Wavelets in Econometrics:
An Application to Outlier Testing

Seth A. Greenblatt*

University of Reading

15th August 1994

ABSTRACT. In recent years, wavelets have become widely used in physics, engineering, and mathematics. They have been used for signal processing, image processing, numerical computation, and data compression. Wavelets have not, however, been used very much in the fields of Economics, Econometrics, and Finance. In this study, We will look at the wavelet transform in the context of multiresolution analysis, discuss its uses in other fields, and present an Econometric application of wavelets to outlier detection.

---

*Postal address for correspondence:
DR. SETH A. GREENBLATT
CENTRE FOR QUANTITATIVE ECONOMICS AND COMPUTING
DEPARTMENT OF ECONOMICS, FACULTY OF LETTERS AND SOCIAL SCIENCES
UNIVERSITY OF READING
PO BOX 218, WHITEKNIGHTS
READING, BERKSHIRE, RG6 2AA, U.K.
E-mail: S.A.Greenblatt@rdg.ac.uk

1. Introduction

Over the past ten years or so, wavelets have become very popular in a wide variety of fields. These fields include engineering, geology, mathematics, meteorology, and signal processing. Economics, econometrics, and finance have been slow to become involved in this rapidly evolving area[1].

Wavelets give us the opportunity to examine data, or functions, on a variety of scales. They also allow us to focus in on local features of a data series in situations where traditional approaches, such as fourier analysis, smear out some of the local detail. Wavelets give us a unifying framework in which to investigate a variety of analyses that appear to be quite different. One such type of analysis is multiresolution analysis, which allows us to take a "microscope" to our data and look at the behaviour at a number of "magnifications".

The remainder of this study is broken up into five main sections. Section 2 will discuss multiresolution analysis. Section 3 will describe the wavelet transform, looking at construction of wavelets and presenting a simple example. The next section will present examples of the application of wavelet analysis to problems

---

I would like to thank seminar participants at the Universities of Reading and Manchester and the IFAC Workshop on Computational Methods in Economics and Finance for constructive comments on an earlier version of this paper. Any errors remaining are my own.

[1] One exception to this is Goffe(1993), a descriptive study using wavelets to look at turning points in macroeconomic data series.

in a variety of fields. Section 5 will present an econometric application of wavelet analysis. Finally, Section 6 will present conclusions and comment upon ongoing and upcoming research.

1.1. Notation. If we allow $\mathbb{R}$ to represent the real numbers and $\mathbb{Z}$ to represent the integers, then we can define $\mathbf{L}^2(\mathbb{R})$ as the vector space of measurable, square-integrable one-dimensional functions $\mathrm{f}(x)$. We can further define the inner product of two functions $\mathrm{f}(x), \mathrm{g}(x) \in \mathbf{L}^2(\mathbb{R})$ as:

$$\langle \mathrm{g}(u), \mathrm{f}(u) \rangle = \int_{-\infty}^{+\infty} \mathrm{g}(u)\, \mathrm{f}(u)\, du.$$

The norm of $\mathrm{f}(x) \in \mathbf{L}^2(\mathbb{R})$ is:

$$\|\mathrm{f}\|^2 = \int_{-\infty}^{+\infty} |\,\mathrm{f}(u)|^2\, du.$$

The convolution of $\mathrm{f}(x), \mathrm{g}(x) \in \mathbf{L}^2(\mathbb{R})$ will be represented by:

$$\mathrm{f} * \mathrm{g}(x) = \int_{-\infty}^{+\infty} \mathrm{f}(u)\, \mathrm{g}(x-u)\, du.$$

The Fourier transform of $\mathrm{f}(x) \in \mathbf{L}^2(\mathbb{R})$ is defined to be:

$$\hat{\mathrm{f}}(\omega) = \int_{-\infty}^{+\infty} \mathrm{f}(x)\, e^{-i\omega x}\, dx.$$

We define $\ell^2(\mathbb{Z})$ as the vector space of square-summable sequences as follows:

$$\ell^2(\mathbb{Z}) = \left\{ (\alpha_i)_{i\in\mathbb{Z}} : \sum_{i=-\infty}^{+\infty} |\alpha_i|^2 < \infty \right\}.$$

2. MULTIRESOLUTION ANALYSIS

A particularly useful way of looking at wavelet decompositions are as multiresolution analyses. In general terms, a multiresolution analysis allows us to look at data or a function at a number of different resolutions. Using multiresolution analysis we can, in effect, take a "microscope" to the data. As with a real microscope, we can use different levels of "magnification" to look at the data at different scales. In this section, we will look at multiresolution analysis, presenting results that will help us to construct wavelets in Section 3.1, below.

In our presentation of multiresolution analysis[2], we follow the development of Mallat(1989b, 1989c). To discuss multiresolution analysis, we must first define an approximation operator A_{2^j} which approximates a signal $\mathrm{f}(x) \in \mathbf{L}^2(\mathbb{R})$ at resolution 2^j. We discuss the properties of this operator to better understand its workings.

- A_{2^j} is a linear projection operator. If we form an approximation $A_{2^j}\,\mathrm{f}(x)$

---

[2] Rioul(1993) presents a elegant approach to multiresolution analysis, as developed in a strictly discrete-time framework. While the approach is of interest in implementation of algorithms for discrete data series, its presentation is beyond the scope of this study.

of f(x) at resolution 2^j, then this approximation is unaffected if we apply the same approximation operator, A_{2^j}, again. In other words, $A_{2^j} \circ A_{2^j} = A_{2^j}$. So we see that A_{2^j} is a projection operator on the the vector space $\mathbf{V}_{2^j} \subset \mathbf{L}^2(\mathbb{R})$. Since we have not specified a particular functional form for f(x), we can see that $\mathbf{V}_{2^j}$ is the set of all possible approximations at resolution 2^j of functions in $\mathbf{L}^2(\mathbb{R})$.[3]

- The approximation that we form by applying A_{2^j} is the best (closest) approximation to the original signal that can be formed at resolution 2^j.

  $$\forall\, \mathrm{g}(x) \in \mathbf{V}_{2^j}, \qquad \|\mathrm{g}(x) - \mathrm{f}(x)\| \geq \|A_{2^j}\,\mathrm{f}(x) - \mathrm{f}(x)\|\,. \tag{1}$$

  This operator is an orthogonal projection on the vector space $\mathbf{V}_{2^j}$.[4]

- Since we lose information every time we go to a lower level of resolution, it is obvious that the approximation of a signal at resolution 2^{j+1} contains all information necessary to formulate an approximation of the same signal at lower resolution 2^j. We can restate this in the following manner[5]:

  $$\forall j \in \mathbb{Z}, \qquad \mathbf{V}_{2^j} \subset \mathbf{V}_{2^{j+1}}. \tag{2}$$

---

[3] See Mallat (1989c), p. 675.

[4] See Rudin (1970), p. 79 for a discussion of orthogonal projections.

[5] The interested reader that looks into the multiresolution literature should be wary of notation. Some studies state (1) as $\mathbf{V}_{2^j} \subset \mathbf{V}_{2^{j-1}}$.

- In a multiresolution analysis framework, the approximations at the various levels of resolution are related to one another in a particular way. The spaces of approximated functions are derived from one another by scaling the approximated functions by the ratio of the respective levels of resolution. We can state this property more concisely as follows:

$$\forall j \in \mathbb{Z}, \qquad \mathrm{f}(x) \in \mathbf{V}_{2^j} \iff \mathrm{f}(2x) \in \mathbf{V}_{2^{j+1}}. \tag{3}$$

- We may characterise the approximation $A_{2^j}\,\mathrm{f}(x)$ of the signal $\mathrm{f}(x)$ by 2^j sampled observations per unit length. If we translate $\mathrm{f}(x)$ by an amount proportional to 2^{-j} length units, then $A_{2^j}\,\mathrm{f}(x)$ is translated by the same amount. Because of the scaling relationship set out above, it will be sufficient for us to fully characterise the translation properties for the resolution $j = 0$ only. We may more succinctly describe the translation properties as follows:
    - We may discretely characterise the problem by stating that there exists an isomorphism $\mathbf{I}$ from $\mathbf{V}_1$[6] onto $\ell^2(\mathbb{Z})$.

---

[6] We are using $\mathbf{V}_1$ and A_1 instead of $\mathbf{V}_{2^0}$ and A_{2^0} to avoid cluttering of notation.

- We may describe a translation of the approximation by:

$$\forall k \in \mathbb{Z}, A_1 \mathrm{f}_k(x) = A_1 \mathrm{f}(x-k), \tag{4}$$

$$\text{where } \mathrm{f}_k(x) = f(x-k). \tag{5}$$

- We may now describe the corresponding translation of the sampled observations as follows:

$$\mathbf{I}(A_1 \mathrm{f}(x)) = (\alpha_i)_{i \in \mathbb{Z}} \iff \mathbf{I}(A_1 \mathrm{f}_k(x)) = (\alpha_{i-k})_{i \in \mathbb{Z}} \tag{6}$$

- As mentioned above, when we approximate at progressively lower resolutions, we lose progressively more information. If we take this to the lower limit where the resolution approaches zero, then the approximated signal contains less and less information until it converges to zero. This can be stated more concisely:

$$\lim_{j \to -\infty} \mathbf{V}_{2^j} = \bigcap_{j=-\infty}^{+\infty} \mathbf{V}_{2^j} = \{\mathbf{0}\} \tag{7}$$

We may also look at the problem as the resolution increases. As the resolution increases to $+\infty$, the approximated signal converges to the original signal. We may restate this as:

$$\lim_{j \to +\infty} \mathbf{V}_{2^j} = \bigcup_{j=-\infty}^{+\infty} \mathbf{V}_{2^j} \text{ is dense}^7 \text{ in } \mathbf{L}^2(\mathbb{R}). \tag{8}$$

Any set of vector spaces which satisfies the above properties may be considered a multiresolution approximation of $\mathbf{L}^2(\mathbb{R})$ and the associated set of operators, A_{2^j} can give an approximation of any function in $\mathbf{L}^2(\mathbb{R})$ at any resolution 2^j.

As we mentioned above, the operator A_{2^j} is an orthogonal projection on $\mathbf{V}_{2^j}$. We now present a theorem that shows that an orthonormal basis of $\mathbf{V}_{2^j}$ may be defined by dilating and translating a unique function $\phi(x)$. This is necessary in order to characterise numerically our orthogonal projection operator.

Theorem 1. [8] *Let* $(\mathbf{V}_{2^j})_{j \in \mathbb{Z}})$ *be a multiresolution approximation of* $\mathbf{L}^2(\mathbb{R})$*. There exists a unique function* $\phi(x) \in \mathbf{L}^2(\mathbb{R})$*, called a scaling function, such that if we set* $\phi_{2^j}(x) = 2^j \phi(2^j x)$ *for* $j \in \mathbb{Z}$*, (the dilation of* $\phi(x)$ *by* 2^j*), then* $(\sqrt{2^{-j}} \phi_{2^j}(x - 2^{-j} n))_{n \in \mathbb{Z}}$ *is an orthonormal basis of* $\mathbf{V}_{2^j}$*.*

We have characterised our scaling function, but we haven't yet discussed how

---

[7] A set $\mathbf{A}$ is said to dense in $\mathbf{B}$ if the closure of $\mathbf{A}$ contains $\mathbf{B}$. The closure $\bar{\mathbf{E}}$ of a set $\mathbf{E} \subset \mathbf{X}$ is the smallest closed set in $\mathbf{X}$ which contains $\mathbf{E}$ (Rudin 1970).

[8] *For proof of this theorem, the interested reader should see Mallat (1989b, 1989c)*

wavelets fit into this framework. There is a very simple and elegant way in which the entire framework falls into place. We discussed above how information is lost as we move to progressively lower levels of resolution. How might we characterise this information that is lost? We may consider this “lost” information as *detailed* information. As we proceed to approximations of lower resolution, we are actually removing level after level of detailed information. This detailed information is merely the difference in information between two levels of resolution.

We may represent this detailed information at resolution 2^j as the the orthogonal complement of $\mathbf{V}_{2^j}$ in $\mathbf{V}_{2^{j+1}}$. If we let $\mathbf{W}_{2^j}$ be the orthogonal complement[9], it has the following properties[10]:

---

[9] Mallat (1989b, 1989c) uses $\mathbf{O}_{2^j}$ for the orthogonal complement.

[10] We can see that the third of the following properties is true because: for $i < j, \quad \mathbf{W}_{2^i} \subset \mathbf{V}_{2^j} \perp \mathbf{W}_{2^j}$.

$$
\begin{aligned}
&\mathbf{W}_{2^j} \perp \mathbf{V}_{2^j}, \\
&\mathbf{W}_{2^j} \oplus \mathbf{V}_{2^j} = \mathbf{V}_{2^{j+1}}, \\
&\mathbf{W}_{2^i} \perp \mathbf{W}_{2^j} \quad \text{for } i \neq j,
\end{aligned}
\tag{9}
$$

where $\perp$ represents orthogonality,

$\oplus$ represents a direct (orthogonal) sum.

We also know that, since the lowest-level approximation is $\{\mathbf{0}\}$[11], we can build up any higher-level of approximation by simply adding the appropriate amount of detail back into the approximation. A more concise way of stating this is:

$$\mathbf{V}_{2^J} = \bigoplus_{j=-\infty}^{J-1} \mathbf{W}_{2^j}. \tag{10}$$

We saw above that $\lim_{J\to+\infty} \mathbf{V}_{2^J}$ is dense in $\mathbf{L}^2(\mathbb{R})$, so we can now see that the limit of this direct sum is also dense in $\mathbf{L}^2(\mathbb{R})$. As we saw above in the case of $\mathbf{V}_{2^j}$, in order for us to be able to numerically compute the orthogonal

---

[11] We can think of this level of approximation as containing *no* detail.

projection of an $\mathbf{L}^2(\mathbb{R})$ function $f(x)$ on $\mathbf{W}_{2^j}$, we must first find an orthonormal basis of $\mathbf{W}_{2^j}$. It can be shown (Mallat 1989b; Mallat 1989c) that the orthogonal wavelet $(\sqrt{2^{-j}}\psi_{2^j}(x - 2^{-j}n))_{n\in\mathbb{Z}}$ corresponding to our scaling function above is an orthonormal basis of $\mathbf{W}_{2^j}$. It then follows that $(\sqrt{2^{-j}}\psi_{2^j}(x - 2^{-j}n))_{(n,j)\in\mathbb{Z}^2}$ is an orthonormal basis of $\mathbf{L}^2(\mathbb{R})$.

This method of looking at wavelets is useful for several reasons. The first reason is computational. The multiresolution framework allows us to develop efficient numerical algorithms for wavelet decomposition and reconstruction. If we consider the original signal to be at resolution $2^0 = 1$, then we begin with the vector space $\mathbf{V}_1$. We know from above that $\mathbf{V}_1 = \mathbf{V}_{2^{-1}} \oplus \mathbf{W}_{2^{-1}} = \mathbf{V}_{\frac{1}{2}} \oplus \mathbf{W}_{\frac{1}{2}}$. We can accomplish this by dilating and translating the scaling function, $\phi(x)$, and the wavelet, $\psi(x)$[12]. In a ideal world with an infinite number of observations, we could carry out this decomposition as many times as we like, but since real-life data sets have a finite number of observations, we may only decompose the series until we can no longer split the observations in the approximation in half[13]. The reconstruction of the series is done by essentially reversing the process. We take

---

[12] Details of the decomposition and reconstruction algorithms may be found in Mallat (1989c) or in Daubechies (1992).

[13] Another problem with finite-length data sets arises at data boundaries, A discussion of this is beyond the scope of this study, but the interested reader should see Cohen, et al.(1993) or Auscher(1992).

the approximation at each level of resolution, add in the detailed information at that level, and we have an approximation at the next higher level of resolution.

The second reason that multiresolution analysis is useful has to do with the intuitive understanding of the decomposed series. If we merely decompose the series as far as is possible given the data, then multiresolution analysis does not give much additional insight beyond that of other ways of looking at the problem. After decomposing we have one observation of approximation (which may be considered a type of mean), and a number of levels of detailed information. If however we only partially decompose the series, then we can look at the series at various levels of resolution. This is an insightful way to put a microscope to the data. Different type of behaviour may become evident at different levels of resolution. We may look at trends, cycles, or extrema in the underlying data generating process. The various levels of detailed information are important, but so are the various levels of approximation. Oftentimes, greater intuitive insight may be gained by examining the approximations than by examining the corresponding detail.

3. A Brief Introduction to Wavelets

Generally, when we look at time series data, we look at it in one of two ways. We look at it in the time order in which it was observed, giving us an intuitive view of various sorts of time-based information. In this form, frequency content is present, but difficult to comprehend. Alternatively, we may compute the Fourier transform of the series, $\hat{\mathrm{f}}(\omega)$, to obtain the spectrum of the series. In this form, frequency content is easily understandable, but time content is hidden away in the phase of $\hat{\mathrm{f}}(\omega)$.

In many situations, such as nonstationarity, we would ideally like to have a combination of the two views of the series. It could often be useful to have a frequency decomposition of $\mathrm{f}(x)$[14] locally in time. Traditional Fourier analysis, working globally over the entire series, is incapable of separating out the local frequency components. Modified forms of the Fourier transform, known as windowed Fourier transforms, have been developed to address this issue, with mixed success[15].

The type of view of the data that we would like to have is known as a

---

[14] We may look at a time series as observation sampled from $\mathrm{f}(x)$, the data generating process for the series.

[15] See Daubechies(1990) for a comparison of wavelets and windowed Fourier transforms.

time frequency representation. An analogy that has been used in the literature (Daubechies 1993) for this type of representation is musical notation. When we read music, we see both the order in which notes should be played (time) and the pitch at which they should be played (frequency). When we perform a wavelet transform on f(x), we end up with a time frequency representation of f(x).

There are a wide variety of of types of wavelet that have been discussed in the literature. In this study, however, we will concentrate on one particular family of wavelets, compactly supported orthonormal wavelets (Daubechies 1988). We concentrate upon these because they have a number of useful properties. Since these wavelets are compactly supported, they can accurately represent local, nonperiodic features with a relatively small number of coefficients. This is in contrast to Fourier analysis, which needs a large number of coefficients to represent local, nonperiodic features.

This family of wavelets also has the property of preserving the $\mathbf{L}^2$ norm during decomposition and reconstruction . There are a number of other nice properties belonging to this family and they are well covered in the literature (Daubechies 1992).

In this section, we will discuss the construction of wavelets, using some of the results from Section 2 above and then we will construct the Haar wavelet as an example.

3.1. Construction of Wavelets. When we think about performing a wavelet transform on a data series, we can look at it as running the data through a smoothing (lowpass) filter. This is, in effect, computing a moving average of the data. The difference between this computation and that of any other moving average is that the weights are chosen in a very particular manner. This lowpass filter is known as a scaling function ($\phi(x)$). Convolving the data with $\phi(x)$ gives us an approximation of the original series, except with some (high frequency) detail filtered out.

If instead we wish to obtain the detailed information, then we must pass the data through a differencing (highpass) filter. This highpass filter is known as the wavelet ($\psi(x)$).

Scaling functions and the corresponding wavelets are defined by the following dilation equations:[16]

$$\begin{aligned}\phi(x) &= 2\sum_{k=0}^{N} c_k\phi(2x-k), & (11)\\ \psi(x) &= 2\sum_{k=N}^{0}(-1)^{k+1}c_k\phi(2x-k). & (12)\end{aligned}$$

where $N+1=p$ is the order of regularity of the wavelet. The higher the order of regularity of the wavelet, the smoother the wavelet is.

These equations are simply difference equations with a difference. The difference is that the equations are on two scales. The two-scale nature of the dilation equations makes them, in all but the most trivial of cases, very difficult to solve analytically. Fortunately, we do not need to solve the equations analytically. We will actually be working with functions that we do not know! This may seem difficult to grasp at first, and we must go into the frequency domain to see why this unusual fact is so.

We take the Fourier transform of (11) and we get:[17]

$$\hat{\phi}(\omega) = \sum_{k=0}^{N} c_k e^{-ik\omega/2}\hat{\phi}\left(\frac{\omega}{2}\right) = \mathrm{m}_0\left(\frac{\omega}{2}\right)\hat{\phi}\left(\frac{\omega}{2}\right), \qquad (13)$$

---

[16] In the literature dilation equations are variously called refinement equations and two-scale difference equations.

[17] Since period is inversely proportional to frequency, $2x$ in space or time becomes $\omega/2$ in frequency space.

where

$$\mathrm{m}_0(\omega) = \sum_{k=0}^{N} c_k e^{-ik\omega} \tag{14}$$

is a transfer function.

If we set $\omega = 0$, then we can see that $\mathrm{m}_0(0) = 1$. This is equivalent to $\sum c_k = 1$. If we introduce the normalisation[18] $\hat{\phi}(0) = \int \phi(x)\, dx = 1$, then we may use recursion to find the infinite product (Strang 1993):

$$\hat{\phi}(\omega) = \mathrm{m}_0\left(\frac{\omega}{2}\right)\hat{\phi}\left(\frac{\omega}{2}\right) = \mathrm{m}_0\left(\frac{\omega}{2}\right)\mathrm{m}_0\left(\frac{\omega}{4}\right)\hat{\phi}\left(\frac{\omega}{4}\right) = \cdots = \prod_{j=1}^{\infty}\mathrm{m}_0\left(\frac{\omega}{2^j}\right). \tag{15}$$

This infinite product contains neither $\phi(x)$ nor $\hat{\phi}(\omega)$ on the right-hand side because, as $2^{-j}\omega$ approaches zero, $\hat{\phi}(2^{-j}\omega)$ approaches one. Using this approach, we can achieve an asymptotic solution for the dilation equation, but even this is unnecessary for our purposes.

Still in the frequency domain, we can derive conditions on the c's. Once we have derived these conditions, we may construct compactly supported orthonormal wavelets of arbitrary regularity simply by solving for the c's from our conditions.

We may now derive the approximation, or moment, condition. If we define

---

[18] This normalisation is equivalent to the denseness of $\{\mathbf{V}_{2^j}\}_{j\in\mathbb{Z}}$ in $\mathbf{L}^2(\mathbb{R})$ as shown in Section 2, above.

$h = 2^{-j}$ as the translation step (mesh width) of the local wavelet functions $\psi(2^j x)$, then we have a scale upon which to measure error in the wavelet approximation (Strang 1989; Strang 1993). This error is of the order h^p, where p refers to the degree of regularity of the wavelet. If we wish to increase the order of accuracy by one (p increases by one), then we must use an additional two coefficients (c's), giving us correspondingly wider (less compact) support.

It can be shown[19] that the Fourier transform $\hat{\phi}(\omega)$ must have zeroes of order p at all points $\omega = 2\pi n$[20]. We may now state this as a condition on $\mathrm{m}_0(\omega)$. As we saw above in equation (15), we may consider $\hat{\phi}(\omega)$ as an infinite product of $\mathrm{m}_0(2^{-j}\omega)$. If we consider all of the values of ω that are integer multiples of 2π, then we can see that we always find factors in the infinite product equal to $\mathrm{m}_0(\pi)$. For example, $\omega = 2\pi$ gives us the first factor as $\mathrm{m}_0(\pi)$. For $\omega = 4\pi$, the second factor is $\mathrm{m}_0(\pi)$. All other appropriate values of ω will reduce down to these two cases because $\mathrm{m}_0(\omega)$ is periodic of period 2π. Since this is a product, a zero value for any factor will produce a zero of $\hat{\phi}(\omega)$.

Since we can say that $\mathrm{m}_0(\omega) = \sum c_k e^{ik\omega}$ has a zero of order p at $\omega = \pi$, we may say that $\mathrm{m}_0^{(m)}(\pi) = 0$[21] for $m = 0, 1, \ldots, p-1$. This is equivalent to saying

---

[19] See Strang(1993) for references and a more complete discussion.

[20] The exception to this rule is at $\omega = 0$ where we find $\hat{\phi}(0) = 1$.

[21] We use the notation $\mathrm{m}_0^{(m)}(\pi)$ for the mth derivative of m_0, evaluated at π.

that the moment condition $\int x^m \psi(x)\, dx = 0$ is satisfied. If we substitute $\omega = \pi$ into $\mathrm{m}_0(\omega)$ and its derivatives of the appropriate orders, then we can arrive at the approximation (moment) condition on the c's as follows:

$$\sum (-1)^k k^m c_k = 0 \qquad m = 0, 1, \ldots, p-1 \tag{16}$$

The other important condition that we must derive is the orthogonality condition. As we showed above in Section 2, $(\phi(x-k))_{k \in \mathbb{Z}}$ is an orthonormal basis of $\mathbf{V}_{2^0}$. This, together with the Poisson summation formula leads to:

$$\sum_{n \in \mathbb{Z}} \left| \hat{\phi}\,(\omega + 2n\pi) \right|^2 = 1. \tag{17}$$

$$= \sum_n \left| \mathrm{m}_0 \left(\frac{\omega}{2} + n\pi \right) \right|^2 \left| \hat{\phi} \left(\frac{\omega}{2} + n\pi \right) \right|^2,$$

splitting into even and odd n,

$$= \sum_k \left| \mathrm{m}_0 \left(\frac{\omega}{2} + 2k\pi \right) \right|^2 \left| \hat{\phi} \left(\frac{\omega}{2} + 2k\pi \right) \right|^2$$

$$+ \sum_k \left| \mathrm{m}_0 \left(\frac{\omega}{2} + (2k+1)\pi \right) \right|^2 \left| \hat{\phi} \left(\frac{\omega}{2} + (2k+1)\pi \right) \right|^2,$$

by periodicity of m_0,

$$= \left| \mathrm{m}_0 \left(\frac{\omega}{2} \right) \right|^2 \sum_k \left| \hat{\phi} \left(\frac{\omega}{2} + 2k\pi \right) \right|^2 + \left| \mathrm{m}_0 \left(\frac{\omega}{2} + \pi \right) \right|^2 \sum_k \left| \hat{\phi} \left(\left(\frac{\omega}{2} + \pi \right) + 2k\pi \right) \right|^2,$$

letting $\omega/2 = \zeta$, by (17),

$$= \left| \mathrm{m}_0 \left(\zeta \right) \right|^2 + \left| \mathrm{m}_0 \left(\zeta + \pi \right) \right|^2 = 1. \tag{18}$$

Using this result, equation (14), and normalising, we see that the orthogonality condition on the c's may be expressed as:

$$\sum_{k=0}^{N} c_k c_{k+2m} = \delta_{m,0}, \qquad m = 0, 1, \ldots, p-1, \tag{19}$$

where $\delta_{m,0} = 1$ for $m = 0$ and 0 otherwise.

3.2. A Simple Example: The Haar Wavelet. In Section 3.1 above, we derived the approximation (moment) and orthogonality conditions that must be satisfied by the c's in order for us to create orthonormal wavelets with compact support and arbitrary order of regularity. As an example, we shall construct the simplest, oldest wavelet that is known, the Haar wavelet. This wavelet is not regular at all, but the technique that we shall use may be used equally well for smoother, more regular wavelets. The more regular the wavelet that we wish to construct, the more c's we must solve for, and hence more involved the computations.

For the Haar wavelet, we let $p = 1$. We take our approximation condition (16) and our orthogonality condition (19) and evaluate them for the $p = 1$ case. The conditions that we must solve for the Haar wavelet coefficients are:

$$c_0 - c_1 = 0 \tag{20}$$

$$c_0^2 + c_1^2 = 1 \tag{21}$$

As we can see from (20), $c_0 = c_1$. Substituting this result into (21), we find that $c_0 = c_1 = 1/\sqrt{2}$[22].

We can see in Figure 1 what the Haar wavelet looks like[23]. As we mentioned above, this wavelet is not at all regular, and we can see here that it even takes

---

[22] The astute reader will have noticed that the sum of the c's is $\sqrt{2}$, not 1 as stated above. We must normalise the c's in this way to allow us to use the same c's for both decomposition *and* reconstruction

[23] All of the figures in this paper were generated by Guy Nason's Wavethresh package in Splus. This software package is discussed in Nason(1993).

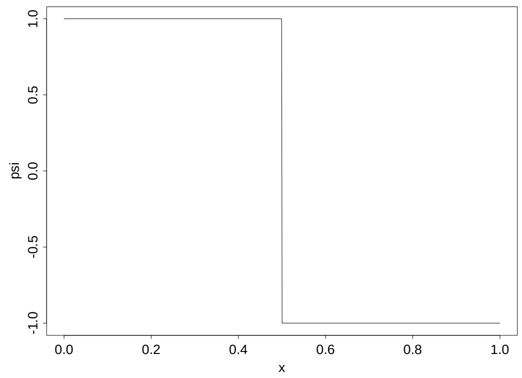

Figure 1. *Haar Wavelet with $p = 1$*

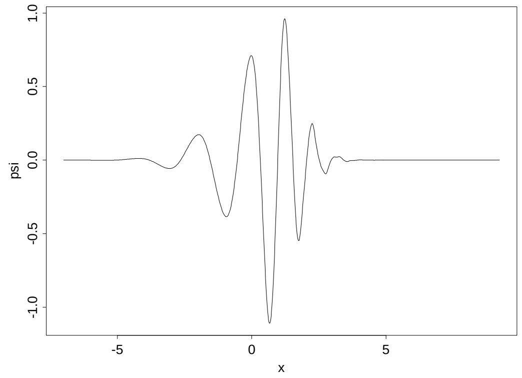

Figure 2. *Compactly supported orthonormal wavelet with $p = 8$*

discontinuous jumps. In contrast, we can see from Figure 2 that a wavelet constructed in the same manner as the Haar wavelet can be much smoother and more regular.

4. Applications of Wavelets to Problems from Other Fields

There are a number of areas where wavelets have been successfully applied. Even though these areas are not all directly applicable to econometrics, they are interesting and they provide useful ideas that *can* be fruitfully applied to econometrics.

4.1. Image Compression. A tremendous literature has developed on the use of wavelets for image processing (Mallat 1989c; Mallat 1989a). Much of this literature describes techniques for recognising various features of images, such as edges, or for noise removal. The area that we will discuss here is the compression of images.

For simplicity, we shall only discuss black-and-white (greyscale) images. A image can thought of as a two-dimensional array (a matrix). Each picture element (pixel) is represented by an element of that matrix. The value of an element represents the intensity, on a scale from white to black, of that particular pixel.

All of our discussion of wavelets thus far has been in a one-dimensional context. Matrices, however, are two-dimensional objects. Fortunately, wavelets can be extended to the two dimensional case. Unfortunately, it is beyond the scale of the present study to discuss the two-dimensional wavelet transform in any detail. There is a paper forthcoming from the author that discusses multidimensional wavelets in detail.

We transform the image using the two-dimensional wavelet transform and we end up with a matrix composed of wavelet coefficients. Just as the vector of one-dimensional wavelet coefficients is partitioned into subvectors by level of detail, the matrix of two-dimensional wavelet coefficients is partitioned into submatrices in a similar manner. Many of these wavelet coefficients are very small. The larger coefficients represent much of the sharp detail in the image. In fact, so much of the important information is contained in a small number of large coefficients that we can set the small coefficients to exactly zero.

The number of coefficients that we should set to zero may be determined in a number of ways. We may wish to set all coefficients of a magnitude smaller than a prespecified threshold to zero. Alternatively, we could zero out the smallest 90% or 95% of all coefficients. In either case, we could end up with a vast majority of the coefficients set to zero. In tests, nearly 95% of the coefficients in a transformed image have been set to zero without losing important detail.

The primary reason for the striking proportion of coefficients containing very little information is the spatial localisation of wavelets. Since the wavelets are localised in space, they are able to efficiently represent local detail in a small number of coefficients. We can contrast this with the use of Fourier transforms. Fourier transforms are *not* localised in space, so they need a large number of coefficients to represent local, non-periodic features. When a large number of Fourier coefficients are set to zero, local features such as edges are smeared out.

Once we have set the small coefficients to zero, we can then put the matrix of wavelet coefficients into a sparse matrix representation. In a sparse representation, we need only store the nonzero coefficients. This allows us to store an image in a small proportion of the original space. When we wish to display the image, we need only perform an inverse wavelet transform to retrieve an image with much of the detail information intact.

4.2. Fast Numerical Linear Algebra. The use of wavelets for fast numerical linear algebra (Alpert 1992; Beylkin, Coifman, and Rokhlin 1991; Beylkin 1993) is similar in spirit to the technique described above for image compression. This technique, however, is more directly applicable to problems in econometrics, where there frequently are large-scale numerical computations involving matrices.

Traditionally, techniques for multiplication or inversion of matrices have attempted to take advantage of the specific structure of the matrices at hand. If the matrices were diagonal, banded, symmetric, or positive definite we would have efficient techniques at our disposal for efficient multiplication or inversion. If the matrices did not have these characteristics, but satisfied certain other conditions, we could cleverly decompose them and take advantage of the new, decomposed structure. If we had no specific structure there were things that we could do to make things a bit more efficient, but not much.

With the wavelet transform, we can take a dense matrix of arbitrary structure and transform it into a matrix of wavelets coefficients, just as we did in the image compression example, above. Again, much of the information is concentrated in a small number of coefficients. A large proportion of the wavelet coefficients are very close to zero. If we set the small coefficients equal to zero, then we have a sparse[24] matrix. Under certain fairly weak conditions, the resulting matrices are not only sparse, but sparse with a special structure. There are a wide variety of fast algorithms that have been designed to deal with this sort of problem.

The question of accuracy must also be considered. Intuitively, it would seem that setting a number of coefficients to zero would introduce serious errors into the computation. In reality, numerical errors are a problem in standard approaches to multiplying and inverting matrices. In fact, for progressively larger numbers of arithmetic operations performed, we get progressively more severe errors. This is where the wavelet approach improves on standard techniques. Since there are drastically fewer elements containing information in the transformed matrix, there are correspondingly fewer arithmetic operations performed. These techniques have been found to be very accurate and very fast compared to standard approaches. The wavelet-based techniques actually make great gains in accuracy *and* speed as compared to standard approaches as the problem size becomes larger.

---

[24] We consider a matrix sparse if it has a large proportion of its elements equal to zero. This is the opposite of a dense matrix.

4.3. Denoising and Smoothing. Recently, there has been a rapidly-growing literature on the use of wavelets for denoising and smoothing (Donoho 1992; Donoho, Johnstone, Kerkyacharian, and Picard 1994a; Donoho and Johnstone 1994; Donoho, Johnstone, Kerkyacharian, and Picard 1994b; Donoho, Johnstone, Kerkyacharian, and Picard 1993; Donoho and Johnstone 1993). Wavelets are well-suited to this type problem because of their properties of time (or space) localisation. When approximating a signal, wavelets can preserve local features (discontinuities, turning points, etc) while still removing noise.

A technique particularly popular in the literature is that of wavelet shrinkage. In this technique, a time series is transformed using a wavelet transform. The standard deviation of the empirical wavelet coefficients is estimated, and this is used to determine a threshold and a shrinkage amount. All wavelet coefficients of magnitude less than the threshold are set to zero. All coefficients larger than the threshold are shrunk toward zero by the amount computed above. Finally, the inverse wavelet transfer is performed to recover the smoothed, denoised signal.

The orthonormality of the wavelet transform tells us that gaussian white noise in the original signal remains gaussian and white. It does, however, get spread over all of the wavelet coefficients. This is why we must shrink coefficients in addition to simple setting the small ones to zero.

4.4. Solution of Differential Equations. Many researchers have seen the advantages of using wavelets in the solution of differential equations[25]. Both ordinary and partial differential equations can be solved using these techniques and we shall attempt to present the general ideas of these techniques without going into technical detail. These techniques could be particularly useful in the numerical analysis of problems in continuous-time finance.

A Galerkin method allows us, after the choice of appropriate basis functions, to approximate the differential equation as system of linear equations, which we may then solve. In a number of areas of science and engineering, the finite element methods is used to choose the basis functions. In the wavelet approach, we use wavelets or functions derived in some way from wavelets as the basis functions.

The matrix that defines this linear system is known as a *stiffness matrix.* In many applications, this matrix can be ill-conditioned. This can be avoided using the wavelet basis, because the stiffness matrix derived in this case may be easily preconditioned to guarantee a uniformly bounded condition number. This makes for a better-behaved and numerically easier to solve system.

---

[25] See Beylkin, et.al.(1991) and Devore and Lucier(1992) for example.

5. AN ECONOMETRIC APPLICATION OF WAVELETS: OUTLIER TESTING

Recently, wavelet-based techniques have been finding their way more and more into the statistics literature. Work has been done in a number of different areas of statistics, such as smoothing and nonparametric regression (Donoho, Johnstone, Kerkyacharian, and Picard 1994a; Donoho and Johnstone 1994; Nason 1994) and long-memory processes (Percival 1993; Percival and Guttorp 1994; McCoy and Walden 1994).

This section is intended to look into an area somewhat different than those specified above. That area is diagnostic testing. Wavelets, when viewed in the context of multiresolution analysis (Mallat 1989c; Mallat 1989b; Rioul 1993) seem to be a natural approach to outlier detection.

Outliers are, by their very nature, localised high-frequency phenomena. If they occurred as anything other than isolated aberrations in the data, they would be considered part of the signal's structure, not an outlier.

Much of the literature on outlier testing (Barnett and Lewis 1994; Hawkins 1980) primarily looks at the problem as one of determining which observations come from the data generating distribution and which ones come from some contaminating distribution. While this is a very good approach that has a great deal of theoretical appeal, in practice it is often difficult and data requirements can be high.

This section looks at the problem in a different way. If, rather than asking a statistician, we were to ask an economist or other scientist that uses statistics what he would consider to be an outlier, he would probably reply that an outlier is an observation that appears to be out of place. It is this intuitive explanation that we use in this study. Along these lines, order statistics have been used to detect outliers (David 1981). Usually these approaches look at the observation with the largest deviation from the mean in a given sample.

In this section we look at the application of an order statistic outlier test to the most detailed level of wavelet coefficient computed using the Haar wavelet. The Haar wavelet is used because it has good local properties and this application has no need for the greater regularity of smoother wavelets. Finally, we present Monte Carlo results comparing the test as applied to data and to detailed wavelet coefficients.

5.1. A Brief Review of Order Statistics. Order statistics have been applied to a wide variety of applications, including robust statistical testing, the statistical theory of reliability, and most importantly for the current study, outlier testing.

Since the literature is immense and rather well developed, we shall only present the briefest of discussions on the use of order statistics[26].

Given random variables $X_1, X_2, \cdots, X_n$, we can sort them into ascending order of magnitude. We may then refer to these variables as

$$X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(n)}$$

where $X_{(i)}$ is known as the ith order statistic.

For our purposes, we can assume that the unordered X_i are independent and identically distributed (*i.i.d.*). Clearly, the ordered $X_{(i)}$ must be dependent because they fulfill the inequality relationship presented above. This special form of dependence causes the distribution theory of order statistics to become nontrivial[27].

Once we have sorted the variables, there are number of things we can look at. We might be interested in the extremes ($X_{(1)}$ and $X_{(n)}$), the range ($W = X_{(n)} - X_{(1)}$), the extreme deviate from the sample mean ($X_{(n)} - \bar{X}$), or any one of a number of statistics based upon these.

---

[26] Interested readers should consult David(1981) for a more detailed and thorough treatment of order statistics

[27] We shall not concern ourselves with distribution theory here. This area is well-addressed by David(1981) and the 1000 references contained therein.

For many of these statistics (depending upon the underlying distribution of the unordered X_i) asymptotic distribution theory and quantiles have been derived. In cases where this information is not available, and when we wish to estimate the finite sample properties of the statistics, we may wish to use Monte Carlo analysis. This is the approach taken in this study.

5.2. The Outlier Test and Monte Carlo Results. The outlier test presented in this study is based upon one familiar from the literature[28]. The difference is that this type of test is normally performed directly upon data. In this section, we will present a test that is performed, not on data, but instead on a subset of the coefficients computed by performing a wavelet transform on the data.

We take our data series and perform a wavelet transform upon it, using the Haar wavelet as our base wavelet. We assume that most, if not all, of the highest-frequency component of the data is either noise or outliers. Keeping this in mind, we need only consider the highest-frequency wavelet coefficients. These coefficients represent the detailed information removed from the data at the first level of wavelet decomposition.

An intuitive way of picturing this level of wavelet coefficient is to picture the

---

[28] See D_2 in David(1981), pg. 221.

true data generating process as a rope lying on a table. It meanders in one direction and then another. If this rope is frayed, then the bristles sticking out from the rope can be considered the noise in the series. If there is a part of the rope where a strand has broken and is sticking out a bit further than the bristles, then we have an outlier. If we then take this rope and pull it straight, then all of the undulations are gone and all that are left are noise bristles and outlier strands. This highest level subset of the wavelet coefficients may be pictured as the rope, pulled straight. The only features that we can see are the noise and any outliers that might be present.

If we continue with our definition that an outlier is an observation that is “out of place”, then we wish to find the coefficient that the furthest away from average. This, after normalising by estimated standard deviation, may be stated as:

$$\max \frac{|w_i - \bar{w}|}{s} \tag{22}$$

where w_i = the ith first-level wavelet coefficient

$\bar{w}$ = the mean of the first-level wavelet coefficients

s = the est. standard deviation of the first-level wavelet coefficients

This order statistic is the same as D_2 on pg. 221 of David(1981) with a minor difference. This statistic is calculated using w, the highest-level wavelet coefficients and D_2 is calculated using x, the actual data.

We performed some Monte Carlo experiments comparing the test statistic as applied to data and as applied to the highest level wavelet coefficients. First we determined the empirical size of the tests by Monte Carlo simulation with $1,000,000$ replications. Having thus determined the 95% and 99% critical values[29], we then proceeded to again generate data, this time introducing one outlier per replication[30].

---

[29] Critical values are available upon request from the author.

[30] Our data generating process for the power tests was $y = \sin(x) + \sin(2x) + \log(1 + x) + \epsilon$ with x varying from 0 to 30π and $\epsilon \sim$ *i.i.d.* N(0, 1). This process was chosen to exhibit trended,

Sample size: 1024

λ	2	3	4	5	6	7
Wavelet(%)	39	73	93	99	99	99
Data(%)	14	41	70	91	99	100

Sample size: 512

λ	2	3	4	5	6	7
Wavelet(%)	40	73	93	99	99	99
Data(%)	15	40	70	91	99	100

Sample size: 256

λ	2	3	4	5	6	7
Wavelet(%)	41	71	92	98	99	99
Data(%)	15	41	71	91	99	100

TABLE 1. *Power of Tests – 99%*

To generate outliers, we found the observation whose error had the largest absolute value and transformed that observation in the following way:

---

cyclical behaviour as many economic series do.

Sample size: 1024

λ	2	3	4	5	6	7
Wavelet(%)	62	88	98	99	99	99
Data(%)	28	58	84	97	99	100

Sample size: 512

λ	2	3	4	5	6	7
Wavelet(%)	64	88	98	99	99	99
Data(%)	28	59	85	97	100	100

Sample size: 256

λ	2	3	4	5	6	7
Wavelet(%)	65	88	97	99	99	99
Data(%)	30	60	85	97	99	100

TABLE 2. *Power of Tests – 95%*

$$x_{maxerr} = x_{maxerr} + \text{signum}(\epsilon_{maxerr} - \mu) * \sigma * \lambda \tag{23}$$

where x_{maxerr} = the observation where the largest error occurred,

ϵ_{maxerr} = the largest error,

μ = the mean of the error generating normal distribution,

σ = the standard deviation of the same distribution,

λ = a multiplier determining the magnitude of the outlier.

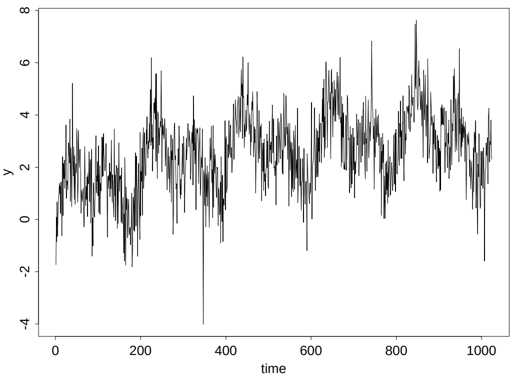

FIGURE 3. *Noisy generated data with outlier at observation 347*

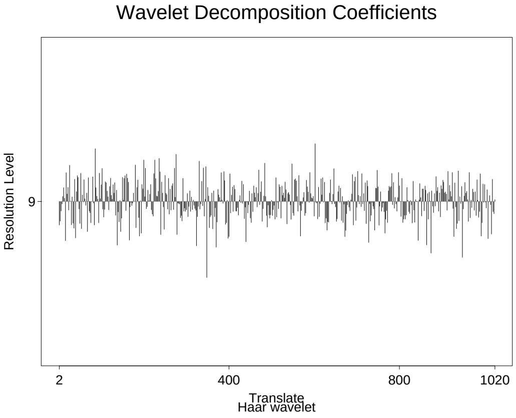

FIGURE 4. *Highest level wavelet coefficients with outlier at observation 347*

Transforming the "outlier" observation in this way allows an observation with an error much larger(smaller) than the mean of the generating distribution to become a positive(negative) outlier[31]. Tables 1 and 2 present the results for the empirical power tests for sample sizes 1024, 512, and 256[32]. As we can see from the tables, the test, as performed on the wavelets, is more powerful on less obvious outliers than the test performed on the data. For outliers obviously far removed from the rest of the data, there is little to distinguish between the two tests.

We can see in Figures 3 and 4 the two representations of the noisy data with an

---

[31] The signum() function returns a -1 for a negative argument, a +1 for a positive argument, and 0 for a zero argument.

[32] We performed the Monte Carlo simulations for the power tests with 10,000, 100,000, and 1,000,000 replications and compared the results. In these cases we found very little sensitivity to the number of replications beyond 10,000, so reported results are for 10,000 replications.

outlier. In Figure 3, we can see the entire noisy data series with a large negative outlier at observation 347 out of 1024. Figure 4 shows the same information as the set of highest-level wavelet coefficients. We used a very large outlier as an example to make it obviously visible in both representations. The outlier shows up in both cases as downward spikes, but since the data moves over a number of different levels, the statistical significance of the outlier is sometimes overwhelmed. The wavelet representation, during our computer experiments, did not seem to suffer from this problem.

6. Conclusion and Directions for Future Research

In this study, we introduced wavelet analysis, making extensive use of the concept of multiresolution analysis. We then discussed examples of applications from fields outside of econometrics. Finally, we presented an econometric application of wavelets.

Early proponents of wavelets in areas such as geophysics and signal processing were so enthusiastic that they made claims that wavelets could do everything from solving a wide range of scientific problems to washing the family car. While I must admit to sharing their enthusiasm, it has become clear that some applications are well-suited to analysis by wavelets, and some are not. While the econometric application (outlier testing) presented in this study is a particularly simple one, we can see the applicability of wavelet analysis to a large number of problems in econometrics, economics, and finance.

We are presently investigating the use of wavelets in the study of nonstationary time series, nonparametric regression, and evaluation of option prices. We are also looking into the use of multivariate wavelets to identify relationships and common features amongst data series. Wavelets look promising in a number of other areas where effective, efficient methods of approximation are needed.

The wavelet literature is large and rapidly growing. The present study does not claim to comprehensively cover the literature. It is, however, intended to present enough of the central concepts and applications to econometricians and economists to whet their appetites and motivate them to become familiar with those parts of the literature that might apply to their own research.

REFERENCES

Alpert, B. K., 1992, Wavelets and other bases for fast numerical linear algebra, See Chui (1992), pp. 181–216.

Auscher, P., 1992, Wavelets with boundary conditions on the interval, See Chui (1992), pp. 217–236.

Barnett, V. and T. Lewis, 1994, *Outliers in Statistical Data* (3rd ed.)., Wiley Series in Probability and Mathematical Statistics. (Wiley, Chichester).

Beylkin, G., 1993, Wavelets and fast numerical algorithms, In I. Daubechies (Ed.), *Different Perspectives on Wavelets*, Volume 47 of *Proceedings of Symposia in Applied Mathematics*, pp. 89–117.

Beylkin, G., R. Coifman, and V. Rokhlin, 1991, Fast wavelet transforms and numerical algorithms I, *Comm. Pure and Appl. Math. 44*, 141–183.

Chui, C. K. (Ed.), 1992, *Wavelets: A Tutorial in Theory and Applications*, (Academic Press, San Diego).

Cohen, A., I. Daubechies, and P. Vial, 1993, Wavelets on the interval and fast wavelet transforms, *ACHA 1*, 54–81.

Daubechies, I., 1988, Orthonormal bases of compactly supported wavelets, *Comm. Pure and Appl. Math. 41*, 909–996.

Daubechies, I., 1990, The wavelet transform, time-frequency localization and signal analysis, *IEEE Trans. on Inf. Theory 36*(5), 961–1005.

Daubechies, I., 1992, *Ten Lectures on Wavelets*, Number 61 in CBMS-NSF Series in Applied Mathematics. (SIAM, Philadelphia).

Daubechies, I., 1993, Wavelet transforms and orthonormal wavelet bases, In I. Daubechies (Ed.), *Different Perspectives on Wavelets*, Volume 47 of *Pro-*

ceedings of Symposia in Applied Mathematics, pp. 1–33.

David, H. A., 1981, *Order Statistics*, Wiley Series in Probability and Mathematical Statistics. (Wiley, New York).

DeVore, R. A. and B. J. Lucier, 1992, Wavelets, *Acta Numerica 1*, 1–56.

Donoho, D. L., 1992, De-noising by soft-thresholding, Technical Report 409, Stanford University, Department of Statistics.

Donoho, D. L. and I. M. Johnstone, 1993, April, Ideal spatial adaptation by wavelet shrinkage, Preprint, Department of Statistics, Stanford University.

Donoho, D. L. and I. M. Johnstone, 1994, Adapting to unknown smoothness by wavelet shrinkage, *Journal of the American Statistical Association*, accepted for publication.

Donoho, D. L., I. M. Johnstone, G. Kerkyacharian, and D. Picard, 1993, April, Density estimation by wavelet thresholding, Preprint, Department of Statistics, Stanford University.

Donoho, D. L., I. M. Johnstone, G. Kerkyacharian, and D. Picard, 1994a, Wavelet shrinkage: Asymptopia?, *Journal of the Royal Statistical Society, B*, (to appear).

Donoho, D. L., I. M. Johnstone, G. Kerkyacharian, and D. Picard, 1994b, April, Wavelet shrinkage: Asymptopia, Preprint, Department of Statistics, Stanford University.

Goffe, W. L., 1993, Wavelets in macroeconomics: An introduction, In D. A. Belsley (Ed.), *Computational Economics and Econometrics*, Volume 2. (Kluwer Academic Press, Amsterdam).

Hawkins, D. M., 1980, *Identification of Outliers*, Monographs on Applied Probability and Statistics. (Chapman and Hall, London).

Mallat, S. G., 1989a, Multifrequency channel decompositions of images and wavelet models, *IEEE Trans. on Acoust. Signal Speech Process. 37*(12), 2091–2110.

Mallat, S. G., 1989b, Multiresolution approximations and wavelet orthonormal bases of $\mathbf{L}^2(\mathbb{R})$, *Trans. Amer. Math. Soc. 315*(1), 69–87.

Mallat, S. G., 1989c, A theory for multiresolution signal decomposition: The wavelet representation, *IEEE Trans. on Patt. Anal. and Mach. Intell. 11*(7), 674–693.

McCoy, E. J. and A. T. Walden, 1994, May, Wavelets and long-memory processes, Technical Report TR-94-07, Imperial College of Science Technology, and Medicine, Statistics Section, Department of Mathematics.

Nason, G. P., 1993, March 26, The discrete wavelet transform in s, Version 2.1 - DRAFT, Department of Mathematics, University of Bristol.

Nason, G. P., 1994, Wavelet regression by cross-validation, Preprint, Department of Mathematics, University of Bristol.

Percival, D. B., 1993, October, An introduction to spectral analysis and wavelets, In *Proceedings of the Workshop on Advanced Mathematical Tools in Metrology, Torino, Italy.*

Percival, D. B. and P. Guttorp, 1994, Long-memory processes, the allen variance and wavelets, In E. Foufoula-Georgiou and P. Kumar (Eds.), *Wavelets in Geophysics*, Wavelet Analysis and Its Applications. (Academic Press, New York), (forthcoming).

Rioul, O., 1993, August, A discrete-time multiresolution theory, *IEEETSP 41*(8), 2591–2606.

Rudin, W., 1970, *Real and Complex Analysis* (International Student Edition ed.)., McGraw-Hill Series in Higher Mathematics. (McGraw-Hill, New York).

Strang, G., 1989, Wavelets and dilation equations: A brief introduction, *SIAM Review 31*(4), 614–627.

Strang, G., 1993, Wavelet transforms versus fourier transforms, *Bull. Amer. Math. Soc. 28*(2), 288–305.