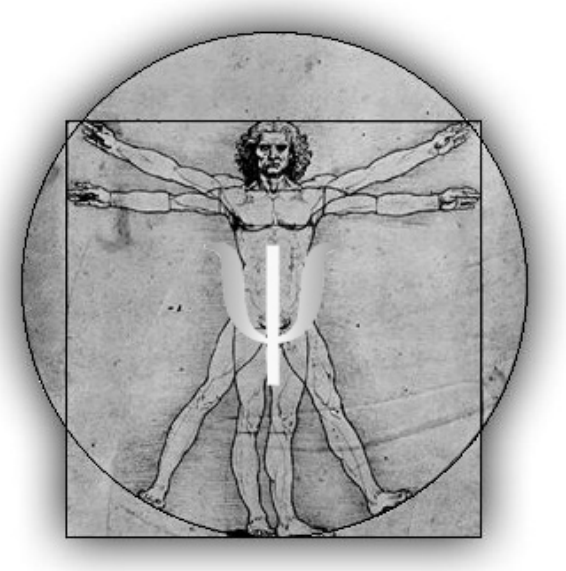
Yuri Hamburger

Wavelet-based Value At Risk Estimation

A multiresolution approach



Master thesis Informatics & Economics
Erasmus University Rotterdam
December 2003

Supervision by dr. ir. J. van den Berg and dr. ir. U. Kaymak

”He who loves practice without theory is like the sailor who boards ship without a rudder and compass and never knows where he may be cast.” – Leonardo da Vinci, 1452-1519

Contents

List of Figures

List of Tables

Preface

A master thesis. A master's piece. At least, that is what it should be. One year ago, the search for an intellectual challenge started and, with the subject of this thesis, an intellectual challenge was found. However, a psychological challenge was also found. Writing a thesis can feel like a mountaineer spotting the top of the mountain, like a sailor on a vessel spotting land. Writing a thesis can also feel like the same mountaineer realizing there's a large valley in between or the captain of the vessel saying it's still a long trip. It can poison your mind, it can haunt you in your dreams and fill them with agony. But in the end, it is inspiring and the accomplishment makes you feel proud.

However, I never could have accomplished this work without the help and support of many. Therefore, I'd like to say thanks to those who gave this help and support.

First of all, I'd like to thank my supervisors, Uzay Kaymak and Jan van den Berg. They were the light in the darkness. They challenged me to look deeper into the theory, view it from a different perspective. They always gave critical and honest feedback to my work, which could take hours, sometimes. I want to thank them in special, as they inspired and helped me to enter the Ph.D. program.

In special, I'd like to thank my parents. Neither could I have started this study, nor could I have found the courage to start the writing of this thesis and the start of an even larger challenge, without their faith and unconditional support.

I would like to thank my friends for unpoisoning my mind in those dark hours.

Finally, I'd like to thank Claudia, for her faith and beautiful smile.

Yuri Hamburger

December 2003.

Chapter 1

Introduction

1.1 Risk

> *Prediction is very difficult, especially of the future*. (Niels Bohr).

Exactly knowing what the future will bring is the dream of every decision maker; it implies certainty. A decision maker having exact knowledge of the future at his disposal, is said to have complete information. Unfortunately, coping with finance or investment related problems and having complete information is a luxury and rather an unlikely situation. Situations characterized by incomplete information can be separated into two groups: situations of uncertainty and situations of *risk* (van den Bergh, Hallerbach, van der Meulen, van der Sar, Schauten, Smit, Soppe, Spronk, Steenbeek and Vorst 1999).

A situation is characterized as uncertain, when the decision maker doesn't have the information at his disposal which allows him to objectively diagnose the probabilities on the states of nature influencing the final result of his decision. Uncertainty can also be a result of not being aware of specific factors and the way they influence the result of a decision. The latter is considered as ignorance (van den Bergh et al. 1999).

We refer to risk, however, when the decision maker has indeed no certainty about specific future states of nature, but is aware of the factors which are on influence on the result of the decision. In addition, the decision maker is able to assign probabilities to each of the possibilities (van den Bergh et al. 1999).

Consider for example the following game which is played with an unbiased dice and a starting balance of €10: when the result of a throw with the dice is lower than 4, the player has to pay 10% of the starting balance for a 1, 20% for a 2 and 30% for a 3. When the result is higher than or equal to four, the player receives 40% of the starting balance for a 4, 50% for a 5 and 60% for a 6. In this situation, the player is aware of the factor influencing the result - the occurrence of each of the six possible outcomes of the throw - and the player is able to assign a probability equal to one-sixth to each of these outcomes.

The expected return of playing this game then equals a weighted average of the possible outcomes:

$$\text{Expected return} = \frac{1}{6} \times -10 + \frac{1}{6} \times -20 + \frac{1}{6} \times -30 + \frac{1}{6} \times 40 + \frac{1}{6} \times 50 + \frac{1}{6} \times 60 = +15\%$$

A way to measure the risk of playing this game is to measure the spread of the returns, or in other words, the variance. The variance summarizes the spread of the possible outcomes. The variance of the return

Outcome	% rate return	Deviation expected return	Squared deviation	Pr	Pr $\times$ squared deviation
1	-10	-25	625	$\frac{1}{6}$	$104\frac{1}{6}$
2	-20	-35	1225	$\frac{1}{6}$	$204\frac{1}{6}$
3	-30	-45	2025	$\frac{1}{6}$	$337.\frac{3}{6}$
4	+40	+25	625	$\frac{1}{6}$	$104\frac{1}{6}$
5	+50	+35	1225	$\frac{1}{6}$	$204\frac{1}{6}$
6	+60	+45	2025	$\frac{1}{6}$	$337\frac{3}{6}$
				Variance =	$1291\frac{2}{3}$

Table 1.1: Estimating the variance of the dice-throwing game

on this game is the sum of expected squared deviations from the expected return. Table 1.1 shows how the variance of the game can be estimated.

The variance here is estimated as $1291\frac{2}{3}$. The way the table estimates the variance can be summarized as follows:

$$\text{variance}(r_{game}) = \frac{\sum_{i=1}^{N}(r_{game,i} - \bar{r})^2}{N} \tag{1.1}$$

Here r_{game} represents the return on playing the game, N is the total number of possible outcomes, $r_{game,i}$ is the return on the game after a throw of the dice with outcome i and $\bar{r}$ is the expected return of the game. Another measure for summarizing risk is the standard deviation. The standard deviation is the square-root of the variance. The variance is mostly denoted as σ^2 and the standard deviation mostly as σ.

Now, assume that the player has to pay 50% of his balance after throwing a one, but he receives 100% of his balance when he throws a six. More specifically, the return is -50% for throwing a one and 100% for throwing a six. The expected return remains the same: 15%. The variance, however, now equals $2758\frac{1}{3}$. The conclusion is that although one expects the same result after playing the game, the game is far more risky as a result of more spread in the possible returns.

1.2 Risk management in finance

Risk does not, although it might seem so, restrict itself to games with dices. An analogous example can be given with a decision maker deciding whether to sell his stock or not. In a similar way, the variance can be estimated for the returns on this stock, plotted in the financial time-series of the past years, corresponding to the stock the decision maker is interested in. The decision maker now disposes over information regarding the spread in the returns over the past years. Based on the spread the decision maker encounters while assessing the historical returns, he can assign an amount of risk to the stock and adapt his decision likewise. A well-informed decision maker has knowledge of the factors influencing the result of his decision and can assign probabilities to the states of nature.

It may seem that a decision maker has to accept risk as an entity falling from the sky. Fortunately, risk is something that can be managed. There are various instruments available to cope with risk. One of these instruments are *insurance*, *hedging* and *Value at Risk*. Schematically, this can be summarized as follows:

- Insurance
- Hedging
  - Futures
  - Forwards
  - Swaps
- Value at Risk

This list is certainly not an exhaustive list. When a company is insured, the risk it faces is transferred to the insurance company and pays a risk premium in return. Hedging involves taking on one risk to neutralize another risk. Tools available for hedging are futures, forwards and swaps and are also referred to as derivatives, as their value depends on the value of another, underlying, asset. For a general discussion regarding risk and risk management, I refer to Brealey and Myers (2001). A detailed discussion regarding the mechanics of derivatives trading is beyond the scope of this thesis. However, for a detailed discussion regarding derivatives, I refer to Hull (2000). The next section discusses the concept of Value At Risk, which is fundamental for this thesis.

1.3 Value at Risk

Futures and other derivatives can be used in order to reduce risk by hedging. However, futures and other derivatives can also be used for speculation. In this case, one enters a futures contract *without* the neutralizing holding of the underlying asset. Speculating with derivatives can be very risky. There are various examples of companies which are brought into trouble as a result of speculation. In 1995, the Baring's bank incurred huge losses as a result of a trader called Nick Leeson. Nick Leeson borrowed a large amount of funds in order to speculate on the Japanese stock market index. Unfortunately, the stock market collapsed as a result of an earthquake. The final losses were $1.4 billion, resulting in the Baring's bank to become insolvent. In addition, Metalgesellschaft entered into an oil futures contract. The final loss incurred was $1.3 billion (Brealey and Myers 2001).

In reaction to the Barings'- and other financial disasters, the need for improved financial control grew and the concept of *Value at Risk* became more important. Value at Risk is one single number for the senior management to express and summarize the total risk of a portfolio with financial assets. Value at Risk measures the worst expected loss over a given horizon under normal market conditions at a given confidence level. Later on, large banks have to base their market risk capital requirements on the VaR number. This requirement was a result of an agreement between U.S. and international banking authorities called the Basel Accord (Jorion 1996).

With the concept of Value at Risk (VaR) models arose to estimate the VaR, varying from models with an analytical, parametric approach to models that use complex, costly simulations. Differences between these models will be discussed later on in this thesis. As capital requirements for banks are based on the VaR figure, it implies that the more accurate a VaR-model is, the more efficient capital can be allocated in order to cover eventual losses.

Time is an important aspect for VaR estimation. In order to estimate the VaR for a given horizon, daily volatility is scaled to h-day volatility using the $\sqrt{h}$-rule for scaling. However, it is a general assumption to assess the historical data in its complete *resolution*; all behavior is assessed, from short-term, high frequent volatility to long-term, low frequent volatility. Despite the fact that financial risk managers might be interested in short-term volatility when estimating a VaR for a short-horizon and long-term volatility when estimating a VaR for a longer horizon, no explicit *distinction* is made between information captured in low(er) and high(er) frequent patterns. The fact that no explicit distinction is made, assumes implicitly that all information in historical data is relevant, regardless the time-horizon chosen, with the result that all this information is encapsulated in the VaR-figure. Under the assumption that a financial risk manager might be interested in the volatility corresponding to the given time-horizon only, only a subset of this information has to be assessed, resulting in a lower VaR-figure. A lower estimated VaR-figure implies that expensive capital can be saved. However, if a VaR-model is too optimistic, risk can be underestimated and reserved capital could be insufficient. The other way round, abundant capital is reserved which could have been invested elsewhere, when the model is too conservative.

The assumption that a financial risk manager is only interested in risk corresponding to his horizon and capital can be saved by assessing the data in the corresponding resolution, makes us in search for a tool. Wavelet analysis, as a tool, is a technique from applied mathematics, able to decompose a given signal or time-series into different timescales by shifting a window variable in both time and scale over the original signal. Each timescale represents a different resolution. The lower scales capture highly frequent behavior, whereas the higher scales capture low-frequent behavior. This technique, which is used in finance and economics during the past few years, might be a contribution to financial risk analysis and more specific VaR estimation, when through a decomposition for each time-horizon an appropriate resolution can be chosen, which is the fundamental hypothesis underlying this thesis.

1.4 Definition of the research question

Summarizing the ideas from the previous sections, the problem is to find a way to define the appropriate resolution for the time-horizon one is interested in, in such a way that the amount of risk can be tailored to ones needs. The research question is therefore: ”Can wavelet analysis help to define the appropriate resolution for the time-horizon one is interested in?”. The following three sub questions, defining the research problem, should be articulated.

First, how can wavelet-analysis, as a tool, decompose a financial time-series into multiple levels, such that each level captures specific information about the behavior of this financial time-series?

Second, when introduced into the domain of Value at Risk, can wavelet analysis help to tailor risk by finding an appropriate resolution for each time horizon, through use of a multi-level decomposition?

Third, can an alternative, wavelet-based model be formulated for Value at Risk estimation which takes such a distinction into consideration and does this model perform well compared to existing competing models?

1.5 Methodology

In order to operationalize the research problem, I took the following steps. As a first step I consulted the available literature on VaR and VaR-estimation in order to understand the general idea and its role in finance and economics and made an overview of the available models for VaR estimation. In order

to acquire intuition for the theory, I worked out numerical examples. First with simple models, later on with more complex, time-varying volatility models.

As a second step I made an overview of the most commonly available wavelet literature and dove into wavelet theory to understand the mathematics behind it. By combining different interpretations of wavelet analysis from different literature, I have defined a personal mathematical framework, consistent with wavelet analysis, using a notation I felt comfortable with.

The third step was to make an overview and comparison of the available software for wavelet analysis. Using this overview, I chose an appropriate software package in order to acquire intuition for the theory and to conduct experiments.

As a fourth step I made the relation between wavelet analysis and VaR estimation by relating the energy in a signal to the variance of a signal. In order to make this relation, I studied the behavior of the variance when applied to a decomposed signal. Based on this relation, I proposed an alternative, wavelet-based model for VaR estimation.

The fifth step was to conduct experiments with the proposed model on artificially generated data and real-life data from the Dow Jones Index. I studied the results and checked whether the model gave an accurate estimation of the VaR, by systematically estimating the worst expected loss given a confidence level and a time-horizon (i.e. the VaR) and checking whether this estimation is exceeded in the given time-horizon. The step is repeated a significant number of times in order to get a reliable notion of the performance.

As a sixth step, I researched how the model performs compared to other existing, more widely used models.

As a final step I validated the results by means of a statistical framework, referred to as backtesting, in order to make sound scientific statements about the performance of the proposed wavelet-based model.

The writing of the final thesis was a recurrent process. It entailed the performing of a literature study to gain a general understanding, experimenting with the theory to get intuition for the theory and reporting the results and findings into small reports. The small reports were bundled into this final thesis. This process was for me the logical order to structure this complex ”to-large-to-handle” problem into smaller solvable problems.

1.6 Structure of the thesis

To support the methodology as presented, the thesis is structured as follows and is shown in figure 1.1. Chapter 1, this chapter, is the introduction. Chapter 2 explains the fundamentals of VaR-estimation. It explains why it is important, how it relates to risk, how it is estimated and what the existing models and their shortcomings are. Chapter 3 explains the fundamentals of wavelet analysis. It explains how wavelet decomposition and reconstruction works and what multiresolution analysis is. In addition, a concise overview of wavelet families and their characteristics is given. Chapter 4 shows how wavelets can be used for analyzing risk and how the correct resolution can be found for a corresponding horizon. Here, a wavelet-based model for VaR-estimation will be formulated. Chapter 5 sets up experiments and demonstrates how the new model performs compared to existing models. Chapter 6 validates these results by discussing and applying models for validation. Chapter 7 concludes the thesis.

Chapter 8, the appendix, gives a detailed overview about the various tools for wavelet analysis currently available.

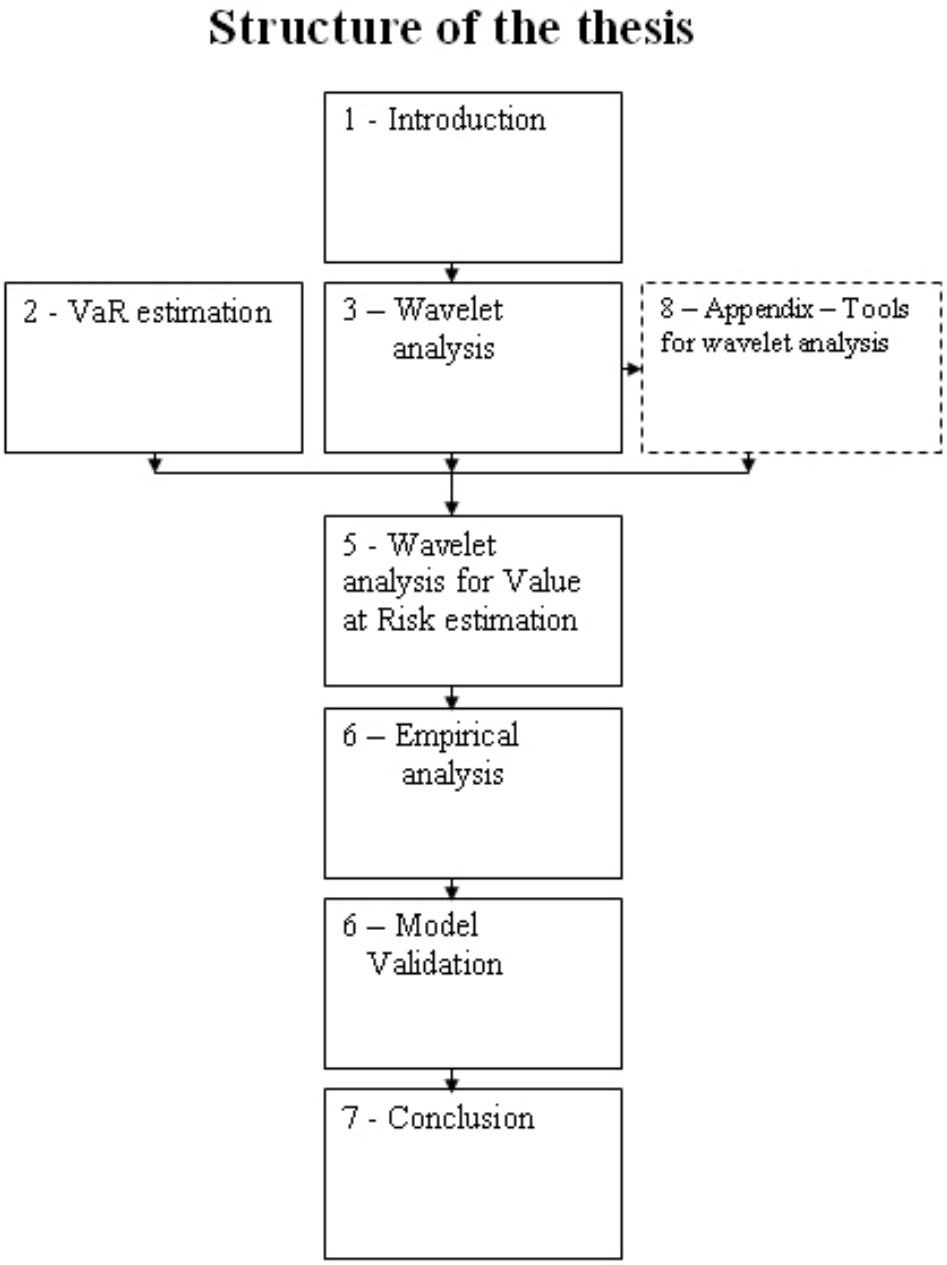


Figure 1.1: Structure of the thesis

Chapter 2

Value at Risk estimation

2.1 Introduction

In the beginning of the 1990's, several examples of financial mismanagement occurred. Examples as Metallgesellschaft, Orange County and Barings underline the lack of and necessity for proper financial management (Huij 2001).

In this period, the interest rate was at such a low position, making Orange County - the government of Florida, USA - borrow a fair amount of money, in order to buy bonds with a higher interest compensation. This position could be very lucrative, given that the interest rate remains at the same position, or decreased even more. Unfortunately, the interest rate increased drastically, resulting in a deteriorated position and a loss of $1.7 billion for Orange County.

Metallgesellschaft, a large German company, created a difficult situation as well by entering into a supply contract for delivering oil to its customers at a predetermined price. Metallgesellschaft hedged the possible risk of an increasing cost price by a futures contract in order to compensate the possible losses of an increased cost price with the revenues of the futures contract. However, the margin calls [1]on the futures contract invoked such a big outflow, that the senior management had to close the hedge position. The final loss was $ 1.3 billion (Hull 2000).

The lessons learned from these disasters is the need for improved *internal control* ; the need for senior management to articulate a clear policy about how derivatives can be traded and the extent to which employees can take positions. With this policy, senior management should ensure that the policy is carried out using controls. And with controls comes the monitoring of risk and the need for improved risk models. Financial institutions for example, usually calculate risk measures as delta, vega and gamma [2] for every market variable it is exposed to. The huge amount of information generated every day is of valuable information to traders, but of limited use for senior management. This is where Value at Risk

[1] A futures trader has to deposit funds in what is called a margin account at the time a futures contract is entered into. The amount is called the initial margin. If the balance of the margin account falls below a specific margin, the maintenance margin (which is lower than the initial margin), the trader has to fill up the space between the maintenance margin and the initial margin. This is called a margin call.

[2] See Hull (2000) for a detailed discussion about delta-gamma-vega analysis.

(VaR) comes in. VaR is one single number for the senior management to express and summarize the total risk of a portfolio with financial assets. The VaR figure is always accompanied with two parameters: the time horizon h and the significance level c, resulting in statements of the form: With c % certainty, the loss on the portfolio will not be more than *VaR* euros in the next period of h days. Losses greater than the VaR are only suffered with a very small specified probability $(1 - c)$. More formally:

Definition 1 *"VaR measures the worst expected loss over a given horizon under normal market conditions at a given confidence level" (Jorion 1996)*

VaR is widely accepted by corporate treasurers, fund managers and financial institutions. In addition, VaR models have been sanctioned for determining market risk capital requirements for large banks by U.S. and international banking authorities through the 1996 Market Risk Amendment to the Basel Accord (Jorion 1996). In fact, under the protection of the Basel Committee on Banking Supervision, the capital requirements for the market risk exposure of large banks are now explicitly based on their VaR estimates for a ten day holding period and 99 percent coverage. In other words, banks now have to set an amount of capital aside based on a VaR estimate in order to cover eventual losses. And, as the definition states, there is only 1 percent chance that these eventual losses exceed the VaR. For a more detailed discussion about these requirements, I refer to Jorion (1996).

The rise of VaR is accelerated as a result of this increased pressure from regulators in order to better control financial risk, the globalization of financial markets leading to exposure to more sources of financial risk and more advanced technology for enterprise wide risk management (Jorion 1996).

The purpose of this chapter is to give insight in the fundamentals of VaR, its relation with risk and the way it is estimated. Furthermore, this chapter should give a clear understanding of models for time-varying risk and its shortcomings. Therefore, the remainder of this chapter is as follows: Section 2.2 gives an overview of VaR in finance and economics. Section 2.3 explains the fundamentals of VaR estimation and section 2.4 explains VaR estimation for a multiple asset portfolio. Then, section 2.5 extends the discussion with the modelling of time-varying risk. The chapter concludes with the shortcomings of these models in section 2.6.

2.2 Value at Risk in finance and economics

Wiener (1997) gives an introduction to Value at Risk where several methods for VaR estimation are discussed. Jorion (1996) explains in his book the various building blocks of VaR measurement. His book not only explains the more analytical parametric methods, but also the more sophisticated and complex simulation methods for VaR estimation. The book is complete in the sense of specifying all the regulatory capital standards. In addition, various methods for validation of the VaR models are discussed, like stress testing and backtesting. Backtesting will be discussed in detail in chapter 6.

The parametric approach, also called variance-covariance approach is more popular than its more complex and sophisticated non-parametric counterpart, the simulation approach. This approach, implemented as either historical simulation or Monte Carlo simulation, is computationally demanding and very costly as well. When the approach is parametric, it is based on the assumption that returns are distributed normally. Historical data is then used to estimate parameters as the mean, variance and correlation. Then, using the parameters to map the risk, the distribution of profits and losses are found for a specific time-horizon. The parametric approach is flexible, easy to understand and widely accepted (Wiener 1997). However, it relies heavily on the assumption of a normal returns distribution. This assumption can be wrong in case when the distribution is "fat-tailed": the frequency of exceptions occurring is higher than when the distribution is assumed to be normal (Hull 2000).

Monte Carlo simulation does not make this assumption. The approach first identifies the set of important market factors. Then, a joint distribution is build from these factors. This distribution is either formed of historical data or data based on specific economic scenarios. The next step is to run simulations, with a large amount of scenarios. Finally, at the end of the period, for each scenario, the profits and losses are measured. If the VaR is to be estimated at a 95% confidence level for example, the results will be ordered from worst losses to best profits and the VaR estimate is then set equal to the 5% quantile of worst losses. Although this approach does not make any assumption regarding the underlying distribution, it converges very slowly, as the complete portfolio has to be revalued many times (Wiener 1997).

Historical simulation also doesn't make any assumption of the returns distribution being normal. The approach observes the historical data of the current portfolio over the last years. Then, the approach begins by measuring daily percentage changes in all market variables. Finally, using a moving window, profits and losses are estimated for every h-day period. Similar to the Monte-Carlo approach, the 5% quantile of the worst results is the VaR estimate after ordering the results (Wiener 1997). Although the approach reflects the probability distribution of the market variables quite accurately, the number of simulation trials is restricted to the number of days in the historical data. In addition, sensitivity analysis is difficult (Hull 2000). Within this thesis, the concentration is focussed on the parametric models.

As already mentioned, each of the different approaches has its advantages and disadvantages. A comparative approach is discussed by Ajassa, Ambrosetti and Giovanni (1998) in their paper. Bredin and Hyde (2001) also discuss the two streams in VaR estimation: parametric and non-parametric VaR estimation, where time-varying risk models as GARCH- and equally weighted moving average (EWMA) models[3] are put next to historical simulation models. They conclude their paper in favor of the EWMA model.

Within the stream of parametric VaR estimation, the methods vary from simple methods that assume the risk (variance) to be constant over the period and the more sophisticated methods that assume that risk varies dynamically over time. Both Hull (2000) and Jorion (1996) give an extensive overview of these models. Engel and Gyzicki (1999) analyzed the effect of these different time-varying risk models on VaR-computation. The conclusion is interesting: firstly, there is considerable variation in variances over time and secondly, simple models such as equally weighted moving averages and EWMA models perform equally well as the more sophisticated GARCH models.

Going deeper into the matter of GARCH models, Drost and Nijman (1993) show that time plays an important role when referring to risk. Not only the assumption that risk varies over time, but also the fact that time has influence on scaling. Their conclusion is that scaling risk over time is not as straightforward as it seems. This is discussed in more detail in section 2.5

Berkowitz and O'Brien (2001) analyzed the accuracy of VaR-models at commercial banks. The results of this research underpins the relevancy of the research in this thesis: at most banks, the VaR-models tended to be conservative, implying that the levels of capital coverage were higher than necessary. For some banks, the models were even inaccurate. Haas (2001) discusses the various methods of backtesting VaR models. Backtesting is discussed in chapter 6.

2.3 Estimating Value at Risk

According to definition 1, VaR measures the worst expected loss over a given time horizon under normal market conditions at a given confidence level. This definition implicitly states that for the estimation of VaR, it is essential to know the time horizon h, a confidence level c and not to forget an initial portfolio, with value W. Before going into details about the horizon h, we first have to know how to define the

[3]These time-varying risk models are discussed in section 2.5

worst expected loss at a single day. The ideas in this section are based on theories from Jorion (1996), Hull (2000) and Brealey and Myers (2001).

Now, consider the value of this portfolio at the end of a particular day, say day n, is W_n. Then, the return u_n is the *continuously compounded return* of this portfolio *during* day n. This can be written as follows:

$$u_n = \ln\left(\frac{W_n}{W_{n-1}}\right) \tag{2.1}$$

Assume now that the initial value of the portfolio is W_0. Then, the value of the portfolio at the end of the day is $W = W_0 e^{u_n} \approx W_0(1 + u_n)$. The expected value (mean) and volatility of u_n are μ and σ. Now assume that the worst possible return on this portfolio is u_n^*. Then, the lowest portfolio value can be defined as: $W^* = W_0(1 + u_n^*)$. Then, VaR can be defined as the loss, relative to the mean μ:

$$\text{VaR}(\mu) = E(W) - W^* = -W_0(u_n^* - \mu) \tag{2.2}$$

which equals

$$\text{VaR}(\mu) = W_0(1 + \mu) - W_0(1 + u_n^*)$$

When VaR is defined as the absolute loss, the following equation holds, filling in 0 for μ:

$$\text{VaR}(0) = E(W) - W^* = -W_0(u_n^*) \tag{2.3}$$

In both equations (2.2) and (2.3), u_n^* corresponds to the *cut-off return*. It is the lowest acceptable rate of return. What is acceptable, depends on the amount of confidence one likes to refer to when making statements or more specific: the confidence level.

Using the notion of the confidence level, VaR can be derived from the probability distribution of the future portfolio $f(w)$. At the given confidence level c, the lowest possible realization of portfolio value W is W^*, such that the probability of exceeding this value, is $1 - c$. Therefore, c can be defined as:

$$c = \int_{W^*}^{\infty} f(w)dw \tag{2.4}$$

The other way round, the probability of a lower realization than W^*, $P(w \leq W^*)$ is $1 - c$:

$$P(w \leq W^*) = \int_{-\infty}^{W^*} f(w)dw = 1 - c \tag{2.5}$$

Figure 2.1 shows the areas c and $1 - c$. At the given confidence level c, the lowest possible realization of portfolio value W, lies at the border of the two areas and is marked with W^*.

This specification is valid for any distribution. However, VaR estimations can be simplified considerably when the underlying distribution is assumed to be normal. When this is the case, the VaR figure can be obtained *directly* from the portfolio standard deviation, using a multiplicative factor which depends on the confidence level and time-horizon. The α, which corresponds to the confidence level c in a probability distribution, can be found as follows:

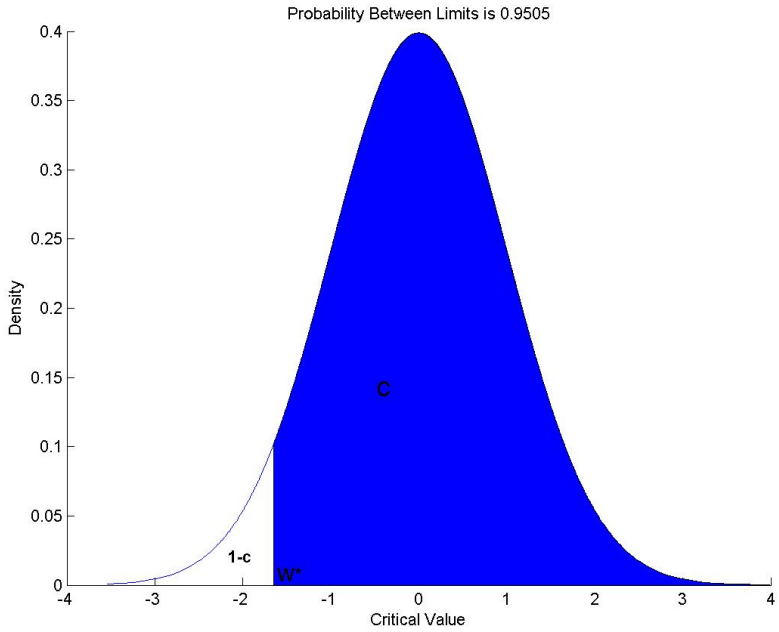


Figure 2.1: The standard normal distribution and the confidence level c

$$-\alpha = \frac{u_n^* - \mu}{\sigma} \tag{2.6}$$

In figure 2.1,which demonstrates a 95% confidence level, we can see that α lies at the point 1.65 below 0.

u_n^* can be written as:

$$u_n^* = -\alpha\sigma + \mu \tag{2.7}$$

Assume that σ and μ are expressed on a daily basis. Now let's introduce the notion of the time horizon. The time horizon to base VaR on is defined as h, in days. However, as σ is expressed on a daily basis, we first have to scale this σ from a 1-day interval to a h-days interval.

A h-days variance can be estimated by multiplying a 1-day variance with h:

$$\sigma_{h-days}^2 = \sigma_{1-day}^2 h$$

Then, by taking square roots at both sides,

$$\sigma_{h-days} = \sigma_{1-day}\sqrt{h}$$

By substituting (2.7) in (2.2), and replacing the 1-day σ with a h-day σ, the following general expression defines the VaR, relative to the mean μ:

$$\text{VaR}(\mu) = -W_0(u_n^* - \mu) = W_0\alpha\sigma\sqrt{h} \tag{2.8}$$

It is custom in VaR calculations that the expected return of an asset is considered as zero. Although this is not exactly true, it is a reasonable assumption, as this expected return is generally small, when compared to the standard deviation of the return. Equation (2.8) also represents the absolute loss.

Example 2.3.1 *Consider a one-stock portfolio based on a €10.000.000,- position in shares of Unilever. The daily volatility is assumed here to be 4%. As the size of the position is €10.000.000,- ,the standard-deviation of the portfolio is €400.000. The VaR of this positions needs to be calculated at a 99% confidence level and a 10-day horizon, yielding the following setting:*

- W_0 =*€10000000*
- σ =*€400000*
- $c = 99\%$, *yielding* $\alpha = 2.33$
- $h = 10$

The VaR corresponding to this setting can then be calculated as: €400000 $\times 2.33\sqrt{10}$ *= €2947242,78.*

Now consider a two-stock portfolio consisting of a €1000000 position in Unilever shares, with a daily volatility of 4% and a €600000 position in KLM with a daily volatility of 2%. The return on the two shares is assumed to have a bivariate normal distribution with a correlation of 0.6. In this case, the portfolio standard deviation needs to be calculated using the following equation:

$$\sigma_{x+y} = \sqrt{\sigma_x^2 + \sigma_y^2 + 2\rho\sigma_x\sigma_y} \tag{2.9}$$

Again, the VaR of this position needs to be calculated at a 99% confidence level given a 10-day horizon, yielding the following setting:

- W_{uni} =*€1000000*
- W_{KLM} =*€600000*
- σ_{uni} =*€1000000* $\times$ *0.04 = €40000*
- σ_{klm} =*€600000* $\times$ *0.02 = €12000*
- $c = 99\%$, $yielding$ $\alpha = 2.33$
- $h = 10$

For a 10-day period, σ_{uni} =*€126491,11 and* σ_{KLM} =*€37947.33.*

Therefore, $\sigma_{uni+KLM} = \sqrt{12649.11^2 + 37947.33^2 + 2 \times 0.6 \times 12649.11 \times 37947.33} = 152.315$

The 10-day VaR at a 99% confidence level for the given portfolio is then: €152315 $\times$ *2.33 = €345895.03.*

The strategy as showed in the example is a classical method for lowering risk through diversification of the portfolio. Diversification however, is not within the scope of this thesis. For a more detailed discussion regarding diversification, I refer to Brealey and Myers (2001) and Hull (2000).

2.4 A linear model for a multiple asset portfolio

The example above showed the VaR calculations of a single-stock and two-stock portfolio. Using the theory from Hull (2000), I want to generalize the ideas above. Consider a portfolio with value W consisting of N assets with an amount m_i invested in asset i $(1 \leq i \leq N)$. As r_i is defined as the absolute return on asset i in one day, then $m_i \triangle r_i$ is the change in the value of the investment in asset i in one day.

$$\triangle W = \sum_{i=1}^{N} m_i \triangle r_i \tag{2.10}$$

is defined as the change in the value of portfolio W in one day. Here is supposed that the changes in the values of the asset prices have a multivariate normal distribution, and therefore $\triangle W$ is considered as normally distributed. For VaR estimations, only the mean and standard deviation of $\triangle W$ needs to be calculated, where the mean was already considered as zero in the previous section.

The standard deviation of $\triangle W$, which will be denoted as σ_W is defined as:

$$\sigma_W = \sum_{i=1}^{N} \sum_{j=1}^{N} \rho_{ij} m_i m_j \sigma_i \sigma_j \tag{2.11}$$

which can also be written as:

$$\sigma_W = \sum_{i=1}^{N} m_i^2 \sigma_i^2 + 2 \sum_{i=1}^{N} \sum_{j<i} \rho_{ij} m_i m_j \sigma_i \sigma_j \tag{2.12}$$

where ρ_{ij} is the correlation between $\triangle r_i$ and $\triangle r_j$ and σ_i is the standard deviation of $\triangle r_i$. Next, σ_{ij} denoting the covariance between $\triangle r_i$ and $\triangle r_j$ and the variance of $\triangle r_i$ if $i = j$, can replace $\rho_{ij}\sigma_i\sigma_j$, resulting in:

$$\sigma_W = \sum_{i=1}^{N} m_i^2 \sigma_i^2 + 2 \sum_{i=1}^{N} \sum_{j<i} m_i m_j \sigma_{ij} \tag{2.13}$$

This can be written in matrix notation as well:

$$\sigma_w^2 = \begin{bmatrix} m_1 & m_2 & \cdots & m_n \end{bmatrix} \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1N} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{N1} & \sigma_{N2} & \cdots & \sigma_{NN} \end{bmatrix} \begin{bmatrix} m_1 \\ m_2 \\ \vdots \\ m_n \end{bmatrix}$$

By defining Σ as the *variance-covariance matrix*, the representation can be written more compactly as:

$$\sigma_w^2 = \mathbf{m^T \Sigma m} \tag{2.14}$$

When finally the portfolio standard deviation using the variance-covariance matrix is calculated, the VaR number can be calculated by using equation (2.8).

2.5 Modelling time-varying risk

As we have seen, VaR estimations are mostly based on a day-by-day basis, implying that we are mostly interested in a 1, 5 or 10-day VaR. However, VaR estimations can be made on a quarterly or even yearly basis. In the previous sections, I assumed the volatilities and correlations which filled the variance-covariance matrix as given. For VaR estimations, we are mostly interested in current volatilities and correlations, as we are assessing the portfolio value over a short time-frame. When estimating VaR, it is important to recognize that portfolio risk is not constant over time, but is subject to a dynamic process. For the modelling of this time-varying risk, there are several models available. In this section, three models will be given attention to, through use of the theories derived from Hull (2000):

- A model which gives equal weight to all observations
- A model given weight to its observations using a GARCH process
- A model which exponentially weights the observations using a moving average.

2.5.1 Volatility estimation

Define u_n as the continuously compounded return during day n, estimated using equation (2.1), of the value of a certain market variable (or portfolio) W, based on the process $W = \{W_1, W_2, \cdots, W_T\}$. For VaR estimation, risk assessment on a daily basis is preferred, in order to transform this into a h-day risk. This return can be estimated using the value of this variable W at days n and $n-1$, using equation (2.1).

Then, the square of the volatility on day n, the variance rate, is estimated as follows: [4]

$$\sigma_n^2 = \frac{\sum_{i=1}^{M} u_{n-i}^2}{M} \tag{2.15}$$

As can be seen, the variance rate is estimated using the M observations before day n. Notice that again is assumed that the expected return of an individual market variable is zero, due to the short time-frame (a day). As can be seen, equation (2.15) gives *equal weight* to all observations. However, it can be said that an observation is more likely to be drawn from the same distribution if the observation is made more recent. A model giving weight to more recent observations is defined by:

$$\sigma_n^2 = \sum_{i=1}^{M} \alpha_i u_{n-i}^2 \tag{2.16}$$

given that

$\alpha_i < \alpha_j$ if $i > j$ and $\sum_{i=1}^{M} \alpha_i = 1$

Again, the variance rate is calculated using the M observations before day n. In this case, there is an amount of weight α_i given to the i^{th} observation before day n. The restriction assures that older observations are given less weight and that the weights sum to unity.

[4] A careful reader can notice that this way of defining the variance is not in line with traditional variance estimation. For instance, the expected return $\bar{u}$ is assumed to be zero. This assumption has very little effect on estimates of the variance, as the expected change of the price of a certain market variable is very small, when compared to the standard deviation of the changes. In addition, a division by M rather than by $M-1$ is made. The estimate of the variance is now a maximum likelihood estimate. Maximum likelihood methods are discussed later on.

2.5.2 ARCH, GARCH(P,Q) and EWMA models

Another model based on the same idea as equation (2.16) extends the model by assuming there is some long-run volatility V. Likewise, this long-run volatility gets a weight γ assigned. Together, this forms the following equation:

$$\sigma_n^2 = \sum_{i=1}^{M} \alpha_i u_{n-i}^2 + \gamma V \tag{2.17}$$

given that

$$\gamma + \sum_{i=1}^{M} \alpha_i = 1$$

This model is commonly known as ARCH, which stands for autoregressive conditional heteroscedasticity. [5] A model that extends this idea is the more general GARCH model, where the weights given to older observations decay exponentially. Bollerslev proposed this model in 1986. In the GARCH model, the estimate of the volatility for day n is not only calculated from u_{n-1} but also from σ_{n-1}. This results in the following equation

$$\sigma_n^2 = \alpha u_{n-1}^2 + \beta \sigma_{n-1}^2 + \gamma V \text{ given that } \alpha + \beta + \gamma = 1 \tag{2.18}$$

From equation (2.18) it is not straightforward to see that the weights decay exponentially. However, as the model is a recurrent model, one can substitute for σ_{n-1}^2 in equation (2.18) and continue by substituting σ_{n-2}^2, etc. Finally, one can see after rewriting that the weight applied to u_{n-i}^2 equals $\alpha\beta^{i-1}$, where the weights decline exponentially at rate β.

In some literature, the long term variance V and its weight γ, are written as ω. This specific GARCH model is known as the GARCH(1,1) model, estimating σ_n^2 from the most recent observation of u^2 and the variance rate. Likewise, GARCH(p,q) models take p observations for u^2 into consideration and q observations for the variance rate in order to estimate σ_n^2.

Based on the same GARCH model, but leaving the long-run volatility V out of the picture, is the exponentially weighted moving average (EWMA).

$$\sigma_n^2 = \lambda \sigma_{n-1}^2 + (1-\lambda) u_{n-1}^2 \text{ with } (0 \leq \lambda \leq 1) \tag{2.19}$$

Here has λ a value between 0 and 1. The EWMA model is a special form of the GARCH(1,1) model, where $\alpha = (1-\lambda), \beta = \lambda$ and $\gamma = 0$. A weight λ is assigned to the most recent observation of the variance rate, σ_{n-1}^2, and a weight $(1-\lambda)$ to u_{n-1}^2. Similarly to the GARCH(1,1) model, the weights decay exponentially in time.

This EWMA model with ($\lambda = 0,94$) is used by JP Morgan's RiskMetrics, which is a sophisticated tool for VaR estimation.

[5] A model exhibits heteroscedasticity when the error-terms do not all have the same variance (Newbold 1995).

2.5.3 Maximum likelihood methods

The parameters of the models as described above, can be estimated using maximum likelihood methods. Using this method, the values of the parameters are chosen in such a way, that the probability of the sequence of observed returns is maximized. The probability for M normally distributed observed returns with mean zero and a time-varying variance σ_n^2 can be estimated using the following equation:

$$\prod_{n=1}^{M} \left[\frac{1}{\sqrt{2\pi\sigma_n^2}} exp\left(\frac{-u_n^2}{2\sigma_n^2} \right) \right] \tag{2.20}$$

Maximizing equation (2.20) is equivalent to maximizing the logarithm of this equation. By maximizing the logarithm of the equation, constant multiplicative factors - 2π in this case - can be ignored. This yields:

$$\sum_{n=1}^{M} \left[-\ln(\sigma_n^2) - \frac{u_n^2}{\sigma_n^2} \right] \tag{2.21}$$

Iteratively, the maximum of these equation can be found, yielding the parameters which can be used for input to the models described above. For a more detailed explanation on the working of maximum likelihood methods, I refer to Hull (2000).

2.5.4 Scaling

In section 2.3, VaR estimation was based on a risk which was constant over time. The previous subsection introduced several models that assumed that risk evolves dynamically over time. In the case of risk which is assumed to be constant over time, scaling a 1-day volatility to a h-day volatility is appropriate with the $\sqrt{h}$-rule. When this is not the case, scaling is not so straightforward.

Diebold, Hickman, Inoue and Schuermann (1997) confirm that the scaling of volatilities with the traditional $\sqrt{h}$-rule is inappropriate when volatilities are updated as occurring with GARCH and other conditional models. This $\sqrt{h}$-rule is only valid when volatility is assumed to be constant over the time period. A reason for this is that scaling rule relies on 1-day returns being *independently identically distributed* (i.i.d.). However, GARCH and other conditional models do not assume that all returns are from the same distribution, i.e. not identically distributed. The current variance rate, which defines the distribution, is updated using a daily updating scheme and is based on the most recent observation of the return and the most recent estimated variance rate. For a more detailed and extensive discussion regarding the appropriate conditions for scaling with $\sqrt{h}$, I refer to Drost and Nijman (1993).

Appropriate scaling is performed by altering the parameters α, β and ω as found by maximizing equation (2.20). Drost and Nijman (1993) study the temporal aggregation of GARCH processes. Suppose we begin with a sample path of a 1-day return series, which follows a standard GARCH process according to equation (2.18). Then Drost and Nijman (1993) show that the corresponding sample path of h-day returns, follows a similar GARCH(1,1) process with:

$$\sigma_{(h)_n}^2 = \omega_{(h)} + \beta_{(h)}\sigma_{(h)_{n-1}}^2 + \alpha_{(h)}u_{(h)_{n-1}}^2 \tag{2.22}$$

where

$$\omega_{(h)} = h\omega \frac{1-(\beta+\alpha)^h}{1-(\beta+\alpha)}$$
$$\alpha_{(h)} = (\beta+\alpha)^h - \beta_{(h)}$$

and $|\beta_{(h)}| < 0$ is the solution of the quadratic equation,

$$\frac{\beta_{(h)}}{1+\beta_{(h)}^2} = \frac{a(\beta+\alpha)^h - b}{a(1+(\beta+\alpha)^{2h}) - 2b}$$

where

$$a = h(1-\beta)^2 + 2h(h-1)\frac{(1-\beta-\alpha)^2(1-\beta^2-2\beta\alpha)}{(\kappa-1)(1-(\beta+\alpha)^2)} + 4\frac{(h-1-h(\beta+\alpha)+(\beta+\alpha)^h)(\alpha-\beta\alpha(\beta+\alpha))}{1-(\beta+\alpha)^2}$$
$$b = (\alpha-\beta\alpha(\beta+\alpha))\frac{1-(\beta+\alpha)^{2h}}{1-(\beta+\alpha)^2}$$

κ is the kurtosis [6] of u_n. As an EWMA model is some sort of a GARCH model, this form of scaling is also appropriate for EWMA models.

2.6 Shortcomings

Huij (2001) discussed shortcomings of time-varying risk models in his thesis. He showed that one of the shortcomings of these models is the impureness after a regime shift. This is the result of the fact that these models give more weight to recent observations. Huij (2001) proposed hidden Markov models for coping with this shortcoming.

When looking at the conventional time-varying volatility models for VaR estimation, as discussed in section 2.5, one can see that these models look at the historical data and assess the time-series with its complete resolution, regardless of the time-horizon chosen. The central question here is: "Does the *complete* resolution yield relevant information at all resolutions?" In other words: do all the high- and low frequency components *hidden* in the complete signal matter for *every* horizon? One might imagine for instance, that risk managers estimating risk for a short horizon are interested in short-term behavior only, whereas risk managers estimating risk for a long horizon are also interested in low frequent behavior. More specifically: the appropriate resolution might entail the right amount of information for a specific horizon. Therefore, a tool is needed which enables the analyst to decompose the signal into all of its components, separating higher frequent behavior from lower frequent behavior, in order to analyze which of these components entail relevant information. Here is where wavelets are introduced.

Wavelet analysis provides such a tool for decomposing a given signal into its different time-scales, in order to give each horizon its own resolution. Chapter 3 will show how wavelets can be utilized to decompose a given time-series into several time-scales. How time-scale decomposition and multiresolution analysis can be exploited in order to assess the amount of risk, can be read in chapter 4.

[6] The kurtosis measures the weight of the tails in a probability density function

Chapter 3

Wavelet analysis

3.1 Introduction

Wavelet analysis is a relatively new tool in the field of applied mathematics. Daubechies (1992), Chui (1992) and Graps (1995) provide the fundamentals of the wavelet theory. Wavelet analysis provides the opportunity to make semi-parametric estimations of highly complex structures without knowing the underlying functional form. Wavelet analysis had its impact on the area of signal processing, data compression and image analysis. The impact on signal processing was reviewed by Donoho (1995). Walker (2000) provides a primer on wavelets and their use in these applications of signal processing, image analysis and data compression.

Wavelet analysis, in contrast to Fourier analysis, gives insight in local behavior, whereas Fourier analysis gives insight in global behavior. The Fourier transform processes time-series by transforming the signal from the time domain into the frequency domain. The new processed signal provides insight in the amount of frequencies and the amount of energy in each frequency existing in this time-series. However, local effects are only visible in the time domain and not in the frequency domain. Wavelet analysis makes use of a fully scalable window, which is shifted along the signal in order to capture local behavior in the time domain. This process is repeated several times with different window-sizes, with a collection of time-frequency representations of the signal as a result. The transformation of the signal into the several resulting wavelet coefficients, each providing information at different scales, is more often referred to as time-scale decomposition. However, as there is no direct connection between the Fourier frequency parameter and the Wavelet parameter, the term scale is preserved for wavelet analysis, whereas the term frequency is preserved for Fourier analysis.

The use of wavelet analysis enables the analysis of non-stationary data, localization in time and time-scale decomposition, which proved to be useful in the analysis of economic and financial data (Ramsey 1999). The application of wavelet analysis in the field of finance and economics is discussed in more detail in section 3.5.

The remainder of this chapter is as follows. Section 3.2 gives an introduction in wavelet analysis with the basics of the discrete wavelet transform. The focus lies on the decomposition. In section 3.3, wavelet reconstruction is discussed in detail, where multiresolution analysis, the heart of wavelet analysis, stands central. Section 3.4 discusses in detail the different types of wavelets used in this research and their properties. The chapter concludes with section 3.5, a literature study regarding wavelets in finance, economics and soft-computing.

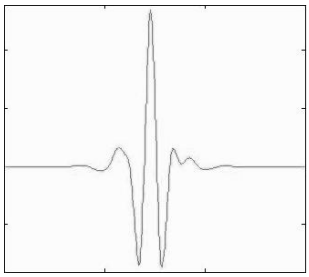


Figure 3.1: Wavelet

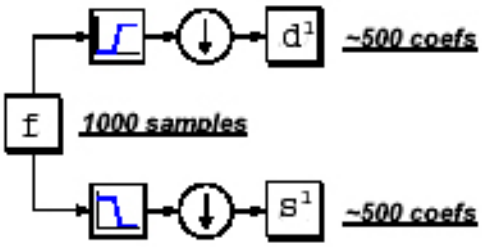


Figure 3.2: Level 1 decomposition

3.2 The wavelet transform - decomposition

Section 2.6 explained the necessity for a tool which can separate high frequent behavior from low frequent behavior. Wavelet analysis, as a tool, is able to perform such a process of separation, which is referred to as the *Discrete Wavelet Transform* or DWT. This section is divided into two subsections. The first subsection gives a general overview of the technique. The second subsection gives a formal mathematical framework.

3.2.1 The Discrete Wavelet Transform - general idea

For many signals, the low-frequent behavior - or low-frequency components - are the part of interest, as it gives the signal its identity. In order to get a good understanding for high-frequency components and low-frequency components, consider the human voice for example. The low-frequency components enables someone to understand what's being said. The high-frequency components however, make the nuances in the voice. Removing these high-frequency components makes the voice to sound different, but still enables one to understand what's being said.

When performing the discrete wavelet transform, two types of filters are applied to the original signal. One type captures the trend of the signal and the other type captures deviations from this trend. The first type is referred to as a scaling filter or *scaling signal* and the second type is referred to as a wavelet filter or *wavelet*. Figure 3.1 shows an example of a wavelet. A scaling signal looks similar, but has different features, in order to capture different behavior. Figure 3.2 shows a schematic illustration of the procedure.

Readers with an electrical and electronics engineering background may relate this procedure to applying low-pass and high-pass filters. However, when looking more carefully at the figure, we can see that the discrete wavelet transform yields two series of approximately 500 *coefficients*, while we started with 1000 samples in the original signal, whereas the application of low- and high-pass filters yields two times

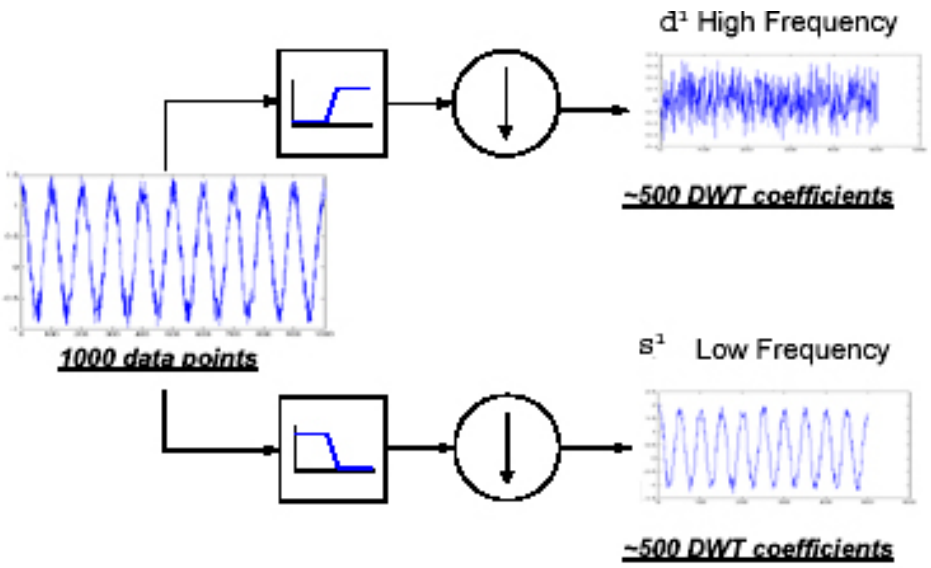


Figure 3.3: Level 1 decomposition of a noisy cosine - trend and detail subsignals

1000 samples. The series capturing the trend of the signal is indicated as s^1 and the series capturing all deviations from the trend is indicated as d^1. The "1" indicates a 1-level DWT. The fact that two series of approximately 500 coefficients are yielded, is due to a process where each coefficient is obtained by taking one value for every two values in the original signal, by averaging successive values. This process is called *downsampling*. The way the averaging is performed depends on the scaling signal and wavelet.

The procedure will be illustrated using a pure cosine with high-frequent added noise. This cosine function contains 1000 samples. Figure 3.3 shows a schematic illustration of the procedure.

When inspecting the figure, we can see that s^1 captures the trend of the signal with much less noise than the original signal and d^1 only captures the high-frequent noise.

Now, let's have a closer look at the "averaging". The averaging is already introduced in the introduction, which stated that local behavior is captured by shifting a window along the signal. In fact, applying a scaling signal or wavelet is equal to shifting a window along the signal. Capturing local behavior is possible as the scaling signals and wavelets are locally defined in time. Each window is obtained by shifting or *translating* the scaling signal or wavelet to the right. In practice this results into multiplying the original signal with a translated version of the original version of the scaling signal or wavelet. How this can be defined mathematically, is discussed in the next subsection. The process is shown in figure 3.4.

This complete process is only a first step. In order to take steps of further refinement, the complete process is iterated. The iteration takes place each time on the previously obtained series capturing the trend, such that at each level a new separation is made between trend and deviation from the trend. Schematically, the process as described above is illustrated in figure 3.5. This is called the wavelet decomposition tree.

Going deeper into the tree entails capturing lower frequent behavior. At the top level, the original signal can be found, including both low- and high-frequent behavior. Going a level deeper, the original signal is separated into a series capturing the trend, s^1, and a series capturing deviations from the trend, d^1, and by doing so the 1-level decomposition is performed. Repeating the same step, but on the series capturing the 1-level trend, s^1, trend is again separated from detail, yielding s^2 and d^2 or a 2-level decomposition. The number of levels in a wavelet decomposition depends on the nature of the signal.

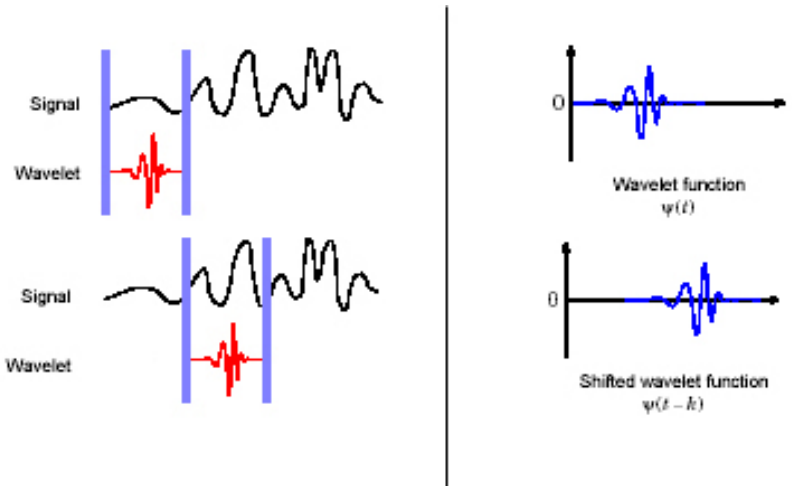


Figure 3.4: Left: shifting a window along the signal. Right: shifting a wavelet function in time.

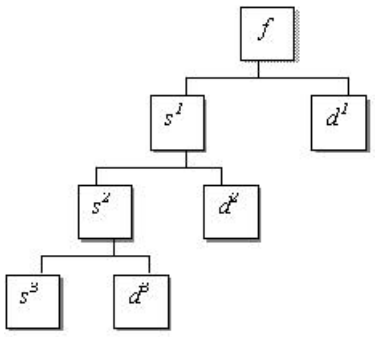


Figure 3.5: Wavelet decomposition tree

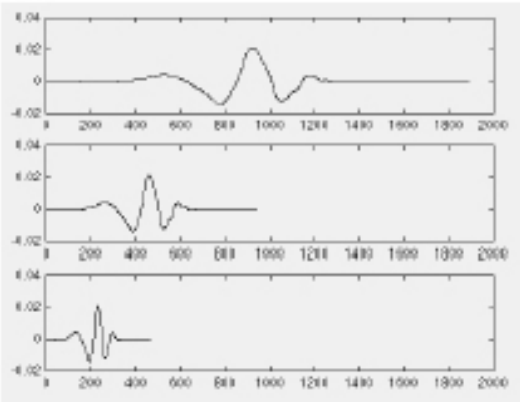


Figure 3.6: Scaling

The next subsection shows that calculating $s^{\mathbf{1}}$ and $d^{\mathbf{1}}$ or $s^{\mathbf{2}}$ and $d^{\mathbf{2}}$ for example, can be *directly* obtained by multiplying the original signal by a stretched or *scaled* version of the original wavelet or scaling signal. The more stretched the wavelet is, the more lower-frequent behavior it captures. The amount of stretching the wavelet or scaling signal depends on the level of decomposition. Before going into the mathematical matter, it is important to mention the notion of scaling here. The mathematical term for scaling is *dilation*. The notion of scaling is illustrated in figure 3.6. The complete collection of translated and dilated wavelets forms a *wavelet family*, where each member is derived from the original wavelet or *mother wavelet*. In the same way, there is also a family of scaling signals.

By using translated and dilated versions of the mother wavelet or scaling signal, a window variable in both time and scale can be shifted along the signal. The translating or shifting ensures that all elements in the signal are captured. Dilating ensures that all behavior in the signal is captured. The more dilated or stretched the wavelet or scaling signal is, the lower frequent behavior it captures. All these ideas can be fit into a mathematical framework. This is discussed in the next subsection.

3.2.2 The Discrete Wavelet Transform - mathematical framework

So far, a general idea of the discrete wavelet transform is given. Although the general idea might seem straightforward, the technique is based on a mathematical framework. This is discussed in this subsection. The theories are derived from Walker (2000), Gençay, Selçuk and Whitcher (2002), Chew (2001), Misiti, Misiti, Oppenheim and Poggi (2002) and Norsworty, Li and Gorener (2000).

Wavelets are functions that satisfy certain properties or requirements. The previous subsection mentioned the idea of the wavelet family. The family of *scaling signals* is defined as:

$$\Phi_{j,k}(t) = \frac{1}{\sqrt{2^j}}\Phi\left(\frac{t - 2^j(k-1)}{2^j}\right)$$

given that

$$\{j = 1, 2...J, k = 1, 2...K | J \leq^2 \log N, K = \frac{N}{2^j}\} \tag{3.1}$$

The wavelet family is defined as:

$$\Psi_{j,k}(t) = \frac{1}{\sqrt{2^j}}\Psi\left(\frac{t - 2^j(k-1)}{2^j}\right)$$

given that

$$\{j = 1, 2...J, k = 1, 2...K | J \leq^2 \log N, K = \frac{N}{2^j}\} \tag{3.2}$$

N equals the number of elements in the original signal.

Although most wavelet families have both scaling signals and wavelets, not all do. Daubechies (1992) gives an extensive discussion about these families.

Each element $\Phi_{j,k}$ from the set as described in (3.1) - which is a scaling signal - is obtained by taking dilations and translations from the scaling signal $\Phi(t)$.

Each element $\Psi_{j,k}$ from the set as described in (3.2) - which is a wavelet - is obtained by taking dilations and translations from the *mother* wavelet $\Psi(t)$.

The term $2^j(k-1)$ in equations (3.1) and (3.2) results into a translation of $2^j(k-1)$ to the right of the support of the wavelet or scaling signal, whereas the term 2^j in the denominator of equations (3.1) and (3.2) results into a dilation of the support of the wavelet or scaling signal. The $(k-1)$ implies that the first wavelet or scaling signal belonging to level j does not translate: it is already in the right position. The other wavelets and scaling signals do translate, with steps based on powers of two. The transform is called discrete, as the scales and positions are dyadic: they are based on powers of two. The multiplication by $\frac{1}{\sqrt{2}}$ is necessary to ensure normality. The term normality is explained later on. For a detailed explanation regarding the particular choice of $\frac{1}{\sqrt{2}}$, I refer to Walker (2000).

Definition 2 *Consider a signal f_m with time-indices m. Then, the support of this signal is the set of time-indices m where $f_m \neq 0$*

The discrete wavelet transform is not the only wavelet transform available. The Continuous Wavelet Transform (CWT) also exists. The difference between these transforms is that the continuous transform does not restrict itself to the use of powers of two, but uses any value for dilating and translating. A result of this is that the continuous transform is a redundant approach, in the sense that using powers of two already enables an exact reconstruction of the signal. Reconstruction is discussed in section 3.3. For the purpose of this thesis, the focus lies on the discrete wavelet transform only, which is basically the most often used approach.

Moreover, scaling signal $\Phi_{j,k}$ denotes the k^{th} scaling signal from the collection of scaling signals *belonging* to level j, for $k = 1, 2...\frac{N}{2^j}$. Likewise, wavelet $\Psi_{j,k}$ denotes the k^{th} wavelet from the collection of wavelets *belonging* to level j, for $k = 1, 2...\frac{N}{2^j}$.

Scaling signals are defined to conform to the restriction:

$$\int \Phi(t)dt = 1, \tag{3.3}$$

whereas mother wavelets are defined to conform to the restriction:

$$\int \Psi(t)dt = 0 \tag{3.4}$$

The name wavelet comes from this restriction, requiring the function to integrate to zero, waving above and below the x-axis. The families are orthogonal, as the scaling signals and mother wavelets are *orthogonal*. *The collection of translated and dilated scaling signals obtained from the function $\Phi(t)$ and the*

collection of translated and dilated wavelets obtained from the function $\Psi(t)$, together form an orthogonal basis in $L^2(\Re)$, which is the space of all square integrable functions. This implies that any element in $L^2(\Re)$ may be represented as a linear combination of these basis functions. The orthogonality property ensures that

$$\int \Phi_{j,k}(t)\Phi_{j,k'}(t) = 0 \text{ for } k \neq k' \tag{3.5}$$

holds for the family of scaling signals and

$$\int \Psi_{j,k}(t)\Psi_{j,k'}(t) = 0 \text{ for } k \neq k' \tag{3.6}$$

holds for the wavelet family.

It also holds *between* families:

$$\forall k, k' : \int \Phi_{j,k}(t)\Psi_{j,k'}(t) = 0 \tag{3.7}$$

For example, the following equations should hold in order to preserve the orthogonality property:

$$\int \Phi_{1,1}(t)\Phi_{1,2}(t) = 0 \text{ and } \int \Phi_{3,2}(t)\Phi_{3,4}(t) = 0$$

but also

$$\int \Psi_{1,1}(t)\Psi_{1,2}(t) = 0 \text{ and } \int \Psi_{3,2}(t)\Psi_{3,4}(t) = 0.$$

and finally

$$\int \Phi_{1,1}(t)\Psi_{1,2}(t) = 0 \text{ and } \int \Psi_{2,2}(t)\Psi_{2,2}(t) = 0.$$

In addition, the term $\frac{1}{\sqrt{2^j}}$ next to equations (3.3) and (3.4) makes the basis not only orthogonal in $L^2(\Re)$, but orthonormal in $L^2(\Re)$ as well. In order to ensure the property of orthonormality, the following equations should hold:

$$\int \Phi_{j,k}(t)\Phi_{j,k}(t) = 1 \tag{3.8}$$

and

$$\int \Psi_{j,k}(t)\Psi_{j,k}(t) = 1 \tag{3.9}$$

An orthogonal basis satisfying normality is said to be orthonormal.

In the previous subsection was mentioned that the scaling *coefficients* $s_{j,k}$ capture the trend of the original signal $f(t)$. Each scaling coefficient is defined as:

$$s_{j,k} = \int \Phi_{j,k}(t) f(t) dt$$

given that

$$\{j = 1, 2...J, k = 1, 2...K | J \leq^2 \log N, K = \frac{N}{2^j}\} \tag{3.10}$$

The wavelet *coefficients* $d_{j,k}$ capture the deviations from the trend. Each wavelet coefficient is defined as:

$$d_{j,k} = \int \Psi_{j,k}(t) f(t) dt$$

given that

$$\{j = 1, 2...J, k = 1, 2...K | J \leq^2 \log N, K = \frac{N}{2^j}\} \tag{3.11}$$

Scaling coefficient $s_{j,k}$ denotes the k^{th} scaling coefficient from the vector of scaling coefficients $s^{\mathbf{j}}$ *belonging* to level j, for $k = 1, 2... \frac{N}{2^j}$. Likewise, wavelet coefficient $d_{j,k}$ denotes the k^{th} wavelet coefficient from the vector of wavelet coefficients $d^{\mathbf{j}}$ *belonging* to level j, for $k = 1, 2... \frac{N}{2^j}$. Therefore, each level j brings $K = \frac{N}{2^j}$ elements of both scaling- and wavelet coefficients.

The vector of scaling coefficients belonging to level $j = 1$ or first series of scaling coefficients $s^{\mathbf{1}}$, is in practice a vector where each element is obtained by applying the scalar product on the original signal and the corresponding level 1 scaling signal. Likewise the second series of scaling coefficients $s^{\mathbf{2}}$ is in practice a vector where each element is obtained by applying the scalar product on the original signal and the corresponding level 2 scaling signal. More general, the j^{th} series of scaling coefficients $s^{\mathbf{j}}$ is a vector where each element is obtained by applying the scalar product on the original signal and the corresponding level j scaling signal. Assume here that both the original function and corresponding scaling signals are row-vectors:

$$\forall t : s^{\mathbf{j}} = (f(t)\Phi^{\mathbf{T}}_{j,1}(t), f(t)\Phi^{\mathbf{T}}_{j,2}(t)...f(t)\Phi^{\mathbf{T}}_{j,\frac{N}{2^j}}(t)) \tag{3.12}$$

This vector of scaling coefficients captures the trend at level j.

The vector of wavelet coefficients belonging to level $j = 1$ or the first series of wavelet coefficients $d^{\mathbf{1}}$ is in practice a vector where each element is obtained by applying the scalar product on the original signal and the corresponding level 1 wavelet. Likewise the second series of wavelet coefficients $d^{\mathbf{2}}$ is in practice a vector where each element is obtained by applying the scalar product on the original signal and the corresponding level 2 wavelet. More general, the j^{th} series of wavelet coefficients $d^{\mathbf{j}}$ is a vector where each element is obtained by applying the scalar product on the original signal and the corresponding level j wavelet. Assume here that both the original function and corresponding wavelets are row-vectors:

$$\forall t : d^{\mathbf{j}} = (f(t)\Psi^{\mathbf{T}}_{j,1}(t), f(t)\Psi^{\mathbf{T}}_{j,2}(t)...f(t)\Psi^{\mathbf{T}}_{j,\frac{N}{2^j}}(t)) \tag{3.13}$$

This vector of wavelet coefficients captures the detail at level j.

So far, we discussed the discrete wavelet transform, which entails in short that the original signal $f(t)$ is decomposed into two series of coefficients: a series containing scaling coefficients, which captures the trend and a series containing wavelet coefficients, capturing all deviations from the trend. These vectors of scaling- and wavelet coefficients, are half the length of the original signal. Then, the complete procedure is repeated on the series capturing the trend. Generally spoken, as there are 2^J elements in the original signal, this signal can be analyzed at most at J levels. More precisely, $d^{\mathbf{j}}$ is a length $\frac{N}{2^j}$ vector associated with changes on a *scale* of length $\lambda_j = 2^{j-1}$ and $s^{\mathbf{j}}$ is a length $\frac{N}{2^j}$ vector associated with changes on a *scale* of length $2^j = 2\lambda_j$. The following example illustrates the discrete wavelet transform, by using a signal of 8 samples. Therefore, a ($^2 \log 8 =$) 3-level decomposition can be performed.

Example 3.2.1 *Consider the signal:*

$$f(t) = \begin{bmatrix} 5 & | & t \in [0,1\rangle \\ 7 & | & t \in [1,2\rangle \\ 10 & | & t \in [2,3\rangle \\ 11 & | & t \in [3,4\rangle \\ 8 & | & t \in [4,5\rangle \\ 5 & | & t \in [5,6\rangle \\ 4 & | & t \in [6,7\rangle \\ 4 & | & t \in [7,8\rangle \end{bmatrix}$$

This signal is denoted as $f(t) = (5, 7, 10, 11, 8, 5, 4, 4)$.

For illustration purposes, we use the simplest wavelets, the Haar wavelets. The Haar wavelet is a step function, taking values 1 and -1 on $[0, \frac{1}{2})$ *and* $[\frac{1}{2}, 1)$ *and 0 elsewhere. The Haar scaling signal takes the value 1 on* $[0, 1)$ *and 0 elsewhere. The Haar wavelet is discussed in more detail in section 3.4, in addition to the other orthogonal wavelets. The 1-level Haar scaling signals for the trend signal are defined as:*

$$\begin{aligned} \Phi_{1,1} &= \left(\tfrac{1}{\sqrt{2}}, \tfrac{1}{\sqrt{2}}, 0, 0, \cdots, 0\right) \\ \Phi_{1,2} &= \left(0, 0, \tfrac{1}{\sqrt{2}}, \tfrac{1}{\sqrt{2}}, 0, 0, \cdots, 0\right) \\ &\vdots \\ \Phi_{1,\frac{N}{2^1}} &= \left(0, 0, \cdots, 0, \tfrac{1}{\sqrt{2}}, \tfrac{1}{\sqrt{2}}\right) \end{aligned}$$

The level-1 Haar wavelets for the detail signal are defined as:

$$\begin{aligned} \Psi_{1,1} &= \left(\tfrac{1}{\sqrt{2}}, \tfrac{-1}{\sqrt{2}}, 0, 0, \cdots, 0\right) \\ \Psi_{1,2} &= \left(0, 0, \tfrac{1}{\sqrt{2}}, \tfrac{-1}{\sqrt{2}}, 0, 0, \cdots, 0\right) \\ &\vdots \\ \Psi_{1,\frac{N}{2^1}} &= \left(0, 0, \cdots, 0, \tfrac{1}{\sqrt{2}}, \tfrac{-1}{\sqrt{2}}\right) \end{aligned}$$

Recall that these wavelets are constructed by taking dilations and translation from the original Haar wavelet, according to equations (3.1) and (3.2). Figure 3.7 shows plots of the scalings signals (left) and wavelets (right) belonging to level 1. Recall as well the special properties of orthogonality and orthonormality. For example:

$$\Phi_{1,1}\Phi_{1,2}^T = \left(\tfrac{1}{\sqrt{2}}, \tfrac{1}{\sqrt{2}}, 0, 0, \cdots, 0\right)\left(0, 0, \tfrac{1}{\sqrt{2}}, \tfrac{1}{\sqrt{2}}, 0, 0, \cdots, 0\right)^T = 0$$

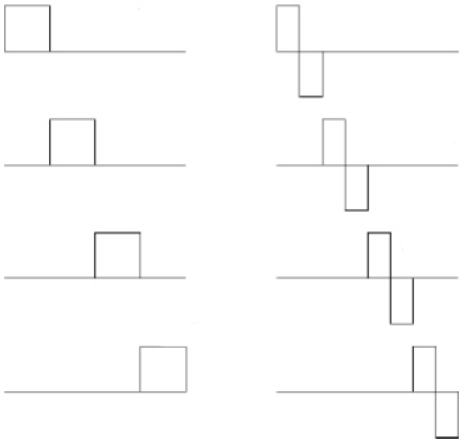


Figure 3.7: Left: scaling signals belonging to level 1. Right: wavelets belonging to level 1.

which verifies orthogonality and

$$\Phi_{1,1}\Phi_{1,1}^T = \left(\tfrac{1}{\sqrt{2}}, \tfrac{1}{\sqrt{2}}, 0, 0, \cdots, 0\right)\left(\tfrac{1}{\sqrt{2}}, \tfrac{1}{\sqrt{2}}, 0, 0, \cdots, 0\right)^T = 1$$

which verifies normality. An orthogonal basis satisfying normality is said to be orthonormal.

Next, when applying equation (3.12) to $f(t)$, the vector of scaling coefficients capturing the trend is:

$$s^{\mathbf{1}} = \left(6\sqrt{2}, \frac{21}{2}\sqrt{2}, \frac{13}{2}\sqrt{2}, 4\sqrt{2}\right)$$

Likewise, after applying (3.13) to $f(t)$, the vector of wavelet coefficients capturing the detail is:

$$d^{\mathbf{1}} = \left(-\sqrt{2}, -\frac{1}{2}\sqrt{2}, \frac{3}{2}\sqrt{2}, 0\right)$$

Notice here that the length of these signals is only half the length of the original signal. For the next two vectors of coefficients, $s^{\mathbf{2}}$ and $d^{\mathbf{2}}$, the level-2 Haar scaling signals and wavelets have to be defined. As already mentioned, the level-2 Haar scaling signals are constructed by translating and dilating the original Haar scaling signals:

$$\begin{aligned}
&\Phi_{2,1} = \left(\tfrac{1}{2}, \tfrac{1}{2}, \tfrac{1}{2}, \tfrac{1}{2}, 0, 0, \cdots, 0\right)\\
&\Phi_{2,2} = \left(0, 0, 0, 0, \tfrac{1}{2}, \tfrac{1}{2}, \tfrac{1}{2}, \tfrac{1}{2}, 0, 0, \cdots, 0\right)\\
&\vdots\\
&\Phi_{2,\frac{N}{2^2}} = \left(0, 0, \cdots, 0, \tfrac{1}{2}, \tfrac{1}{2}, \tfrac{1}{2}, \tfrac{1}{2}\right)
\end{aligned}$$

Likewise, the level-2 Haar wavelets for the detail signal are defined as:

$$\begin{aligned}
&\Psi_{2,1} = \left(\tfrac{1}{2}, \tfrac{1}{2}, \tfrac{-1}{2}, \tfrac{-1}{2}, 0, 0, \cdots, 0\right)\\
&\Psi_{2,2} = \left(0, 0, 0, 0, \tfrac{1}{2}, \tfrac{1}{2}, \tfrac{-1}{2}, \tfrac{-1}{2}, 0, 0, \cdots, 0\right)\\
&\vdots\\
&\Psi_{2,\frac{N}{2^2}} = \left(0, 0, \cdots, 0, \tfrac{1}{2}, \tfrac{1}{2}, \tfrac{-1}{2}, \tfrac{-1}{2}\right)
\end{aligned}$$

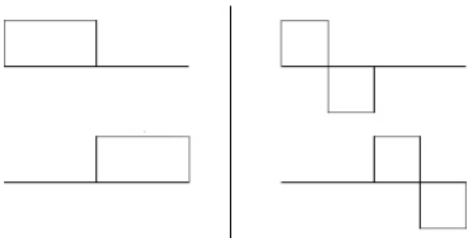


Figure 3.8: Left: scaling signals belonging to level 2. Right: wavelets belonging to level 2.

Figure 3.8 shows plots of the scaling signals and wavelets belonging to level 2.

When applying equation (3.12) to $f(t)$, the vector of scaling coefficients capturing the level-2 trend is:

$$s^{\mathbf{2}} = \left(\frac{33}{2}, \frac{21}{2}\right)$$

When applying equation (3.13) to $f(t)$, the vector of wavelet coefficients capturing the level-2 detail is:

$$d^{\mathbf{2}} = \left(\frac{-9}{2}, \frac{5}{2}\right)$$

Again, these signals are a quarter of the length of the original signal, or half the length of $s^{\mathbf{1}}$ and $d^{\mathbf{1}}$. In a similar fashion, $s^{\mathbf{3}}$ and $d^{\mathbf{3}}$ can be found:

$$s^{\mathbf{3}} = \left(\frac{27}{\sqrt{2}}\right)$$

$$d^{\mathbf{3}} = \left(\frac{6}{\sqrt{2}}\right)$$

Example 3.2.1 shows the way of obtaining these vectors of coefficients by applying equation (3.1) and (3.2), which is the formal mathematical procedure. However, this is the case with the Haar wavelet. When using other wavelets, for example a Daubechies wavelet, one has to redefine the scaling signals and wavelets for each level.

When looking carefully, one can see what really happens; the first vector of scaling coefficients capturing the trend, $s^{\mathbf{1}}$, can be obtained by systematically taking the average of two successive elements of the original signal, and multiplying these by $\sqrt{2}$. In a nearly similar fashion, $d^{\mathbf{1}}$ can be obtained by taking half the difference of two successive elements and multiplying the result by $\sqrt{2}$. The second pair of scaling- and wavelet coefficients, $s^{\mathbf{2}}$ and $d^{\mathbf{2}}$ are obtained by repeating the same process, but on $s^{\mathbf{1}}$. Finally, $s^{\mathbf{3}}$ and $d^{\mathbf{3}}$ are obtained by repeating the same process on $s^{\mathbf{2}}$. This explains why there exists only half the amount of elements in the signal of the next higher level. When using different wavelets, the idea of averaging remains the same. In most cases, the only difference is that more values are taken for calculating the average. The complete procedure remains the same as well.

To see this, consider tables 3.1, 3.2 and 3.3. We can see here that $s^{\mathbf{1}}$ is obtained by systematically taking the average of two successive elements of the original signal and multiplying this by $\sqrt{2}$. Table 3.2 shows

$\mathbf{f}$	5	7	10	11	8	5	4	4
$\mathbf{s^1}$	$6\sqrt{2}$		$\frac{21}{2}\sqrt{2}$		$\frac{13}{2}\sqrt{2}$		$4\sqrt{2}$	
$\mathbf{d^1}$	$-\sqrt{2}$		$\frac{-1}{2}\sqrt{2}$		$\frac{3}{2}\sqrt{2}$		0	

Table 3.1: Decomposition process - step 1

$\mathbf{s^1}$	$6\sqrt{2}$	$\frac{21}{2}\sqrt{2}$	$\frac{13}{2}\sqrt{2}$	$4\sqrt{2}$
$\mathbf{s^2}$	$\frac{33}{2}$		$\frac{21}{2}$	
$\mathbf{d^2}$	$\frac{-9}{2}$		$\frac{5}{2}$	

Table 3.2: Decomposition process - step 2

that the same idea is repeated on the previously obtained $s^{\mathbf{1}}$ to get $s^{\mathbf{2}}$. Table 3.3 shows the repetition on $s^{\mathbf{2}}$ to get $s^{\mathbf{3}}$. $d^{\mathbf{2}}$ however, is obtained by systematically taking half the difference of two successive elements of $s^{\mathbf{1}}$ and multiplying this by $\sqrt{2}$. The same holds for $d^{\mathbf{3}}$, which is obtained by systematically taking half the difference of two successive elements from $s^{\mathbf{2}}$ and multiplying this by $\sqrt{2}$. Notice here that taking half the difference of a pair and multiplying this by $\sqrt{2}$, is the same as taking a difference of a pair and dividing this by $\sqrt{2}$.

3.3 The wavelet transform - reconstruction

Section 3.2 showed how the discrete wavelet transform can be performed in order to analyze, or decompose, signals. This process is called decomposition or analysis. The other half of the story is how those components can be assembled back into the original signal without loss of information. This process is called reconstruction, or synthesis. The mathematical manipulation that effects synthesis is called the inverse discrete wavelet transform (IDWT). This discussion thus shows how discrete signals are synthesized by beginning with a very low resolution signal and successively adding on details to create higher resolution versions, ending with a complete synthesis of the signal at the finest resolution.

With respect to the 1-level Haar wavelet transform, as performed previously in example 3.2.1, the original signal $f(t)$ can be mapped back via the following equation:

$$f(t) = \left(\frac{s_{1,1} + d_{1,1}}{\sqrt{2}}, \frac{s_{1,1} - d_{1,1}}{\sqrt{2}}, \cdots, \frac{s_{1,\frac{N}{2^1}} + d_{1,\frac{N}{2^1}}}{\sqrt{2}}, \frac{s_{1,\frac{N}{2^1}} - d_{1,\frac{N}{2^1}}}{\sqrt{2}} \right) \tag{3.14}$$

In this equation, $s_{1,1}$ denotes the first element of $s^{\mathbf{1}}$ and $d_{1,1}$ denotes the first element of $d^{\mathbf{1}}$.

Equation (3.14) can also be written as:

$\mathbf{s^2}$	$\frac{33}{2}$	$\frac{21}{2}$
$\mathbf{s^3}$	$\frac{27}{\sqrt{2}}$	
$\mathbf{d^3}$	$\frac{6}{\sqrt{2}}$	

Table 3.3: Decomposition process - step 3

$$f(t) = A^{\mathbf{1}} + D^{\mathbf{1}}, \tag{3.15}$$

where

$$A^{\mathbf{1}} = \left(\frac{s_{1,1}}{\sqrt{2}}, \frac{s_{1,1}}{\sqrt{2}}, \frac{s_{1,2}}{\sqrt{2}}, \frac{s_{1,2}}{\sqrt{2}} \cdots, \frac{s_{1,\frac{N}{2^{\mathbf{1}}}}}{\sqrt{2}}, \frac{s_{1,\frac{N}{2^{\mathbf{1}}}}}{\sqrt{2}} \right)$$

and

$$D^{\mathbf{1}} = \left(\frac{d_{1,1}}{\sqrt{2}}, \frac{-d_{1,1}}{\sqrt{2}}, \frac{d_{1,2}}{\sqrt{2}}, \frac{-d_{1,2}}{\sqrt{2}} \cdots, \frac{d_{1,\frac{N}{2}}}{\sqrt{2}}, \frac{-d_{1,\frac{N}{2^{\mathbf{1}}}}}{\sqrt{2}} \right)$$

$A^{\mathbf{1}}$ is in the literature often referred to as the level-1 approximation of the original signal, whereas $D^{\mathbf{1}}$ is often referred to as the level-1 detail of the original signal.

When recalling equations (3.12) and (3.13), equation (3.15) is expressible as:

$$A^{\mathbf{1}} = \left(s_{1,1}\Phi_{1,1} + s_{1,1}\Phi_{1,2}, \cdots, +s_{1,\frac{N}{2^{\mathbf{1}}}}\Phi_{1,\frac{N}{2^{\mathbf{1}}}} \right) \tag{3.16}$$

$$D^{\mathbf{1}} = \left(d_{1,1}\Psi_{1,1} + d_{1,1}\Psi_{1,2}, \cdots, +d_{1,\frac{N}{2^{\mathbf{1}}}}\Psi_{1,\frac{N}{2^{\mathbf{1}}}} \right) \tag{3.17}$$

The signal $f(t)$ is expressible as equation (3.15), when this signal is analyzed at level 1. If the signal is analyzed at level 2, $f(t) = A^{\mathbf{2}} + D^{\mathbf{2}} + D^{\mathbf{1}}$ and $f(t) = A^{\mathbf{3}} + D^{\mathbf{3}} + D^{\mathbf{2}} + D^{\mathbf{1}}$ if the signal is analyzed at level 3. More general, if the number N of signal values is divisible J times by 2, then a J-level multiresolution-analysis:

$$f(t) = A^{\mathbf{J}} + D^{\mathbf{J}} + D^{\mathbf{J-1}}+, \cdots, +D^{\mathbf{2}} + D^{\mathbf{1}} \tag{3.18}$$

can be performed on $f(t)$. Multiresolution analysis (MRA) is the heart of wavelet analysis. Here, the signal is built up from a low resolution signal $A^{\mathbf{J}}$ and detail signals $D^{\mathbf{J}}+, \cdots, +D^{\mathbf{2}} + D^{\mathbf{1}}$ are added to form the original signal.

Finally, the discrete wavelet approximation, generalizing all of the ideas as described above, can be expressed as:

$$f(t) \approx \underbrace{\sum_{k=1}^{K} s_{J,k}\Phi_{J,k}(t)}_{\mathrm{A^J}} + \underbrace{\sum_{k=1}^{K} d_{J,k}\Psi_{J,k}(t)}_{\mathrm{D^J}} + \underbrace{\sum_{k=1}^{K} d_{J-1,k}\Psi_{J-1,k}(t)}_{\mathrm{D^{J-1}}} + \cdots + \underbrace{\sum_{k=1}^{K} d_{1,k}\Psi_{1,k}(t)}_{\mathrm{D^1}}, \tag{3.19}$$

given that

$$\{J \leq^2 \log N, K = \frac{N}{2^j}\}$$

which is most common in the wavelet literature.

Example 3.3.1 *Consider again the signal* $f(t) = (5, 7, 10, 11, 8, 5, 4, 4)$. *Through use of equations (3.18) and (3.19),* $A^{\mathbf{3}}$ *can be calculated as:*

$$A^{\mathbf{3}} = \frac{27}{\sqrt{2}}\Phi_{3,1} = \frac{27}{\sqrt{2}} \left(\frac{1}{\sqrt{8}}, \frac{1}{\sqrt{8}}, \frac{1}{\sqrt{8}}, \frac{1}{\sqrt{8}}, \frac{1}{\sqrt{8}}, \frac{1}{\sqrt{8}}, \frac{1}{\sqrt{8}}, \frac{1}{\sqrt{8}} \right)$$

$A^{\mathbf{3}}$	6.75	6.75	6.75	6.75	6.75	6.75	6.75	6.75
$D^{\mathbf{3}}$	1.5	1.5	1.5	1.5	-1.5	-1.5	-1.5	-1.5
$D^{\mathbf{2}}$	-2.25	-2.25	2.25	2.25	1.25	1.25	-1.25	-1.25
$D^{\mathbf{1}}$	-1	1	-0.5	0.5	1.5	-1.5	0	0
$f(t)$	5	7	10	11	8	5	4	4

Table 3.4: results of summation of individual components

$$A^{\mathbf{3}} = \left(6\frac{3}{4}, 6\frac{3}{4}, 6\frac{3}{4}, 6\frac{3}{4}, 6\frac{3}{4}, 6\frac{3}{4}, 6\frac{3}{4}, 6\frac{3}{4}\right)$$

Similarly, $D^{\mathbf{3}}$ can be calculated as:

$$D^{\mathbf{3}} = \frac{6}{\sqrt{2}}\Psi_{3,1} = \frac{6}{\sqrt{2}}\left(\frac{1}{\sqrt{8}}, \frac{1}{\sqrt{8}}, \frac{1}{\sqrt{8}}, \frac{1}{\sqrt{8}}, \frac{-1}{\sqrt{8}}, \frac{-1}{\sqrt{8}}, \frac{-1}{\sqrt{8}}, \frac{-1}{\sqrt{8}}\right)$$

$$D^{\mathbf{3}} = \left(1\frac{1}{2}, 1\frac{1}{2}, 1\frac{1}{2}, 1\frac{1}{2}, -1\frac{1}{2}, -1\frac{1}{2}, -1\frac{1}{2}, -1\frac{1}{2}\right)$$

Likewise, $D^{\mathbf{2}}$ and $D^{\mathbf{1}}$ can be calculated, yielding:

$$D^{\mathbf{2}} = \left(-2\frac{1}{4}, -2\frac{1}{4}, 2\frac{1}{4}, 2\frac{1}{4}, 1\frac{1}{4}, 1\frac{1}{4}, -1\frac{1}{4}, -1\frac{1}{4}\right)$$

$$D^{\mathbf{1}} = \left(-1, 1, \frac{-1}{2}, \frac{1}{2}, 1\frac{1}{2}, -1\frac{1}{2}, 0, 0\right)$$

Here can be noticed that each of the components has the same length as the original signal. Finally, by adding up all of the individual components - Excel proved to be useful - one can see in table 3.4, that the sum equals the original signal f(t). This final example verifies all of the ideas and confirms that dyadic steps which are taken in the DWT, as mentioned in the previous section, are enough for exact reconstruction of the signal. This indicates again that the CWT is a redundant transform.

3.4 Wavelet families

The variety of wavelets is not restricted by the Haar wavelet. There are many more wavelets, each characterized by its amount of symmetry, vanishing moments, support width and regularity. For the purpose of this research, there will be only focused on the family of orthogonal, compactly supported wavelets. For an extensive overview of the wavelet families, the mathematical background and their properties, I refer to Daubechies (1992).

3.4.1 Exploiting wavelet characteristics

Each wavelet has its own characteristics which makes the wavelet more useful in some situations. Four important characteristics can be distinguished:

- Support width
- vanishing moments
- Symmetry
- Regularity

Essential here, is to know when to use which wavelet. Therefore, each of these characteristics will be discussed in more detail in the following sections.

Support width

All of the wavelets as described above, are compactly supported wavelets, implying that the wavelet has a "rapid fall-off". This means that the basis functions are non-zero only on a finite interval. The compact support of the wavelet basis functions allows the wavelet transform to efficiently represent functions or signals which have localized features. Compact support width allows for reduced computational complexity, better time-resolution but poorer frequency resolution (Misiti et al. 2002).

Vanishing moments

The m^{th} moment of a wavelet is defined as:

$$\int t^m \Psi(t) dt \tag{3.20}$$

If the first M moments of a wavelet are zero, then all polynomial type signals of the form

$$x(t) = \sum_{m=0}^{M} c_m t^m$$

have (near) zero wavelet coefficients (Misiti et al. 2002). If we use a wavelet with enough number of vanishing moments, M, to analyze a polynomial with a degree less than M, then all detail coefficients will be zero which allows for an excellent compression ratio. The compression ratio is therefore useful in another domain of wavelet analysis, called compression.[1] Compression can be used for compressing the information captured in sounds and images, in order to decrease the amount of space they use when stored on a hard drive or other medium. Compression through wavelet analysis however, lies not in the scope of this thesis and is therefore not further discussed. For a discussion regarding compression through wavelet analysis however, I refer to Walker (2000). Although most functions are not polynomial, each function can locally be expressed as a polynomial when expanded as a Taylor-series.

[1] Compression can be done by performing a wavelet transform and removing all wavelet coefficients below a certain threshold. When reconstructing the signal, no significant loss is found.

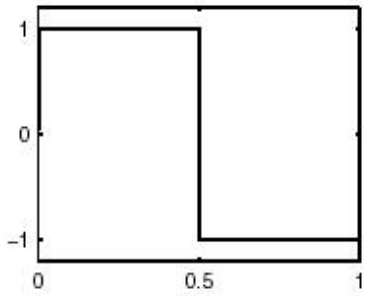


Figure 3.9: The Haar wavelet

Symmetry

Symmetry implies that a wavelet or scaling signal is symmetrically formed. Symlets for example, are the most symmetrical wavelets. Symmetry can be useful in the area of image processing. Image processing, which is two dimensional wavelet analysis, is not within the scope of this thesis and is therefore not discussed. For a discussion regarding image processing through wavelet analysis however, I refer to Walker (2000).

Regularity

Regularity or smoothness can be seen as the number of times a function can be differentiated at any given point, which is closely related to the number of vanishing moments a wavelet has. The more vanishing moments a wavelet has, the smoother the wavelet is. Smoothness provides;

- numerical stability
- better reconstruction properties
- necessary for certain applications, such as the solution of diff. equations

3.4.2 Haar wavelets

The Haar wavelet is the oldest and simplest wavelet. The Haar wavelet resembles a step function, taking the value 1 on $[0, \frac{1}{2})$ and the value -1 $[\frac{1}{2}, 1)$.

The Haar wavelet has one vanishing moment and a support width of one. As illustrated before, the Haar transform is performed by computing running averages and differences with scaling signals and wavelets. Figure 3.9 displays the Haar wavelet.

3.4.3 Daubechies wavelets

These wavelets are proposed by Ingrid Daubechies, a familiar name in wavelet research. The daubechies transform are defined in the same way as the Haar transform, by computing running averages and differences. However, these averages and differences are computed with different scaling signals and wavelets, and, as the support of these wavelets is larger, these averages and differences are computed by using

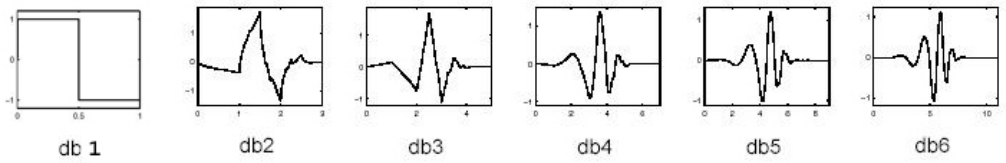


Figure 3.10: Daubechies wavelets

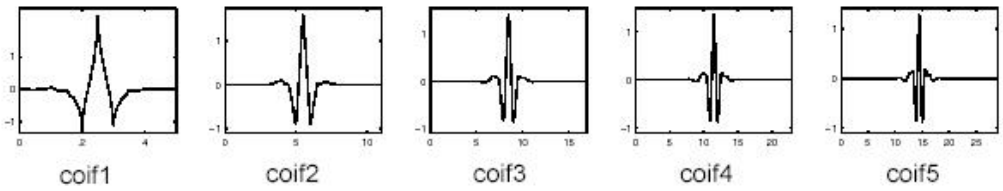


Figure 3.11: Coiflets

more values of the signal, providing a tremendous improvement in the capabilities of these transforms. They provide a set of tools including compression and noise removal for audio signals and for images (Walker 2000). The Daubechies family is a family with compactly supported wavelets, with the highest number of vanishing moments given this support width. The daubechies wavelets are indicated by their name (i.e. db or Daub) and order. The order indicates the support width. The db1 wavelet is similar to the Haar wavelet. The db$N$ wavelet has a support width of $2N - 1$ and N vanishing moments. Some authors use $2N$ instead of N. Daubechies wavelets are not symmetric. Figure 3.10 displays a series of Daubechies wavelets.

3.4.4 Coiflets

This family of compactly supported wavelets are designed for the purpose of maintaining a close match between the trend values and the original signal values. Coiflets are constructed by Daubechies as well, following suggestions of Coifman (Misiti et al. 2002). Coiflets have a support width of $6N - 1$ and have the highest number of vanishing moments for both the scaling signal $(2N - 1)$ and the wavelet $(2N)$. Some authors use $2N$ instead of N These wavelets are more symmetric than daubechies wavelets. Figure 3.11 displays a series of Coiflets.

3.4.5 Symlets

Symlets are, as the name indicates, nearly symmetrical wavelets, proposed by Daubechies as a modification on the Daubechies wavelets. These compactly supported wavelets have N vanishing moments and a support width of $2N - 1$. Figure 3.12 displays a series of Symlets

3.5 Wavelets in finance, economics and soft-computing

Ramsey (1999) gives an overview of the contribution of wavelets to the analysis of economic and financial data. The ability to represent highly complex structures without knowing the underlying functional form proved to be a great benefit for the analysis of these time-series. In addition, wavelets facilitate the

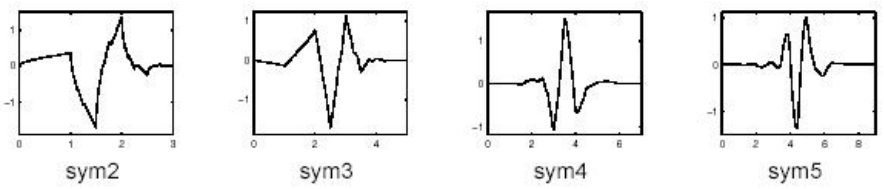


Figure 3.12: Symlets

precise location of discontinuities and the isolation of shocks. Furthermore, the process of smoothing found in the time-scale decomposition facilitates the reduction of noise in the original signal, by first decomposing the signal into the wavelet components, then eliminating all values with a magnitude below a certain threshold and finally reconstructing the original signal with the inverse wavelet transform (Walker 2000). Stevenson (2000), for example, used wavelet analysis for the filtering of spot electricity prices in the deregulated Australian electricity market. By examining both the demand and price series at different time locations and levels of resolution, Stevenson was able to reveal what was signal and what was noise. Von Sachs and McGibbon researched the contribution of wavelets to non-stationarity and complex functions. They derived the bias and variance for wavelet coefficient estimators under the assumption of local stationarity.

Ramsey and Lampart (1998) used wavelet analysis for time-scale decomposition. They researched both the relationships between consumption and income and money and GDP. The time-scale decomposition yielded a new transformed signal built up from the several wavelet coefficients representing the several scales. At each scale, a regression was made between the two variables. This research yielded three conclusions: First, the relationship between economic variables vary across different scales. Second, the decomposition resolved anomalies from the literature. Third, the research made clear that the slope relating consumption and income declines with scale. In this context, the role of real interest was strong in the consumption-income relation. Chew (2001) researched the relationship between money and income, using the same technique of wavelet-based time-scale decomposition as Ramsey and Lampart (1998) did. This research yielded a greater insight in the money versus income nexus in Germany. Arino (1996) used wavelet-based time-scale decomposition for forecasting applications. The approach used was to apply forecasting methods on each of the resulted coefficients from the time-scale decomposition. After applying forecast methods on each of these coefficients, the final forecast of the complete series was obtained by adding up the individual forecasts.

Aussem and Murtagh (1997) used neural networks to examine the individual coefficients. The trained neural network with its approximated variables in the target function, was used for the final forecast. In the area of finance, multi-resolution analysis appears useful, as different traders view the market with different time resolutions, for example hourly, daily, weekly or monthly. The shorter the time-period, the higher the frequency. Different types of traders create the multi-scale dynamics of time-series.

Struzik (2001) applied the wavelet-based effective Holder exponent to examine the correlation level of the Standard & Poor's index locally at arbitrary positions and resolutions (time and scale).

Norsworty et al. (2000) applied wavelets to analyze the relationship between the return on an asset and the return on the market portfolio, or investment alternative. Similar to other researches in the field of finance and economics, they applied wavelet-based time-scale decomposition to investigate whether there are changes in behavior for different frequencies. The research indicated that the effect of the market return on an individual asset's return will be greater in the higher frequencies than in the lower. Truly important for this thesis is the fact that the market's influence is in the higher frequency, suggests that filtering out the market noise permits the investigator to focus on longer-term movements specific to the company and its industry.

Wavelets are also applied in the area of soft-computing, especially in the area of neural networks and fuzzy systems. Tan and Yu (1999) researched the complementarity and equivalence relationships between convex fuzzy systems and wavelets. After discussing the fundamentals of both fuzzy- and wavelet systems, the authors conclude that there is a complementarity relationship. As a result, fuzzy systems and wavelets can be combined together, to represent linguistic and numerical knowledge.

In the case of function approximation, a fuzzy quantization and fuzzy rules can be constructed to effectively represent the function. In addition, wavelets can be used for further improvement of the approximation accuracy, by capturing the fine features. Furthermore, there is equivalence between multi-scale fuzzy systems and wavelets. This means that any result obtained from a multi-scale fuzzy approximation, can have its direct interpretation in an equivalent wavelet approximation. Fuzzy rules can be generated from wavelet coefficients.

Hoa, Zhang and Xu (2001) proposed a fuzzy wavelet network for function approximation. Such a FWN is built forth on wavelet neural networks. These networks are a combination of feedforward neural networks and wavelets. The main issue here is that wavelets are used as the transformation function in the hidden layer, whereas this was a sigmoid- or hyperbolic tangent function traditionally. The network is then trained by adapting the translation and dilation parameters in the wavelet function, in the variable wavelet variant. The FWN consists of four layers: input, fuzzification, inference and defuzzification layers. This is inspired by the traditional neuro-fuzzy systems, as described in Jang, Sun and Mizutani (1997). However, the difference between the FWN and the neuro-fuzzy systems is that the defuzzification-layer is built-up from a number of sub-wavelet neural networks, instead of using constants or linear equations. The FWN uses both globalized and localized approximation of the function, yielding better local accuracy and faster convergence. The input and fuzzification layer of the FWN construct the antecedents of the fuzzy rules from the input space. The sub-wavelet neural networks create the consequents of these rules, by making linear combinations of a finite set of wavelets, based on the same inputspace. Later on, in the defuzzification, the antecedents and consequents are combined to form the final output. For learning, the extended Kalman filter and least squares estimation is used. FWN's proved to be very accurate for function approximation.

Chapter 4

Wavelet analysis for Value at Risk estimation

4.1 Introduction

Chapter 2 explained the fundamentals of Value at Risk estimation and offered an overview of various time-varying volatility models currently available. Although it pointed out the importance of scale in case of different time-horizons using the Drost-Nijman equation for scaling, these models use one resolution only for all horizons. The chapter ended with section 2.6 and came with the central question: ”Does the *complete* resolution yield relevant information at all resolutions?” In other words: do all the high- and low frequency components *hidden* in the complete signal matter for *every* horizon? The section concludes with the necessity for a tool which enables the decomposition of a time-series in order to facilitate the making of distinction.

Chapter 3 explained the fundamentals of wavelet analysis and showed how a signal can be decomposed to multiple resolutions. Each resolution gives insight into behavior at different scales.

This chapter merges the two discussions into a new domain: Wavelet-based risk analysis for Value at Risk estimation, using a multiresolution approach. It shows how wavelets can be used for analyzing risk and how the correct resolution can be found for a corresponding horizon. As mentioned before, low frequent behavior might not be of interest for risk managers estimating VaR with a short horizon. At the end of the chapter, it is clear that in specific situations, not all information is relevant for the risk manager, depending on his or her specific conditions.

Chapter 1 started with a general discussion regarding risk and how risk relates to the variance of a financial time-series and finally how risk relates to Value at Risk by using the notion of risk management. It states that we refer to risk, when the decision maker has no certainty about specific future states of nature, but is aware of the factors which are on influence on the result of the decision. The spread in the outcomes of the states of nature indicate the amount of risk and is summarized as the variance. In order to link risk to wavelets, we are interested in the risk corresponding to a specific time-horizon or scale. Analogous to estimating the variance of a financial time-series in its complete resolution, we would expect that it is possible to estimate the variance at a specific resolution of the financial time-series as well. Fortunately, it is possible to estimate the variance for each level j. However, the fact that it is possible to estimate the variance for each level is due to a specific property of wavelet analysis: *the preservation of energy*. Preservation of energy ensures that the variances of the individual levels can be summed up to the variance of the original time-series, similar to the components in an MRA summing up

to the original signal. The fact that the preservation of energy ensures that the variances of the individual levels can be summed up to the variance of the original time-series, is the result of a relation between the energy in a time-series and the variance of a time-series. This relation is discussed in more detail in section 4.3.

When the variance of a specific level relates to the risk isolated to the time-horizon corresponding to this specific level, a risk manager can decide whether to include this amount of risk in his assessment, which is the fundamental idea underlying this thesis. Through the notion of energy, risk and variance can be introduced into the domain of wavelet analysis.

In order to merge these discussions, it is necessary to answer the following questions:

- Is the energy in a signal preserved when decomposing it into multiple levels?
- If the energy remains preserved, does this also hold for the risk?
- If the variance remains preserved, how can we estimate this for a specific level?
- How is the energy distributed over the levels of decomposition? More specifically: How does risk behave over the different levels?

When we have insight in the way risk behaves over the different levels, we can see which amount of risk might be of interest.

Therefore, the remainder of this chapter is structured as follows. Section 4.2 shows that energy is preserved over multiple levels. Next, section 4.3 shows how the variance can be estimated for a specific level and how this is distributed over the different levels. Finally, section 4.4 exploits the knowledge of the way risk behaves by using the wavelet energy distribution for VaR-estimation.

4.2 Preservation of energy

This section introduces wavelets in the domain of risk, by examining the variance at multiple scales, as multi-scale behavior of the variance, can bring new insights. As pointed out in chapter 3, a vector of wavelet coefficients is associated with changes at a particular scale. For example, the level j wavelet coefficients $d_{j,k}$ are associated with the scale $\lambda_j = 2^{j-1}$, essentially meaning that each wavelet coefficient was constructed using a difference of two weighted averages, each one of length λ_j. Thus, when the discrete wavelet transform is applied to a stochastic process, it produces a decomposition on a scale-by-scale basis. Therefore, the variance of the underlying process can be decomposed on a scale-by-scale basis as well. The mathematical theory of this section and section 4.3 is derived from Gençay et al. (2002).

In order to decompose the variance at multiple levels, it is necessary to show that the discrete wavelet transform is energy preserving, that is showing that the sum of all squared wavelet coefficients and the sum of squared scaling coefficients at scale J equals the sum of squared elements from the underlying process. More formally:

$$\| \mathbf{w} \|^2 = \sum_{j=1}^{J} \sum_{k=1}^{\frac{N}{2^j}} d_{j,k}^2 + s_{J,0}^2 = \sum_{t=0}^{N-1} f_t^2 = \| \mathbf{f} \|^2 \tag{4.1}$$

where $\mathbf{w}$ is defined as the vector containing wavelet coefficients of levels $j = 1, 2...J$ and the scaling coefficients of level J:

$$\mathbf{w} = \begin{bmatrix} d_{1,1} \\ d_{1,2} \\ \vdots \\ d_{1,\frac{N}{2^1}} \\ d_{2,1} \\ \vdots \\ d_{2,\frac{N}{2^2}} \\ \vdots \\ d_{J,k} \\ s_{J,k} \end{bmatrix}$$

Now, consider an $N \times N$ orthonormal matrix $\mathbf{W}$ defined as:

$$\mathbf{W} = \begin{bmatrix} \mathbf{W_1} & \mathbf{W_2} & \cdots & \mathbf{W_J} & \mathbf{V_J} \end{bmatrix}$$

with

$$\mathbf{W_j} = \begin{bmatrix} \Psi_{j,1}^T, & \Psi_{j,2}^T, & \cdots & \Psi_{j,\frac{N}{2^j}}^T \end{bmatrix}$$

for $j = 1, 2, ..., J$ and

$$\mathbf{V_j} = \begin{bmatrix} \Phi_{j,1}^T, & \Phi_{j,2}^T, & \cdots & \Phi_{j,\frac{N}{2^j}}^T \end{bmatrix}$$

for $j = J$

Then, the reconstruction of the original signal as performed using equation (3.19) can also be performed using a matrix operation:

$$\mathbf{f} = \mathbf{W}\mathbf{w} \tag{4.2}$$

This however, can be quite an exercise to see without an example.

Example 4.2.1 *Consider again the signal* $f(t) = (5, 7, 10, 11, 8, 5, 4, 4)$ *of length* 8

Now consider an 8×8 *orthonormal matrix* $\mathbf{W}$*, where the first four columns represent the 4 Haar wavelets belonging to level 1, where column 5 and 6 represent the 2 Haar wavelets belonging to level 2, where column 7 represents the Haar wavelet belonging to level 3 and where column 8 represents the Haar scaling signal belonging to 3. These Haar wavelets and Haar scaling signals are the same as used in example 3.2.1.*

Consider also a vector $\mathbf{w}$ *containing the wavelet coefficients of levels 1, 2 and 3 and the scaling coefficient of level 3. Then, using equation (4.2) the original signal f(t) can be reconstructed as follows:*

$$\mathbf{f} = \mathbf{W}\mathbf{w} = \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} \\ \frac{-1}{\sqrt{2}} & 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} \\ 0 & \frac{1}{\sqrt{2}} & 0 & 0 & \frac{-1}{2} & 0 & \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} \\ 0 & \frac{-1}{\sqrt{2}} & 0 & 0 & \frac{-1}{2} & 0 & \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} \\ 0 & 0 & \frac{1}{\sqrt{2}} & 0 & 0 & \frac{1}{2} & \frac{-1}{\sqrt{8}} & \frac{1}{\sqrt{8}} \\ 0 & 0 & \frac{-1}{\sqrt{2}} & 0 & 0 & \frac{1}{2} & \frac{-1}{\sqrt{8}} & \frac{1}{\sqrt{8}} \\ 0 & 0 & 0 & \frac{1}{\sqrt{2}} & 0 & \frac{-1}{2} & \frac{-1}{\sqrt{8}} & \frac{1}{\sqrt{8}} \\ 0 & 0 & 0 & \frac{-1}{\sqrt{2}} & 0 & \frac{-1}{2} & \frac{-1}{\sqrt{8}} & \frac{1}{\sqrt{8}} \end{bmatrix} \begin{bmatrix} -\sqrt{2} \\ \frac{-1}{2}\sqrt{2} \\ \frac{3}{2}\sqrt{2} \\ 0 \\ \frac{-9}{2} \\ \frac{5}{2} \\ \frac{6}{\sqrt{2}} \\ \frac{27}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} 5 & 7 & 10 & 11 & 8 & 5 & 4 & 4 \end{bmatrix} = \mathbf{f}$$

When considering table 3.4, we can see that the first element of $f(t)$ is a 5. The table also shows that 5 is the sum of -1, -2.25, 1.5 and 6.75. This -1 is the product of the first row of $\mathbf{W}$ and $\mathbf{w}$. Due to the ingenuity of orthonormality, $\mathbf{W}$ has zero values on exactly the right places. To get a good understanding of how everything fits together, try to relate the results in table 3.4 to the results of the product $\mathbf{f} = \mathbf{W}\mathbf{w}$.

Due to the orthonormality of $\mathbf{W}$, the following equation also holds:

$$\mathbf{W}\mathbf{W}^T = \mathbf{W}^T\mathbf{W} = \mathbf{I} \tag{4.3}$$

Then, with this knowledge and matrix algebra, it can be proven that the discrete wavelet transform preserves energy:

$$\| \mathbf{f} \|^2 = \mathbf{f}^\mathbf{T}\mathbf{f} = (\mathbf{W}\mathbf{w})^\mathbf{T}\mathbf{W}\mathbf{w} = \mathbf{w}^\mathbf{T}\mathbf{W}^\mathbf{T}\mathbf{W}\mathbf{w} = \mathbf{w}^\mathbf{T}\mathbf{w} = \| \mathbf{w} \|^2$$

In terms of a scale-by-scale decomposition, the preservation of energy can be written as:

$$\| \mathbf{f} \|^2 = \sum_{\mathbf{j}=\mathbf{1}}^{\mathbf{J}} \| \mathbf{d}^\mathbf{j} \|^2 + \| \mathbf{s}^\mathbf{J} \|^2 \tag{4.4}$$

Equation (4.4) does not only hold for the coefficients being a result of the discrete wavelet transform, but also for the coefficients of the MRA:

$$\| \mathbf{f} \|^2 = \sum_{j=1}^{J} \| \mathbf{D}^\mathbf{j} \|^2 + \| \mathbf{A}^\mathbf{J} \|^2 \tag{4.5}$$

This can be checked through a simple example, which is shown in table 4.1 First, recall table 3.4 showing that the 3-level Haar MRA sums up to the original signal, as indicated by equation (3.19). The same table is copied with the squared results below. In the final column is shown that the sum of the energies of the individual components sums up to the sum of the energy of the signal.

4.3 Estimating the wavelet variance

After demonstrating that the discrete wavelet transform is energy preserving, we can see that the sum of squared elements in a time-series equals the sum of squared coefficients in the vector $\mathbf{w}$ for that time-series. Each coefficient is obtained by averaging the original time-series by use of a scaling function

A^3	6.75	6.75	6.75	6.75	6.75	6.75	6.75	6.75	
D^3	1.5	1.5	1.5	1.5	-1.5	-1.5	-1.5	-1.5	
D^2	-2.25	-2.25	2.25	2.25	1.25	1.25	-1.25	-1.25	
D^1	-1	1	-0.5	0.5	1.5	-1.5	0	0	
$f(t)$	5	7	10	11	8	5	4	4	**sum:**
$\parallel \mathbf{A^3} \parallel^2$	45.56	45.56	45.56	45.56	45.56	45.56	45.56	45.56	**364.5**
$\parallel \mathbf{D^3} \parallel^2$	2.25	2.25	2.25	2.25	2.25	2.25	2.25	2.25	**18**
$\parallel \mathbf{D^2} \parallel^2$	5.06	5.06	5.06	5.06	1.56	1.56	1.56	1.56	**26.5**
$\parallel \mathbf{D^1} \parallel^2$	1	1	0.25	0.25	2.25	2.25	0	0	**7**
$\parallel \mathbf{f} \parallel^2$	25	49	100	121	64	25	16	16	**416**

Table 4.1: results of summation of individual components - squared

or wavelet. The variance, in addition, is obtained by calculating the sum of squared deviations and dividing the result by the number of elements, according to equation (1.1). Apparent is that both the energy of a time-series and the variance of a time-series are based on the same principle: the summation of squared elements. Gençay et al. (2002) show in fact that through the sum of squared wavelet- and scaling coefficients, an unbiased estimator can be found for the wavelet variance based on the DWT. For the derivation of this unbiased estimator, I refer to Gençay et al. (2002). For the purpose of this thesis, it is sufficient to use the following equations, also based on the theories of Gençay et al. (2002).

We can define the wavelet variance associated with scale λj for the underlying process $f(t)$ as follows:

$$\sigma_f^2(\lambda_j) = \frac{1}{2\lambda_j}\text{var}(d_{j,k}) \tag{4.6}$$

In addition, the wavelet variance decomposes the variance on a scale-by-scale basis, using:

$$\sum_{j=1}^{\infty} \sigma_f^2(\lambda_j) = \text{var}(f(t)) \tag{4.7}$$

Recall from chapter 3 that each level j brings $\frac{N}{2^j}$ wavelet coefficients, implying that the variance is estimated over a decreasing number of coefficients when the level j increases. Therefore it is necessary to multiply with $\frac{1}{2\lambda_j}$ in order to make the variance at each level proportional to the number of coefficients belonging to that level.

The wavelet variance can also be calculated on the basis of an MRA:

$$\sigma_f^2(\lambda_j) = \text{var}(D^{\mathbf{j}}) \tag{4.8}$$

This time, there is no need to multiply by $\frac{1}{2\lambda_j}$, as the MRA is already reconstructed over the same time-domain as $f(t)$.

As a demonstration of these ideas, consider the signal $u(t)$ as the returns on the Dow-Jones index measured from 22/02/1972 to 22/05/2003. The returns are calculated using equation (2.1), so that the log-difference between the closing and opening of a single trading-day is taken for all days. The complete time-series represents 7889 observations, which is shown in figure 4.1. In addition, figure 4.2 shows

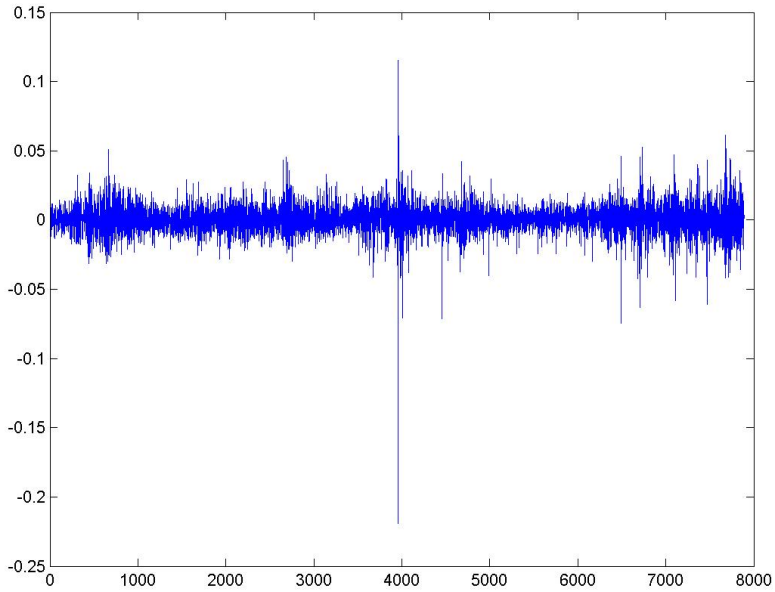


Figure 4.1: Dow Jones Index - Continuously compounded returns 1972 - 2003

a level-12 MRA, based on a Symlet6 wavelet, for illustration purposes. This wavelet is chosen for its smoothing characteristics. A level 12 decomposition is chosen as the series can be decomposed maximally into 12 levels ($=^2 \log(7890)$). When performing a visual inspection on this decomposition, it is immediately evident that the level-12 trend is nearly flat, implying that returns remained mostly constant over the time-period. The outlier in the middle of the series, is clearly visible in the original signal, but dampened in each of the detail coefficients. However, tracks of this outlier are visible in the level-5 and level-6 details of the signal. In addition, the very long run differences - the level-11 and level-12 details - show the periodic properties of the signal.

For this demonstration, the wavelet variance is based on two different level-12 decompositions: a Symlet6 decomposition and a Daubechies4 decomposition. Each of these wavelet variances is calculated for level 1,2...12 using equation (4.6) for each of the two decompositions. Thus, this variance is based on the wavelet coefficients, rather than the reconstructed coefficients as used in equation (4.8). Table 4.2 shows the results for the Symlet6 decomposition and table 4.3 shows the results for the Daubechies4 decomposition.

When inspecting these tables, it becomes immediately evident that:

- The variances remain preserved and sum up to the variance of the original returns series, up to a small approximation error.
- Regardless of the wavelet chosen, equation (4.7) holds, up to a small approximation error. Although the energy remains preserved, as a result of the different wavelet type, the energy is distributed differently.

In addition to the wavelet-decomposition, an MRA can be constructed as well. Similarly, the MRA is constructed based on a Symlet6 wavelet and a Daubeches4 wavelet. The variances are estimated using equation (4.8). Table 4.4 shows the results for the MRA Symlet6 decomposition and table 4.5 shows the results for the MRA Daubechies4 decomposition.

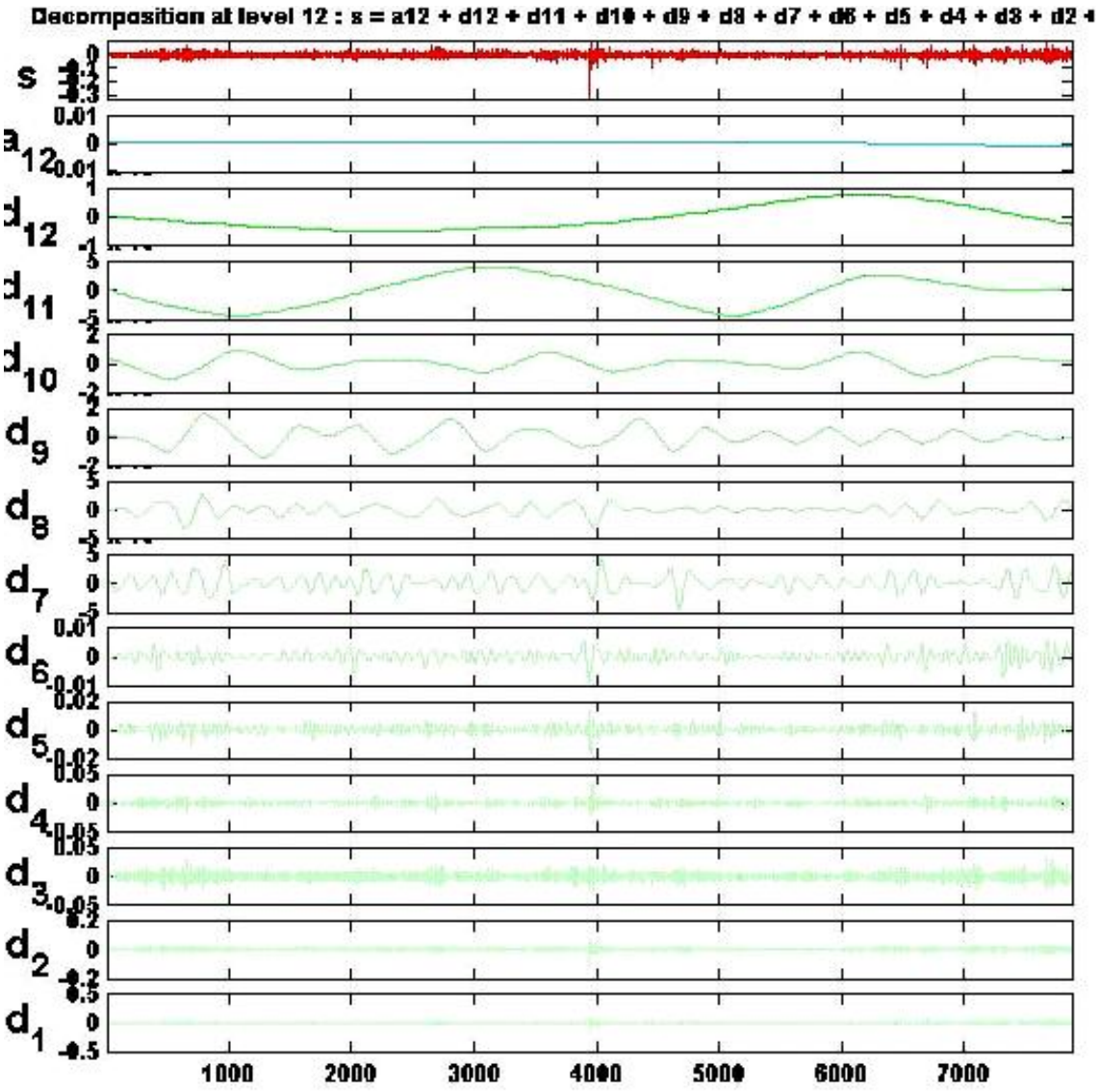


Figure 4.2: Dow Jones Index - level-12 Symlet6 MRA 1972 - 2003

level	**Variance**	**distribution**
d^1	0.000097743	47.56%
d^2	0.000053624	26.09%
d^3	0.000028985	14.10%
d^4	0.000011508	5.60%
d^5	0.000006187	3.01%
d^6	0.000002730	1.33%
d^7	0.000001202	0.58%
d^8	0.000000582	0.28%
d^9	0.000000211	0.10%
d^{10}	0.000000119	0.06%
d^{11}	0.000000041	0.02%
d^{12}	0.000000070	0.03%
s^{12}	0.000002507	1.22%
sum:	**0.000205508**	**100%**
true variance:	**0.000204252**	

Table 4.2: Variance decomposition of the Dow-Jones index using Symlet6 wavelets

level	**Variance**	**distribution**
d^1	0.000097576	47.58%
d^2	0.000053002	25.85%
d^3	0.000028629	13.96%
d^4	0.000012806	6.25%
d^5	0.000005805	2.83%
d^6	0.000003275	1.60%
d^7	0.000001364	0.67%
d^8	0.000000392	0.19%
d^9	0.000001026	0.50%
d^{10}	0.000000192	0.09%
d^{11}	0.000000135	0.07%
d^{12}	0.000000146	0.07%
s^{12}	0.000000715	0.35%
sum:	**0.000205065**	**100%**
true variance:	**0.000204252**	

Table 4.3: Variance decomposition of the Dow-Jones index using Daubechies4 wavelets

level	**Variance**	**distribution**
D^1	0.000097817	47.82%
D^2	0.000053689	26.25%
D^3	0.000028965	14.16%
D^4	0.000011695	5.72%
D^5	0.000006389	3.12%
D^6	0.000002913	1.42%
D^7	0.000001369	0.67%
D^8	0.000000741	0.36%
D^9	0.000000343	0.17%
D^{10}	0.000000190	0.09%
D^{11}	0.000000064	0.03%
D^{12}	0.000000165	0.08%
A^{12}	0.000000219	0.11%
sum:	**0.000204560**	**100%**
true variance:	**0.000204252**	

Table 4.4: MRA Variance decomposition of the Dow-Jones index using Symlet6 wavelets

level	**Variance**	**distribution**
D^1	0.000097647	47.70%
D^2	0.000053092	25.94%
D^3	0.000028515	13.93%
D^4	0.000012935	6.32%
D^5	0.000005969	2.92%
D^6	0.000003384	1.65%
D^7	0.000001429	0.70%
D^8	0.000000467	0.23%
D^9	0.000000682	0.33%
D^{10}	0.000000130	0.06%
D^{11}	0.000000123	0.06%
D^{12}	0.000000182	0.09%
A^{12}	0.000000152	0.07%
sum	**0.000204710**	**100%**
true variance:	**0.000204252**	

Table 4.5: MRA Variance decomposition of the Dow-Jones index using Daubechies4 wavelets

When inspecting these tables, it becomes immediately evident that:

- The variances remain preserved and sum up to the variance of the original returns series, up to a small approximation error.
- Regardless of the wavelet chosen, equation (4.7) holds, up to a small approximation error. Although the energy remains preserved, as a result of the different wavelet type, the energy is distributed differently.
- The magnitude and energy-distributions are nearly similar to the variance decomposition based on the wavelet coefficients.

4.4 Value at Risk estimation using the wavelet energy distribution

It is shown in section 4.2 how the energy distributes itself over the different levels of decomposition. Section 4.3 showed that in fact the variance remains preserved over the various levels. As the Discrete Wavelet Transform is an energy preserving decomposition, the energy thus remains preserved and therefore the variance remains preserved as well. The way the variance is distributed over the levels is more or less the same as the distribution of energy (i.e. the sum of squared coefficients). [1]

Now, let us formulate an alternative model - a wavelet-based model -, which uses the wavelet energy distribution for estimating the risk, as essential for VaR-estimation. This model will use the wavelet energy distribution for determining which amount of risk is essential for the horizon the VaR is based on, and which amount of risk can be neglected.

Recall from chapter 3 that the Discrete Wavelet Transform is dyadic: it is based on powers of two. Therefore, level-1 detail using an arbitrary wavelet corresponds to a 2-day horizon, whereas level-2 and level-3 details correspond to 4 and 8 day horizons respectively, if all data-elements are days of course.

Then, if the trading-horizon is h days, the maximum level j to base our risk on is a level j such that $(j \leq^2 \log h)$, *under the assumption that the financial risk manager is only interested in the risk associated with his trading horizon he bases his VaR estimation on*.

Next, recall from section 4.2 that

$$\| \mathbf{f} \|^2 = \sum_{j=1}^{J} \| \mathbf{D^j} \|^2 + \| \mathbf{A^J} \|^2$$

holds.

Then, we can express the relative amount of energy associated with detail level j as:

$$\varepsilon_{rel,j} = \frac{\| \mathbf{D^j} \|^2}{\| \mathbf{f} \|^2} \tag{4.9}$$

And, if we assign the value of the level j such that $(j \leq^2 \log h)$ to m, where m is the maximum depth, the equation which summarizes the relative amount of energy over the relevant horizon is:

[1] Here I say more or less, as depending on the wavelet type, a higher variance can be kept in the level-J approximation (or level-J set of scaling coefficients) yielding automatically less variance in levels $j = 1...J-1$ for the detail (or sets of wavelet coefficients).

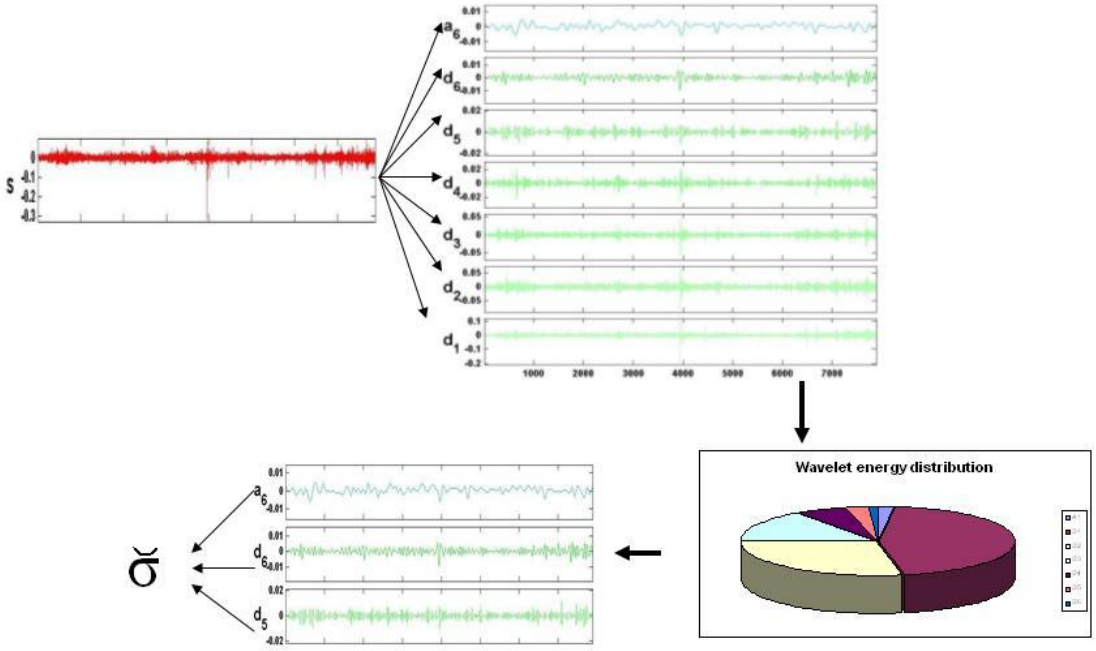


Figure 4.3: From signal through the wavelet-energy distribution to risk

$$\varepsilon_{rel} = \sum_{j=1}^{m} \alpha_j \frac{\| \mathbf{D^j} \|^2}{\| \mathbf{f} \|^2} \tag{4.10}$$

The weight, α_j can be used to assign different weights to different levels. For the purpose of this thesis, α_j is set to 1. When increasing the weight for a specific level, the variance increases, resulting in a higher estimated VaR figure. A higher estimated VaR figure results less often in exceeding this figure. The following chapter explains the relevance of the number of times this VaR figure is exceeded, as this number should lie in a certain interval. Fine-tuning the weight assigned to a level can therefore help to keep the number of times the VaR is exceeded, in a certain interval. However, more on these certain intervals in chapter 6.

Then, the total relative energy ε_{rel} can be applied to the original variance, σ, to obtain a moderated $\breve{\sigma}$:

$$\breve{\sigma} = \varepsilon_{rel}\sigma \tag{4.11}$$

Then, the final equation for estimating the VaR, given the α corresponding to the confidence level, the initial portfolio value W_0 and horizon h equals:

$$\text{VaR} = W_0 \, \alpha\breve{\sigma}\sqrt{(h)} \tag{4.12}$$

This equation is directly related to equation (2.8). The idea of all this is shown in figure 4.3.

The following experiment will apply this idea on a position worth €10000000 in assets based on the Dow-Jones index. The daily volatility at a given moment is 2% and a financial risk manager needs to estimate a 10-day VaR, given a 95% confidence level. For any given returns series, the original series

Horizon	Percentage
2 days	43.67%
4 days	24.24%
8 days	13.25%
sum	**81.16%**

Table 4.6: Relative energy

shows a daily projection of the returns process. A level-1 detail using an arbitrary wavelet, corresponds to a 2-day horizon, whereas level-2 and level-3 details correspond to 4 and 8 day horizons respectively. One might imagine that only these horizons are of use for the financial risk manager, who is only interested for these horizons, and may omit the 16, 32 and higher day horizons.

Now, assume that the distribution of variance is as indicated in table 4.6 for the 2, 4 and 8 day horizons. Equal weight is assigned to each of the trading horizons below.

The total amount of variance equals 81.16% in this case. Applied to the daily volatility, this equals 1.62%, rather than the original 2%. In this case, VaR is estimated as: 10.000.000 x 0,0162 x 2,33 x $\sqrt{10}$ = 1193633.33, whereas this figure was normally estimated as 1473621,39.

Chapter 5

Empirical analysis

5.1 Introduction

Chapter 3 explained the fundamentals of Value at Risk estimation and the various volatility models available which are the basis for Value at Risk (VaR) estimation. Chapter 4 explained the fundamentals of wavelet analysis and the basis for time-scale decomposition. Chapter 5 elaborated on the idea of time-scale decomposition and introduced an alternative model for VaR estimation based on wavelets.

In this chapter, several experiments are performed in order to compare the various models for VaR estimation. This chapter begins with an overview of the hard- and software used in the experiments. In addition, a detailed description of the dataset is given. Then, experiments and their results of VaR estimation based on a standard variance are explained. Next, experiments and their results of VaR estimation based on GARCH and EWMA models are reported. Finally, experiments and their results of Value at Risk estimation based on wavelets are mentioned.

5.2 Hardware and software

The hardware used for the experiments is a personal computer with the following specifications:

- *AMD Athlon 2200 Processor,*
- *512 MB memory,*
- *80GB HD,*
- *Asus K7V333 Motherboard (KT333 chipset),*
- *Asus Geforce4 Ti4400 128mb graphics card* .

The software used for the experiments is the following:

- *Microsoft Windows XP operating system,*
- *Microsoft Excel 2003,*

- *Matlab 6.5 Release 13,*
- *Wavelet toolbox 2.2,*
- *Optimization Toolbox 2.2,*
- *Statistics toolbox 4.0,*
- *Signal Processing Toolbox 6.1 .*

The Wavelet Toolbox is a comprehensive tool that supports exploration of wavelet properties and applications. The Wavelet Toolbox is useful in 1-D and 2-D applications such as speech and audio processing, image and video processing, communications, geophysics, finance, and medicine. The toolbox provides a graphical user interface and a broad collection of Matlab routines (Misiti et al. 2002).

The GARCH Toolbox enables the user to model volatility of financial and economic time series, and provides a framework for the analysis of serial dependence in conditional variances. The toolbox offers several variants of GARCH models, including standard ARCH/GARCH models, as well as asymmetric EGARCH and Glosten, Jagannathan and Runkle (GJR)GARCH models designed to capture leverage effects in asset returns. The other toolboxes are required in order to support the Wavelet- and GARCH toolbox.

5.3 Description of datasets

Two datasets were used for the empirical analysis: an artificial dataset and a real-life dataset. While using an artificial dataset, dataset-specific features can be better isolated. Each dataset consists of continuously compounded returns and a corresponding price-series, either artificially generated or obtained from real-life data.

The artificial dataset contains daily returns which are artificially generated by a GARCH(1,1) process. The idea of generating artificial data by using a GARCH(1,1) process is derived from van den Berg, Kaymak and van den Bergh (2003). The process of obtaining the artificial returns consists of 4 steps:

- Each return is drawn from a normal distribution with mean $\mu = 0$ and a standard deviation equal to the estimated local volatility - which is the square-root of the variance rate - at day n, which is $\sigma_n : u_n \sim \mathcal{N}(\mu, \sigma_n)$.
- Each day, the estimate of the variance rate is updated according to equation (2.18), which is repeated here:
$$\sigma_n^2 = \alpha u_{n-1}^2 + \beta \sigma_{n-1}^2 + \gamma V \text{ given that } \alpha + \beta + \gamma = 1$$
- The parameter values for the GARCH(1,1) process are also derived from van den Berg et al. (2003) and are in line with the parameters empirically found in stock return series, such that: $\alpha = 0.2, \beta = 0.78$, $\gamma = 0.02$ and $V = 0.03$.
- The price-series itself is calculated using the returns obtained from the GARCH(1,1) process, with an initial price $S_0 = 100$:
$$S_n = S_{n-1} e^{u_n} \tag{5.1}$$

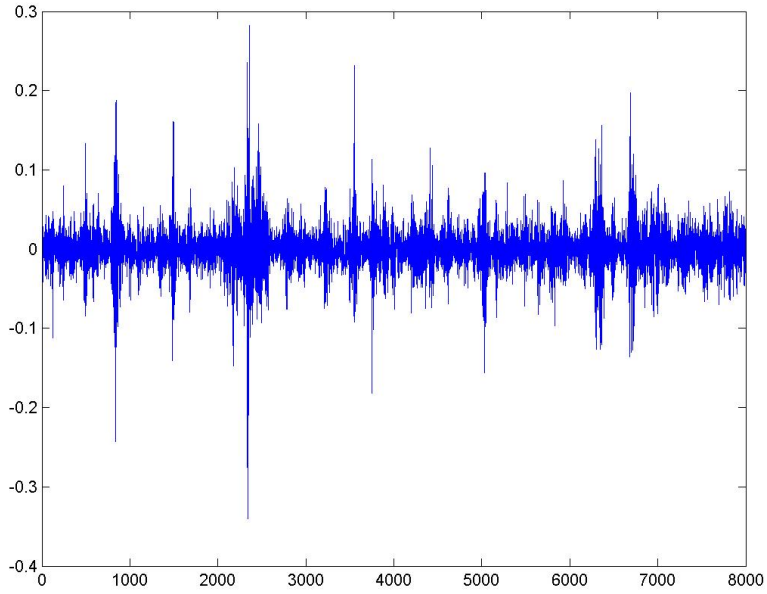


Figure 5.1: Artificially generated returns - 8000 observations

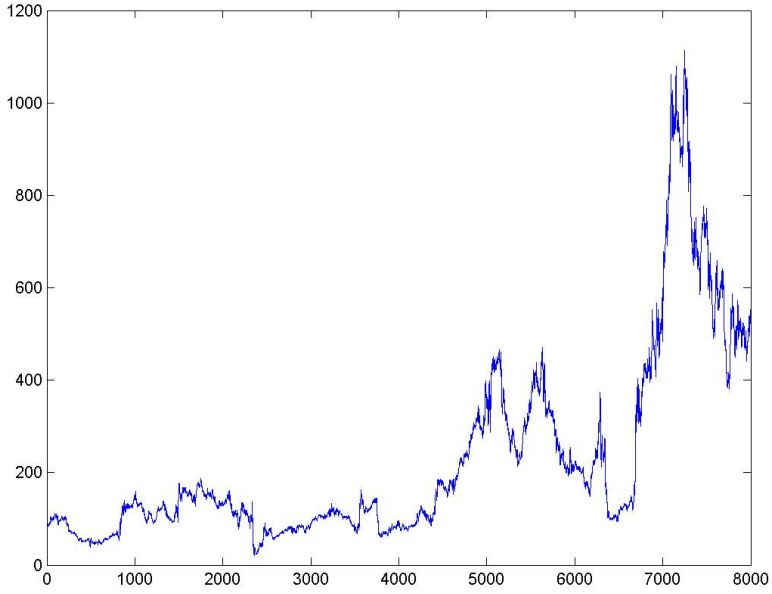


Figure 5.2: Price-series corresponding to the artificially generated returns - 8000 observations

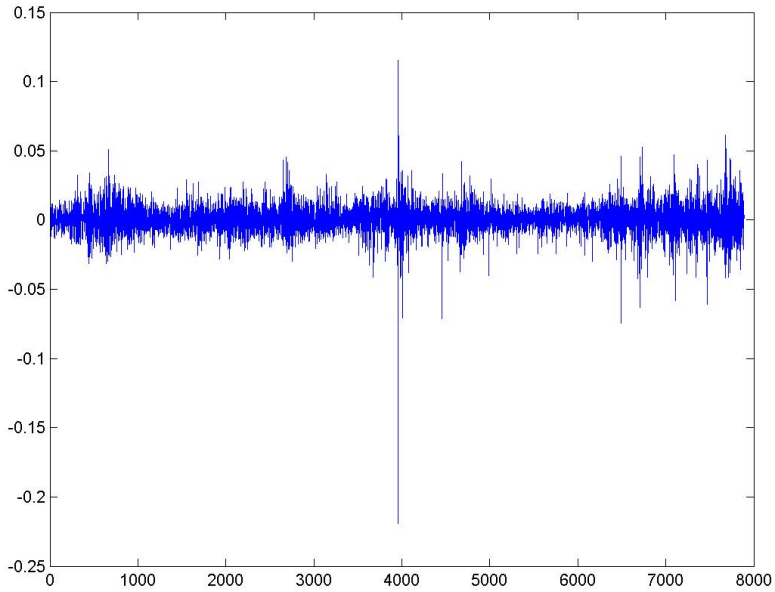


Figure 5.3: Dow Jones Index - Continuously compounded returns 1972 - 2003

The artificially generated returns are shown in figure 5.1. The corresponding price-series is shown in figure 5.2.

The real-life dataset is the dataset called *returns*, and contains real-life continuously compounded returns obtained from the Dow-Jones Index and the series of the Dow Jones Index itself. The index starts at May 5th 1972 and ends February 5th 2003. The 7889 elements of the returns series are obtained using equation (2.1).

Both the Dow Jones Index series and the Dow Jones Index returns series are shown in figures 5.4 and 5.3 respectively.

5.4 Value at Risk estimation based on a standard constant variance

Each experiment for VaR estimation has to give answer to several questions. For this research, one question stands central: *What is the performance on future returns?* When the VaR estimation is good, the significance level should more or less correspond to the number of *exceptions*, i.e. the number of times an observation of a *real* return exceeds the previously estimated VaR estimation. For example, for a 5 percent significance level, 1 exception should occur every 20 observations. If the number of exceptions is too low, the model is too conservative. If the number of exceptions is too high, the model is to optimistic. In both cases, the model needs to be adjusted.

This experiment, where the VaR-figure is estimated using a standard, constant variance, is conducted in order to compare the results of the other approaches. In this way, this simple, though widely used approach can be compared to the more sophisticated GARCH/EWMA approach. In addition, this approach can be compared to the wavelet-based approach, which is similar in the sense that it assumes the variance to be constant over time.

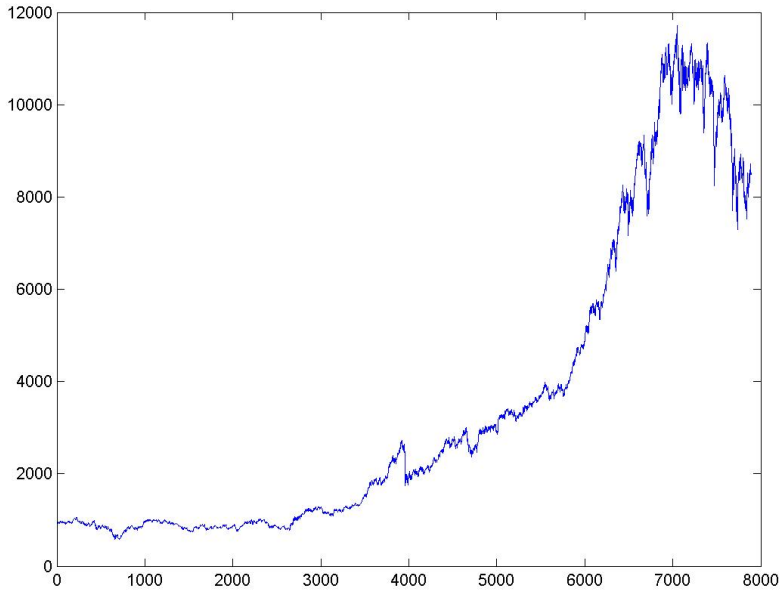


Figure 5.4: Dow Jones Index - 1972 - 2003

5.4.1 Experiment Design

The process of generating results with a constant variance based VaR estimation can be explained as follows. The experiment is conducted first on the artificial dataset and then repeated on the real-life dataset:

- First, the dataset is split up in a trainingset and a testset. In the case of constant variance based VaR estimation, the trainingset is used completely for estimating the variance. No sliding windows is used for the estimation of the variance The testset will be used solely for testing the estimated VaR estimations on future returns.
- The variance is calculated over the trainingset using equation (2.15).
- The scaling to a h-day horizon will be performed with the $\sqrt{h}$-rule. In this specific case, this rule is appropriate, as the variance is assumed to be constant over the horizon.
- When the correct scaling is done, VaR estimates can be compared with future returns. If the horizon is *h* days, then for every *h* days, a new VaR is estimated. If the real return over the coming period of *h* days ”exceeds” the estimated VaR *h* days ago, an exception is found.
- Finally, the number of exceptions is expressed as a proportion of the total number of observations, which is called the *failure rate*. The notion of failure rate is discussed in more detail in chapter 6.

For this experiment, the following assumptions are made:

- Returns over a *h* -day-period - for *h* -day VaR estimates - are only measured at the beginning and the end of this period. No attention is paid to individual trading days.

Dataset:	Artificial dataset				Real-life dataset			
confidence level	5 days	10 days	15 days	20 days	5 days	10 days	15 days	20 days
95%	0.0278	0.0275	0.0317	0.0267	0.0420	0.0345	0.0385	0.0364
97.5%	0.0165	0.0191	0.0197	0.0141	0.0268	0.0204	0.0235	0.0238
99%	0.0118	0.0119	0.0109	0.0113	0.0147	0.0119	0.0121	0.0137

Table 5.1: VaR estimation based on a standard, constant variance - Failure rates

- The portfolio is frozen over the *h* -day horizon. No additional trades are made in the period, nor are their additional fees earned. The return on such a portfolio is called a *hypothetical return*, which is used in the experiments.

5.4.2 Results

As already mentioned, the VaR estimation is based on a standard, constant variance. In order to gain insight in the VaR forecasting ability of the model, the following experiments are done:

first:

- VaR estimation given a 5, 10, 15 and 20 day horizon;
- A 95, 97,5 and 99 percent confidence level for each horizon
- A ”position” on the index of the artificial dataset worth €10000000 [1]
- A split on day 4000, resulting in a testSet of 4000 days.

then:

- VaR estimation given a 5, 10, 15 and 20 day horizon;
- A 95, 97,5 and 99 percent confidence level for each horizon
- A position on the Dow Jones Index worth €10000000
- A split on day 4000, resulting in a testSet of 3890 days.

Table 5.1 shows the results.

When comparing the results of the experiments on the artificial dataset and the real-life dataset, it is apparent that the results for the artificial dataset show a systematical overestimation of the VaR, resulting in too few exceptions, especially on the 95% and 97.5% confidence levels. However, the model in its simplicity performs quite accurate on the 99% confidence level on both datasets and also accurate on the 97.5% confidence level on the real-life dataset. However, for the 95%confidence level, it might be so that not enough exceptions are found for the model to be accepted, when the model is validated in chapter 6.

[1]The financial position is used in the experiments in order to give it a ”real-life” sense. It is not necessary for obtaining results.

5.5 Value at Risk estimation based on GARCH- and EWMA models

The following experiments also have to give answer to the question *What is the performance on future returns?* with respect to the models for VaR-estimation based on a GARCH and EWMA approach. The experiments are twofold. Firstly, experiments are conducted using scaling according to Drost and Nijman (1993). Secondly, experiments are conducted using the traditional $\sqrt{h}$ scaling. These experiments should bring light in the discussion whether $\sqrt{h}$ scaling is truly so inappropriate. Both types of experiments are conducted on the artificial dataset and on the real-life dataset.

5.5.1 Experiment design

The steps which have to be taken in order to calculate the number of exceptions are the following. The experiment is conducted first on the artificial dataset and then repeated on the real-life dataset:

- First, the dataset is split up in a trainingset and a testset. The parameters α, β and ω need to be estimated conform the maximum likelihood equations (2.20) using the corresponding fitting routines in the GARCH toolbox, over the trainingSet. The fitting routine in the GARCH toolbox generates a GARCH(1,1)-model. The model consists of the parameters α, β and ω. The routine also generates the complete set of daily volatilities generated by this GARCH(1,1) model.

- For the EWMA model, the parameter estimation process is skipped; the important parameter λ is set to 0.94, which is the JPMorgan's riskmetrics standard.

- Next, the parameters are scaled to the appropriate horizon using equation (2.22), as discussed in section 2.5. When one is interested in e.g. a 10-day horizon, the daily volatility has to be scaled to a 10-day volatility.

- As an EWMA model is some sort of a GARCH model, this form of scaling is appropriate as well for EWMA models. When the correct scaling is done, VaR estimates can be compared with future returns. If the horizon is *h* days, then for every *h* days, a new VaR is estimated. If the real return over the coming period of *h* days "exceeds" the estimated VaR *h* days ago, an exception is found.

  In addition, scaling using the $\sqrt{h}$ is tried out as well, in order to compare its working with the Drost-Nijman equations for scaling.

- Finally, the number of exceptions are expressed as a proportion of the total number of observations, which is called the *failure rate*.

For this experiment, the following assumptions are made:

- Returns over an *h* -day-period - for *h* -day VaR estimates - are only measured at the beginning and the end of this period. No attention is paid to individual trading days.

- The portfolio - a stock position on the Dow Jones Index in this case - is frozen over the *h* -day horizon. No additional trades are made in the period, nor are their additional fees earned. The return on such a portfolio is called a *hypothetical return*, which is used in the experiments.

α	0.0679
β	0.9201
ω	1.32775168084179e-006

Table 5.2: Parameters of the GARCH(1,1) model

$1 - \lambda(= \alpha)$	0.06
$\lambda(= \beta)$	0.94

Table 5.3: Parameters for the EWMA model

5.5.2 Results

After constructing the GARCH(1,1) model for the real-life dataset, the parameters are estimated as shown in table 5.2. For the artificial dataset, the parameters are equal to the parameters as proposed in section 5.3 For the EWMA model, the parameters are shown in table 5.3 and are equal for the artificial and real-life dataset.

In order to gain insight in the VaR forecasting ability of the models, the following experiments are conducted:

first:

- VaR estimation given a 5, 10, 15 and 20 day horizon;
- VaR estimation for both models with $\sqrt{h}$ scaling and Drost-Nijman scaling.
- A 95, 97,5 and 99 percent confidence level for each horizon
- A ”position” on the index of the artificial dataset worth €10000000
- A split on day 4000, resulting in a testSet of 4000 days.

then:

- VaR estimation given a 5, 10, 15 and 20 day horizon;
- VaR estimation for both models with $\sqrt{h}$ scaling and Drost-Nijman scaling.
- A 95, 97,5 and 99 percent confidence level for each horizon
- A position on the Dow Jones Index worth €10000000
- A split on day 4000, resulting in a testSet of 3890 days.

Table 5.4 shows the results for VaR estimation based on a GARCH(1,1) model with Drost-Nijman scaling and table 5.6 shows the results for VaR estimation based on a GARCH(1,1) model with $\sqrt{h}$ scaling.

One can see clearly that the results of the experiments based on $\sqrt{h}$ scaling are far more better than the results of the experiments based on Drost-Nijman scaling. The results of the experiments based on $\sqrt{h}$

Dataset:	Artificial dataset				Real-life dataset			
confidence level	5 days	10 days	15 days	20 days	5 days	10 days	15 days	20 days
95%	0.1455	0.1367	0.1174	0.1031	0.1632	0.1547	0.1430	0.1287
97.5%	0.1061	0.1009	0.0807	0.0625	0.1259	0.1194	0.1185	0.1031
99%	0.0735	0.0667	0.0530	0.0430	0.0971	0.0923	0.0891	0.0770

Table 5.4: VaR estimation based on a GARCH(1,1) model with Drost-Nijman scaling - Failure rates

Dataset:	Artificial dataset				Real-life dataset			
confidence level	5 days	10 days	15 days	20 days	5 days	10 days	15 days	20 days
95%	0.2119	0.27342	0.2955	0.3378	0.2044	0.2361	0.2499	0.2598
97.5%	0.1765	0.2474	0.2675	0.3079	0.1730	0.2098	0.2297	0.2396
99%	0.1540	0.2129	0.2434	0.2818	0.1393	0.1846	0.2086	0.2205

Table 5.5: VaR estimation based on a EWMA model with Drost-Nijman scaling - Failure rates

Dataset:	Artificial dataset				Real-life dataset			
confidence level	5 days	10 days	15 days	20 days	5 days	10 days	15 days	20 days
95%	0.0450	0.0529	0.0489	0.0550	0.0424	0.0445	0.0428	0.0457
97.5%	0.0266	0.0275	0.0282	0.0295	0.0296	0.0270	0.0252	0.0258
99%	0.0128	0.0131	0.0131	0.0141	0.0175	0.0159	0.0160	0.0129

Table 5.6: VaR estimation based on a GARCH(1,1) model with $\sqrt{h}$ scaling - Failure rates

Dataset:	Artificial dataset				Real-life dataset			
confidence level	5 days	10 days	15 days	20 days	5 days	10 days	15 days	20 days
95%	0.0770	0.0777	0.0791	0.0786	0.0661	0.0541	0.0588	0.0566
97.5%	0.0529	0.0614	0.0584	0.0553	0.0450	0.0381	0.0400	0.0361
99%	0.0353	0.0401	0.0405	0.0364	0.0275	0.0216	0.0240	0.0232

Table 5.7: VaR estimation based on a EWMA model with $\sqrt{h}$ scaling - Failure rates

scaling are quite accurate as well: on a 95% confidence interval, 5% exceptions are expected. We can see that this is nearly the case.

When we compare the results of the experiments on the artificially generated data and the real-life data, both results are nearly the same. With respect to the 95% confidence level however, the results of the experiments on the artificially generated data converge slightly more to the desired failure rate.

Table 5.5 shows the results for VaR estimation based on an EWMA model with Drost-Nijman scaling and table 5.7 shows the results for VaR estimation based on an EWMA model with $\sqrt{h}$ scaling.

The same conclusion can be drawn for the EWMA model: the results of the experiments based on $\sqrt{h}$ scaling are far more better than the results of the experiments based on Drost-Nijman scaling.

When we compare the results of the experiments on the artificially generated data and the real-life data, the results of the experiments on the real-life data are better.

However, the GARCH(1,1) model performs better on both types of experiments than the EMWA model. A first possible explanation could be the fact that the value of 0.94 assigned to λ is chosen, whereas the parameters for the GARCH model are estimated using a sophisticated fitting routine. Choosing a different value for λ can yield better results.

5.6 Value at Risk estimation based on wavelets

Similarly to the experiments for GARCH/EWMA-based VaR estimation, the wavelet-based VaR estimation has to give answer to the question: *How accurate is this estimation when compared to future returns?* In addition, the experiments should make clear *if* a different choice of a wavelet has impact on the results and *if* a different wavelet has impact, which wavelet produces the best results. These experiments are twofold. Firstly, experiments are conducted where the amount of energy as proposed in equation (4.12). Secondly, experiments are conducted where the amount of energy includes one extra level to the basic amount as proposed in equation (4.12). This should bring insight in the effect of including one extra level of information on the failure rate. Both types of experiments are conducted on the artificial dataset and the real-life dataset.

5.6.1 Experiment design

The process of generating results with wavelet-based VaR estimation - which is different in a significant number of steps - can be explained as follows. First, the experiments are performed on the artificial dataset. Then, the experiments are repeated on the real-life dataset.

- First, the dataset is split up in a trainingset and a testset. In the case of wavelet-based VaR estimation, the trainingset is used completely for estimating the wavelet coefficients and the corresponding energy distribution. The testset is used solely for testing the estimated VaR estimations on future returns.
- Next, the maximum decomposition level is calculated, as explained previously in chapter 3.
- After calculating the maximum decomposition level, the returns series is decomposed in the corresponding number of levels. This decomposition is based on this maximum decomposition level and the type of wavelet chosen.

- Next, the energy distribution is extracted from the wavelet decomposition. As explained in chapter 4, this energy distribution corresponds to the distribution of the variance of the returns series over the different levels after decomposing.
- Then, given the time-horizon - e.g. 5, 10, 15 or 20 days - the maximum amount of levels are calculated. If for instance the time-horizon is 20 days, then the first power of 2 smaller than 20 is used, which is $2^4 = 16 < 20$. In this case, only the first 4 levels are involved in the estimation of the risk. In addition, the experiment is extended with one extra level, in order to compare the effects of adding this extra level.
- The variance is calculated over the trainingset and the energy distribution as found in the previous step is applied.
- The scaling to a h-day horizon is performed with the $\sqrt{h}$-rule. In this specific case, this rule is appropriate, as the variance is assumed to be constant over the horizon.
- When the correct scaling is done, VaR estimates can be compared with future returns. If the horizon is *h* days, then for every *h* days, a new VaR is estimated. If the real return over the coming period of *h* days "exceeds" the estimated VaR *h* days ago, an exception is found.
- Finally, the number of exceptions is expressed as a proportion of the total number of observations, which is called the *failure rate*.

For the experiments a Haar-wavelet, a Daubechies4 wavelet and a Symlet6 are used. The Haar wavelet is the most simplest wavelet. The Daubechies4 wavelet is already smoother than the Haar-wavelet, but not symmetric. The Symlet6 is smoother than both the Haar wavelet and the Daubechies4 wavelet, and it is symmetric.

For this experiment, the following assumptions are made:

- Returns over a *h* -day-period - for *h* -day VaR estimates - are only measured at the beginning and the end of this period. No attention is paid to individual trading days.
- The portfolio - a stock position on the Dow Jones Index in this case - is frozen over the *h* -day horizon. No additional trades are made in the period, nor are their additional fees earned. The return on such a portfolio is called a *hypothetical return*, which is used in the experiments.

5.6.2 Results

As already mentioned, the VaR estimation is based on three different types of wavelets: the Haar wavelet, the Daubechies4 wavelet and the Symlet6. In order to gain insight in the VaR forecasting ability of the models, the following experiments are done:

first:

- VaR estimation based on the three types of wavelet models given a 5, 10, 15 and 20 day horizon;
- A 95, 97,5 and 99 percent confidence level for each horizon
- A "position" on the index of the artificial dataset worth €10000000
- A split on day 4000, resulting in a testSet of 4000 days.

Dataset:	Artificial dataset				Real-life dataset			
confidence level	5 days	10 days	15 days	20 days	5 days	10 days	15 days	20 days
95%	0.0493	0.0310	0.0331	0.0251	0.0904	0.0493	0.0553	0.0462
97.5%	0.0317	0.0215	0.0218	0.0138	0.0645	0.0305	0.0387	0.0299
99%	0.0192	0.0132	0.0117	0.0100	0.0431	0.0207	0.0224	0.0202

Table 5.8: VaR estimation based on wavelets (Haar) without extra level of information - Failure rates

Dataset:	Artificial dataset				Real-life dataset			
confidence level	5 days	10 days	15 days	20 days	5 days	10 days	15 days	20 days
95%	0.0926	0.0691	0.0597	0.0549	0.0497	0.0532	0.0576	0.0492
97.5%	0.0643	0.0426	0.0419	0.0341	0.0632	0.0350	0.0422	0.0329
99%	0.0415	0.0270	0.0296	0.0247	0.0213	0.0224	0.0256	0.0215

Table 5.9: VaR estimation based on wavelets (Daubechies4) without extra level of information - Failure rates

- The experiments are performed using the wavelet-based model without and with the extra level of information, as proposed in the previous subsection.

then:

- VaR estimation based on the three types of wavelet models given a 5, 10, 15 and 20 day horizon.
- A 95, 97,5 and 99 percent confidence level for each horizon.
- A position on the Dow Jones Index worth €10000000.
- A split on day 4000, resulting in a testset of 3890 days.
- The experiments are performed using the wavelet-based model without and with the extra level of information, as proposed in the previous subsection.

Table 5.8 shows the results for the experiments with the Haar wavelet, table 5.9 shows the results for the experiments with the Daubechies4 wavelet and table 5.10 shows the results for the experiments with the Symlet6 wavelet. These results are *without* the extra level of information.

Table 5.11 shows the results for the experiments with the Haar wavelet, table 5.12 shows the results for the experiments with the Daubechies4 wavelet and table 5.13 shows the results for the experiments with the Symlet6 wavelet. These results are *with* the extra level of information.

Dataset:	Artificial dataset				Real-life dataset			
confidence level	5 days	10 days	15 days	20 days	5 days	10 days	15 days	20 days
95%	0.0718	0.0478	0.0464	0.0366	0.0865	0.0513	0.0556	0.0475
97.5%	0.0480	0.0293	0.0316	0.0236	0.0612	0.0318	0.0387	0.0309
99%	0.0307	0.0210	0.0215	0.0138	0.0418	0.0211	0.0240	0.0202

Table 5.10: VaR estimation based on wavelets (Symlet6) without extra level of information - Failure rates

Dataset:	Artificial dataset				Real-life dataset			
confidence level	5 days	10 days	15 days	20 days	5 days	10 days	15 days	20 days
95%	0.0330	0.0270	0.0296	0.0211	0.0564	0.0422	0.0478	0.0420
97.5%	0.0232	0.0185	0.0178	0.0113	0.0398	0.0250	0.0315	0.0283
99%	0.0137	0.0110	0.0102	0.0087	0.0233	0.0181	0.0175	0.0185

Table 5.11: VaR estimation based on wavelets (Haar) with extra level of information - Failure rates

Dataset:	Artificial dataset				Real-life dataset			
confidence level	5 days	10 days	15 days	20 days	5 days	10 days	15 days	20 days
95%	0.0706	0.0594	0.0519	0.0457	0.0622	0.0454	0.0494	0.0452
97.5%	0.0478	0.0371	0.0368	0.0299	0.0421	0.0276	0.0344	0.0296
99%	0.0305	0.0240	0.0261	0.0168	0.0272	0.0194	0.0185	0.0198

Table 5.12: VaR estimation based on wavelets (Daubechies4) with extra level of information - Failure rates

Dataset:	Artificial dataset				Real-life dataset			
confidence level	5 days	10 days	15 days	20 days	5 days	10 days	15 days	20 days
95%	0.0533	0.0403	0.0401	0.0339	0.0573	0.0425	0.0484	0.0433
97.5%	0.0325	0.0258	0.0288	0.0211	0.0402	0.0256	0.0325	0.0283
99%	0.0227	0.0175	0.0173	0.0113	0.0243	0.0188	0.0178	0.0192

Table 5.13: VaR estimation based on wavelets (Symlet6) with extra level of information - Failure rates

When comparing the results of the experiments based on the two datasets, we can see that the results for the artificial dataset have a lower magnitude for nearly all Haar- and Symlet6 cases. However, this is rather the contrary for the magnitudes of the results of the Daubechies4 wavelet. The Haar-based model performs quite accurate on the 99% level, but is rather conservative on the 95% level.

Purely based on visual inspection, the results might suggest that the variant with the extra amount of information performs better: the magnitudes of the results are more in line with the confidence level. With this extra level, an additional amount of variance is included, resulting in a higher VaR figure, which causes the number of exceptions to fall. The variant with the extra amount of information performs quite accurate. In case of the Haar wavelet, the 10 days VaR with 97.5% confidence has exactly 2.5% exceptions. Unfortunately, for a 5 day horizon, the VaR is underestimated, resulting in too much exceptions. Purely based on visual inspection, the results might suggest that the Haar- and Symlet6 wavelet perform better than the Daubechies4 wavelet. However, no statements about true performance can be made without the use of statistical tests, which is discussed in chapter 6

5.7 Discussion

When looking at the results of the GARCH and EWMA experiments, it becomes clear that the proportion of failures is not in line with the confidence level when performed with Drost-Nijman scaling: the failure-rate is too high. This high failure rate is a direct result of systematically underestimating the future losses; the VaR-figure is too optimistic in this case. However, when inspecting the failure rates with respect to the GARCH-based VaR-model, a positive trend is visible.

A possible explanation for systematically underestimating future losses can be linked to the Drost-Nijman equation (2.22). When looking closer to the formula, the scaled volatility is lower than the original volatility in most cases: risk scaled to a h-day horizon remains in a certain ”area” and has less freedom to take a variety of values. However, intuitively the risk over 10 days should be larger than the risk over 1 day. When the risk is estimated lower, the VaR is estimated too low, resulting in many exceptions.

When inspecting the results of the GARCH and EWMA experiments with $\sqrt{h}$ scaling, the results are far more accurate: they are far more in line with the confidence level. This is an unexpected result, given that the theory by Drost and Nijman (1993) states that scaling by $\sqrt{h}$ is explicitly inappropriate. The theory has a strong conflict with the results. In addition can be seen that the GARCH model however, performs better than the EWMA-model.

The same holds for the wavelet-based model and the model with a constant variance. Both types of models have results which are nearly in line with the confidence level. With respect to the wavelet-based model, the results of the Symlet6 experiment are slightly more accurate than the two other types. The added-value of the wavelet-based model lies in the fact that the wavelet-based model is more optimistic than the model with constant variance, resulting in a VaR which is lower, as less risk is used. The model with the standard, constant variance assigns conservatively the full amount of risk, resulting in too few exceptions. Although these experiments give some intuitive insight in the behavior of the models, no real conclusions can be drawn as long as the models and their results are not validated yet. This is done in chapter 6.

Chapter 6

Model validation

6.1 Introduction

Although experimenting is an important step in research, concluding statements can only be made after a thorough process of validation. The goal of this chapter is therefore to check whether the results as gathered in chapter 5 are valid. More specific: chapter 5 conducted experiments, where every experiment yielded a number of failure rates, which were shown in the various tables. The failure rates express the number of failures as a proportion to the number of observations; they express the number of times the VaR figure is exceeded on the total of observations. Sometimes a failure rate is 4.3%, for example, while it is 5.3% the other time, given that the confidence level is 95%. Neither of these failure rates are exactly 5% and neither of these results are bad results. Rejecting or accepting a model based on just inspecting these results is not a justified action. Therefore, validation procedures are necessary to make scientifically justified statements about these results.

The remainder of this chapter is structured as follows: Section 6.2 discusses a general framework for the validation of VaR-models: backtesting. It discusses why it is important, how it works and which models are available for making those scientifically justified statements. After reading this section, it should be clear when to reject and when to accept a VaR-model. Finally, in section 6.3, the same VaR-models which are used in 5 are validated based on their results, using one of the models as discussed in section 6.2.

6.2 Validating Value at Risk models - backtesting

VaR models are useful in one way only: if they predict the risk well. Model validation is the process of checking whether a model performs adequately, which can be done by a set of tools. One of these tools is backtesting. Backtesting is a tool which verifies whether projected losses are in line with actual losses, in the form of a statistical framework (Jorion 1996). This entails systematically comparing the history of VaR forecasts with the corresponding portfolio returns. For VaR users and risk managers, these checks are essential for examining whether their model is well calibrated. If not, the model needs to be reexamined, in terms of parameters, assumptions and ways of modelling. The theory and corresponding models regarding backtesting, are derived from Jorion (1996) and Hull (2000).

Chapter 2 mentioned that under supervision of the Basel Committee, large banks have to base their capital requirements for market risk exposure explicitly on VaR models with 99% significance and an

horizon of 10 days. In addition, backtesting plays an important role in the decision the Basel Committee made for allowing internal VaR models for capital requirements.

When the model is perfectly calibrated, the number of observations falling outside VaR should be in line with the confidence level.

Definition 3 *In this thesis, an observation is a moment where the actual return over an horizon of h days is compared with the forecasted VaR number for this same horizon.*

The number of observations exceeding the VaR is also known as the number of exceptions. With too many exceptions, the model underestimates the risk. With too few exceptions, the model is in fact conservative, leading to an inefficient allocation of capital.

As already mentioned, backtesting involves systematically comparing the history of VaR forecasts, with the actual, subsequent returns. With perfectly calibrated model, the number of exceptions should be in line with the confidence level. With a 95% confidence level, one expects 5% exceptions. However, in practice this is not the case: a greater percentage - 6%, 7% or 8% percent can also be found, as a result of bad luck. A smaller amount of course, can be found as well. At some point, there must be a decision made to accept or reject the current model. The decision can be made based on the results of a statistical test. These statistical tests are discussed in the next section.

6.2.1 Exception-based model backtesting

One way to verify the accuracy of the model, is to examine the failure rate of the model.

Definition 4 *The failure rate is the proportion of times the VaR figure is exceeded in a given sample. As at a given moment, a VaR number is specified given a certain confidence level c for a total of T observations, N is defined as the number of exceptions - i.e. the number of observations where the actual loss exceeds the VaR - and $\frac{N}{T}$ is the failure rate (Jorion 1996).*

In this situation, the user wants to know at a given confidence level, whether N is too small or too large under the null hypothesis that the probability of an exception to occur $p = P$ (for instance, 0.01, 0.02 or 0.05) in a sample size of T, without making any assumption about the returns distribution. A setup for this test is the framework called the Bernouilli trials, which is used for testing sequences of success and failure. The number of exceptions N follows a binomial probability distribution:

$$f(x) = \binom{T}{N} p^N (1-p)^{T-N} \tag{6.1}$$

with $E(N) = pT$ and $\sigma^2 = p(1-p)T$.

When T is large, the central limit theorem[1] can be used to approximate the binomial distribution by the normal distribution - as a shortcut:

$$z = \frac{N - PT}{\sqrt{p(1-p)T}} \sim \mathcal{N}(0,1) \tag{6.2}$$

This binomial distribution can be used to test whether the number of exceptions is acceptably small.

[1] the central limit theorem is also depending on the size of p. When $p(1-p)T > 9$, (6.2) is a good approximation. However, when for instance T is large but p is small (yielding a figure < 9), then the approximation can be improved by using *continuity correction*, i.e. replacing N by $N + 0,5$.

Probability level p	VaR confidence level	T = 255 days	T = 510 days	T = 1000 days
0.01	99%	$N < 7$	$1 < N < 11$	$4 < N < 17$
0.025	97.5%	$2 < N < 12$	$6 < N < 21$	$15 < N < 36$
0.05	95%	$6 < N < 21$	$16 < N < 36$	$37 < N < 65$
0.075	92.5%	$11 < N < 28$	$27 < N < 51$	$59 < N < 92$
0.10	90%	$16 < N < 36$	$38 < N < 65$	$81 < N < 120$

Table 6.1: Model backtesting - 95% confidence test confidence regions

Example 6.2.1 *Suppose we have 250 observations. Suppose we have a VaR model with a significance level of 99%. We'd like to know the probability that more than 4 exceptions occur. Then the situation is as follows:*

- $T = 250$
- $p = 0.01$
- $N = 4$

After substituting the right figures into the right equation, it tells us that more than 4 exceptions can be found in 10.8% of the time, and more than 3 exceptions 24.2% of the time.

Another model for failure-rate based backtesting is the test created by Kupiec, in 1995 (Jorion 1996). He developed 95% confidence regions for such a test. These regions are shown in table 6.1. It is important to keep in mind that the 95% confidence level for a VaR *model* being correct should not be confused with the confidence level of the VaR *figure* itself. Therefore, the table can help with making a decision whether to accept the model or not, depending on if the number of exceptions lies in a specific region, given the number of observations T and the VaR confidence level.

Each of these regions are defined by the tailpoint of the log-likelihood ratio:

$$LR_{uc} = -2\ln\left[(1-p)^{T-N}p^N\right] + 2\ln\left\{\left[1-\left(\frac{N}{T}\right)\right]^{T-N}\left(\frac{N}{T}\right)^N\right\} \tag{6.3}$$

Again, the variable N is the number of exceptions, where the variable T is the total number of observations. The variable p is the probability of an exception to occur. This ratio is asymptotically χ^2-distributed, with 1 degree of freedom, under the null hypothesis that p is the true probability. To see how this works, consider example 6.2.2.

Example 6.2.2 *Suppose we have a VaR model with a 99% percent coverage. We have observations of 2 years of data, which is 510 observations. Recall from definition 3 that an observation can be made over any horizon, not necessarily a day.*

When looking up the value not to be exceeded in a statistical table of χ^2-values, the value corresponding to a 95% confidence level and 1 degree of freedom is 3.84. When inserting $T = 510$ and $p = 0.01$ into equation (6.3), each value of N between 6 and 11 yields an outcome of LR_{uc} smaller than or equal to 3.84. Every value of N smaller than 2 or greater than 10 will yield a value of LR_{uc} larger than 3.84. And, if the value is larger than 3.84, the model will be rejected. Kupiec created the table by systematically checking the values of N yielding a value of LR_{uc} smaller than or equal to 3.84. [2]

[2]Again, Microsoft Excel provides an easy way to check the table. Easily can be seen that LR_{uc} remains under the 3.84

6.2.2 Distribution forecast-based backtesting

Exception based methods are involving only one quantile as opposed to the whole distribution. Crnkovic and Drachman (1996) state that institutions should involve the entire probability density function and should assess the quality of the forecast likewise. Their method extends the idea of the exception-based approach as follows:

- A range of probabilities p between 0 and 1 needs to be selected (for instance, 0.01, 0.02, 0.03 etc.)
- Each day, a riskmanagement system reports a set of VAR numbers for each of the different confidence levels, say VaR_1, VaR_2 etc.
- The day after, the riskmanagers records whether the observed profit or loss is below $ValueAtRisk_1$, $ValueAtRisk_2$ etc.
- At the end of the period, the total number of observations below $ValueAtRisk_n$ are compiled as N_n.

Finally, the total proportion of observations $\frac{N_i}{T} = \widehat{F}(p_i)$ for each probability level is reported. This is in fact similar to the previously mentioned exception tests, but for a range of values. With perfect calibration, the empirical distribution should match p. The following test checks whether the actual distribution of profits and losses is consistent with its predicted shape:

$$K = \max_i \left[\widehat{F}(p_i) - p_i\right] + \max_i \left[p_i - \widehat{F}(p_i)\right] \tag{6.4}$$

This is a Kuiper test statistic and has a known distribution.

6.3 Validating the various models

In this thesis, the Kupiec test as described in section 6.2 is used for the validation for the various models. Section 6.2 also showed a table Kupiec constructed for backtesting purposed.

However, as the amount of observations used in this thesis exceeds the amount used in the table, a first step is to construct a confidence region manually.

A practical way to do this is to use make a graph of equation (6.3) using various values of N while holding T and p constant and inspect the nonrejection region visually. An example of such a graph is shown in figure 6.1.

The graph shows two lines. The pink line represents the log-likelihood ratio defined by equation (6.3). The blue line represents the threshold of 3.84: the value in the χ^2-distribution table with a 95% confidence level and 1 degree of freedom. The area where the log-likelihood ratio comes under this blue line, is the non-rejection region.

With (more or less) 3000 observations and a VaR-confidence level of 95%, it gives a nonrejection region for values of N between 128 and 176. In the following subsections, all of the failures corresponding to a model and its parameters are validated using the Kupiec test. This is done by checking if the amount of failures lie in the nonrejection area, defined by equation (6.3)

when entering values for N corresponding to the region.

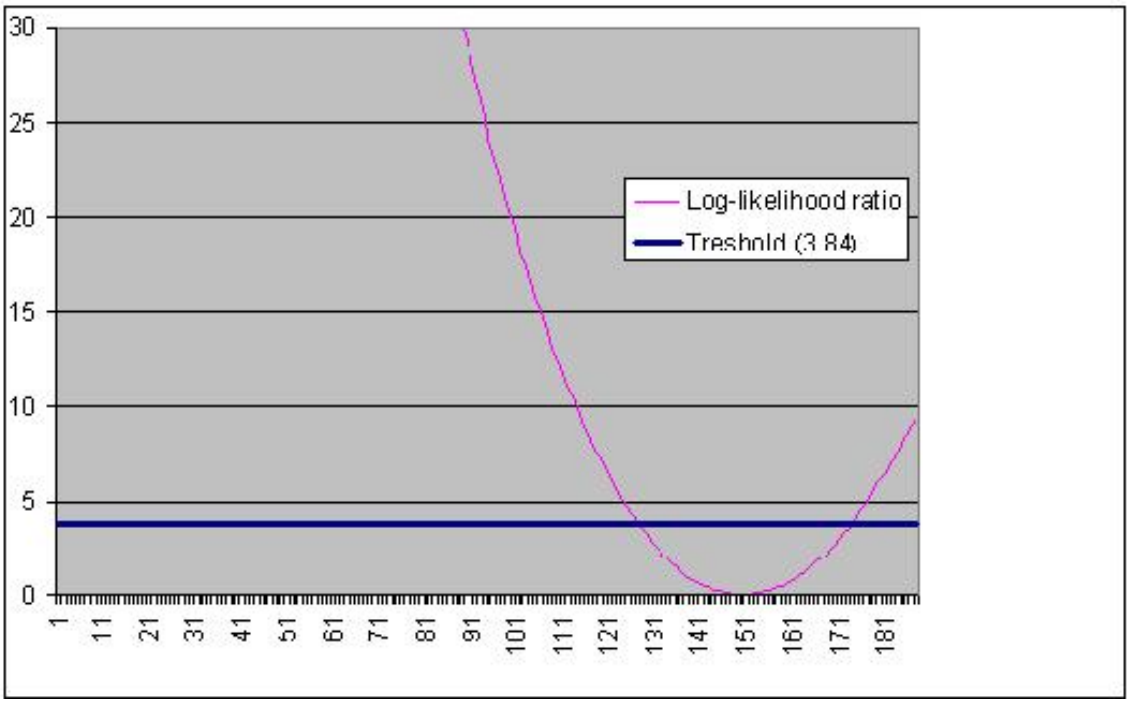


Figure 6.1: Equation (6.3) with $T = 3000$ and $p = 0.05$ for various values of N

6.3.1 Validating the various models - GARCH/EWMA

Chapter 5 reported the failure rates for the VaR models with a standard constant variance approach, an EWMA/GARCH approach, and a wavelet-based approach. The failure rates for a 5-, 10-, 15- and 20-day horizon were reported, given a 95%, a 97.5% and a 99% confidence level. The experiments were conducted on both an artificial and a real-life dataset.

Table 6.2 displays the failures for VaR confidence levels 95%, 97.5% and 99% for 10-, 15- and 20-day horizons. These failures are expressed as the number of failures N, rather than the proportion of failures on the total of observations, $\frac{N}{T}$. *These failures are obtained from the results of the real-life dataset*. For the experiments based on a GARCH/EWMA model, the results for both $\sqrt{h}$-scaling and Drost-Nijman scaling are shown. For the wavelet-based models, the results are shown for the regular model and the variant with one extra level of information included.

The first column represents the type of model and the confidence level. Columns 2-7 represent the amount of failures for a 10-, 15- and 20-days horizon, given the type of model and the confidence level. Column 8 represents the non-rejection region, or the interval the number of failures should lie in, in order to be accepted. These intervals are calculated using equation (6.3).

For 3000 observations and a VaR-confidence level of 95%, the failure rates for the GARCH-based model with Drost-Nijman scaling fluctuate between the 353 and 418. The Kupiec test accepts the model when these failures fall in the nonrejection region (128-176), implying that the tests reject the GARCH-based model with Drost-Nijman scaling in this setting.

The same holds for the EWMA model with Drost-Nijman scaling, with failures fluctuating between the 600 and 750, which are far too high. These failures are to high for accepting the EWMA-based model with Drost-Nijman scaling, which could be accepted if return failures between 128 and 176.

When inspecting the results, the GARCH-based VaR-model with $\sqrt{h}$-scaling is accepted in nearly all sit-

Model	**10 days**	**15 days**	**20 days**	**10 days**	**15 days**	**20 days**	nonrejection region
% conf.	Constant variance						
95%	107	126	118				$128 < N < 176$
97.5%	**66**	**77**	**80**				$62 < N < 95$
99%	44	**41**	47				$21 < N < 44$
% conf.	GARCH - $\sqrt{h}$-scaling			GARCH - Drost-Nijman-scaling			
95%	126	**132**	**143**	418	397	353	$128 < N < 176$
97.5%	**75**	**77**	**74**	322	322	281	$62 < N < 95$
99%	49	**43**	**37**	242	234	195	$21 < N < 44$
% conf.	EWMA - $\sqrt{h}$-scaling			EWMA - Drost-Nijman-scaling			
95%	186	203	214	733	792	792	$128 < N < 176$
97.5%	126	148	157	656	731	728	$62 < N < 95$
99%	90	95	98	578	651	669	$21 < N < 44$
% conf.	Wavelet(Haar)-without extra level			Wavelet(Haar)-with extra level			
95%	**152**	**170**	**142**	**130**	**147**	**129**	$128 < N < 176$
97.5%	**94**	119	**92**	**77**	**94**	**87**	$62 < N < 95$
99%	64	69	62	56	54	57	$21 < N < 44$
% conf.	Wavelet(Daubechies4)-without extra level			Wavelet(Daubechies4)-with extra level			
95%	**164**	179	**151**	**140**	**152**	**139**	$128 < N < 176$
97.5%	108	130	101	**85**	106	**91**	$62 < N < 95$
99%	69	79	66	60	57	61	$21 < N < 44$
% conf.	Wavelet(Symlet6)-without extra level			Wavelet(Symlet6)-with extra level			
95%	**158**	**171**	**146**	**131**	**149**	**133**	$128 < N < 176$
97.5%	98	119	**94**	**79**	100	**87**	$62 < N < 95$
99%	65	74	62	58	55	59	$21 < N < 44$

Table 6.2: Failures of the various models - boldfaced results indicate an accepted model

uations. The EWMA-based model with $\sqrt{h}$-scaling however, is not. The model is slightly too optimistic, generating failures which are outside the non-rejection region.

6.3.2 Validating the various models - constant variance

The model based on a constant variance performs well. Table 6.2 shows that for the 97.5% region, all of the values lie in the non-rejection region, implying that the Kupiec test accepts all of the models, for all horizons. However, for the 95% VaR-confidence level, the model is quite conservative. As all risk is included, the VaR is estimated too high, resulting in too few exceptions for all horizons. For the 99% VaR-confidence level, the model is a little bit too optimistic. Given the 99% confidence level, it is exceeded for nearly all horizons.

6.3.3 Validating the various models - wavelets

Table 6.2 shows the effect of including one extra level of information on the results. More models are accepted when using this variant. In case of the Haar-based model, the model is accepted in nearly all cases, except for the 99% confidence level. This implies that for the default setting, which is the 95% confidence interval, the model outperforms the GARCH-based model. Including this extra amount of

information causes the VaR-figure to increase, resulting in less exceptions. The results suggest that using this variant is preferable.

However, as already mentioned, all wavelet-based models are rejected for the 99% confidence level. When inspecting the magnitudes of the failures, the results suggest that the Haar- and Symlet6 wavelet facilitate a better performance than the Daubechies4 wavelet, due to less exceptions. The difference in performance can be linked directly to the distribution of energy corresponding to the wavelet, as a result of the different features each wavelet has.

Chapter 7

Conclusions and future research

7.1 Summary

This thesis gives an answer to the three questions articulated in the introduction. Firstly, an analysis indicates that wavelet-analysis, as a tool, can decompose a financial time-series into multiple levels. Decomposing a financial time-series reveals the behavior of this time-series at different time-scales. Interesting is to see that the Discrete Wavelet Transform can decompose a signal into multiple levels and that the signal can be perfectly reconstructed. Decomposition and reconstruction using scaling and translation based on dyadic steps is sufficient for perfect reconstruction, rather than steps of any length as the Continues Wavelet Transform does. In addition, the thesis shows that wavelets and the concept of time-scale decomposition in particular, already play an important role in finance and economics. The thesis also gives an overview of VaR-estimation and discusses the available models for VaR estimation and their shortcomings.

Secondly, the fact that wavelets are designed to have a special property, orthonormality, shows that the decomposition is energy preserving. Through use of this notion, the thesis shows that the energy of a time-series and the variance of a time-series are related. Experiments indicated that the variance (or volatility) remains preserved over all levels and all variances captured in the individual levels sum up to the original variance captured in the time-series. In addition, these experiments give insight in the distribution of energy over the individual levels.

Through use of the wavelet-energy distribution, insight can be given in the distribution of volatility over the various levels of decomposition. Based on this insight, an alternative, wavelet-based model for Value At Risk estimation is formulated. The model decomposes the original signal by performing a multiresolution analysis and makes a distinction between relevant and irrelevant information in terms of the considered time-scale. Therefore, it assesses only the levels corresponding to the relevant time-horizon, yielding an amount of risk tailored to a specific horizon.

Thirdly, the experiments have shown that the formulated wavelet-based model and a variant on this model are a decent alternative to the time-varying volatility models (GARCH/EWMA). The experiments also showed that the complex method of scaling using the Drost-Nijman equations does not yield an accurate VaR-model. The traditional method of scaling with $\sqrt{h}$ yields quite accurate models, though not in line with the theory. The fact that Drost-Nijman scaling does not yield satisfactory results, can be a result of the dataset, the result of wrongfully made assumptions or perhaps the application in the field of VaR estimation might not be appropriate at all. However, the fact still remains that scaling with $\sqrt{h}$ in an independently, identical distributed environment indeed yields satisfactory results, which has a strong

conflict with the theory Drost and Nijman (1993) proposed. As mentioned in chapter 5, lowering risk over a longer period rather than amplifying it is counterintuitive, which is in fact the result of the Drost and Nijman equations. Lowering the risk, yields a smaller Value At Risk estimate, resulting directly in a higher number of failures to happen.

Backtesting models show that for a VaR-model to be accepted, the number of the failures should lie in a specific, predetermined region, depending on the number of observations and level of (VaR) significance. The thesis shows how validation of the results gathered in chapter 5, through use of such a particular backtesting model, referred to as the Kupiec test.

Through use of this test and the corresponding findings, table 6.2, scientifically justified statements can be made. The findings indicate that the wavelet-based model, using the Haar wavelet, is in fact a decent alternative to the GARCH-based model with $\sqrt{h}$-scaling, but does not outperform this model: it underestimates the VaR on the 99% confidence level. The GARCH/EWMA models with Drost-Nijman scaling are rejected for all horizons and all confidence levels. This also holds for the EWMA model with $\sqrt{h}$-scaling. However, the performance with $\sqrt{h}$-scaling yields far more preferable results than with Drost-Nijman scaling.

The model with a constant variance underlines the added-value of the wavelet-based model: as all risk is entirely used, the model conservatively overestimates the VaR, resulting in too few exceptions, causing the Kupiec test to reject some instances of this model. The wavelet-based model indicates that by taking into consideration the appropriate amount of information, the model generates not too much and not too few failures on the real-life dataset.

7.2 Discussion

The thesis has shown that the wavelet-based model outperforms the model with constant variance; the wavelet-based model assesses only the ”relevant” information. Therefore, the VaR was estimated lower, causing more (the exact amount of) failures to occur. Interesting here, is to discuss the fact that ”relevant” information, as we may call it so, concentrates in the high-frequent regions. One might expect though, that this should rather concentrate in the lower frequent behavior: the introduction of chapter 3 already mentioned that low-frequent behavior gives the signal its identity.

In addition, another discussion can discuss the fact that a model, assuming the historical variance to be constant, performs equally well - and better in some cases - compared to the more complex models which have the assumption that volatility behaves dynamically over time. Finally, the fact that scaling with $\sqrt{h}$ yields far more accurate VaR-models than the models which use Drost-Nijman scaling, which is most certainly not in line with the theory, should be discussed as well.

Future research should most certainly assess the possibilities for extending the wavelet-based model with weights different than 1. Chapter 4 already mentioned this idea. Using the notion of weights, the weights can increase or decrease the amount of risk corresponding to a level, resulting in the VaR estimate to be higher or lower, resulting in a higher or lower number of failures to happen. Weights therefore, can help to fine tune the model. This idea can be further extended to the arena of computational intelligence. The weights can be estimated such that the model performs maximally at the trainingset. This weight estimation can be performed by learning: for each observation, an estimate of the VaR and the return over this period are available. Then, a value can be assigned to the weights, such that for each observation, the VaR estimate is close to the realized return over the same period. Algorithms such as neural networks are available for an implementation of this idea. For a more detailed discussion regarding neural networks, I refer to Jang et al. (1997).

This research considered Value at Risk estimation for a portfolio with stocks. I truly recommend research which extends the wavelet-based model for more complex, perhaps derivative-based, financial portfolios. Looking at risk from the perspective of time and scale can yield new insights, not only in the domain of finance and economics.

Bibliography

Ajassa, G., Ambrosetti, S. and Giovanni, N.: 1998, VaR (value-at-risk) models: A comparative analysis of parametric and non parametric approaches.

Arino, A.: 1996, Forecasting time series via the discrete wavelet transform, *Computing in Economics and Finance* .

Aussem, A. and Murtagh, F.: 1997, Combining neural network forecasts on wavelet-transformed time-series., *Connection Sci.* **9**, 113–121.

Berkowitz, J. and O'Brien, J.: 2001, How accurate are value-at-risk models at commercial banks?, *Technical report*.

Brealey, R. and Myers, S.: 2001, *Principles of Corporate Finance, 7th Edition*, McGraw Hill, New York.

Bredin, D. and Hyde, S.: 2001, Forex risk: Measurement and evaluation using value-at-risk, *Technical report*, Central Bank of Ireland, research department and Manchester School of Accounting and Finance.

Chew, C.: 2001, The money and income relationship of european countries by time scale decomposition using wavelets, *Preliminary paper*, New York University.

Chui, C.: 1992, *An introduction to wavelets*, Academic Press, San Diego.

Daubechies, I.: 1992, Ten lectures on wavelets, *CBMS-NSF Regional Conf. Ser. On Applied Mathematics*, Vol. 61.

Diebold, F., Hickman, A., Inoue, A. and Schuermann, T.: 1997, Converting 1-day volatility to h-day volatility: Scaling by $\sqrt{h}$ is worse than you think, *Technical report*, Pennsylvania.

Donoho, D.: 1995, Adapting to unknown smoothness using wavelet shrinkage, *Journal Of The American Statistical Association* **90**, 1200–1224.

Drost, F. and Nijman, T.: 1993, Temporal aggregation of GARCH processes, *Econometrica* **61**, 909–927.

Engel, J. and Gyzicki, M.: 1999, Value at risk: On the stability and forecasting of the variance-covariance matrix, *Research discussion paper - 1999-04*, Research and Finance Studies, Australian Prudential Regulation Authority.

Fleming, B., Yu, D., Harrison, R. and Jubb, D.: 2000, Wavelet-based detection of coherent structures and self-affinity in financial data, *The European Physical Journal B* **20**, 543–546.

Gençay, R., Selçuk, F. and Whitcher, B.: 2002, *An introduction to Wavelets and Other Filtering Methods in Finance and Economics*, Academic Press, San Diego, USA.

Gençay, R., Selçuk, F. and Whitcher, B.: 2003, Multiscale systematic risk.

Graps, A.: 1995, An introduction to wavelets, *IEEE Computational Science and Engineering* **2**.

Haas, M.: 2001, New methods in backtesting, *Research discussion paper*, Financial Engineering Research Center Caesor, Bonn.

Hembregts, P., Herzberg, A., Kalbfleisch, H., Traves, W. and Witla, J.: 1995, An introduction to wavelets with an application to andrews' plots, *Journal of Computational and Applied Mathematics* **64**, 41–56.

Hoa, D., Zhang, P.-A. and Xu, J.: 2001, Fuzzy wavelet networks for function learning, *IEEE Transactions on Fuzzy Systems* **9**.

Huij, J.: 2001, Analyse en beheer van financieel risico van aandelen, *Technical report*, Faculty of Economics, Department of Computer Science, Erasmus University Rotterdam.

Hull, J.: 2000, *Options, Futures and other Derivatives, Fourth Edition*, Prentice Hall, inc., Upper Saddle River.

Jang, J.-S., Sun, C.-T. and Mizutani, E.: 1997, *Neuro-fuzzy and Soft Computing*, Prentice Hall International.

Jorion, P.: 1996, *Value at Risk: The New Benchmark for Controlling Market Risk*, Irwin, USA.

Kirstein, R.: 2001, The new basle accord, internal ratings, and the incentives of banks, *German Working Papers in Law and Economics* **2001-1-1017**.

Misiti, M., Misiti, Y., Oppenheim, G. and Poggi, J.-M.: 2002, *Wavelet Toolbox for use with Matlab - User's Guide*, The Mathworks, Inc.

Newbold, P.: 1995, *Statistics for business and economics, fourth edition*, Prentice Hall International.

Norsworty, J., Li, D. and Gorener, R.: 2000, Wavelet-based analysis of time series: an export from engineering to finance, *IEEE International Engineering Management Society Conference*, Albuquerque, New Mexico.

Ramsey, J.: 1999, The contribution of wavelets to the analysis of economic and financial data, *Philosophical Transactions of the Royal Society of London* **357**.

Ramsey, J. and Lampart, C.: 1998, Decomposition of economic relationships by time scale using-wavelets: Money and income, *Macroeconomic Dynamics* **2**, 49–71.

Shin, T. and Han, I.: 2000, Optimal signal multi-resolution by genetic algorithms to support artificial neural networks for exchange-rate forecasting., *Expert Systems with Applications* **18**, 257–269.

Stevenson, M.: 2000, Filtering and forecasting spot electricity prices in the increasingly deregulated australian electricity market, *Technical report*, School of Finance and Economics, University of Technology Sydney, Sydney.

Struzik, Z.: 2001, Wavelet methods in (financial) time-series, *Physica A* **296**, 307–319.

Tan, S. and Yu, Y.: 1999, Complementarity and equivalence relationschips between convex fuzzy sets with symmetry restrictions and wavelets., *International Journal of Fuzzy Sets and Systems* **101**, 423–438.

van den Berg, J., Kaymak, U. and van den Bergh, W.-M.: 2003, Financial markets analysis by probabilistic fuzzy modelling, *ERIM Report Series in Management - ERS-2003-036-LIS* .

van den Bergh, W., Hallerbach, W., van der Meulen, J., van der Sar, N., Schauten, M., Smit, J., Soppe, A., Spronk, J., Steenbeek, O. and Vorst, A.: 1999, *Financiering en Belegging, tiende druk*, Rhobeta consultants.

Walker, J.: 2000, *Wavelets and their scientific applications*, Chapmann and Hall/CRC.

Wiener: 1997, Introduction to value at risk, *Risk Management and Regulation in Banking*, Jerusalem, Israel.

Chapter 8

Appendix - Tools for wavelet analysis

For the analysis of signals and images, various software tools are available. Some tools are available in a stand-alone version, others are available as an extension on existing (mathematical) software packages. For the purpose of this research, Mathworks' Matlab - a commonly used software package for research purposes - is used as the basic software package. Version 6.5 is the latest version. Matlab is very useful, as all of the operations are vector-based. In addition, Matlab facilitates the use of external, pre-compiled programs for faster calculations, which seriously reduces the time of computations. Currently, three commonly used toolboxes for wavelet analysis are available. Each of these toolboxes are designed for use in combination with Matlab.

These toolboxes are:

- Mathworks' Wavelet Toolbox - version 2.2
- Rice Wavelet Toolbox - version 2.4
- Wavelab - version 8.02

In the following sections, each of these toolboxes are reviewed in terms of functionality, ease-of-use and documentation.

8.1 Tools for wavelet analysis - Mathworks' Wavelet Toolbox 2.2

The Wavelet Toolbox is an open, customizable environment that lets you easily develop wavelet based algorithms. The Wavelet Toolbox builds on the numeric and visualization capabilities of Matlab to provide point-and-click graphical tools and command-line functions for analysis, synthesis, denoising, and compression of signals and images (Misiti et al. 2002).

8.1.1 Wavelet Toolbox - overview

The Wavelet Toolbox is a comprehensive tool that supports interactive exploration of wavelet properties and applications. The Wavelet Toolbox is useful in 1-D and 2-D applications such as speech and audio

processing, image and video processing, communications, geophysics, finance, and medicine. The toolbox provides an innovative graphical user interface and an extensive collection of Matlab routines (Misiti et al. 2002).

This toolbox features:

- Continuous wavelet transforms
- Discrete wavelets 1D - transforms,inverse transforms, reconstruction, etc.
- Discrete wavelets 2D - transforms, inverse transforms, reconstruction, etc.
- Discrete stationary wavelets
- Wavelet Packets
- Tree management utilities
- A library of standard wavelet families, including Daubechies wavelet filters, complex Morlet and Gaussian wavelets and real reverse biorthogonal and discrete Meyer wavelets
- A graphical user interface

Although the toolbox offers a vast amount of functions, for the purpose of this research only the most important functions are reviewed. These functions are mostly related to the discrete 1D wavelets, in the form of the discrete wavelet transform. This transform is discussed before in previous sections and is commonly used in finance and economics. Other applications such as 2D wavelets for image analysis, denoising and compression are beyond the scope of this research. The toolbox provides the opportunity to choose either to use the command-line or the graphical user interface (GUI). The installation of the software package is straightforward and intuitive. The installation of the Wavelet Toolbox is in this case integrated with the installation of the Matlab package. It is only a matter of choosing the appropriate installation location. A very detailed and easy-to-read manual (Misiti et al. 2002) accompanies with the toolbox as well. This manual provides a complete overview of the package, with a detailed description of wavelet analysis in general, and its applications. Each of the functions is described in detail, mostly accompanied with some numerical examples or real-life cases. Most examples are both demonstrated for use with the command-line as for use with the GUI. The Wavelet Toolbox provides several functions in the 1D wavelet domain. Each of these functions with a corresponding explanation is shown in 8.1. Note that functions are indicated with "()", regardless of the number of arguments the function accepts

It is immediately apparent that most operations as illustrated with the numerical example from chapter 3 - multilevel decomposition, reconstruction, the 1-level transform - are available. For the remainder of this section, use of the command-line is illustrated by making the same calculations based on the previously used signal from chapter 3. Next, the graphical user interface is demonstrated by analyzing the Euro-Dollar index.

For illustration purposes, the signal $f(t) = (5, 7, 10, 11, 8, 5, 4, 4)$ is considered again. To perform a single level decomposition on this signal, using the same Haar wavelet, the `dwt()` function can be used, which can be directly entered in the Matlab command-line:

```
[cA1, cD1] = dwt(f,'Haar');
```

cA1 corresponds to the set of wavelet coefficients capturing the trend, whereas cD1 corresponds to the set of wavelet-coefficients capturing the detail. This statement yields the following values:

Discrete Wavelets 1D	
`appcoef()`	Extract 1-D approximation coefficients
`detcoef()`	Extract 1-D detail coefficients
`dwt()`	Single level discrete 1-D wavelet transform
`dwtmode()`	Discrete wavelet transform extension mode
`idwt()`	Single level inverse discrete 1-D wavelet transform
`upcoef()`	Direct reconstruction from 1-D wavelet coefficients
`upwlev()`	Single level reconstruction of 1-D wavelet decomposition
`wavedec()`	Multilevel 1-D wavelet decomposition
`waverec()`	Multilevel 1-D wavelet reconstruction
`wenergy()`	Energy for 1-D wavelet decomposition
`wrcoef()`	Reconstruct single branch from 1-D wavelet coefficients

Table 8.1: Overview of Wavelet Toolbox - Discrete Wavelet Analysis 1D functionality

cA1 = 8.4853 14.8492 9.1924 5.6569

cD1 = -1.4142 -0.7071 2.1213 0

Note here that these two sets of wavelet-coefficients correspond exactly with the values from the numerical example in 3, though not written in squareroot-form. To construct the level-1 approximation and detail ($A^{\mathbf{1}}$ and $D^{\mathbf{1}}$) from the coefficients cA1 and cD1, the following statements can be used:

```
A1 = upcoef('a',cA1,'Haar',1,l_f);
```

```
D1 = upcoef('d',cD1,'Haar',1,l_f);
```

Here, `l_f` corresponds to the length of the signal $f(t)$, whereas `'a'` and `'d'` indicate that the reconstructed signal corresponds to either approximation or detail. Execution of this statement yields:

A1 = 6.0000 6.0000 10.5000 10.5000 6.5000 6.5000 4.0000 4.0000

D1 = -1.0000 1.0000 -0.5000 0.5000 1.5000 -1.5000 0 0

Notice again that D1 corresponds to the result of $D^{\mathbf{1}}$ in the numerical example of 3. In a nearly similar fashion, a level-3 decomposition using the Haar wavelet can be performed on the signal $f(t)$. Nearly, as the wavedec() routine, rather than the dwt() routine, has to be invoked for a multilevel decomposition:

```
[C,L] = wavedec(f,3,'Haar');
```

Although this statement doesn't yield any direct information, each of the wavelet coefficients corresponding to each of the levels can be extracted individually. These coefficients can be extracted using the following statement for the wavelet coefficients capturing the trend subsignal:

```
cA3 = appcoef(C,L,'Haar',3);
```

and the following statements to extract the wavelet coefficients capturing the detail signal:.

```
cD3 = detcoef(C,L,3);
```

```
cD2 = detcoef(C,L,2);
```

```
cD1 = detcoef(C,L,1);
```

These statements yield the following results, equal to the results of example 3.2.1:

cA3 =19.0919

cD1 = -1.4142 -0.7071 2.1213 0

cD2 =-4.5000 2.5000

cD3 =4.2426

Finally, the level 3 approximation and the level-1, 2 and 3 detail of the original signal can be reconstructed directly from the results of the decomposition, `[C,L]`. To reconstruct the level 3 approximation from C, the following statement can be entered:

```
A3 = wrcoef('a',C,L,'Haar',3);
```

The following statements can be entered to reconstruct the level-1, 2 and 3 detail:

```
D1 = wrcoef('d',C,L,'Haar',1);

D2 = wrcoef('d',C,L,'Haar',2);

D3 = wrcoef('d',C,L,'Haar',3);
```

The results of these statements are the following:

A3 = 6.7500 6.7500 6.7500 6.7500 6.7500 6.7500 6.7500 6.7500

D3 = 1.5000 1.5000 1.5000 1.5000 -1.5000 -1.5000 -1.5000 -1.5000

D2 = -2.2500 -2.2500 2.2500 2.2500 1.2500 1.2500 -1.2500 -1.2500

D1 = -1.0000 1.0000 -0.5000 0.5000 1.5000 -1.5000 0 0

Which counts up to the original signal, as demonstrated in 3

8.1.2 The Wavelet Toolbox - Graphical User Interface

In addition to the command line, the Wavelet Toolbox provides the opportunity to analyze signals and images using a graphical user interface, instead of entering the individual statements. Using the GUI yields immediate results, though reduces the amount of control one has on the process of analysis. The GUI offers graphical tools for basic 1D and 2D analysis. In addition, the GUI offers specialized 1D and 2D tools and tools to analyze and examine wavelets and their corresponding wavelet families and wavelet packets. Finally, the GUI provides tools for signal and image extension. For the purpose of this research, this section concentrates solely on the basic 1D wavelet tools. As an example, the index of the Euro/Dollar ratio will be analyzed using wavelets. This index starts at January 5th 1999 and ends at March 1st 2000. The signal is presented in figure 8.1.

The signal can be analyzed with a variety of different wavelets; Haar wavelets, Daubechies wavelets, Symlets, Coiflets or Biorthogonal wavelets. Each of these wavelets can be configured with different properties. The signal can be analyzed at different levels. As mentioned earlier, each class of wavelets has its own properties. The first decomposition is based on the Haar wavelet, as can be seen in figure 8.2. Although the level-1 detail signal resembles the level-1 detail signal of the Haar decomposition strongly, it's obvious that the trend signal A^3 is a coarse and 'stepwise' approximation of the original signal. This can be justified with the fact that the Haar wavelets have a small support, computing averages and differences based on a smaller amount of values.

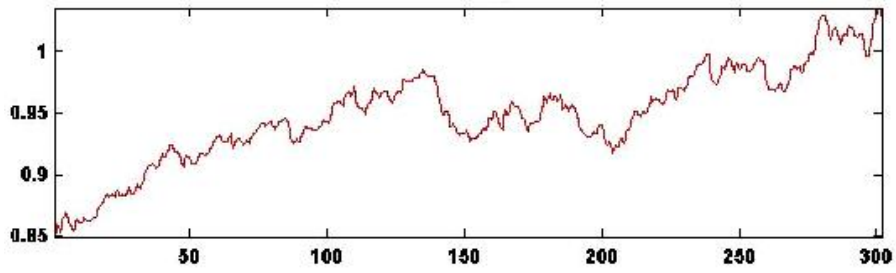


Figure 8.1: Euro-Dollar index (January 5th 1999 - March 1st 2000)

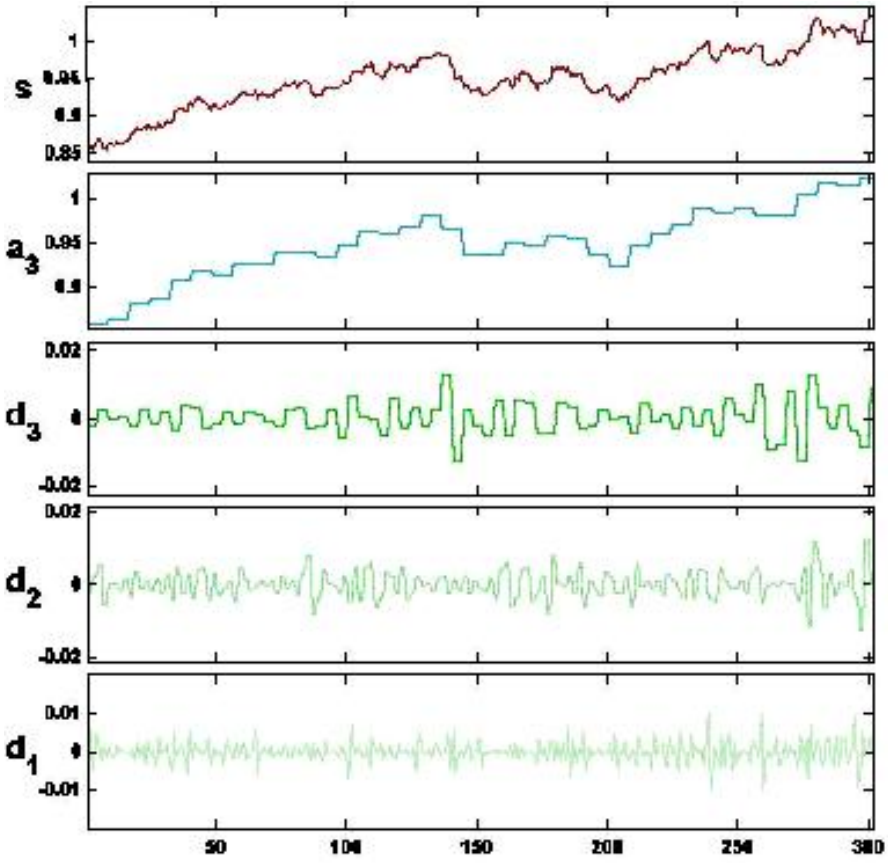


Figure 8.2: Haar decomposition of the Euro-Dollar index

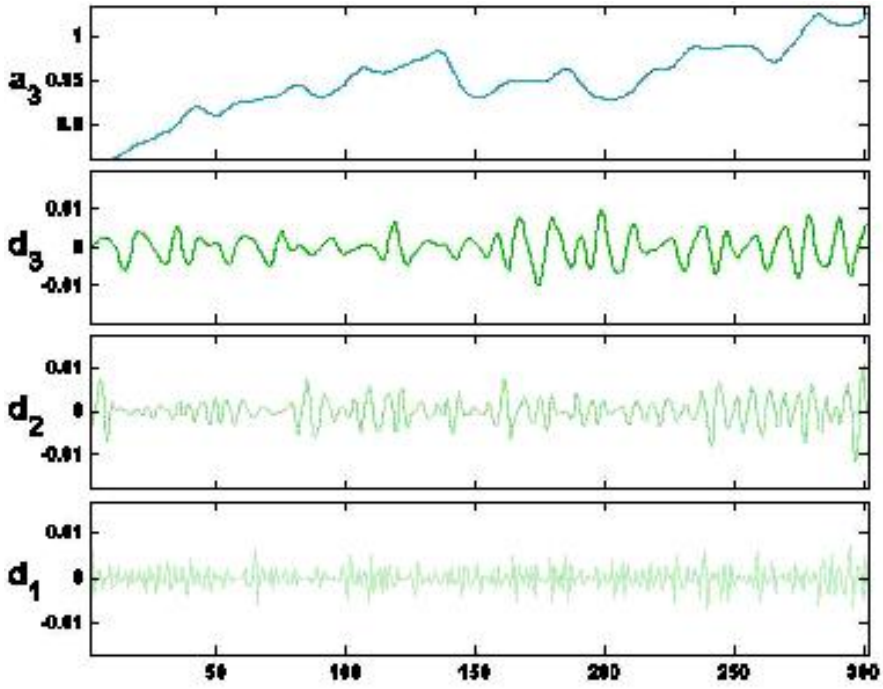


Figure 8.3: Symlet5 decomposition of the Euro-Dollar index

The second decomposition is based on the Symlet5 wavelet, as can be seen in 8.3. It is immediately obvious that the Symlet5 decomposition yields a finer trend signal A^3, due to its large support width. In the final period of the original signal, heavier oscillations are evident. As can be seen, this is visible in the level-2 and level-3 detail signal, D^2 and D^3.

8.1.3 The Wavelet Toolbox - conclusion

To conclude with, it can be stated that Mathworks' Wavelet Toolbox is an extensive, comprehensible and easy-to-use environment for wavelet analysis. The toolbox is complete in its functionality, and the use of the functions is intuitive and straightforward, as it corresponds strongly with the available literature. In addition, the documentation gives complete and detailed insight in how the toolbox works, accompanied with detailed information on wavelet-analysis in general. The command line gives insight in the operations regarding wavelet-analysis, whereas the graphical user interface gives the user visual insight in the process of decomposition, reconstruction, denoising, compression and more. In short, the Wavelet Toolbox scores very high on ease-of-use, functionality and documentation.

8.2 Tools for wavelet analysis - Rice Wavelet Toolbox

The Rice Wavelet Toolbox is a collection of Matlab routines, developed for wavelet decomposition and wavelet denoising and the toolbox is developed by members of the DSP group, from the Rice University in the USA.

Wavelet transforms	
`mdwt()`	Discrete orthogonal wavelet transform using Mallat alg. (1D and 2D)
`midwt()`	Inverse discrete orthogonal wavelet transform
`mrdwt()`	Undecimated (redundant) discrete wavelet transform (1D and 2D)
`mirdwt()`	Inverse undecimated discrete wavelet transform
`daubcqf()`	Daubechies filter coefficients
Wavelet domain processing	
`denoise()`	Denoise signal and images by thresholding wavelet coefficients
`HardTh()`	Hard thresholding
`SoftTh()`	Soft thresholding
Other	
`makesig()`	Create Donoho-Johnstone test signals
`compile()`	Compile the Rice Wavelet toolbox (could be necessary to run the first time)

Table 8.2: Overview of Wavelet Toolbox - Discrete Wavelet Analysis 1D functionality

8.2.1 Rice Wavelet Toolbox - overview

In contrast to the previously reviewed toolboxes, the Rice Wavelet Toolbox (RWT) is a very simple toolbox, as it has only functionality implemented for wavelet decomposition and denoising, and some supporting functionality. The installation of the toolbox is easy and straightforward. The documentation however, is rather scarce, as it is restricted to a description of the functions. The help of the functions is good, as each function is described using small examples for both 1D and 2D applications and each parameter is described completely. The toolbox provides the functions as shown in table 8.2.

In the remainder of this section, use of the Rice Wavelet Toolbox is illustrated through an example of wavelet transformation. Next, an example of wavelet-based denoising is given by using the Euro-Dollar index.

Similar to the other toolboxes, RWT provides the functionality to perform a decomposition based on the original signal. For illustration purposes, the signal $f(t) = (5, 7, 10, 11, 8, 5, 4, 4)$ is considered. To decompose a signal, the `mdwt()` function has to be invoked. To perform a single level decomposition on this signal, using the Haar wavelet, the `mdwt()` function can be used, which can be directly entered in the Matlab command window:

```
[y,L] = mdwt(x,h,L);
```

here `y` corresponds to a new vector containing the wavelet coefficients, with the coefficients arranged in the order of increasing frequency. `L` corresponds to the depth of the decomposition and `x` corresponds to the original signal and `h` corresponds to the filter. To invoke the statement correctly, a filter needs to be created. This can be done by the statement:

```
[h_0,h_1] = daubcqf(N,TYPE);
```

where holds that `h_0` corresponds to the Daubechies scaling filter and `h_1` corresponds to the Daubechies wavelet filter. `N` corresponds to the length of the filter (i.e. the order of the Daubechies wavelet). `TYPE` corresponds to an optional parameter that distinguishes the minimum phase, maximum phase and mid-phase solutions. Finally, the statement

```
h = daubcqf(2, 'min');
```

can be used. Recall that a Daubechies-2 wavelet is similar to the Haar wavelet. `Min` is the default setting. With this statement, the following statement can be made:

```
[y,L] = mdwt(f, h, 1)
```

yielding:

```
y = 8.4853 14.8492 9.1924 5.6569 -1.4142 -0.7071 2.1213 0
```

Note here that these two sets of wavelet-coefficients correspond exactly with the values from the numerical example in chapter 3, though not written in squareroot-form. Furthermore, a level-2 wavelet decomposition can be performed on the signal by invoking the following statement

```
[y,L] = mdwt(f, h, 2)
```

yielding:

```
y = 16.5000 10.5000 -4.5000 2.5000 -1.4142 -0.7071 2.1213 0
```

Similarly, a statement for the level-3 decomposition yields

```
y = 19.0919 4.2426 -4.5000 2.5000 -1.4142 -0.7071 2.1213 0
```

The original signal can be reconstructed through the inverse wavelet transform, using the midwt function:

```
[x,L] = midwt(y,h,L)
```

Recall that a j-level reconstruction can be based only on a j-level decomposition. For the level-3 reconstruction of the original signal, the result is:

```
x = 5.0000 7.0000 10.0000 11.0000 8.0000 5.0000 4.0000 4.0000
```

which equals the original signal $f(t)$. Although the RWT toolbox can perform a decent decomposition of the signal, this can only be performed using wavelets of the Daubechies family. In addition, the signal can only be reconstructed to the original signal with the inverse wavelet transform. There is no functionality implemented for wavelet reconstruction yielding a $j-th$ approximation and $j-th$ detail of the original signal. As a result, this has to be done by performing equation (3.19) manually.

8.2.2 Rice Wavelet Toolbox - command line denoising

RWT has implemented functionality for wavelet-based denoising. Denoising is equal to performing a $j-$level wavelet decomposition on the original signal, eliminating all values below a certain threshold and reconstructing the original signal back using the inverse wavelet transform. For illustration purposed, the previously used Euro-Dollar index will be used here. In RWT, this can be done by invoking the statement:

```
[xd,xn,option] = denoise(x,h,type,option);
```

where `xd` corresponds to the original signal and `xn` corresponds to the noise term $xd - x$. The variable `option` corresponds to a vector of actual parameters used by the program. The variable `x` corresponds to the original signal and `h` corresponds to the wavelet filter. The variable `type` corresponds to the type of transform (DWT or undecimated DWT) and `option` corresponds to an option vector. After invoking this function, the original and denoised signal can be plotted in one particular window, which makes the effect of the denoising clear. The result is shown in figure 8.4.

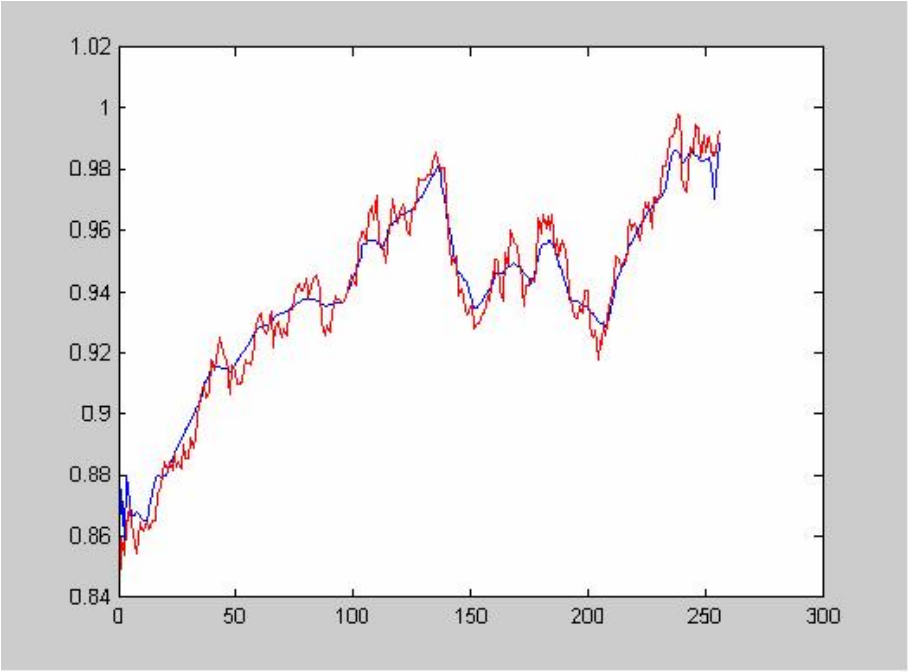


Figure 8.4: Euro-Dollar index - original (red) and denoised (blue)

8.2.3 Rice Wavelet Toolbox - conclusion

To conclude, it can be stated that the RWT is a toolbox which can be used for the smaller wavelet applications. The wavelet decomposition yields the correct results, but lacks functionality for complete reconstruction, which is a pity. Wavelet based denoising is complete as well, as the user can choose between several thresholding techniques and parameters. Although the help in Matlab is complete, the documentation accompanied with the toolbox is rather scarce. In this context, RWT scores well on ease-of-use, but scores low on both functionality and documentation.

8.3 Tools for wavelet analysis - Wavelab 8.02

Another toolbox available for wavelet analysis is Wavelab. The current release of Wavelab is version 8.02. It is a library of Matlab routines for wavelet-analysis, wavelet packet analysis, cosine-packet analysis and matching pursuit. This library has been used in teaching courses on wavelet transforms and time-frequency transforms at Berkely and Stanford University.

8.3.1 Wavelab - overview

The toolbox features:

- Continuous wavelet transforms
- Discrete wavelets 1D - transforms, inverse transforms, reconstruction,
- Discrete wavelets 2D - transforms, inverse transforms, reconstruction,

- Discrete stationary wavelets,
- Wavelet packet analysis,
- Cosine packet analysis,
- Matching pursuit,
- Time-frequency analysis,
- Fractals.

Wavelab offers a wide variety of functionality. Most of the functionality as described above, is implemented with the support of a substantial amount of functions. However, some of the functionality, especially the discrete 1D/2D wavelet transformation is not as richly implemented as the wavelet packet functionality. Again, for the purpose of this research, the focus lies upon discrete 1D wavelet analysis, as this is most commonly used in the domain of finance and economics. Although there is some browser functionality for signals, Wavelab doesn't offer a graphical user interface as extensive as the Mathworks' toolbox. In addition, the installation of Wavelab is relatively difficult and requires in-depth knowledge of Matlab, inspite of the accompanied documentation. With respect to this documentation, the manual of Wavelab is restricted to an overview of the toolbox, installation information and a summation of the available functions and the corresponding input-output parameters, which is basically the same as invoking the help command on these functions from the Matlab command-line. For the remainder of this chapter, use of Wavelab is demonstrated through an example of wavelet decomposition.

8.3.2 Wavelab - command line decomposition

Similar to other toolboxes, Wavelab provides the functionality to perform a decomposition based on the original signal. For illustration purposes, the signal $f(t) = (5, 7, 10, 11, 8, 5, 4, 4)$ is considered. To decompose a signal, the `fwt_po` function has to be invoked, which is a forward wavelet transform (fwt), periodized and orthogonal (po). To perform a single level decomposition on this signal, using the Haar wavelet, the `fwt_po` function can be used, which can be directly entered into the Matlab command-line. Wavelab highlights in the help function that `fwt_po` is the standard function for discrete orthogonal wavelet decomposition, which is similar to the decompositions illustrated in the other toolboxes. The usage of the function is:

```
wc = FWT_PO(x,L,qmf);
```

where **wc** corresponds to a vector containing wavelet coefficients and **x** corresponds to the original signal. **L** corresponds to the level and **qmf** corresponds to the quadrature mirror filter, which is basically the wavelet filter the decomposition is based on. Similar to RWT, the filter has to be constructed first. In this context, the statement:

```
qmf = makeONfilter('Haar');
```

suffices to construct the filter. Finally, the following statement yields the level-1 decomposition:

```
wc = FWT_PO(f,2,qmf);
```

```
wc = 8.4853 14.8492 9.1924 5.6569 1.4142 0.7071 -2.1213 0
```

Note here that although the magnitudes of the values correspond exactly to the values of the numerical examples and the examples of the other toolboxes, the signs of the values don't. This can be a result of a different interpretation of the Haar-wavelet.

Featured functionality	Wavelet Toolbox	Wavelab	Rice Wavelet Toolbox
Continues wavelets	✓	✓	
Wavelet denoising	✓	✓	✓
Wavelet compression	✓	✓	
Wavelet packets	✓	✓	
Discrete stationary wavelets	✓	✓	
Graphical User Interface	✓		
Wavelet families, descriptions and operations	✓		
Wavelet demonstrations	✓	✓	
Tree management utilities	✓	✓	
Time-frequency tools		✓	
Fractals		✓	
Discrete wavelets - 1D	✓	✓	✓
Discrete wavelets - 2D	✓	✓	✓
Cosine packet analysis		✓	
Matching pursuit		✓	

Table 8.3: Overview of functionality - all toolboxes

8.3.3 Wavelab - conclusion

The Wavelab toolbox is very functional in terms of a large collection of functions. Each of the applications, with exception of decomposition, is supported with a substantial amount of functions. However, the installation is rather complicated. In addition, using the functions demands sincere knowledge of both Matlab and wavelet analysis in general. The documentation is rather scarce as well, for it is a print-out of the help corresponding to all of the functions.

8.4 Comparison of the toolboxes

This section discusses the three toolboxes in the form of a brief comparison. Mathworks' Wavelet Toolbox is a very functional toolbox and covers nearly the same functionality as Wavelab. However, each of these functions are better documented and are easier to use. In addition, the extensive GUI gives tremendous insight in the use of wavelets and the effect of different properties on the decomposition. In addition, Mathworks' Wavelet toolbox facilitates the wavelet reconstruction on different levels, yielding approximation and detail components A^j and D^j of the original signal, both available in the command-line and in the GUI, which is extremely useful in financial time-series analysis. Finally, the documentation is extensive and complete and gives an excellent overview of the concept of wavelet analysis and all of the functionality, including numerical examples. It is important that no knowledge of wavelet analysis at all is demanded in order to start working with the Wavelet Toolbox, where a sincere knowledge of both Matlab and wavelet analysis is demanded for working with Wavelab. The installation of Wavelab was not straightforward. It demanded in-depth knowledge of Matlab in order to get the toolbox working. The Rice Wavelet toolbox scores, as already mentioned, low on both functionality and documentation. However, the toolbox is straightforward and easy-to-use. With respect to functionality, an overview is given for each of the toolboxes. This overview is shown in table 8.3. This comparison facilitates the final conclusion, which is shown in table 8.4.

Criteria	Wavelet Toolbox	Wavelab	Rice Wavelet Toolbox
Functionality	++	++	-
Ease-of-use	++	-	+
Documentation	++	-	-

Table 8.4: Tools for wavelet analysis - scoreboard