J. King Saud Univ., Vol. 21, *Admin. Sci.* (1) (2), pp. 1-11, Riyadh (2010/1431H.)

The Wavelet Transformation and Decomposition of ARIMA Models

Shaban A. Shaban* and Ayman Orabi**
**Institute of Statistical Studies & Research*
Cairo University, Cairo, Egypt
drshabanshaban @yahoo.com
***College of Administrative Sciences*
King Saud University, Riyadh, Saudi Arabia
Orabi_issr@yahoo.com

(Received 10/1/1431H.; accepted for publication 24/5/1431H.)

Keywords. Wavelet transformation; Threshold; Autoregressive, moving-average, ARIMA model; Sum square error.

Abstract. The objective of this article is to study the effect of wavelet filter on the time series data. By using wavelet transformation and hard thresholding technique, the ARIMA model decomposed into sum of two ARIMA models and the relation between sum of square errors due to the ARIMA model and sum of two ARIMA models will be discussed. . By using wavelet transformation and soft thresholding technique, the ARIMA model decomposed into sum of three ARIMA models and the relation between sum of square errors due to the ARIMA model and sum of three ARIMA models will be discussed.

1. Introduction

The wavelet transform is a powerful mathematical tool that is receiving more and more attention by the statistical community. While most work is being done in the engineering and physical sciences, wavelet transforms have already proven useful in well established statistical fields such as nonparametric regression, classification, and time series analysis. The ground_breaking work of Donoho and co-workers (Donoho 1993; Donoho and Johnstone1994; Donoho 1995; Donoho, Johnstone, Kerkyacharian, and Picard 1995) introduced statisticians to wavelet transforms in the context of signal estimation and wavelet shrinkage.

The past two decades have witnessed the development of wavelet analysis, Donoho and Johnstone (1995), Johnstone and Silverman (1997), Nason and von Sachs (1999), Priestley (1996), Percival, and Walden, (1999), have applied wavelet theory to the estimation of the functions whose observations are contaminated by noise as well as time series analysis either in the time domain or in the observation are equally spaced and independent. Most of these applications are based on the assumption that the observations are equally spaced and independent. However, for time series data the observations are likely to be dependent.

In **Section** 2, a short background on wavelet will be introduced. In **Section** 3, Bayesian Wavelet Shrinkage and Thresholding will be represented. In **Section** 4, we study the effect of wavelet transformation, hard thresholding technique and soft thresholding technique on the sum of square error for the ARIMA model. In **Section** 5, an application is illustrated by example.

2. A short background on wavelets

In this section we give a brief overview of some relevant material on the wavelet series expansion and a fast wavelet transform that we will need latter.

2.1. The wavelet series expansion

The term wavelets is used to refer to a set of orthonormal basis functions generated by dilation and translation of a compactly supported scaling function (or father wavelet), ϕ, and a mother wavelet, ψ, associated with an r -regular multiresolution analysis of $L^2(R)$. A variety of different wavelet families now exist that combine compact support with various degrees of smoothness and numbers of vanishing moments (see, Daubechies (1992)), and these are now the most intensively used wavelet families in practical applications in statistics. Hence, many types of functions encountered in practice can be sparsely (i.e. parsimoniously) and uniquely represented in terms of a wavelet series. Wavelet bases are therefore not only useful by virtue of their special structure, but they may also be (and have been!) applied in a wide variety of contexts.

For simplicity in exposition, we shall assume that we are working with periodized wavelet bases on [0, 1] (see, for example, Mallat (1999)), letting the periodized wavelet denote as:

$$\phi_{jk}^{p}(t)=\sum_{l\in Z}\phi_{jk}(t-l) \text{ and}$$

$$\psi_{jk}^{p}(t)=\sum_{l\in Z}\psi_{jk}(t-l), \quad \text{for} \quad t\in[0,1],$$

Where

$$\phi_{jk}(t)=2^{j/2}\phi\left(2^{j}t-k\right)$$

$$\text{and} \quad \psi_{jk}(t)=2^{j/2}\psi\left(2^{j}t-k\right).$$

For any $j_0 \geq 0$, the collection

$$\left\{\phi_{j_0k},\ k=0,1,\ldots,2^{j_0}-1;\ \psi_{j_0k},\ \mathrm{j}\geq \mathrm{j}_0\geq 0,\ k=0,1,\ldots,2^{j}-1\right\}$$

is then an orthonormal basis of $L^2([0,1])$.

The idea underlying such an approach is to express any function $g\in L^2([0,1])$ in the form

$$g(t)=\sum_{k=0}^{2^{j_0}-1}\alpha_{j_0k}\phi_{j_0k}(t)$$

$$+\sum_{j=j_0}^{\infty}\sum_{k=0}^{2^{j}-1}\beta_{jk}\phi_{jk}(t),\quad \mathrm{j}_0\geq 0,\quad t\in[0,1],$$

where

$$\alpha_{j_0k}=\langle g,\phi_{j_0k}\rangle=\int_0^1 g(t)\phi_{j_0k}(t)dt,\quad \mathrm{j}_0\geq 0,$$

$$k=0,1,\ldots,2^{j_0}-1$$

and

$$\beta_{jk}=\langle g,\psi_{jk}\rangle=\int_0^1 g(t)\psi_{jk}(t)dt,\quad \mathrm{j}\geq \mathrm{j}_0\geq 0,$$

$$k=0,1,\ldots,2^{j}-1$$

An usual assumption underlying the use of periodic wavelets is that the function to be expanded is assumed to be periodic. However, such an assumption is not always realistic and periodic wavelets exhibit a poor behavior near the boundaries (they create high amplitude wavelet coefficients in the neighborhood of the boundaries when the analyzed function is not periodic). However, periodic wavelets are commonly used because the numerical implementation is particular simple. While, as Johnstone (1994) has pointed out, this computational simplification affects only a fixed number of wavelet coefficients at each resolution level, we will also present later on an effective method, developed recently by Oh and Lee (2005), combining wavelet decompositions with local polynomial regression, for correcting the boundary bias introduced by the inappropriateness of the periodic assumption.

2.2. The discrete wavelet transform

In statistical settings we are more usually concerned with discretely sampled, rather than continuous, functions. It is then the wavelet analogy to the discrete Fourier transform which is of primary interest and this is referred to as the discrete wavelet transform (DWT). Given a vector of function values $g=(g(t_1),\ldots,g(t_n))'$ at equally spaced points t_i, the discrete wavelet transform of g is given by:

$$d = Wg \tag{1}$$

where d is an $n\times 1$ vector comprising both discrete scaling coefficients, c_{j_0k}, and discrete wavelet coefficients, d_{jk}, and W is an orthogonal $n\times n$ matrix associated with the orthonormal wavelet basis chosen. The c_{j_0k} and d_{jk} are related to their continuous counterparts α_{j_0k} and β_{jk} (with an approximation error of order n^{-1}) via the relationships:

$$c_{j_0k}\approx\sqrt{n}\,\alpha_{j_0k} \quad \text{and} \quad d_{jk}\approx\sqrt{n}\,\beta_{jk}.$$

The factor $\sqrt{n}$ arises because of the difference between the continuous and discrete orthonormality conditions. This root factor is unfortunate but both the definition of the DWT and the wavelet coefficients are now fixed by convention, hence the different notation used to distinguish between the discrete wavelet coefficients and their continuous counterpart. Note that, because of orthogonality of W, the inverse DWT (IDWT) is simply given by:

$$g = W^T d \qquad (2)$$

where W^T denotes the transpose of W.

If $n = 2^J$ for some positive integer J, the DWT and IDWT may be performed through a computationally fast algorithm developed by Mallat (1989) that requires only order n operations. In this case, for a given j_0 and under periodic boundary conditions, the DWT of g results in an n-dimensional vector $\mathbf{d}$ comprising both discrete scaling coefficients $c_{j_o k}$, $k = 0,1,...,2^{j_0} - 1$ and discrete wavelet coefficients d_{jk}, $j = j_0,...,J-1$; $k = 0,1,...,2^j - 1$.

We do not provide technical details here of the order n DWT algorithm mentioned above. Essentially the algorithm is a fast hierarchical scheme for deriving the required inner products which at each step involves the action of low and high pass filters, followed by a decimation (selection of every even member of a sequence). The IDWT may be similarly obtained in terms of related filtering operations. For excellent accounts of the DWT and IDWT in terms of filter operators we refer to Nason & Silverman (1995), Strang & Nguyen (1996), or Burrus, Gonipath & Guo (1998).

2.3. Classical threshold schemes

Since the wavelet representation of many kinds of function is very economical, it is reasonable to assume that there are a few large value wavelet coefficients concentrated near the areas of major spatial activity, e.g. discontinuities, but the majority of wavelet coefficients are small. Also, owning to the fact that the wavelet transform is orthogonal, if the ε_i are assumed to be independent Gaussian noise, then the wavelet coefficients will also be contaminated with independent Gaussian noise. So in this case, the empirical wavelet coefficients can be written as

$$\tilde{d}_{jk} = d_{jk} + \varepsilon_{jk} \qquad (3)$$

and $\tilde{d}_{jk}$ is distributed as:

$$\tilde{d}_{jk} \sim N\left(d_{jk}, \sigma^2\right)$$

Based on these assumptions, Donoho and Johnstone (1994, 1995) suggested two types of thresholding methods: hard and soft thresholding. Hard thresholding sets all the wavelet coefficients to be 0 if their absolute values are below a certain threshold $\lambda \geq 0$:

$$\hat{d}_{jk} = \eta_\lambda\left(\tilde{d}_{jk}\right) = \tilde{d}_{jk} I\left(\left|\tilde{d}_{jk}\right| > \lambda\right)$$

(hard thresholding) (4)

Soft thresholding shrinks the wavelet coefficients that are larger than the threshold by λ:

$$\hat{d}_{jk} = \eta_\lambda\left(\tilde{d}_{jk}\right) = \operatorname{sgn}\left(\tilde{d}_{jk}\right)\max\left(0, \left|\tilde{d}_{jk}\right| - \lambda\right)$$

(soft thresholding)

Hard and soft thresholdings are illustrated in Fig. (1).

2.4. Choices of threshold

Too large a threshold might cut off important parts of the true function underlying the data, whereas too small a threshold may excessively retains noise in the reconstruction. Universal Threshold Donoho and Johnstone (1994) proposed the universal threshold:

$$\lambda = \sigma\sqrt{2\log(n)}$$

When σ is unknown, σ may be replaced by a robust estimate $\hat{\sigma}$, such as the median absolute deviation (MAD) of the wavelet coefficients at the finest level $J = \log(N) - 1$ divided by 0.6745 and can be expressed as

$$\hat{\sigma} = MAD\left\{d_{jk}, k = 1,...,2^J\right\}/0.6745$$

3. Bayesian Wavelet Shrinkage and Thresholding

Various Bayesian approaches for thresholding and non-linear shrinkage in general have been proposed recently. See for example Chipman et al. (1997), Abromovich and Sapatinas (1999), Abramovich et al. (2000), Clyde and George (1999, 2000) and Johnstone and Silverman (1998, 2005). These methods have been shown to be effective. In these approaches, a prior distribution is imposed on the wavelet coefficients, which is designed to capture the sparseness of the wavelet expansions that is common to most applications. The function can then be estimated by applying a suitable Bayesian rule to the resulting posterior distribution of wavelet coefficients. In general, a Bayesian rule $\eta(x)$ is a shrinkage rule if and only if η is antisymmetric and increasing on $(-\infty,\infty)$ and $0 \le \eta(x) \le x$ for all $x \ge 0$. The family of shrinkage rules $\eta(x,t)$ will be a thresholding rule with threshold t if and only if

$$\eta(x,t) = 0 \text{ if and only if } |x| \le t$$

A popular prior model for each wavelet coefficient d_{jk} is a mixture of one normal distribution and a point mass at zero. The normal distribution with large variance represents the significant coefficients while a point mass at zero represents the negligible ones. A hierarchical model can be expressed as:

$$d_{jk} \,\Big|\, r_j \sim r_j N(0,\tau_j^2) + (1-r_j)\delta(0) \qquad (5)$$

where $r_j \sim Bernoulli(p_j)$ for different resolution level j and $\delta(0)$ is a point mass at zero. The binary random variable r_j determines whether the relevant wavelet coefficient is nonzero $(r_j = 1)$, and comes from an $N(0,\tau_j{}^2)$ distribution, or zero $(r_j = 0)$, and arises from a point mass at zero. From (3), the posterior cumulative distribution of d_{jk} conditional on the empirical wavelet coefficient $\hat{d}_{jk}$ and σ^2 is given by

$$d_{jk} \,\Big|\, \hat{d}_{jk},\sigma^2 \sim \Pr(r_{jk} = 1 \,|\, \hat{d}_{jk},\sigma^2) N\left(\frac{\tau_j^2}{\sigma^2+\tau_j^2}\hat{d}_{jk}, \frac{\tau_j^2\sigma^2}{\sigma^2+\tau_j^2}\right)$$
$$+(1-\Pr(r_{jk} = 1 \,|\, \hat{d}_{jk},\sigma^2)\delta(0) \qquad (6)$$

The posterior probabilities can be expressed as

$$\Pr(r_{jk} = 1 \,|\, \hat{d}_{jk}) = \frac{1}{1+O_{jk}(\hat{d}_{jk},\sigma^2)} \qquad (7)$$

$$O_{jk}(\hat{d}_{jk},\sigma^2) = \frac{1-p_j}{p_j} \cdot \frac{(\sigma^2+\tau_j^2)^{1/2}}{\sigma} \exp\left(-\frac{\tau_j^2\hat{d}_{jk}^2}{2\sigma^2(\sigma^2+\tau_j)}\right) \qquad (8)$$

3.1. Shrinkage estimates using posterior mean approaches

Clyde *et al.* (1998) obtained wavelet shrinkage estimates by considering the posterior mean. Assuming that an accurate estimate of the noise variance is available, the closed form expressions for the posterior mean of wavelet coefficient d_{jk} conditionally on $\hat{d}_{jk}$ and σ^2, can be derived from (6) and (7)) as

$$E(d_{jk} \,|\, \hat{d}_{jk},\sigma^2) = \frac{1}{1+O_{jk}(\hat{d}_{jk},\sigma^2)} \cdot \frac{\tau_j^2}{\sigma^2+\tau_j^2}\hat{d}_{jk} \qquad (9)$$

4. Wavelet and ARIMA model

In this Section, we study the effect of wavelet transformation on the sum of square error for the ARIMA model.

Theorem1:

Suppose that the time series Z_t has constant variance and non seasonality and it is contaminated by correlated noise, using wavelet transformation and hard threshold technique, this time series is decomposed into two time series as following:

$$Z_t = X_t + Y_t$$

where X_t and Y_t denote the signal and noise, respectively, then the sum of the two sum square residuals from ARIMA model of X_t pluse the sum square residual from ARIMA model of Y_t is less than or equal to the sum square residual from ARIMA model of Z_t under the condition that the analysis of each time series Z_t, X_t and Y_t done individually.

Proof

The proof of the previous theorem depends on the proofs given in following steps:

Step (1) X_t and Y_t are uncorrelated

Proof

From equation (1) the DWT of Z is:

$$\mathbf{d} = WZ$$

by using hard threshold in equation (2) and let

$$\mathbf{dl}_i = \begin{cases} \mathbf{d}_i & \text{if } \mathbf{d}_i > |\lambda| \\ 0 & \text{otherwise} \end{cases}$$

and

$$\mathbf{d2}_i = \begin{cases} \mathbf{d}_i & \text{if } \mathbf{d}_i < |\lambda| \\ 0 & \text{otherwise} \end{cases}$$

it is clear that **d1** and **d2** are uncorrelated use the inverse discrete wavelet transformation (IDWT)

$$X = W^T \mathbf{d}1 \text{ and } Y = W^T \mathbf{d}2$$

where the matrix W is orthonormal, then

$$X^T Y = \mathbf{d}1^T W W^T \mathbf{d}2 = \mathbf{d}1^T \mathbf{d}2$$

Therefore

X_t and Y_t are uncorrelated

Step (2) proof that $\Delta^s Z_t = \Delta^s X_t + \Delta^s Y_t$, where Δ^s is the difference backward operator of order s.

Proof

The first difference backward of Z_t is:

$$\begin{aligned} \Delta Z_t &= Z_t - Z_{t-1} \\ &= X_t + Y_t - X_{t-1} - Y_{t-1} \end{aligned}$$

then

$$\Delta Z_t = \Delta X_t + \Delta Y_t$$

also, the second difference backward of Z_t is:

$$\begin{aligned} \Delta^2 Z_t &= Z_t - 2Z_{t-1} + Z_{t-2} \\ &= X_t + Y_t - 2(X_{t-1} + Y_{t-1}) + X_{t-2} + Y_{t-2} \\ &= \Delta^2 X_t + \Delta^2 Y_t \end{aligned}$$

and so on. The s^{th} difference backward is:

$$\Delta^s Z_t = \Delta^s X_t + \Delta^s Y_t$$

Step (3) $\hat{Z}_{t\,z} = \hat{X}_{t\,z} + \hat{Y}_{t\,z}$, where $\hat{Z}_{t\,z}, \hat{X}_{t\,z}$ and $\hat{Y}_{t\,z}$ are the forecasting of Z, X and Y respectively, at time t by using ARMA model on Z

Proof

Now we consider Z as a stationary time series. Firstly we give the proof for AR(p) case. In this case $\hat{Z}_{t\,z}$ is given by

$$\hat{Z}_{t\,z} = \sum_{s=0}^{p} \hat{\Phi}_{zs} Z_{t-s}$$

where

$\hat{\Phi}_{zs}$'s are the estimated parameters of AR(p) due to the analysis of time series Z

then

$$\begin{aligned} \hat{Z}_{t\,z} &= \sum_{s}^{p} \hat{\Phi}_{zs} (X_{t-s} + Y_{t-s}) \\ &= \sum_{s}^{p} \hat{\Phi}_{zs} X_{t-s} + \sum_{s}^{p} \hat{\Phi}_{zs} Y_{t-s} \\ &= \hat{X}_{t\,z} + \hat{Y}_{t\,z} \end{aligned}$$

Secondly we give the proof for the MA(q) case:

Let q =1, then

$$Z_t = \varepsilon_t - \theta \varepsilon_{t-1}$$

set $\varepsilon_0 = 0$

$$\varepsilon_t = Z_t + \sum_{s=1}^{t-1} (\theta)^s Z_{t-s}$$

set $\varepsilon_t = 0$

then

$$\hat{Z}_{t\,z} = -\sum_{s=1}^{t-1} (\hat{\theta}_z)^s Z_{t-s}$$

$$\hat{Z}_{t\,z} = -\sum_{s=1}^{t-1} (\hat{\theta}_z)^s (X_{t-s} + Y_{t-s})$$

$$\begin{aligned} \hat{Z}_{t\,z} &= -\sum_{s=1}^{t-1} (\hat{\theta}_z)^s X_{t-s} - \sum_{s=1}^{t-1} (\hat{\theta}_z)^s Y_{t-s} \\ &= \hat{X}_{t\,z} + \hat{Y}_{t\,z} \end{aligned}$$

where $\hat{\theta}_z$ is the estimated parameters of MA(1) due to the analysis of the time series Z.

Let q =2, from equation (6)

$$Z_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2}$$

set $\varepsilon_0 = \varepsilon_1 = 0$

then

$\varepsilon_2 = Z_2$, and

$$\varepsilon_t = Z_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2}$$

$$= Z_t + \theta_1 \left(Z_{t-1} + \theta_1 \varepsilon_{t-2} + \theta_2 \varepsilon_{t-3} \right)$$
$$+ \theta_2 \left(Z_{t-2} + \theta_1 \varepsilon_{t-3} + \theta_2 \varepsilon_{t-4} \right)$$

$$= Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2} + \theta_1^2 \varepsilon_{t-2} + 2\theta_1\theta_2 \varepsilon_{t-3} + \theta_2^2 \varepsilon_{t-4}$$
$$= Z_t + \theta_1 Z_{t-1} + (\theta_2 + \theta_1^2) Z_{t-2} + (2\theta_1\theta_2$$
$$+ \theta_1^3)\varepsilon_{t-3} + (\theta_2^2 + \theta_1^2\theta_2)\varepsilon_{t-4}$$

and so on

we can rewrite the error ε_t in the following form

$$\varepsilon_t = \sum_{s=0}^{t-1} f_{t-s}(\theta_1, \theta_2) Z_{t-s}$$

where

$f_t(\theta_1, \theta_2) = 1$

$f_{t-1}(\theta_1, \theta_2) = \theta_1$

$f_{t-2}(\theta_1, \theta_2) = \theta_2 + \theta_1^2$

and so on

set $\varepsilon_t = 0$, then

$$\hat{Z}_{t\,z} = -\sum_{s=1}^{t-1} f_{t-s}(\hat{\theta}_1, \hat{\theta}_2) Z_{t-s}$$
$$= -\sum_{s=1}^{t-1} f_{t-s}(\hat{\theta}_1, \hat{\theta}_2)(X_{t-s} + Y_{t-s})$$
$$= -\sum_{s=1}^{t-1} f_{t-s}(\hat{\theta}_1, \hat{\theta}_2) X_{t-s} - \sum_{s=1}^{t-1} f_{t-s}(\hat{\theta}_1, \hat{\theta}_2) Y_{t-s}$$
$$= \hat{X}_{t\,z} + \hat{Y}_{t\,z}$$

by the same way we can prove the MA(q) case as above.

Combining firstly and secondly proofs we can prove step 3 as follows:

let

$$\hat{Z}_{t\,z} = \sum_{s=1}^{t-1} g_{t-s}(\hat{\Phi}_{1z}, \hat{\Phi}_{2z}, \ldots, \hat{\Phi}_{pz}, \hat{\theta}_{1z}, \hat{\theta}_{2z}, \ldots, \hat{\theta}_{qz}) Z_{t-s}$$

where

$g_{t-s}(\Phi_{1z}, \Phi_{2z}, \ldots, \Phi_{pz}, \theta_{1z}, \theta_{2z}, \ldots, \theta_{qz})$ is function of ARMA(p,q)'s parameters on Z

then

$$\hat{Z}_{t\,z} = \sum_{s=1}^{t-1} g_{t-s}(\hat{\Phi}_{1z}, \hat{\Phi}_{2z}, \ldots, \hat{\Phi}_{pz}, \hat{\theta}_{1z}, \hat{\theta}_{2z}, \ldots, \hat{\theta}_{qz}) X_{t-s}$$
$$+ \sum_{s=1}^{t-1} g_{t-s}(\hat{\Phi}_{1z}, \hat{\Phi}_{2z}, \ldots, \hat{\Phi}_{pz}, \hat{\theta}_{1z}, \hat{\theta}_{2z}, \ldots, \hat{\theta}_{qz}) Y_{t-s}$$
$$= \hat{X}_{t\,z} + \hat{Y}_{t\,z}$$

Step (4) proof that $\sum \hat{X}_{t\,z} Y_t = 0$, where $\hat{X}_{t\,z}$ is the forecasting of X at time t by using ARMA model of Z

Proof

From properties of threshold the time series Y represent as the noise has mean zero i.e.

$\sum Y_t = 0$, since X and Y are uncorrelated

Therefore, $\hat{X}$ and Y are uncorrelated too, and the covariance between $\hat{X}$ and Y equal zero i.e.

$$\operatorname{cov}(\hat{X}, Y) = \frac{1}{n}\sum \hat{X}_t Y_t - \left(\frac{\sum \hat{X}}{n}\right)\left(\frac{\sum Y}{n}\right) = 0$$
$$= \sum \hat{X} Y = 0$$

Step (5) $\sum \hat{Y}_{t\,z} X_t = 0$, where $\hat{Y}_{t\,z}$ is the forecasting of Y at time t by using the ARMA model of Z

Proof

From step 4

$\sum Y_t = 0$

then, $\sum \hat{Y}_t$ must be approximately equal zero

Since X and Y are uncorrelated. Then, X and $\hat{Y}$ are uncorrelated i.e.

$$\operatorname{cov}(X, \hat{Y}) = \frac{1}{n}\sum X_t \hat{Y}_t - \left(\frac{\sum X}{n}\right)\left(\frac{\sum \hat{Y}}{n}\right) = 0$$
$$= \sum X \hat{Y} = 0$$

Step (6) $\sum \hat{Y}_{t\,z} \hat{X}_t = 0$

Proof

From steps 1, 4 and 5, we obtain;

$$\operatorname{cov}(\hat{X}, \hat{Y}) = \frac{1}{n}\sum \hat{X}_{tz} \hat{Y}_{tz} - \left(\frac{\sum \hat{X}_{tz}}{n}\right)\left(\frac{\sum \hat{Y}_{tz}}{n}\right) = 0$$
$$= \sum \hat{X}_{tz} \hat{Y}_{tz} = 0$$

Step (7)

$$\sum(Z_t - \hat{Z}_{t\,z})^2 = \sum(X_t - \hat{X}_{t\,z})^2 + \sum(Y_t - \hat{Y}_{tz})^2$$

Proof

$$\sum(Z_t - \hat{Z}_{t\,z})^2 = \sum(X_t + Y_t - \hat{X}_{t\,z} - \hat{Y}_{tz})^2$$

$$= \sum(X_t - \hat{X}_{t\,z})^2$$

$$+ \sum\left(Y_t - \hat{Y}_{tz}\right)^2 + 2\sum\left((X_t - \hat{X}_{t\,z})\left(Y_t - \hat{Y}_{tz}\right)\right)$$

from steps 1, 3, 4 and 5

$$\sum\left((X_t - \hat{X}_{t\,z})\left(Y_t - \hat{Y}_{tz}\right)\right) = 0$$

then

$$\sum(Z_t - \hat{Z}_{t\,z})^2 = \sum(X_t - \hat{X}_{t\,z})^2 + \sum(Y_t - \hat{Y}_{tz})^2$$

Step(8) finally,

$$\sum(X_t - \hat{X}_{t\,x})^2 + \sum(Y_t - \hat{Y}_{t\,y})^2 \le \sum(Z_t - \hat{Z}_{t\,z})^2$$

, where $\hat{X}_{t\,x}$ is the forecasting of X at time t by using ARMA model on X and $\hat{Y}_{t\,y}$ is the forecasting of Y at time t by using ARMA model on Y.

Proof

From step 3

$$\hat{X}_{t\,x} = \sum_{s=1}^{t-1} g_{t-s}(\hat{\Phi}_{1x}, \hat{\Phi}_{2x}, \ldots, \hat{\Phi}_{px}, \hat{\theta}_{1x}, \hat{\theta}_{2x}, \ldots, \hat{\theta}_{qx})X_{t-s}$$

where

$g_{t-s}(\Phi_{1x}, \Phi_{2x}, \ldots, \Phi_{px}, \theta_{1x}, \theta_{2x}, \ldots, \theta_{qx})$ is a function of ARMA(p,q)'s parameters on X

then

$$\sum(X_t - \sum_{s=1}^{t-1} g_{t-s}(\Phi_{1x}, \Phi_{2x}, \ldots, \Phi_{px}, \theta_{1x}, \theta_{2x}, \ldots, \theta_{qx})X_{t-s})^2$$

is minimum

Therefore

$$\sum(X_t - \sum_{s=1}^{t-1} g_{t-s}(\Phi_{1x}, \Phi_{2x}, \ldots, \Phi_{px}, \theta_{1x}, \theta_{2x}, \ldots, \theta_{qx})X_{t-s})^2$$

$$\le \sum(X_t - \sum_{s=1}^{t-1} g_{t-s}(\Phi_{1z}, \Phi_{2z}, \ldots, \Phi_{pz}, \theta_{1z}, \theta_{2z}, \ldots, \theta_{qz})X_{t-s})^2$$

i.e.

$$\sum(X_t - \hat{X}_{tx})^2 \le \sum(X_t - \hat{X}_{tz})^2 \qquad (10)$$

Similarly

$$\sum(Y_t - \hat{Y}_{ty})^2 \le \sum(Y_t - \hat{Y}_{tz})^2 \qquad (11)$$

then, by addition inequalities (10) and (11) we have

$$\sum(X_t - \hat{X}_{tx})^2 + \sum(Y_t - \hat{Y}_{ty})^2 \le \sum(X_t - \hat{X}_{tz})^2$$

$$+ \sum(Y_t - \hat{Y}_{tz})^2$$

Finally, from step (3) we find that:

$$\sum(X_t - \hat{X}_{tx})^2 + \sum(Y_t - \hat{Y}_{ty})^2 \le \sum(Z_t - \hat{Z}_{tz})^2 . \qquad (12)$$

This complete the proof.

Theorem2:

Suppose that the time series Z_t has constant variance and non seasonality and it is contaminated by correlated noise, using wavelet transformation, soft threshold technique and Bayesian rule, this time series is decomposed into tree time series as following:

$$Z_t = X_t + Y_t$$

where X_t and Y_t denote the signal and noise, respectively, and

$$X_t = X1_t + X2_t \qquad (13)$$

where X1 is the time series of X multiple by the shrinkage factor and X2 is the time series of remaining term, then the total sum square of the residuals from three ARIMA models of X1 , X2 and Y less than or equal to the total sum square residuals from the two ARIMA models of X_t and Y_t under the condition that the analysis of each time series $Z_t, X1_t, X2_t$ and Y_t done individually.

Proof

From treorem1 we will need prove then the total sum square residuals from two ARIMA models less than or equal sum square residuals ARIMA model of X

From (9)

$$X1_t = \eta_t X_t \quad (14)$$

and

$$X2_t = (1-\eta_t)X_t \quad (15)$$

where η_t the shrinkage factor

then

$$X2_t = (1-\eta_t)/\eta_t X1_t \quad (16)$$

and

$$X_t = [1+(1-\eta_t)/\eta_t]X1_t$$

i.e.

$$X_t = X1_t/\eta_t \quad (17)$$

Let

$$e = X - \hat{X}, \quad (18)$$

$$e1 = X1 - \hat{X}1, \quad (19)$$

and

$$e2 = X2 - \hat{X}2, \quad (20)$$

where $\hat{X}_t$, $\hat{X}1_t$ and $\hat{X}2_t$ are the forecasting of X, X1 and X2 at time t by using ARMA model on X , respectively
from eq. 16, .17,18,20

$$e_t = e1_t/\eta_t$$

$$e2_t = (1-\eta_t)/\eta_t\, e1_t \quad (21)$$

then

$$\sum_t e_t^2 = \sum_t (e1_t + e2_t)^2 = \sum_t e1_t^2/\eta_t^2 \quad (22)$$

Let

$$e'1 = X1 - \hat{X}'1,$$

and

$$e'2 = X2 - \hat{X}'2,$$

where $\hat{X}1_t$ and $\hat{X}2_t$ are the forecasting of X, X1 and X2 at time t by using ARMA model on X1 and ARMA model on X2, respectively
then

$$\sum_t (e'1_t + e'2_t)^2 = \sum_t e'1_t^2 + \sum_t e'2_t^2 + 2\sum_t e'1_t e'2_t$$

but
$e'1$ and $e'2$ are independent, then
$\sum_t e'1_t e'2_t = 0$

Therefore,

$$\sum_t (e'1_t + e'2_t)^2 = \sum_t e'1_t^2 + \sum_t e'2_t^2 \quad (23)$$

Since

$$\sum_t e'1_t^2 \le \sum_t e1_t^2 \quad (24)$$

And

$$\sum_t e'2_t^2 \le \sum_t e2_t^2 \quad (25)$$

by addition 24 and 25

$$\sum_t e'1_t^2 + \sum_t e'2_t^2 \le \sum_t e1_t^2 + \sum_t e2_t^2$$

From 25

$$\sum_t e'1_t^2 + \sum_t e'2_t^2 \le \sum_t e1_t^2 + \sum_t [(1-\eta_t)/\eta_t]^2 e1_t^2$$

i.e.

$$\sum_t e'1_t^2 + \sum_t e'2_t^2 \le \sum_t [\eta_t^2 + (1-\eta_t)^2] e1_t^2/\eta_t^2$$

i.e.

$$\sum_t e'1_t^2 + \sum_t e'2_t^2 \le \sum_t e1_t^2/\eta_t^2 + \sum_t [2\eta_t^2 - 2\eta_t] e1_t^2/\eta_t^2$$

i.e

$$\sum_t e'1_t^2 + \sum_t e'2_t^2 \le \sum_t e1_t^2/\eta_t^2 + 2\sum_t [1 - 1/\eta_t] e1_t^2$$

i.e.

$$\sum_t e'1_t^2 + \sum_t e'2_t^2 \le \sum_t e1_t^2/\eta_t^2 - 2\sum_t [1/\eta_t - 1] e1_t^2$$

but

$0 \le 1/\eta_t - 1$

so

$0 \le 2\sum_t [1/\eta_t - 1] e1_t^2$

then

$\sum_t e'1_t^2 + \sum_t e'2_t^2 \le \sum_t e1_t^2 / \eta_t^2$

from 22

$\sum_t (e'1_t + e'2_t)^2 \le \sum_t e^2$

(5) Application example

In this section, we will analyze the numbers of tourists whom coming to Egypt monthly through the period (1990 – 2006) as the time series by using wavelet technique (Ministry of tourisms in Egypt is the source of data). We use the MathCAD 14 and Minitab 13 for the analysis. The output results are displayed in the following:

5.1. ARIMA Model for Z

$$\Delta\hat{Z}_t = -0.01854 - 1.3229\,\Delta\hat{Z}_{t-1} - 0.9949\,\Delta\hat{Z}_{t-2}$$
$$+1.2879\,\varepsilon_{t-1} + 0.9853\,\varepsilon_{t-2}$$

where ΔZ_t is the first back difference of Z_t and $\varepsilon_t = \Delta Z_t - \Delta\hat{Z}_t$. The sum of square error from t=3 to t=192 as the following :

$$\sum_{t=3}^{192} (z - \hat{z})^2 = 5.44$$

where Z is analyze the numbers of tourists whom coming to Egypt monthly through the period (1990 – 2006) divided on 1000, take the natural logarithm and take the deference 12.

5.2. ARIMA Model for Y

$$\Delta\hat{Y}_t = 0.0000938 + 0.7407\,\varepsilon_{t-1} + 0.2455\,\varepsilon_{t-2}$$

where ΔY_t is the first back difference of Y_t and $\varepsilon_t = \Delta Y_t - \Delta\hat{Y}_t$.

5.3. ARIMA Model for X

$$\Delta\hat{X}_t = -0.004421 + 0.1522\,\Delta\hat{X}_{t-1}$$

where ΔX_t is the first back difference of X_t. Then we can use model

$$\hat{Z}_t = \hat{X}_t + \hat{Y}_t$$

Therefore

$$\hat{Z}_t = -0.0043272 + X_{t-1} + 0.1522\,\Delta X_{t-1} + Y_{t-1}$$
$$+0.7407\,\varepsilon y_{t-1} + 0.2455\,\varepsilon y_{t-2}$$

where $\varepsilon y_t = \Delta Y_t - \Delta\hat{Y}_t$. The sum of square error from t=3 to t=192 as the following

$$\sum_{t=3}^{192} (z - \hat{X} - \hat{Y})^2 = 4.52$$

The efficiency of using wavelet relative to without wavelet as the following:

$$\left[\frac{\sum_{t=3}^{192} (z - \hat{z})^2}{\sum_{t=3}^{192} (z - \hat{X} - \hat{Y})^2}\right] * 100 = 122.8\%$$

5.4. ARIMA Model for X1

$$\Delta^2\hat{X}1_t = 0.0002013 - 0.3702\,\varepsilon_{t-1} - 0.6103\,\varepsilon_{t-2}$$

where $\Delta X1_t$ is the second back difference of $X1_t$ and $\varepsilon_t = \Delta^2 X1_t - \Delta^2\hat{X}1_t$

5.5. ARIMA Model for X2

$$\Delta\hat{X}2_t = -0.0001167 \quad -0.8541\,\varepsilon_{t-1} - 0.1355\,\varepsilon_{t-2}$$

where $\Delta X2_t$ is the first back difference of $X1_t$ and $\varepsilon_t = \Delta X2_t - \Delta\hat{X}2_t$

Then we can use model

$$\hat{Z}_t = \hat{X}_t 1 + \hat{X}_t 2 + \hat{Y}_t$$

Therefore

$$\hat{Z}_t = 0.0001784 - + X1_{t-2} + 0.3702\varepsilon x1_{t-1}$$
$$-0.6103\varepsilon x1_{t-2} + X2_{t-1} - 0.8541\varepsilon x2_{t-1}$$
$$-0.1355\varepsilon x2_{t-2} + Y_{t-1} + 0.7407\varepsilon y_{t-1} + 0.2455\varepsilon y_{t-2}$$

where

$\varepsilon x1_t = \Delta^2 X1_t - \Delta^2 \hat{X}1_t$, $\varepsilon x2_t = \Delta X2_t - \Delta\hat{X}2_t$ and $\varepsilon y_t = \Delta Y_t - \Delta\hat{Y}_t$. The sum of square error from t=3 to t=192 as the following:

$$\sum_{t=3}^{192}(z - \hat{X}1 - \hat{X}2 - \hat{Y})^2 = 3.72$$

The efficiency of using wavelet and Bayesian rule relative to without wavelet and Bayesian rule as follows:

$$\left[\frac{\sum_{t=3}^{192}(z-\hat{z})^2}{\sum_{t=3}^{192}(z - \hat{X}1 - \hat{X}2 - \hat{Y})^2}\right] * 100 = 149.2\%$$

Where Z, X ,X1, X2 and Y represents the variables as the above theorem1 and theorem2.

Note that, the total sum of squares for X and Y ARIMA models less than the sum of squares for Z and the total sum of squares for X1, X2 and Y ARIMA models less than the total sum of squares for X and Y ARIMA models.

6. References

Abramovich, F., Bailey, T. C. and Sapatinas, T. (2000). Wavelet analysis and its statistical applications. The Statistician, 49, 1-29.

Abramovich, F. and Sapatinas, T. (1999). Bayesian approach to wavelet decomposition and shrinkage. In Bayesian Inference in Wavelet-Based Models (Lecture Notes in Statistics, Vol. 141), edited by P. M¨uller and B. Vidakovic, 33–50. Springer–Verlag, New York.

Burrus, C., Gonipath, R. and Guo, H. (1998). Introduction to Wavelets and Wavelet Transforms: A Primer. Englewood Cliffs, Prentice Hall.

Chipman, H. A., Kolaczyk, E. D. and McCulloch, R.E. (1997). Adaptive
Bayesian wavelet shrinkage. J. Am. Statist. Assoc., 92, 1413–1421.

Clyde, M. and George, E. I. (1999). Empirical Bayes estimation in wavelet nonparametric regression. In Bayesian Inference in Wavelet-Based Models (Lecture Notes in Statistics, Vol. 141), edited by Mller, P. and Vidakovic, B. Oppenheim, 309–322. Springer–Verlag, New York.

Clyde, M. and George, E. I. (2000). Flexible empirical Bayes estimation for wavelets. J. R. Statist. Soc. B, 62, 681–698.

Daubechies, I. (1992). Ten Lectures on Wavelets. Philadelphia: Society for Industrial and Applied Mathematics.

Donoho, D. L. (1993) Non-linear wavelet methods for recovery of signals, densities and spectra from indirect and noisy data. Proc. Symp. Appl. Math., 47, 173-205.

Donoho, D. L. (1995) Nonlinear solution of linear inverse problems by wavelet-vaguelette decomposition. Appl. Comput. Harm. Anal., 2, 101-126.

Donoho, D. L. and Johnstone, I. M. (1994) Ideal spatial adaptation by wavelet shrinkage. Biometrika, 81, 425-455.

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

Donoho, D. L., Johnstone, I. M., Kerkyacharian, G. and Picard, D. (1995) Wavelet shrinkage: asymptopia (with discussion). J. R. Statist. Soc. B, 57, 301-369.

Johnstone, I. (1994). Minimax bayes, asymptotic minimax and sparse wavelet priors. In Statistical Decision Theory and Related Topics, V (S. Gupta and J. Berger, eds.). Springer-Verlag, New York, 303–326.

Johnstone, I. M. and Silverman, B. W. (1997) Wavelet threshold estimators for data with correlated noise. J. R. Statist. Soc. B, 59, 319-351.

Johnstone, I. M. and Silverman, B. W. (1998). Empirical Bayes approaches to mixture problems and wavelet regression. Technical Report, School of Mathematics, University of Bristol, Bristol.

Johnstone, I. M. and Silverman, B. W. (2005). Empirical Bayes selection of wavelet thresholds. Annals of Statistics, 33(4), 1700–1752.

Mallat, S. G. (1989) A theory for multiresolution signal decomposition: the wavelet representation. IEEE Trans. Pattn Anal. Mach. Intell., 11, 674-693.

Mallat, S. (1999). A Wavelet Tour of Signal Processing. 2nd ed. Academic Press, San Diego.

Nason, G. and Silverman, B. (1995). The stationary wavelet transform and some statistical applications. In Wavelets and Statistics (A. Antoniadis and G. Oppenheim, eds.), vol. 103 of Lect. Notes Statist. Springer-Verlag, New York, 281–300.

Nason, G. P. and von Sachs, R. (1999) Wavelets in time series analysis. Phil. Trans. R. Soc. Lond. A, 357, 2511-2526.

Priestley, M. B. (1996) Wavelets and time-dependent spectral analysis. J. Time Ser. Anal., 17, 85-104.

Oh, H.-S. and Lee, T. C. M. (2005). Hybrid local polynomial wavelet shrinkage: wavelet regression with automatic boundary adjustment. Comput. Statist. Data Anal., 48 809–819.

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

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

تحول وتكسير المويجات باستخدام نموذج أريما

شبعان عبد الحميد شعبان
معهد الدراسات الإحصائية
جامعة القاهرة

أيمن عرابي عبد اللطيف
كلية علوم الإدارة
جامعة الملك سعود

(قدم للنشر في ١٠/ ١/ ١٤٣١هـ؛ وقبل للنشر في ٢٤/ ٥/ ١٤٣١هـ)

الكلمات المفتاحية: تحول المويجات، والحدود، الارتداد التلقائي، متوسط التحرك، نموذج أريما، مجموع مربع الخطأ.

ملخص البحث. الهدف من هذا البحث هو دراسة تأثير مفلتر المويجات على بيانات سلسلة الزمن. باستخدام تقنية تحول المويجات والحدود الصعبة. وقد تم تحليل نموذج أريما إلى نموذجين، كما تمت مناقشة العلاقة بين حصيلة مربع الخطأ الناتج عن نموذج أريما ومحصلة النموذجين باستخدام تقنية تحول المويجات والحدود السهلة، كما تم تحليل نموذج أريما إلى ثلاث نماذج ومناقشة العلاقة بين حصيلة مربع الخطأ الناتج عن نموذج أريما ومحصلة الثلاث نماذج.