# - CHAPTER 3 -

# RESEARCH METHODOLOGY

## 3.0 Overview

Based on the literature review in chapter 4, there seems to be no or little research done to analyze the three currency pairs' (EURO/USD, GBP/USD and YEN/USD) behavior during the period one hour before and after 1PM GMT. Furthermore, the use of wavelet method for high frequency data analysis is very much limited. However in recent years, financial analysts are beginning to see the benefits of wavelet methods in analyzing financial data. That is it's capable of de-noising the data and analyzed in various time scale.

Starting part of this chapter shall describe the theory behind discrete wavelet transform and Haar filter used in the analysis of high frequency data captured. The results generated from the analysis are presented.

## 3.1 Discrete Wavelet Transform

The discrete wavelet transform (DWT) is a mathematical tool that projects a time series onto a collection of orthonormal basis functions (wavelets) to produce a set of wavelet coefficients. These coefficients capture information from the time series at different frequencies at distinct times. For a function $f$

defined on the entire real line, a suitably chosen mother wavelet function *ψ* can be used to expand *f* as

$$f(t)=\sum_{j=-\infty}^{\infty}\sum_{k=-\infty}^{\infty} w_{jk}\,2^{j/2}\psi(2^{j}t-k) \qquad (1)$$

Where the functions ψ $(2^{j}t - k)$ □□□ are all orthogonal to one another.

The coefficient $w_{jk}$ conveys information about the behavior of the function f concentrating on effects of scale around $2^{-j}$ near time k $\times 2^{-j}$.

The DWT can effectively compress a wide range of signals – a large proportion of DWT coefficients can actually be set to zero without appreciable loss of information. It can deal well with heterogeneous and transient behavior that makes it so attractive for financial time series analysis. However, one problem associated with the application of the DWT for time series analysis is that it suffers from a lack of translation invariance. This means that circularly shifting a time series will not necessarily shift its DWT coefficients in a similar manner.

### 3.2 Maximum Overlap Discrete Wavelet Transform (MODWT)

MODWT is an enhanced version of DWT whereby the main difference between both is that MODWT is able to take in any sample size, where else DWT is limited by sample size of multiple of 2s.

This problem can be tackled by means of a highly redundant non-orthogonal transform called the maximal overlap discrete wavelet transform (MODWT). For a redundant transform like the MODWT, an N samples input time series will have an N samples resolution scale for each resolution level. Therefore, the features of wavelet coefficients in a multi resolution analysis (MRA) will be lined up with the original time series in a meaningful way.

For a time series X with arbitrary sample size N, the $j^{th}$ level MODWT wavelet ($\widetilde{W}_j$) and scaling ($\widetilde{V}_j$) coefficients are defined as

$$\widetilde{W}_{j,t} \equiv \sum_{l=0}^{L_j-1} \widetilde{h}_{j,l} X_{t-l \bmod N} \tag{2a}$$

$$\widetilde{V}_{j,t} \equiv \sum_{l=0}^{L_j-1} \widetilde{g}_{j,l} X_{t-l \bmod N} \tag{2b}$$

where $\widetilde{h}_{j,l} \equiv h_{j,l} / 2^{j/2}$ are the MODWT wavelet filters, and $\widetilde{g}_{j,l} \equiv g_{j,l} / 2^{j/2}$ are the MODWT scaling filters.

For a time series *X* with *N* samples, the MODWT yields an additive decomposition or MRA given by,

$$X = \sum_{j=1}^{J_0} \widetilde{D}_j + \widetilde{S}_{J_0} \tag{3}$$

Where

$$\widetilde{D}_{j,t} = \sum_{l=0}^{N-1} \widetilde{h}^{\circ}_{j,l} \widetilde{W}_{j,t+l \bmod N} \qquad (4a)$$

$$\widetilde{S}_{j,t} = \sum_{l=0}^{N-1} \widetilde{g}^{\circ}_{j,l} \widetilde{V}_{j,t+l \bmod N} \qquad (4b)$$

According to Equation (3), at a scale j, a set of coefficients $\{D_j\}$ is obtained, each with the same number of samples (N) as in the original signal (X). These are called wavelet "details" and they capture local fluctuations over the whole period of a time series at each scale. The set of values $S_{J0}$ provide a "smooth" or overall "trend" of the original signal. Adding $D_j$ to $S_{J0}$, for j = 1, 2,…, $J_0$, gives an increasingly more accurate approximation of the original signal. This additive form of reconstruction enables the prediction of each wavelet sub-series ($D_j$, $S_{J0}$) separately and adds the individual predictions to generate an aggregate forecast.

Since the intention is to make one-step-head predictions, MODWT should be performed in such a way that the wavelet coefficients (for each level) at time point n should not be influenced by the behavior of the time series beyond point n. That is to perform the MODWT incrementally where a wavelet coefficient at a position n is calculated from the signal samples at positions less than or equal to n, but never larger. This will gives the flexibility of dividing the wavelet coefficients for making one-step-ahead predictions in the same way as it would for the original signal.

### 3.3 Haar Filter and Wavelet

A set of times series data that undergoes wavelet transformation using Haar Filter creates a Haar wavelet, which is the simplest type of wavelet, as explained by Walker (1999). In discrete form, Haar wavelets are related to a mathematical operation called Haar transform. The Haar transform serves as a prototype for all other wavelet transforms. Like all wavelet transform, the Haar transform decomposes a discrete signal into two sub-signals of half its length. One sub-signal is a running average or trend. The other sub-signal is a running difference or fluctuation.

The Haar wavelet transform has a number of advantages such as:

i) It is conceptually simple and will be explained later.

ii) It is fast and memory efficient because it can be calculated in place without a temporary array.

iii) It is exactly reversible without the edge effects that are a problem with other wavelet transforms.

However, there are also limitations, which can be a problem with for some applications. In generating each of averages for the next level and each set of coefficients, the Haar transform performs an average and difference on a pair of values. Then the algorithm shifts over by two values and calculates another average and difference on the next pair. The high frequency coefficient spectrum should reflect all high frequency changes. The Haar window is only two elements wide. If a big change takes place from an even value to an odd value, the change will not be reflected in the high frequency coefficients. So Haar wavelet transform is not useful in compression and noise removal of audio signal processing.

Conceptually, wavelet transform is an inner product of the time series with the scaled and translated wavelet $\varphi(x)$ usually an $n^{th}$ derivative of a smoothing kernel $\theta(x)$. The scaling and translation actions are performed by two parameters; the scale parameter ‘s’ adapts the width of the wavelet to the microscopic resolution required, thus changing its frequency contents and the location of the analyzing wavelet is determined by the parameter ‘b’:

$$Wf(s,b) = < f, \varphi > (s,b) = \frac{1}{s}\int_{s} dx f(x)\varphi(\frac{x-b}{s}) \qquad (5)$$

Where $s,b \in R$ and $s > 0$ for the continuous version (CWT) or is taken on a discrete, usually hierarchical grid of value $s_i$, $b_i$ for discrete version (DWT). $\Omega$ is the support of the f(x) or the length of the time series.

The choice of smoothing kernel $\theta(x)$ and related wavelet $\varphi(x)$ depends on the application and on the desired properties of the wavelet transform. In the data analysis of this thesis, a simple block smoothing function was carried out to have the optimal localization both in frequency and position of the related wavelets.

$$\theta(x) = \begin{cases} 1 - for & 0 \le x \le 1 \\ 0 - otherwise \end{cases} \qquad (6)$$

The wavelet obtained from this kernel is defined on finite support and are referred as Haar:

1 for $0 \le x \le \frac{1}{2}$

$$\theta(x) = \begin{cases} -1 & \text{for} \quad \frac{1}{2} \le x \le 1 \\ 0 & \text{otherwise} \end{cases} \qquad (7)$$

For a particular choice of rescaling and position shift parameters (dyadic pyramidal scheme) the Haar system constitutes and orthonormal basis:

$$\varphi_{m,n}(x) = 2^{-m}\varphi(2^{-m}x - n) \qquad (8)$$

Where $m > 0$ and $n = 0 \dots 2^m$

Assume an arbitrary time series f = $\{f_i\}$, i = 1…$2^N$ on the normalized support $\Omega(f)=[0,1]$. Using the orthonormal basis described, the function f can be represented with the linear combination of Haar wavelets:

$$f = f^0 + \sum_{m=1}^{N}\sum_{l=0}^{2^m} c_{m,l}\varphi_{m,l} \qquad (9)$$

Where $f_0$ is the coarsest approximation of the time series $f^0 = \langle f,\theta \rangle$, and each coefficient $c_{m,l}$ of the representation can be obtained as $c_{m,l} = \langle f,\varphi_{m,l} \rangle$. The $f^j$ of the time series $f$ with the smoothing kernel $\theta_{j,k}$ forms a ladder of multi-resolution approximations:

$$f^{j-1} = f^j + \sum_{k=0}^{2j} \langle f,\varphi_{j,k} \rangle \varphi_{j,k} \qquad (10)$$

Where $f^j = \langle f,\theta_{j,k} \rangle$ and $\theta_{j,k} = 2^{-j}\theta(2^{-j}x-k)$.

Thus it is then possible to move from one approximation level j-1 to another level j by simply adding the detail contained in the corresponding wavelet coefficient $c_{j,k}$, where k = 0... $2^j$.

### 3.4 Data Collection

For all the three currency pair, minute by minute data for every Monday starting from start of June 2008 to end of December 2008 is used for the wavelet analysis. This period of time is chosen because it was the period where the sub-prime crisis in USA peaked and investors' decision making went hay wire causing over selling and over buying.

The minute by minute trading data is collected from an online forex historical data repository, http://www.forexrate.co.uk. As the site is a valid and certified forex trading portal, therefore it can be safe to make the assumption that the data is reliable. The initial source for data was from Bloomberg but they cannot provide the minute by minute data on the time period that required.

Using this portal, minute by minute data for every Monday from June 2008 to December 2008 from one hour before 1pm GMT to one hour after 1pm GMT is captured. As the wavelet analysis is be done up to the $7^{th}$ level, the total data entries requires is $2^7$ which is 128 entries.

### 3.5 Data Cleansing

After the collection of data, an application is written to check to ensure that data entry is available for every minute for every currency pair. Entries (minute) that do not have data will be manually filled up using the average value based on the value of before and after that particular minute. For example, if data is missing in minute X, then the value will be calculated based on

$$\text{Data in Minute X} = \left( \ [(\text{Data in Minute X} + 1) \ (\text{Data in Minute X} - 1)] \ / 2 \ \right)$$

## 3.6 MODWT Wavelet Toolbox for Matlab (Software)

The MODWT Wavelet Toolbox for Matlab is add-on software module which is downloaded from www.atmos.washington.edu/~wmtsa/. It is an implementation of the wavelet-based techniques for the analysis of time series documented in a book by Percival, D. B. and A. T. Walden (2000) called Wavelet Methods for Time Series Analysis (WMTSA).

## 3.7 Analysis Procedures

Once MODWT and MATLAB are setup, the data collected as explained in earlier chapter are then imported using MATLAB and the test is done using the MODWT wavelet toolbox. A program script is written to load the data and parse it to the MODWT wavelet toolbox using MATLAB.

The procedures in chronological order:

i. Decompose the forex data using MODWT toolkit up to the 7th level (D7, which requires 128 observations of data). This is done using the toolkit's modwt () with provided by the toolkit. (**Note**: The definition of the time period in the analysis is D1: 1st to 2nd minute, D2: 2nd to 4th minute, D3: 4th to 8th minute, D4: 8th to 16th minute, D5:16th to 32nd minute, D6: 32nd to 64th minute and D7: 64th to 128th minute)

ii. Execute Multi Resolution Analysis (MRA) on the decomposed data. This is done using the toolkit's imodwt_mra ().

iii. The results of the decomposition are then stored in an excel file according to the date of data. This is then used for the correlation analysis and is done using the toolkit's modwt_wcor ().

iv. The correlation results are then stored in an excel file according to the date of the data.

v. Using the correlation results, the graphs are plotted for further analysis.

The steps in the procedure above are repeated for following three conditions/tests:

i. Spread of each currency pair

ii. Volume of each currency pair

Using the MODWT results from the test done, the correlation analysis is done to investigate the following:

i. Spread of the currency pair combinations.

ii. Volume of the currency pair combinations.

iii. Spread and volume of each currency pair.

Note: "Currency pair combinations" is referring to "EURO/USD and GBP/USD", "GBP/USD and YEN/USD" and "EURO/USD and YEN/USD"

Charts for the correlation analysis are documented in chapter four. Descriptive statistics for the data can be found in Appendix A. The Matlab application script used for this analysis is in Appendix D.

### 3.8 Summary

Using MATLAB and the MODWT toolkit, the procedures produced the wavelet output for D1 to D7. Using the output, correlation analysis was done for all individual currency pairs and currency pair combinations. In order to have a clearer view on the correlation pattern, average values are considered and used to plot charts. From these charts, the correlation value and trend are more interpretable and is present in chapter four.