Maximum Overlapping Discrete Wavelet Transform in Forecasting Banking Sector

S. Al Wadi, Abdulkareem Hamarsheh and Hazem Alwadi

Department of mathematics, Faculty of Science
Al Hussien bin Talal University, Ma'an, Jordan
sadam_alwadi@yahoo.co.uk, abkmsaleh@yahoo.com

School of Social Science
Jordan University of Science and Technology, Irbid, Jordan
hazemalwadi@yahoo.com

Copyright © 2013 S. Al Wadi et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

In this paper, we present the advantages of Maximum overlapping Discrete Wavelet Transform (MODWT) in improving the forecasting accuracy financial time series data. Amman stock market (ASE) in Jordan was selected as a tool to show the ability of MODWT in forecasting financial time series, using Banking sector. Experimentally, this article suggests a novel technique for forecasting the banking data based on MODWT and ARIMA model. Daily return data from 1993 until 2009 is used for this study.

Keywords: Fluctuations, MODWT, ARIMA model, forecasting, Banking data

1. Introduction

During the last three decades the banking sector was very stable in Jordan, very few number of banks have exited from the market. Since this sector has a very nice environment for Growing, well capitalized, liquid and profitable and privately owned, Open to external investors. Moreover, this sector has Comprehensive banking services: retail, corporate, Islamic, and e-banking and

These reasons motivate the researches to focus in analyzing the banking sector in Jordan as in [18], and improving the forecasting accuracy using MODWT as in this article. In recent years, stock markets forecasting is required for the investors and it has got very high attention in financial time series and financial researchers. The accurate forecasting of financial prices is an important issue in investment decision making.

However, financial time series data appears noisy and non-stationary [6,12]. The noise characteristic indicates the unavailability of complete information from past behavior of financial markets to fully capture the dependency between future and past prices. The information that is excluded in the forecasting model is considered as noise while the non-stationary characteristic indicates the distribution of financial time series changing over time. Therefore, financial time series forecasting is considered as one of the most challenging tasks of time series analysis.

There are many forecasting models that have been used in the forecasting literature, such as; simple moving average, linear regression, neural network, ARMA model and ARIMA model. In order to provide estimates for the future, these models analyze the historical data. Usually time series are not deterministic series. In this regard, MODWT will be used as an effective method to forecast the future events in banking sector in ASE, then selecting the best WT function in forecasting processes.

This paper is organized as follows. The next section describes the mathematical and literature review. Section 3 provides a description of data set. Section 4 describes the methodology. The experimental results are presented to demonstrate the effectiveness of MODWT in using banking data will be presented in section 5. In Section 6 we summarize our contributions and mention the conclusion.

2. Mathematical and Literature Review

2.1 Wavelet Transform
Wavelet analysis is a mathematical model that transforms the original signal (especially with time domain) into a different domain for analysis and processing [13,16]. This model is very suitable with the non-stationary data, i.e. mean and autocorrelation of the signal are not constant over time, that is well know, most of the financial time series data is non-stationary, that is why we applied MODWT.

In the MODWT and discrete wavelet transform (DWT) case, consider that the time domain is the original domain. Although, these models of transformation process from time domain to time scale domain, these processes are known as signal decomposition because a given signal is decomposed into several other
Maximum overlapping discrete wavelet transform

signals with different levels of resolution. These processes allow recovering the original time domain signal without losing any information. MODWT has reverse process which is called the inverse MODWT or signal reconstruction [4].

The MODWT and DWT is implemented using a multiresolution pyramidal decomposition technique. In fact, a recorded digitized time signal \( S(n) \) can be analyzed into its detailed \( cD1(n) \) and smoothed (approximations) \( cA1(n) \) signals using high-pass filter (HiF-D) and low-pass filter (LoF-D), respectively. High-pass filter has a band-pass response. Consequently, the filter signal \( cD1(n) \) is a detailed coefficient of \( S(n) \) and contains higher frequency components. While the approximation signal \( cA1(n) \) has a low-pass frequencies filter response. The decomposition of \( S(n) \) into \( cA1(n) \) and \( cD1(n) \) is the first scale decomposition. Inversely, that is possible to perform the original signal from the approximations and details coefficients.

In this paper we will focus in the most famous types of discrete DWT which are HWT (Haar wavelet transform) and dWT (Daubechies wavelet transform), then compare them with MOWDT dynamic. The wavelets having compact support or narrow window function are suitable for local analysis of the signal. dWT, HWT and MODWT are compactly supported orthonormal [1].

Definition: [2,7] DWT can be defined by the following function:

\[
\psi_{j,k}(t) = 2^j \psi(2^j t - k),
\]

\( j, k \in \mathbb{Z}; z = \{0,1,2,....\}. \)

where \( \psi \) is a real valued function having compactly supported, and \( \int_{-\infty}^{\infty} \psi(t)dt = 0 \)

Generally, the MODWT were evaluated by using dilation equations, given as:

\[
\phi(t) = \sqrt{2} \sum_{k} l_k \phi(2t - k), \quad \psi(t) = \sqrt{2} \sum_{k} h_k \phi(2t - k).
\]

Father and mother wavelets were defined by the last two equations where \( \phi(2t-k) \) represents the father wavelet, and \( \psi(t) \) represents the mother wavelet. Father wavelet defines the lowpass filter coefficients \( (h_k) \) and high pass filters coefficients \( (l_k) \) are defined as [3].

\[
l_k = \sqrt{2} \int_{-\infty}^{\infty} \phi(t)\phi(2t-k)dt, \quad h_k = \sqrt{2} \int_{-\infty}^{\infty} \psi(t)\psi(2t-k)dt.
\]

HWT is the oldest and simplest example in the DWT and is defined as:
\[
\psi^H(t) = \begin{cases} 
1, & 0 \leq t \leq \frac{1}{2} \\
-1, & \frac{1}{2} \leq t \leq 1 \\
0, & \text{Otherwise}
\end{cases}
\]

where \( \psi_1(\omega) \) presents the HWT. HWT is the simplest and oldest DWT; it was improved by dWT in 1992. He developed the frequency – domain characteristics of the HMODWT. However, we do not have a specific formula for this method of MODWT. So, we tend to use the square gain function of their scaling filter, the square gain function was defined as [9].

\[
g(f) = 2 \cos^l(\pi f) \sum_{l=0}^{\frac{l-1}{2}} \left( \sum_{l=1}^{l+1} \sin^2(\pi f) \right)
\]

where \( l \) : Positive number and represents the length of the filter, for more details and examples about the MODWT mathematical model and its applications refer [2,8, 9].

As critically review. First, According to the past decade, few researchers has paying attention on the application of MODWT to solve financial issues such as study the banking data and forecasting in ASE. Second, for the last 10 years a number of comparative studies using different methodologies have been carried out using various MODWT functions alone as well as in combination with other models. However, no research conducted which is comparing between DWT and MODWT in content of forecasting area using banking data. Therefore, the forecasting model will be used is ARIMA model, since this model is easy to combine with other models and can be used for long run of dataset.

2.2. ARIMA model
ARMA is a suitable model for the stationary time series data, although most of the software uses least square estimation which requires stationary. To overcome this problem and to allow ARMA model to handle non-stationary data, the researchers investigate a special class for the non-stationary data. This model is called Auto-regressive Integrated Moving Average (ARIMA). This idea is to separate a non-stationary series one or more times until the time series becomes stationary, and then find the fit model. ARIMA model has got very high attention in the scientific world. This model is popularized by George Box and Gwilym Jenkins in 1970s [4]. There are a huge number of ARIMA models; generally there are ARIMA (p, q, d) where: P: order of autoregressive part (AR), d: degree of first differentiation (I) and q: order of the first moving part (MA). Note that, if there is no differencing been done (d = 0), Then ARMA model can be got from ARIMA model [5].

The general mathematical ARIMA model can be defined as [11]:
Maximum overlapping discrete wavelet transform

\[ W_t = \mu + \frac{\beta(v)}{\varepsilon(v)} a_t. \]

Where:
- \( t \): Indexes time.
- \( W_t \): The response series \( Y_t \) or a difference of the response series.
- \( \mu \): The mean term.
- \( \nu \): The backshift operator; that is, \( \nu X_t = X_{t-1} \).
- \( \varepsilon(v) \): The autoregressive operator.
- \( \beta(v) \): The moving-average operator
- \( a_t \): The independent disturbance, also called the random error.

The model building process involves the following steps; Model identification, Model parameter estimation, Model Diagnostics and Forecasting [15].

3. Data Description

In order to illustrate the effectiveness of MODWT and HWT and dWT the Amman Stock Market data sets are selected for discussion. We consider a daily return data for the time period from April 1993 (the days when stock exchanges were open) until December 2009 with a total of 4096 observations. In HWT and dWT the total number of observations for mathematical convenience is suggested to be divisible by \( 2^j \). It means that the data should satisfy the condition of observations= \( 2^j \), whereas this condition is not necessary for MODWT [7,17].

4. Methodology

We will present the criteria which have been used to make a fair comparison, and then the framework comparison will be presented with more details.

4.1 Prediction accuracy criteria

We have been adopted to compare the performance of the models within two types of accuracy criteria [14]:

1- Root mean squared error (RMSE).

\[ \text{RMSE} = \sqrt{\frac{\sum_{i=1}^{N} (\text{actual value-predicted value})^2}{N}}. \]

2- Mean absolute percentage error (MAPE).

\[ \text{MAPE} = \frac{1}{N} \sum_{i=1}^{N} \left| \frac{\text{actual value-predicted value}}{\text{actual value}} \right| \times 100\% \]

Where \( N \) represents the number of observations used for analysis.

4.2. Comparison framework:
The MODWT converts the data into two sets; approximation series (CA1 (n)) and details series (DA1 (n)). These two series present a better behavior. i.e. More stable in variance and no outliers than the original price series, then, they can be predicted more accurately. The reason for the better behavior of these two series is the filtering effect of the MODWT. In this paper the Approximation series has been used since this series behave as the main component of the transform. The procedure explained in this paper is as follows:
First, decompose through the MODWT, HWT and dWT the available historical return data. Second, Use a specific ARIMA model fitted to each one of the Approximation series to make the forecasting. Third, this technique is compared with an ARIMA model used directly to forecast the return data series by using the above criteria. Fourth, comparing the results for the entire model used in this paper.

5. Experimental Results

In this paper, the minimum value of MSE, RMSE and MAE is considered to select the best ARIMA model of the daily return data. All choices of ARIMA models for the return data are included in this test between (0,0,0) and (2,2,2). If we choose more than two, then there are more complicated conditions that should be satisfied. Also, if \( p \) and \( q \) are more than two, then Autocorrelation function (ACF) and partial Autocorrelation function (PACF) will be presented as an exponential decay. This means that ARIMA model becomes worthless and there is no importance.

<table>
<thead>
<tr>
<th>Statistical fit</th>
<th>Value after transform via MODWT</th>
<th>Value after transform via HWT</th>
<th>Value after transform via dWT</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSE</td>
<td>0.4</td>
<td>0.88</td>
<td>0.67</td>
</tr>
<tr>
<td>MAE</td>
<td>3.5%</td>
<td>10.44%</td>
<td>8.7%</td>
</tr>
</tbody>
</table>

The banking data for Amman stock market has been used as a case study. Price forecasting is performed using daily data. Moreover, for the sake of fair comparison the same sample data is selected. (From 1993-2009). The fit ARIMA model with HWT for the original return data is considered with RMSE equal to 0.88 as presented in Table 1, which is the worst model in this case study. While the fit ARIMA model for the transform data by using dWT is selected with RMSE
equal to 0.67 as presented in Table 1 also. Whereas, the forecasting is the best using MODWT since it has RMSE equal to 0.4. In order to further corroborate our findings and hence conclusions, this study also used another statistical criterion to analyze the small difference in RMSE, which is, MAPE. Table 1 shows the results using MAPE. When the results of these criteria (RMSE and MAPE) are taken collectively, the forecasting accuracy is improved using MODWT combined with a suitable ARIMA model compared to the forecasting accuracy using ARIMA model with HWT and dWT respectively.

6. Conclusion

This study implemented MODWT on ASE banking data. The success application of this study is in removal the outliers and irregular data. Therefore, in this empirical study the sample data set was experimentally tested in terms of forecasting accuracy and decomposition levels. The purpose of doing so was to find out, whether MODWT combined with a suitable ARIMA model produces more accurate forecasting compared to HWT and DWT with ARIMA model. And also to find out the event that occurred in banking sector in ASE during past 20 years. The findings using statistical prediction criteria RSME and MAPE have revealed that MODWT combined with ARIMA model is more accurate at forecasting. Therefore, this particular finding means that the forecasting can be improved using this modern model (MODWT).

References


Received: April 21, 2013