. Fourier series expansions are defined over the space $L^2(0, 2\pi)$, with infinite energy, but finite power, when extended to the whole real line. The decomposition that is achieved using Fourier analysis is in the frequency domain. Further, recall that a single disturbance in the time domain affects all frequencies and that a single disturbance is treated by Fourier analysis as an event of period T, where T is the length of the observed series.

In Fourier analysis all frequencies are obtained by "rescaling" the fundamental frequency; that is, we project the signal onto a sequence of functions of the form:

$$\{e^{-in\omega_0}\} \tag{1}$$

where ω_0 is the fundamental frequency and "n" provides the scaling.

In contrast wavelets are functions defined over Besov spaces and provide a

basis for functions that are defined in such spaces. In particular, each wavelet is compact and therefore must be indexed in the time domain. More precisely, each basis function is expressible as:

$$g(t) = \frac{1}{\sqrt{s}} g\left(\frac{t-k}{s}\right) \tag{2}$$

where k is the time domain index and s is the scale at which g(.) is evaluated. Consequently, we have defined in the time domain a sequence of functions that are doubly indexed, once by location in the time domain, once by the scale; the preceding division by $\sqrt{s}$ in equation 2 ensures that the norm of g(.) is one. Each function g(.) is centered at k with a scale, or dilation, of s. Scaling in Fourier analysis is in terms of frequency, scaling in wavelet analysis is in terms of time. This distinction is important for applications.

Because of the compactness of the wavelets and their time indexing, all projections of a signal onto the wavelet space are essentially local. In contrast, in Fourier analysis the projections are essentially global, although localization can be achieved by convolving the observed series with a filter that is centered at a given point, t_0, with rapid decrease in the modulus of the weights on either side of t_0.

Wavelets can either be defined in terms of a sequence of pairs of filters, to be discussed below, or in terms of functions created through splines that satisfy certain properties. The choice of wavelet to be used in the analysis of actual data will depend on the weights that the researcher places on the various criteria. This is a great strength of wavelets as a tool because one can choose that class of wavelet function that is most suitable to the properties of the function to be represented.

Symmetry of the function g(.) is one such criterion and is the one that is seldom satisfied, except approximately; the Haar wavelet is an example of an orthogonal symmetric wavelet. Symmetry of the wavelet is useful for representing functions that exhibit local symmetries, but the most important benefit of symmetry is that phase shifts are not introduced into the coefficients created by the projection operation. The effect of a phase shift is to displace events along the time axis; that is, if a maximum in the original series occurs at time t_0, the phase shifted maximum occurs at time $t_0 + \delta$, where δ may be positive or negative. Non-symmetric filters introduce phase shifts, see for example, Percival and Walden (2000) and Gencay et al. (2002).

Orthogonality is a very useful property of wavelets as it is of any transformation procedure, but does not hold universally for all wavelet classes. While orthogonality does not in the least ensure zero-phase transformations, recent efforts have been made to modify procedures so as to approximate zero-phase filters, see for example, McCoy et al (1995) and Hess-Nielsen and Wickerhauser (1996). However, even non-orthogonal wavelets still provide a basis for the space into which the function is to be projected. Maximum overlap discrete wavelet transforms

(MODWT) and biorthogonal wavelets are examples, Bruce and Gao (1996), or Percival and Walden (2000).

Smoothness is sometimes an important property for a wavelet basis. If the function to be represented is thought to be smooth, then smoothness is a desirable property. However, there are many functions for which smoothness is definitely not a characteristic in which case, one does not want to impose smoothness. For example, the Haar function is the least smooth of all the wavelet classes and is therefore useful in representing the time path of Poisson processes. The degree of smoothness is measured by the number of continuous derivatives of the basis function.

Below I will define precisely "father" and "mother" wavelets; father wavelets integrate to one and are used to represent the very long scale smooth component of the signal; mother wavelets integrate to zero and represent the deviations from the smooth components. Father wavelets generate what are known as the "scaling coefficients" and the mother wavelets generate the differencing coefficients. An alternative way to view the difference is that the father wavelet acts as a low pass filter, whereas the mother wavelets act as high pass filters. Different scales translate into different frequency bands that are passed; this interpretation is elaborated below.

The number of vanishing moments indicates yet another important characteristic of wavelets. If a wavelet basis is said to "have m vanishing moments," that means that an m'th order polynomial will be passed through by the mother wavelets; the projection integrates to zero and the polynomial component of the signal will be captured solely by the father wavelet. If a signal contains a polynomial component together with more complex elements, using a wavelet with the appropriate number of vanishing moments is clearly very useful in decomposing the signal.

For any suitable choice of function $\Phi(.)$, we can define the corresponding father and mother wavelets:

$$\Phi_{J,k} = 2^{-\frac{J}{2}}\Phi\left(\frac{t - 2^J k}{2^J}\right) \tag{3}$$

$$\int \Phi(t)\,dt = 1$$

and

$$\Psi_{j,k} = 2^{-\frac{j}{2}}\Psi\left(\frac{t - 2^j k}{2^j}\right), j = 1,..J \tag{4}$$

$$\int \Psi(t)\,dt = 0$$

$\Phi_{J,k}$ is the father wavelet and $\Psi_{j,k}$ is the mother wavelet. For facilitating the

mathematical development and analysis, the above statements have restricted the scale parameter "s" to the dyadic scale 2^j. Given this family of basis functions, we can define a sequence of coefficients that represent the projections of the observed function onto the proposed basis. We define:

$$s_{J,\,k} = \int f(t)\Phi_{J,\,k} \tag{5}$$

and

$$d_{j,\,k} = \int f(t)\Psi_{j,\,k}; j = 1,.....J$$

where the $s_{J,k}$ are the coefficients for the father wavelet at the maximal scale, 2^J, known as the "smooth coefficients," and the $d_{j,\,k}$ are the detail coefficients obtained from the mother wavelet at all scales from 1 to J, the maximal scale. Given the coefficients the function f(.) can be represented by:

$$
\begin{aligned}
f(t) \quad &= \quad \sum_k s_{J,\,k}\Phi_{J,\,k}(t) + \sum_k d_{J,\,k}\Psi_{J,\,k}(t) + ... \\
&\quad ... + \sum_k d_{j,\,k}\Psi_{j,\,k}(t)... + \sum_k d_{1,\,k}\Psi_{1,\,k}(t)
\end{aligned}
\tag{6}
$$

or f(t) can be represented as:

$$f(t) = S_J + D_J + D_{J-1} + ...D_j + ...D_1 \tag{7}$$

where

$$
\begin{aligned}
S_J \quad &= \quad \sum_k s_{J,\,k}\Phi_{J,\,k}(t) \tag{8} \\
D_j \quad &= \quad \sum_k d_{j,\,k}\Psi_{j,\,k}(t), \; j = 1,...J
\end{aligned}
$$

The easiest way to visualize the above is to consider a sequence of topographical maps; S_J provides a smooth outline and each D_j in turn provides a higher level of detail. Equation 7 indicates that the complete function will be obtained by the multiresolution of the signal, but one can also obtain less detailed representations by examining only:

$$
\begin{aligned}
S_j \quad &= \quad S_J + D_J + ...D_{j+1} \tag{9} \\
or \; S_j \quad &= \quad S_{j+1} + D_{j+1} \tag{10}
\end{aligned}
$$

An alternative way to think about wavelets is in terms of solutions to sets of

equations defined by low and high pass filters. Thus, we define $\Phi(t)$ and $\Psi(t)$ by:

$$\Phi(t) = \sqrt{2}\sum_{k=0}^{N} l(k)\Phi(2t-k) \tag{11}$$

$$\Psi(t) = \sqrt{2}\sum_{k=0}^{N} h(k)\Phi(2t-k) \tag{12}$$

given two linear filters $l(k)$ and $h(k)$. $l(k)$ is a lowpass filter and $h(k)$ is a high pass filter. Correspondingly, the low and high pass filters can be obtained from the father and mother wavelets as follows:

$$l(k) = \frac{1}{\sqrt{2}}\int \Phi(t)\Phi(2t-k)dt \tag{13}$$

$$h(k) = \frac{1}{\sqrt{2}}\int \Psi(t)\Phi(2t-k)dt \tag{14}$$

$$or\ h(k) = (-1)^{k}l(k) \tag{15}$$

The simplest example is provided by the Haar wavelet for N = 2

$$l(k) = \{\frac{1}{\sqrt{2}},\frac{1}{\sqrt{2}}\} \tag{16}$$

$$h(k) = \{\frac{1}{\sqrt{2}},-\frac{1}{\sqrt{2}}\} \tag{17}$$

The low pass filter averages, the high pass filter differences.

Strang and Nguyen (1996) and Percival and Walden (2000)develop the relationship between wavelets and filter banks, that is, sequences of pairs of high and low pass filters; these are excellent references for this approach, especially the first chapter of Strang and Nguyen (1996). The analysis indicates that one can approach the analysis of the properties of wavelets either through wavelets themselves or through the properties of the filter banks. Both approaches are useful and informative. Many new classes of wavelets are now generated by specifying properties for the filter banks.

The development of wavelet analysis using filter banks helps to clarify the relationship between wavelet analysis and Fourier analysis. Because wavelet transforms can be obtained through a cascade of low and high pass filters one can obtain the transfer function of the filters, and if one assumes stationarity in the time series, one can determine the frequency ranges of the series that will be captured by the filter banks. However, the validity of this interpretation depends on the assumption of stationarity of the signal. Where the stationarity assumption is violated, the frequency interpretation can only be approximate and local.

The introduction of filter banks reveals clearly the difficulty of dealing with boundary conditions that must be faced by any transformation that relies on filters and two sided filters in particular. This problem is of course a general problem and is not one peculiar to wavelets; note for example the difficulties in this regard for Fourier analysis. In the standard literature of wavelet analysis there are several approaches to the problem. Zero padding at the ends is one solution that is also a standard approach in Fourier analysis. Other solutions involve capitalizing on any periodicity in the data, or the use of polynomials to capitalize on any regularities that can be captured by polynomial approximation at either end of the observed series. Yet another procedure that has been used involves reflection; that is, extend the data by a reflection and assume a periodic boundary. Clearly, the differences between these various boundary rules depend upon differences in assumptions as to the fundamental nature of the series under examination.

In economic analysis one seldom wants to use the periodic approach, but often the polynomial approximation is useful. One can perhaps devise new procedures that capture special aspects of the data. However one deals with the issue, there is inevitably an aspect of *ad hoc* justification for the adjustment and one must be very careful how one interprets the coefficients estimated at the ends of the time series. One neglected issue is that the optimal choice of boundary rule may well vary with the scale.

The basics of the distribution theory of wavelets are easily derived when the model is one of a signal observed with random error. We can summarize the above development by representing the wavelet transform by:

$$w = Wy \tag{18}$$

where y is the vector of N observations on a signal, w is the N dimensional vector of wavelet coefficients, and W is an N×N orthonormal matrix that summarizes the transformations listed above. The first elements of w are the coefficients s_J, the last elements of w are the coefficients d_1. If the number of observations, N, is divisible by 2^J, a mathematical convenience, the number of coefficients of each type is given by:

- $\frac{N}{2}$ coefficients $d_{1,\,k}$

- $\frac{N}{2^2}$ coefficients $d_{2,\,k}$

- ....

- $\frac{N}{2^J}$ coefficients $d_{J,\,k}$

- $\frac{N}{2^J}$ coefficients $s_{J,\,k}$

- $\frac{N}{2} + \frac{N}{2^2} + ... \frac{N}{2^J} + \frac{N}{2^J} = N$

Consider the signal given by:

$$y = f + u \tag{19}$$

where f is a signal that may well be highly erratic and u is an error term with zero mean and constant variance. Following equation 18, we easily see that the coefficient vector w is now:

$$w = Wy = Wf + Wu \tag{20}$$

so that we see that the noise affects every coefficient, at least in theory. If we define Z = Wu, the vector of transformed error terms, the covariance of the error is:

$$Cov(Z) \quad = \quad V = WGW' \tag{21}$$
$$G \quad = \quad E\{uu'\} \tag{22}$$

If u is multivariate Gaussian, then so is Z. More importantly, even if u is not multivariate Gaussian, the wavelet coefficient estimates are asymptotically unbiased and approximately Gaussian under reasonably weak conditions. Further, the correlation of coefficient estimators across scales approaches zero as the scales separate under equally weak conditions; see for example, Percival and Walden (2000), Brillinger (1994) or Brillinger (1996), or Ramsey (1999a). Finally, the coefficient vector w is sparce; that is, many of the coefficients are zero or near zero. From a functional approximation perspective, this pragmatic result is important in that wavelets provide excellent data compression as will be demonstrated below.

The atomistic decomposition of waveform dictionaries using projection pursuit methods introduces another aspect to the general class of wavelet transforms and an alternative way to explore the properties of wavelets. An important shift in approach is introduced using these methods in that one includes the wavelet coefficients in terms of an ordering on the size of their moduli. This is in contrast to the methods stressed so far that presented the wavelets either according to scale for a given time instant or according to time for a given scale. The new approach is designed predominately for the exploratory phase in empirical analysis in which one begins with very little information on the nature of the signal being analyzed.

We begin by specifying a structural "atom" $g_\gamma(t)$:

$$g_\gamma(t) = \frac{1}{\sqrt{s}} g\left(\frac{t-k}{s}\right) e^{i\zeta t} \tag{23}$$

where g(.) is a mother or father wavelet from one of the classes defined above. But the basis function has now been augmented by multiplying the fundamental function g(.) by the transform, $e^{i\zeta t}$. In this formulation, the basis function $g_\gamma(.)$,

as is g(.) itself, is centered at k and its energy is concentrated in a neighborhood of k that is proportional to s, the scale factor. Correspondingly, the Fourier transform of $g_\gamma(.)$, is centered at ζ with energy in a neighborhood of ζ that is proportional to $1/s$, see Mallat (1989) and Mallat and Zhang (1993). This condition clearly indicates that one cannot achieve high resolution in both the frequency and the time domains, a gain in one necessitates a decline in resolution of the other. Wavelets allow one to chose the appropriate trade-off between resolution in the time and frequency domains, Fourier stresses resolution in the frequency domain at the expense of the time domain.

The initial analytic procedure is the same in that we determine the coefficients of the projection on to the basis created by the sequence of functions $g_\gamma(.)$. We define, using the symbol $\langle . \rangle$ to indicate the inner product:

$$\alpha_n = \left\langle f, g_{\gamma_n}(t) \right\rangle \tag{24}$$

and

$$f(t) = \sum_{n=1}^{N} \alpha_n g_{\gamma_n}(t) \tag{25}$$

where γ_n is the nth. selection from a discretized set, $\gamma_n = \{s_n, k_n, \zeta_n\}$; and s_n is defined in terms of the usual dyadic expansion. N represents the total number of observations that for mathematical convenience is assumed to be divisible by 2^J.

There are two major elements of this formulation that require discussion. First, we note that waveform dictionaries are able in the context of a single basis function class to represent both highly non-differentiable functions, certainly non-smooth functions, and functions that are naturally representable by combinations of harmonic series. The second distinction to the previous analytical approach is that the expansion is now ordered in terms of the maximum modulus of the coefficients irrespective of scale or location. Implicitly in the discussion of the discrete wavelet transform, attention was focused either on the analysis of the signal at a given scale over varying locations, or at a particular location over varying scales. Waveform dictionaries are particularly good at detecting harmonic based signals, as well as those signals best amenable to analysis by projection onto compact supports. Wavelets are good at detecting dirac delta functions, chirps (that is, signals that represent bursts of energy that vary in frequency and amplitude), as well as signals that exhibit phase shifts and isolated discontinuities. In short, when basic prior information about the nature of a signal is sparse, waveform dictionaries provide an excellent exploratory tool.

The main type of wavelet discussed so far is known as the "discrete wavelet transform, DWT, or "atoms" in the waveform dictionary expansions. There are in fact many generalizations that we could examine; such as the maximum overlap discrete wavelet transform, MODWT, also known as the "non-decimated wavelet

transform," or the "stationary DWT," or the "translation invariant DWT," that gains in resolution of a signal, but loses the property of orthogonality. The MODWT transform can provide a multiresolution analysis of the data, its coefficients are associated with zero phase filters, it is translation invariant as one of its pseudonyms states, and the transform can be applied to data sets whose length is not divisible by 2^J.

We can also consider the continuous wavelet transform, see for example, Percival and Walden (2000), pages 9,10. We can define,

$$W(\lambda, t) = \int_{-\infty}^{\infty} \psi_{\lambda,t}(u)x(u)du \tag{26}$$

$$where\ \psi_{\lambda,t}(u) \equiv \frac{1}{\sqrt{\lambda}}\psi\left(\frac{u-t}{\lambda}\right) \tag{27}$$

The main difference is that in equation 26 we consider continuous variations in the scale component, λ, and in the time component, t. For example, we might define:

$$\psi\left(\frac{u-t}{\lambda}\right) = \begin{cases} -\frac{1}{\lambda} & if & t-\lambda < u \le t \\ \frac{1}{\lambda} & if & t < u \le t+\lambda \\ 0 & if & otherwise \end{cases} \tag{28}$$

The definition of "wavelet packets," that is, the discrete wavelet packet transform, DWPT, enables one to capture periodic components within a series, see Percival and Walden (2000). While the DWT decomposes the frequency range [0,1/2] into adjacent intervals of frequency bands, the DWPT coefficients can be localized to a particular narrow band of frequencies and at a particular interval of time; thereby creating a "time-frequency" decomposition. The DWPT provides a link between time-scale and time-frequency decompositions; Percival and Walden (2000), page 206.

In particular we might mention in passing the 2-D wavelet transforms that may well in the future prove to be very useful in the analysis of economic data, especially in the context of economic relationships that are defined with respect to geographic regions or social groups.

3.   Wavelets and Insight

3.1.   *Noise smoothing versus de-noising*

The standard statement of a noisy signal tends to mask crucial assumptions about the relationship between the noise and the signal. For example, the usual statement is:

$$y_t = f_t + \varepsilon_t \tag{29}$$

where y_t is the observed signal, f_t is the actual signal, and ε_t is the noise. In economics and finance, the often unstated assumption is that the signal, f_t is smooth, but the noise is definitely non-smooth. Under these assumptions, the generic answer to approximate f_t is to use some form of averaging over the noise. However, when the signal is either non-smooth itself, or contains discontinuities, or regime shifts; the smoothing recommendation is counter productive in that the very structure of the signal that one is trying to capture will be lost, or at least, distorted by the process. Such circumstances require de-noising, not smoothing. The principle idea behind "de-noising" is that one can define a "noise threshold" such that variations in the data below the threshold are to be regarded as noise, whereas variations greater than the threshold are regarded as "signal." De-noising and smoothing are complementary approaches and which is used depends on the characteristics of the data generating mechanism. Economists should be aware of the benefits from using the de-noising approach in the context of models with regime shifts and other forms of discontinuities or points of non-differentiability.

Recall the wavelet transform in its orthonormal matrix version:

$$w = Wy = Wf + W\varepsilon \tag{30}$$

What is required is Wf, but what we observe is Wy that is contaminated by $W\varepsilon$. If it is reasonable to assume that the modulus of Wf is large relative to the variance of $W\varepsilon$, and if we can obtain a reasonable estimate of the variance of $W\varepsilon$, we can calculate a noise floor. We can proceed by shrinking the observed coefficients and setting those below the noise floor to zero. There are several ways to achieve this result. One pair of ways is that recommended by Donoho and Johnson in a series of path breaking articles, Donoho and Johnstone (1995), Donoho and Johnstone (1998), Donoho et al (1995). They defined soft shrinkage by:

$$\delta_s(w) = \begin{cases} 0 & if \quad |w| \le c \\ sgn(w)(|w| - c) & if \quad |w| > c \end{cases} \tag{31}$$

and hard,$\delta_h(w)$, by:

$$\delta_h(w) = \begin{cases} 0 & if \quad |w| \le c \\ w & if \quad |w| > c \end{cases} \tag{32}$$

$\delta_s(w)$, soft shrinkage, returns the amount by which $|w|$ exceeds the threshold, c, zero otherwise. $\delta_h(w)$ returns w itself, or zero in the same circumstances. Yet another alternative is to use Breiman's Garrote, Breiman (1995):

$$\delta_g(w) = \begin{cases} 0 & if \quad |w| \le c \\ w - \frac{c^2}{w} & if \quad |w| > c \end{cases} \tag{33}$$

In all three cases, the value returned if the estimated coefficient is below the threshold is the same, zero. The three cases differ in how they treat the value of the coefficient when the threshold is exceeded.

The main difficulty is that some judgement is needed to settle on the appropriate value for the threshold given an estimate for the variance of the noise terms, see for example, Johnstone and Silverman (1997). As a pragmatic matter, it is customary to regard the "coefficients" at the d_1 level as representing almost entirely noise; often this assumption is extended to the d_2 coefficients. The concept is that at the d_1 level the signal to noise ratio is close to zero, whereas at higher scales the ratio is increasing rapidly. Johnstone and Silverman (1997) have recommended that in many cases one should calculate the noise threshold scale by scale.

In any event, the sparce matrix characteristic of W is enhanced by this approach so that the coefficient vector w contains even more zero terms, especially at the lower scales. The distinction between zero and non-zero coefficients is sharpened by these procedures with a potential bias towards eliminating actual, but very small coefficients. However, the overall impact on the accuracy of the multiresolution approximation will not be affected to any appreciable degree.

The implications for insight into the analysis of non-homogeneous signals is very important as the graphs presented by Donoho et al (1995) demonstrate. In every case the essential and distinctive elements of the signals would have been lost by smoothing. In economics and finance much attention has been paid to the issue of regime shifts and threshold models, see for example, Frances and Dijk (2000). Clearly, these are situations in which error thresholding of some type is essential and the thoughtless use of averaging procedures masks, at least distorts, the signal characteristics being sought.

A related aspect to detecting regime shifts and discontinuities is the choice of wavelet to use. Given the discussion above one choice would be the Haar wavelet. In general, wavelets that are zero phase, not very smooth, and have very few vanishing moments would be most suitable for detecting regime shifts and discontinuities. One might well consider different optimal wavelet properties at different scales. For example, it might well be the case that at the longest scales the series is very smooth, but that at periods representing the business cycle, the series is non-smooth, and at even smaller scales the series is both smooth and periodic. Wavelet analysis can easily be structured to deal with these cross scale complexities.

A use of wavelets that may well prove to be exceedingly beneficial in understanding economic, but more particularly financial data, is the decorrelation of long memory processes and the consequent improvement in the estimation of long-memory parameters. While no striking insights have yet been obtained in financial data; the promise is there; see for example an intriguing series of papers, Fan and Whitcher (2001), Jensen (1999), Jensen (1999a), McCoy and Walden (1996),

Whitcher (2000), and Whitcher (2001)

3.2.   Time scale decompositions of economic relationships

The dependence of physical relationships on the concept of time scale is wide spread and fundamental in all sciences. In Ramsey and Lampart (1998a) and Ramsey and Lampart (1998) the concept was reintroduced into the analysis of economic and financial relationships. Consider for example any discussion of the term structure of interest rates, or the elementary statements of micro economic analysis, Henderson and Quandt (1980), or the distinction between permanent and transitory shocks, or the distinction between equilibrium and the efficiency of dynamic adjustment, Blanchard and Fischer (1992); all involve the notion, at least implicitly, of time scale. The situation can be easily illustrated. Imagine a sequence of traders that make decisions over different horizons; for example, one can visualize traders operating minute by minute, or hour by hour, or day by day, or month by month, or year by year. Or consider the difference in time horizon and its effects on bond holdings between short term money managers and those determining the investment portfolio for an insurance company. Alternatively, imagine an individual deciding on the purchase of a house, a car, groceries, a chocolate bar. In these examples, it is clear that those variables that would receive the most attention, or weight in the decision process, and likely the structure of the relationship itself will vary over the different time scales that are implicitly defined by the different decision making horizons. For the consumer as one varies the time scale from longest to shortest the relevant time horizon for decisions shortens, the terms of the relevant interest rates shorten, the "permanence" of income shortens. In general one would expect the nature of the relationship between the relevant variables to vary with scale, not to mention the size of the corresponding coefficients. In short, markets are composed of the actions of a variety of agents that are operating at each moment at different time scales. Indeed, each agent operates on many scales simultaneously. These distinctions have been long enshrined in economics in terms of the short and long runs. While pedagogically useful, the reality is that there are many time scales and there may well be a continuum of scales. Consequently, the time and cross-sectionally aggregated time series that constitute the economists' observations are an amalgam of sub-relationships that are defined over different time scales. An early application of this is in Serroukh and Walden (1998), where the idea that the cross covariances between scales might differ was successfully explored using data on pack ice in the Beaufort Sea. Another example is Whitcher et al (2000) in which wavelet cross-correlations were used to investigate short and long term atmospheric relationships. See also Gencay et al. (2002) for examples in economics and finance.

In Ramsey and Lampart (1998a) and Ramsey and Lampart (1998) two relationships were examined; that between expenditure and income and that between money and income. With respect to the former, the claim that the relationship would vary and that the relevant variables would differ across scales was confirmed. The real interest rate was discovered to be a significant variable only for the longest time scales and only for durable goods. For both durable goods and for non-durable goods, the degree of fit and the strength of the relationship declined monotonically as the scale decreased. At certain scales the relationship between expenditure and income was seemingly more complex than a simple linear relationship; this will be discussed in full in the next section.

For the money income relationship, the question of major interest is whether "money causes income" or "income causes money" in the Granger sense. The result of the empirical analysis by Ramsey-Lampart of this long debated and inconclusively resolved issue is that at the shortest scale, income causes money, that at intermediate scales, money causes income, and that at the longest scales, there is a feedback mechanism. All of this not only accords with theory, but parts of this result are aspects of conventional wisdom in the literature. Recently, in Gencay et al. (2002) the empirical results on the money income relationship were confirmed by employing the same techniques, but using slightly different data definitions, for the U.S., United Kingdom, Japan, and Austria. For all countries, except Austria, the Ramsey-Lampart results obtained were almost exactly the same qualitatively. Austria was different in that the first three scales had money causing income, but at the longest scales the feedback mechanism prevailed. This substantial confirmation is a remarkable result in econometric analysis.

Gencay et al. (2002) also examined the relationship between money growth and inflation. In this example as well as the previous ones the extant empirical literature is plagued by the fact that the empirical relationship between money and the price level "changes according to the sampling period, the level of monetary aggregation, and the methodology employed to filter the price level and money," Gencay et al. (2002), page 157. Six countries were examined; Argentina, Brazil, Chile, Israel, Mexico, and Turkey. For all countries at the highest available time scale, 32 months, the relationship is a feedback mechanism. The empirical results are clearest for Argentina and Turkey for which countries, money Granger causes inflation at the shortest scales and is a feedback mechanism at the longest scales. The results are mixed for Israel and Mexico and little can be said for Chile at the lowest scales. Nevertheless, given the known difficulties with the data and their limited extent, these are remarkable results.

We can reasonably conclude from this analysis that decomposing macro-economic time series into their time scale components is a very successful strategy in trying to unravel the relationship between economic variables. Relationships that are at best problematical using standard methods and aggregated data are revealed in a

consistent manner using time scale decompositions.

But preliminary results indicate that the success of time scale analysis using stock and foreign exchange market data is less persuasive to the economics profession. Presumably this is due to the much greater degree of "mixing" that is inherent in these data due to the rapid and extensive arbitrage activities that characterize financial markets so that the obvious results obtained for macro variables are not as prevalent.

However, this is not to claim that wavelet analysis does not have much to contribute; indeed there is very much to be learned as is clearly demonstrated in some papers by Gencay and coauthors. Interesting examples are represented in a recent series of papers, Gencay et al (2001), Gencay et al (2001a), and Gencay et al (2002a) In these papers the authors demonstrate the benefits of wavelet analysis in evaluating variations in foreign exchange volatility by time scale. They also demonstrate considerable improvements in the estimation of measures of systematic risk within the CAPM model and provide insight into the process that can be obtained through deconstructing the time series by scale. As one might anticipate, the effect of beta and its relevance to the analysis of portfolios increase with time scale.

3.3.   Delays as functions of the state space

Mentioned above was the observation concerning both Ramsey and Lampart (1998a) and Ramsey and Lampart (1998) that at certain scales the relationship between the variables expenditure and income and that between money and income were surprisingly complex. Further examination of these data revealed an interesting and potentially important insight into the nature of the relationship between economic variables over time. The graphs in Ramsey and Lampart (1998a)and Ramsey and Lampart (1998)demonstrated that at some scales the phase relationship between expenditure and income and that between money and income varies over time. Recall that two series are in phase if peaks match peaks and troughs match troughs in time. There is a fixed "delay," if the time difference between peaks and between troughs remains constant. In the expenditure-income case the observed historical period is characterized by the two series being out of phase and moving into phase, whereas for the latter pair of variables the phase relationship varies continuously.

The standard unchallenged assumption in economics is that if there are delays between two variables, that delay is fixed. What was discovered in the papers cited above is that the delays at certain scales are functions of time and are likely to be functions of the state space. This is an interesting and challenging research opportunity and one that opens a number of intriguing possibilities. At the very least one is encouraged by these results to speculate about the reasons for the time varying delays. Cursory introspection indicates that the "timing" of action

by economic agents is a neglected aspect of behavior. Recent events illustrate this remark. For example, the 2001 push by auto-manufacturers to lower the purchase price on cars presumably had two effects; one is undoubtedly increased quantity demanded in reaction to an implicit price decline, but the other was to shorten the delay between income variation and its assessment and actual expenditure. Alternatively, pre-announced price changes will induce both demanders and suppliers to alter the timing of their market transactions. For example, some years back the Federal Energy Office, F.E.O. as it was known then, announced on a particular date that it was going to mandate a new higher price a few days later. Gas stations were swamped with consumers and gas stations were most reluctant to supply gas until the new prices were in effect; most of this change in excess demand stemmed from the change in timing of supply and demand.

A more macro aspect of the phase difference between variables is illustrated in terms of leading and lagging indicators and the inevitable shift in the lead-lag relationship, at least in terms of the timing of the relationship that is involved, see Diebold (1998). Diebold discusses the relationship between housing starts and housing completions, Diebold (1998), page 325-6. It is easy to recognize that while on average there may well be a standard lag between starts and completions, variations in economic conditions will impact the lag; either lengthening it or shortening it. Similar variations between leading and lagging variables is illustrated for cattle and hogs. The supply of piglets and calves anticipates the future supply of hogs and cattle for slaughter. Nonetheless, it is well known that variations in economic conditions can easily alter the timing of bringing hogs and cattle to market, Nerlove et al (1979). All of these examples indicate the opportunity for varying the lead-lag relationship between variables as a function of the state of the system. A simple regression between the leading and lagging components that does not recognize the dependence of phase on economic conditions will inevitably discover needless complexities in the relationships.

The general equation that we should consider; perhaps at only certain time scales, is:

$$Y_t = \beta_0 + \beta_1 X_{t-d_1(Z)} + \beta_2 X_{t-d_2(Z)} + \varepsilon_t \tag{34}$$

where the $t - d_1(Z)$ and $t - d_2(Z)$ represent the state space dependent delays and Z represents the component of the state space that affects the timing of the delay. An immediate research agenda that arises from this observation is to determine the extent to which time varying delays are present in economic and financial data. The next task of course is to discover the economic mechanism that generates these variations. In particular, one might well question why at a time scale of 32 months, income and expenditure have historically, at least for the recent past, been related by time varying phase. A corresponding question is why this problem is revealed at 32 months, not at other scales, given of course, that the result is robust to subsampling and re-estimation using similar, but different data.

An excellent exploratory tool for discovering time varying delays is to examine the multi-resolution decomposition of economic signals. By comparing the scale decompositions of time series for variables that are theoretically linked, one could discover whether at some time scales phase variation is involved. Under some circumstances, Fourier analysis could be used to determine the phase relationship. Indeed, exactly this has been done with some aggregated macro time series using oblate spheroid wave functions to determine the phase drift, see for example, Ramsey and Thomson (1999). Any regression analysis that does not allow for phase variation will fail to reveal what might well be a simple, even linear, relationship.

3.4.   Atomistic decomposition using waveform dictionaries of financial data

In a previous section I discussed the notion of atomistic decompositions of a time series using waveform dictionaries. This is an exploratory tool for examining the structure of financial data in the sense that one would examine the extent to which the data can be represented by a mixture of wavelets and frequency components. In particular, the approach enables one to evaluate the extent to which frequency components of a signal can be detected when one allows the frequency part of the signal to oscillate in amplitude, including disappearing altogether at times. In Ramsey et al. (1995) the scaling properties of daily observations of the Standard and Poor market index were analyzed and the random walk nature of the data questioned. These data were discovered to contain properties that indicated that the data were far more complex than previously assumed. Further in Ramsey and Zhang (1996) 16384 daily observations of the Standard and Poor's 500 index were analyzed by atomistic decomposition using waveform dictionaries and in Ramsey and Zhang (1997) one year of tic by tic data on world wide foreign exchange data were observed between the U.S. dollar and the Deutschmark for the time period October 1992 to September 1993. The data were obtained from Olsen Assocs. in Zurich and were also analyzed by atomistic decomposition.

There are some similarities and some differences in the analytical results using these two similar, but different data sources. The first general conclusion is that a relatively small number of wavelet coefficients are needed to describe the observed time series; there is, in short, considerable data compression available using wavelets. For example, using the Standard and Poor's log first differenced data only about 1000 coefficients were needed to fit the data points very closely, whereas performing the same procedures on random data would require 3000 coefficients for the same degree of fit. The results are even more impressive for the foreign exchange data where only 100 wavelet coefficients provide a very close fit to the data. However, notwithstanding the very good fits, the ability to forecast is minimal. Wavelets provide an explanation for a widespread phenomenon in economic and financial research in that while good fits can often be obtained, our ability to forecast is very poor. For both sets of data, all the power is in chirps, Dirac

delta functions, and for the foreign exchange data, a few frequencies that come and go, seemingly at random. Further, there is no pattern in the distribution of the occurrence of the chirps and of the Dirac delta functions in either case.

In short, while there seems to be in any given historical period a relatively simple dynamical system as represented by a small number of wavelet coefficients, the ability to forecast is minimal as there is no repeatable structure to the distribution of the chirps and Dirac delta functions. In essence, with these data we seem to have a sequence of random selections of relatively simple models. I believe that this circumstance underlies much of the explanation for good fits, but poor forecasts.

Another common aspect of these two data sets is that while the signs of differences over most sub-periods seem to be Binomial with a probability of 0.5, there is some persistence in the absolute values of the differences; that is, there is persistence only in the magnitudes of the changes. These results indicate a common aspect to the financial markets. Instead of markets smoothly converging towards a new temporary equilibrium generated by new information, the markets instead react with bursts of intense activity covering a narrow band of frequencies; the adjustment process is itself oscillatory. Further, the structure in the data that is observed using the levels data is removed by differencing, but that the observance of weak structure is reintroduced when the differenced data are converted to moduli.

3.5.   *Waxing and waning of frequencies*

The financial data discussed in the previous section produced a further important insight. The traditional approach to the application of spectral techniques to economic and financial data, has focussed on the discovery of complex, but stable, frequency components; this research has discovered little evidence of stable frequencies in financial data and not much more in economic data. What was discovered in the financial data using wavelets is that there may be some evidence of low frequency components that wax and wane in strength, so that an analysis that seeks to discover the global presence of frequencies is likely to fail. This result opens up a new and potentially exciting avenue of research as we seek to determine the factors that give rise to the oscillation in the strength of the signals and to seek an explanation for the distribution for the observed frequencies. The other aspect that is of interest is that the occurrence of the frequencies is concentrated in the lowest frequencies; indeed, the relative concentration increases as the frequency is reduced.

3.6. *Forecasting and structural change*

In a previous section, I introduced a key aspect of the forecasting problem in economics and even more significantly in finance. Wavelets clarify and put into stark relief the nature of the problem that while time series fits over historical data periods are uniformly good, indeed, often very good, the ability to forecast past the very near future is very bad. We seldom anticipate major events in either the financial markets or in the economy generally. The relatively reasonable average GNP forecast results for projections one quarter, or even one year ahead, are I claim due primarily to the inevitable inertia in the economy; witness the difficulty in beating naive forecasts that is so prevalent in economics and finance.

The key difficulty lies in the forecasts for the intermediate to long term period. While I have argued elsewhere, see Ramsey (1996), that on average our forecast errors for intermediate term forecasts lie well outside any reasonable confidence interval, the question is why. One obvious answer is that the conditions needed for the forecast error bounds to be applicable may well have changed since the forecast was made. For example, the assumption that certain variables would remain constant, or that estimated relationships would remain fixed, is violated so that the estimated model no longer applies. The wavelet results for financial data in particular indicate that the problem is more severe. This observation was summarized in Ramsey (1996) by stating that regression, linear or not, is local, whereas forecasting is global in structure. Good fits do not, indeed cannot, guarantee good forecasts.

The main result from the wavelet analysis is that while a very small number of wavelet coefficients are needed to provide very good fits to any historical data, (there is very good data compression), the set of relevant coefficients varies period by period. This is seen more clearly when one recognizes where the power in the wavelet analysis resides; the power is in dirac delta functions, chirps, and sporadic occurrences of harmonic signals. There is no observable pattern to the temporal distribution of the coefficients. Contrast this situation with that obtained by estimating even a complex combination of harmonic signals. Forecasting is easy in this alternative case if the presumption of the continuance of the signal as observed over the historical period holds, whereas in the former situation, there is nothing forecastable. Expressing the matter this way shifts the question from why are forecasts so poor, to how do we manage to fit the data with so few wavelet coefficients. The key is that while the economy is continually evolving and suffers from random noise effects, the average observed structure at any one point in time is quite simple, thereby giving rise to a relatively few coefficients needed to describe the current system. The source of this observation requires some deep analysis. Inertia obviously plays a role in that the evolution from one state to another requires time and the system is robust to small perturbations. However, this is not likely to be the entire explanation. A guess is that market clearing imposes suffi-

cient constraints on the system that fluctuations are not entirely random, but are at least partly constrained. An alternative view is that over any historical period, our models represent (usually) linear approximations to a non-linear manifold. It requires only a relatively small change in the fundamentals to induce a substantial change in the linear approximation.

Notwithstanding the negative position taken above, there are useful examples of gains in forecasting economic variables using wavelets. There are two sets of examples. The first is by Arino, see Arino (1996) in which the aggregate time series is broken down into a seasonal and a trend component by a wavelet decomposition and traditional means are used to forecast the individual components; an aggregate forecast can be obtained by recombining the component series. The results, using Spanish car sales figures, are encouraging. An alternative approach also uses the idea of decomposing the signal to be forecast into its time scale components, but instead uses neural networks on each component to produce an overall forecast after recombining the components, see Aussem and Murtagh (1997) and Aussem et al (1998); the former forecasts the sun spot series and the latter forecasts five day ahead closing equity prices. Another example that illustrates the forecasting process using MODWT on the Dow Jones Industrial averages is Fryzlewicz et al.(2002). A very recent paper indicating the benefits for forecasting using a wavelet approach is Li and Hinich (2002).

While these approaches are very promising, the basic problem still remains that forecasting is global, whereas fitting is local. To use the latter to determine the former requires strong untestable assumptions; untestable that is before the occurrence of the forecast events. There is little evidence that the economy, or financial markets, are any more representable as steady states than heretofore.

A criticism that might well be levelled against this approach of separating the time series into its time scale components is that for the longest time scales in particular, there will be a serious problem generated by the truncation of the data series. However, this is a problem with forecasting any series using any non-causal filter; it is not a problem restricted to wavelets. What may not be appreciated is that the wavelet approach stimulates innovative solutions to what is a generic problem, see for example an article on the persistence of output in the business cycle by time scale; Jensen and Liu (2002). Indeed, in the wavelet approach, one is not restricted to a single solution for the entire series; one can separate the series into its time-scale components and apply the appropriate procedure to each component. For example, if one component is seasonal, one can extend the series by appending projections of the seasonal component, where there is a trend factor that can often be approximated by a polynomial.

4. Summary and Speculations on Potential Gains from Wavelet Analysis

Clearly the overwhelming gain in insight into the operation of economic forces stems from time-scale decomposition. The analysis has confirmed that time-scale matters; that allowing for time-scale resolves some anomalies in the literature, and that the results to date are in accord with basic economic theory.

The empirical results indicate new avenues for theoretical research. Either in terms of a market or with an individual, the analysis of agents operating on several scales simultaneously requires theoretical development in order to reconcile the various scale components; that is, some degree of coordination between scales is needed. This challenge is particularly acute in the context of a singe agent who is operating on several time scales simultaneously. The empirical results indicate that the individual's decision making process is more complex than economists have allowed so far.

Another stimulus to theoretical and empirical research is the discovery that the "delays" between variables in economic relationships may well be functions of the state space and certainly there is evidence that they vary over time. Most time series research assumes that the delays are fixed constants and that if the maximum delay is "k," then all delays are present from 1 to k. This latter assumption was questioned and empirically refuted in Ramsey and Anderson (2002) where the authors discovered using the time series version of multivariate adaptive regression splines, MARS, that many of the delays between 1 and the maximum value were zero. Now we recognize that any delay may well vary over the sample period. This discovery also stimulates interesting theoretical and further empirical research. First, can we discover the mechanisms that explain the variation in delays, does the presence of time varying delays resolve many puzzles about the lack of stability in economic and financial relationships? To what extent are the delays controllable through policy variations.

In a similar vein, the discovery in the context of financial data that there is evidence of very low frequencies, but that these frequencies wax and wane in amplitude provides an important insight. Standard procedures are unlikely to discover these time varying amplitudes. Nevertheless, the empirical result indicates that there is once again both a theoretical and an empirical challenge in that we need to confirm the extent of this phenomenon and then to elucidate the theoretical reasons that explain the occurrence.

The distinction between noise suppression and noise averaging is a very important innovation in econometric analysis. Increasingly, economists are paying more careful attention to discontinuities and regime shifts. The usual approach obfuscates the discontinuous changes that are being sought, whereas the wavelet

approach of noise suppression by determining a noise threshold facilitates detecting the precise location of the regime shift or discontinuity in the relationship. If not wavelets, certainly the use of noise thresholds, should become the standard for analysis in models that involve discontinuities and regime shifts.

Lastly, the empirical results on forecasting introduce some novel ideas, but the overall benefit is open to debate. As argued above, the difficulty is not so much with the use of wavelets, but the difficulties in forecasting in the context of continuously evolving and constantly randomly shocked economic systems. The major innovation in forecasting is to recognize that by deconstructing the total series into its constituent time scales, it is possible to tailor specific forecasting techniques to each time scale series and thereby gain in efficiency of forecast. This approach will be particularly beneficial in those circumstances where the relative amplitudes of the constituent time scales are varying over time. Wavelets provide yet another advance on the standard approach in that the wavelet procedure facilitates the construction of approximations to the formation of forecasts that are based on non-causal filters. That is, in such cases one is faced with the problem that one has to "forecast" half the length of a non-causal filter merely to obtain a one-step-ahead forecast. This is a common problem; the advantage of the wavelet approach is that one can tailor "optimum" extensions using time-scale deconstruction.

There are other avenues created by the wavelet approach. One interesting research line that has not yet been explored is to examine the standard assumption that the important variations in the economy are in terms of the business cycle and that the period of the business cycle is between three and six years. If this were true, the wavelet approach should detect very strong amplitudes for time-scale relationships at those periods. For non-durable goods production, we know that this is not true, for example, see Ramsay and Ramsey (2002) where there is reported a detailed examination of the seasonal and "business cycle" components of non-durable goods production. One could consider broadening the scope of the enquiry to examine the extent to which there are important resonating time-scales of any period. Preliminary analysis indicates a more complicated picture in that there seems to be considerable oscillation in the amplitudes of the higher time-scales. This result in itself, if confirmed, indicates interesting opportunities for research in that at the least the business cycle is highly variable in length and that the variation in relevant time-scale may well be subject to economic analysis.

An area of application of wavelets that should provide extensive and deep insights is in the analysis of the term structure of interest rates. Nowhere else is the role of the horizon of the decision maker on market outcomes so clearly indicated. Further, the available data on interest rates and bond maturities encapsulate the distinctions between decision makers with different time horizons. Wavelet analysis should provide insight into the relationship across time scales and therefore across bond maturities as well as the dependence of the term structure on economic

conditions.

The comments above have only begun to indicate the opportunities for research that are introduced by the wavelet approach. There are many other avenues that I have not introduced in this short article. For example, nothing has been said about bivariate wavelets that might well prove to be very useful in economic analysis, especially in those areas examining the dependence of economic relationships on geographic regions, or on the effects of social contiguity.

It is safe to claim that we will be impressed by the innovations that will be stimulated by the use of wavelets in empirical economic research over the next decade.

References

Arino, Miguel A. (1996}: "Forecasting Time Series via Discrete Wavelet Transform," working paper, Univ. de Navarra, Barcelona.

Aussem, A., and Murtagh, F. (1997): "Combining Neural Network Forecasts on Wavelet-transformed Time Series," *Connection Science*, **9**, No. 1, 113-121.

Aussem, A., Campbell, J., and Murtagh, F. (1998): "Wavelet-based Feature Extraction and Decomposition Strategies for Financial Forecasting," *Journal of Computational Intelligence in Finance*, March/April, 5-12.

Blanchard, O.J. and Fischer, S.(1992): *Lectures on Macroeconomics,* MIT Press, Cambridge.

Breiman, L.(1995): "Better subset regression using the nonnegative garrote," *Technometrics*, **37**, 373-384.

Brillinger, D.R.(1994): "Uses of cumulants in wavelet analysis," *Proceedings of SPIE Conference on Advanced Signal Processing*, **2296:** 2-18.

Brillinger, D.R. (1996): "Some uses of cumulants in wavelet analysis," *Journal of Nonparametric Statistics*, **6**, 93-114.

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

Carmona,R., Hwang, W., and Torresani, B (1998): *Practical Time-Frequency Analysis*, Academic Press, New York.

Chui, Charles K. (1992): *An Introduction to Wavelets*, Academic Press, New York.

Diebold, F.X. (1998): *Elements of Forecasting,* South Western, Cincinnati.

Donoho, D.L., and Johnstone, I.M. (1995): "Adapting to unknown smoothness via wavelet shrinkage," *J. Amer. Statist. Ass.*, **90**, 1200-1224.

Donoho, D.L., and Johnstone,I.M. (1998): "Minimax estimation by wavelet shrinkage," *Ann. Statist.*, **26**, 879-921.

Donoho, D.L., and Johnstone,I.M., Keryachariam, G. and Picard, D. (1995): "Wavelet shrinkage: asymptopia?" *J. R. Statis. Soc.* **57**, 301-369.

Fan, Y. and Whitcher, B. (2001): "A wavelet solution to the spurious regression of fractionally differenced processes," Economics Working paper, Vanderbilt Univ.

Frances, P.H., and Dijk, D.van (2002): *Non-linear time series models in empirical finance*, Cambridge Press, Cambridge.

Fryzlewicz, P., van Bellegem, S., and von Sachs, R. (2002): "Forecasting non-stationary time series by wavelet process modelling," Technical Report 0208, Institute de Statistique, Univ. Catholique de Louvain.

Gencay,R., Selcuk, F., Whitcher, B. (2001): "Differentiating intraday seasonalities through wavelet multi-scaling," *Physica A,* **289**, 543-556.

Gencay,R., Selcuk, F., Whitcher, B. (2001a): "Scaling properties of foreign exchange volatility," *Physica A,* **289**, 249-266.

Gencay,R., Selcuk, F., Whitcher, B. (2002): *An Introduction to Wavelets and Other Filtering Methods in Finance and Economics,* Academic Press, New York.

Gencay,R., Selcuk, F., Whitcher, B. (2002a): "Robustness of systematic risk across time scales," informal University of Windsor working paper.

Hardle, C., Kerkyacharian, G., Picard, D., and Tsybakov, A. (1998): *Wavelets, Approximation, and Statistical Applications*, Springer Verlag, New York.

Henderson, J.M., and Quandt, R.E.(1980): *Micro-economic Theory,* McGraw-Hill, New York.

Hess-Nielsen, N. and Wickhauser, B. (1996): "Wavelet and time-frequency analysis," *Proceedings of the IEEE,* **84(4)**, 523-540.

Jensen, M.(1999): "An Approximate wavelet MLE of short and long memory parameters," *Studies in Nonlinear Dynamics and Economics,* Academic Press, San Diego.

Jensen, M.(1999a): Using wavelets to obtain a consistent ordinary least squares estimator of he long-memory parameter," *Journal of Forecasting*, **18(1)**, 17-32.

Jensen, M. J., and Liu, M. (2002): "Do Long Swings in the Business Cycle Lead to Strong Persistence in Output," Univ. of Missouri Dept.of Economics Working paper.

Johnstone, I.M., and Silverman, B. (1997): "Wavelet Threshold Estimators for Data with Correlated Noise," *J. R. Statist.Soc., B.* **59,** 319-351.

Li, T.H., and Hinich, M. (2002): "A filter bank approach for modeling and forecasting seasonal patterns," *Technometrics,* **44(1)**, 1-14.

Mallat, S. (1989): "A Theory of Multiresolution Signal Decomposition: The Wavelet Representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* **11**, 674-693.

Mallat,S. and Zhang, Z. (1993): "Matching Pursuits with Time-Frequency Dictionaries," *IEEE Transactions on Signal Processing,* **41**, 3397-3415.

"On the phase of least-asymmetric scaling and wavelet filters," Techinical Report TR-95-15, Dept. of Mathematics, Imperial College of Science, Technology, and Medicine.

McCoy, E.J., and Walden, A.T. (1996): "Wavelet analysis and synthesis of stationary long-memory processes," *Journal of Computational and Graphical Statistics*, **5(1)**, 26-56.

Nerlove, M., Grether, D.M., and Carvalo, J.L. (1979): *Analysis of Economic Time Series,* Academic Press, New York.

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

Ramsey, James B. (1996): "If Nonlinear Models Cannot Forecast, What Use Are They?", *Studies in Nonlinear Dynamics and Econometrics (SNDE)*, **1**, No. 2, July, 65-86.

Ramsey, J. B., Zaslavsky, G., and Usikov, D. (1995): "An Analysis of U. S. Stock Price Behavior Using Wavelets," *Fractals*, Vol. 3, No. 2, 377-389.

Ramsey, J. B. and Zhang, Z. (1996): "The Application of Waveform Dictionaries to Stock Market Index Data," *Predictability of Complex Dynamical Systems*, edited by Yurii A.Kravtsov and James Kadtke, Springer-Verlag, 189-208.

Ramsey, J. B. and Zhang, Z. (1997): "The Analysis of Foreign Exchange Rates Using Waveform Dictionaries," *Journal of Empirical Finance*, **4**, 341-372.

Ramsey,J.B., and Lampart, C. (1998): "The Decomposition of Economic Relationships by Time Scale Using Wavelets: Money and Income," *Macroeconomic Dynamics,* **2**, 49-71.

Ramsey, J. B., and Lampart,C.(1998a): "The Decomposition of Economic Relationships by Time Scale Using Wavelets: Expenditure and Income," *Studies in Nonlinear Dynamics and Econometrics*, **3**, No. 4, 23-42.

Ramsey,J. B. (1999): "The contribution of wavelets to the analysis of economic and financial data," *Phil. Trans. R. Soc.Lond.A,* **357**, 2593-2606; reprinted in *Wavelets*, edt. Silverman B.W. and J.C. Vassilicos, Oxford Univ. Press, Oxford, 1999, Chapter 12, 221-236.

Ramsey,J. B. (1999a): "Regression over Time Scale Decompositions: A Sampling Analysis of Distributional Properties," *Economic Systems Research,* **11**, No. 2, 163-183.

Ramsey,J. B., and Thomson, D. (1999): "A Reanalysis of the Spectral Properties of Some Economic and Financial Time Series," *Nonlinear Time Series Analysis of Economic and Financial Data*, edt. Philip Rothman, Kluwer Academic Press, Boston, 45-85.

Ramsey,J. B., and Anderson, H. (2002): "U.S. and Canadian industrial production indices as coupled oscillators," *Journ. of Econ. Dynamics and Control,* **26**, 33-67.

Ramsay, J. O., and Ramsey, J. B. (2002): "Functional data analysis of the dynamics of the monthly index of nondurable goods production," *Journ. of Econometrics*, **107**, 327-344.

Serroukh, A., and Walden, A. T. (1998): "The Scale Analysis of Bivariate Non-Gaussian Time Series via Wavelet Cross-Covariance," Statistics Section Tech. Report, TR-98-02, Imperial College of Science, London.

Strang, G. and Nguyen, T. (1996): *Wavelets and Filter Banks*, Wellesley-Cambridge Press, Cambridge.

Whitcher, B. (2000): "Wavelet-based estimation for seasonal long-memory processes," Technical Report 00-14, Eurandom, The Netherlands.

Whitcher, B. (2001): Simulating Gaussian stationary processes with unbounded spectra," *Journal of Computational and Graphical Statistics,* **10(1),** 112-134.

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