



An empirical study of hedge-ratios for corporate bonds

Vincent Herman Tony Van der Sanden

Erasmus School of Economics

Master Thesis Economics & Informatics

Computational Economics
Student: Vincent H.T. Van der Sanden
Student ID: 261523
Supervisor: Prof.Dr.Ir. U. Kaymak
Co-Reader: Dr. Ir. M.P.E. Martens
June 2010



An empirical study of hedge-ratios for corporate bonds

Vincent Van der Sanden

Abstract

Investments in corporate bonds, or credits, are subject to two important sources of market risk: interest rate risk and credit risk.

1. Interest rate risk: uncertainty in risk-free interest rate (government bonds)
2. Credit risk: uncertainty of potential downgrades and default

These two market risks provide an opportunity for the active investor to gain additional returns or avoid losses. Active credit investors are specialized in evaluating and comparing the credit risk of corporate bonds. Given their analysis of credit risk these investors buy a portfolio of bonds of which they expect higher returns than the benchmark which usually consists of all available bonds in the market. However, an investment in credits may also lead to an unintended exposure to interest rate risk. Since the specialized credit investor does not have the skill to predict the risk-free interest rate he or she does not want to have such an exposure. Therefore, the investor wants to hedge unintended interest rate risk in the portfolio.

The duration is a well known measure for interest rate sensitivity and is used to set up a hedge against the unintended interest rate risk in the portfolio. This hedge should immunize the loss and profit resulting from interest rate changes. De Backer (2006) [5] shows that this method of hedging is accurate for bonds with a low spread, but for bonds with higher spreads this is not an accurate way to immunize the interest rate risk as is shown by Ben Dor et al (2004) [1]. For high yielding bonds the hedge does not reduce the risk on the contrary it has even increased it during specific sub-periods in history. Therefore the hedging of corporate bonds to interest rate risk is an interesting field of investigation. Nevertheless, not much academic research has been published in this area which is further underlining the need to do new research.

De Backer (2006) [5] proposes a new measure, Empirical Duration. Empirical Duration (ED) is the sensitivity of a credit investment to the interest rate risk. De Backer finds that the ED is depending on the credit spread and the maturity of the corporate bond. By combining 5 different intervals of the maturity and 6 different intervals of the credit spread a so called Maturity-Spread Table (MS-Table) is created with 30 distinct Maturity-Spread buckets. Each bucket has a different ED value, the ED of a specific corporate bond is determined by matching its maturity and spread to the buckets. The ED of a bond portfolio is the market value weighted average of the individual ED values of

the bonds. When the ED of the portfolio is different from the ED value of the benchmark the credit investor is able to hedge out the undesired interest rate risk exposure.

Within the initial MS-Table standard intervals were used without very well thought-out considerations. The interval values of the MS-Table have a very big influence on the ultimate table of Empirical Durations which will be used to hedge against unintended interest rate risk. Therefore the interval decision of the MS-Table is highly important.

We use a genetic algorithm to evaluate and improve the quality of the MS-Table that has been developed by De Backer. We are able to reduce the number of maturity-spread buckets from 30 to 15. In addition the ED values of the new buckets are optimized. The new MS-Table has a 52% better fitness function than the MS-Table from De Backer. The new MS-Table delivers a better hedge of undesired interest rate to a credit investor.

Contents

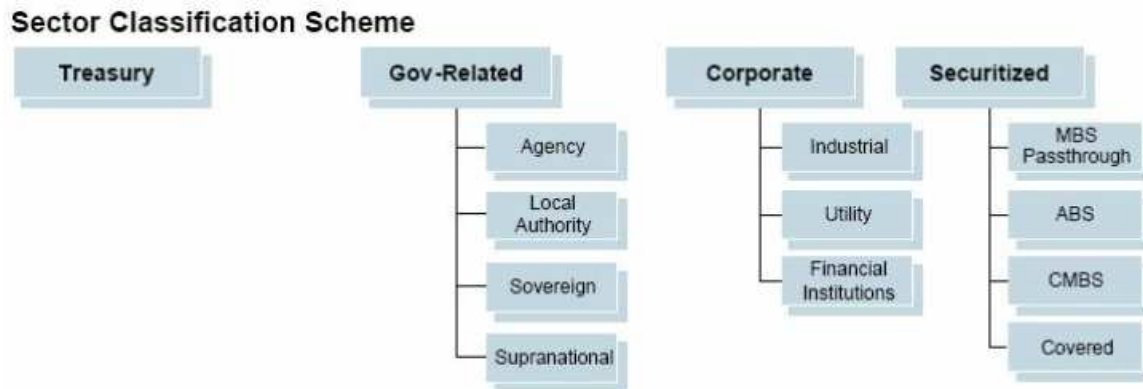
Chapter 1 - Introduction.....	7
1.1 Fixed income landscape.....	7
1.2 Example Coca Cola.....	7
1.3 Professional credit investor.....	11
1.4 Literature study.....	12
1.5 Goal.....	13
1.6 Research Questions.....	14
1.7 Research Setup.....	14
1.8 Outline.....	15
Chapter 2 - Corporate Bonds.....	17
2.1.1 Short introduction to Bonds.....	17
2.1.2 Factors that Affect the Price.....	18
2.1.3 Price and Interest Rates Typically Move in Opposite Directions.....	19
2.1.4 Bond pricing and Excess Return.....	20
2.1.5 Separation of Responsibilities in practice.....	21
2.2 General Introduction to bonds.....	21
2.3 Risks and interest rate risk of bonds.....	23
2.4 Modified duration.....	27
2.5 Approximation for the percentage change in bond price.....	30
2.6 Importance of a duration neutral portfolio.....	31
Chapter 3 - Genetic Algorithms.....	35
Outline.....	35
3.1 What is a Genetic Algorithm?.....	36
3.2 Genetic Algorithm Terminology.....	37
3.3 Outline of the Algorithm.....	37
3.4 Genetic Representation.....	38
3.5 Fitness Function.....	40
3.6 Used settings.....	41
Chapter 4 – The Hedge-ratio Model and Bucket Classification.....	43
4.1 Hedge-ratio Model.....	43
4.2.1 Data.....	44
4.2.2 Data analysis.....	46
4.2.3 Graphs of data analysis – Visualization of the data.....	47
4.2.4 Data fields.....	49
4.2.5 Monthly change in government yield.....	49
Chapter 5 – Research Setup.....	53
5.1 Research Steps.....	53
5.1.1 Outline.....	53
5.2 Experiment Description.....	54
5.2.1 Experiment 1 - Analyze the dynamics of the ED-space over time.....	54
5.2.2 Experiment 2 - Fitness Function Evaluation for MS-Tables.....	55
Chapter 6 – Results.....	59
Outline.....	59

6.1	Experiment 1 - How does the Empirical Duration model behave through time	59
6.1.1	Experiment 1.1 - 1 bucket, 1 time frame	59
6.1.2	Experiment 1.2 - 1 bucket, 131 time frames	61
6.1.3	Experiment 1.3 - 30 buckets - 131 time frames	68
6.2	Experiment 2 - Fitness Function Evaluation.....	70
6.4	Overview of Results.....	76
6.4.1	Experiment 1, empirical dependencies	76
6.6	Conclusions.....	77
Chapter 7	– Follow-up research	79
Appendix A)	Matlab Code.....	81
Glossary	91
Bibliography	93

Chapter 1 - Introduction

1.1 Fixed income landscape

The current Fixed Income landscape consists of Treasury, Government Related, Corporate and Securitized bonds. A Treasury bond is a debt instrument issued by a government. Government Related bond have direct guarantees of a timely payment of interest and principal from central governments. Bonds, whose interest and principal payments are backed by the cash flows from a portfolio or pool of other assets, are called securitized bonds. Bonds that are issued by corporations are called Corporate bonds or in short Corporates. Figure 1.1 shows the Sector Classification Scheme of Barclays Capital which shows the global split up of different types of bonds within Fixed Income.



Source: Barclays Capital

Figure 1.1

In this research project we focus on corporate bonds and mainly on the interest rate sensitivity of corporate bonds.

1.2 Example Coca Cola

Corporate Bonds have two important sources of market risk: interest rate risk and credit risk. The (risk-free) interest rate risk is due to changes of market prices of government bonds. These interest rate changes have an impact on the price of corporate bonds. Therefore, government bonds and corporate bonds have a common risk factor: the interest rate risk. The credit risk which is the risk of potential downgrades and default has its own impact on the price of a corporate bond, but may be small relative to the impact of the interest rate risk. We start with an illustrative example to demonstrate the two risks for a bond of Coca Cola Enterprises:

We compare two different bonds that mature in February 2012. The first bond is a corporate bond that is issued by Coca Cola Enterprises. The second bond is a US Treasury bond which is issued by the US Government. The details of both bonds on 30 November 2007 are shown in Table 1.1.

CCE 8.5% 02/12**UST 4.875% 02/12**

Issuer	Coca Cola Enter.
Currency	USD
Amount outstanding	0.25 billion
Issue date	2/1/1992
Maturity date	2/1/2012
Coupon	8.5%
Price on 30/11/2007	114.642
Yield-To-Maturity	4.59%
Credit spread	1.28%

Issuer	US Treasury
Currency	USD
Amount outstanding	24.8 billion
Issue date	2/15/2002
Maturity date	2/15/2012
Coupon	4.875%
Price on 30/11/2007	106.09375 (106-03)
Yield-To-Maturity	3.31%

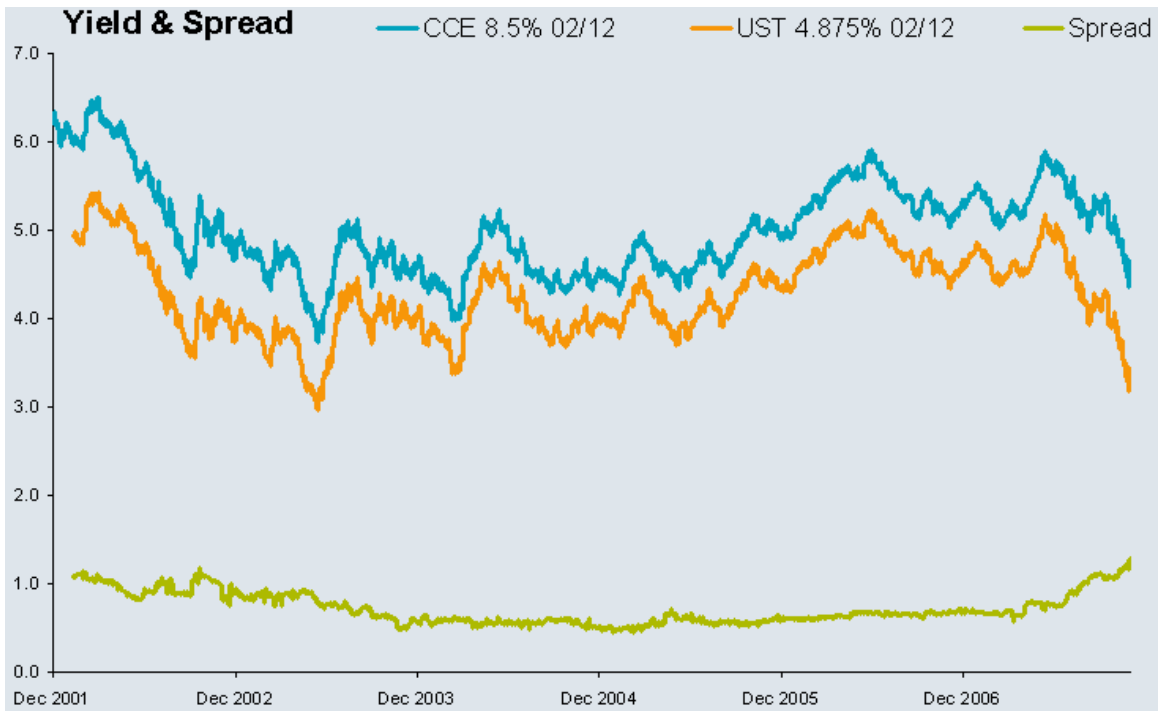
Table 1.1 - Details of the Coca Cola and US Treasury bonds.

The Yield-To-Maturity of the Coca Cola bond is 4.59% and the Yield-To-Maturity of the US Treasury bond is 3.31%. The difference between both yields is 1.28%, which is the Credit spread. This extra yield is required by the market because the perceived risk of investing in a corporate like Coca Cola is higher than investing in a US Treasury and the compensation for this risk is 1.28% or 128 basis points (bps).

To illustrate how both prices and yields have evolved through time I have included two graphs. The first graph (Graph 1.1) shows the price from December 2001 until December 2007. The second graph (Graph 1.2) shows the yield from both instruments and the differences between both yields (Spread). The period starts with the first observation of the Treasury bond that was issued in February 2002. The period ends with the start of the credit crisis. I have omitted the most recent period to be able to show the behavior of a safe corporate bond in a more tranquil period.



Graph 1.1 - Price graph



Graph 1.2 - Yield & Spread graph

The price graph shows that the price of the Coca Cola corporate bond and the US Treasury bond generally move in the same direction. In the yield & spread graph you can clearly see that the yield moves in the opposite direction of the price.

As mentioned earlier, there are two important sources of market risk for a credit investment: interest rate risk and credit risk. The interest rate risk is clearly visible in the large yield movements in the Yield & Spread graph where the US Treasury bond movements and the Coca Cola bond seem to be highly correlated. In this example the interest risk is the largest source of risk. The difference between in yield between the government bond and the credit bond is plotted in Graph 1.2 as the spread, which is relatively stable during this period. In the recent credit crisis the spreads of credit investments have been much more volatile.

The spread offers the credit investor a reward for the additional credit risk. A professional credit investor mainly focuses on the spread of a corporate bond and is less interested in the yield of the bond. The reason behind this has to do with the clients that hire a credit investor to manage their money. The first decision of the client is to invest the money in bonds. The client can choose between government bonds and corporate bonds. Both generally move in the same direction but investing in a corporate bond adds credit risk. Although the investment contains more risk than investing in a government bond, there is also a reward for taking this extra risk. Because the client already made the decision to invest in bonds a credit investor tries to mainly add extra value by selecting the best corporate bonds.

When the risk free interest rate changes, a bond with a higher maturity will have a bigger price change than a bond with a lower maturity. When one corporate bond is sold and another corporate bond with a different maturity is bought back the interest rate sensitivity of the portfolio will change. This is an unintended change because the reason one bond was sold and another bond was bought was because the investor was intending to change the credit risk, not the interest rate risk. In the Coca Cola example above the interest rate risk would probably affect the final return much more than the credit risk. To be able to add value on selecting the right credits without a disturbance from interest rate risk the interest rate risk should be hedged.

In June 2003, the bottom of the stock market after the IT bubble burst, the price of both bonds are at the peak with Coca Cola trading at 134.81 and the US Treasury bond at 114.06 and the spread at this time was 78 basis points (bps). The following year yields went up and prices dropped for the US Treasury bond and for the Coca Cola bond. Investing in one of both instruments would have led to a negative return on the price of the instrument. Though, the spread decreased in this period from 78 bps to 59 bps in June 2004. The opposite happened in the last period in the graph, from June 2007 forward the prices of both bonds go up and both yields decline. At the same time the spread, or the difference in yields between both instruments, increased from 76 bps to 128 bps. It is a known phenomenon that in periods when the yield increases there is a tendency that spreads decline, and in periods where yields decrease the spread goes up. There is a negative correlation between yield movements and spread movements. The reason behind this is because when yields go down the economy is usually in a recession and credit risk is increasing because of an increased level of rating downgrades and defaults. The opposite is also true when yields go up the economy is recovering and credit risk is generally decreasing. This negative relationship between treasury yields and corporate bond yield spreads is reported by Duffee (1998) [6]. In addition Jarrow and Turnbull (2000) [9] find that high yield corporate bonds (credit rating lower than Baa) have a smaller interest rate risk than low yield bonds (like Coca Cola which currently has an Aa credit rating).

The Coca Cola example shows the difference between the interest rate risk and the credit risk. From June 2003 forward the economy improving and therefore the yields of government and credit bonds went up. With the more positive economic environment the quality of Coca Cola was perceived better because of a more stable cash income for the company and the required extra reward compared to investing in a government bond decreased.

This clearly shows that although an investment in Coca Cola would led to a negative return in absolute terms because of the change in interest rates, the decision to buy Coca Cola instead of a US treasury bond was a favorite one and in fact earned money in a relative sense. A credit investor is looking at many different bonds and is mainly focused on the spread level. The credit investor's goal is to buy the bonds from which the spread goes down the most or increases less than the spreads of other corporate bonds. Although buying Coca Cola performed better than buying the US Treasury bond, if you could have invested in other bonds than Coca Cola, buying a bond of which the spread

went down even more would have been a decision even more favorable than investing in Coca Cola.

1.3 Professional credit investor

To give a concrete example of the specialization level of the professional credit investor I briefly describe how Robeco is organized. Within the Robeco Fixed Income department there is a split in responsibilities and skill-sets regarding the investments. Fixed income investment decisions are taken by a separate rates team and credit team. The rates team has expertise in trading on government bonds and in government bond yield curves and decide on the duration exposure of the investment portfolios. The credit team consists of analysts that investigate companies to determine their credit risk. Given the market spread and expected credit risk the portfolio managers of the credit team decide which corporate bonds to buy. Because of the split in responsibilities the credit team does not have a clear view on interest rate movements. This is out of their scope, their work is to generate performance on the change in credit risk. The complete portfolio is usually compared to a benchmark that contains all investable bonds within a certain country, currency or sector. The difference between the performance of the portfolio and benchmark is called outperformance. The goal is to achieve a positive outperformance for the client.

When the credit investor from a credit team decides to sell one bond, and buy back a new bond from the proceeds of the sale, the main focus is on the spread of the new bond. The credit investor expects that the spread of the new bond will decrease more than spread of the bond that has been sold. The more a spread decreases, the more return the credit team will add to the investment portfolio. Though, if the new bond has a different maturity than the bond that was sold the position leads to an unintended exposure to interest rate risk. The reason is that if you have a bond with a longer maturity the price of this bond will move harder when the risk-free interest curve moves, in other words, the bond has a larger duration. This risk is an unintended risk that may jeopardize the potential positive return from a decrease of the spread, as shown in the Coca Cola example. Therefore the investor would like to hedge out the interest rate risk.

The correct or 'optimal' hedge is hard to identify. Currently at Robeco the duration is used to set up the hedge. Though, recent studies show that you should not use 100% of the duration to hedge the interest rate risk but less, which means that we need to put effort in identifying the correct hedge-ratio. In order to get a good feel for the 'optimal' hedge-ratio or hedge-ratio matrix we will take a close look at many periods in the past and will investigate what would have been the optimal hedge-ratio in a certain period. The problem is even more complex because the 'optimal' hedge-ratio is different for bonds with different characteristics. Those characteristics chosen to split up the universe of bonds in buckets are the maturity and the credit spread of a bond. We developed a Maturity-Spread Table (MS-Table) to split up bonds in groups and identify each group separately. To find an improved MS-Table that performs best in most periods, we generated different MS-Tables and tested how good the performance of such an MS-Table would have been over time. The quality of one MS-Table was measured by a

fitness-function which we carefully designed based on our ideas on which table would fit our needs best. Our goal is not to generate a model that predicts the future but our initial goal is to find the best MS-Table in the past and qualitatively try to learn from this.

We start with an iterative search to generate and evaluate many MS-Tables and enhance this by using genetic algorithms to find the best MS-Table. The genetic algorithm helps us with finding an improved MS-Table and gives us valuable insight in structure of a powerful MS-Table. The found MS-Table has a 52% better fitness function than the MS-Table from De Backer. The new MS-Table delivers a better hedge of undesired interest rate to a credit investor.

1.4 Literature study

Not much academic research has been published on hedging credits which underlines the need to do new research. Jarrow (1978) [8] claims to be the first to extend the duration risk measure for corporate bonds.

Belongia, M.T., and G.J. Santoni., 1984, Hedging Interest Rate Risk with Financial Futures: Some Basic Principles [14] explain the basic principles behind hedging and how the duration gap and interest rate can be hedged with futures. They show calculation examples and show some illustrative examples with clear tables with expected streams of receipts and payments of futures and of assets and explain how to implement such a hedge.

Fridson, M., and J. Kenney., How do Changes in Yields Affect Quality Spreads?, *Extra Credit* (July/August 1994), pp. 4–13. [12] Describes a relation between yields and spread levels and how this affects the institutional investor. There is a negative correlation between yield movements and spread movements. The reason behind this is because when yields go down the economy is usually in a recession and credit risk is increasing because of an increased level of rating downgrades and defaults. The opposite is also true when yields go up the economy is recovering and credit risk is generally decreasing. This negative relationship between treasury yields and corporate bond yield spreads is reported by Duffee (1998) [6]. In addition Jarrow and Turnbull (2000), Empirically returns on high yield bonds have a higher correlation with equity index returns and a lower correlation with Treasury bond index returns than do low yield bonds., [14] show the declined interest rate sensitivity of credits for higher spreads (high yield). Mentink, A., 2005, *Essay on Corporate Bonds* [15] also talks about this relationship and mention that central banks and governments know this relationship and act upon them. He also talks about the structural approach of credit risk models and the reduced form approach. Black and Scholes (1973) [16] and Merton (1974) [17] have pioneered the structural approach as Dai and Singleton (2003) [18] and Litterman and Iben (1991) [19] have pioneered the reduced form approach. In the structural approach they assume endogenously determined default risks, as the reduced form approach assumes exogenously default rates at a historical default rate or at a conditional stochastic process. Houweling and Vorst (2005) [20] have also used this type of model for pricing

and hedging of portfolios of credit securities. This piece shows an example of credit default swaps.

M. Lin, and J.C. Curtillet, Another Look at the Relation Between Credit Spreads and Interest Rates (2007) [13] state that there is a relation between spread and interest rate, but they take another look at this relationship and will break down the credit spread in 3 risk components: default risk, downgrade risk, and liquidity risk. They provide a model where the liquidity risk layer is taken into account and state that the relation is weaker than suggested in other theoretical models.

Ben Dor et al (2007) [3] further explores a measure called the DTS (Duration Times Spread) to estimate the ex-ante credit risk. But on a practical hedge-ratio for credits not much research has been published. De Backer (2006) [5] is the most relevant study we could find.

We use a genetic algorithm to improve and simplify the empirical duration methodology from De Backer.

1.5 Goal

The goal is to come up with the best possible hedge against the interest rate risk. In the study of De Backer an MS-Table was used to calculate the Empirical Duration. This MS-Table is very important for the outcome of the study though it was not optimized. In order to make a good step in finding the optimal hedge against the interest rate risk we decided to find an MS-Table which is able to capture the interest rate risk even better.

To find the best MS-Table we start to be able to determine a fitness score to rate one MS-Table. A certain MS-Table will result in an Empirical Duration-Table (ED-Table) and we would like to find the best ED-Table. The best ED-Table would be one, when used for hedging purposes, which would result in the minimum unwanted performance because of interest rate risk. We designed a method to calculate the fitness of an ED-Table based on our needs. This score would directly help us with finding the best possible ED-Table by simply finding the ED-Table with the lowest fitness score. The fitness score measures the size of the total unwanted performance made within a fixed period for a big group of bonds assuming that the ED-Table was used to implement a hedge against the interest rate risk.

We continue with testing different MS-Tables and calculate their fitness. The MS-Table that has the lowest fitness is the preferred table.

It is also important that an MS-Table is stable through time which increases our trust in the robustness of the solution and its stability. For this we take a detailed look at the first table and its behavior. By using many periods in history and because of the generic split up of the bond universe the risk of coming to an over fit solution is minimized.

When we have found the best historical MS-Table we try to draw conclusions from this table.

1.6 Research Questions

This thesis focuses on the following questions:

1. What is the best possible hedge ratio matrix for a portfolio of corporate bonds and how can generate this table?

Follow up questions are:

2. What is the best Maturity Spread-Table to use in this hedging process?

In order to design better' MS-Tables we have to define what we see as the best MS-Table and what our exact goals are in the hedging process.

1.7 Research Setup

In order to find the best possible hedge-ratio-matrix we will start with taking a close look at the stability of the empirical hedge-ratio through time using just one MS-Table. We would like to learn how stable those results are and if they would have been the same in other periods in the past. If we would find stable results of hedge ratios through time with one MS-Table or split up of the universe, this would give us more comfort in generating many MS-Tables and use them for hedging purposes. This stability through time for one table is important because we would like to prevent to come up with a final table that is over fit and only gives good results for one period in the history but not for others. To prevent to come up with a solution that is over-fit we investigated 180-periods instead of one. Using this approach, there is more risk that the table is too general than too specific as the economic environment can change through time. Knowing the historical relations is important for several reasons.

		Spread level (basis points)					
		0 - 50	50 - 100	100 - 150	150 - 250	250 - 400	400+
Maturity (years)	0 - 3	1*	2	3	4	5	6
	3 - 5	7	8	9	10	11	12
	5 - 7	13	14	15	16	17	18
	7 - 10	19	20	21	22	23	24
	10+	25	26	27	28	29	30

Table 1.2: Maturity/Spread buckets. This table shows the Maturity-Spread split up of the bond universe. Within this work this is called an MS-Table. *This is the first bucket. The table contains.30 buckets in total.

In Table 1.2 we can see that there are 5 maturity buckets and 6 spread buckets and the universe is split up in 30 different groups of bonds.

We used a proprietary database with Lehman data in a period of 190 months (1988 till 2004) and started at 180 spots. This resulted in 180 times in a hedge-ratio for each bucket, resulting in (180 periods *30 buckets) 5400 hedge-ratios.

Following on this first step, we created a measure to determine the quality of the above MS-Table in a quantitative number, for which we created the fitness. After creating this fitness measure our goal is find the MS-Table with a better, a lower fitness. We looking for the table with the lowest fitness, so the table that would result in the smallest amount of performance based on interest rate movements.

The fitness-function defines the quality and performance of an MS-Table. A certain MS-Table will result in an Empirical Duration-Table (ED-Table) and we would like to find the best ED-Table. If this ED-Table would have been used for hedging purposes, our goal is that the hedges would have resulted in the minimum unwanted exposure to interest rate risk.

In order to come up with better MS Tables we used an iterative approach where we evaluated the fitness of 200 different tables. We will also create a Genetic Algorithm that creates is aimed to generate different tables with a low fitness in a directional way in order to find the best possible MS Table.

We will use bond data from 1988 till 2004, a total of 190 months are available with a total of 25,000 bonds. The available variables are: the price of the bond (P), the total return (TR), the option adjusted spread (OAS), the option adjusted duration (OAD), the maturity date (M), the yield (y), yield to worst (ytw), sector and the rating of the bond. As we use a lot of jargon in finance, a glossary is included in the end. Maturity date is the date on which a debt becomes due for payment . Spread level is the extra yield a bond gives above a government bond with the same maturity date. A larger OAS implies a greater return for greater risks.

1.8 Outline

Within Chapter 2 we will start with an introduction to Corporate Bonds and explain the relationship between interest rate changes and spread changes and how this could be hedged. In Chapter 3 we give a short introduction to Genetic algorithms. In Chapter 4 we will describe the used econometric model to determine the empirical duration. In Chapter 5 we will explain the Research Setup and in Chapter 6 we will discuss the Results and write down our main conclusions. In Chapter 7 we will discuss possible future research.

Chapter 2 - Corporate Bonds

2.1.1 Short introduction to Bonds

Bonds are issued by public authorities (government), credit institutions, companies and supranational institutions in the primary markets. When you buy a bond, you are lending money to the issuer. The bond is a legal promise to pay you interest for the use of your money and to repay you the original amount you paid for the bond (the principal) at the maturity date.

The price of a fixed income instrument is the present value of the future cash flows. The future cash flows are the coupon payments and the repayment of the principal at maturity. In calculating the present value of a fixed income instrument you will need to know future interest rates. This means that interest rates in the future will affect the price of a bond today. If the interest rate curve changes, the predicted future cash flows will change and this will affect the present value and the price of all bonds. The formula to calculate the price of a bond:

$$\text{Bond Price} = \frac{C}{(1+i)} + \frac{C}{(1+i)^2} + \dots + \frac{C}{(1+i)^n} + \frac{M}{(1+i)^n}$$

C = coupon payment

n = number of payments

i = interest rate, or required yield

M = value at maturity, or par value

The following figure shows the Present Value of an Ordinary Annuity:

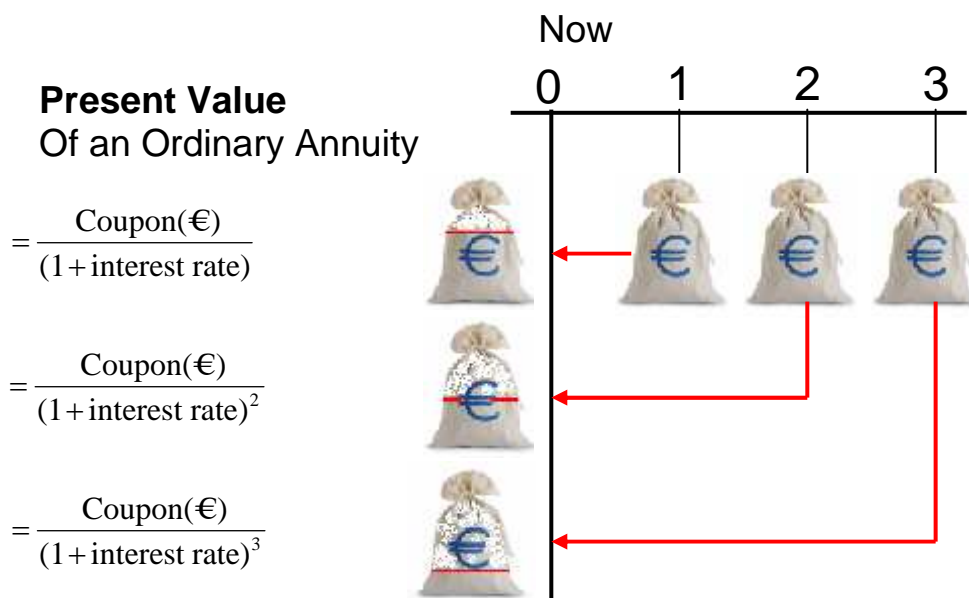


Figure 2.1 - Present Value of an Ordinary Annuity

The sum of the above equals the present value of an annuity formula:

$$PV = \frac{C}{i} \times [1 - (1+i)^{-n}]$$

Where:

PV = Present Value

C = coupon payment

i = Interest rate

n = Number of periods

If you buy a new bond and plan to keep it to maturity, changing prices, interest rates, and yields do not affect you, unless the bond is called. Callable bonds are bonds where the buyer has the right to buy the bond back at a specific price. This is an option and is not an obligation. The price investors are actually willing to pay for the bond may fluctuate through time. This price may fluctuate because of several reasons. Although there is a legal promise to pay the bond holder, it could happen that the issuer does not want to or is not able to (for instance in the case of bankruptcy) meet those legal conditions. In general, changes in the perceived probability of a different pay off structure than promised in the contract of the bond will be the trigger for price changes. The pay-off structure of a bond will be different when the company goes into default or when companies do not comply with the rules in the contract.

Interest payments are guaranteed and the principal is safe as long as you hold the security to maturity, the time at which the government or company agrees to pay back the principal. However, if you sell the security before maturity, you risk losing some of the principal if interest rates have risen or the risk climate changed.

2.1.2 Factors that Affect the Price

There are various types of risk associated with bonds. Some examples of those types of risk are: interest rate risk, default risk, liquidity risk, credit risk, and exchange rate risk.

The price investors are willing to pay for a bond is influenced by many risk factors. The price is affected largely by prevailing interest rates, because this usually has the biggest impact on the future cash flows.

Another important factor is the default risk. When a company that issued a bond goes into default, you will not receive all the coupon payments and the principal. The bigger the perceived probability that a company defaults the lower the price of a Bond will be. More generally if the payments that are promised in the contract are more uncertain or unstable the price of a bond will much likely be lower.

The macro economic environment creates a certain level of risk perception and this influences the amount of risk people are willing to take. The average level of risk perception influences the requested yield on a corporate bond. There is a causal relation between the current state of the economy and the interest rate, because interest rates can and are actively changed by central banks who use the current state of the economy as an input to determine their overnight policy interest rates. This process results in a negative correlation between government bond yield and spread. The reasons behind this will be explained in detail later.

2.1.3 Price and Interest Rates Typically Move in Opposite Directions

When Interest Rates move up, prices of Bonds go down. As a bond's price increases, yield decreases. The following figures show this relationship:

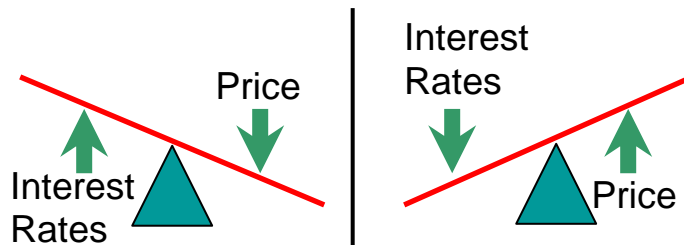


Figure 2.2 - Relation between Interest Rates and Price

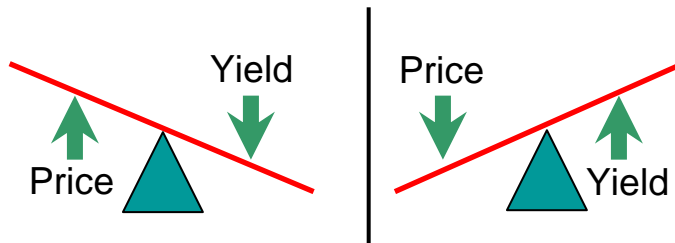


Figure 2.3 - Relation between Price and Yield

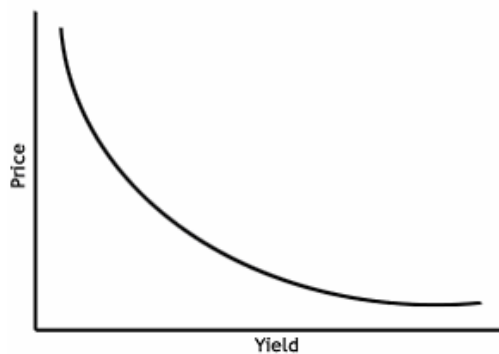


Figure 2.4 - Price-Yield Graph.

2.1.4 Bond pricing and Excess Return

In credits, bonds are usually priced using credit spread. Spread is the spread between the conventional yield on the security and the same interpolated point on the government yield curve based on the bond's maturity. Spread is usually measured in basis points. One percent equals 100 basis points.

Duration, also known as Macaulay duration is the weighted average maturity of the security's cash flows, where the present values of the cash flows serve as the weights. The greater the duration of a security, the greater its percentage price change given change in yield.

Excess Return is a security's return minus the return from a risk less security during the same time period. The "risk-free rate of return" is generally the rate of return of a national government issue.

In the fixed income market a credit portfolio manager is given the task to invest in credit bonds (credits). The goal is to buy the best bonds that eventually do not default or buy the bonds where the spread will tighten more than the average of the benchmark. This way outperformance will be made. Although interest rate changes generate the most profits/losses made in Fixed Income investments, this part of the performance is not used when calculating the performance of a Fixed Income portfolio manager. The reason is that this part of the performance cannot be influenced directly. Excess return is a measure that measures performance without the influence of interest rate changes.

The changes in interest rates are filtered out by using the Excess Return as a measure on how good a portfolio manager managed a portfolio. The Excess Return of the portfolio is compared to the Excess return of the benchmark. If the weighted sum of all excess returns of the portfolio is higher than the benchmark then the portfolio manager made outperformance. The performance made by the portfolio manager is better than the performance of the benchmark. In this scenario the portfolio manager added value to the portfolio.

The problem is that Excess return assumes that spreads do not move when interest rates moves. Empirical research showed that this is an incorrect assumption. This results in an undesirable risk factor in the portfolio: interest rate risk.

When interest rates rise the price of the bond will go down. Though you will earn on your hedge based on the duration, which on general is more than what you lose because spreads tend to tighten when government bond yields go up resulting in outperformance without your direct influence. In the opposite case when interest rates go down, we will see that the hedge based on duration does not immunize all the losses made because of this moment. This is an undesirable situation where you would like to have a smaller hedge. In both scenarios the hedge is too big, to solve this problem we would like to use Empirical duration instead of the duration. This turns out to be a better measure of reducing the interest rate risk.

2.1.5 Separation of Responsibilities in practice

Portfolio managers in the Investment Grade market, bonds with a rating bigger than 'BB', should have a clear view on which bonds to buy and which bonds to sell based on other arguments than the interest rate movements. A portfolio manager could make the active decision to go overweight in one sector versus another sector without making an active call on interest rate movements. Both are independent decisions and only one is taken. The active decision is the sector bet and if interest rate exposure is created when implementing this decision this is an undesired exposure. As interest rate movements can generate an undesirable return the portfolio managers would like to hedge out that risk.

As we said, there are various types of risks (or performance drivers) involved when you buy bonds. In this thesis we will look at the interest rate risk of bonds. In the coming chapters we go into detail about what bonds are, what kind of different bonds there are and what risks are associated with them. We will also explain why we would like to reduce some of those risks and how we can do this.

The Excess return of a bond is the return over duration neutral treasuries. This is a security's return minus the return from a no-risk security during the same period. At the moment a no-risk security is a treasury bond. The OAD is used in this calculation and the Excess Return of a security is based on a theoretical calculation. There is a mismatch between this calculated excess return and the 'real' or empirical return based on interest rate changes for a bond with a certain OAD. Meaning, a portfolio manager is still exposed to an undesirable risk factor on which he has no influence. The average OAD of a portfolio is not always an active bet and usually the result of issuer selection and about idiosyncratic parameters about a company which the portfolio managers likes or dislikes/ The duration exposure which is the result is an undesirable one. Because the excess return calculation contains fundamental or structural mistakes, this can work either in favor or against this Portfolio manager. When it works against him, he would like to be able to undo this with an active position or hedge in order to have a more stable result and on which he has more influence. As it is not always known when those interest rate movements will work in favor or against you, the decision is made to hedge out the risk as good as possible. The sub-optimal approach in calculation Excess returns using the OAD is true for both the portfolio and benchmark and therefore has no direct impact on the performance of a fund which is close to its benchmark. When the differences are bigger the results are also of more influence. A portfolio manager who wants to be hedged against the interest rate risk should only hedge the difference in interest rate risk between the portfolio and the benchmark.

2.2 General Introduction to bonds

A bond is a certificate of debt. There are *government bonds* and *corporate bonds*. In this thesis we will concentrate on *corporate bonds*. When you buy a bond, you are lending money to the government or to a corporate, respectively. The party that borrows the amount of money is called the *issuer*. In return for the loan the issuer will pay a

specified rate of interest (the *coupon rate*) during the life of the bond, and a specified amount (the *principal* or the *par value*) at a specific time in the future, the *maturity date*. Corporate bonds are part of a class of bonds that are called *credits*. Credits are by definition all bonds that are not issued by the government in the home currency.

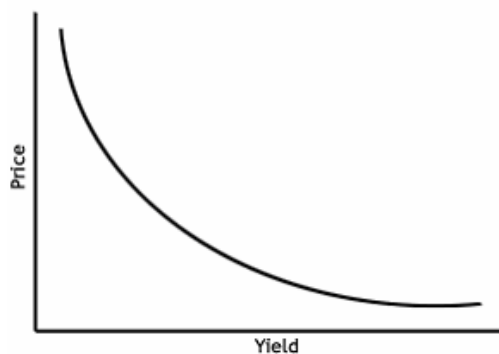
A key feature of any bond is its *term-to-maturity*, the number of years during which the borrower has promised to meet the conditions of the debt. A bond's term-to-maturity is the date on which the debt will cease and the borrower will redeem the issue by paying the face value, or principal. In practice, the words *maturity*, *term*, and *term-to-maturity* are used interchangeably to refer to the number of years remaining in the life of a bond. Technically, however, *maturity* denotes the date the bond will be redeemed, and either *term* or *term-to-maturity* denotes the number of years until that date.

A bond's coupon is the periodic interest payment made to owners during the life of the bond. The coupon is always cited, along with the maturity, in any quotation of a bond's price.

Let's assume the interest rate is 5%. If you have 100 and put this on a bank account, this will be worth 105 after one year. Let's assume there is a bond with a par-value of 100, a coupon of 5% and a term-to maturity of one year. The price of this bond should be exactly 100, because you are indifferent in putting your money on a savings account or buying that bond for 100. This is true if there is no credit risk.

The price of a financial instrument is equal to the present value of the expected future value. Therefore, when the coupon rate, the principal and the time to maturity are known, it is possible to compute the price of the bond. The cash flows are the periodic interest payments to the maturity date and the par value that will be received at maturity.

The *yield* (or *internal rate of return*) of a bond is the percentage return on an investment at an annual rate. The higher the yield, the lower the price of the bond will be. There exists a convex relationship between the required yield and the bond price [2, page 91].



Graph 2.2: Convex relationship between yield and bond price

Now we are able to write the price of a bond as follows. Let $P = P(t, y(t))$ be the

price of the bond as function of time t and yield y , where the yield $y = y(t)$ is itself a function of time. The price of the bond at time $t = 0$ and yield $y(t) = y$ can be written as

$$P(0, y) = \sum_i \frac{CF_i}{(1+y)^{T_i}}, \quad (2.1)$$

where CF_i is cash flow i and T_i is the time of the i^{th} coupon payment.

Using (2.1) we are able to compute the price of a bond when we know the cash flows and the required yield.

The required yield is an important factor for the price of a bond. When the required yield goes up, the price of a bond goes down (Figure 2.1). For a corporate bond it is possible that the required yield is high. This might be true because there is an increased chance this company goes bankrupt and will not be able to meet the conditions of the dept. Therefore the buyers are requiring a higher yield and the price of the bond will drop.

2.3 Risks and interest rate risk of bonds

There are many risks associated with investing in fixed income securities. The different types of risk that an investor in fixed income securities is exposed to are as follows [2, page 18]:

- *Market or interest-rate risk.*
- Reinvestment risk.
- Timing, or call, risk.
- *Credit, or default, risk.*
- Yield-curve, or maturity, risk.
- Inflation, or purchasing power, risk.
- Marketability, or liquidity, risk.
- Exchange rate, or currency, risk.
- Volatility risk.
- Political or legal, risk.
- Event risk.
- Sector risk.

This list gives a good idea of the risks an investor in fixed income securities is exposed to. We are going to take a closer look at the *market or interest-rate risk*. There is a relationship between the interest rate and the price of a bond. The price of a typical fixed income security moves in the opposite direction of the change in interest rates. As interest rates rise (fall), the price of a fixed income security will fall (rise). To see that this holds true, suppose you have a bond; it will have a certain coupon interest. When the government bond yield rises, it becomes less attractive to buy this bond because in this

new situation you can purchase other bonds with higher coupon rates. Therefore, the price of the bond will decline.

For an investor who plans to hold a fixed income security to maturity, the change in its price before maturity is not of concern; however, for an investor who may have to sell the fixed income security before maturity date, an increase in interest rates will mean the realization of a capital loss. The risk is referred to as *market risk*, or *interest-rate risk*, which is by far the biggest risk faced by an investor in the fixed income market [7].

Because there is a relationship between the interest rate and the price of a bond there is a risk involved when you have bonds. If interest rates change, the value of the bonds you have will change, so you could make losses or profit when this happens. It is possible to protect yourself against those undesirable losses or profits. This is called hedging. Hedging is a strategy designed to minimize exposure to an undesirable risk, in this case interest rate risk, while still allowing to profit from an investment activity [4].

A characteristic of the interest rate risk is that the interest rate risk rises with maturity. This can be explained using the definition of duration. When the maturity is greater, the duration will also be greater, and by definition this means that there will be larger changes in price for the same change in government bond yield. So there is a higher interest rate risk.

A corporate bond also is exposed to the interest rate risk. The important difference between government bonds and corporate bonds is the fact that a corporate has an extra source of risk: *credit risk*. Credit risk is the risk that an issuer will be unable to make interest or principal payments when they are due; this is called *default*. A firm could go bankrupt; in such a case the bond holder will not get back the invested amount of money.

Because of this extra risk the (required) yields for corporate bonds will be higher than they are for government bonds. So, the yield for a corporate consists of the government bond yield plus an additional interest, which is called the *(credit)spread*. Besides the probability of default, the credit spread also depends on a second factor, this is the *recovery rate*. Usually you can get a certain percentage of the par value back in case of default; this is the recovery rate.

The result is a higher yield for corporate bonds compared to government bond. The yield will be larger because of the larger risk that you are exposed to.

The probability that a firm will default is hard to measure. It's a very important factor for investors in bonds because this probability influences the price of a bond. Because of this, we would like to have a credit rating for a company. The probability that a firm will default can be expressed by the *rating* of the firm. The ratings are determined by rating agencies, the most important agencies are Moody's and Standard&Poor's. Firms can be separated into two categories: *investment grade* and *high yield*. Investment grade firms are very creditworthy and have ratings AAA (*Aaa*), AA (*Aa*), A or BBB (*Baa*) for S&P (Moody's). High yield firms are less creditworthy (and therefore investing in

high yield bonds means that you also take a higher risk) and have ratings *BB (Ba)*, *B*, *CCC (Caa)*, *CC (Ca)* and *C*. A rating of *D* means default. The lower the rating the bigger the chance a firm will default.

There exists an important link between the government bond yield and the spread: there exists a negative correlation between government bond yields and credit spreads. This means that when government bond yield rises, the spread tends to fall and when the government bond yield declines, the spread tends to rise.

An explanation for this fact can be given as follows. When a country is going into economic decline then the following situation may occur:

1. The government bond yield will be low.
2. The spread will be high: in poor times the credit risk will be higher (the probability that a firm will default is higher), so to still buy corporate bonds a higher yield will be 'demanded' to compensate. So the spread will be higher.

Putting these two interest rates together we can see that a lower government bond yield will be (partly) compensated by a higher spread, making the net change in yield for a corporate smaller. The opposite also holds: when the economy recovers, the government bond yield will rise, but the spreads are likely to fall at the same time.

For government bonds we have already seen that when the government bond yield rises, the prices of bonds will fall. Now in the case of corporate bonds the situation will be different. Because of the above effect, when the government bond yield rises, the net change in yield for a corporate bond will be smaller. Therefore, the change in price will also be smaller.

From this point on we will use the following notation: by *interest rate duration* we mean the price sensitivity of the bond to changes in the *government bond yield*.

Now immediately we can see the effect the above will have on interest rate duration: the duration for government bonds will be larger than for corporate bonds.

A distinction can be made between investment grade bonds and high yield bonds. Because investment grade bonds are bonds issued by (very) creditworthy firms they will have a very low spread. When the government bond yield rises, the change in spread will be very small and so the net change in yield will be nearly equal to the change in government bond yield, and so the interest rate duration of investment grade bonds will also be almost equal to the interest rate duration for government bonds.

Now we turn to high yield bonds. The lower the rating of a firm, the higher the spread will be. Further, empirical observations have showed the following: the higher the spread, the larger the change in spread (in opposite direction) as the government bond yield changes.

We can also put this in terms of interest rate duration. The interest rate duration is negatively correlated with the spread. The lower the rating, the higher the spread, and the lower the interest rate duration tends to be. The change in spread may even compensate the government bond yield completely, in which case the interest rate duration will be

equal to zero. So for high yield bonds the interest rate risk gets smaller and smaller when the spread gets higher. This means that a change in government bond yield is less important for high yield bonds and the credit risk is larger.

An example of how the interest rate could create performance in a portfolio where the entire OAD is used to generate a hedge:

	Maturity					Empirical Duration	
	0	1	2	3	4	Using 80% of the Duration	
BM - %	20%	20%	20%	20%	20%		
Portfolio - %	20%	20%	20%	20%	30%		
Difference	0%	0%	0%	0%	10%		
Interest Rates go up 1%, price effect							
Total Outperformance	0%	0%	0%	0%	-4%		-4.00%
Hedge	0%	0%	0%	0%	4%		3.20%
Total	0%	0%	0%	0%	0%		-0.80%
2nd order Effect: Spreads tighten	0%	0%	0%	0%	0.80%		0.80%
Total Effect after Second Round Effect	0%	0%	0%	0%	0.80%		0.00%
	Maturity					Empirical Duration	
	0	1	2	3	4	Using 80% of the Duration	
BM	20%	20%	20%	20%	20%		
Portfolio	20%	20%	20%	20%	30%		
Difference	0%	0%	0%	0%	10%		
Interest Rates go down 1%, price effect							
Total Outperformance	0%	0%	0%	0%	4%		4.00%
Hedge	0%	0%	0%	0%	-4%		-3.20%
Total	0%	0%	0%	0%	0%		0.80%
2nd order Effect: Spreads tighten	0%	0%	0%	0%	-0.80%		-0.80%
Total Effect after Second Round Effect	0%	0%	0%	0%	-0.80%		0.00%

Table 2.1 - performance effect of interest rate change, including a hedge

There can be given an explanation to show why this negative correlation holds true. Consider an extreme case, suppose you expect a firm to go bankrupt in a few years. If you would decide to buy this bond, you know (expect) that you will not receive all the coupon payments and the par value. With this in mind, it seems likely that you are not willing to pay the 'theoretical' price of the bond, but instead you will demand a lower price (that will depend on the recovery rate), which will be independent of the maturity date (so all the bonds of this firm will have the same price, irrespective of the maturity). From this price you can calculate the yield, which will be high, and so the spread will be high (with this high yield you will earn your investment back faster). Now the government bond yield may fall, but considering the above, this will not have an impact on the price because you are just not willing to pay more for such a bond. So the government bond yield has no (or little) influence anymore; the credit risk has determined its price.

Also in less extreme cases the probability of default (and the recovery rate) is taken into consideration; the bond's price is affected by this, and the influence of the government bond yield declines.

As I experienced in practice when spreads go up people tend to look at the price of an instrument instead of looking and trading using the price. When more and more people look at the price instead of spread, the relation between interest rate movements and price movements of a bond compared to the related government bond disappear.

We know that a firm with low rating will have a high spread. Should we measure the credit risk of a firm according to the rating or to the spread? Research by Robeco has shown that it is better to take the spread as measure. This is because the spread gives a better impression of the creditworthiness of a firm than the rating. For example, consider a firm with a certain rating that has shown good performance for considerable time. Then it is likely that the spread has decreased, because the risk has reduced. However, it may well be possible that the rating of this firm is still the same. This makes clear that the spread is a better measure for the risk. Furthermore there are 41 different rating categories and spread is a continuous variable. Because the spread has more values it is a measure with a higher significance level. In practice the spread can change on a daily basis where the rating only changes periodically. The rating is known to almost all persons (traders) who trade in bonds. This information is also taken into account determining the price of a bond which strengthens the research that the spread is a better measure.

2.4 Modified duration

We would like to have a measure than can be used to quantify the price sensitivity of a bond. Such a measure should take into account all factors that affect the price volatility. Characteristics that determine the price volatility of a bond are the time to maturity, the yield, the coupon rate and the yield volatility. *Duration* is a measure that takes these factors into consideration.

First we will explain another (closely related) measure, the *Macaulay Duration* [10]. Bierwag et al (1983) [4] describes the practical use of the duration measure in investment management. The Macaulay Duration is defined as a weighted average time to maturity of the bond's cash flows, with weights equal to the present value of each cash flow as a percentage of the present value of all the cash flows of the bond (so, as a percentage of the price of the bond). We can write:

$$\begin{aligned}
 & \text{Macaulay Duration (in periods)} \\
 & = \\
 & \frac{t_1 * PVCF_1 + t_2 * PVCF_2 + \dots + t_n * PVCF_n}{PVTCF}
 \end{aligned} \tag{2.2}$$

where

t_i = period of cash flow i

$PVCF_i$ = present value of cash flow i

$PVTCF$ = present value of the total cash flow of the bond (= $P(0, y)$)

n = number of periods

Since we know that the present value of the total cash flow of the bond, $PVTCF$, is equal to the price of the bond, we can write formula (2.2) as

$$\text{Macaulay Duration (in periods)} = \frac{1}{P(0, y)} \sum_{i=1}^n \frac{T_i CF_i}{(1+y)^{T_i}},$$

where

y = bond yield

T_i = period of cash flow i

CF_i = cash flow i

Properties of the Macaulay Duration are

1. The Macaulay Duration is always less than its maturity (except for a zero coupon bond, in which case it is equal to the time to maturity).
2. The greater the maturity, the greater the duration.
3. The lower the coupon, the greater the duration.
4. The greater the yield, the lower the duration.

Now the *Modified Duration* is the measure defined by

$$\text{Modified Duration} = \frac{\text{Macaulay Duration}}{1+y}, \quad (2.3)$$

where y is the yield to maturity.

From these definitions we can see that the Modified Duration is determined by:

1. *The time to maturity*: the longer the time to maturity, the higher the duration. (Therefore, when you expect interest rates to fall (increase), you should hold bonds with long (short) maturities)
2. *The coupon rate*: The lower the coupon rate, the higher the duration.
3. *The yield*: the higher the yield, the lower the duration.

Now the link between bond price volatility and duration is as follows:

$$\text{Approximate percentage change in price} = - \text{Modified Duration} * \text{yield change} \quad (2.4)$$

Often Modified Duration is referred to as the percentage of change in price per *100-basis-point* change in yield. It is a measure of the price sensitivity to interest rate movements. The higher the duration, the higher the risk (because a change in market interest rate will have a larger impact on the bond's price).

To end this section, we will show that the following holds:

$$\frac{\partial P}{\partial y}(0, y_0) = -MD * P(0, y_0), \quad (2.5)$$

where MD is the modified duration.

From formula (2.1) for the bond price and from equations (2.2) and (2.3) it can be seen that the modified duration can be written as the following sum:

$$MD = \frac{\text{MacaulayDuration}}{1+y} = \frac{1}{1+y} \frac{1}{P(0, y)} \sum_i \frac{T_i CF_i}{(1+y)^{T_i}} \quad (2.6)$$

Now we can take the partial derivative with respect to y in (2.1) and by using (2.6) this results in

$$\begin{aligned} \frac{\partial P}{\partial y}(0, y) \Big|_{y=y_0} &= \frac{-1}{(1+y_0)} \sum_i \frac{T_i CF_i}{(1+y_0)^{T_i}} \\ &= \frac{P(0, y_0)}{P(0, y_0)} \frac{-1}{(1+y_0)} \sum_i \frac{T_i CF_i}{(1+y_0)^{T_i}} = -MD * P(0, y_0). \end{aligned} \quad (2.7)$$

2.5 Approximation for the percentage change in bond price

In this section we will derive an approximation for the percentage change in bond price to get a better feeling of the used model used in our examination.

Let $P = P(t, y(t))$ be the price of the bond as function of time t and yield y , where the yield $y = y(t)$ is itself a function of time. For notational convenience, denote $y(t)$ by y_t . The percentage change in bond price is given by

$$\frac{P(T, y_T) - P(0, y_0)}{P(0, y_0)} = \frac{P(T, y_T)}{P(0, y_0)} - 1. \quad (2.8)$$

We approximate $P(T, y_T)$ by a Taylor expansion around the point y_0 ,

$$P(T, y_T) = P(T, y_0) + (y_T - y_0) \frac{\partial P}{\partial y}(T, y_0). \quad (2.9)$$

Substituting (2.9) in (2.8) results in

$$\frac{P(T, y_T)}{P(0, y_0)} - 1 = \frac{P(T, y_0)}{P(0, y_0)} + \frac{1}{P(0, y_0)} (y_T - y_0) \frac{\partial P}{\partial y}(T, y_0) - 1. \quad (2.10)$$

Now we will rewrite the terms $\frac{P(T, y_0)}{P(0, y_0)}$ and $\frac{\partial P}{\partial y}(T, y_0)$ in this expression.

For the first term, we know that when the interest rate remains unchanged during the period between $t = 0$ and $t = T$, the bond price will increase by the amount of interest that is earned over the period, so the term $P(T, y_0)$ can be written as

$$P(T, y_0) = P(0, y_0)(1 + y_0)^T. \quad (2.11)$$

For the second term, we take the partial derivative in (2.11) with respect to y :

$$\frac{\partial P}{\partial y}(T, y_0) = \frac{\partial P}{\partial y}(0, y_0)(1 + y_0)^T + \frac{P(0, y_0)T(1 + y_0)^T}{(1 + y_0)}.$$

We can go one step further by using the fact that $\frac{\partial P}{\partial y}(0, y_0) = -MD * P(0, y_0)$, where MD is the modified duration. Then we have

$$\frac{\partial P}{\partial y}(T, y_0) = -MD * P(0, y_0)(1 + y_0)^T + \frac{P(0, y_0)T(1 + y_0)^T}{(1 + y_0)}. \quad (2.12)$$

Now we are able to substitute (2.11) and (2.12) into (2.10) to arrive at

$$\begin{aligned} \frac{P(T, y_T)}{P(0, y_0)} - 1 &= (1 + y_0)^T + (y_T - y_0) \left\{ -MD(1 + y_0)^T + \frac{T(1 + y_0)^T}{(1 + y_0)} \right\} - 1 \\ &= (1 + y_0)^T \left\{ 1 - MD(y_T - y_0) + \frac{T(y_T - y_0)}{(1 + y_0)} \right\} - 1. \end{aligned}$$

Because we have periods of months, we have $T = \frac{1}{12}$. A commonly used approximation is $(1 + y_0)^{\frac{1}{12}} \approx 1 + \frac{1}{12} y_0$ and this gives us

$$\begin{aligned} \frac{P(T, y_T)}{P(0, y_0)} - 1 \Big|_{T=\frac{1}{12}} &= \left(1 + \frac{1}{12} y_0 \right) \left\{ 1 - MD \left(y_{\frac{1}{12}} - y_0 \right) + \frac{\frac{1}{12} \left(y_{\frac{1}{12}} - y_0 \right)}{(1 + y_0)} \right\} - 1 \\ &= \frac{1}{12} y_0 - MD \left(y_{\frac{1}{12}} - y_0 \right) \left(1 + \frac{1}{12} y_0 \right) + \frac{1}{12} \left(y_{\frac{1}{12}} - y_0 \right) \frac{1}{(1 + y_0)} \left(1 + \frac{1}{12} y_0 \right). \end{aligned}$$

Finally, using the approximation $1 + y_0 \approx 1$ (and therefore also $1 + \frac{1}{12} y_0 \approx 1$) we obtain the result:

$$\frac{P(T, y_T) - P(0, y_0)}{P(0, y_0)} \Big|_{T=\frac{1}{12}} = \frac{1}{12} y_0 - \left(MD - \frac{1}{12} \right) \left(y_{\frac{1}{12}} - y_0 \right). \quad (2.13)$$

The result obtained in (2.13) is an approximation for the percentage change in price of a bond.

2.6 Importance of a duration neutral portfolio

For bonds with different characteristics, different relationships exist with the interest rates. Therefore the bonds can be separated into different buckets (separated by certain characteristics) and the effect of interest rate changes for the chosen bucket can be measured. For this bucket a strong relationship can be found between the interest rate changes and the total return of corporate bonds. For another bucket a different relationship could be found.

It is important to know the interest rate risk a bond carries because portfolio managers might want to minimize the interest rate risk of their portfolio. They might have expertise in the profits or losses they can make by buying specific bonds in specific sectors, but they might not want to be exposed to the interest rate risk. You could compare this with buying a stock of a company that trades in USD without the desire to be exposed to USD, in this situation you would want to hedge this exchange rate risk.

Portfolio managers sometimes prefer to have a *duration-neutral* portfolio. They don't want to take interest rate risk but they do want to take other risks. In order to reduce the interest rate risk they could use the Empirical Duration (variable for interest rate sensitivity) to hedge this risk. In practice the Empirical Duration is not equal to the real Duration (interest rate sensitivity). The real duration differs for different spread levels and maturity dates. The larger the OAS (Option Adjusted Spread) the greater the return will be for greater risks.

Another strong argument for hedging is that it reduces variance over the course of years. It is good to perform every year and this is preferred above a very good result one year and a bad result another year because of the interest rate influence.

In practice you will only hedge a fraction of the interest rate risk because a portfolio usually follows a benchmark and only the difference between the portfolio and the benchmark will be hedged.

When the interest rate risk of a corporate bond has been determined, it can be used for two purposes. The interest rate risk can be used for *hedging purposes* and also for *risk measurement*.

We can hedge against interest rate risk with futures on government bonds. When the interest rate risk of a corporate bond is known, it can be used to hedge the interest rate risk of a portfolio that contains several corporate bonds. This can be done by using the *hedge-ratio*. In the current practice there is a difference between investment grade bonds and high yield bonds. Investment grade bonds are actually treated as government bonds, i.e. the modified duration for these bonds is assumed to be equal to the interest rate duration. For high yield bonds there was no hedge applied at all until recently. This means that it is assumed that there is no interest rate risk for these bonds. Currently this is not the case anymore; now half of the modified duration is used for hedging purposes.

The hedge-ratio can be viewed as a measure of the influence of the interest rate risk on the price change of a bond. It can be used to determine the part of the analytical duration (OAD) that should be taken into account in hedging the risk. The analytical duration is the bond price sensitivity to changes in the total yield of the bond (instead of only the government bond yield) and it can be calculated from the present value of the cash flows of the bond. The hedge-ratio takes values between zero and one. A hedge-ratio of 1 means that you should take the full analytical duration (i.e., there is no credit risk, only interest rate risk), a hedge-ratio of 0 means that you do not have to take any of the analytical duration into consideration, i.e., that the government bond yield changes have

no influence, in this situation there would be no interest rate risk. The hedge-ratio can be calculated by dividing the empirical duration (interest rate duration) by the analytical duration.

In the study of De Backer a significant relation between the hedge-ratio and the spread level was found. The hedge-ratio was high for low spread levels and low for high spread levels.

There are several models to calculate the hedge-ratio, but we chose the bucket approach based model. This model gives a different hedge-ratio for each bucket. We also chose this model because it gave the highest R-squared compared to other models which means that we can have confidence in quality of the regression performed in this model. This is important if you want to compare the outcomes of successive regressions.

We are looking at the stability of this hedge-ratio through time and analyze the motion of this hedge-ratio in order to get a better feeling for the optimal update frequency of the hedge-ratio and the influence of using a different length of history in determining the hedge-ratio. We will also try to determine the best MS-Table which will lead to the best possible hedges.

Chapter 3 - Genetic Algorithms

Outline

In our goal to find the best possible hedge ratio matrix we decided to find out what would have been the best possible hedge ratio matrix in the past. As there is not one hedge-ratio but the hedge ratio is different for bonds with different maturity and spread characteristics the universe has to be split up in clusters. For each cluster we will determine the hedge ratio based on the model presented in Chapter 4 and calculate the accuracy of this implemented hedge. A perfect hedge would cover 100% of all performances due to interest rate changes. The sum of all performances for all different clusters is something we would like to minimize.

Information about which MS-Table performed best historically will give us valuable insight in how we should cluster the bonds today and what MS-Table to use in our current process.

Because we know the percentage of market value (the size of each cluster), the impact on performance of each cluster (a higher duration has more impact than a low duration cluster) we are able to calculate the sum of 'unwanted' performance using a certain MS-Table.

A possible way to find the optimal MS-Table would be to manually try an many, or all, different possible MS-Tables and calculate the historical result. For this we would have to make millions of different tables and this would take an enormous amount of time. Because each iteration would consume several minutes, this approach would consume an too much time to ever finish calculating. Many MS-tables that look much alike would be tested and this is would generate a lot of redundant calculations because in the end we would like to be able to generalize.

Taking those arguments into account, we tried to find an algorithm that would be flexible and adjust based on the results of a previous run. We designed a method use a gene to define an MS-Table. Because a Genetic algorithm uses a fitness function or a pain function attached to the gene and this feedback is used to generate a new gene (a new MS-table) this would significantly reduce the time to do this research. Using a genetic algorithm we can efficiently go through a very big search-space and solve a problem that would normally take many more calculation rounds and it could even be that without simplifications we could never find a solution at all. This approach is innovative and might be new in finance but using this or similar technology we can find high-level parameters of a financial models in a flexible and fast way. This high-level fine tuning and optimizing of models can very well have a high impact on the total performance of a model. Therefore this is an interesting field of investigation.

With in this chapter we would like to explain what a Genetic Algorithm is and how we have implemented the Genetic Algorithm for our purposes and we will use this technique in the Research Setup later.

3.1 What is a Genetic Algorithm?

A genetic algorithm is a search technique used in computing to find the exact or approximate solution to optimization and search problems. A genetic algorithm can be used for both constrained and unconstrained optimization problems that are based on natural selection, the process that drives biological evolution. The genetic algorithm repeatedly modifies a population of individual solutions. At each step, the genetic algorithm selects individuals at random from the current population to be parents and uses them to produce the children for the next generation. Over successive generations, the population "evolves" toward an optimal solution. You can apply the genetic algorithm to solve a variety of optimization problems that are not well suited for standard optimization algorithms, including problems in which the objective function is discontinuous, nondifferentiable, stochastic, or highly nonlinear. More details about Genetic Algorithms are found in [11].

The main steps within a genetic algorithm are the following: Selection, Reproduction (Crossover and Mutation) and Termination.

Selection

During each successive generation, a proportion of the existing population is selected to breed a new generation. Individual solutions are selected through a fitness-based process, where fitter solutions are typically more likely to be selected. Certain selection methods rate the fitness of each solution and preferentially select the best solutions. Other methods rate only a random sample of the population, as this process may be very time-consuming. Most functions are stochastic and designed so that a small proportion of less fit solutions are selected. This helps keep the diversity of the population large, preventing premature convergence on poor solutions.

Reproduction: Crossover and Mutation

The next step is to generate a second generation population of solutions from those selected through genetic operators: crossover and/or mutation. For each new solution to be produced, a pair of 'parent' solutions is selected for breeding from the pool selected previously. By producing a 'child' solution using the above methods of crossover and mutation, a new solution is created which typically shares many of the characteristics of its parents. New parents are selected for each new child, and the process continues until a new population of solutions of appropriate size is generated. These processes ultimately result in the next generation population of chromosomes that is different from the initial generation. Generally the average fitness will have increased by this procedure for the population, since only the best organisms from the first generation are selected for breeding, along with a small proportion of less fit solutions.

Termination

The generation process is repeated until a termination condition has been reached. Common terminating conditions are that the solution found satisfies the minimum criteria or a maximum number of generations. Other stop criteria are that a time limit is reached, a fitness limit is exceeded, or the change of the total fitness is less than a certain number.

A genetic algorithm usually starts with an *initialization* phase where many individual solutions are randomly generated to form an initial population. The population size depends on the nature of the problem. Occasionally the solutions may be 'seeded' in areas where optimal solutions are likely to be found.

In order to be able to use a genetic algorithm the first requirement is that there is a *genetic representation* of the solution. The second requirement is that there needs to be a *fitness function* to evaluate the solution.

We have chosen to use the Genetic Algorithm Toolbox which in Matlab and in the next paragraph some terminology will be explained.

3.2 Genetic Algorithm Terminology

The **Fitness Function** is the function you want to optimize. For standard optimization algorithms, this is known as the objective function. Within a research it is very important to have a well defined Fitness Function because it defines your optimum.

An **Individual** is any point to which you can apply the fitness function. The value of the fitness function for an individual is its score.

A **population** is an array of individuals. At each iteration, the genetic algorithm performs a series of computations on the current population to produce a new population. Each successive population is called a new generation.

Diversity refers to the average distance between individuals in a population. A population has high diversity if the average distance is large; otherwise it has low diversity.

To create the next generation, the genetic algorithm selects certain individuals in the current population, called **parents**, and uses them to create individuals in the next generation, called **children**. Typically, the algorithm is more likely to select parents that have better fitness values.

3.3 Outline of the Algorithm

The following outline summarizes how the genetic algorithm works:

1. The algorithm begins by creating a random initial population
2. The algorithm then creates a sequence of new populations. At each step, the algorithm uses the individuals in the current generation to create the next population. To create the new population, the algorithm performs the following steps:
 - a. Scores each member of the current population by computing its fitness value.
 - b. Scales the raw fitness scores to convert them into a more usable range of values
 - c. Select members, called parents, based on their fitness.
 - d. Some of the individuals in the current population that have lower fitness are chose as *elite*. These elite individuals are passed to the next population.
 - e. Produces children from parents. Children are produced either by making random changes to a single parent-mutation-or by combining the vector entries of a pair of parents-crossover.
 - f. Replaces the current population with the children to form the next generation.
3. The algorithm stops when one of the stopping criteria met.
 - a. Generations - The algorithm stops when the number of generations reaches the maximum value of Generations
 - b. Time limit - The algorithm stops after running for an amount of time in seconds equal to Time limit.
 - c. Fitness limit - The algorithm stops when the value of the fitness function for the best point in the current population is less than or equal to Fitness limit.
 - d. Stall time limit - The algorithm stops if there is no improvement in the objective function during an interval of time in seconds equal to Stall time limit.
 - e. Function Tolerance - The algorithm runs until the weighted average change in the fitness function value is less than Function tolerance.

3.4 Genetic Representation

In order to be able to use a genetic algorithm the first requirement is that there is a *genetic representation* of the solution. Our initial MS-Table looks the following (Table 3.1):

		Spread level (basis points)					
		0 - 50	50 - 100	100 - 150	150 - 250	250 - 400	400+
Maturity (years)	0 - 3	1*	2	3	4	5	6
	3 - 5	7	8	9	10	11	12
	5 - 7	13	14	15	16	17	18
	7 - 10	19	20	21	22	23	24
	10+	25	26	27	28	29	30

Table 3.1 - Maturity-Spread Table *This is the first bucket. The table contains.30 buckets in total.

In order to be able to use a genetic algorithm to test different MS-Tables we needed a genetic representation of an MS-Table combined with a fitness function.

We had to quantify the boundaries that were used in table Table 3.2 in order to be able to create a genetic representation of this table. We quantified the boundaries and replaced the different spread and maturity intervals with s1 till s5 and m1 till m4.

		Spread level (basis points)					
		0 - s1	s1 - s2	s2 - s3	s3 - s4	s4 - s5	s5+
Maturity (years)	0 - m1	1*	2	3	4	5	6
	m1 - m2	7	8	9	10	11	12
	m2 - m3	13	14	15	16	17	18
	m3 - m4	19	20	21	22	23	24
	m4+	25	26	27	28	29	30

Table 3.2 - Maturity-Spread Table with dynamic intervals for the Maturity and for the Spread. m1 till m4 are dynamic boundaries for the maturity and s1 till s4 are dynamic boundaries for the spread.

The first interval in Table 3.2 within spread is 0 - s1. The second interval in spread is s1 - s2 and so forth. For the maturity we see the first interval 0 - m1, and the second interval to be m1-m2. The size of the first spread bucket is s1 - 0 and the size of the second spread bucket is s2 - s1. The size of a bucket always has to be equal or bigger than 0.

The first idea was to represent the gene with values of the boundaries. So the representation of the entire table within a gene would look the following:

[s1, s2, s3, s4, s5, m1, m2, m3, m4]

For Table 3.1 this would mean that the gene would look the following genetic representation of the table:

[50, 100, 150, 250, 400, 3, 5, 7, 10]

After providing this gene to the code, the code would create the MS-Table. Though there were some problems with this representation because when in the reproduction process a gene was created where for instance s2 was smaller than s1 the code wouldn't be able to create a good table.

To avoid this problem we came up with a different representation of the problem, we defined the elements of the gene as the size of a bucket. This means that the first number provided was s1 - 0, and the second number provided was s2-s1. When a gene would contain negative numbers we used 0 as the size of a bucket. This resulted in the following representation:

[size s1, size s2, size s3, size s4, size s5, size m1, size m2, size m3, size m4]

By definition the last spread buckets was after the sum of al the previous deltas till infinity and the same for the maturity.

To new definition of Table 3.1 we would use the following gene:

[50, 50, 50, 100, 150, 3, 2, 2, 3]

The first split up will be from 0 to 50, the second to 50+50 so till 100, and then 50 more so till 150 then 100 more till 250 and the last 150 makes it 400 for the last boundary and the last bucket contains values with a spread above 400.

3.5 Fitness Function

The fitness function should help us to find the best possible MS-Table. An MS-Table will generate an Empirical Duration-Table (ED-Table) and we would have the ED-Table that would have resulted in the best hedges over time. The definition of the best hedge has to be chosen by the professionals that would like to use those hedges to hedge against the interest rate risk. After several meetings and many tests to come to this fitness we came up with the following definition of the fitness function or in our case a pain function because the lower the outcome the better, as you will see by the following elements of the function:

- Risk-Free Interest Rate Change: The bigger the interest change in a bucket for a certain time frame is the more important it is to have the correct hedge-ratio for this bucket. If there would be no interest rate change at all within a bucket the impact of a bucket to the fitness or pain function would be 0, the bigger the interest rate change within a bucket, the more important it is to have a good fit on the hedge-ratio within this bucket.
- Duration: Duration has the same effect, the higher the duration the higher the impact. A duration of 0 would mean no interest rate sensitivity and the higher the duration of a bucket, the bigger the impact when hedge-ratios are not 'optimal'
- Market Value: The total market value weight is important for the impact of a bucket. If there would be no market value at all in a bucket the effect of the ED-Table bucket value would be zero. The bigger the market value invested in this bucket, the more important it is to give this bucket more importance in fitness.
- Worst case Hedge-ratio Difference (maximum minus minimum). The maximum pain within one bucket is the maximum minus the minimum. This is an important one, because we used 180 periods instead of one, we found more than one hedge-ratio for each bucket. Each of the 180 periods had its own hedge-ratio. Though, we looked at the difference between the maximum and minimum found hedge-ratio and this difference is an indication of how stable a hedge-ratio would be over time. The size of this difference is seen as a lack of accuracy, which would create a possible less good hedge for this instrument group. You can also see this as that we would constantly make the maximum mistake in taking the wrong hedge-ratio.

Interpreting the fitness function we did, the fitness or pain-function represents the mini-max regret and the total formula is as follows:

$$TF = \sum_{p=1}^p \sum_{b=1}^b |\Delta r_{p,b}| \cdot OAD_{p,b} \cdot MV - \%_{p,b} \cdot |\Delta HR|_b \quad (3.1)$$

where

TF = Total Fitness

$|\Delta r|$ = absolute delta R, monthly change in government yield

OAD = Option Adjusted Duration

$MV - \%$ = Market Value-% of this bucket

$|\Delta HR|$ = Maximum Hedge-ratio minus the Minimum Hedge-ratio for a bucket

p = period

b = bucket

Within the research we try to minimize the fitness of the individual MS-Table because in fact our fitness function defined as the pain. The higher the absolute interest rate differences within a certain period, for a certain bucket, the more pain this bucket will contribute to the total fitness. If there would be no changes in interest rate in a certain bucket this bucket would never contribute to the total pain. The same is true for the OAD. The higher the OAD, the more performance impact there is when the interest rate changes. The higher the MV-% of a bucket the more impact this bucket can have to the total pain. The last factor, $|\Delta HR|$, assumes that the worst possible hedge is done based on the worst empirical predictions of the MS-Table. For this we take the difference between the maximum and the minimum hedge ratio for a bucket for all periods and assume that this is the pain for a certain bucket. The multiplication of all those variables will, for the initial table, 5400 pain values and we are searching for the MS-Table that would generate the least amount of pain.

3.6 Used settings

We have taken a close look at each of the different genetic algorithm settings. We used a population size of 9 in this research and a function tolerance of 0.01 and a time limit of 3 days. The time limit was almost never reached and we had to put a lot of effort in optimizing the Matlab code in order to keep the running time as low as possible. Therefore we had to save variables in a database instead of to recalculate them each iteration and to use incremental tricks for calculating averages and so forth.

One call of the fitness function in Matlab would look the following:

TableToFit([20 20 20 20 20 1.25 1.25 1.25 1.25])

The TableToFit function will generate Table 3.3:

		Spread level (basis points)					
		0 - 20	20 - 40	40 - 60	60 - 80	80 - 100	100+
Maturity (years)	0 – 1.25	1*	2	3	4	5	6
	1.25 – 2.5	7	8	9	10	11	12
	2.5 – 3.75	13	14	15	16	17	18
	3.75 – 5	19	20	21	22	23	24
	5+	25	26	27	28	29	30

Table 3.3 - Table dynamically generated from the following genetic representation:
 TableToFit([20 20 20 20 20 1.25 1.25 1.25 1.25])

When the function (3.1) TableToFit is called with this gene, the result will be the fitness of this MS-Table.

Chapter 4 – The Hedge-ratio Model and Bucket Classification

4.1 Hedge-ratio Model

There are different ways to calculate a hedge-ratio. There are different linear regression models that could be used. We chose a model from the study of De Backer [5]. We chose this model because it describes the relation between the interest rate changes on the excess return or yield change in bonds.

We will concentrate on the hedge-ratio. We have established that for low ratings or low spread levels the hedge-ratio will be high, meaning that the interest rate risk should be taken into account. Ben Dor et al (2005) [2] computes the hedge-ratio directly by performing the following regression:

$$TR - \frac{y^0}{12} = \beta_1 + \beta_2 * \Delta r * OAD + \beta_3 * \Delta r * OAS * OAD . \quad (4.1)$$

The hedge-ratio is given by

$$HR = -\beta_2 - \beta_3 * OAS . \quad (4.2)$$

where

TR = Total Return

$\frac{y^0}{12}$ = Monthly yield to worst, of the previous month

β_1 = Coefficient 1

β_2 = Coefficient 2

Δr = delta R, monthly change in government yield

OAD = Option Adjusted Duration

β_3 = Coefficient 3

OAS = Option Adjusted Spread

HR = Hedge-ratio

The left part of the equation shows the difference in expected excess return and realized expected return in a certain period. We will try to explain this difference by performing a regression and to find the relationship between this difference and the change in the government yield (Δr). Analyzing the error term made showed that we would do even better to include the last term which catches the different relationship for different oas-regimes. This decision can also be explained qualitatively because of the probability of default for companies behind high yielding bonds.

Though, for a fixed bucket will have approximately a constant spread level for all months, so the final term in equation (3.1) may be omitted because it will be captured in the coefficient of the second term, β_2 . So for a fixed bucket we will perform the regression given by

$$TR - \frac{y_0}{12} = \beta_1 + \beta_2 * \Delta r * OAD. \quad (4.3)$$

Because there is a negative relation between the change in government yield and the change in expected return we will take the negative value of β_2 to determine the hedge-ratio. It's common practice to take the positive value of a hedge-ratio when you talk about a hedge-ratio.

One of the reasons we created maturity/spread buckets was to be able to eliminate the last part of the equation.

Using equation (43) the hedge-ratio for a bucket will be equal to a constant, $HR = -\beta_2$.

4.2.1 Data

The Excess Return that is widely used in the market is calculated using the duration.

September - 1988	October - 1988	November - 1988	August - 1993	September - 1993	October - 1993	July - 1998	August - 1998	September - 1998	August - 1999	September - 1999	October - 1999	July - 2004	Time Frame
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Time Frame 1
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Time Frame 2
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Time Frame 3
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Time Frame 60
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Time Frame 131

Table 4.1 - Visual representation of time frames. Each time frame represents a period of 60 months

We would like to use all periods of 5 years starting from October 1988 – September 1993, the next analysis contains November 1988 – October 1993 and so on until August - 1999 – July 2004.

In this thesis we are going to generate the hedge-ratios for all different buckets and time frames. We will graph the results and interpret them.

We will also do a research that shows the month on month relationship between government yield changes and corporate yield changes. This was because of new insights found because of the done research.

In experiment 1 we are going to look at one specific group (Table 1.2), group 1. This group contains all bonds with a maturity of 0-3 years and a spread of 0-50 basis points. We chose for this group (cluster) because the outcomes were reliable in this bucket. Most buckets with a spread till 150 had reliable outcomes for the hedge-ratio in De Backer (2006) [5]. We are going to follow this group over time (Table 1.3). We have data of 190 months; the length of the period is 60 months. We calculated 131 hedge-ratios.

We have used monthly US bond data for the period September 30, 1988 until July 31, 2004; this amounts to a total of 190 months (Table 4.1). The dataset contains bond data of 25,000 bonds for this period. All bonds existed for a certain period, which is noted with the p in Table 3.1.

	3-10-1988	1-11-1988	1-12-1988	1-12-1992	4-1-1993	1-2-1993	1-4-2002	1-5-2002	1-4-2004	3-5-2004	1-6-2004	1-7-2004	2-8-2004
bond1	-	-	p	p	-	-	-	-	-	-	-	-	-
bond2	-	-	p	p	p	p	-	-	-	-	-	-	-
bond3	-	-	-	-	-	-	-	p	p	-	-	-	-
bond4	-	-	p	p	-	-	-	-	-	-	-	-	-
..	-	-	-	-	-	-	p	p	p	p	p	p	p
bond25000	-	p	p	-	-	-	-	-	-	-	-	-	-

Table 4.2 - Visualization of data. From the start till the end a 'p' represents a period that a bond existed

It is important to note that it is possible that a bond will be used in the regression of one month could be out of the dataset the month after. Also new bonds could be included. Each bond has its own characteristics and the average characteristics of all bonds in the universe can change over time. Though, those changes go slowly as the universe consists of 25,000 bonds and only a fraction is included and excluded month on month.

The dataset will be divided into maturity/spread buckets shown in Table 1.2. First of all, for every date we calculated the corresponding government bond yield for every maturity available. We did not create buckets using the rating of a bond, but instead we used the *spread level* and *maturity date* because recent research showed that this is a better fragmentation. We used the spread level because we assume a direct relation between the spread level and the correlation between spread and interest rate changes and because spread is omitted in the regression and we would like the spread level to be relatively constant for a group. We have done this for fixed buckets. As shown in Table 4.3 we divided the data in 30 maturity/spread buckets.

		Spread level (basis points)					
		0 - 50	50 - 100	100 - 150	150 - 250	250 - 400	400+
Maturity (years)	0 - 3	1	2	3	4	5	6
	3 - 5	7	8	9	10	11	12
	5 - 7	13	14	15	16	17	18
	7 - 10	19	20	21	22	23	24
	10+	25	26	27	28	29	30

Table 4.3: Maturity/Spread buckets. This was the originally used MS-Table from the work of De Backer

For each month, every available bond is put in its corresponding maturity/spread bucket. Then for each bucket the market value weighted average of a number of properties will be calculated: the spread level, the total return, the excess return, the OAD, the government bond yield, the change in government bond yield and finally the yield to worst.

4.2.2 Data analysis

Before we started the research we did a lot of work to gather and analyze the used data.

The monthly US bond data for the period September 30, 1988 until July 31, 2004, this amounts to a total of 190 months came from Lehman Brothers. The dataset contains bond data of 25,000 bonds for this period.

Each row in the database represents a bond. At a certain point in time the bond will be created. We enriched the data with the difference between variables from month to month. It's important to note that the first month we will find data from a bond the difference cannot be shown compared to the previous month because there were no values before. (explain clearly how this impacts available data and risks in a econometric model.) This also holds true for the month that a bond will mature. We had to write Matlab code to prevent data errors. This way we calculated the difference in spread and the difference in duration and excess return for a all bonds. For some we had to interpolate and compare to an interest rate curve which was a calculation intensive process.

We also used a JP Morgan Interest Rate curve and we used a linear interpolation method. This is standard market practice instead of making assumptions of the shape of a curve. For each security in the benchmark we looked up the 'risk free interest rate' on the JP Morgan Interest Rate table. We also calculated the monthly change in yield within the risk-free space and stored this for each bond. This was a calculation intensive process.

We also created many graphs and interpreted those graphs in the search for the best fitting model/spread regimes/duration regimes and knowledge of market practice was used to generate the best possible buckets.

4.2.3 Graphs of data analysis – Visualization of the data

In order to get a good feel of how the 25,000 bonds in the universe were distributed we decided to make a visualization of the data. We did this to find gaps in the data and common errors.

Different graphs of the data:

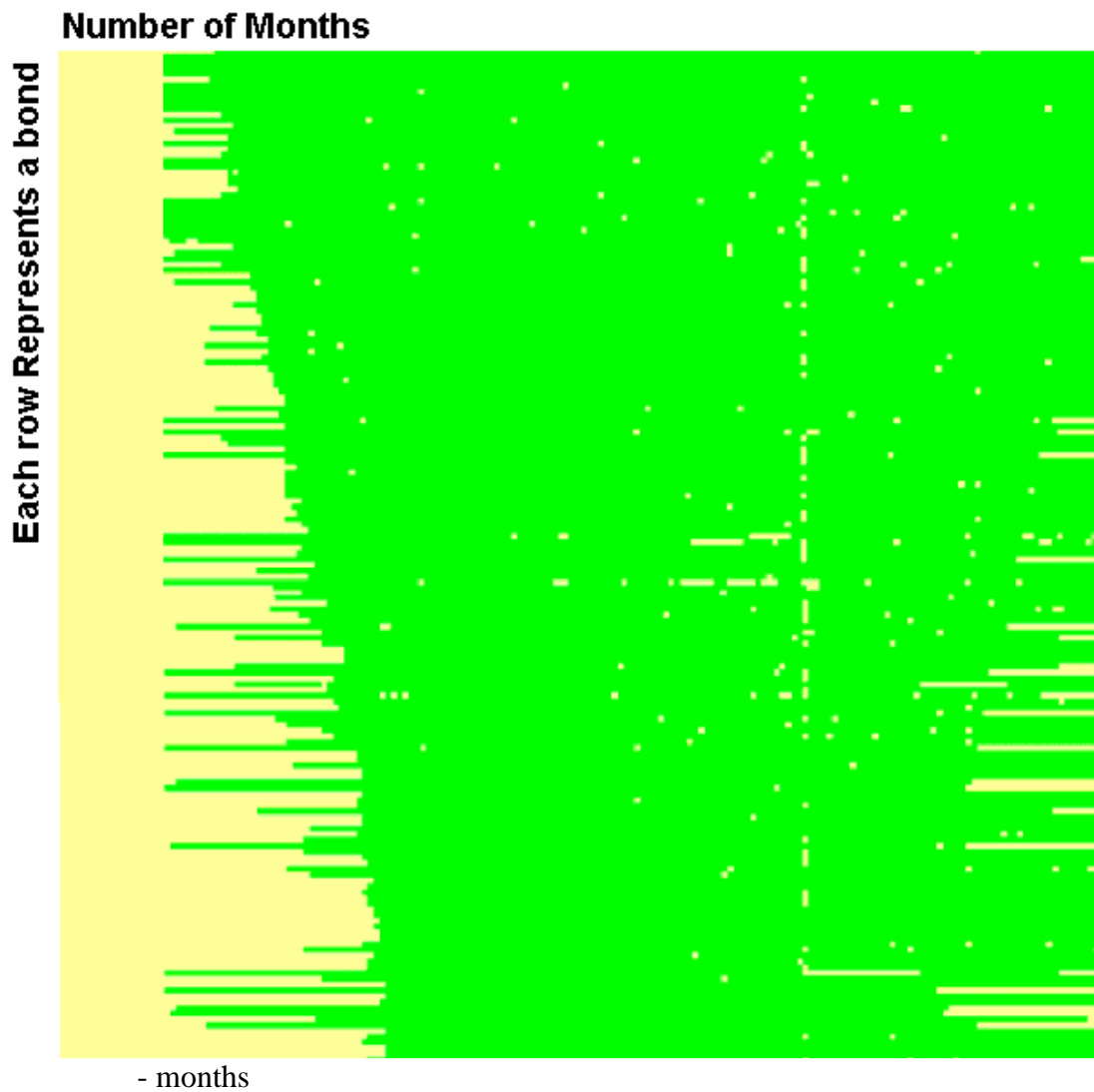


Figure 4.1 - Visualization of a larger part of the AIDA Database: Horizontal; months, Vertical; each row represents a bond. We that for bonds that exist for a longer period there are some missing datapoints are missing. Data is only used in a regression when there is a complete history of 60 months.

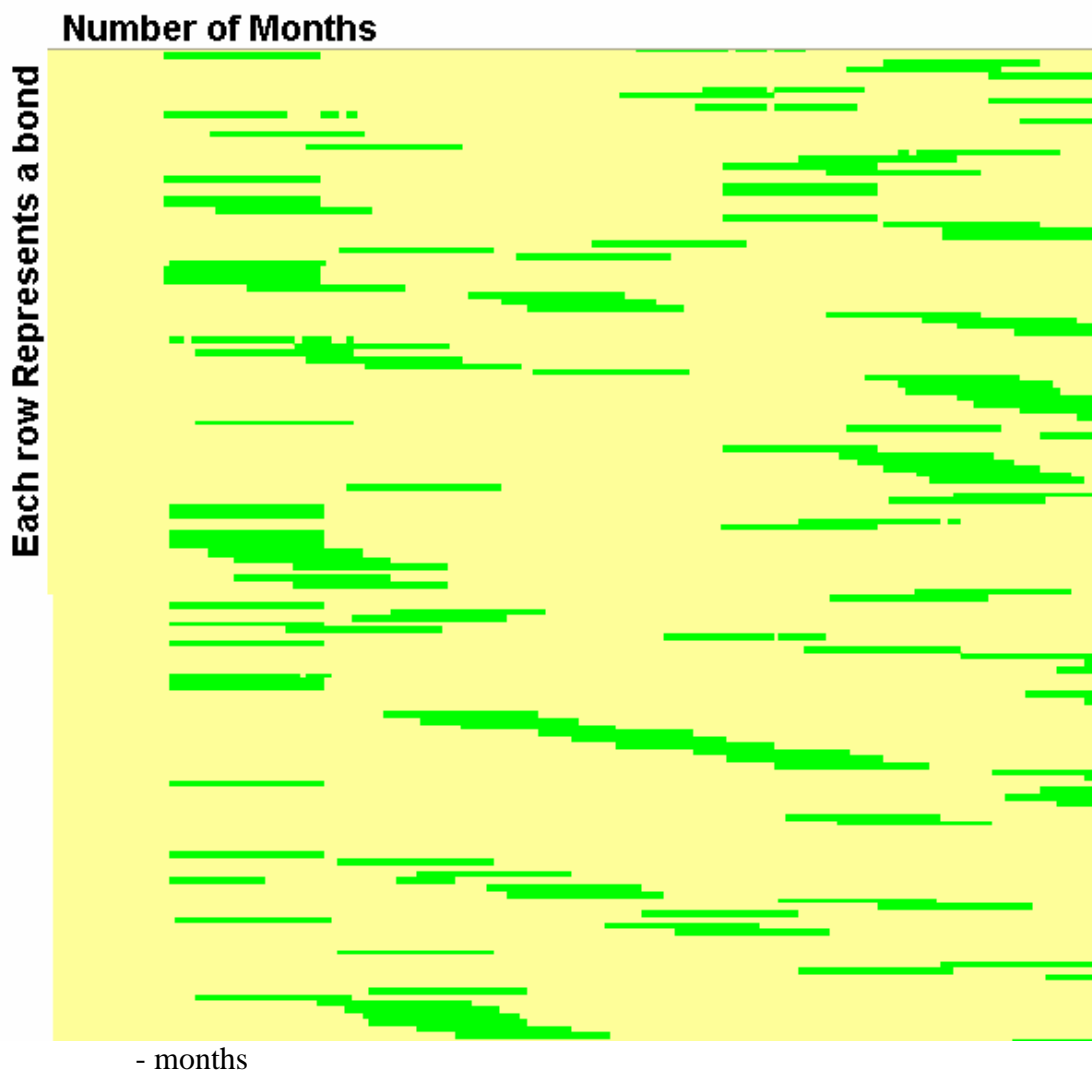


Figure 4.2 - Visualization of a part of the AIDA Database: The data of other bonds which exist for a shorter period of time are relatively complete

Green stands for data for this bond and yellow means that there is no data. At a certain point in time a bond is issued. From that point onwards you expect a green line to emerge in as shown above. Though, if there is a yellow dot this means that there is no data at a certain period for this bond. In the first screenshot we see some yellow points in a vertical line which means that the data quality was low for that month and we see some other gaps, but in general the quality of the data is good.

The reason we did this test is to visually inspect the data and to see if we could find a pattern in missing data. If for instance bonds that would exist for five years would always have several datapoints missing and this wouldn't be the case for bonds that existed 3 years we could say something about a certain bias of the total universe. We were not find such structural mistakes in the database with our visual inspection.

The information in the AIDA database is carefully selected to represent the bond market in the past. It should be representative on many average values in most big benchmarks in each period. This is on average spread, on average duration in a certain period in time. This is proprietary data from Barclays and the data itself cannot be published directly. A lot of work has been done to generate this database.

Important to notice is that data is only used in our model if all required fields about this bond were available.

4.2.4 Data fields

For a total of 25,000 bonds the following data fields were available:

Coupon	Rating
Excess Return	Sector
Market Value	Source
Maturity	Spread
Option Adjusted Duration	Total Return
Price	Yield

Table 4.4 - Available attributes for the bonds in the AIDA database

Within the research we used the Maturity and the Spread to assign bonds a certain bucket in the MS-Table. Market Value and Option Adjusted Duration were used to calculate the fitness and in calculating the hedge-ratio.

4.2.5 Monthly change in government yield

The monthly change in government yield (Δr) is determined with JP Morgan data on US Treasury Bonds. This data was on a monthly basis. We create an interpolation curve to be able to determine the interest rate change for a certain point in history and for a certain maturity. This data is used in the research for each bond at each maturity. Combining the government yield change at a certain point in history for a certain maturity was a calculation intensive process. Table 4.5 shows a snapshot of the US Treasury Yields.

Date	3 mnd	6 mnd	1yr	2yr	3yr	5yr	10yr	30yr
2-Jan-89	8.4	8.7	8.8	9.0	7.8	9.0	9.0	8.9
2-Jan-90	8.0	7.9	7.8	7.9	6.8	7.9	7.9	8.0
2-Jan-91	6.6	6.6	6.5	7.1	6.2	7.6	8.0	8.1
2-Jan-92	4.2	3.9	4.0	4.8	5.1	6.0	6.8	7.5
4-Jan-93	3.2	3.4	3.5	4.5	5.0	5.9	6.6	7.3
3-Jan-94	3.2	4.1	3.6	4.3	4.6	5.3	5.9	6.4
2-Jan-95	5.9	6.0	6.9	7.7	7.8	7.8	7.8	7.9
2-Jan-96	5.2	5.2	5.2	5.2	5.2	5.4	5.6	6.0
2-Jan-97	5.2	5.3	5.6	6.0	6.1	6.3	6.5	6.8
2-Jan-98	5.3	5.4	5.4	5.6	5.6	5.6	5.7	5.8
4-Jan-99	4.5	4.6	4.6	4.6	4.7	4.6	4.7	5.2
3-Jan-00	5.4	5.8	6.0	6.4	6.4	6.5	6.6	6.6
2-Jan-01	5.8	5.6	5.2	4.9	4.9	4.8	4.9	5.4
2-Jan-02	1.7	1.8	2.2	3.2	3.8	4.5	5.2	5.6
2-Jan-03	1.2	1.2	1.4	1.8	2.1	3.0	4.0	5.0
2-Jan-04	0.9	1.0	1.3	1.9	2.4	3.3	4.4	5.2

Table 4.5 - JP Morgan data snapshot with yields of US Treasury bonds.

This table was used to determine the monthly change of each point, shown in Table 4.6 and Table 4.7:

Date	3 mnd	6 mnd	1yr	2yr	3yr	5yr	10yr	30yr
3-Oct-88	7.5	7.8	8.1	8.4	7.2	8.6	8.8	9.0
1-Nov-88	7.6	7.8	8.0	8.2	7.3	8.4	8.7	8.8
1-Dec-88	8.1	8.4	8.6	8.8	7.5	8.9	9.0	9.0
2-Jan-89	8.4	8.7	8.8	9.0	7.8	9.0	9.0	8.9
1-Feb-89	8.6	8.9	9.0	9.1	8.2	9.0	9.0	8.8
1-Mar-89	9.0	9.2	9.4	9.5	7.9	9.4	9.3	9.2
3-Apr-89	9.2	9.4	9.5	9.6	8.2	9.4	9.2	9.1
1-May-89	9.0	9.2	9.3	9.3	8.3	9.2	9.1	9.0
1-Jun-89	8.9	8.9	8.9	8.8	7.2	8.6	8.6	8.6
3-Jul-89	8.2	8.1	8.0	8.0	6.8	8.0	8.1	8.1
1-Aug-89	7.9	7.9	7.6	7.5	6.5	7.5	7.7	7.8
1-Sep-89	8.1	7.8	8.2	8.3	7.0	8.1	8.2	8.1
2-Oct-89	8.1	7.8	8.5	8.5	7.3	8.3	8.3	8.2

Table 4.6 - Yield values of US Treasuries on a monthly basis.

Date	Interest Rate Change									
	3 mnd	6 mnd	1yr	2yr	3yr	5yr	10yr	30yr		
3-Oct-88										
1-Nov-88	0.08	0.03	- 0.07	- 0.17	0.08	- 0.21	- 0.16	- 0.20		
1-Dec-88	0.53	0.59	0.59	0.62	0.13	0.53	0.34	0.27		
2-Jan-89	0.28	0.22	0.19	0.14	0.36	0.07	- 0.01	- 0.10		
1-Feb-89	0.28	0.22	0.19	0.14	0.36	0.07	- 0.01	- 0.10		
1-Mar-89	0.33	0.30	0.37	0.41	- 0.28	0.39	0.36	0.35		
3-Apr-89	0.20	0.22	0.12	0.02	0.34	- 0.04	- 0.14	- 0.11		
1-May-89	- 0.21	- 0.21	- 0.20	- 0.23	0.03	- 0.22	- 0.08	- 0.07		
1-Jun-89	- 0.05	- 0.32	- 0.43	- 0.52	- 1.02	- 0.51	- 0.52	- 0.40		
3-Jul-89	- 0.69	- 0.78	- 0.83	- 0.80	- 0.48	- 0.67	- 0.51	- 0.53		
1-Aug-89	- 0.33	- 0.17	- 0.42	- 0.55	- 0.24	- 0.49	- 0.36	- 0.24		
1-Sep-89	0.21	0.09	0.60	0.86	0.44	0.67	0.44	0.30		
2-Oct-89	- 0.03	- 0.04	0.27	0.14	0.36	0.18	0.12	0.10		

Table 4.7 - Yield changes compared to the previous period.

To illustrate, at 1-Dec-88 the interest rate at the 6 month part of the curve was 8.4 and 1-Nov-88 this was 7.8. This shows an interest increase of 0.59 (0.6) for this treasury. In order to calculate the excess return of a corporate bond with the same maturity the yield change of this bond should be compared to this increase of 0.59. If the yield of a corporate increases the same amount, there is no spread impact. If the yield increases more there is a spread widening and if it increases less there is a spread tightening. This effect can be taken into consideration when using the duration to hedge the interest rate risk.

Chapter 5 – Research Setup

5.1 Research Steps

5.1.1 Outline

In order to answer the research questions we wanted to start with investigating one bucket of one table and see how stable those results were through time. This should be seen as the first part of the research.

Within this first part, Experiment 1, we will take a close look at the Empirical Durations through time. We will start to look at 1 bucket in 1 time frame to be able to show the outcome of this research. After this, we will look at one bucket for 131 time frames and show graphs about how the different values would have changed over time. This is very important because we would like to have a good feeling about stable different parameters are over time. Within the graph we will highlight the values of bucket 1, the bucket with bonds with a spread lower than 50 and a maturity till 3 years. We make different graphs, where an important one is the graph of the hedge-ratio over time. The knowledge from this experiment was also used to determine the fitness function.

When we understand and feel comfortable with the outcomes of the previous experiments we go on the second experiment, Experiment 2, where we use the acquired knowledge and designed fitness function to determine the fitness of different MS-Tables. We start with giving the existing MS-Table a fitness score.

At this moment we are able to rate each MS-Table and we will start with an iterative search over different MS-Tables to determine their fitness. The next step within the second experiment is to use the genetic algorithm to go to the search space and come up with the MS-Table with the lowest fitness. The last step was to use the lowest fitness scores found by the iterative search and use those as the start point for the genetic algorithm in order to search at relevant spaces within the search space.

This is the main part of the research and this Genetic Algorithm and this fitness function is what this thesis is about and where we try to add value.

We will eventually evaluate those results and draw conclusions from them.

An overview of the experiments:

Experiment 1 - How does the Empirical Duration model behave through time

Experiment 1.1 - 1 bucket, 1 time frame

Experiment 1.2 - 1 bucket, 131 time frames

Experiment 1.3 - 30 buckets - 131 time frames

Experiment 2 - Fitness Function Evaluation

Experiment 2.1 - Determining the Fitness of the Currently used Table (Table 1.2)

- Experiment 2.2 - Iterative Evaluation of different Tables
- Experiment 2.3 - Genetic Algorithm Free Search
- Experiment 2.4 - Genetic Algorithm Search - Starting points from Experiment 2.2

5.2 Experiment Description

In this chapter we will describe the details and reasoning behind several experiments.

5.2.1 Experiment 1 - Analyze the dynamics of the ED-space over time

Experiment 1.1 - 1 bucket, 1 time frame

We created a Matlab script to calculate (Formula 4.3) the above described model.

The variables that are calculated are described in Table 5.1:

Average OAS (Option Adjusted Spread)	Number of Corporate Bonds in a bucket
Average deltaR (change in risk-free rate)	Durbin Watson
Average ER (Excess Return)	
Average OAD (Option Adjusted Duration)	
Average TR (Total Return)	
Average TRminusYo (Total Return minus Risk free Return)	
Average Yield	
Variable 1 Coefficient	
Variable 1 Coefficient-statistic	
Variable 1 t-probability	
Variable 2 Coefficient (=Hedge-ratio)	
Variable 2 Coefficient-statistic	
Variable 2 t-probability	

Table 5.1: Calculated Variables in the First Experiment

We are going to perform the regression given in equation (4.3) for one specific spread/maturity bucket, group 1 (Table 4.3). We chose group one because the R-squared was high and this group contains enough bonds (data) for the entire period. We can also see that for low spread levels and a low maturity the hedge-ratio may be approximated by a straight line in Figure 6.1. This is also a reason why we chose for Group 1. We will start with calculating this model for bucket 1 and for the period of October-88 till September-93. This is a period of 60 months.

Experiment 1.2 - 1 bucket, 131 time frames

The second enhancement to the first experiment is to run it for 131 different start dates (Table 4.1) and we chose the length of a time frame of 60 months for all start dates.

	From	Till
Analysis 1	October-88	September-93
Analysis 2	November-88	October-93
.....
Analysis 130	July-99	June-04
Analysis 131	August-99	July-04

Table 5.2: This table shows the different time frames of 60 months

In the second experiment enables us to take a closer look at one specific bucket through time, in this case Group 1, all bonds with a maturity of 0-3 years and a spread of 0-50 basis points. In total of 131 hedge-ratios (Table 3.3) will be calculated for this specific group.

Experiment 1.3 - 30 buckets - 131 time frames

We performed the entire research for each bucket (30) and for each period (131) which results in 3.960 regressions. The outcome of all those regressions is a set of variables containing the r-squared, the average oad, average and the calculated hedge-ratio within this period.

The output will be ordered in tables, visualized and attached to this thesis. We will analyze the output in detail.

It took a lot of calculation power and time to calculate those results

5.2.2 Experiment 2 - Fitness Function Evaluation for MS-Tables

Experiment 2.1 - Fitness Function for the Currently used table

The total fitness of the used MS-Table used in the research of De Backer will be determined.

Experiment 2.2 - Iterative Evaluation of different Tables

We calculated the fitness for many different tables with a constant size of each maturity bucket and each spread bucket. We incremented the spread with 10 basispoint for each iteration and the maturity with 0.25 for each iteration and did this until the first bucket contained all bonds. It took a very long time to do this experiment and we had to make many parts of the Matlab code as efficient as possible in order to be able to do this.

Within this experiment we calculated the difference in the maximum and the minimum hedge-ratio that we found using the model for 131 time frames which we used as the Hedge-ratio Difference. This means that from the found 131 values we determined the highest and the lowest value and determined the difference between them. This is a measure for how stable and trustworthy the outcome of this value is for one period.

We wrote a function TableToFit.m which accepted different MS-Tables as a gene. The gene is a representation of the table. As the table has 6 columns for the spread and 5 rows for the maturity, there are 5 split points for the spread and 4 splitting points for the duration. The gene therefore consists of 9 values. To give an example, Table 5.1 is generated from a gene.

		Spread level (basis points)					
		0 - 100	100 - 200	200 - 300	300 - 400	400 - 500	500+
Maturity (years)	0 - 5	1	2	3	4	5	6
	5 - 10	7	8	9	10	11	12
	10 - 15	13	14	15	16	17	18
	15 - 20	19	20	21	22	23	24
	20+	25	26	27	28	29	30

Table 5.1 - Example Maturity-Spread Table generated in the Iterative Process - generated from the following gene: [100 100 100 100 100 5 5 5 5]

Eventually when we use this MS-Table this table will have a certain fitness based on our definition of fitness. In the results we will show the different tables and the different fitness values and will try to explain why the results are the way they are.

We let the spread difference go from 20 to 300 with steps of 5, which will generate 56 different Spread split-ups and the duration from 1 to 6 with steps of 0.25. This results in $((300-20)/5+1)*((6-1)/0.25+1)=1197$ different MS-Tables that we generated and calculated the fitness for.

Experiment 2.3 - Genetic Algorithm Free Search

Following on Experiment 2.1, we changed the method of MS-Table generation to a method that generates a table totally free based on a genetic algorithm. We started with the standard genetic algorithm settings and started with a many random start positions and let the algorithm come up with the best MS-Table. We used the same fitness function as in the previously described experiments.

Many different tables are possible because the table is not limited to fixed distances in Spread or Maturity. The number of buckets can also be less than 30 as the algorithm can create tables that contain less than 30 buckets because an interval can be zero and therefore be empty and contain no bonds, and therefore it will not contribute to the total pain. Important to notice is that if there are empty groups, other groups will contain more bonds and more market value so the pain will be shifted to those groups. If the difference in hedge ratio is bigger in those buckets than it would have been in others buckets this could increase the total pain.

Experiment 2.4 - Genetic Algorithm Search - Starting points from Experiment 2.2

Within this experiment we decided to run the genetic algorithm an extra 24 times. For each run we used the best outcomes of experiment 2.1 as the start positions of the genetic algorithm. The reason behind this decision was because the search space was very big and calculation time was an issue we preferred start positions with a low fitness value and would like to see if they could be even further optimized.

Chapter 6 – Results

Outline

Within this chapter we will show and describe the generated results. Within Experiment 1 we will discuss the results for one bucket and the results for several time frames. When we have gone through this in detail and understand how one MS-Table will behave and how stable the resulting Empirical Duration is, we will continue with the results from Experiment 2 where we will try to find the best possible MS-Table.

6.1 Experiment 1 - How does the Empirical Duration model behave through time

6.1.1 Experiment 1.1 - 1 bucket, 1 time frame

The results are summarized in Table 6.1.

maturity bucket	Spread bucket	Hedge-ratio	R-squared (%)
0 - 3	0-50	0.89	0.82
	50-100	0.88	0.81
	100-150	0.78	0.61
	150-250	0.69	0.33
	250-400	0.25	0.01
	400+	-0.79	0.01
3 - 5	0-50	0.86	0.80
	50-100	0.88	0.81
	100-150	0.80	0.73
	150-250	0.67	0.49
	250-400	0.35	0.09
	400+	-0.27	0.01
5 - 7	0-50	0.90	0.84
	50-100	0.86	0.80
	100-150	0.82	0.76
	150-250	0.67	0.51
	250-400	0.37	0.12
	400+	-0.19	0.01
7 - 10	0-50	0.92	0.85
	50-100	0.87	0.82
	100-150	0.83	0.76
	150-250	0.68	0.51
	250-400	0.32	0.10
	400+	-0.04	0.00
	0-50	1.00	0.80

10+	50-100	0.88	0.81
	100-150	0.82	0.74
	150-250	0.70	0.57
	250-400	0.47	0.19
	400+	0.04	0.00

Table 6.1: Hedge-ratio for every bucket. (Aug-94 – Jul-99)

In Figure 6.1 a plot of the hedge-ratio for every maturity bucket is shown.

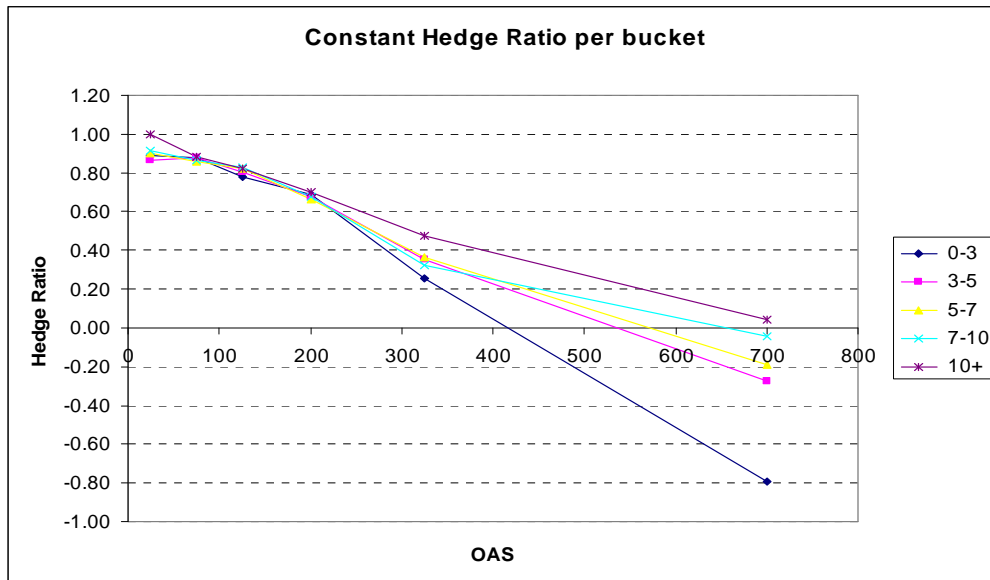


Figure 6.1: Hedge-ratio at average bucket spread levels, for every maturity bucket.

As we look at Figure 6.1 we can observe that the longer the maturity the higher the Hedge-ratio for a fixed spread bucket. For the spread regime 0-50 the distance between the hedge-ratios is clearly visible, the distance between the hedge-ratios becomes smaller for higher spreads and for higher spreads the distance increases again. For very high spreads we see negative Hedge-ratios.

In Figure 6.2 and 6.3 the value of R-squared and sigma squared (the standard deviation of the estimated error) respectively are displayed per bucket. We can see that for every maturity bucket the value of R-squared decreases when the spread level rises. The higher the spread level, the less of the variation in total return can be explained by variation in the independent variables (duration times changes in the government bond yield).

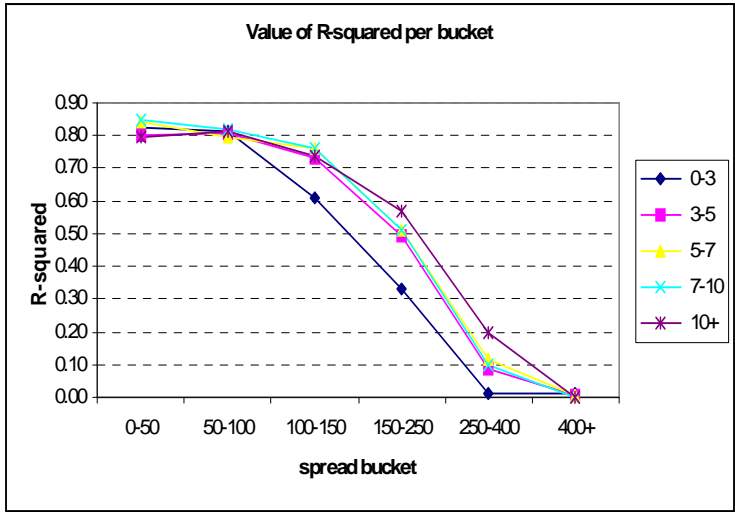


Figure 6.2:- R-Squared distribution for different Spread/Maturity Buckets

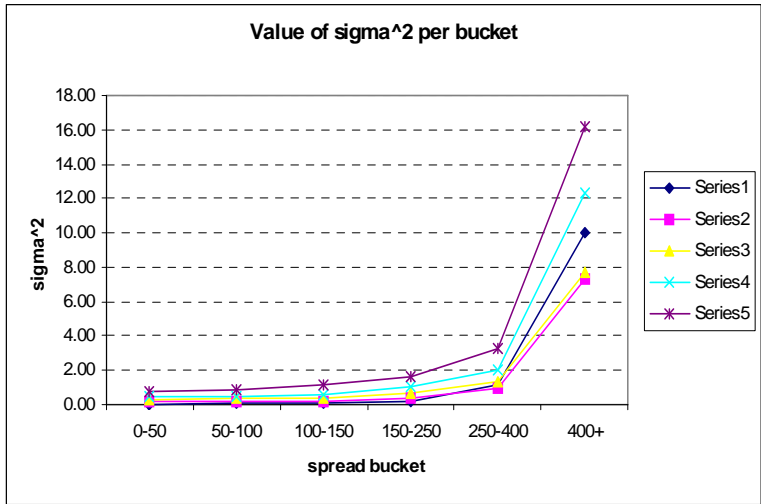


Figure 6.3- Sigma-Squared distribution for different Spread/Maturity Buckets

This first experiment shows us that for bonds with a long maturity and low spreads the model works very well, and that the same model performs less well for bonds with a higher spread and with a lower maturity.

6.1.2 Experiment 1.2 - 1 bucket, 131 time frames

The second experiment performs the same analysis done in the first experiment and we will analyse the data for a several periods and we will compared the outcomes between periods. We will try to find conclusions based on those results in the conclusion section.

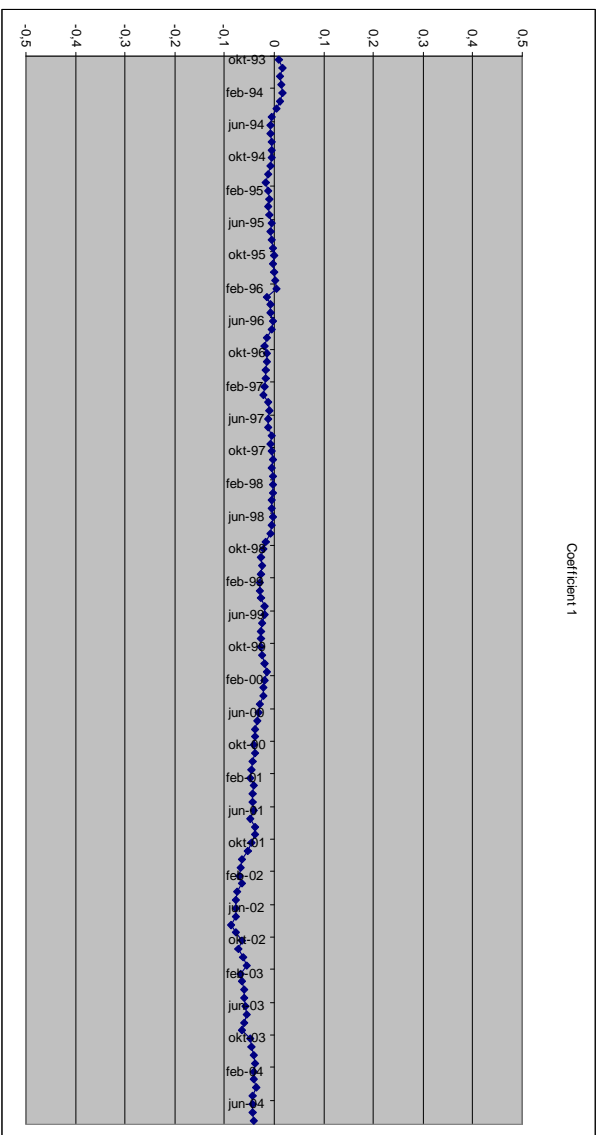


Figure 6.4 - β_1 - Coefficient 1

it becomes slightly negative but in general we can state that coefficient 1 is close to 0 for the entire period. β_1 is the constant in Formula 4.1 and is expected to be close to 0.

$$TR - \frac{y^0}{12} = \beta_1 + \beta_2 * \Delta r * OAD + \beta_3 * \Delta r * OAS * OAD$$

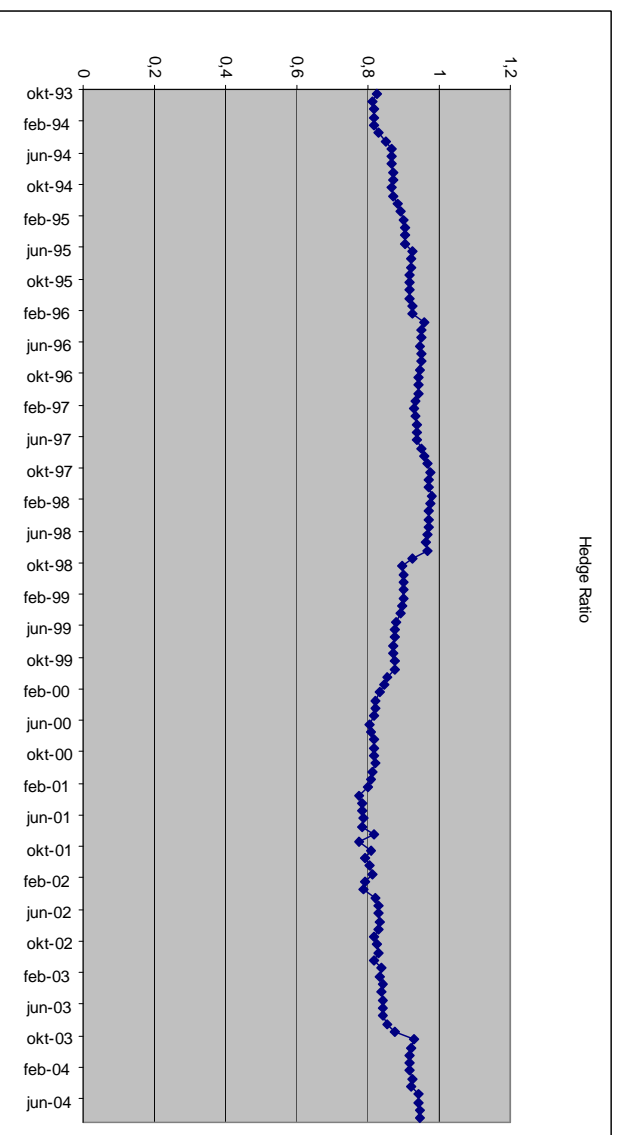


Figure 6.5 - Hedge-ratio - β_2

The hedge-ratio is the main variable of this research. As we have seen in Table 6.1 the Hedge-ratio is different for each separate group. Within Figure 6.5 we show the hedge ratio from Group 1 which contains bonds with a maturity between 0 and 3, and spread between 0 and 50. The hedge-ratio started in 1993 at 0.8 and moved between 0.8 and 1. The found hedge-ratios are created with 60 months of history for each data point. This variation between 0.8 and 1 is out of the scope of this research. We also looked at other hedge ratios and although they did vary through time the graphs had a similar form as Figure 6.5. We can investigate the exact trend further in future research as it is out of the scope of this research.

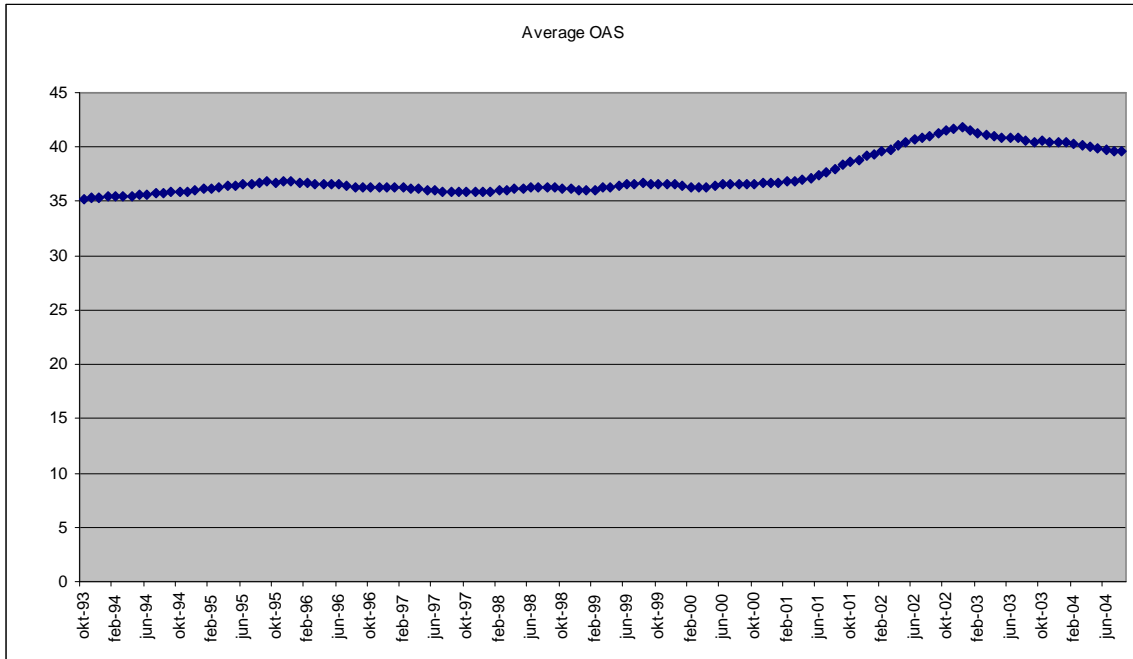


Figure 6.6 - Average OAS

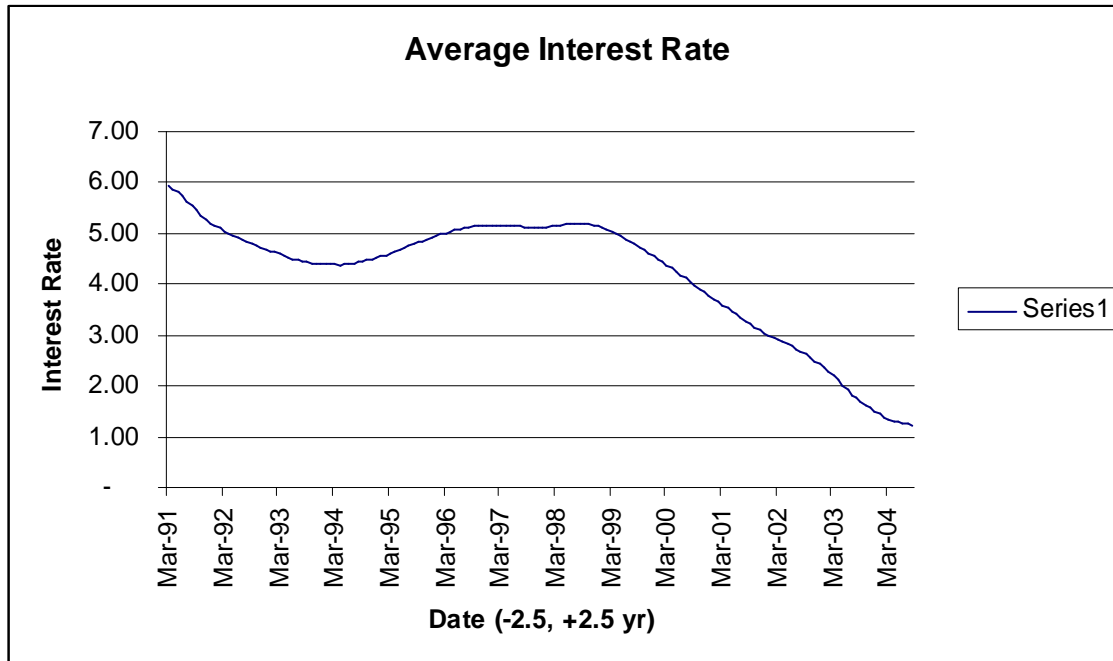


Figure 6.7 - Interest Rates

The OAS was mostly stable the first period, which gives us comfort that the algorithm which splits up the universe was working fine and that the average spread within a bucket was stable. The hike in average spread after this period is still small and those 7 basispoints movement are low compared to the lower interest rates in that period. Therefore this can be seen as a minor movement. The stability of spread is important for the interpretation and to be able to compare adjacent results in experiment 2. We can take a closer look at the influence of the change of OAS and the boundary choices in future research. When interest rates declined spreads increased, this seems logical and according to the known theory (Figure 6.7).

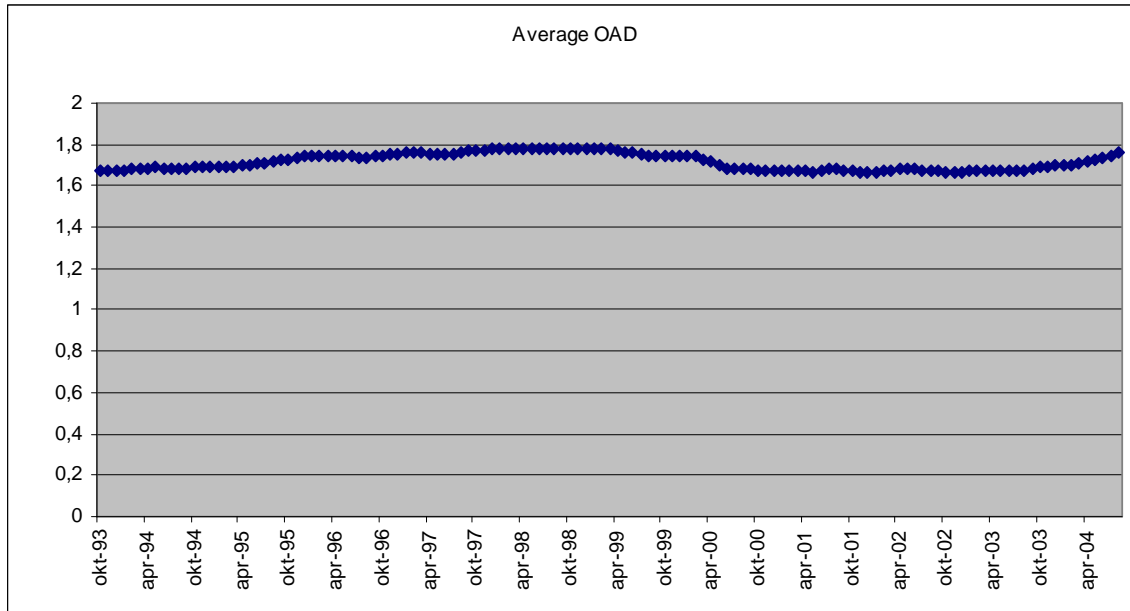


Figure 6.7 - Average OAD

When the Average OAD increases we also see an increase in the Hedge-ratio for the period. Both variables seem to be highly correlated. The Empirical Duration which is part of the OAD comes closer to the OAD when the OAD is in increasing on average.

The total change from 1.7 to 1.8 in Average OAD is a 5% maximum. The OAD moved in the same direction as the Hedge-ratio with a 5% movement with is different from the 25% movement of the Hedge-ratio. In future research we could take a closer look at other buckets and the influence of the Average OAD to the hedge-ratio. Though the OAD is relatively stable through time.

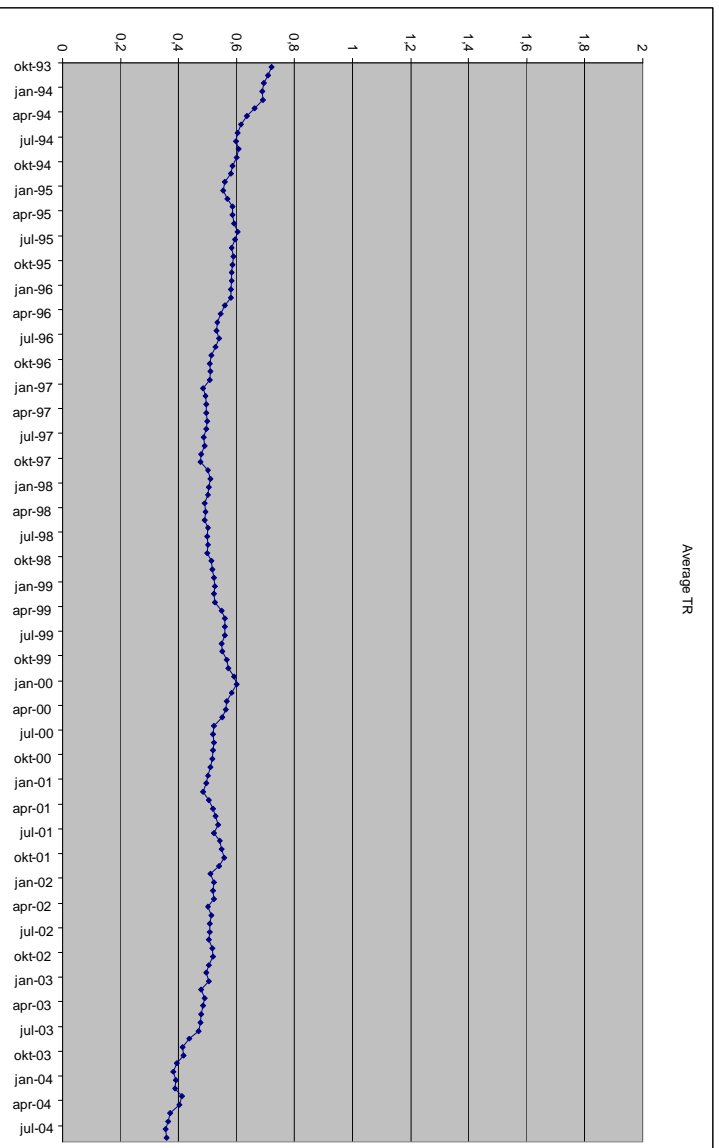


Figure 6.8 - Average TR

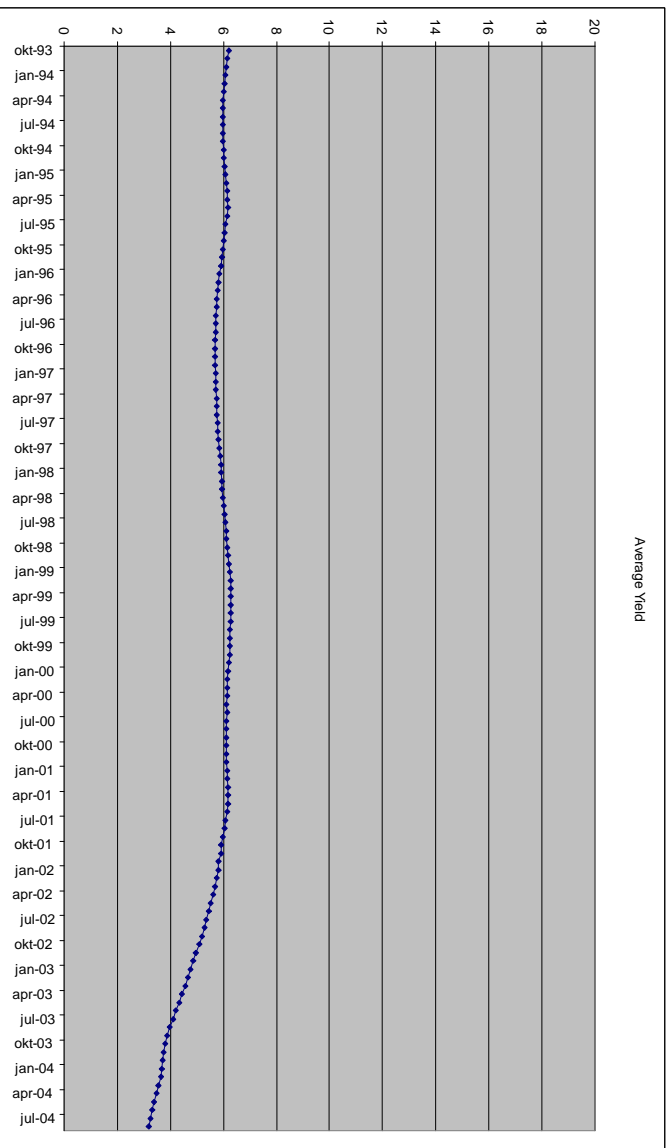


Figure 6.9 - Average Yield

The average total return declines when the yield declines.

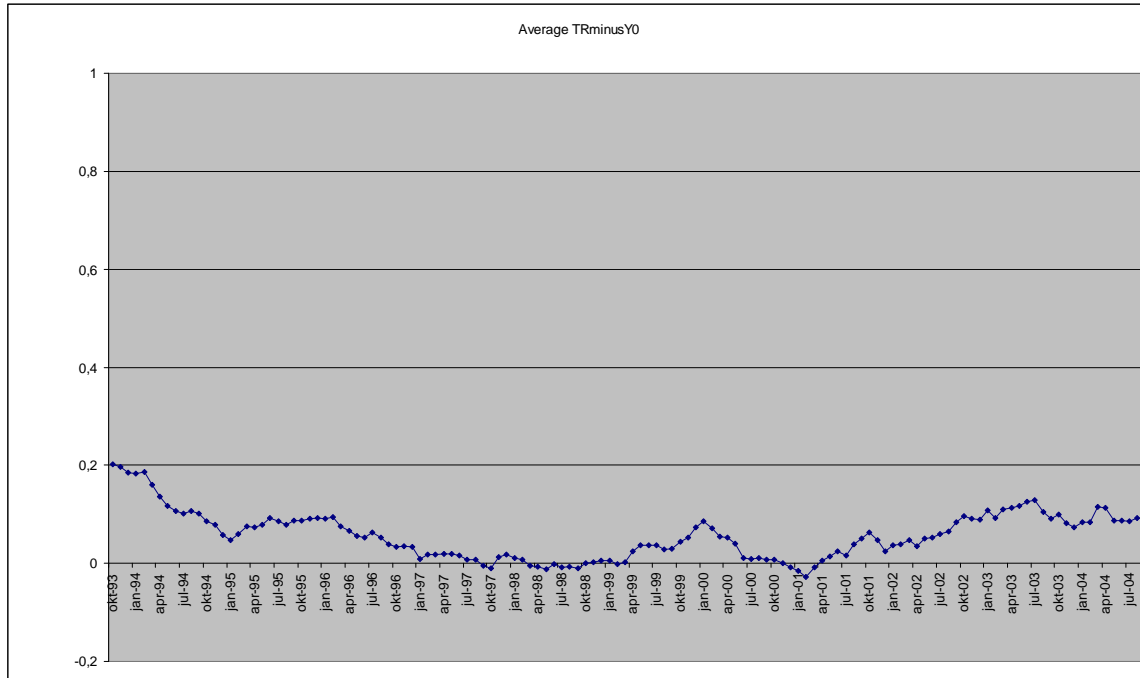


Figure 6.10 – Average TR minus Yo (Excess Return)

The difference between the values in Figure 6.8 and Figure 6.9 gives Figure 6.10. Figure 6.10 shows us the average Excess Return ($TR - \frac{y_0}{12}$) which is used in (4.3) to calculate the hedge-ratio. It is interesting to see that when the interest rates decline and the total yield declines that the excess return (8.10) is positive with low interest rates.

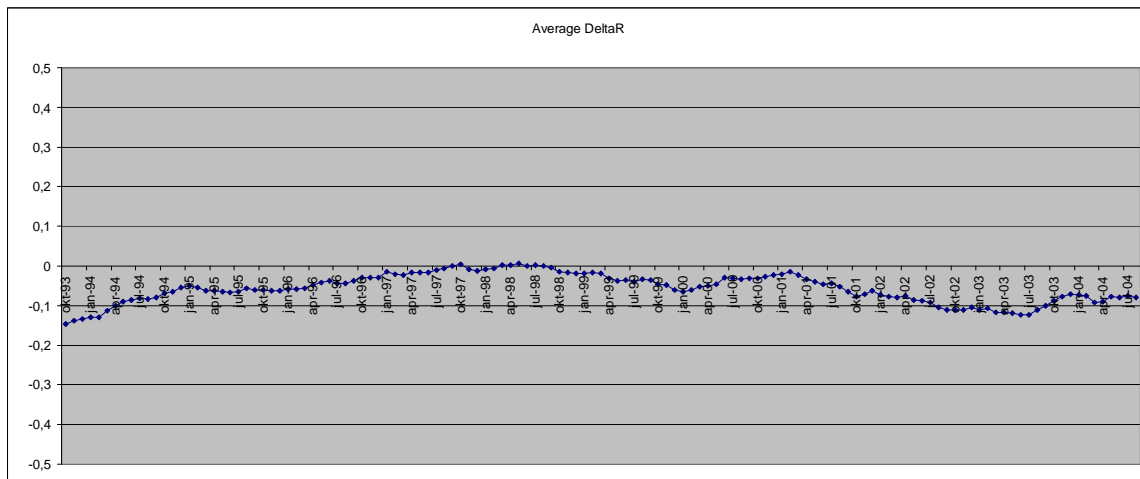


Figure 6.11 – Average Delta R

When the values in Average TR minus Yo (Figure 6.10) are close to 0, the values in Figure 4.11 are also close to 0. Very interesting too see is that when the changes in interest rates are very low, the excess returns of bonds are also very low.

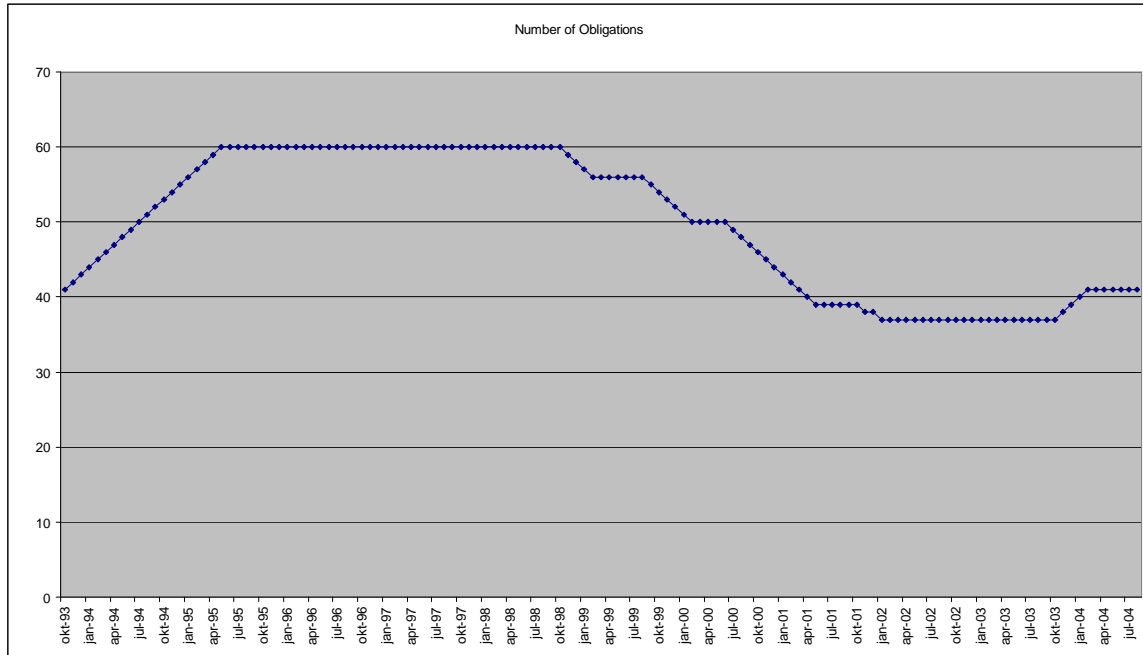


Figure 6.12 - *Number of bonds*

In the latter part of the simulations there were less bonds in the bucket. The minimum number was 38 and it's maximum was 60. The amount of bonds are still enough to generate a good regression and test the model on this data.

6.1.3 Experiment 1.3 - 30 buckets - 131 time frames

Because of the many results found in Experiment 1.3 we had to make several decisions in how to summarize those results in order to analyze them.

We chose for several ways to summarize the results in a way we could compare them. We wanted to have a good idea of averages through time and the difference between the maximum and minimum value and the distance from the average.

The variables that are calculated are described in Table 6.2:

Average Hedge-ratio
Average R-Squared
Maximum Hedge-ratio
Minimum Hedge-ratio
Max Hedge-ratio – Min Hedge-ratio
(Max – Min) / Min
Max – Avg.
Avg – Min.
Bucket Name
Rank

Maturity / Spread	Average of Hedge-ratio	Average of R-squared	Max of Hedge-ratio	Min of Hedge-ratio	Max - Min	(Max - Min)/Min	Max - Avg	Avg-Min	Bucket	
									Bucket Name	Bucket
0-3 / 0-50	0.88	0.83	0.98	0.77	0.20	26%	10%	11%	0-3 / 0-50	1
0-3 / 50-100	0.87	0.82	0.98	0.76	0.22	30%	11%	11%	0-3 / 50-100	2
0-3 / 100-150	0.78	0.61	1.05	0.44	0.61	139%	28%	34%	0-3 / 100-150	3
0-3 / 150-250	0.68	0.35	0.98	0.36	0.62	172%	30%	32%	0-3 / 150-250	4
0-3 / 250-400	0.45	0.10	1.21	0.03-	1.24	-4440%	75%	48%	0-3 / 250-400	5
0-3 / 400-inf	0.87-	0.05	0.95	2.27-	3.22	-142%	182%	140%	0-3 / 400-inf	6
3-5 / 0-50	0.86	0.81	1.05	0.62	0.43	70%	19%	24%	3-5 / 0-50	7
3-5 / 50-100	0.87	0.82	0.97	0.76	0.21	27%	10%	10%	3-5 / 50-100	8
3-5 / 100-150	0.80	0.74	0.94	0.66	0.27	41%	14%	13%	3-5 / 100-150	9
3-5 / 150-250	0.68	0.50	0.85	0.49	0.36	73%	17%	19%	3-5 / 150-250	10
3-5 / 250-400	0.36	0.14	0.55	0.10	0.45	452%	18%	27%	3-5 / 250-400	11
3-5 / 400-inf	0.10-	0.07	0.67	1.28-	1.96	-152%	77%	119%	3-5 / 400-inf	12
5-7 / 0-50	0.89	0.82	1.14	0.48	0.66	138%	26%	41%	5-7 / 0-50	13
5-7 / 50-100	0.84	0.80	0.95	0.71	0.24	34%	11%	13%	5-7 / 50-100	14
5-7 / 100-150	0.81	0.77	0.93	0.69	0.24	36%	12%	12%	5-7 / 100-150	15
5-7 / 150-250	0.66	0.50	0.79	0.52	0.27	53%	13%	14%	5-7 / 150-250	16
5-7 / 250-400	0.40	0.17	0.61	0.12	0.49	407%	21%	28%	5-7 / 250-400	17
5-7 / 400-inf	0.09-	0.05	0.61	0.97-	1.58	-163%	70%	88%	5-7 / 400-inf	18
7-10 / 0-50	0.92	0.85	1.02	0.76	0.26	34%	10%	16%	7-10 / 0-50	19
7-10 / 50-100	0.85	0.81	0.94	0.71	0.23	33%	9%	14%	7-10 / 50-100	20
7-10 / 100-150	0.82	0.75	0.92	0.67	0.25	37%	11%	14%	7-10 / 100-150	21
7-10 / 150-250	0.66	0.51	0.80	0.48	0.32	67%	14%	18%	7-10 / 150-250	22
7-10 / 250-400	0.32	0.14	0.56	0.06	0.50	845%	24%	26%	7-10 / 250-400	23
7-10 / 400-inf	0.05	0.05	0.88	0.77-	1.65	-214%	83%	81%	7-10 / 400-inf	24
10-inf / 0-50	1.01	0.80	1.17	0.75	0.42	57%	16%	26%	10-inf / 0-50	25
10-inf / 50-100	0.87	0.80	0.95	0.72	0.22	31%	8%	14%	10-inf / 50-100	26
10-inf / 100-150	0.81	0.72	0.93	0.61	0.32	51%	12%	20%	10-inf / 100-150	27
10-inf / 150-250	0.68	0.56	0.82	0.47	0.35	76%	14%	22%	10-inf / 150-250	28
10-inf / 250-400	0.49	0.25	0.68	0.27	0.41	151%	19%	22%	10-inf / 250-400	29
10-inf / 400-inf	0.06	0.05	0.64	0.73-	1.37	-189%	58%	79%	10-inf / 400-inf	30
Grand Total	0.58	0.51	1.21	2.27-						

Table 6.2: Calculated Variables in Experiment 1.3

This table above gives us a lot of insight in the averages and the difference between the maximum value and the minimum value of the hedge ratio for each bucket. It seems that

there are differences but for the buckets with lower spreads the differences are smaller than for groups with higher spreads which seem to be more volatile. For the first 3 spread buckets we differences in the maximum and minimum value compared to the averages that are about 14% . It seems that we can continue with experiment 2 with enough insight in how to generate a fitness function based on those results which are relatively stable.

6.2 Experiment 2 - Fitness Function Evaluation

Experiment 2.1 - Determining the Fitness of the Currently used Table

As already explained in Chapter 3.5 we came up wit the following Fitness Function

$$TF = \sum_{p=1}^p \sum_{b=1}^b | \Delta r_{p,b} | \cdot OAD_{p,b} \cdot MV - \%_{p,b} \cdot | \Delta HR |_b$$

where

TF = Total Fitness

$| \Delta r |$ = absolute delta R, monthly change in government yield

OAD = Option Adjusted Duration

$MV - \%$ = Market Value-% of this bucket

$| \Delta HR |$ = Maximum Hedge-ratio minus the Minimum Hedge-ratio for a bucket

p = period

b = bucket

When we applied this function to the MS-Table 1.2

		Spread level (basis points)					
		0 - 50	50 - 100	100 - 150	150 - 250	250 - 400	400+
Maturity (years)	0 - 3	1*	2	3	4	5	6
	3 - 5	7	8	9	10	11	12
	5 - 7	13	14	15	16	17	18
	7 - 10	19	20	21	22	23	24
	10+	25	26	27	28	29	30

Maturity/Spread buckets. This table shows the Maturity-Spread split up of the bond universe. Within this work this is called an MS-Table. *This is the first bucket. The table contains.30 buckets in total.

we found out that the this MS-Table has a fitness value of 18.9%. We can interpret this that the sum of all absolute performances when we would have used this table to determine our hedges would be 18.9%. The closer to 0%, the better the resulting hedge would be.

Experiment 2.2 - Iterative Evaluation of different Tables

Iterative Search results:

An example of the results are the following:

Spread	Duration	TOTAL PAIN	PAIN CLEAN	Recall
100	1.00	0.158	0.158	TableToFit([100 100 100 100 100 1 1 1 1])
100	1.25	0.200	0.200	TableToFit([100 100 100 100 100 1.25 1.25 1.25 1.25])
100	1.50	0.174	0.174	TableToFit([100 100 100 100 100 1.5 1.5 1.5 1.5])
100	1.75	0.159	0.159	TableToFit([100 100 100 100 100 1.75 1.75 1.75 1.75])
100	2.00	0.142	0.142	TableToFit([100 100 100 100 100 2 2 2 2])
100	2.25	0.155	0.155	TableToFit([100 100 100 100 100 2.25 2.25 2.25 2.25])
100	2.50	0.165	0.165	TableToFit([100 100 100 100 100 2.5 2.5 2.5 2.5])
100	2.75	0.157	0.157	TableToFit([100 100 100 100 100 2.75 2.75 2.75 2.75])
100	3.00	0.165	0.165	TableToFit([100 100 100 100 100 3 3 3 3])
100	3.25	0.182	0.182	TableToFit([100 100 100 100 100 3.25 3.25 3.25 3.25])
100	3.50	0.162	0.162	TableToFit([100 100 100 100 100 3.5 3.5 3.5 3.5])
100	3.75	0.171	0.171	TableToFit([100 100 100 100 100 3.75 3.75 3.75 3.75])
100	4.00	0.161	0.161	TableToFit([100 100 100 100 100 4 4 4 4])
100	4.25	0.166	0.166	TableToFit([100 100 100 100 100 4.25 4.25 4.25 4.25])
100	4.50	0.140	0.140	TableToFit([100 100 100 100 100 4.5 4.5 4.5 4.5])
100	4.75	0.135	0.135	TableToFit([100 100 100 100 100 4.75 4.75 4.75 4.75])
100	5.00	0.174	0.174	TableToFit([100 100 100 100 100 5 5 5 5])
100	5.25	0.167	0.167	TableToFit([100 100 100 100 100 5.25 5.25 5.25 5.25])
100	5.50	0.136	0.136	TableToFit([100 100 100 100 100 5.5 5.5 5.5 5.5])
100	5.75	0.123	0.123	TableToFit([100 100 100 100 100 5.75 5.75 5.75 5.75])
100	6.00	0.131	0.131	TableToFit([100 100 100 100 100 6 6 6 6])

Table 6.4 - This table shows several results of the total fitness found when performing the Iterative Search based on certain fixed spread and duration intervals starting with 0 for both characteristics.

The total result table consists of 1197 different fitness values and to get a grip on this amount we looked at where the optimum was based on the spread split up and where this was based on the maturity split up.

The following average fitness was found for the different MS-tables clustered for different duration intervals:

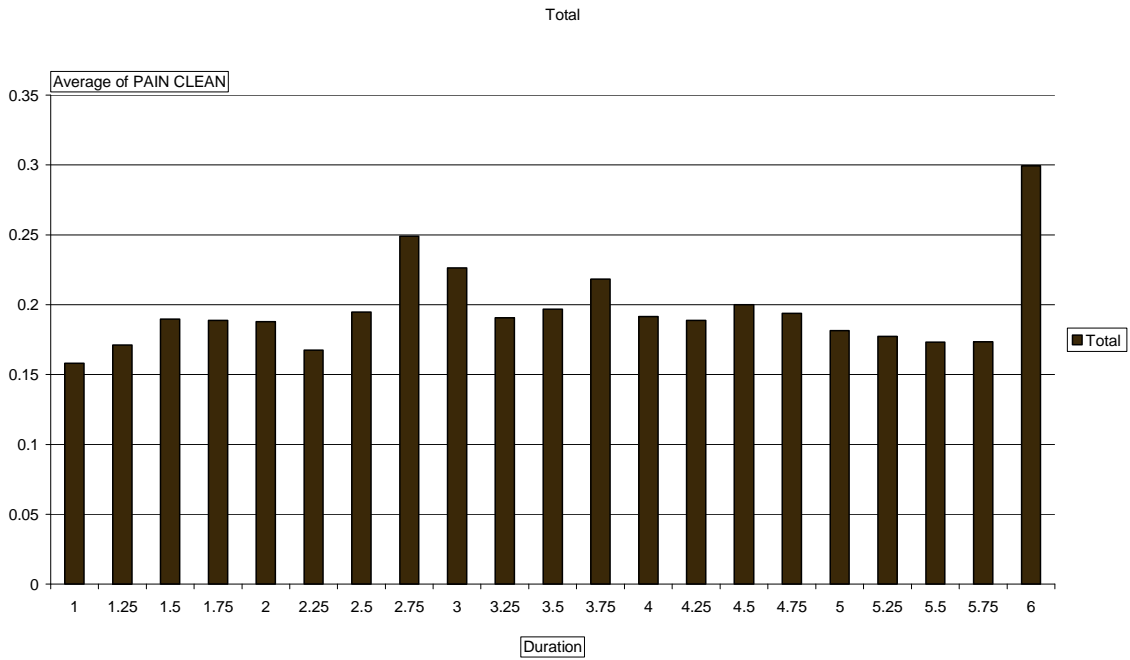


Figure 6.13 - Average fitness for different duration break ups for the buckets

The Average was found for durations of 2.25 but the best individual had a duration step of 5.25. There is not a real pattern.

Looking at spread intervals we find the following table:

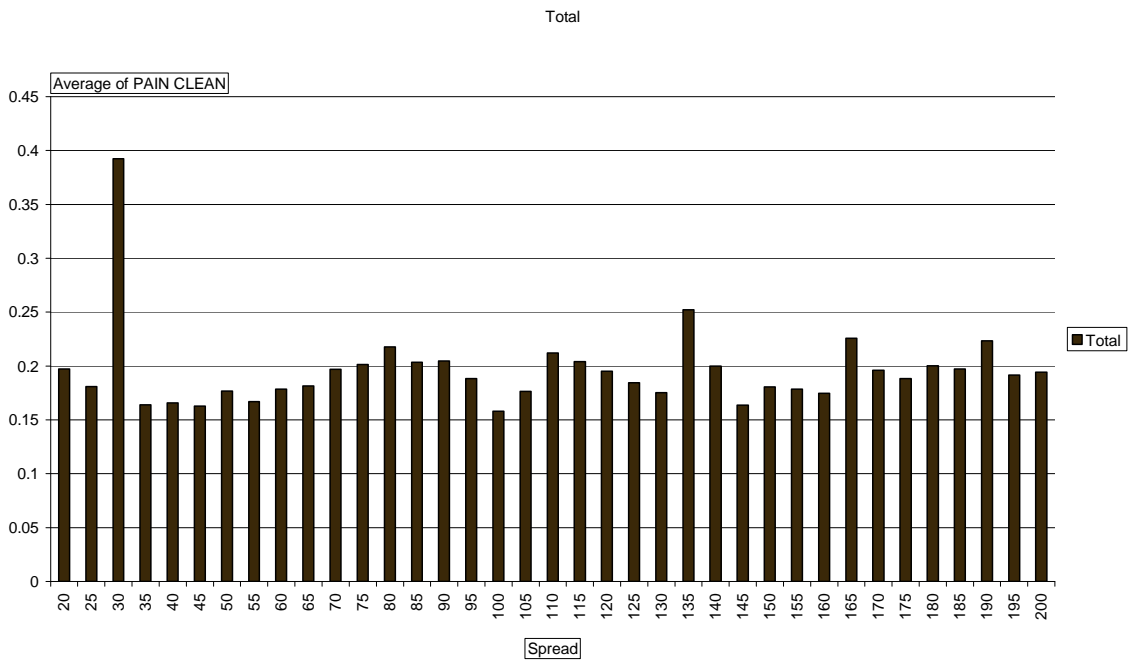


Figure 6.14 - Average fitness found for different fixed spread break ups for the buckets

The minimum average is found at 100.

The above 2 figures, Figure 6.13 and Figure 6.14, do not lead to any direct conclusions. It seems that both characteristics are needed to split up the universe combining both characteristics.

The table with the lowest fitness was with a fixed spread interval of 105 and a fixed maturity interval of 5.25. So (the rounded) found best MS-table was:

		Spread level (basis points)					
		0 - 105	105 - 210	210 - 315	315 - 420	420 - 525	525+
Maturity (years)	0 - 5	1	2	3	4	5	6
	5 - 11	7	8	9	10	11	12
	11 - 16	13	14	15	16	17	18
	16 - 21	19	20	21	22	23	24
	21+	25	26	27	28	29	30

Table 6.5 - The best possible spread table found with fixed divisions for the spread/maturity buckets. Fitness: 0.1179

The fitness of this table was 0.1179. This is a 37.6% better fitness compared to the found 18.9% with the initial MS-Table.

Within experiment 2.3 we found even better tables, this means that the GA search method was able to find tables with an even better fitness.

Experiment 2.3 - Genetic Algorithm Free Search

The total free GA run came with a fitness of 0.098. This was a very interesting result as the algorithm found a solution that was 16.9 percent points better than the best Iterative Solution. This gives a good indication that we use good GA settings and that our GA is capable of going through the search space and finding solutions. In the next Run we will use the start-points of the Iterative search.

The found table in Experiment 2.3:

		Spread level (basis points)					
		0 - 0	0 - 136	136 - 195	195 - 255	255 - 508	508+
Maturity (years)	0 - 8	1	2	3	4	5	6
	8 - 16	7	8	9	10	11	12
	16 - 26	13	14	15	16	17	18
	26+	19	20	21	22	23	24

Table 6.6 - Table found with the Free Search - Fitness: 0.098 - 48% better than the original table -

Experiment 2.4 - Genetic Algorithm Search - Starting points from Experiment 2.2

All different MS-tables used starting the fittest genes found in the iteration:

Spread 0	Spread 2	Spread 3	Spread 4	Spread 5	Spread 6	Maturity 1	Maturity 2	Maturity 3	Maturity 4	Maturity 5	Fitness
0 - 105	105 - 210	210 - 315	315 - 420	420 - 525	525+	0 - 5	5 - 11	11 - 16	16 - 21	21+	0.120
0 - 0	0 - 97	97 - 199	199 - 292	292 - 510	510+	0 - 0	0 - 10	10 - 11	11 - 15	15+	0.102
0 - 96	96 - 241	241 - 501	501 - 688	688 - 848	848+	0 - 1	1 - 1	1 - 9	9 - 9	9+	0.100
0 - 105	105 - 210	210 - 315	315 - 420	420 - 531	531+	0 - 4	4 - 4	4 - 8	8 - 8	8+	0.113
0 - 105	105 - 105	105 - 210	210 - 632	632 - 945	945+	0 - 0	0 - 5	5 - 5	5 - 16	16+	0.106
0 - 0	0 - 121	121 - 188	188 - 203	203 - 456	456+	0 - 7	7 - 10	10 - 10	10 - 16	16+	0.095
0 - 105	105 - 210	210 - 315	315 - 480	480 - 688	688+	0 - 5	5 - 11	11 - 16	16 - 21	21+	0.109
0 - 114	114 - 147	147 - 217	217 - 479	479 - 630	630+	0 - 5	5 - 5	5 - 9	9 - 9	9+	0.103
0 - 105	105 - 239	239 - 424	424 - 529	529 - 608	608+	0 - 0	0 - 3	3 - 10	10 - 14	14+	0.108
0 - 147	147 - 339	339 - 413	413 - 489	489 - 522	522+	0 - 7	7 - 10	10 - 13	13 - 14	14+	0.120
0 - 4	4 - 104	104 - 293	293 - 492	492 - 653	653+	0 - 0	0 - 0	0 - 0	0 - 9	9+	0.108
0 - 110	110 - 220	220 - 337	337 - 447	447 - 615	615+	0 - 8	8 - 13	13 - 13	13 - 13	13+	0.099
0 - 89	89 - 204	204 - 386	386 - 403	403 - 518	518+	0 - 7	7 - 10	10 - 16	16 - 19	19+	0.107
0 - 111	111 - 207	207 - 446	446 - 446	446 - 630	630+	0 - 0	0 - 5	5 - 5	5 - 10	10+	0.094
0 - 116	116 - 210	210 - 335	335 - 515	515 - 630	630+	0 - 7	7 - 13	13 - 21	21 - 29	29+	0.124
0 - 115	115 - 230	230 - 295	295 - 410	410 - 562	562+	0 - 1	1 - 8	8 - 14	14 - 20	20+	0.113
0 - 130	130 - 260	260 - 390	390 - 526	526 - 526	526+	0 - 2	2 - 4	4 - 7	7 - 9	9+	0.112
0 - 135	135 - 270	270 - 445	445 - 580	580 - 715	715+	0 - 1	1 - 3	3 - 4	4 - 5	5+	0.115
0 - 104	104 - 311	311 - 311	311 - 492	492 - 701	701+	0 - 0	0 - 0	0 - 1	1 - 8	8+	0.107
0 - 106	106 - 252	252 - 726	726 - 731	731 - 786	786+	0 - 18	18 - 22	22 - 26	26 - 26	26+	0.110
0 - 118	118 - 267	267 - 493	493 - 663	663 - 749	749+	0 - 0	0 - 0	0 - 10	10 - 10	10+	0.096
0 - 117	117 - 177	177 - 344	344 - 506	506 - 642	642+	0 - 6	6 - 6	6 - 8	8 - 20	20+	0.102
0 - 0	0 - 145	145 - 286	286 - 431	431 - 576	576+	0 - 5	5 - 9	9 - 18	18 - 18	18+	0.115
0 - 107	107 - 195	195 - 507	507 - 507	507 - 618	618+	0 - 0	0 - 8	8 - 8	8 - 19	19+	0.090

Table 6.7 - Genetic Algorithm Free Search - 24 fittest Starting Points from the Iterative Search - Highest fitness: 0.090362 (52% better than the original table)

The best found MS-Table looks the following:

		Spread level (basis points)					
		0 - 107	107 - 195	195 - 507	507 - 507	507 - 618	618+
Maturity (years)	0 - 0	1	2	3	4	5	6
	0 - 8	7	8	9	10	11	12
	8 - 8	13	14	15	16	17	18
	8 - 19	19	20	21	22	23	24
	19+	25	26	27	28	29	30

Table 6.8 - The best performing Maturity Spread table - Fitness: 0.090

This is an interesting table, because the GA finds a table where there is added value in removing 2 maturity split ups for the first and the third bucket for the maturity. We end up splitting up in 3 maturity buckets instead of 5 different maturity buckets. The same is true for the spread.

The found final table contains less buckets for spread and for duration and is more accurate than the previously table based on our definition of fitness definition:

		Spread level (basis points)				
		0 - 107	107 - 195	195 - 507	507 - 618	618+
Maturity (years)	0 - 8	1	2	3	4	5
	8 - 19	6	7	8	9	10
	19+	11	12	13	14	15

Table 6.9 - The best performing Maturity Spread table simplified - Fitness: 0.090

This is an amazing result as we reduced the complexity of the table with 50%, from 30 buckets to 15 buckets. and the fitness improved from 18.9% to 9.04% which is an reduction of 52.2% in hedging quality and this is a very good result. The last group (19+) is does not contain a lot of bonds and market value, but it was needed. If we would omit this small group and combine it with the second duration group we would be left with mainly 10 active groups instead of the 30 we started with. This is a very general solution which we could put effort in investigating further as this might really add value to the total hedging process.

6.4 Overview of Results

6.4.1 Experiment 1, empirical dependencies

The table below shows the average hedge-ratio over time.

We found out that the highest R-squared was found for groups with a low spread. The first 2 columns that have the lower R-squared values and the spread seems the dominant factor in splitting up the universe. When spread increases we see that R-squared is decreasing and that hedge-ratios also decrease for those groups. The real big decline in the hedge ratio occurs when the spread reaches a level over 150 – 250 so bigger than 225 on average. Table 6.1 is used to come up with Table 6.10

Bucket Number

Maturity	Spread level (basis points)					
	0 - 50	50 - 100	100 - 150	150 - 250	250 - 400	400+
0 – 3	1	2	3	4	5	6
3 – 5	7	8	9	10	11	12
5 – 7	13	14	15	16	17	18
7 – 10	19	20	21	22	23	24
10+	25	26	27	28	29	30

R-Squared Rank

Maturity	Spread level (basis points)					
	0 - 50	50 - 100	100 - 150	150 - 250	250 - 400	400+
0 – 3	2	5	15	20	25	27
3 – 5	7	4	13	19	24	26
5 – 7	3	8	11	18	22	30
7 – 10	1	6	12	17	23	28
10+	9	10	14	16	21	29

Top 10	Next 5	Next 5	Next 5	Next 5
--------	--------	--------	--------	--------

Table 6.10 - visualization of an MS-Table

One of the conclusions that can be drawn from those results is the following is that for the lower spreads the model works better. This is true for one MS-Table but we can also draw this conclusions when generating many tables. Therefore we might for a future research not try to optimize the total universe but draw a line somewhere between 150 and 250 and investigate the form and fitness of the resulting hedges.

6.4.2 Experiment 2 - MS-Tables

MS-Table results:

It is interesting to see that table 6.11:

		Spread level (basis points)					
		0 - 4	4 - 104	104 - 293	293 - 492	492 - 653	653+
Maturity (years)	0 - 9	1	2	3	4	5	6
	9+	7	8	9	10	11	12

Table 6.11 - A Maturity Spread table with only 12 buckets. Clean but still has a very low fitness.

Table 6.11 results in a fitness of 10.8 (42.9% increase of performance of the hedge) although only has one split-up in Maturity. The split-up made in spread is a pattern found in almost every table. This means that the split-up in spread does have an important role, which we already found in our previous conclusions in experiment 1.

6.6 Conclusions

We used a genetic algorithm to evaluate and improve the quality of the maturity-spread table (MS-Table) that has been developed by De Backer. We are able to reduce the number of maturity-spread buckets from 30 to 15. In addition the effective duration values of the new buckets are optimized. The new MS-Table has a 52% better fitness function than the MS-Table from De Backer. The new MS-Table is expected to deliver a better hedge of undesired interest rate to a credit investor. This is a short but promising conclusion.

Considering the fitness increase compared to the previously used MS-Table the chosen approach to change the form of the MS-Table seems to have a big added value. The best table we came up with is Table 6.9 the following table:

		Spread level (basis points)				
		0 - 107	107 - 195	195 - 507	507 - 618	618+
Maturity (years)	0 - 8	1	2	3	4	5
	8 - 19	6	7	8	9	10
	19+	11	12	13	14	15

The best performing Maturity Spread table simplified - Fitness: 0.090

It would be worth while to investigate the exact amount of basispoints we could earn using this hedging approach compared to the currently used hedge approach.

Chapter 7 – Follow-up research

We have developed an improved model and tested this model for several periods in time. On basis of the research we have done several follow-up research ideas arise:

- Compare results for even more buckets. We could even analyze those results and try to find regimes which could be linked to other variables within the economy right now to come to a multi regime model.

- Simulate the construction of investment portfolios in the past and analyze how well the hedge ratios performs. We could create excess return numbers and use real historical data to really test the above tables and publish and analyze the number of basispoints that could be won with this new approach.

- Robustness analysis of the history length and simulate with different results to find the best workable hedge ratio. Should this be a historical length 60 months or would a shorter period come with better hedging tables?

- Government bond yields and credit spreads are know to have a negative correlation (Duffee (1998)). The influence of this negative correlation on the hedge ratio is an interesting aspect to further investigate and combine with the above research.

- Determine the optimal update frequently of a hedge-ratio compared to the reduction of Duration risk is generates.

- Investigate different MS-Tables for different markets.

- Would the Genetic Algorithm produce even more important MS-tables when we would only take bonds with a spread lower than a certain spread, for instance a spread of 200 basis points.

- Finally, we could enhance the search method by making the search space more flexible and fuzzy instead of the currently chosen crisp approach.

Appendix A) Matlab Code

```
% Manual Steps before the run of this script (at Robeco)
% Remove Q path
% Save
% Done.
%
clear;

load 'bonddata-clean';

%de variabelen die in regressie waarden worden opgeslagen
variabele_namen =
{'start_date';'end_date';'mat_spread_bucket(1,1)';'mat_spread_bucket(1,2)';'mat_spread_bucket(2,
1)';'mat_spread_bucket(2,2)';'R-squared';'Rbar-squared';'Sigma^2';'Durbin
Watson';'Nobs';'Nvars';'Variable 1 Coefficient';'Variable 1 Coefficient-statistic';'Variable 1 t-
probability';'Variable 2 Coefficient';'Variable 2 Coefficient-statistic';'Variable 2 t-
probability';'Average OAS';'Average TR';'Average ER';'Average OAD';'Average
DeltaR';'Average Yield';'Average TRminusY0'};

for loop = 1:1:131
    %kies een start datum. 71 staat voor de 71e maand. (dus voor 1-8-94)
    start_date = loop;

    %end_date = cols(available_mat);
    % kies een eind datum. 130. 130 staat voor 2-8-1999
    end_date = start_date+59;
    %bereken het aantal maanden gegevens dat gebruikt gaat worden in de analyse

    disp('start simulation :')

    disp('start date')
    disp(start_date)
    %[SUCCESS,MESSAGE] = xlswrite('matlaboutput.xls',start_date,[int2str(start_date) '-'
int2str(end_date)],'B1');

    disp('end date')
    %[SUCCESS,MESSAGE] = xlswrite('matlaboutput.xls',end_date,[int2str(start_date) '-'
int2str(end_date)],'B2');
    disp(end_date)

    % rename the matrices
    OAS = SPREAD_Matrix;
    OAD = OADUR_Matrix;
    %ExcessReturn = EXRET_Matrix;
    Maturity = MATURITY_Matrix;
    Rating = RATING_Matrix;
    Sector = SECTOR_Matrix;
```

```

Source = SOURCE_Matrix;
TotalReturn = TOTRET_Matrix;
Yield = YIELD_Matrix;
MarketValue = MV_Matrix;

% tel voor elke maand hoeveel HY bonds er zijn en hoeveel IG bonds, de
% grens ligt op een spread van 500

%zet alle tellers op 0
%hyteller=zeros(size((Rating),2),1);
hyteller=zeros(size((Rating),2),1);
%igteller=zeros(cols(Rating),1);
igteller=zeros(size((Rating),2),1);

for j=1:size((Rating),2) % voor elke maand
    i=1;
    while i<= size((Rating),1) % ga elke rij af
        if isnan(Rating(i,j))==1 % als er niks staat in een rij, dan ga je naar de volgende rij
            i=i+1;
        else % er staat een rating, boven de 500 is het hy, onder de 500 is het ig
            if Rating(i,j) >= 500
                hyteller(j)=hyteller(j)+1;
                i=i+1;
            else if Rating(i,j) < 500
                igteller(j) = igteller(j)+1;
                i=i+1;
            end
        end
    end
end
end
end

% verwijder oude gegevens, ze zijn hernoemd tot betere namen
%clear MATURITY_Matrix EXRET_Matrix RATING_Matrix SECTOR_Matrix ...
% SOURCE_Matrix TOTRET_Matrix MV_Matrix YIELD_Matrix;

% matrix met de rente inlezen vanuit Excel. (mbv de maturity matrix)
% In deze matrix zitten 8 kolommen en 191 rijen. 191 maanden, en per maand
% de rente van: 3mnd, 6mnd, 1jr, 2jr, 3jr, 5jr, 10jr en 30jr

RENTE_Matrix = xlsread('RENTE JPM.xls',1,'N2:U192');

%Rente aanmaken in de goede size
Rente = Maturity;
%deltaRente ook aanmaken
deltaRente = Maturity;

%voor elke maand
for j = 1:size((Maturity),2)
    %stel de looptijd bucket in
    looptijd_bucket = [0,0.25,0.5,1,2,3,5,10,30,inf];

```

```

% maak voor deze maand een rente_buck variabele
rente_buck = [RENTE_Matrix(j+1,1),RENTE_Matrix(j+1,1:8),RENTE_Matrix(j+1,8)];
%interpoleer over de maturity om de Rente_maand_j te vullen
Rente_maand_j = interp1(looptijd_bucket,rente_buck,Maturity(:,j));
% vul de Rente matrix
Rente(:,j) = Rente_maand_j;

% vanaf de tweede maand kan deltaRente worden gevuld.
if j > 2
    deltaRente(:,j) = Rente(:,j) - Rente(:,j-1);
else % voor de eerste maand komt overal NaN te staan
    deltaRente(:,j) = NaN;
end
end

% -----
%           Maturity/Spread Buckets maken
% -----

% aantal_maanden = cols(available_mat);
% tel het aantal bonds dat in de database zit
aantal_bonds = size((available_mat),1);

aantal_maanden = end_date-start_date+1;
% de maturity bucket indeling
maturity_bucket = [0,3,5,7,10,inf];
% de coarse spread bucket indeling (grove indeling)
coarse_spread_bucket = [0,50,100,150,250,400,inf];
% other bucket distribution for extra analysis
% coarse_spread_bucket = [0,25,50,100,150,250,350,450,650,850,1000,inf];
% coarse_spread_bucket =
[0,50,100,150,200,250,300,350,400,450,500,550,600,650,700,750,800,850,900,inf];
% de coarse spread bucket indeling (fijne indeling)
fine_spread_bucket = [0:25:700 inf];
oas_bucket = 1; % 1 = coarse, 2 = fine

% wanneer gekozen is voor 1 (coarse)
if oas_bucket == 1
    % het aantal buckets berekenen. Het gaat om het aantal groepen, dus -1
    % van het aantal grenzen
    totaal_aantal_buckets = (length(coarse_spread_bucket)-1) * (length(maturity_bucket)-1);
else % wanneer gekozen is voor 2
    totaal_aantal_buckets = (length(fine_spread_bucket)-1) * (length(maturity_bucket)-1);
end

% variabele om te registreren of er een oas beschikbaar is
oas_beschikb = ones(190,1);
% variabele om te registreren of er een oad beschikbaar is
oad_beschikb = ones(190,1);

% momenteel: (30*60) --> 1800 bij 8

```

```

gegevens_buckets = ones(totaal_aantal_buckets*aantal_maanden,8);
aantal_in_laatste_spreadbucket = ones(190,1);
for j = start_date:end_date
    aantal_in_laatste_spreadbucket(j) = 0;
    % maak voor elke maand vectoren met de beschikbare gegevens
    % available_mat wordt dus aangepast naar beschikbaarheid van oad
    for ll = 1:length(available_mat(:,j))
        if available_mat(ll,j)==1 && isnan(OADUR_Matrix(ll,j))==1
            available_mat(ll,j)=0;
        end
    end
end

%een vector met enen op plekken welke gegevens beschikbaar zijn
%clear beschikbaar_maand_j;
beschikbaar_maand_j = available_mat(:,j)==1;

%beperking van beschikbare gegevens voor bepaalde sectoren en source
%sector moet zijn sector, 4, 11, 15, 18, 22, 23 (Electrical Utilities,
%Finance, Other Utilities, Telephone Utilities, Transportation) --> ask PH why??
% --> ask PH why??

%source: 2, 4 (US IG, US HY, so US research)
%rating: NOT 1000 or 2000
for ii = 1:length(beschikbaar_maand_j)
    if ((beschikbaar_maand_j(ii) == 1) && ...
        (Sector(ii,j) == 4 || Sector(ii,j) == 11 || Sector(ii,j) == 15 || ...
        Sector(ii,j) == 18 || Sector(ii,j) == 22 || Sector(ii,j) == 23) && ...
        (Source(ii,j) == 2 || Source(ii,j) == 4) && (Rating(ii,j) ~= 2000 && Rating(ii,j) ~=
1000))
        beschikbaar_maand_j(ii) = 1;
    else
        beschikbaar_maand_j(ii) = 0;
    end
end
%ook moet de deltaRente beschikbaar zijn voor die maand anders is het
%geen geldig datapunt
for jj = 1:length(beschikbaar_maand_j)
    if ((beschikbaar_maand_j(jj) == 1) && (isnan(deltaRente(jj,j)) == 1))
        beschikbaar_maand_j(jj) = 0;
    end
end

%eindelijk hebben we een selectie gemaakt van data die voldoet aan de
%eisen

% nu vullen we de variabelen voor deze maand
% vul OAS van deze maand
maandOAS = OAS(beschikbaar_maand_j,j);
%bekijk hoeveel oas data NaN is.
oas_beschikb(j) = sum(isnan(maandOAS));
% vul OAD van deze maand

```

```

maandOAD = OAD(beschikbaar_maand_j,j);
% bekijk hoeveel gegevens deze maand niet beschikbaar zijn
oad_beschikb(j) = sum(isnan(maandOAD));
% maandExcessReturn = ExcessReturn(beschikbaar_maand_j,j);
% vul Maturity van deze maand
maandMaturity = Maturity(beschikbaar_maand_j,j);
% vul Yield, TotalReturn, MarketValue, Rente, DeltaR van deze maand
maandYield = Yield(beschikbaar_maand_j,j);
maandTotalReturn = TotalReturn(beschikbaar_maand_j,j);
maandMarketValue = MarketValue(beschikbaar_maand_j,j);
maandRente = Rente(beschikbaar_maand_j,j);
maandDeltaR = deltaRente(beschikbaar_maand_j,j);
% bepaal het aantal gegevens van maandOAS.
aantal_beschikbaar_in_maand = length(maandOAS);

% deel voor deze maand de buckets in:

% voor elke maturity bucket...
for k = 1:length(maturity_bucket)-1

    % bepaal het aantal elementen van de bucket en vul spread_bucket
    % met deze gegevens

    % dit stuk code kan buiten de loop
    if oas_bucket == 1
        % disp('use coarse spread buckets')
        length_bucket = length(coarse_spread_bucket);
        spread_bucket = coarse_spread_bucket;
    else
        % disp('use fine spread buckets')
        length_bucket = length(fine_spread_bucket);
        spread_bucket = fine_spread_bucket;
    end

    % voor elke spread bucket
    for l = 1:length_bucket-1

        % (71e maand) dus  $70 \cdot 30 = 2100 + 1 + 0 = 2101 = n$ ,  $30 \cdot 190$  max,
        % dus een teller om aan te geven welke bucket het is (per 30
        % naar een nieuwe maand)

         $n = 1 + (k-1) \cdot (\text{length}(\text{spread\_bucket}) - 1) + (j-1) \cdot \text{totaal\_aantal\_buckets};$ 

        % kijk voor elke bond in die maand of deze in bucket (k,l) hoort
        bucket = ones(length(beschikbaar_maand_j),9);
        m = 0;
        for i = 1:aantal_beschikbaar_in_maand
            % als de maturity tussen (0 en 3 ligt als k = 1, 3 en 5 bij
            % k = 2, tussen 5 en 7 bij k = 3, tussen 7 en 10 bij k = 4,
            % en groter dan 10 bij k = 5.
            % EN

```

```

% de OAS valt ook binnen (0 50 100 150 250 400 99999)
% OF NU coarse_spread_bucket =
[0,25,50,100,150,250,350,450,650,850,1000,inf];
if ((maandMaturity(i) < maturity_bucket(k+1) && maandMaturity(i) >=
maturity_bucket(k)) && ...
(maandOAS(i) < spread_bucket(l+1) && maandOAS(i) >= spread_bucket(l)))
% valt binnen deze bucket
% disp('valt binnen bucket')
m = m+1;
%bucket(m,1:9) =
[maandMaturity(i),maandOAS(i),maandTotalReturn(i),maandExcessReturn(i),...
%
maandOAD(i),maandRente(i),maandDeltaR(i),maandYield(i),maandMarketValue(i)];

%ik sommeer alles in de bucket vector wat in deze
%bucket valt. bucket is uiteindelijk een m bij
%9 vector met alles binnen deze bucket
bucket(m,1:9) = [maandMaturity(i),maandOAS(i),maandTotalReturn(i),1,...

maandOAD(i),maandRente(i),maandDeltaR(i),maandYield(i),maandMarketValue(i)];
end
end
%verklein bucket (is eigenlijk niet nodig denk ik)
bucket = bucket(1:m,1:9);
%het aantal bonds in deze specifieke bucket
aantal_bonds_in_bucket = m;

% het aantal in de laatste spread bucket bijhouden van maand j
if l == length_bucket-1
aantal_in_laatste_spreadbucket(j) = aantal_in_laatste_spreadbucket(j) +
aantal_bonds_in_bucket;
end

% nu kunnen we voor deze bucket de gem looptijd, spread, TR etc berekenen.
% rij n: gem waarden van bucket n
% n=1:30 - alle buckets voor de 1e maand (de werkelijke 1e
% maand, niet de eerste maand van het onderzoek)
% n=31:60 - alle buckets voor de 2e maand etc

% bereken marktwaarde gewogen gemiddelden

Total_Bucket_MarketValue = sum(bucket(:,9)); % de som van alle marketvalues
gewichten = bucket(:,9)/Total_Bucket_MarketValue; % de procentuele gewichten
% minimum aantal bonds moet 10 zijn om meegenomen te kunnen
% worden in de analyse
if aantal_bonds_in_bucket >= 10

% bereken de marktwaarde gewogen gemiddelden
% average_mat = wmean(bucket(:,1),gewichten);
average_mat = sum(bucket(:,1) .* gewichten);
% average_OAS = wmean(bucket(:,2),gewichten);

```

```

        average_OAS = sum(bucket(:,2) .* gewichten);
    %average_TR = wmean(bucket(:,3),gewichten);
        average_TR = sum(bucket(:,3) .* gewichten);
    %average_ER = wmean(bucket(:,4),gewichten);
        average_ER = sum(bucket(:,4) .* gewichten);
    %average_OAD = wmean(bucket(:,5),gewichten);
        average_OAD = sum(bucket(:,5) .* gewichten);
    %average_rente = wmean(bucket(:,6),gewichten);
        average_rente = sum(bucket(:,6) .* gewichten);
    %average_deltaR = wmean(bucket(:,7),gewichten);
        average_deltaR = sum(bucket(:,7) .* gewichten);
    %average_yield = wmean(bucket(:,8),gewichten);
        average_yield = sum(bucket(:,8) .* gewichten);

    %average_MV = sum(bucket(:,8))/aantal_bonds_in_bucket;

    gegevens_buckets(n,1:8) = [average_mat, average_OAS,average_TR, average_ER,...
        average_OAD, average_rente, average_deltaR, average_yield];
    else
        gegevens_buckets(n,1:8) = [NaN,NaN,NaN,NaN,NaN,NaN,NaN,NaN];
    end
end
end
end
% De matrix gegevens_bucket is nu gevuld voor maand j.
% per bucket zijn de karakteristieken bekend van verschillende ter zake
% doende variabelen.
% dit is nu bekend voor achtereenvolgende buckets.
end
A=gegevens_buckets;
% }

% mat 1-3, spre 0-50: rij 1,31,61, etc.
% mat 1-3, spr 50-100: rij 2, 32, 2 etc.

% -----
%                Regressies uitvoeren
% -----

% kol1: maturity   kol2: OAS    kol3: TR    kolom 4: ER
% kol5: OAD        kol6: rente   kol7: deltaR  kolom 8: yield to worst

%
% 3a.  $TR - y_0/12 = B1 + B2 \cdot \text{deltaR} \cdot \text{OAD}$ 

disp('regressie 3a, hedge-ratio:  $TR - y_0/12 = B1 + B2 \cdot \text{deltaR} \cdot \text{OAD}$ ')
beta2 = ones(totaal_aantal_buckets,1);
R_squared = ones(totaal_aantal_buckets,1);
t_stat = ones(totaal_aantal_buckets,1);
sigma_squared = ones(totaal_aantal_buckets,1);
average_bucketOAS = ones(totaal_aantal_buckets,1);
average_bucketTR = ones(totaal_aantal_buckets,1);

```

```

average_bucketER = ones(totaal_aantal_buckets,1);
average_bucketOAD = ones(totaal_aantal_buckets,1);
average_bucketDeltaR = ones(totaal_aantal_buckets,1);
average_bucketYield = ones(totaal_aantal_buckets,1);

lengte = ones(totaal_aantal_buckets,1);
lhs = ones(175,totaal_aantal_buckets);

%% regressie waarden in een grote matrix zetten om vervolgens in excel
%% weg te kunnen schrijven per simulatie
regressie_waarden = ones(25,totaal_aantal_buckets);

for i = 1:totaal_aantal_buckets
    % disp('maturity/spread bucket: ')
    mat_spread_bucket = choose_bucket(i,maturity_bucket,spread_bucket);

    % vul regressie waarden
    % i gaat hier van 1 tot totaal_aantal_buckets
    regressie_waarden(1,i) = start_date;
    regressie_waarden(2,i) = end_date;
    regressie_waarden(3,i)= mat_spread_bucket(1,1);
    regressie_waarden(4,i)= mat_spread_bucket(1,2);
    regressie_waarden(5,i)= mat_spread_bucket(2,1);
    regressie_waarden(6,i)= mat_spread_bucket(2,2);

    % data = [OAS,TR,ER,OAD,deltaR,yield]
    data = make_data_matrix(i,gegevens_buckets,start_date,end_date);

    if length(data(:,1)) ~= 0
        average_bucketOAS(i) = sum(data(:,1))/length(data(:,1));

        % extra
        average_bucketTR(i) = sum(data(:,2))/length(data(:,2));
        average_bucketER(i) = sum(data(:,3))/length(data(:,3));
        average_bucketOAD(i) = sum(data(:,4))/length(data(:,4));
        average_bucketDeltaR(i) = sum(data(:,5))/length(data(:,5));
        average_bucketYield(i) = sum(data(:,6))/length(data(:,6));

        TRminusY0 = data(:,2) - data(:,6)/12;

        OASi = data(:,1);
        deltaR_times_OAD = data(:,5) .* data(:,4);
        deltaR_times_OAS_times_OAD = deltaR_times_OAD .* OASi;

        lhs(1:length(TRminusY0),i) = -TRminusY0 ./ deltaR_times_OAD;
        lhsi = lhs(1:length(TRminusY0),i);
        lengte(i) = length(lhsi);

        x1 = [ones(length(deltaR_times_OAD),1) deltaR_times_OAD];
        % x2 = [ones(length(deltaR_times_OAD),1) deltaR_times_OAD
        deltaR_times_OAS_times_OAD];

```



```

y = floor(4*(length(TRminusY0)/100)^(2/9));
%%%result3a = ols(TRminusY0,x); prt(result3a)

result3a = nwest(TRminusY0,x1,y);

% if result3a.rsqr ~= -Inf
  if result3a.rsqr >= -10

      % vul regressie waarden
      regressie_waarden(7,i)= result3a.rsqr;
      regressie_waarden(8,i)= result3a.rbar;
      regressie_waarden(9,i)= result3a.sige;
      regressie_waarden(10,i)= result3a.dw;
      regressie_waarden(11,i)= result3a.nobs;
      regressie_waarden(12,i)= result3a.nvar;

      regressie_waarden(13,i)= result3a.beta(1);
      regressie_waarden(14,i)= result3a.tstat(1);
      regressie_waarden(15,i)= 99;
      regressie_waarden(16,i)= result3a.beta(2);
      regressie_waarden(17,i)= result3a.tstat(2);
      regressie_waarden(18,i)= 99;

      regressie_waarden(19,i)= average_bucketOAS(i);
      regressie_waarden(20,i)= average_bucketTR(i);
      regressie_waarden(21,i)= average_bucketER(i);
      regressie_waarden(22,i)= average_bucketOAD(i);
      regressie_waarden(23,i)= average_bucketDeltaR(i);
      regressie_waarden(24,i)= average_bucketYield(i);

      regressie_waarden(25,i)= sum(TRminusY0(:,1))/length(TRminusY0(:,1));

      %%%result3b = nwest(TRminusY0,x2,y); prt_reg(result3b)
      beta2(i) = result3a.beta(2);
      R_squared(i) = result3a.rsqr;
      sigma_squared(i) = result3a.sige;
      t_stat(i) = result3a.tstat(2);
      % end of zero test
    end

  end

end

[SUCCESS,MESSAGE] =
xlswrite('matlaboutput_run1.xls',regressie_waarden,[int2str(start_date) '-' int2str(end_date)],'B2');
end

```


Glossary

Yield:

The yield on a bond is its annual return, affected by the price the buyer pays for it.

Spread:

The spread is the number of basis points the government curve would have to shift for the present value of the cash flows to equal the security's price. Spread is usually measured in basis point which is one-hundredth of a percent. If the yield of a bond is one percent higher than a comparable government bond the spread will be 100 basis points.

Duration:

Also known as Macauley duration is the weighted average maturity of the security's cash flows, where the present values of the cash flows serve as the weights. The greater the duration of a security, the greater its percentage price change given change in yield.

Excess Return:

A security's return minus the return from a no-risk security during the same time period. The "risk-free rate of return" is generally the rate of return of a national government issue.

Maturity data:

Maturity date is the date on which a debt becomes due for payment.

Commonly used Abbreviations:

Spread – OAS

Duration – OAD

Maturity – M

Yield – y

Yield to Worst – ytw

Hedge-ratio – HR

Price – P

Total Return – TR

Period – t

Bibliography

- [1] Ben Dor, A., Dynkin, L., Konstantinovskiy, V., 2004, Empirical Duration of High Yield Credit, Report Quantitative Portfolio Strategy, Lehman Brothers.
- [2] Ben Dor, A., Dynkin, L., Hyman, J., 2005, Empirical Duration of Credit Securities: Dependence on Spread, Report Quantitative Portfolio Strategy, Lehman Brothers.
- [3] Ben Dor, A., Dynkin, L., Hyman, J., Houweling, P., Van Leeuwen, E. and O. Penninga, 2007. DTS (Duration Times Spread). *Journal of Portfolio Management*, Winter 2007, Vol. 33, No. 2, 77–100.
- [4] Bierwag, G.O., Kaufman, G.G. and A. Toevs, Duration: Its Development and Use in Bond Portfolio Management. *Financial Analysts Journal*, Vol. 39, No. 4 (Jul. - Aug., 1983), pp. 15-35.
- [5] De Backer, B., 2006, Interest Rate Risk of Corporate Bonds. Final Report Internship at Quantitative Strategies Department of Robeco.
- [6] Duffee, G.R., 1998. The Relationship Between Treasury Yields and Corporate Bond Yield Spreads. *Journal of Finance* 53 (6), 2225-2241.
- [7] Fabozzi, F.J., 2000, *The Handbook of Fixed Income Securities*, McGraw-Hill, 6th edition (August 24, 2000)
- [8] Jarrow, R.A., 1978. The Relationship Between Yield, Risk, and Return of Corporate Bonds, *Journal of Finance* 33 (1978), 1235-1240.
- [9] Jarrow, R.A., Turnbull, S., 2000. The Intersection of Market and Credit Risk. *Journal of Banking & Finance* 24 (2000) 271-299.
- [10] Macaulay, Frederick R., 1938, *Some Theoretical Problems Suggested by the Movements of Interest Rates, Bond Yields, and Stock Prices in the United States since 1856*, New York, NY, National Bureau of Economic Research, 1938. (Partially reprinted in Leibowitz 1980).
- [11] The Mathworks, 2009, *Genetic Algorithm and Direct Search Toolbox 2 - User's Guide*.
- [12] Fridson, M., and J. Kenney., “How do Changes in Yields Affect Quality Spreads?” *Extra Credit* (July/August 1994), pp. 4–13.
- [13] Lin, M., and J.C. Curtillet., Another Look at the Relation Between Credit Spreads and Interest Rates, *Journal of Fixed Income* (2007) 17, 59-71.

[14] Belongia, M.T., and G.J. Santoni., 1984, Hedging Interest Rate Risk with Financial Futures: Some Basic Principles.

[15] Mentink, A., 2005, Essay on Corporate Bonds, , 22-24.

[16] Black, F. and Scholes, M., 1973, The pricing of options and corporate liabilities, *Journal of Political Economy* 81, 637-654

[17] Marton, R.C., 1974, On the pricing of corporate debt: The risk structure of interest rates, *Journam of IFinance* 29 (2), 449-469

[18] Dai, Q., and Singleton, K.J., 2003, Term structure dynamics in theory and reality, *Review of Financial Studies* 16 (3), 631-679

[19] Litterman, R. and Iben, T., 1991, Corporate bond valuation and the term structure of credit spreads, *Journal of Portfolio Management*, 52-64

[20] Houweling, P., Mentink, A. A. and Vorst, A.C.F., 2004, Valuing euro rating-triggered step-up telecom bonds, *The Journal of Derivatives* 11 (3), 63-80