Wavelets: A Personal Retrospective of a Decade's Research

James B. Ramsey
New York University, New York

Abstract

After stressing the importance of analyzing the various basis spaces, the exposition evaluates the alternative bases available to wavelet researchers I demonstrate the impact of choice of basis for the projection of the regressand which forms a linear space. This development is followed by a very brief overview of articles using wavelet tools. The comparative advantage of wavelets relative to the alternatives considered is stressed.

Keywords: Wavelets, Fourier series, B-splines, basis, unit roots, co-integration, outliers, forecasting, instrumental variables, errors in the variables, structural equations, thresholding.

I wish to thank, as will the reader, my colleague Marco Gallegati for his contribution to the paper for without it the paper would not have been completed on time.

Keynote talk for the Bank of Finland meetings in Helsinki, October 20th, 2011.

1 Introduction

In this paper, I will restrict my evaluation of this past decade's research to avenues of thought in which I have had some personal interest. This is not to deny the worth of the literature not cited, but reflects my reaction to assessing the overall enormous impact of wavelet analysis during the past decade; others will, I am sure, fill the gap. To those not cited my apologies.

The paper begins with a review of the main features of wavelet analysis which are contrasted with other analytical procedures, mainly Fourier, splines, and linear regression analysis. A review of Percival and Walden (2000), Bruce and Gao (1996), Crowley (2007), the excellent review by Gencay *et al.* (2002), or the Palgrave entry for Wavelets by Ramsey before proceeding would be beneficial to the neophyte wavelet researcher.

The next section contains a non-rigorous development of the theory of wavelets and contains discussions of wavelet theory in contrast to the theory of Fourier series and splines. The third section discusses succinctly the practical use of wavelets and the last section concludes.

Before proceeding the reader should note that all the approximating systems are characterized by the functions that provide the basis vectors. For projections, the orthogonal complement is the null space. For a regular regression framework, the basis is the standard Euclidean space, E_N. For the Fourier projections we have the frequency scaled sine and cosine functions that produce a basis of infinite power, high resolution in the frequency domain; e.g.

$$\begin{aligned}&\mathrm{Re}^{i2\pi ft} \text{ or alternatively expressed :}\\ &\qquad 1, \sin(k\omega t),\ \cos(\mathrm{k}\omega t),\ k = 1,2,3,...\end{aligned}$$

are highly differential, but are not suitable for analyzing signals with discrete changes and discontinuities.

The basis functions for splines are polynomials that are also differential and are defined over a grid determined by the knots; various choices for the differentiability at the knots determine the flexibility and smoothness of the spline approximation and the degree of curvature between knots.

Obviously, the analysis of any signal involves choosing both the approximating function and the appropriate basis vectors generated from the chosen function.

2 Functional Representation and Basis Spaces

2.1 An Overview of Bases in Regression Analysis

Relationships between economic variables are characterized by two universal components. Either the variable is a functional defined by an economic equation as a function of itself lagged to its past, i.e. is autoregressive; or is a function of time, i.e. is a "time series"; or it is a projection onto the space spanned by a set of functions, labeled, "regressors" each of which in turn may be autoregressive, or a vector of "time series". The projection of the regressand on the regressors provides a relationship between the variables, which is invariant to permutations of the indexing of the variables:

$$Y = X\beta + u$$
$$Y_{perm} = X_{perm}\beta + u_{perm}$$

where Y is the regressand,Y_{perm} the permuted values of Y, X_{perm}, represents a conformable permutation of the rows of X, and $u_{perm,}$ a conformable permutation to $Y_{perm.}$. However, if the formulation of the model involves an "ordering" of the variables over space, or over time, the model is then not invariant to permutation of the index of the ordering. It is known, but seldom recognized as a limitation of the projection approach, that least squares approximations are invariant to any permutation of the ordering. Consequently, the projection approach omits the information within the ordering in the space spanned by the residuals, which is, of course, the null space.

The other distinguishing characteristic is that added to the functional development of the variable known as the "regressand" is an unobserved random variable, "u", which may be represented by a solitary pulse, or may have a more involved stochastic structure. Complicating the situation is the presence of unobservable "error terms" in the observation of the regressors. In the former case, the regressand vector is contained in the space spanned by the regressors, whereas in the latter case the regressand is projected onto the space spanned by the regressors.

The usual practice is to represent regressors and the regressand in terms of the standard Euclidean N dimensional space; i.e. the ith component of the basis vector is "1", the remaining entries are zero; in this formulation, we can interpret the observed terms, $\{x_i\}$, $\{y_i\}$, i=1,2,...k; as N dimensional vectors relative to the linear basis space, E_N.

The key question the analyst needs to resolve is to derive an appropriate procedure for determining reasonable values for the unknown parameters and coefficients of the system; i.e. estimation of coefficients and forecasting of declared regressands. Finally, if the postulated relationship is presumed to vary over space or time, special care will be needed to incorporate those changes in the relationship over time or over the sample space.

Consider as a first example, a simple non-linear differentiable function of a single variable x, $f(x|\theta)$, which can be approximated by a Taylor's series expansion about the point $a_1$in powers of x:

$$y = f(x|\theta) = f(a_1|\theta) + f^1(a_1|\theta)(x - a_1) + \frac{f^2(a_1|\theta)(x-a_1)^2}{2!} \qquad (1)$$
$$+\frac{f^3(a_1|\theta)(x-a_1)^3}{3!} + \frac{R(\xi)}{4!}$$

for some ξ value. This equation approximately represents the variation of y in terms of powers of x. Care must be taken in that the derived relationship is not exact as the required value for ξ in the remainder term will vary for different values for a_1, x, and the highest derivative used in the expansion. Under the assumption that $R(\xi)$ is approximately zero, the parameters θ and the coefficient $a_1$can be estimated by least squares using N independent drawings on the regressand's error term, assuming the regressors are observed error free; one has:

$$\min_{\theta, a_1}\{\Sigma_1^N (y_i - f(x_i|\theta))^2\} \tag{2}$$

A single observation,i, on this simple system is:

$$y_i \ \{x_i, x_i^2 \ x_i^3\} \tag{3}$$

$$i = 1, 2, 3....N \tag{4}$$

This model is easily extended to differential functions which are themselves functions of multivariate regressors. The key aspect of the above formulation is that the estimators are obtained by a projection onto the space spanned by the regressors. Other, perhaps more suitable spaces, can be used instead. The optimal choice for a basis, as we shall see, is one that reduces significantly the required number of coefficients to represent the function $\{y_i\}$ with respect to the chosen basis space. Different choices for the basis will yield different parameterizations, the research analyst is interested in minimizing the number of coefficients; actually the dimension of the supporting basis space.

An alternative, ancient, procedure is provided by the monomials:

$$\{1, t^2, t^3, t^4, t^5, ...\} \tag{5}$$

that is, we consider the projection of a vector y on the space spanned by the monomials,$\{t^0 \cdot t^1,...t^k\}$, or as became popular as a calculation saving device, one considered the projection of y on the **orthogonal components** of the sequence in equation (5), see Kendall & Stuart, Vol 2 (1961).

These first two procedures indicate that the underlying concept was that insight would be gained if the projections yielded approximations that could be specified in terms of very few estimated coefficients. Further very little structure was imposed on the model, either in terms of the statistical properties of the model or in terms of the restrictions implied by the underlying theory. The results obtained by these procedures are best described as "exploratory."

2.2 Exponential and Power Bases

Two other simple basis spaces are the exponential:

$$\{e^{\lambda_1 t},\ e^{\lambda_2 t},\ e^{\lambda_3 t}.....e^{\lambda_k t}\} \tag{6}$$

and the power base:

$$\{t^{\lambda_1},\ t^{\lambda_2},\ t^{\lambda_3},\ ....t^{\lambda_k}\} \tag{7}$$

The former is most useful in modeling differential equations, the latter in modeling difference equations.

2.3 Fourier Bases

The next procedure in terms of longevity of use is Fourier analysis. The basis for the space spanned by Fourier coefficients is given by:

$$1, \sin(k\omega t), \cos(k\omega t), \tag{8}$$
$$k = 1, 2, 3, ...$$

The approximating sequences are given most simply by:

$$y = f(t) \cong \sum_{k=1}^{K} c_k \phi_k \tag{9}$$

where the sequence $\{c_k\}$specifies the coefficients chosen to minimize the squared errors between the observed sequence and the known functions shown in equation (9), ϕ_k is the basis function as used in equation (8). And the coefficients are given by:

$$c_k = \int f(t)\phi_k(t)dt \tag{10}$$

We note three important aspects of this equation. We gain in understanding if the number of coefficients are few in number; *i.e.* K is "small". We gain if the function "f" is restricted to functions of a class that can be described in terms of the superposition of trigonometric functions and their derivatives. The fit for functions that are continuous, but not every where differential can only be approximated using many basis functions. The equations generating the basis functions, ϕ_k, based on the fundamental frequency, ω, are re-scaled versions of that fundamental frequency. The concept of re-scaling a "fundamental" function to provide a basis will occur in many guises.

Fourier series are useful in fitting global variation, but respond to local variation only at very high frequencies thereby substantially increasing the required number of Fourier coefficients. For example, consider fitting a Fourier basis to a "Box function", any reasonable degree of fit will require very many terms at high frequency; see Bloomfield (1976).

Economy of coefficients can be obtained for local fitting by using windows; that is, instead of:

$$\hat{h}(\omega) = \frac{1}{2\pi} \sum_{-\infty}^{\infty} \hat{R}(s) \cos(s\omega)$$

where $\hat{R}(s)$ is the sample covariance at lag "s". We consider:

$$\hat{h}(\omega) = \frac{1}{2\pi} \sum_{-M}^{M} \lambda(s) \hat{R}(s) \cos(s\omega) \tag{11}$$

where $\lambda(s)$ is given by:

$$\lambda(s) = \begin{cases} 1, |s| \leq M \\ 0, |s| > M \end{cases} \tag{12}$$

$\lambda(s)$ is the "window function" which has maximum effect at s = 0, +-2kπ, k=1,2,3,...Distant correlations are smoothed, the oscillations of local events are enhanced; see Bloomfield (1976).

2.4 Spline Bases

A very versatile basis class is defined by the spline functions. A standard definition of a version of the spline basis , the B-spline, S(t), is:

$$S(t) = \sum_{k=1}^{m+L-1} c_k B_{k,(t,\tau)} \tag{13}$$

where S(t) is the spline approximation, {c$_k$}, are the coefficients of the projection, $B_{k,(t,\tau)}$ is the B-spline function at position k, with knot structure, τ. The vector τ designates the number of knots, L, and their position which defines the subintervals that are modeled in terms of polynomials of degree m. At each knot the polynomials are constrained to be equal in value for polynomials of degree 1, agreement for the first derivative for polynomials of degree 2, etc. Consequently, adjacent spline polynomials line up smoothly.

B-Splines are one of the most flexible basis systems, so that it can easily fit locally complex functions.

An important use of splines is to interpolate over the grid created by the knots in order to generate a differential function, or more generally, a differential surface. Smoothing is a local phenomenon, see de Boor (2001).

2.5 Wavelets

Wavelets provide one of the most flexible sets of basis functions.

The methodology is based on an analysis that enables us to decompose any time series into their time scale and periodic components. Wavelets are based on particular types of function $\Psi(t), \Phi(t)$ that are localized both in time and frequency domain and used to decompose a function $f(t)$ (*i.e.* a signal, a surface, a series, etc.) into more elementary functions. Unlike the Fourier transform, which uses the sum of certain basis functions (sines and cosines) to represent a given function may be seen as a decomposition on a frequency-by-frequency basis, the wavelet transform utilizes some elementary functions (father, Φ and mother wavelets,Ψ) that, being well-localized in both time and scale, provide a decomposition on a scale-by-scale basis as well as on a frequency basis. The inner product Φ with respect to f is essentially a low pass filter that produces a moving average; indeed we recognize the filter as a linear time-invariant operator. The corresponding wavelet filter is a high pass filter that produces moving differences; Strang and Nguyen (1996). Separately, the low pass and high pass filters are not invertible, but together they separate the signal into frequency bands. Corresponding to the low pass filter there is a continuous time scaling function $\phi(t)$. Corresponding to the high pass filter is a wavelet $w(t)$.

Consider the following equation:

$$w(t) = \sqrt{2} \sum d(k)\phi(2t - k)$$

where $d(k)$ are the high pass coefficients.

This gives wavelets a distinct advantage over a purely frequency domain analysis. Because Fourier analysis presumes that any sample is an independent drawing, Fourier analysis requires "covariance stationarity", whereas wavelet analysis may analyze both stationary and long term non-stationary signals. This approach provides a convenient way to represent complex signals. Expressed differently, spectral decomposition methods perform a global analysis whereas wavelet methods act locally in both frequency and time. Fourier analysis can relax local non-stationarity by windowing the time series as was indicated above. The problem with this approach is that the efficacy of this approach depends critically on making the right choice of window and, more importantly, presuming its constancy over time.

For wavelet analysis however, as we have observed, there are two basic wavelet functions, father and mother wavelets, $\phi(t)$ and $\psi(t)$. The former integrates to 1 and reconstructs the smooth part of the signal (low frequency), while the latter integrates to 0 (similarly to sine and cosine) and can capture all deviations from the smooth trend (high frequency). The mother wavelets, as said above, play a role similar to sines and cosines in the Fourier decomposition. They are compressed or dilated, in the time domain, to generate cycles to fit actual data. The approximating wavelet functions $\phi_{J,k}(t)$ and $\psi_{J,k}(t)$ are generated from father and mother wavelets through scaling and translation as follows:

$$\phi_{J,k}(t) = 2^{-\frac{J}{2}}\phi\left(\frac{t-2^J k}{2^J}\right) \tag{14}$$

and

$$\psi_{J,k}(t) = 2^{-\frac{J}{2}}\psi\left(\frac{t-2^J k}{2^J}\right) \tag{15}$$

where j indexes the scale, so that 2^j is a measure of the scale, or width, of the functions (scale or dilation factor), and k indexes the translation, so that $2^j k$ is the translation parameter.

Given a signal $f(t)$, the wavelet series coefficients, representing the projections of the time series onto the basis generated by the chosen family of wavelets, are given by the following integrals:

$$\begin{aligned} d_{j,k} &= \int \psi_{j,k}(t)f(t)dt \\ s_{J,k} &= \int \phi_{J,k}(t)f(t)dt \end{aligned} \tag{16}$$

where $j = 1, 2, ..., J$ is the number of scales and the coefficients d_{jk} and s_{Jk} are the *wavelet transform coefficients* representing, respectively, the projection onto mother and father wavelets. In particular, the detail coefficients $d_{Jk}, ..., d_{2k}, d_{1k}$ represent progressively finer scale deviations from the smooth behavior (thus capturing the higher frequency oscillations), while the smooth coefficients s_{Jk} correspond to the smooth behavior of the data at the coarse scale 2^J (thus capturing the low frequency oscillations).

Finally, given these wavelet coefficients, from the functions

$$S_{J,k} = \sum_k s_{J,k}\phi_{J,k}(t) \text{ and } D_{j,k} = \sum_k d_{J,k}\psi_{J,k}(t) \tag{17}$$

we may obtain what are called the *smooth* signal, $S_{J,k}$, and the *detail* signals, $D_{j,k}$, respectively. The sequence of terms $S_J, D_J, ..D_j, ..., D_1$ for $j = 1, 2, ..., J$ represents a set of signal components that provide representations of the original signal $f(t)$ at different scales and at an increasingly finer resolution level.

It is useful to note that whether we are examining wavelets or sinusoids or Gabor functions we are in fact approximating $f(t)$ by "atoms"; a collection of atoms is a "dictionary". We seek to obtain the best M atoms for a given $f(t)$ out of a dictionary of P atoms. There are three standard methods for choosing the M atoms in this over sampled situation. The first is matching pursuit in which the M atoms are chosen one at a time; this procedure is referred to as greedy and sub-optimal. An alternative method is the best basis algorithm which begins with a dictionary of bases. The third method is known as basis pursuit where the dictionary is still over complete. The synthesis of $f(t)$ in terms of $\phi_i(t)$ is under-determined. This brief discussion indicates that the essential objective is to choose a good basis. A good basis depends upon the resolution of two characteristics; *linear independence* and *completeness*. Independence represents uniqueness of representation and completeness ensures that any $f(t)$ is represented. Adding vectors will destroy independence, removing vectors will destroy completeness. If every vector v or function $v(t)$ can be represented uniquely as:

$$\begin{gathered} v = \sum b_i v_i \\ or \\ v(t) = \sum b_i v_i(t) \end{gathered} \tag{18}$$

If the coefficients b_i satisfy:

$$A||v||^2 \leq \sum |b_i|^2 \leq B||v||^2 \; with \; A > 0. \tag{19}$$

This is the defining property of a Riesz basis.

Much of the usefulness of wavelet analysis has to do with its flexibility in handling a variety of nonstationary signals. Indeed, as wavelets are constructed over finite intervals of time and are not necessarily homogeneous over time, they are localized in time and scale.

Any moderately experienced "Waveletor" knows to choose his wavelet function so as to maximize the "number of zero moments", to ascertain the number of continuous derivatives, and to worry about symmetry of the underlying filters. The more experienced Waveletor knows also to consider the shape of the function at zero scale. While many times the choice of wavelet function makes little or no difference there are times, when such considerations are important for the analysis in hand; for example, the inappropriate use of the Haar function for resolving continuous smooth functions, or using smooth functions to represent samples of discontinuous paths. Wavelets provide a vast array of alternative wavelet functions; e.g. Gaussian, Gaussian 1st derivative, Mexican hat, the Daubechies series, the Mallat series, and so on. The key to the importance of the differences lies in choosing the appropriate degree and nature of the oscillation within the supports of the wavelet function. I have previously stated that at each scale the essential operation is one of differencing using weighted sums; the alternative rescalable wavelet functions provide an appropriate basis for such differences. The astute listener would have noted that the concept of re-scaling the primitive function to provide a linear space with respect to which the function can be represented is not restricted to wavelets, but is also critical in constructing a basis space for Fourier series.

It is very useful to view the use of wavelets in "regression analysis" in greater generality than as a simple exercise in "least squares fitting." As indicated above the use of wavelets involves the properties of the implicit filters used in the construction of the wavelet function. Such an approach to determining the properties of wavelet analysis provides for a structured, but highly flexible system, that is characterized by a "scarce transformation matrix;"that is, most coefficients in the transformed space are zero. Indeed, the source of the benefit from creating a spanning set of basis vectors, both for Fourier analysis and wavelets, is the reduction in degrees of freedom from N, in the given Euclidean space, to K in the transformed space, where K is very much smaller than N; simple linear regression models illustrate the same situation and perform a similar transformation.

The argument so far, has compared wavelets to splines and to Fourier series or integrals. A discussion of the differences is required. Splines are easily dealt with in that the approximations implied by the spline procedure is to interpolate smoothly a sequence of observations from a smooth differential signal. The analysis is strictly local, even though most spline algorithms average over the whole sample space. The fit is almost entirely determined by the observed data points, so that, little structure is imposed on the process. What structure is predetermined is generated by the position of the knots.

Fourier series, or Fourier integrals, are strictly global over time or space, notwithstanding the use of windows to obtain useful local estimates of the coefficients. Wavelets, however, can provide a mixture of local and global characteristics of the signal, are easily modified to incorporate restrictions of the signal over time or space. Wavelets generalize Fourier integrals and series in that each frequency band, or octave, groups together, frequencies separated by the supports at each scale. A research analyst can incorporate the equivalent of a windowed analysis of Fourier integrals and incorporate time scale variations as in Ramsey and Zhang (1996) and Ramsey and Zhang (1997). Further, as illustrated by cosine wave packets, see Bruce and Gao (1996), and the wide choice for low and high pass filters, see Strang and Nguyen (1996), considerable detail can be captured, or suppressed and basic oscillations can be incorporated using band pass filters to generate oscillatory wavelets.

3 Some Examples of the Use of Wavelets

While it is well recognized that wavelets have not been as widely used in Economics as in other disciplines, I hope to show that there is great scope for remedying the situation. The main issue involves the gain in insight to be stimulated by using wavelets; quite literally, the use of wavelets encourages researchers to generalize their conception of the problem at hand.

3.1 Foreign Exchange and Waveform Dictionaries

A very general approach using time frequency atoms is especially useful in analyzing financial markets. Consider the equation $g_\gamma(t)$ where $\gamma = (s, u, \xi)$:

$$g_\gamma(t) = \frac{1}{\sqrt{s}} g(\frac{t-u}{s}) e^{i\xi t} \tag{20}$$

We impose the conditions ||g||=1 where ||g|| is L^2 where $g(0) \neq 0$. For any scale parameter s, frequency modulation ξ and translation parameter u: the factor $1/\sqrt{s}$ normalizes the norm of $g(t)$ to 1; $g(t)$ is centered at the abscissa u and its energy is concentrated in the neighborhood of u, size is proportional to s; the Fourier transform is centered at the frequency ξ and its energy is concentrated in the neighborhood of ξ and size is proportional to 1/s. Matching pursuit was used to determine the values of the coefficients. Raw tick by tick data on three foreign exchange rates were obtained from October 1, 1992 to September 30, 1993. The waveform analysis indicates that there is efficiency of structure but only at the lowest frequencies equivalent to periods of two hours with little power. There are some low frequencies that wax and wane in intensity. Most of the energy of the system seems to be in the localized energy frequency bursts.

The frequency bursts provide insights into market behavior. One can view the dominant market reaction to news as a sequence of short bursts of intense activity that are represented by narrow bands of high frequencies. For example, only the first one hundred structures provides a good fit to the data at all but the highest frequencies. Nevertheless the isolated bursts are themselves unpredictable, see Ramsey and Zhang (1997).

3.2 Instrumental Variables and "Errors in the Variables"

To begin the discussion the "errors in variables" problem is still as nearly unstructured as it has always been; that is, we endeavor to search for a strong instrumental variable. However it is very difficult to recognize one when shown a plausible variable. Further, it is as difficult to recognize a weak instrument that if used would yield worse results. I have labeled this approach "solution by assumption" since one has in fact no idea if a putted variable is, or is not, a useful instrumental variable.

Wavelets can resolve the issue; see Ramsey *et al.* (2010) and Gencay and Gradojevic (2009) for an extensive discussion of this critical problem. The task is simple: use wavelets to decompose the observed series into a "noise" component and a structural component, possibly refined by thresholding the coefficient estimates, Ramsey *et al.* (2010). The benefits from recognizing the insights to be gained from this approach are only belatedly coming to be realized. If all the variables in a system of equations can be factored into a structural component, {itself decomposable into a growth term and a oscillation term}, and into a noise term; e.g.

$$
\begin{aligned}
y_i &= y_i^* + \varepsilon_i \\
x_i &= x^* + \eta_i \\
z_i &= z_i^* + \omega_i
\end{aligned}
$$

where the starred terms are structural and the terms $\varepsilon_i,\ \eta_{i,}\ \omega_i$ are random variables either modeled as simple pulses or have a far more complex stochastic structure, including having distributions that are functions of the structural terms. If we wish to study the structure of the relationships between the variables, we can easily do so; see Silverman (2000) and Johnstone (2000). In particular, we can query the covariance between the random error terms, select suitable instrumental variables, solve the simultaneous equation problem, and deal effectively with persistent series.

Using some simulation exercises Ramsey *et al.* (2010) demonstrated how the structural components revealed by the wavelet analysis yield nearly ideal instrumental variables for variables observed with error and for co-endogenous variables in simultaneous equation models. Indeed, the comparison of the outcomes with current standard procedures indicates that as the nonparametric approximation to the structural component improves, so does the convergence of the near structural estimates.

While I have posed the situation in terms of linear regression, the benefits of this approach are far greater for non-linear relationships. The analysis of Donoho and Johnstone (1995) indicates that asymptotic convergence will yield acceptable results and convergence is swift.

3.3 Structural Breaks and Outlier Detection

Most economic and financial time series evolve in a nonlinear fashion over time, are non-stationary and their frequency characteristics are often time-dependent, that is, the importance of the various frequency components is unlikely to remain stable over time. Since these processes exhibit quite complicated patterns like abrupt changes, jumps, outliers and volatility clustering, a locally adaptive filter like the wavelet transform is particularly well suited for evaluation of such models.

An example of the potential role to be played by wavelets is provided by the detection and location of outliers and structural breaks. Indeed, wavelets can provide a deeper understanding of structural breaks with respect to standard classical analysis given their ability to identify the scale as well as the time period at which the inhomogeneity occurs. Specifically, based on two main properties of the discrete wavelet transform (DWT), *i.e.* the energy preservation and approximate decorrelation properties, a wavelet-based test for homogeneity of variance (see Whitcher, 1998, and Whitcher *et al.*, 2002) can be used for detecting and localizing regime shifts and discontinuous changes in the variance[1].

Similarly, structural changes in economic relationships can be usefully detected by the presence of shifts in their phase relationship. Indeed, although a standard assumption in economics is that the delay between variables is fixed, Ramsey and Lampart (1998a and 1998b) have shown that the phase relationship (and thus the lead/lag relationship) may well be scale dependent and vary continuously over time. Therefore examining scale-by-scale overlaid graphs between pairs of variables can provide interesting insights into the nature of the relationship between these variables and their evolution over time (Ramsey, 2002). A recent example of this approach is provided in Gallegati and Ramsey (2011) where the analysis of such variations in phase is proven to be useful for detecting and interpreting structural changes in the form of smooth changes, for example, the q-relationship proposed by Tobin.

Finally, wavelets provide a natural way to seek for outliers in that wavelets allow for local distributions at all scales and outliers are at the very least a "local" phenomenon (see for a very brief introduction,

[1]See Gallegati *et al.* (2011) for the application to two well known events characterizing the US economy in the early 80s: the so-called Great Moderation and the change of monetary policy conduct after 1979 (*i.e.* Volcker disinflation period).

Wei, Z. *et al.*, , and Greenblatt, 1996). The idea of thresholding (see for example, Bruce and Gao, 1996, and Nason 2008), is that the noise component is highly irregular, but with a modest amplitude of variation, which is dominated by the variation of the structural component. Naively, outliers are observations drawn from a different distribution; intuitively one tends to consider observations for which the modulus squared is very large relative to the modulus of the remainder of the time series, or cross-sectional data. But outliers may be generated in far more subtle ways and not necessarily reveal themselves in terms of a single large modulus, but in terms of a temporary shift in the stochastic structure of the error terms. In these cases, "thresholding, in particular soft thresholding," Bruce and Gao (1996), Nason (2008), will prove to be very useful especially in separating the coefficient values of "structural components" from noise contamination.

3.4 Time scale relationships

The separation of aggregate data into different time scale components by wavelets can provide considerable insights into the analysis of economic relationships between variables. Indeed, economics is an example of a discipline in which time scale matters just because different agents' decisions are likely to have different time scales. Consider, for example, traders operating in the market for securities: some, the fundamentalists, may have a very long view and trade looking at firms' or market' fundamentals; some others, the chartists, may operate with a time horizon of weeks or days. A corollary of this assumption is that different planning horizons are likely to affect the structure of the relationships themselves, so that such relationships might vary over different time horizons or hold at several time scales, but not at others.

Although the concepts of the "short-run" and of the "long-run" are central for modeling economic and financial decisions, variations in the relationship across time scales are seldom discussed in economics and finance. We should begin by recognizing that for each variable postulated by the underlying theory we admit the possibility that:

$$y_s = g_s(y_{j,s}\, x_{i,s}),$$

where y_s is the dependent variable at scale "s," $\mathrm{g}_s(.)$ are arbitrary functions specified by the theory, which might differ across scales, $\mathrm{y}_{j,s}$ represents the codependent variables at scale "s", and $\mathrm{x}_{i,s}$ represents exogenous variables x_i at scale "s"; that is, the relationships between economic variables may well be scale dependent.

Following Ramsey and Lampart (1998a, 1998b) many authors have confirmed that allowing for different time scales of variation in the data can provide a fruitful understanding of the complex dynamics of economic relationships among variables with non-stationary or transient component variables. For example, relationships that are veiled when estimated at the aggregate level, may be consistently revealed after allowing for a decomposition of the variables into different time scales. In general, the results indicate that by using wavelet analysis it is possible to uncover relationships that are at best puzzling using standard regression methods (e.g. .Ramsey and Lampart (1998a, 1998b), Gallegati *et al.* ,2009, 2011).and that ignoring time and frequency dependence between variables when analyzing relationships in economics and finance can lead to erroneous conclusions.

3.5 Comments on Forecasting

The standard concerns about forecasting carry over to the use of wavelets, but as might have been anticipated wavelets incorporate a degree of refinement and flexibility not available using conventional methods, see for example, Diebold (1998). With wavelets, one can choose the scale at which the forecast is to be made, treating each scale level as a separate series for forecasting purposes. Secondly, one should note that at any given point in time, the "forecast",will depend on the scales at which one wishes to evaluate the forecast; for example, at all scales for a point in time, t_0, or for a subset of scales at time t_0. Further, one might well choose to consider, at a given minimum scale whether to forecast a range, given the chosen minimum scale, or to forecast a point estimate at time t_0.

These comments indicate a potentially fruitful line of research and indicates that the idea of "forecasting" is more subtle than has been recognized so far. Forecasts need to be expressed conditional on the relevant scales, and that the usual forecasts are special cases of a general procedure. Indeed, one concern that is ignored in the conventional approach is to recognize across scales the composition of the variance involved in term of the variances at each scale level. For examples, see Gallegati *et al* (2011a), Yousefi *et al* (2005) and Wei *et al.*(2006).

3.6 Some Miscellaneous Examples

Fan and Gencay (2007) have explored the gain in efficiency in discovering unit roots and applying tests for cointegration using wavelet procedures. Further, using MODWT multi-resolution techniques the authors demonstrate a significant gain in power against near unit root processes. In addition, the wavelet approach leads to a novel interpretation of Von

Neumann variance ratio tests.

Gallegati *et al* (2011) reviewed the literature on the "wage Phillips curve" using U.S. data. The most significant result of the multiscale analysis is that in the long run there is a one to one relationship between wage and price inflation and the close relationship of nominal changes with unemployment rate at business cycle scales. Over all, the paper suggests that allowing for different time scales of variation in the data can provide a richer understanding of the complex dynamics of economic relationships between variables. Relationships that are puzzling when tested using standard methods can be consistently estimated and structural properties revealed using timescale analysis. The authors note with some humor that Phillips himself can be considered as the first user of wavelets in Economics!

One of the most cogent rationalizations for the use of wavelets and timescale analysis is that different agents operate at different timescales. In particular, one might examine the behavior of central banks to elucidate their objectives in the short and long run. This is done in Aquiar-Conraria *et al* (2008) in assessing the relationship between central bank decision-making and government decision-making. The authors confirm that the macro relationships have changed and evolved over time.

In Rua and Nunes (2009) and Rua (2010), interesting results are obtained in both papers, which concentrate on the role of wavelets in the analysis of the co-movement between international stock markets. In addition, the authors generalize the notion of co-movement across both time and frequency. In Samia *et al* (2009), a wavelet approach is taken in assessing values for VaR's and compared favorably to the conventional ARMA-GARCH processes.

3.7 Conclusions

The functional representation of regression functions projected onto basis spaces was elucidated. The first step began with standard Euclidean N space and demonstrated a relationship to Taylor's series approximations, monomials, exponential and power bases. Fourier series were used to illustrate the relationship to wavelet analysis in that both versions included a concept of rescaling a fundamental function to provide a basis. Spline bases were also defined and related to wavelets. In the discussion and development of wavelets a number of aspects not normally considered were discussed and the concept of atoms was introduced. One can characterize the research analyst's objective as seeking to obtain the best M atoms for a given f(t) out of a dictionary of P atoms; the overall objective is to choose a good basis which depends upon the resolution of two characteristics; linear independence and completeness. Independence guarantees uniqueness of representation and completeness ensures that any f(t) is represented. Adding vectors will destroy independence, removing vectors will destroy completeness. The generality of wavelet analysis is enhanced by the choices available of functional forms to suit specific characteristics of the vector space in which the function resides; for example Haar, Gaussian, Gaussian first derivative, Mexican Hat and so on. In addition, further generalization of the approximation provided by wavelets is illustrated in terms of the Waveform dictionary which uses a triplet of parameters to represent translation, scaling, and is centered around a fundamental frequency $e^{i\xi t}$.

While the discussion above has demonstrated the wide usefulness of the wavelet approach, one might speculate that many more insights are liable to occur as the implications of this unique space are explored. Not enough attention has yet been expended on the wide variation in the formation of wavelet forms and their application in practical problems. In short, attention may well be concentrated in the future on capturing variation within the function's supports and thereby providing alternative determinations of very short run behavior. The implied flexibility of wavelets provides deconvolution of very short run phenomena as well as the intermediate run and long run phenomena.

The paper also contains brief reviews of a variety of applications of wavelets to economic examples which are of considerable importance to economists interested in accurate evaluations of policy variables A wide variety of data sources have been examined, including both macroeconomic and financial data. In these models the problem of errors in the variables is critical, but wavelets provide the key to resolving the issue. Some papers examine data for structural breaks and outliers. Comments on forecasting were presented. These thoughts indicate that forecasting is more subtle than is currently believed in that forecasts require to be calculated conditional on the scales involved in the forecast. Some forecasts might well involve only a particular subset of the time scales included in the system.

4 References

Aguair-Conraria, L., Azevedo, N., Aoares, M. (2008) "Using Wavelets to Decompose the Time-Frequency Effects of Monetary Policy." Physica A 387, 2863-2878.

Bloomfield, R. (1976) *Fourier Analysis of Time Series: An Introduction.* John Wiley and Sons.

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

Crowley P., (2007), "A guide to wavelets for economists", *Journal of*

Economic Surveys, 21 (2), 207-267.

deBoor, C. (2001) *A Practical Guide to Splines.* Springer

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", *Journal of the American Statistical Association, 90*, 1200-1224.

Fan, Y. and Gencay, R., (2007) "Unit root and cointegration tests with wavelets" Working Paper.

Gallegati, Marco, Gallegati, Mauro, Ramsey, J.B. and Semmler, W. (2009), "The US wage Phillips curve over different time horizons", Giornale degli Economisti e Annali di Economia, 68 (2), 113-148.

Gallegati, Marco, Gallegati, Mauro, Ramsey, J.B. and Semmler, W. (2011), "The US wage Phillips curve across frequencies and over time", *Oxford Bulletin of Economics and Statistics*,

Gallegati,M., Ramsey, J.G. and Semmler, W. (2011a) "Forecasting output using interest rate spreads: a wavelet decomposition analysis", Working Paper.

Gallegati Marco and J.B. Ramsey (2011), "Testing the (in)stability of the q-relationship:", *mimeo*

Gençay, R., Selçuk, F., and Whitcher, B., (2002), *An Introduction to Wavelets and Other Filtering Methods in Finance and Economics*, Academic Press, San Diego.

Gençay, R. and N. Gradojevic, (2009) "Errors-in-Variables Estimation with No Instruments," Working Paper Series 30-09, Rimini Centre for Economic Analysis.

Greenblatt, S.A., (1996), "Wavelets in econometrics: an application to outlier testing", in Gilli, M. (eds.), *Computational economic systems: Models, methods, and econometrics, advances in computational economics*, 5, 139 160, Kluwer Academic Publishers, Dordrecht.

Johnstone, I.M.(2000) "Wavelets and the theory of non-parametric function estimation". *Wavelets,* Edts: Silverman, B.W. and Vassilicos, J.C. 89-110. Oxford University Press

Kendall, M.G. and Stuart, A. (1961) *The Advanced Theory of Statistics,Vol 2.* Charles Griffin and Co.

Nason G.P. (2008), *Wavelet methods in statistics with R*, Springer.

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

Ramsey J.B. and Zhang, Z. (1996), "The application of waveform dictionaries to stock market index data", in *Predictability of Complex Dynamical Systems,* edited by Kravtsov, Y.A. and Kadtke, J. Springer-Verlag.

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

Ramsey, J.B. and Lampart, C. (1998a), "The decomposition of economic relationship by time scale using wavelets: money and income", *Macroeconomic Dynamics and Econometrics*, 2, 49-71.

Ramsey, J.B. and Lampart, C. (1998b), "The decomposition of economic relationship by time scale using wavelets: expenditure and income", *Studies in Nonlinear Dynamics and Econometrics*, Vol. 3, No. 4, 23-42.

Ramsey, J.B. (2002), "Wavelets in economics and finance: Past and future", *Studies in Nonlinear Dynamics & Econometrics*, vol.6 (3), 1-29.

Ramsey, J.B., Gallegati, Marco, Gallegati, Mauro, and W. Semmler, (2010), "Instrumental variables and wavelet decomposition", *Economic Modeling*, 27 (6), 1498-1513.

Rua, A. and Nunes, L. C. (2009) "International comovement of stock market returns: a wavelet analysis." Journal of Empirical Finance 16, 632-639.

Rua, A. (2010) "Measuring comovement in the time-frequency space." Journal of Macroeconomics 32, 685-691.

Samia, M., Dalenda, M., Saoussen, A., (2009) "Accuracy and conservatism of VaR models: a wavelet decomposed VaR approach versus standard ARMA-GARCH method" International Journal of Economics and Finance. Vol 1, No. 2 174-184.

Silverman, B.W. (2000) "Wavelets in statistics: beyond the standard assumptions." Edts: Silverman, B.W. and Vassilicos, J.C., 71-88. Oxford University Press

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

Wei, Z., Bo, L., Tao, Z.X., Xiong, X., Yue, K. (2006) "Application of the wavelet based multi-fractal for outlier detection in financial high-frequency time series data", IEEE.

Whitcher, B.J, (1998), *Assessing Nonstationary Time Series Using Wavelets*, PhD Thesis, University of Washington.

Whitcher, B.J, Byers, S.D., Guttorp, P. and D.B. Percival (2002), "Testing for Homogeneity of Variance in Time Series: Long Memory, Wavelets and the Nile River", *Water Resources Research*, 38.

Yousefi, S., Weinreich, I., Reinarz, D. (2005) "Wavelet-based prediction of oil prices", Chaos, Solitons and Fractals 25. 265-275.