



Research Report
Department of Statistics
Göteborg University
Sweden

**Some aspects of wavelet
analysis in time series**

Abdullah Almasri

**Research Report 2000:4
ISSN 0349-8034**

Mailing address:	Fax	Phone	Home Page:
Dept of Statistics	Nat: 031-773 12 74	Nat: 031-773 10 00	http://www.stat.gu.se/stat
P.O. Box 660	Int: +46 31 773 12 74	Int: +46 31 773 10 00	
SE 405 30 Göteborg			
Sweden			

Some Aspects of Wavelet Analysis in Time Series

Abdullah Almasri

Department of Statistics, Göteborg University, P.O. Box 660, SE-40530 Göteborg, Sweden

This thesis consists of two papers dealing with the methodology of wavelet and its application in time series.

We give here a brief summary of the contents of the two papers in the thesis.

The first paper describes an alternative approach for testing the existence of trend among time series. The test method has been constructed using wavelet analysis which have the ability to decompose a time series into low frequencies (trend) and high frequencies (noise) components. Under the normality assumption the test is distributed as F . However, the distribution of the test is unknown under other conditions, like non-normality. To investigate the properties of the test statistic under wide conditions, empirical critical values for the test have been generated using Monte Carlo simulations. The results are then compared with those results obtained by applying the OLS method for testing the trend.

A number of cases have been studied regarding the size of the test, where the number of observations, long memory parameter, different types of wavelets and the distribution of the errors have been varied. For each case 10,000 replications have been performed and four different nominal sizes have been studied. For the power calculations the strength of the trend parameter has been varied.

In this study, we find that using the tabulated critical values from the F distributions, the test tends to overreject under the null hypothesis, while when using generated critical values, the test performs satisfactorily. The Harr wavelet has shown to exhibit the highest power among the other wavelet's types, but the power is still lower than the OLS under some conditions.

The methodology here has been applied to real temperature data in Sweden for the period 1850-1999. The results indicate a significant increasing trend which agrees with the "Global warming" hypothesis during the last 100 years.

The second paper, (jointly written with Ghazi Shukur), presents an illustration of the use of wavelet analysis and the importance of time scale decomposition in determining the causality relation between two important macro variables. The relation between government spending and revenue has been studied using three methods, the conventional test, a bootstrap simulation approach and a multivariate Rao's F -test. These test methods have shown different results when using quarterly and monthly data. The wavelet decomposition of the time series into different time scales of variation helped in determining the causality direction between these two macro variables.

Acknowledgment

I would like to thank my supervisor, Associate Professor Ghazi Shukur who guided me and encouraged me during my work with this thesis. I would also like to thank Professor Marianne Frisén, Dr. Shakir Hussain, Erik Järpe, and all other colleagues at the Department of Statistics for fruitful discussions and comments on the draft manuscript.

Paper's titles

This thesis consists of two papers, which are referred to by their Roman numbers I and II.

I. An Alternative Approach for Testing the Significance of Trend in the Presence of Long Memory Process, Using Wavelet Analyses

II. An Illustration of the Causality Relation between Government Spending and Revenue Using Wavelets Analysis on Finish Data

An Alternative Approach for Testing the Significance of Trend in the Presence of Long Memory Process, Using Wavelets Analysis

Abdullah Almasri

Department of Statistics
University of Göteborg
P. O. Box 660
SE- 405 30 Göteborg
Sweden

ABSTRACT

This paper describes an alternative approach for testing the existence of trend among time series. The test method has been constructed using wavelet analysis which has the ability to decompose a time series into low frequencies (trend) and high frequencies (noise) components. Under the normality assumption the test is distributed as F . Using generated empirical critical values the properties of the test statistic have been investigated under different conditions and different types of wavelet. The results are then compared with those results obtained by applying the OLS method for testing the trend. The Harr wavelet has shown to exhibit the highest power among the other wavelet types, but the power is still lower than the OLS under some conditions.

In contrast with the OLS estimate the wavelet estimate has a localisation in time and frequency property which leaves the estimated trend to vary with time.

The methodology here has been applied to real temperature data in Sweden for the period 1850-1999. The results indicate a significant increasing trend which agrees with the "Global warming" hypothesis during the last 100 years.

1. INTRODUCTION

Most of the time series of aggregated variables exhibit a steadily increasing or decreasing pattern, known as a trend. For example, recent concerns about the possibility of climate change have focused attention on temperature series. A crucial question raised by these data is

whether the temperature rise of around 0.5 C^0 is the start of a systematic warming or if it is an effect of natural variability, for details, see Bloomfield and Nychka (1992). A plausible statistical model for such series can consist of a mean level, a possible trend plus a stationary time series. ($X_t = u + T_t + Z_t$), where X_t is the annual mean temperature, T_t is the trend due to human-made effects and Z_t is a (mean zero) random variable expressing the natural variation in global temperature. Now, for a stationary Gaussian process, the dependence among the random components can be described by the autocovariances, $E(Z_t Z_{t+h}) = \gamma(h)$. It is this dependence between successive values of Z_t that makes the identification of the T_t difficult.

The same is true for the availability of high frequency long time series from e.g. returns of speculative asset. Much research has been devoted to the study of long range behaviour of financial data. A common finding in much of the empirical literature is that the returns themselves contain little dependence, which is in agreement with the efficient market hypotheses.

However, to simplify the interpretation of variability of a trend estimate we assume in this paper that the trend (T_t) can be approximated by a polynomial linear trend. The ideas, are of course are not restricted to this parametric form.

The most difficult problem when testing for linear trend is the presence of dependence among the residuals. Because the residuals are dependent, tests for trend based on the classical ordinary least squares (OLS) regression are inappropriate. Although it is well known that this problem exists, estimation and testing for the existence of the trend is still done by the OLS.

Many economic and physical phenomena are modelled by so called long-memory processes. The autocovariance functions for such processes used to exhibit a slow decay. The wavelet analysis, however, has been extensively used for such purposes, since it suitably matches the structure of these processes. The autocovariance function of the wavelet transformed series exhibits different behaviour, in the sense that autocovariance functions of the transformed series decay hyperbolically fast at a rate much faster than the original process. In general, the series that are correlated in the time domain become almost uncorrelated in the wavelet domain, see McCoy and Walden (1996).

The wavelet estimate has the advantage of the localisation in time and frequency, which means that the detail of the estimate is seen to vary with t in contrast to the usual parametric

methods or kernel estimates. This property gives us an additional result of variability on different scales in the case that we have a long memory process.

In this study we use the wavelet analysis to construct a test statistic in order to test for the existence of trend in the series. Since we are also interested in the presence of the long-memory process among the data, we study the properties of our test statistic under a variety of such conditions. We will compare the results when using the wavelet analysis with results that are obtained by applying the OLS method under the same conditions. To demonstrate our testing approach seasonally adjusted monthly temperature data for Stockholm during the period 1850-1999 are modelled.

The paper is organised as follows: Section 1 gives an introduction. Section 2 presents the methodology. In Section 3 we introduce the wavelet based estimation and testing for the trend. Section 4 describes the Monte Carlo design we used, while Section 5 presents the estimated results. In Section 6, we present an empirical application. Finally, we give a short summary and conclusions in Section 7.

2. THE WAVELET METHODOLOGY

The wavelet transform has been expressed by Daubechies (1992) as “a tool that cuts up data or functions into different frequency components, and then studies each component with a resolution matched to its scale”. Thus, with wavelet transform one can analyse series with heterogeneous (unlike Fourier transform) or homogeneous information at each scale. Unlike the Fourier transform, which uses only sines and cosines as basis functions, the wavelet transform can use a variety of basis functions. We hereby will give a brief presentation of this decomposition methodology.

2.1 The Discrete Wavelet Transform

Let $\mathbf{X} = (X_0, X_2, \dots, X_{N-1})'$ be a column vector containing N observations of a real-valued time series, where we assume that N is an integer multiple of 2^M , where M is a positive integer. The discrete wavelet transform (DWT) of level J is an orthonormal transform of \mathbf{X} defined by

$$\mathbf{d} = (\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_j, \dots, \mathbf{d}_J, \mathbf{s}_J)' = \mathbf{W}\mathbf{X}, \quad (1)$$

where \mathbf{W} is an orthonormal $N \times N$ real-valued matrix, i.e., $\mathbf{W}^{-1} = \mathbf{W}'$ so $\mathbf{W}'\mathbf{W} = \mathbf{W}\mathbf{W}' = \mathbf{I}_N$, and called the wavelet matrix. $\mathbf{d}_j = \{d_{j,k}\}$, $j=1,2,\dots,J$, are $N/\lambda_j \times 1$ real-valued vectors of wavelet coefficients at level j associated with scale λ_j and location k , where $\lambda_j = 2^j$. The real-valued vector $\mathbf{s}_J = \{s_{J,k}\}$ is made up of $N/2^J$ scaling coefficients. Thus, the first $N - N/2^J$ elements of \mathbf{d} are wavelet coefficients and the last $N/2^J$ elements are scaling coefficients, where $J \leq M$. Notice that the length of \mathbf{X} does coincide with the length of \mathbf{d} (length of $\mathbf{d}_j = 2^{M-j}$, and $\mathbf{s}_J = 2^{M-J}$).

The wavelet coefficients tell us how much a weighted average changes from a particular time period of effective length λ_j to the next. The $N/2^J$ scaling coefficients are associated with variations on scales λ_j and higher.

In practice, the DWT is applied without exhibiting the matrix \mathbf{W} , and we therefore use a fast filtering algorithm of order $O(n)$ based on so called quadrature mirror filters that uniquely correspond to the wavelet of interest, see Mallat (1989).

In what follows, we will merely consider the wavelet in terms of filters. Now, let $\{h_1\} \equiv \{h_{1,0}, \dots, h_{1,L-1}\}$ denote the wavelet filter coefficients of a Daubechies compactly supported wavelet of width L , where $L < N$, and let $\{g_1\} \equiv \{g_{1,0}, \dots, g_{1,L-1}\}$ be the corresponding scaling filter coefficients, defined via the quadrature mirror relationship,

$$h_l = (-1)^l g_{L-1-l} \quad \text{for} \quad l = 0, \dots, L-1$$

The Haar wavelet, is a filter of width $L = 2$, that can be defined either by its wavelet coefficients (“mother” wavelet filter),

$$h_0 = \frac{1}{\sqrt{2}}, \text{ and } h_1 = -\frac{1}{\sqrt{2}},$$

or, equivalently, by its scaling coefficients (“father” wavelet filter),

$$g_0 = g_1 = \frac{1}{\sqrt{2}}.$$

The Haar wavelet is special since it is the only compactly supported (zero outside a finite interval) orthogonal wavelet that is symmetric. Daubechies (1992) derived other types of compactly supported orthogonal wavelets, often called Symmlets ($S(L)$), which are “least asymmetric”. The number in the name of the wavelet indicates the width of the filter.

Note that wavelets with a small L are narrower and less smooth, while wavelets with a large L , are relatively wide and smooth. The scaling coefficients defining Daubechies families of wavelet filters of varying length can be found in Daubechies (1992).

The filters are applied to any sequence $a = \{a_N\}$ through the operators \mathbf{H} and \mathbf{G}

$$(\mathbf{H}a)_k = \sum_N h_{N-2k} a_N ; \quad (\mathbf{G}a)_k = \sum_N g_{N-2k} a_N \quad (2)$$

An application of operator \mathbf{H} and \mathbf{G} corresponds to one step in the discrete wavelet transformation. The complete discrete wavelet transformation is a process that recursively applies equation (2).

The algorithm starts by applying the filters to the data vector \mathbf{X} and obtains the sub-vector of wavelet coefficients $\mathbf{d}_1 = \mathbf{H}\mathbf{X}$ together with the corresponding smooth coefficients $\mathbf{s}_1 = \mathbf{G}\mathbf{X}$ at level $j=1$. The procedure continues by applying the operators again to \mathbf{s}_1 to obtain $\mathbf{d}_2 = \mathbf{H}\mathbf{s}_1 = \mathbf{H}\mathbf{G}\mathbf{X}$ and $\mathbf{s}_2 = \mathbf{G}^2\mathbf{X}$, and so on until reaching the last scale J , noting that $\mathbf{d} = \mathbf{H}\mathbf{G}^{J-1}$ and $\mathbf{s}_J = \mathbf{G}^J$ contain only one coefficient. The wavelet decomposition of the vector \mathbf{X} can be represented as the vector \mathbf{d} of the same size, given by

$$\mathbf{d} = (\mathbf{H}\mathbf{X}, \mathbf{H}\mathbf{G}\mathbf{X}, \mathbf{H}\mathbf{G}^2\mathbf{X}, \dots, \mathbf{H}\mathbf{G}^{J-1}\mathbf{X}, \mathbf{G}^J\mathbf{X})' .$$

In the case that $N = 2^M$, \mathbf{d}_1 has $N/2$ coefficients, \mathbf{d}_j has $N/2^j$ coefficients and \mathbf{s}_j has $N/2^j$.

For more details about the wavelet transform in terms of operators, see Strang and Nguyen (1996).

Now, we can write (1) in terms of \mathbf{H} and \mathbf{G} at $J = 1$ as follow:

$$\mathbf{W} = \begin{bmatrix} \mathbf{H} \\ \mathbf{G} \end{bmatrix} = \begin{bmatrix} h_0 & h_1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & h_0 & h_1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & h_0 & h_1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & h_0 & h_1 \\ g_0 & g_1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & g_0 & g_1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & g_0 & g_1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & g_0 & g_1 \end{bmatrix}$$

For more details about \mathbf{W} matrix, see Vidakovic (1999). Note that when the length of the filter equals 2 (Haar wavelet), the wavelet does not exhibit problems with boundaries. An illustrative example of the DWT is given in the appendix of this paper.

The way of going back to time domain from the wavelet domain is by reconstructing \mathbf{X} with the inverse wavelet transformation. In matrix form the inverse wavelet transformation is performed by $\mathbf{X} = \mathbf{W}'\mathbf{d}$ which is again equivalent to applying a fast reconstruction algorithm using mirror filters.

2.2 Multiresolution Analysis

The concept of multiresolution analysis was first introduced by Mallat (1989). The multiresolution analysis of the data leads to a better understanding of wavelets. The idea behind multiresolution analysis is to express $\mathbf{W}'\mathbf{d}$ as the sum of several new series, each of which is related to variations in \mathbf{X} at a certain scale. Now, since the matrix \mathbf{W} is orthonormal, we can reconstruct our time series from the wavelet coefficients \mathbf{d} by using

$$\mathbf{X} = \mathbf{W}'\mathbf{d}.$$

We partition the columns of \mathbf{W}' commensurate with the partitioning of \mathbf{d} to obtain

$$\mathbf{W}' = [\mathbf{W}_1 \mathbf{W}_2 \dots \mathbf{W}_J \mathbf{V}_J], \quad .$$

where \mathbf{W}_j is an $N \times N/2^j$ matrix and \mathbf{V}_j is an $N \times N/2^j$ matrix. Thus, we can define the multiresolution analysis of a series by expressing $\mathbf{W}' \mathbf{d}$ as a sum of several new series, each of which is related to variations in \mathbf{X} at a certain scale:

$$\mathbf{X} = \mathbf{W}' \mathbf{d} = \sum_{j=1}^J \mathbf{W}_j \mathbf{d}_j + \mathbf{V}_J \mathbf{s}_J = \sum_{j=1}^J \mathbf{D}_j + \mathbf{S}_J. \quad (3)$$

The terms in (3) constitute a decomposition of \mathbf{X} into orthogonal series components \mathbf{D}_j (detail) and \mathbf{S}_j (smooth) at different scales and the length of \mathbf{D}_j and \mathbf{S}_j coincides with the length of \mathbf{X} ($N \times 1$ vector). Because the terms at different scales represent components of \mathbf{X} at different resolutions the approximation is called multiresolution decomposition, see Percival and Mofjeld (1997).

The smooth scale \mathbf{S}_j gives a smooth approximation to \mathbf{X} . Adding the detail scale \mathbf{D}_j yields \mathbf{S}_{j-1} , a scale 2^{j-1} approximation to \mathbf{X} . The \mathbf{S}_{j-1} approximation is a refinement of the \mathbf{S}_j approximation. Similarly, we can refine further to obtain the scale 2^{j-1} approximations

$$\mathbf{S}_{j-1} = \mathbf{S}_j + \mathbf{D}_j.$$

The collection $\mathbf{S}_J, \mathbf{S}_{J-1}, \dots, \mathbf{S}_1$ provides a set of multiresolution approximations of \mathbf{X} . Analogous to kernel smoothing, which has a parameter called bandwidth or smoothing parameter, we called the index J a wavelet smoothing parameter. Increasing the smoothing parameter J allows less detail in the smooth approximation of \mathbf{X} , while a small J allows additional detail in the smooth approximation of \mathbf{X} , see Ogden (1997).

Note that when $J = M$ the \mathbf{S}_J represents a “final smoothing” in the data and in the Haar case this is just the sample mean. An illustrative example of the multiresolution analysis is given in the appendix of this paper.

3. Wavelet-based estimation of the trend

A common issue in time series analysis is decomposing the different components of variations. In some applications it is important to decompose a time series into low frequencies (trend), and high frequencies (noise) components. A multiresolution analysis is a convenient setting for decomposing and describing the different scales of variation in the data. Yajima (1988) considered a polynomial regression, which consists of a polynomial trend and a stationary process with long memory. Based on the decomposition statistics we consider the following model for a time series data $\{ X_t \}$:

$$X_t = \mu + T_t + Z_t, \quad t = 0, \dots, N-1, \quad (4)$$

where μ is a constant term, T_t is an unknown deterministic polynomial trend function of order r :

$$T_t = \sum_{j=0}^r a_j t^j,$$

In this study, we consider only the first order of the polynomial function.

Z_t is a residual term which is a long-memory process defined by

$$(1-B)^\delta Z_t = \varepsilon_t, \quad (5)$$

where, $\delta \in (0, 0.5)$ is the long memory parameter, $\{\varepsilon_t\}$ is a Gaussian white noise process with mean zero and $\sigma_\varepsilon^2 > 0$. Here, B , is the backward shift operator, and

$$(1-B)^\delta = \sum_{\nu=0}^{\infty} b_\nu (\delta) B^\nu,$$

with

$$b_\nu (\delta) = (-1)^\nu \frac{\Gamma(\delta+1)}{\Gamma(\nu+1)\Gamma(\delta-\nu+1)}.$$

The spectral density function of $\{ Z_t \}$ is given by

$$S(\omega) = \frac{\sigma_\varepsilon^2}{2^\delta (1 - \cos(\omega))^\delta},$$

where $\omega = \frac{2\pi i}{N}$, $i = 0, \dots, N/2 - 1$, and $0 \leq \delta < \frac{1}{2}$.

Based on the δ two different types of processes can be defined:

1. When $\delta = 0$, the process is a white noise.
2. When $0 < \delta < 1/2$ the spectral density has a pole at zero, in which case the process exhibits slow hyperbolic decay of the autocovariances and satisfies the long memory process.

The idea of the wavelet estimation procedure is simple: we replace the unknown wavelet and scaling coefficients in (1) by estimates, which are based on the observed data.

Many authors have used the same approach in the last ten years to estimate density functions, see Brillinger (1994, 1996).

In section 2.1 it was seen that the DWT coefficients can be written into two different types: scaling coefficients $\{s_{j,k}\}$, and wavelet coefficients $\{d_{j,k}\}$.

Now, since

$$\mathbf{d} = (\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_j, \mathbf{s}_j)'$$

we can write \mathbf{d} :

$$\mathbf{d} = \mathbf{d}_w + \mathbf{d}_s,$$

where \mathbf{d}_w is an $N \times 1$ vector containing the wavelet coefficients and zeros at all other locations, and \mathbf{d}_s is an $N \times 1$ vector containing the scaling coefficients and zeros at all other locations. Since $\mathbf{X} = \mathbf{W}'\mathbf{d}$, we can write

$$\mathbf{X} = \mathbf{W}'\mathbf{d} = \mathbf{W}'\mathbf{d}_s + \mathbf{W}'\mathbf{d}_w = \hat{\mathbf{T}} + \hat{\mathbf{Z}}. \quad (6)$$

The wavelet estimator for \mathbf{X} is simply

$$\mathbf{X} = \mathbf{W}'\mathbf{d} = \mathbf{W}'\mathbf{d}_s + \mathbf{W}'\mathbf{d}_w = \hat{\mathbf{T}} + \hat{\mathbf{Z}},$$

where $\hat{\mathbf{T}}$ is an estimator of the polynomial trend \mathbf{T} at level J , while $\hat{\mathbf{Z}}$ is a tapered 'version' of \mathbf{X} .

The issue of choosing the level of our estimate depends on the goal of our application. In some applications, for example, if we want to “zooming” to local trends and cycles, we should choose J to be small. In other applications, we will want to set J to be large, if our aim is to detect the global trend. It is important to use a suitable level.

As we mentioned earlier, the wavelet estimate has the advantage of the localisation in time and frequency, which means that the detail of the estimate is seen to vary with t in contrast to the usual parametric methods or kernel estimates. This property gives us additional results of variability on different scales (different J) in the case when we have a long memory process, because such processes appear to be local trends and cycles, which are, however, spurious and disappear after some time, see Hosking (1984) and Beran (1994).

Note that, in the case of Haar wavelet ($L=2$), the estimate of the trend at level ($J = M-1$) will be simply the average of the first half series for the first $N/2$ observations and the average of the second half for the second $N/2$ observations.

An important issue is how to choose the wavelet filter. A central factor to use a particular wavelet is to match the characteristics of the series we are analysing. The Haar wavelet, which is a piecewise constant function, preserves the discontinuities, and therefore it is most suitable to identify a structural break in the data. By contrast, other wavelets with $L > 2$ are smoother and tend to blur the discontinuities.

In general, the wavelets with a wider support (L is big) are smoother but spatially less localised, while the wavelets with a narrow support (L is small) are more spatially localised but less smooth.

3.1. ANALYSIS OF VARIANCE

The DWT decompose the variance of the time series into quantities, that measure the fluctuation separately, scale by scale, see Percival and Mofjeld (1997). The orthonormality of the matrix \mathbf{W} implies that the DWT is an energy preserving transform so that

$$\|\mathbf{d}\|^2 = \|\mathbf{X}\|^2 \equiv \sum_{t=0}^{N-1} X_t^2 . \quad (7)$$

Given the structure of the wavelet coefficients, the energy in \mathbf{X} is decomposed, on a scale by scale basis, via

$$\|\mathbf{X}\|^2 = \sum_{j=1}^J \|\mathbf{D}_j\|^2 + \|\mathbf{S}_J\|^2, \quad (8)$$

where, $\|\mathbf{D}_j\|^2$ represents the contribution to the squared norm of \mathbf{X} due to change at scale λ_j , whereas $\|\mathbf{S}_J\|^2$ represents the contribution due to variations at scale λ_J . Now we can write the estimated variance of the time series in terms of wavelet and scaling coefficients:

$$\hat{\sigma}_X^2 \equiv \frac{1}{N} \sum_{t=0}^{N-1} (X_t - \bar{X})^2 = \frac{1}{N} \|\mathbf{d}\|^2 - \bar{X}^2 = \frac{1}{N} \sum_{j=1}^J \|\mathbf{d}_j\|^2 + \frac{1}{N} \|\mathbf{s}_J\|^2 - \bar{X}^2 = \sum_{j=1}^J \hat{\nu}_X^2(\lambda_j) + \hat{\sigma}_{s_J}^2,$$

where $\hat{\nu}_X^2(\lambda_j)$ is the estimated variance of the wavelet coefficients at scale λ_j , and $\hat{\sigma}_{s_J}^2$ is the estimated variance of the trend.

3.2. THE TEST STATISTIC

Let X_0, X_1, \dots, X_{N-1} be a sequence of independent normal random variables with zero mean and variance $\sigma_X^2 = 1$. We would like to test the null hypothesis $H_0 : Trend = 0$. A test statistics that can discriminate between this null hypothesis and the alternative hypothesis $H_1 : Trend \neq 0$ is defined as follows:

$$G = \frac{\hat{\sigma}_{s_J}^2}{\sum_{j=1}^J \hat{\nu}_X^2(\lambda_j)}. \quad (9)$$

To determine the distribution of G we require the distribution of the estimated variance of the trend and the distribution of the estimated wavelet variance. Since the distribution of the coefficients of \mathbf{d} are uncorrelated normal random variables with zero means and $\{\eta_{j,k}\}$ variances, where $\eta_{j,k} = \sigma_X^2$ (under H_0), see Brillinger (1996), then the squared coefficients will be distributed as a $\sigma_X^2 \chi_1^2$. Since the Haar wavelet has no boundary effects the test statistics $(N - N/2^J - J)/(N/2^J - 1)G$ can be shown to follow an F distributed with $(N/2^J - 1)$ and $(N - N/2^J - J)$ degrees of freedom.

On the other hand, when applying the Symmlets wavelets, there exist boundary effects that make the test statistics in (9) be only asymptotically distributed as F with the same degrees of freedom as in the case of Haar.

To see whether the distribution of G follows the F distribution, a Monte Carlo simulation with 10,000 replications has been performed and compared with the related F distribution. Figure 1 shows quantile-quantile plot for the $(N - N/2^J - J)/(N/2^J - 1) G$, i.e. $(1012/3) G$, against F distributions with 3 and 1012 degrees of freedom.

The distribution of the test statistic is unknown, however, in situations when the errors are not normally distributed and when they exhibit some form of dependency. It is therefore important to generate empirical critical values in such situations in order to investigate the properties of the test static. This will be done by means of simulation experiments.

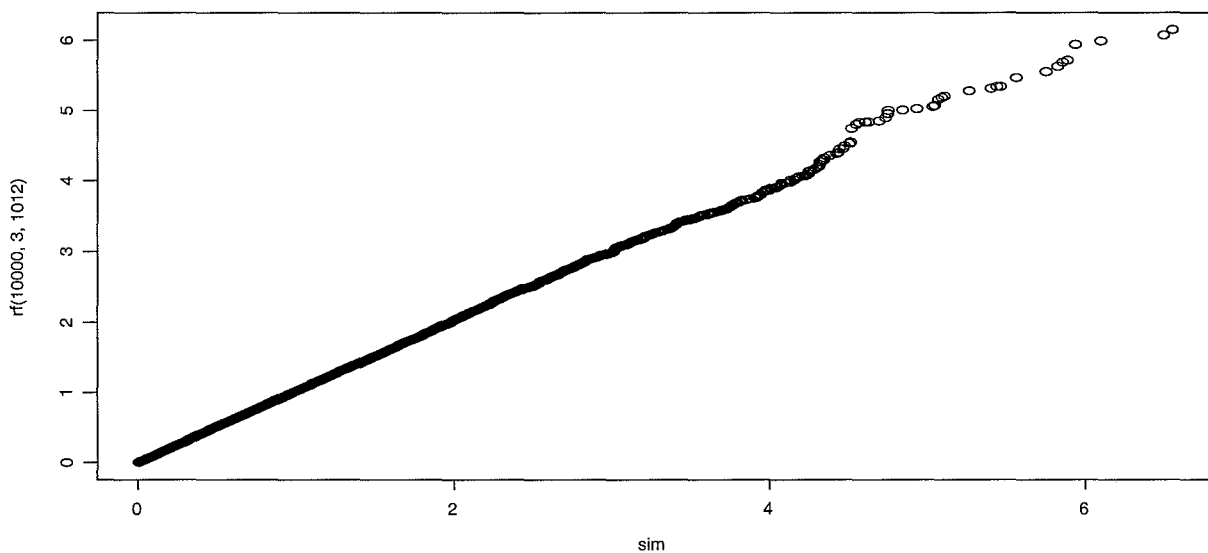


Figure 1. Quantile-quantile plot for the simulated G , using the Haar wavelet filter, against an F distribution with 3 and 1012 degrees of freedom.

4. MONTE CARLO DESIGN

In the absence of exact results it is necessary to investigate the finite sample performance of the statistics by means of simulation experiments. The design of a good Monte Carlo study is dependent on (i) what factors are expected to affect the properties of the tests under investigation and (ii) what criteria are being used to judge the results. The first question will be treated directly below, and we will return to the values used in our experiment shortly in Table I. However, the second question will be dealt with later.

Factors That Affect the Properties of the G Test

A number of factors can affect the size and/or power of the G test. The sample size (N), the long memory parameter (δ), the type of wavelet, and the trend coefficient (a_j) are four such factors. In this paper we will study the consequences of varying these factors.

A number of other factors can also affect the properties of the G test. The distribution of Z_t (and thus X_t) is an obvious candidate to examine. We have chosen first a normal distribution in our experiment, but using a fat-tailed distribution could affect the properties of the tests. However, regarding the possibility of climate change, we do not have any reason not to accept a normal approximation. On the other hand, evidence of fat-tailed distributions can often be found in empirical econometrics and / or time series, e.g. finance, demand analysis, price expectations. If a test performs well under these conditions (i.e., when the underlying assumption that the error terms are normally and identically distributed may not hold), then it is usually referred to as robust. It is hence important to study the effect of fat tailed and/or the combined effect of fat tailed and dependent errors on the properties of the G test. The class of non-normal or contamination distributions is fairly large, and the effects of these distributions mostly render the error terms heavy-tailed. More specifically, the error terms in this study will also be generated by the t-distribution with degrees of freedom equal to three and five. Proceeding in this manner we cover a reasonable degree of the fatness in the tails of the errors.

A similar point is our assumption of stationary errors; if Z_t is generated by a random walk, this could also affect the results, but this is out of the scope of this study at present.

Note that the distribution of the test statistic we use is either known only asymptotically or not known at all (e.g., in situations when the errors are not normally distributed and when they exhibit some form of dependency). As a result, the test can very likely not exhibit the correct size and inferential statements and judgements based on them might be misleading. The same is true even when using the OLS regression. Using the tabulated critical values, the t-statistic for testing the significance of the trend coefficient can also very likely not exhibit the correct size, especially under conditions of non-normally distributed errors or when they exhibit some form of dependency. However, by using the Monte Carlo simulations we can produce our critical values in different situations so that the true size of the test approaches its nominal value. Here, we use 10,000 replications for the calculation of the critical values.

In a Monte Carlo study we calculate the *estimated size* by simply observing how many times the null is rejected in repeated samples under conditions where the null is true. The Monte Carlo experiment has been performed by generating data according to (4), using the wavelet decomposition on the data, and calculating the test statistic defined in (9) and then reject the null hypothesis if the test exceeds the respective simulated critical value. For each model we have performed 10,000 replications for the calculation of the sizes and the power functions. In Tables I we present a summary of the experimental design we used in this paper.

TABLE I. Values of Factors that Vary for Different Models - Size and Power Calculations

Factor	Symbol	Design
Number of observations	N	256, 512, 1024
Nominal size	π_0	1%, 5%, 10%, 20%
Long parameter for the errors	δ	0, .15, .30, .45
Trend parameter for X (only for power calculations)	a	1/N, 3/N, 5/N

To judge the reasonability of the results, we require that the estimated size should lie within the 95% confidence interval of the actual size. Note that this bound is rather wide, but here we are trying to get at the difference between a test which grossly over- or under-rejects, and one that only deviates by about .5% or less. For example, if we consider a nominal size

of 5% we define a result as “no bias” if the estimated size lies between 4.56% and 5.44%. An approximate 95% confidence interval for the actual size (π), however, can be given as

$$\hat{\pi} \pm 2\sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{NR}}, \quad (10)$$

where $\hat{\pi}$ is the estimated size and NR is the number of replications. However, since we are mainly interested in the behaviour of the distributions in the tails, only results using the conventional 5% significance level have been analysed.

Note that most of the factors we discussed earlier are shown to affect the performances of the test. In the following section we display some important results regarding the estimated size of the test in our tables. As regards the estimated power of the test we have mainly compared them graphically.

5. RESULTS

In this section we present the results of our Monte Carlo experiment concerning the size and power of the G test. The G test has been performed to test for the existence of trend. We use different types of wavelets (i.e., the Haar and the Symmlets: S4, S6 and S8) for constructing the G test. The results are then compared with results obtained by applying the OLS method for the same purpose.

5.1. ANALYSIS OF THE SIZE

Here we present our most important results along with results of the main dominating effects of our Monte Carlo experiment regarding the size of the G test. Using the tabulated critical values from the t and F distributions, the test methods have shown to reject correctly under the null hypothesis and when there is no dependence among the errors. The tests overreject, however, when there is some form of dependency among the errors. When applying the G statistic and the t statistic (by applying the OLS method), these tests have shown to overreject under the null hypothesis by about 20% when the dependency parameter is equal to 0.15, and about 40% when it is equal to 0.45. On the other hand, when using the simulated critical values the results show that all the methods perform well regarding the size of the tests.

The simulated limits have shown to be robust when the errors are generated from normal distributions with different means and variances and even when they are generated from the t distribution. However, the limits have shown to be very sensitive to the existence of dependency among the errors. We find that higher dependence among the errors causes the limits also to be higher, see Table II.

To summarise, our analysis revealed three factors that affect the performance of the size of the G and t test statistics, namely

- 1) The choice of the critical values.
- 2) The strength of the dependency structure in the series (only affects the critical limits).
- 3) The sample size (only when using the tabulated critical values).
- 4) Type of wavelet.

Table II. The simulated critical limits, 10,000 replications.*

N = 256.

	S4	S6	S8	Haar	OLS
$\delta = 0$.0317	.0319	.0316	.0327	1.66
$\delta = .15$.0873	.0885	.0873	.0841	2.96
$\delta = .30$.2290	.2263	.2302	.2262	5.29
$\delta = .45$.5985	.5830	.5974	.6272	7.92

N = 512.

	S4	S6	S8	Haar	OLS
$\delta = 0$.0159	.0159	.0156	.0161	1.62
$\delta = .15$.0526	.0553	.0557	.0524	3.21
$\delta = .30$.1751	.1813	.1756	.1854	5.99
$\delta = .45$.4671	.4679	.4745	.4717	10.85

N = 1024.

	S4	S6	S8	Haar	OLS
$\delta = 0$.0077	.0077	.0075	.0077	1.64
$\delta = .15$.0336	.0347	.0352	.0325	3.68
$\delta = .30$.1251	.1225	.1256	.1292	7.69
$\delta = .45$.4440	.4355	.4476	.4284	14.60

* In the case of wavelets, the critical limits are based on $J = 6, 7$ and 8 for the $N = 256, 512$ and 1024 , receptively.

5.2. ANALYSIS OF THE POWER

In this sub-section we discuss the most interesting results of our Monte Carlo experiment, which regards the power of the G test using different types of wavelets. The results are also compared with those obtained by applying the t statistic. The power results of the tests are estimated by calculating rejection frequencies from 10,000 replications for trend parameters given by $1/N$, $3/N$ and $5/N$ and different sample sizes, 256, 512 and 1024, See Tables III - VI. Even if a correctly given size is not sufficient to ensure well performance of a test, it is a prerequisite. As we mentioned in the previous sub-section, when the errors are dependent the different tests accurately estimate the size only when we use the appropriate simulated critical limits. Hence, all the estimated powers in this paper have been calculated using critical values estimated from the size experiments, i.e. the powers have been size corrected.

Table III. Estimated powers

$N = 256$, trend = $1/N$.

	S4	S6	S8	Haar	OLS
$\delta = 0$.89	.85	.86	.97	1
$\delta = .15$.40	.33	.37	.60	.80
$\delta = .30$.16	.15	.15	.23	.34
$\delta = .45$.09	.09	.09	.11	.22

$N = 256$, trend = $3/N$.

	S4	S6	S8	Haar	OLS
$\delta = 0$	1	1	1	1	1
$\delta = .15$	1	1	1	1	1
$\delta = .30$.82	.74	.77	.96	.99
$\delta = .45$.31	.26	.27	.50	.76

$N = 256$, trend = $5/N$.

	S4	S6	S8	Haar	OLS
$\delta = 0$	1	1	1	1	1
$\delta = .15$	1	1	1	1	1
$\delta = .30$.99	.98	.99	1	1
$\delta = .45$.64	.53	.58	.91	.99

Table IV. Estimated powers

$N = 512$, trend = $1/N$.

	S4	S6	S8	Haar	OLS
$\delta = 0$	1	.99	.99	1	1
$\delta = .15$.60	.51	.55	.80	.95
$\delta = .30$.19	.152	.184	.25	.50
$\delta = .45$.11	.10	.10	.11	.20

$N = 512$, trend = $3/N$.

	S4	S6	S8	Haar	OLS
$\delta = 0$	1	1	1	1	1
$\delta = .15$	1	1	1	1	1
$\delta = .30$.92	.83	.87	.99	1
$\delta = .45$.40	.31	.35	.58	.74

$N = 512$, trend = $5/N$.

	S4	S6	S8	Haar	OLS
$\delta = 0$	1	1	1	1	1
$\delta = .15$	1	1	1	1	1
$\delta = .30$	1	1	1	1	1
$\delta = .45$.77	.61	.67	.95	.98

Table V. Estimated powers

$N = 1024$, trend = $1/N$.

	S4	S6	S8	Haar	OLS
$\delta = 0$	1	1	1	1	1
$\delta = .15$.85	.75	.79	.94	.99
$\delta = .30$.25	.22	.24	.35	.54
$\delta = .45$.08	.07	.08	.11	.20

$N = 1024$, trend = $3/N$.

	S4	S6	S8	Haar	OLS
$\delta = 0$	1	1	1	1	1
$\delta = .15$	1	1	1	1	1
$\delta = .30$.98	.95	.97	1	1
$\delta = .45$.34	.27	.30	.58	.88

$N = 1024$, trend = $5/N$.

	S4	S6	S8	Haar	OLS
$\delta = 0$	1	1	1	1	1
$\delta = .15$	1	1	1	1	1
$\delta = .30$	1	1	1	1	1
$\delta = .45$.75	.60	.66	.95	.99

Looking at Table III, i.e. when the sample size is equal to 256 observations and when the trend parameter is equal to $1/N$ (i.e. low power), it is clear that the OLS method exhibits the highest power when compared with the different wavelet methods. However, the Haar wavelet has shown to have higher power than the other Symmlets types. The power of the Haar is almost close to the OLS in the case of no dependency and, in general, the power deteriorates when the dependency structure becomes stronger.

When the trend parameter is equal to $3/N$ (i.e. medium power), all the methods exhibit high power (almost equal one) when there is no or low dependency. With very high dependence, the OLS is still best, but the power is lower. Finally, when we have high power (i.e. trend parameter equal $5/N$), the OLS and the Haar perform very similarly with powers equal to one, even in such situations when the dependence structure is very high.

This is almost the case when the sample size is equal to 512 observations. The estimated powers become somewhat higher in general except in the case of the OLS, with medium trend parameter. The power in this case was a bit lower than in the case of 256 observations (see Table IV).

Finally, in Table V we present the estimated powers when the sample size is equal to 1024 observations. The powers were also a bit higher in this case. With $1/N$ trend parameter the results of the Haar are almost similar to those of the OLS, except with high dependency structure in the error, in this case the OLS has higher power than the Haar but it is still around 20%. On the other hand, when the trend parameters are equal to $3/N$ and $5/N$, with no, low or medium dependence in the error structure, all the methods exhibit the highest power value of one.

Regarding the fat tailed distribution, we in Table VI present results of the estimated power when the errors are generated from the $t_{(3)}$ distribution. When the errors are generated from the $t_{(5)}$ distribution, the results have shown to be fairly similar to those in the case of normally distributed errors, and are hence not included in this paper. Looking at Table VI, we can see that almost all the methods exhibit an estimated power value equal to one in all situations. The only exception is when the sample size equals 256 and 512 and when the trend parameter equals $1/N$. The OLS and the Haar have shown to exhibit higher power than the others, and

the powers are generally rather low when comparing with those obtained when the errors are generated from the normal distribution with zero dependency.

Table VI. Estimated powers using the $t_{(3)}$ distribution,

N = 256

	S4	S6	S8	Haar	OLS
Trend = 1/N	.48	.40	.43	.69	.86
Trend = 3/N	.99	.99	.99	1	1
Trend = 5/N	1	1	1	1	1

N = 512

	S4	S6	S8	Haar	OLS
Trend = 1/N	.73	.65	.69	.88	.98
Trend = 3/N	1	1	1	1	1
Trend = 5/N	1	1	1	1	1

N = 1024

	S4	S6	S8	Haar	OLS
Trend = 1/N	.97	.94	.96	.99	1
Trend = 3/N	1	1	1	1	1
Trend = 5/N	1	1	1	1	1

The factors that affect the power of the G and t test statistics proved to be rather similar to those that affect the size. The number of observations and the strength of the dependency structure in the series had a considerable effect on the estimated powers. Using the $t_{(3)}$, the power is rather low in the case when the trend parameter is equal to 1/N, i.e. very low trend.

Generally, we find that when using the Haar wavelet, the G test exhibits more power than when using the Symmlets type wavelets. The OLS exhibits the highest power among the other methods when the trend is very low, while for medium and high trends the G test based on the Haar and OLS exhibits fairly similar power. All methods perform similarly in large samples and high trend. Note that the good performance of the OLS method is almost expected since this method is known to be best when testing for linear trends. However, in the near future we intend to investigate situations with non linear trends which may give another picture of the test's performances.

6. EMPIRICAL STUDY

In this section we apply our approach to real data. Knowledge of natural climatic variations on century scales is essential to the search for a man-made or a natural activity on the global climate. The foremost question is whether there is evidence for global warming. Statistically, global warming can be interpreted as an increasing deterministic trend. However, the interpretation of trends in observed temperature data is complicated, because the measured trends are different for different time periods.

Smith (1993) showed that the global temperature data have a steady rise from about 1910 to 1940, but there followed a period when temperatures were static or even decreasing until 1975, which could mean that the temperature data follow a long memory process. Such dependency can very likely affect the trend estimates and, therefore, should be taken into account when estimating and testing for trends.

Our data set consists of 1792 observations of seasonally adjusted monthly temperatures for Stockholm during the years 1850-1999. The global temperature series is plotted in Figure 2 and suggests the possible presence of an increasing trend (global warming). Using the maximum likelihood method we get an estimated value of the long memory parameter δ equal to 0.15.

We now apply the methodology developed in Section 3 to Stockholm's temperature data. The critical values of the temperature data have been calculated by means of Monte Carlo simulations using 10,000 replications with sample size equal to the real series (i.e., 1792 observations) and $\hat{\delta} = 0.15$. We applied the DWT at $J = 7$ and 8 , using the Haar wavelet filter and the Symmlets wavelet filter ($L = 8$). We then calculated the wavelet based test statistic G and the OLS based test statistic t , and compared them with the respective simulated critical values, see Table VII below. The tests results indicated the existence of the trend among the data at least at the 10% and 5% levels confirming the hypothesis of the "global warming". Figure 3 shows the wavelet-based estimation of the trend for the temperature data, for $J = 7$ and $J = 8$.

Table VII: Result for the G and t test statistics, along with their simulated critical limits.

	Test statistics	10%	5%	1%
Haar				
$J = 7$	0.056	0.041	0.046	0.057
$J = 8$	0.036	0.026	0.031	0.039
S8				
$J = 7$.046	0.042	0.048	0.060
$J = 8$.038	0.026	0.031	0.043
OLS	7.42	3.09	3.90	5.69

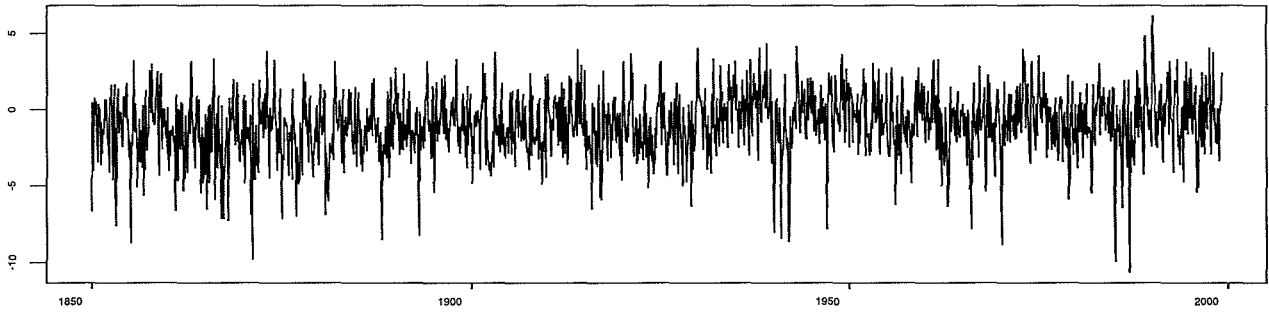


Figure 2. Stockholms temperature data.

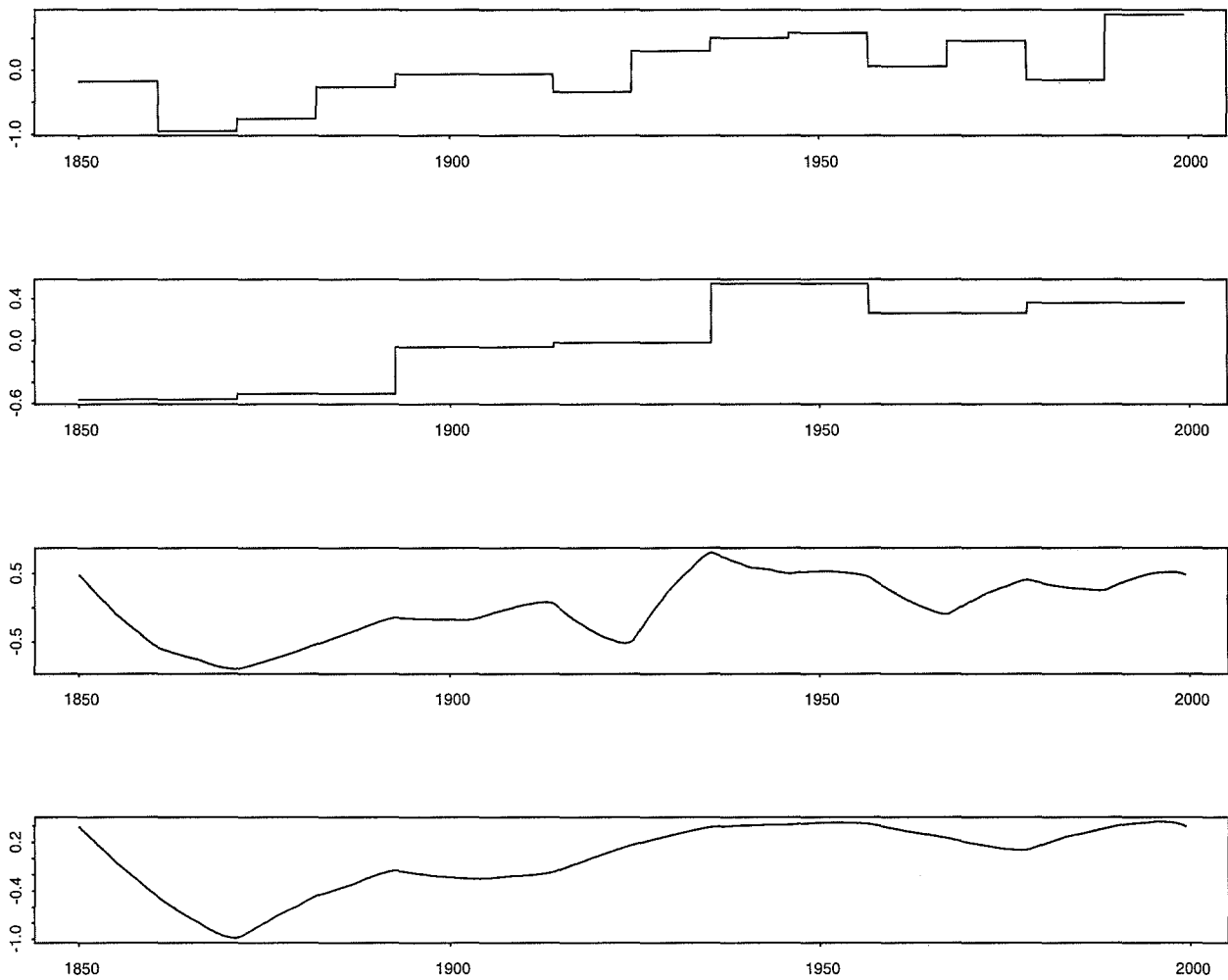


Figure 3. The estimates of the trend from a Haar and $S(8)$ based on the DWT at $J = 7$ and 8 , respectively.

7. CONCLUSIONS

In this paper, using the wavelet analysis, we introduce an alternative approach for testing for the existence of trends. Under the normality assumption the tests can be shown to be asymptotically distributed as F . The distribution of the test is unknown, however, under other conditions, e.g. non-normality. We therefore generated empirical critical values in order to investigate the properties of the test. The investigation has been carried out using Monte Carlo simulations. A number of cases were studied where the number of observations, long memory parameter and the distribution of the errors have been varied. For each case we have performed 10,000 replications and studied four different nominal sizes. The power properties have also been investigated using 10,000 replications per case, where in addition to the properties mentioned above the strength of the trend parameters has also varied. We then compared the results from the wavelet based test statistic with the OLS method for testing for linear trend.

When using the tabulated critical values from the F or t distributions, the tests have shown to overreject the null hypothesis when it is true. On the other hand, when using the simulated critical values both the test's methods perform well regarding the size of the test.

Regarding the power of the tests, the OLS has shown to have higher power than the wavelet based test when the trend parameter is very low and the long term parameter is high. The powers in these cases, however, are only around 20% even for the OLS based test. The power of the tests became higher with large samples, larger trend parameter and low long memory parameter. Under such conditions, almost all the methods behave similarly.

When the errors are generated from a fat tailed distribution we found that the Haar wavelet and the OLS have very similar and very high powers, but somewhat less than in the cases when the errors are normally distributed without dependency.

The methodology in this paper has been applied to an empirical study regarding the temperature data in Sweden for the period of 1850-1999. The results indicate a significant increasing trend in the temperature data which agree with the hypothesis of the "global warming" during the last 100 years.

Wavelets estimates have been found to take on a simple form in the Haar wavelet, and the advantage of the localisation in time and frequency leaves the estimate to vary with time in contrast to the OLS. This property gives us additional results of variability on different scales, and testing the significance of trend on different scales.

REFERENCES

- Beran, J. (1994). *Statistical Methods for Long Memory Processes*. Chapman & Hall, London.
- Bloomfield, P. and Nychka, D. W. (1992). Climate Spectra and Detecting Climate Change. *Climatic Change* 21, 275-287.
- Brillinger, D. R. (1994). Some river wavelets, *Environmetrics* 5, 211-220.
- Brillinger, D. R. (1996). Some uses of cumulants in wavelet analysis, *Nonparametric Statistics* 6, 93-114.
- Daubechies, I. (1992). *Ten Lectures on Wavelets*, Volume 61 of *CBMS-NFS Regional Conference Series in Applied Mathematics*. Philadelphia: Society for Industrial and Applied Mathematics.
- Hosking, J.R.M. (1984). Modelling persistence in hydrological time series using fractional differencing. *Water Resources Res.*, 20, 1898-1908.
- Mallat, S. G. (1989). A Theory for Multiresolution Signal Decomposition: The Wavelet Representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 11, 674-693.
- McCoy, E. J. and A. T. Walden (1996). Wavelet analysis and synthesis of stationary long-memory processes. *Journal of Computational and Graphical Statistics* 5(1), 26-56.
- Ogden, R. T. *Essential Wavelets for Statistical Applications and Data Analysis*. Birkhäuser, Boston, 1997.
- Percival, D. B. and H. O. Mofjeld (1997). Analysis of subtidal coastal sea level fluctuations using wavelets. *Journal of the American Statistical Association* 92(439), 868-880.
- Smith, R.L. (1993) Long-range dependence and global warming. *Statistics for the Environment*, V. Barnettv and F. Turkman (eds.), Wiley, Chichester.
- Vidakovic, B. (1999). *Statistical Modeling by Wavelets*. New York: John Wiley & Sons.
- Yajima, Y. (1988) On estimation of regression model with long memory stationary errors. *Ann. Statist.*, 791-807.

APPENDIX

- The Discrete Wavelet Transform (DWT):

Figure 4 shows a simple 4 elements data set $\mathbf{X} = \{6,2,4,3\}$ and their decomposition algorithm at level $J=2$, using the Haar wavelet ($h_0=1/\sqrt{2}$, $h_1=-1/\sqrt{2}$, and $g_0=g_1=1/\sqrt{2}$).

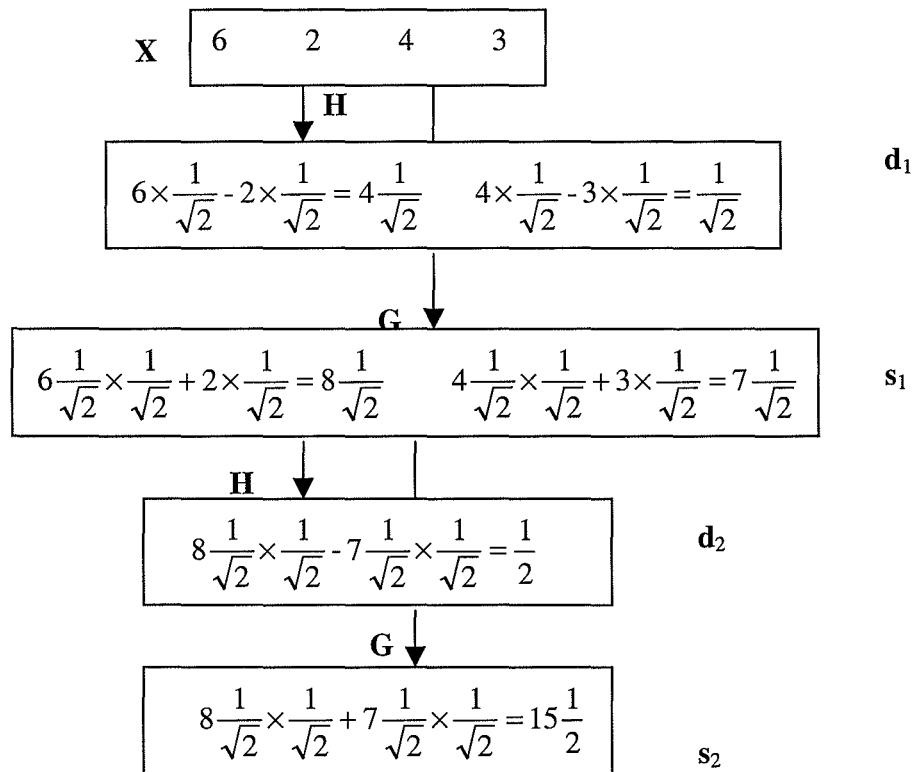


Figure 4. An illustration of a decomposition procedure.

Alternatively, we can represent \mathbf{d}_1 and \mathbf{s}_1 in Figure 4 in term of matrices as follows:

$$\mathbf{d} = \mathbf{W} \mathbf{X} \quad (11)$$

$$\begin{Bmatrix} 4\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 8\frac{1}{\sqrt{2}} \\ 7\frac{1}{\sqrt{2}} \end{Bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 & 0 \\ 0 & 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & 0 \\ 0 & 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{Bmatrix} 6 \\ 2 \\ 4 \\ 3 \end{Bmatrix}$$

We call (11) the discrete wavelet transform at level 1 (i.e. $J=1$).

Traditionally, only the detail coefficients at all scales, and the smooth coefficients at the last scale are considered a complete set of wavelet coefficients. In our example of Figure 4, the complete set is \mathbf{d}_1 , \mathbf{d}_2 and \mathbf{s}_2 .

The discrete wavelet transform at level 2 (i.e. $J=2$) which are \mathbf{d}_1 , \mathbf{d}_2 and \mathbf{s}_2 in Figure 4, can be presented as follows:

$$\mathbf{d} = \mathbf{W} \mathbf{X} \quad (12)$$

$$\begin{Bmatrix} 4\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ \frac{1}{2} \\ 15\frac{1}{2} \end{Bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 & 0 \\ 0 & 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \end{bmatrix} \begin{Bmatrix} 6 \\ 2 \\ 4 \\ 3 \end{Bmatrix}$$

- **The MultiResolution Analysis (MRA):**

$\mathbf{X}=\mathbf{W}'\mathbf{d}$ (\mathbf{W} is an orthonormal matrix, i.e. $\mathbf{W}^{-1}=\mathbf{W}$). A multiresolution analysis for (11) is:

$$\begin{Bmatrix} 6 \\ 2 \\ 4 \\ 3 \end{Bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} & 0 \\ -\frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} & 0 \\ 0 & \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ 0 & -\frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{Bmatrix} 4\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 8\frac{1}{\sqrt{2}} \\ 7\frac{1}{\sqrt{2}} \end{Bmatrix} \quad (13)$$

$$= \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 \\ -\frac{1}{\sqrt{2}} & 0 \\ 0 & \frac{1}{\sqrt{2}} \\ 0 & -\frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \mathbf{d}_1 \\ \left\{ 4\frac{1}{\sqrt{2}} \right\} \\ \left\{ \frac{1}{\sqrt{2}} \right\} \end{bmatrix} + \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{2}} & 0 \\ 0 & \frac{1}{\sqrt{2}} \\ 0 & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \mathbf{v}_1 \\ \mathbf{s}_1 \\ \left\{ 8\frac{1}{\sqrt{2}} \right\} \\ \left\{ 7\frac{1}{\sqrt{2}} \right\} \end{bmatrix} = \begin{bmatrix} 2 \\ -2 \\ 1 \\ -1 \end{bmatrix} + \begin{bmatrix} 4 \\ 4 \\ \frac{7}{2} \\ \frac{7}{2} \end{bmatrix} \quad (14)$$

We call (14) the multiresolution analysis at level 1 (i.e. $J=1$).

We now have two series \mathbf{D}_1 and \mathbf{S}_1 and the S_{11} and S_{12} are the averages of the first two observations while the S_{13} and S_{14} are the averages of the last two observations.

In the same way we can compute the multiresolution analysis at level 2 (i.e. $J=2$) from (12). Note that the S_{21} , S_{22} , S_{23} and S_{24} are the average of the \mathbf{X} .

An Illustration of the Causality Relation between Government Spending and Revenue Using Wavelets Analysis on Finish Data.

Abdullah Almasri and Ghazi Shukur

Department of Statistics
Göteborg University
P.O. Box 660

SE - 405 30 Göteborg, Sweden

ABSTRACT

Quarterly data for the period 1960:1 to 1997:2, conventional tests, a bootstrap simulation approach and a multivariate Rao's F -test have been used to investigate if, the causality between government spending and revenue in Finland have been changed at the beginning of 1990 due to future plans to create the European Monetary Union (EMU). The results indicated that during the period before 1990, the government revenue Granger caused spending, while the opposite has happened after 1990, which agrees better with Barro's tax smoothing hypothesis. However, when using monthly data instead of quarterly data for almost the same sample period, totally different results have been noted.

The general conclusion is that the relationship between spending and revenue in Finland is still not completely understood. The ambiguity of these results may well be due to the fact that there are several time scales involved in the relationship, and that the conventional analyses may be inadequate to separate out the time scale structured relationships between these variables. Therefore, to empirically investigate the relation between these variables we attempt to use the *wavelets* analysis that enables to separate out different time scales of variation in the data. We find that time scale decomposition is very important for analysing these economic variables.

Key words : Wavelets; Timescale; Causality tests; Spending; Revenue; EMU

JEL Classification: C32, H62

1. INTRODUCTION

Wavelet is a fairly new approach in analysing data (e.g. Daubechies 1992) that is becoming increasingly popular for a wide range of applications (e.g. statistics, time series analyses). This subject is not really familiar in econometrics, however, and very few studies have used the wavelets in econometric applications (e.g. Ramsey and Lampart 1998, Goffe, W.L. 1993).

Ramsey and Lampart (1998) have used the wavelet analysis and found it useful in studying the relationship between money and income. The idea was based on the fact that the time period (time scale) of the analysis is very crucial for determining those aspects of decision making that are relatively more important, and those that are relatively less important. Moreover, they stated that in econometrics one can envisage a cascade of time scales within which different levels of decisions are being made. Some decisions are taken with long horizons, others with short horizons. The authors used the US Federal Reserve Board as an example to show that the choice of the time scale determines not only the length of the period over which one requires forecasts of future events, but also the very choice of the variables that are to be the focus of the decision maker's processing of information. They empirically investigated the relation between money and income by using wavelet analysis that enables to separate out different time scales of variation in the data. Shortly speaking, they investigated the role of time scale in economic relationships in terms of money and income relationship.

In this paper we use the wavelets analysis in studying the relationship between government spending and revenue in Finland. There is some controversy about the nature of the relationship between spending and revenue and the extent to which the relevant theory is supported by the empirical evidence.

The issue of curtailing budget deficits is one of the central themes of economic policy in many member countries of the European Union (EU) and is one of the key convergence criteria of the European Monetary Union (EMU) membership. Correcting fiscal imbalances is a necessary precondition for the EMU membership. As a matter of fact, government spending has often exceeded government revenues in almost all member counties of the EU.

The question of interest focuses on the causal nexus of government spending and revenue in the new member countries of the EU. Hence, it is important to investigate whether the

political system first decides how much to spend and then decides how much to bring in as revenue. In other words, we are investigating whether the decisions regarding the amount of spending in these countries precede the decisions regarding the amount of taxes, or if connection is the other way around, or if these decisions are taken simultaneously.

Another question of interest is how the convergence criteria of the EMU affect the causality nexus? That is, has the causality changed at the beginning of 1990 because of the future plans of creating the EMU.

Shukur and Hatemi-J (1998), and Hatemi-J and Shukur (1999), investigated this subject and tried to answer analytically these questions regarding government financial policy in Finland. In Shukur and Hatemi-J (1998), the authors applied a VAR model and a VECM on quarterly data and found that only government revenue Granger causes spending for the sample period 1960:1 to 1997:2. This result did not accord with Barro's (1979) tax smoothing hypothesis, which assumes that causality runs from government spending to revenue. This hypothesis takes the path of government spending to be exogenous and taxes are adjusted to minimise distortion, while the budget is balanced intertemporally. However, in order to answer the question whether the causality changed at the beginning of 1990 because of the future plans of creating EMU, the authors split the sample into two subsamples, before and after 1990, and separately performed tests for Granger causality between spending and revenue in each subsample. They found that the causality nexus proved to exist from spending to revenue in the last subperiod, which agrees better with Barro's tax smoothing hypothesis. That is, the causality has changed direction at the beginning of 1990. One can, of course, think about this result as if the change might have happened due to the future plans of creating the EMU.

In Hatemi-J and Shukur (1999) the authors used different test methods for the same purpose. In addition to the single equation Likelihood Ratio (LR) test for causality they used two other tests, the systemwise Rao's *F*-test (Rao, 1973), developed by Shukur and Mantalos (2000), and the bootstrap test developed by Mantalos (1998). The results from this study have been shown to be similar to those found by Shukur and Hatemi-J (1998). The Rao's *F*-test has been found to work very well in integrated cointegrated VAR systems, while the bootstrap test proved to work well even in such situations where the systems are not cointegrated. Note that in the Shukur and Mantalos test (2000) the authors use the Ordinary Least Squares (OLS) method, while we in this study use Zellner's Iterative Seemingly Unrelated Regression (ISUR)

method. The ISUR technique provides parameter estimates that converge to the maximum likelihood parameter estimates which are unique.

In this paper, however, when using monthly data instead of quarterly data for almost the same sample period different results have been noted. The results obtained by using the monthly data have shown that there exist feedback relations (i.e. in two directions) between these variables over the entire sample period, 1960:01 to 1998:09. When splitting the data into two subsamples, 1960:01 to 1989:12 and 1990:01 to 1998:09, similar results have been noted. These results are obtained by applying the three different test methods, i.e. the LR test, systemwise Rao's F-test and the bootstrap test.

Therefore, the general conclusion is that the relationship between spending and revenue in Finland is still not understood completely. The ambiguity of these results may well be due to the fact that there are several time scales involved in the relationship, and that the conventional analysis may be inadequate to separate out the time scale structured relationships between the variables.

Here we attempt to shed light on this issue by separating the empirical analysis of the relationship between the variables into that between the variables after separation into time scale components. Instead of considering the net relationship over all time scales as in the conventional analysis, we (as in Ramsey and Lampart, 1998) consider a set of relationships, one for each time scale.

The paper is organised as follows: Section 1 gives an introduction. In Section 2 we introduce the wavelets analysis, while in Section 3 we present the data used in this study. In Section 4 we present the model and the methodology. Section 5 describes estimation and testing procedures, and Section 6 presents the estimated results. Finally, we give a short summary and conclusions in Section 7.

2. WAVELET ANALYSIS

The wavelet transform has been expressed by Daubechies (1992) as “a tool that cuts up data or functions into different frequency components, and then studies each component with a resolution matched to its scale”. Thus, with wavelet transform one can analyse series with heterogeneous (unlike Fourier transform) or homogeneous information at each scale. Unlike the Fourier transform, which uses only sines and cosines as basis functions, the wavelet transform can use a variety of basis functions.

The wavelet decomposition in this paper is made with respect to the so called Symlets basis, and we will hereby give a brief presentation about this decomposition methodology. Let $\mathbf{X} = (X_0, X_2, \dots, X_{T-1})'$ be a column vector containing T observations of a real-valued time series, and assume that T is an integer multiple of 2^M , where M is a positive integer. The discrete wavelet transform of level J is an orthonormal transform of \mathbf{X} defined by

$$\mathbf{d} = (\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_j, \dots, \mathbf{d}_J, \mathbf{s}_J)' = \mathbf{W}\mathbf{X},$$

where \mathbf{W} is an orthonormal $T \times T$ real-valued matrix, i.e. $\mathbf{W}^{-1} = \mathbf{W}'$ so $\mathbf{W}'\mathbf{W} = \mathbf{W}\mathbf{W}' = \mathbf{I}_T$.

$\mathbf{d}_j = \{d_{j,k}\}$, $j=1,2,\dots,J$, are $T/2^j \times 1$ real-valued vectors of wavelet coefficients at scale j and location k .

The real-valued vector \mathbf{s}_J is made up of $T/2^J$ scaling coefficients. Thus, the first $T - T/2^J$ elements of \mathbf{d} are wavelet coefficients and the last $T/2^J$ elements are scaling coefficients, where $J \leq M$. Notice that the length of \mathbf{X} does coincide with the length of \mathbf{d} (length of $\mathbf{d}_j = 2^{M-j}$, and $\mathbf{s}_J = 2^{M-J}$).

In practice, the discrete wavelet transform is applied without exhibiting the matrix \mathbf{W} , and we therefore use a fast filtering algorithms of order $O(n)$ based on so called quadrature mirror filters that uniquely correspond to the wavelet of interest, see Mallat (1989). In what follows, we will merely consider the wavelet in terms of filters.

Now, let $\{h_l\} \equiv \{h_{1,0}, \dots, h_{1,L-1}\}$ denote the wavelet filter coefficients of a Daubechies compactly supported wavelet of width L , where $L < T$, and let $\{g_l\} \equiv \{g_{1,0}, \dots, g_{1,L-1}\}$ be the corresponding scaling filter coefficients, defined via the quadrature mirror relationship,

$$h_l = (-1)^l g_{L-1-l} \quad \text{for} \quad l = 0, \dots, L-1.$$

It is important to note that the first kind of wavelet filter is called the Haar wavelet, (Haar 1910), which is a filter of width $L = 2$, that can be defined either by its wavelet coefficients,

$$h_0 = \frac{1}{\sqrt{2}}, \text{ and } h_1 = -\frac{1}{\sqrt{2}},$$

or, equivalently, by its scaling coefficients,

$$g_0 = g_1 = \frac{1}{\sqrt{2}}.$$

The Haar wavelet is special since it is the only compactly supported (zero outside a finite interval) orthogonal wavelet that is symmetric.

However, Daubechies (1992) developed a finite number of filter coefficients that are not only orthonormal, but also have compact support, i.e. the Daublets 'D(L)' and the Symmlets 'S(L)'. Note that the Daublets are quite asymmetric while Symmlets were constructed to be as nearly symmetric as possible.

The filters can be applied to any sequence $a = \{a_T\}$ through the operators \mathbf{H} and \mathbf{G}

$$(\mathbf{H}a)_k = \sum_T h_{T-2k} a_T; \quad (\mathbf{G}a)_k = \sum_T g_{T-2k} a_T.$$

An application of operator \mathbf{H} and \mathbf{G} corresponds to one step in the discrete wavelet transformation. The complete discrete wavelet transformation is a process that recursively applies to the above equation.

The algorithm starts by applying the filters to the data vector \mathbf{X} and obtains the sub-vector of wavelet coefficients $\mathbf{d}_1 = \mathbf{H}\mathbf{X}$ together with the corresponding smooth coefficients $\mathbf{s}_1 = \mathbf{G}\mathbf{X}$ at level $j=1$. The procedure continues by applying the operators again to \mathbf{s}_1 to obtain

$\mathbf{d}_2 = \mathbf{H}\mathbf{s}_1 = \mathbf{G}\mathbf{H}\mathbf{X}$ and $\mathbf{s}_2 = \mathbf{G}^2\mathbf{X}$, and so on until reaching the last scale J . The wavelet decomposition of the vector \mathbf{X} can be represented as the vector \mathbf{d} of the same size, given by

$$\mathbf{d} = (\mathbf{H}\mathbf{X}, \mathbf{G}\mathbf{H}\mathbf{X}, \mathbf{G}\mathbf{H}^2\mathbf{X}, \dots, \mathbf{G}\mathbf{H}^{J-1}\mathbf{X}, \mathbf{G}^J\mathbf{X})' .$$

For more details about the wavelet transform in terms of operators, see Strang and Nguyen (1996). The multiresolution analysis of the data leads to a better understanding of wavelets. The idea behind multiresolution analysis is to express $\mathbf{W}'\mathbf{d}$ as the sum of several new series, each of which is related to variations in \mathbf{X} at a certain scale.

Now, since the matrix \mathbf{W} is orthonormal we can reconstruct our time series from the wavelet coefficients \mathbf{d} by using

$$\mathbf{X} = \mathbf{W}'\mathbf{d} .$$

We partition the columns of \mathbf{W}' commensurate with the partitioning of \mathbf{d} to obtain

$$\mathbf{W}' = [\mathbf{W}_1 \ \mathbf{W}_2 \ \dots \ \mathbf{W}_J \ \mathbf{V}_J] ,$$

where \mathbf{W}_j is a $T \times T/2^j$ matrix and \mathbf{V}_J is a $T \times T/2^J$ matrix. Thus, we can define the multiresolution analysis of a series by expressing $\mathbf{W}'\mathbf{d}$ as a sum of several new series, each of which is related to variations in \mathbf{X} at a certain scale:

$$\mathbf{X} = \mathbf{W}'\mathbf{d} = \sum_{j=1}^J \mathbf{W}_j \mathbf{d}_j + \mathbf{V}_J \mathbf{s}_J = \sum_{j=1}^J \mathbf{D}_j + \mathbf{S}_J .$$

The terms in the previous equation constitute a decomposition of \mathbf{X} into orthogonal series components \mathbf{D}_j (detail) and \mathbf{S}_j (smooth) at different scales, and the length of \mathbf{D}_j and \mathbf{S}_j coincides with the length of \mathbf{X} ($T \times 1$ vector). Because the terms at different scales represent components of \mathbf{X} at different resolutions, the approximation is called a multiresolution decomposition, see Percival and Mofjeld (1997).

As we mentioned earlier the wavelet decompositions in this paper will be made with respect to the Symlets basis. This has been done by using the S-plus Wavelets package produced by

StatSci of MathSoft that was written by Bruce and Gao (1996). Figure 2 and Figure 3 show the multiresolution analysis of order $J = 6$ based on S(8) wavelet filter.

When choosing a specific kind of wavelets several factors should be taken into consideration. Two such important factors are the smoothness and the spatial localisation of the wavelet. In general the wavelets with a wider support (L is big) are smoother but spatially less localised, while the wavelets with a narrow support (L is small) are more spatially localised but less smooth. To get a reasonable degree of smoothness without loosing the property of spatial localisation we use quite a moderate size of L , i.e. $L = 8$. The Wavelet filter coefficients for the Symmlets of length 8, i.e. S(8), are given as follows:

$$h_0 = 0.07576571, h_1 = -0.02963553, h_2 = -0.4976187, h_3 = 0.8037388, \\ h_4 = -0.2978578, h_5 = -0.09921954, h_6 = 0.01260397, h_7 = 0.0322231.$$

Recall that the scaling filter is related to the wavelet filter via the quadrature mirror filter relationship given by equation (1).

Note that Ramsey and Lampart (1998) have used $L = 10$. Here, to investigate whether the use of other sizes of L has any impact on the results of the study, $L = 6$ and $L = 10$ have been used in some experiments, but we did not find any noticeable effects on our inferential statements.

3. DATA

The investigation of the causal relationship between government spending (S) and government revenue (R) is performed by using monthly data that are drawn from the *International Monetary Found* (IMF), and cover the period 1960:01 through 1998:9. The variables are chosen to be in logarithmic form, and hereafter will be referred to as $\ln S$ and $\ln R$, respectively.

As mentioned in Section 2 wavelet can help us in decomposing the original series into a set of orthogonal series components that provide representations of the original series. Usually a series is decomposed into six different components, e.g. D1, D2, . . . , D6, that stand for different frequency intensities in the original series, and a last component (S6) which stands

for the long run trend in the series. To explain, the time scale D1 stands for the finest level in the series and represents the highest frequency that occurs at the one-month scale. In the same manner, the D2 can stand for the next finest level in the series and represents the two-month scale, D3 for the four-month-scale, D4 for the eight-month scale, D5 for the sixteen-month scale and finally, D6, which may stand for the 32-month scale.

In addition to the monthly data we examined the relationship between the revenue and spending when the variation in each variable has been restricted to a specific scale, i.e., when the variables are transformed by wavelet transformation into the different time scales, D1 to D5 (see Figures 1 and 2). We did not use the D6 scale since it was difficult for us to find a useful interpretation of a 32-month scale. Note that the S6, in these figures, stands for the log term trend and has not been considered in our causality analysis.

Using wavelet, we reconstructed the above mentioned time series by time scales and the relations between the variables at each time scale. We then examined and illustrated how the extent to which an allowance for different effects by scale and variations in the relationships over time leads to insight into the total variation of the signal over time.

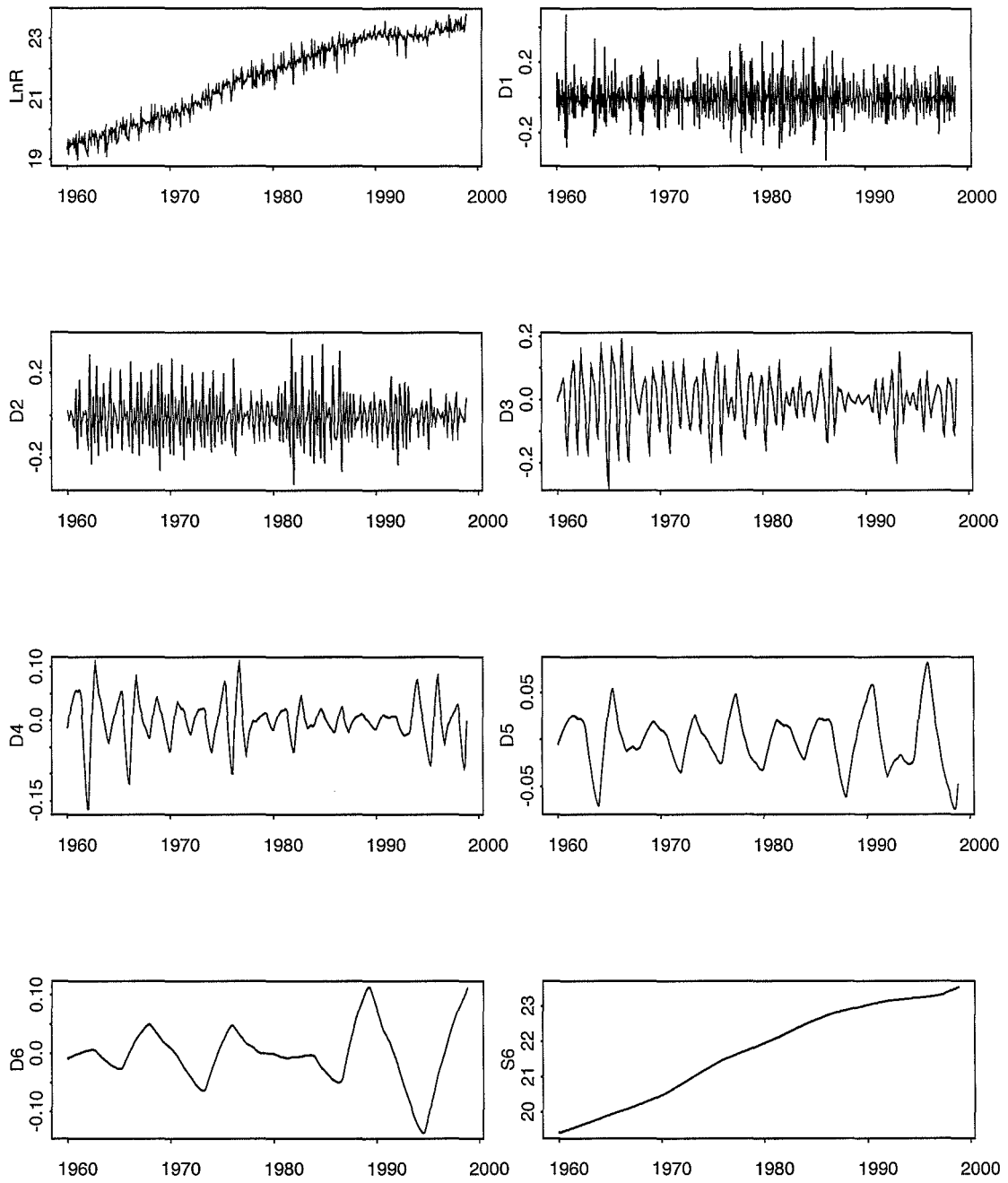


Figure 1. Time series plots of data for LnR and their different scales.

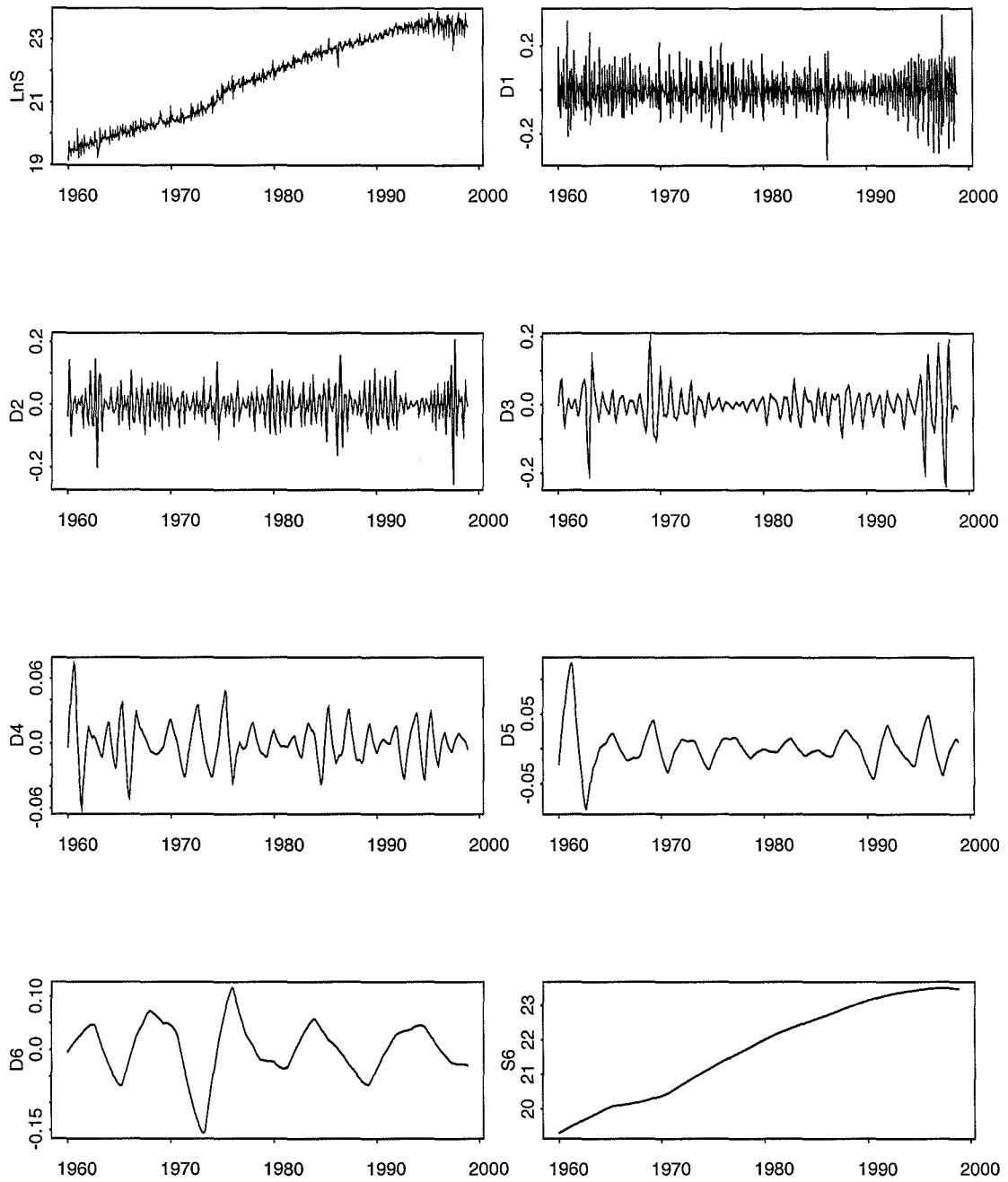


Figure 2. Time series plots of data for LnS and their different scales.

4. MODEL SPECIFICATION AND TESTING METHODOLOGY

By causality we mean causality in the Granger (1969) sense. That is, we would like to know if one variable precedes the other variable or if they are contemporaneous. The Granger approach to the question whether $\ln S$ causes $\ln R$ is to see how much of the current value of the second variables can be explained by past values of the first variable. $\ln R$ is said to be Granger-caused by $\ln S$ if $\ln S$ helps in the prediction of $\ln R$, or equivalently, if the coefficients of the lagged $\ln S$ are statistically significant in a regression of $\ln R$ on $\ln S$. Empirically, one can test for causality in Granger sense by means of the following vector autoregressive (VAR) model:

$$\ln R_t = a_0 + \sum_{i=1}^k a_i \ln R_{t-i} + \sum_{i=1}^k b_i \ln S_{t-i} + e_{1t} , \quad (1)$$

$$\ln S_t = c_0 + \sum_{i=1}^k c_i \ln R_{t-i} + \sum_{i=1}^k f_i \ln S_{t-i} + e_{2t} , \quad (2)$$

where e_{1t} and e_{2t} are error terms, which are assumed to be independent white noise with zero mean. The number of lags, k , will be decided by using the Schwarz (1978) information criteria, the Hanna and Quinn (1971) criteria and the systemwise likelihood ratio (LR) test. In order to see if the variables are cointegrated (i.e. if there exists any long run relationship between the variables) we first test for integration of each variable. A variable is integrated of order d , denoted $I(d)$, if it must be differenced d times to achieve stationarity. We use the augmented Dickey-Fuller (1979, 1981), in what follows referred to as ADF, tests for deciding the integration order of each aggregate variable. The distinction between stationary $I(0)$ and non-stationary $I(1)$ processes is a first step in analysis of time series. Several authors take the first difference to remove nonstationarity, while others are restrictive against differencing believing that information will be lost. Note that, to achieve stationarity, Ramsey and Lampart 1998 have used the data in logarithmic differenced form. Hence, to avoid any eventual loss of information, we in this study use the original series but in logarithmic form.

In the rest of this section we will present the different approaches tests for causality that we use in this study, i.e. the conventional singlewise (LR) test, and the two recommended, Rao's F -test and the Bootstrap test mentioned in Shukur and Mantalos (2000) and Mantalos (1998), respectively.

4.1. Conventional Causality Test (singlewise LR test)

According to Granger and Newbold (1986) we can test for causality in the following way:

We construct a joint F -tests for the inclusion of lagged values of $\ln S$ in (1) and for the lagged values of $\ln R$ in (2). The null hypothesis for each F -test is that the added coefficients are zero and therefore the lagged $\ln S$ does not reduce the variance of $\ln R$ forecasts (i.e. b_i in (1) are jointly zero for all i), or that lagged $\ln R$ does not reduce the variance of $\ln S$ forecasts (i.e. f_i in (2) are jointly zero for all i). If neither null hypothesis is rejected, the results are considered as inconclusive. On the other hand, if both of the F -tests rejected the null hypothesis, the result is labelled as a feedback mechanism. A unique direction of causality can only be indicated when one of the pair of F -tests rejects and the other accepts the null hypothesis.

4.2. The Systemwise Rao's F -test

In this subsection we present the Granger-causality test by using the multivariate Rao's F -test. Consider the following VAR(p) process:

$$y_t = \eta + A_1 y_{t-1} + \dots + A_p y_{t-p} + \varepsilon_t, \quad (3)$$

where $\varepsilon_t = (\varepsilon_{1t}, \dots, \varepsilon_{kt})'$ is a zero mean independent white noise process with nonsingular covariance matrix Σ_ε and, for $j = 1, \dots, k$, $E|\varepsilon_{jt}|^{2+\tau} < \infty$ for some $\tau > 0$. The order p of the process is assumed to be known. Now, by partitioned y_t in (m) and $(k-m)$ dimensional sub-vectors y_t^1 and y_t^2 and A_i matrices partitioned conformably then y_t^2 does not Granger-cause the y_t^1 if the following hypothesis:

$$H_0 = A_{12,i} = 0 \text{ for } i = 1, \dots, p-1. \quad (4)$$

is true.

Let us define:

$$Y = (y_1, \dots, y_T) \quad (k \times T) \text{ matrix,}$$

$$B = (v, A_1, \dots, A_p) \quad (k \times (kp + 1)) \text{ matrix,}$$

$$Z_t: = \begin{bmatrix} 1 \\ y_t \\ \vdots \\ y_{t-p+1} \end{bmatrix} \quad ((kp + 1) \times 1) \text{ matrix,}$$

$$Z: = (Z_0, \dots, Z_{T-1}) \quad ((kp+1) \times T) \text{ matrix, and}$$

$$\delta: = (\varepsilon_1, \dots, \varepsilon_T) \quad (k \times T) \text{ matrix.}$$

By using these notations, for $t = 1, \dots, T$, the VAR (p) model including a constant term (v) can be written compactly as:

$$Y = BZ + \delta. \quad (5)$$

We first estimate model (5), equation by equation, using the OLS method. The whole VAR system is then estimated using Zellner's Iterative Seemingly Unrelated Regression (ISUR) method. As we previously mentioned, the ISUR technique provides parameter estimates that converge to the maximum likelihood parameter estimates which are unique.

Let us denote by $\hat{\delta}_U$, the $(k \times T)$ matrix of estimated residuals from the *unrestricted* regression (3), and by $\hat{\delta}_R$ the equivalent matrix of residuals from the *restricted* regression with H_0 imposed. The matrix of cross-products of these residuals will be defined as $S_U = \hat{\delta}_U' \hat{\delta}_U$ and $S_R = \hat{\delta}_R' \hat{\delta}_R$ respectively. The Rao's F -test can be then written as:

$$RAO = (\phi/q)(U^{1/s} - 1) \quad (6)$$

$$\text{where } s = \sqrt{\frac{q^2 - 4}{k^2(G^2 + 1) - 5}}, \quad \phi = \Delta s - r, \quad \Delta = T - (k(kp+1) - Gm) + \frac{1}{2}[k(G-1) - 1],$$

$r = q/2 - 1$, and $U = \det S_R / \det S_U$. $q = Gm^2$ is the number of restrictions imposed by H_0 , where G is the p restriction in (3) and m is the dimension of the sub-vector y_t^1 . RAO is approximately distributed as $F(q, \phi)$ under the null hypothesis, and reduces to the standard F statistic when $k = 1$.

4.3. The Bootstrap Testing Approach

In this subsection we present the Bootstrap testing procedure (Efron, 1979). Generally, the distributions of the test statistics we use are known only asymptotically, which means that the tests may not have the correct size, and inferential comparisons and judgements based on them could be misleading. However, several studies (e.g. Horowitz, 1994; Mantalos and Shukur, 1998; and Shukur and Mantalos, 1997), have shown the robustness of the bootstrap critical values.

From regression (5), a direct residual resampling gives:

$$Y^* = \widehat{B}Z^* + \delta^* \quad (5a)$$

where δ^* are i.i.d. observations $\delta_1^*, \dots, \delta_T^*$, drawn from the empirical distributions (\widehat{F}_δ) putting mass $1/T$ to the adjusted OLS residuals $(\widehat{\delta}_i - \overline{\delta})$, $i = 1, \dots, T$. The basic principle of the Bootstrap testing is to draw a number of Bootstrap samples from the model under the null hypothesis, calculate the Bootstrap test statistic (T_s^*). The Bootstrap test statistic (T_s^*) can then be calculated by repeating this step N_b number of times. We then take the (α) :th quantile of the bootstrap distribution of T_s^* and obtain the α -level "bootstrap critical values" (c_{α}^*). We then calculate the test statistic (T_s) which is the estimated test statistic. Finally, we reject the null hypothesis if $T_s \leq c_{\alpha}^*$.

As regards N_b , the number of the bootstrap samples used to estimate bootstrap critical value, Horowitz (1994) used the value of $N_b=100$. In this study we follow the recommendation in Davidson and Mackinnon (1996) and use $N_b=1000$ to estimate the P-value.

5. ESTIMATION AND TESTING PROCEDURES

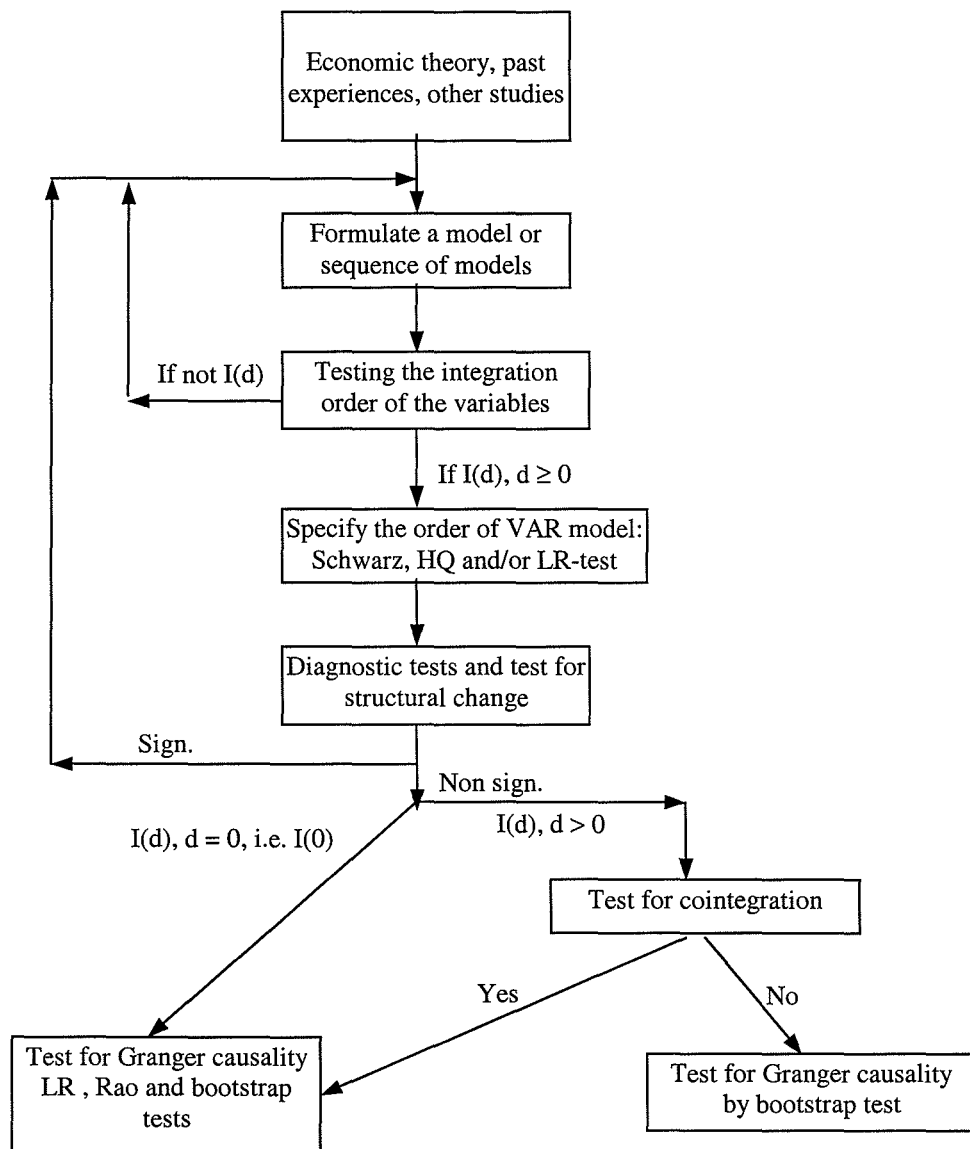
In this paper, we intend to study the causal nexus of government spending and revenue in Finland by constructing a vector autoregressive (VAR) model that allows for causality test in the Granger sense. For this purpose, we propose a simple strategy for how to select an appropriate model by successively examining the adequacy of a properly chosen sequence of models, using both single equation and systemwise tests. Note that the methodology used for misspecification testing in this paper follows the ideas described in Godfrey (1988). We apply his line of reasoning to the problem of autocorrelation, and then extend it to other forms of misspecification. Systemwise Rao's F -test have been adopted to test the adequacy of the model. If the systemwise misspecification tests lead to rejection, single equations tests will be conducted to identify specific equation/s that may lead to misspecification.

Our aim is to find a well-behaved model, which satisfies its underlying statistical assumptions, and which at the same time agrees with theoretical restrictions of economic theory. Given such a model, we then test for the presence and direction of the causality, and draw some conclusions about the study. In this study we use the standard program packages EViews, RATS, CATS, S-plus, and Gauss.

In Figure 3, we present an outline of our strategy for how to solve issues regarding specification of models and choice of proper ways to tackle situations that can arise with non-stationary series.

In order to construct a model to fit a specific data set, model builders make use of prior information derived from economic theory and previous experiences. In our study of the relationship between government spending and revenue in Finland we try to construct, estimate, test, and analyse a VAR model that adequately represents the relationship and mimic the true data-generating process. Given that such a model exists, we will then study other theoretical restrictions imposed by economic theory, and draw some important inferential statements.

Figure 3: Model Selection and Testing Strategy Outline



First, we use the ADF test to consider the integration nature of the variables included in the VAR model. If the variables are stationary, i.e. $I(0)$, we then apply the VAR model and carry out our estimation procedure. If the variables are non-stationary, any regression between them may be spurious. Accordingly, a test for cointegration has to be performed. Note that, according to the Granger (1969) representation theorem, the variables that are cointegrated have an error correction model (ECM) representation, and vice versa. Hence, another possibility for estimation is also available, i.e. the VECM. In this study, however, we only concentrate on applying the VAR model.

Second, we determine the suitable degree of the VAR process by considering a number of VAR models. We begin by estimating these models equation by equation using the OLS method.¹ We will then specify the order of the VAR model. There are two common ways, to do that, either using the LR test, or using some model selection criteria (e.g., the Schwarz (1978), Hannan and Quinn (1979) information criteria, in what follows referred to as SC and HQC respectively). The model that minimises these criteria will be selected.

Third, we apply a battery of diagnostic tests to ensure the appropriateness of the model and if the statistical assumptions are indeed satisfied. Now, if a model is subjected to several specification tests, one or more of the test statistics may be so large (or the p-values very small) that the model is clearly unsatisfactory. At that point one either has to modify the model or search for an entirely new one. For example, if the residuals appear to be autocorrelated, then we can reestimate the model using lag lengths that yield serially uncorrelated residuals. Hence, we may face a sequence of candidate models that are almost free from autocorrelation. In this case we recommend starting from the beginning, using the models selection criteria, and to perform other diagnostic tests to choose the proper model.

Fourth, if the chosen model is shown to be fairly adequate, we reconsider the integration nature of the included variables and proceed as follows:

When the variables are stationary, we will continue our procedure by testing the variables included in the selected VAR model for causality, in the Granger sense, using any of the previously mentioned tests. Here, we can verify if there is any causal nexus between the variables. On the one hand, if the variables are not stationary, we test for cointegration between them using the Johansen (1988) procedure. If there is indication of cointegration, we can still use any of the tests to investigate the causality relationship between the variables. On the other hand, in the case of non-cointegration between the variables, then only the bootstrap test approach can support us with reliable results for causality. By following the strategy outlined here, one can avoid inadequate models that might lead to misleading results and inferences.

¹ Note that the OLS estimates are both consistent and asymptotically efficient and that the Seemingly Unrelated Regressions (SUR) do not add to the efficiency of the estimation procedure since each equation in the VAR model contains the same right-hand-side variables. The Iterative Seemingly Unrelated Regressions (ISUR) method should be used, however, if cross-equation restrictions are imposed. In such case the ISUR technique provides parameter estimates that converge to the maximum likelihood parameter estimates, which are unique.

6. RESULTS

In this section we present the estimation results of applying the following testing and estimation procedure. Firstly, we test the variables (i.e. those of the monthly data and those that are produced by the wavelets transformation) for stationarity by applying the ADF test. Secondly, we determine the order of the VAR process by using the SC, HQC information criteria and LR test. Thirdly, cointegration analysis, according to the Johansen (1988) procedure, is performed. Fourthly, and finally, tests for Granger causality are carried out on the selected VAR model. Note that this procedure has been applied once for the entire sample period, i.e. 1960:01 to 1998:9, and then separately for each of the two subperiods, 1960:01 to 1990:01 and 1990:02 to 1998:9, respectively.

When looking at the whole sample period of the monthly data, the ADF test results indicate that each variable is integrated of the same order one, i.e., $I(1)$. For the time scales that are produced by the wavelet transformation, the test results indicate that all the series are to be considered as stationary variables. This is even the case for the separated sample periods. Using the above mentioned SC, HQC information criteria and LR test, we find the selected orders of the VAR process for each time scale. Repeating this procedure for each subsample, we find that the order of the VAR process remain the same. The same is right even when testing for the cointegration between the variables. The results from the Johansen test procedure have shown that the VAR systems for the integrated variables in the monthly data are cointegrated for all the sample periods.

When determining the manner of presentation, some account has to be taken to the results obtained. Our original intention was to present the results for the three time periods separately, but since the stationarity and cointegration results have shown to be the same for all the periods, and also, to make it easier to follow all the results from the study, we present our results in one overall table, Table I. The results from the different causality tests have shown to be fairly similar, and are also presented in this table. In what follows, we analyse the results starting by the monthly data and then the different wavelet's time scales.

When looking at the entire sample period and using the monthly data, the evidence indicates strongly feedback effect between the variables, i.e., that LnS and LnR are shown to strongly

TABLE I. Overall Estimated Results, *P*-values.

Time scale	<i>LnS LnR</i>		VAR order	Nature of VAR	Test methods for causality	1960-1998			1960-1990			1990-1998		
						Results	Null Hypotheses		Results	Null Hypotheses		Results	Null Hypotheses	
						<i>LnS ⇔ LnR</i>	<i>LnS ≠ LnR</i>	<i>LnR ≠ LnS</i>	<i>LnS ⇔ LnR</i>	<i>LnS ≠ LnR</i>	<i>LnR ≠ LnS</i>	<i>LnS ⇔ LnR</i>	<i>LnS ≠ LnR</i>	<i>LnR ≠ LnS</i>
Monthly data	I(1) I(1)	(5)	Cointegrated	Single-LR:	<i>LnS ⇔ LnR</i>	0.0000	0.0010	<i>LnS ⇔ LnR</i>	0.0000	0.0063	<i>LnS ⇔ LnR</i>	0.0004	0.0173	
				Rao <i>F</i> -test:	<i>LnS ⇔ LnR</i>	0.0000	0.0004	<i>LnS ⇔ LnR</i>	0.0000	0.0025	<i>LnS ⇔ LnR</i>	0.0020	0.0009	
				Bootstrap:	<i>LnS ⇔ LnR</i>	0.0000	0.0006	<i>LnS ⇔ LnR</i>	0.0000	0.0004	<i>LnS ⇔ LnR</i>	0.0023	0.0016	
D1	I(0) I(0)	(6)	Stationary	Single-LR:	<i>LnS ⇒ LnR</i>	0.0012	0.0687	inconclusive	0.1565	0.0753	<i>LnS ⇒ LnR</i>	0.0000	0.0981	
				Rao <i>F</i> -test:	<i>LnS ⇒ LnR</i>	0.0018	0.0653	inconclusive	0.1565	0.0653	<i>LnS ⇒ LnR</i>	0.0001	0.0671	
				Bootstrap:	<i>LnS ⇒ LnR</i>	0.0020	0.0630	inconclusive	0.1490	0.0680	<i>LnS ⇒ LnR</i>	0.0000	0.0820	
D2	I(0) I(0)	(4)	Stationary	Single-LR:	inconclusive	0.0700	0.9529	inconclusive	0.2980	0.9333	<i>LnS ⇒ LnR</i>	0.0166	0.8855	
				Rao <i>F</i> -test:	inconclusive	0.0719	.9500	inconclusive	0.2988	0.9334	<i>LnS ⇒ LnR</i>	0.0155	0.7579	
				Bootstrap:	inconclusive	0.0710	0.8900	inconclusive	0.2960	0.9380	<i>LnS ⇒ LnR</i>	0.0160	0.7930	
D3	I(0) I(0)	(2)	Stationary	Single-LR:	inconclusive	0.7520	0.5292	inconclusive	0.7705	0.6953	inconclusive	0.3298	0.6021	
				Rao <i>F</i> -test:	inconclusive	0.8672	0.4599	inconclusive	0.7710	0.6960	inconclusive	0.6954	0.5153	
				Bootstrap:	inconclusive	0.7080	0.5290	inconclusive	0.7910	0.7230	inconclusive	0.7140	0.5910	
D4	I(0) I(0)	(4)	Stationary	Single-LR:	inconclusive	0.1556	0.3736	inconclusive	0.3631	0.5758	<i>LnS ⇒ LnR</i>	0.0141	0.3869	
				Rao <i>F</i> -test:	inconclusive	0.1300	0.3491	inconclusive	0.3638	0.5765	<i>LnS ⇒ LnR</i>	0.0149	0.5153	
				Bootstrap:	inconclusive	0.1360	0.3340	inconclusive	0.3630	0.5660	<i>LnS ⇒ LnR</i>	0.0110	0.4750	
D5	I(0) I(0)	(5)	Stationary	Single-LR:	inconclusive	0.5296	0.3714	inconclusive	0.1311	0.5279	inconclusive	0.9733	0.4320	
				Rao <i>F</i> -test:	inconclusive	0.5299	0.3717	inconclusive	0.1314	0.5283	inconclusive	0.3463	0.8320	
				Bootstrap:	inconclusive	0.5900	0.5090	inconclusive	0.1660	0.5080	inconclusive	0.8720	0.7110	

Granger cause each other. We obtained similar indications even for the separate subperiods, 1960:01 to 1990:01 and 1990:02 to 1998:9. This may mean that the decisions regarding the amount of spending and the decisions regarding the amount of taxes are taken simultaneously.

Looking at the very finest time scale, D1, the evidence from the whole sample period, indicates that LnS strongly Granger cause LnR . When analysing the first subperiod, however, the results indicate inconclusive causality nexus between the variables. When analysing the second subperiod, however, the results have shown that causality nexus may exist from LnS to LnR , which agrees better with Barro's tax hypothesis and those results found by Shukur and Hatemi-J (1998), using the quarterly data.

At the next finest time scale, D2, when looking at the entire period and the first subperiod, the results indicate inconclusive causality relation, while we find that LnS strongly Granger cause LnR during the second subperiod. The results from the second subperiod support those of D1 that are obtained from the same subperiod.

At the first intermediate time scales, the second intermediate time scales and the highest level of time scale, i.e., D3, D4, and D5, respectively, the results indicate inconclusive causality relation between the two variables in almost all cases. The only exception is for D4 during the second subperiod, the results in this case have shown that only LnS Granger causes LnR .

Note that when considering Figure 1 and 2, we can see how the wavelet transformations can successfully and clearly zoom out the high frequency variations in the data, which is not so clear when considering the original monthly data. A clear consideration of, for example, the second subperiod of the time scales D1 and D2 can show that the variations in the LnS series are much higher than the variations in the LnR series. This may give an alternative indication of the influence of the LnS on the LnR .

The results are fairly plausible. When analysing the monthly data, the relation between these variables indicate a feedback mechanism in all cases, which can imply that the decisions regarding the amount of spending and the decisions regarding the amount of taxes are taken simultaneously. On the other hand, when looking at the finest, and second intermediate scales, D1, D2 and D4, respectively, we generally find that strong causality nexus may exist from LnS to LnR during the second time period. This can imply that the decisions regarding the

amount of spending precede the decisions regarding the amount of taxes, which agrees with Barro's tax hypothesis. In other words, at those time scales or at low frequencies, the political system in Finland, during the second subperiod, first decides how much to spend and then decides how much to bring in as revenue by taxes. This can be a result of the future plans of creating EMU.

7. CONCLUSIONS

In this paper, we empirically investigate the relation between government spending and revenue by using wavelet analysis that enables to separate out different time scales of variation in the data, i.e. we investigate the role of time scale in economic relations in terms of government spending and revenue.

When investigating the causal nexus of government spending and revenue in Finland, using monthly data and quarterly data, different results have been obtained. This may be due to the fact that there are several time scales involved in the relationship, and that the conventional analysis may be inadequate to separate out the time scale structured relationships between the variables.

Hence, we used the wavelet analysis in investigating the causal nexus of government spending and revenue in Finland during the period 1960:01 through 1998:09. Different test methods have been used to investigate the causality nexus between these variables. The most interesting result is that when looking at the finest, and second intermediate scales, D1, D2 and D4, respectively, tests results indicate that strong causality nexus may exist from LnS to LnR during the second time period, i.e., 1990 to 1998. This can imply that the decisions regarding the amount of spending precede the decisions regarding the amount of taxes, i.e. the political system in Finland, first decides how much to spend and then decides how much to bring in as revenue by taxes. The above result agrees with Barro's tax hypothesis and can be a result of the future plans of creating EMU.

REFERENCES

- Barro, R. L. (1979): "On the determination of the public debt" *Journal of political economy*, **87**, 940-941.
- Bruce, A. G. and Gao, H.-Y. (1996). *Applied Wavelet Analysis Through S-Plus*, New York: Springer-Verlag.
- Davidson, R and J.G. MacKinnon (1996): "The size Distortion of Bootstrap Tests" *Working Paper*, Department of Economics, University of Queen's, Canada.
- Daubechies, I. (1992). *Ten Lectures on Wavelets*, Volume 61 of *CBMS-NFS Regional Conference Series in Applied Mathematics*. Philadelphia: Society for Industrial and Applied Mathematics.
- Dickey, D. A., and Fuller, W. A. (1979): "Distribution of the Estimators for Autoregressive Time Series With a Unit Root," *Journal of the American Statistical Association*, **74**, 427-431.
- _____, (1981): "The Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root," *Econometrica*, 1057-72.
- Efron, B. (1979): "Bootstrap methods: Another look at the jack-knife," *Annals of Statistics* **7**, 1-26.
- Godfrey, L. G. (1988): "Misspecification Tests in Econometrics," Cambridge: Cambridge University Press.
- Goffe, W. L. (1993): "Wavelets in macroeconomics: An introduction," in D. A. Belsley (Ed.), *Computational Techniques for Econometrics and Economic Analysis*, Kluwer Academic Publishers.
- Granger, C. W. J. and P. Newbold (1986): "*Forecasting Economic Time Series*," 2nd ed. New York: Academic Press.
- Granger, C. W. J. (1969): "Investigating Causal Relations by Econometric Models an Cross-Spectral Methods," *Econometrica*, **37**, 24-36.
- Hannan, E. J., and Quinn, B.G. (1979): "The Determination of the Order of an Autoregressive," *Journal of the Royal Statistical Society*, **B41**, 190-195.
- Hatemi-J, Abdulnasser and Shukur, Ghazi (1999), "The Causal Nexus of Government Spending and Revenue in Finland: A Bootstrap Approach," *Applied Economics Letters*, **4**, 641-644.
- Horowitz, J. L. (1994): "Bootstrap-based critical values for the information matrix test," *Journal of Econometric*, **61**, 395-411.
- Johansen, S., (1988): "Statistical Analysis of Cointegration Vectors," *Journal of Economic Dynamics and Control*, **12**, 231-54.
- Mallat, S. G. (1989): "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **11**, 674-693.
- Mantalos, P. (2000): "A Graphical Investigation of the Size and Power of the Granger-Causality Tests in Integrated-Cointegrated VAR Systems," To appear in *Studies in Nonlinear Dynamics & Econometrics*, **4**, nr 1, 2000.

- Mantolos, P. and Shukur, G. (1998): "Size and Power of the Error Correction Model Cointegration Test. A Bootstrap Approach," *Oxford Bulletin of Economics and Statistics*, **60**, 249-255.
- Percival, D. B. and H. O. Mofjeld (1997): "Analysis of subtidal coastal sea level fluctuations using Wavelets," *Journal of the American Statistical Association*, **92**(439), 868-880.
- Ramsey, J. B. and C. Lampart (1998): "Decomposition of Economic relationships by Timescale Using Wavelets, Money and Income," *Macro economic Dynamics*, **2**, p. 49-71.
- Rao, C. R. (1973): "Linear Statistical Inference and Its Application," Second edition. New York: Wiley.
- Schwarz, G. (1978): "Estimation the Dimension of a Model," *Annals of Statistics*, **6**, 461-464.
- Shukur, G. and Hatemi-J A. (1998): "The Causal Nexus of Government Spending and Revenue in Finland," *Working paper* 1998:3, Department of Statistics and Department of Economics, Lunds University, Sweden.
- Shukur, G., and P. Mantalos (1997): "Size and Power of the RESET Test as Applied to Systems of Equations: A Bootstrap Approach," Department of Statistics, *Working paper* 1997:3, Lunds University, Sweden.
- Shukur, G. and Mantalos, P. (2000): "A Simple Investigation of the Granger-causality Test in Integrated-Cintegrated VAR Systems," Accepted for publication, *Journal of Applied Statistics*.
- Strang, G and T. Nguyen (1996): *Wavelets and Filter Banks*. Wellesley-Cambridge Press, Wellesley, MA.

Acknowledgment

The authors are very grateful to Professor Douglas Hibbs for interesting and helpful comments.

Research Report

- | | | |
|---------|---------------------------------------|---|
| 1999:5 | Mantalos, P. & Shukur, G.: | Testing for cointegrating relations- A bootstrap approach. |
| 1999:6 | Shukur, G.: | The effect of non-normal error terms on the properties of systemwise RESET test. |
| 1999:7 | Järpe, E. & Wessman, P.: | Some power aspects of methods for detecting different shifts in the mean. |
| 1999:8 | Johnsson, T.: | On statistics and scientific thinking. |
| 1999:9 | Afsarinejad, K.: | Trend-free repeated measurement designs. |
| 1999:10 | Carlquist, A. m.fl. | The impact of season and climate on growth during early childhood in different socio-economic groups in Lahore, Pakistan. |
| 1999:11 | Carlquist, A, Erling, V. & Frisé, M.: | Longitudinal methods for analysis of the influence of breastfeeding on early child in Pakistan. |
| 1999:12 | Carlquist, A.: | Longitudinal methods for analysis of early child health in Pakistan. |
| 1999:13 | Frisé, M.: | Characterization of methods for surveillance by optimality. |
| 2000:1 | Hatemi-J, A. & Shukur, G.: | Multivariate based causality tests of twin deficits in the US. |
| 2000:2 | Petzold, M.: | Aspects on tolerance limit estimation - some common approaches and flexible modeling. |
| 2000:3 | Andersson, L.: | Statistical test of the existence of a turning point. |