

# Stochastic Scenario Generation for the Term Structure of Interest Rates

Arngrímur Einarsson

Supervisors:  
Jens Clausen  
Kourosh M. Rasmussen

Kongens Lyngby 2007

Technical University of Denmark  
Informatics and Mathematical Modelling  
Building 321, DK-2800 Kongens Lyngby, Denmark  
Phone +45 45253351, Fax +45 45882673  
[reception@imm.dtu.dk](mailto:reception@imm.dtu.dk)  
[www.imm.dtu.dk](http://www.imm.dtu.dk)

# Summary

---

In models of risk and portfolio management in the fixed income security market as well as in models of pricing of interest rate sensitive derivatives one should model the most likely future movements of the whole term structure of interest rates. A lot of work has been done on modeling interest rates for derivative pricing purposes. But when it comes to generating interest rate scenarios for managing the risk and return of fixed income securities the amount of work done is less developed. In particular when using multi stage stochastic programming the bottle neck in many cases seems to be capturing the interest rate uncertainty properly in accordance with the state of the art economic and financial assumptions.

The objective is therefore to construct a model capable of capturing the interest rates in order to generate interest rate scenarios.

The term structure of interest rates is modeled by using historical term structures. This historical data has several dimensions which will be reduced to a few key factors of the term structure using factor analysis.

When we have recognized these factors they are used to construct a stochastic factor model capable of describing the future movement of the term structure of interest rates. The model used for that purpose is a vector autoregression model.

Finally the Factor model is used as an input into a scenario generating system to generate scenarios and make some general observations and experiments on them.



# Preface

---

This thesis was written at Informatics and Mathematical Modeling at the Technical University of Denmark in partial fulfillment of the requirements for acquiring the degree, Master of Science in Engineering.

The project was completed under the supervision of Jens Clausen and Kourosh Marjani Rasmussen and was carried out in the period from February the 2nd to November the 30th 2007 and is credited for 35 ECTS points.

Lyngby, November 2007



# Acknowledgments

---

I thank my supervisors on this project; Kourosch Marjani Rasmussen and Jens Clausen, for the guidance in this work. Especially Kourosch who acted as my main guide in this work.

I would also like to thank Snorri Páll Sigurðsson for providing the code for drawing scenario trees along with reviewing of my writing and Sverrir Grímur Gunnarsson for reviewing my writing as well.





# Contents

---

<b>Summary</b>	<b>i</b>
<b>Preface</b>	<b>iii</b>
<b>Acknowledgments</b>	<b>v</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Stochastic Programming and Scenario Generation . . . . .	1
1.2 Available Data . . . . .	5
1.3 Outline of the Thesis . . . . .	6
<b>2 Factor Analysis</b>	<b>9</b>
2.1 The Term Structure of Interest Rates . . . . .	10
2.2 Factor Analysis of the Term Structure . . . . .	13
2.3 Application of Factor Analysis . . . . .	19
2.4 Conclusion . . . . .	23

---

<b>3</b>	<b>Normality of Interest Rates</b>	<b>27</b>
3.1	Introduction . . . . .	28
3.2	Normality Inspection . . . . .	32
3.3	Normality Versus Log-normal . . . . .	42
3.4	Conclusion . . . . .	51
<b>4</b>	<b>Vector Autoregression</b>	<b>53</b>
4.1	Stationary, Invertability and White Noise . . . . .	54
4.2	The Vector Autoregression Process . . . . .	57
4.3	Choosing the Factors in the VAR Model . . . . .	60
4.4	Analyzing the Order and Stability of a VAR Model . . . . .	61
4.5	Construction of a VAR Model . . . . .	64
4.6	Estimation of the Parameters in a VAR Model . . . . .	66
4.7	Conclusion . . . . .	68
<b>5</b>	<b>Scenario Tree Generation</b>	<b>69</b>
5.1	Scenarios and Scenario Trees . . . . .	70
5.2	The Quality of the Scenario Tree . . . . .	72
5.3	A Scenario Generation Model . . . . .	74
5.4	Test Case of Scenario Generation . . . . .	78
5.5	Conclusion . . . . .	90
<b>6</b>	<b>Conclusion</b>	<b>91</b>
6.1	Main findings . . . . .	91

---

6.2	Future work . . . . .	92
<b>A</b>	<b>Further Results</b>	<b>95</b>
A.1	PCA, the Eigenvectors for 1995–2006 . . . . .	95
A.2	PCA, Adding One Year at a Time . . . . .	97
A.3	PCA, Performed Before and After the Changeover to Euro . . . . .	100
A.4	Roots of $p$ th Order Difference Equations . . . . .	101
A.5	The Parameters Estimated For Scenario construction . . . . .	102
A.6	Tree Plots, $i$ Period, August 2005 . . . . .	106
A.7	Tree Plots, 1 Period, August 2007 . . . . .	114
<b>B</b>	<b>Code</b>	<b>123</b>
B.1	Data read In . . . . .	123
B.2	Principal Component Analysis . . . . .	126
B.3	Normality Inspection . . . . .	130
B.4	Vector Auto Regression . . . . .	134
B.5	Simple Arbitrage test . . . . .	144



# Introduction

---

## 1.1 Stochastic Programming and Scenario Generation

Managing portfolios of financial instruments is in essence managing the tradeoff between risk and return. Optimization is a well suited and frequently used tool to manage this tradeoff. Financial risks arise due to the stochastic nature of some underlying market parameters such as interest rates. So it is necessary to include stochastic parameters in optimization for portfolio managing, turning portfolio optimization into stochastic optimization or stochastic programming (SP). A vital part of SP in portfolio management is scenario generation which is the main subject of this thesis. In the next two sections a short overview is given of stochastic programming and scenario generation for the term structure of interest rates and the relations between them.

### 1.1.1 Stochastic programming

Whereas a deterministic optimization problem contains only known parameters, stochastic programming is an optimization problem containing uncertain parameters. When formulating a SP problem the uncertain parameters can either

be described by stochastic distributions, when working with a single period, or stochastic processes when working with multiple periods. As an example of a formulation of a SP problem we give a single-period SP formulation, taken from Kall & Wallace (1994):

$$\begin{aligned} \text{Min } & f_0(\mathbf{x}, \tilde{\boldsymbol{\xi}}) \\ \text{s.t. } & f_i(\mathbf{x}, \tilde{\boldsymbol{\xi}}) \leq 0, i = 1, \dots, m \\ & \mathbf{x} \in \mathbf{X} \subset \mathbb{R}^n. \end{aligned} \quad (1.1)$$

In 1.1 above  $f_0(\mathbf{x}, \tilde{\boldsymbol{\xi}})$  denotes the objective function,  $f_i(\mathbf{x}, \tilde{\boldsymbol{\xi}})$  denotes the constraints and  $\boldsymbol{\xi} = \boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \dots, \boldsymbol{\xi}_T$  is a vector of random parameters, over the time  $t = (1, 2, \dots, T)$ , whose distribution is independent of the vectors of the decision parameters  $\mathbf{X}$ . Note however that this formulation is incomplete for it neither specifies the constraints needed nor is the meaning of Min specified.

With the exception of some trivial cases formulation 1.1 can not be solved using a continuous distribution to describe the random parameters. That is due to the fact that in continuous setting the decisions parameters become functions, making the problem a functional optimization problem, which cannot be solved numerically as it is. The usual way of reducing the problem so it can be solved, is to restrict it to a discrete-state problem, so that the random vector  $\boldsymbol{\xi} = \boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_T$  take only finitely many values, i.e. the decision functions are reduced to decision vectors with finitely many values. This discrete distribution, containing limited number of possible outcomes, is called *scenarios*.

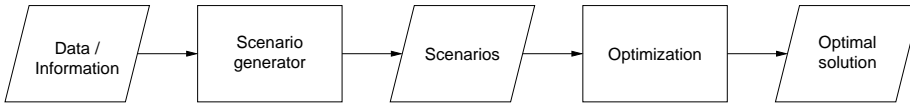


Figure 1.1: A digram showing the steps involved in the solving of a discrete stochastic programming optimization problem.

Solving a SP problem using scenarios is a multi step process. Figure 1.1 shows an abstract overview of that process. The input is some information relevant to the problem, usually in the form of some sort of data, but it can just as well be some other kind of information, such as an expert opinion, for example. Given the input the scenario generator is some sort of system which processes the input and returns the scenarios as an output. The scenarios then serve as a stochastic input into the optimization model, possibly along with some deterministic data, which finally returns an optimal solution of the problem.

Now if we treat the optimization part of the process shown in figure 1.1 as a black box device, and make the assumption that it finds the global optimal solution for *given* scenarios, then it is quite obvious that the optimal solution found is only as good as the scenarios generated allow it to be. Put differently, the

quality of the output of the optimization is directly dependent on the quality of the input or the scenarios generated.

Therefore the benefits of using a good scenario generator is quite obvious. Construction of a scenario generator, intended for generating scenarios of interest rates which could be beneficial for use in portfolio management in the the fixed income marked is the main subject of this work.

### 1.1.2 Scenario trees

Interest rate scenarios are usually displayed with so called *scenario trees*, an example of such a scenario tree can be seen in figure 1.2 which shows a multi period, scenario tree. In the figure the nodes represent the possible stages at each period and the arcs represent the relations of the stochastic variables. Each path through a scenario tree is a Scenario and a definition of a scenarios taken from *Practical Financial Optimization* (2005) is:

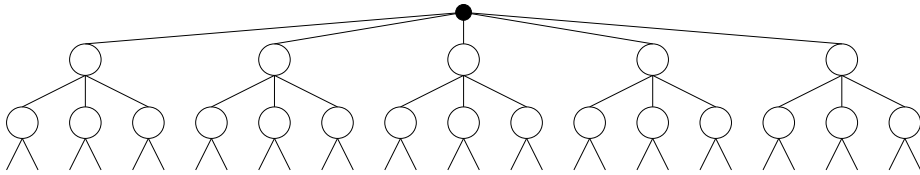


Figure 1.2: An example of a scenario tree.

#### DEFINITION 1.1 Scenarios.

A scenario is a value of a discrete random variable representing data together with the associated probability  $p^l \geq 0$ . Each scenario is indexed by  $l$  from a sample set  $\Omega$ , and the probabilities satisfy  $\sum_{l \in \Omega} p^l = 1$ .  $\diamond$

### 1.1.3 Overview of scenario generation methods

A general approach to generate scenarios is to take some information, believed to be representative of the problem which the aim is to model, and use them to generate scenarios. A typical form of information used are historical data observations. For our purposes historical data of interest rates are an obvious choice as a source of information.

It should be noted that there exists no general scenario generation approaches which can be applied for all stochastic programming models. Scenario generation is usually rather problem specific and therefore it is difficult to compare

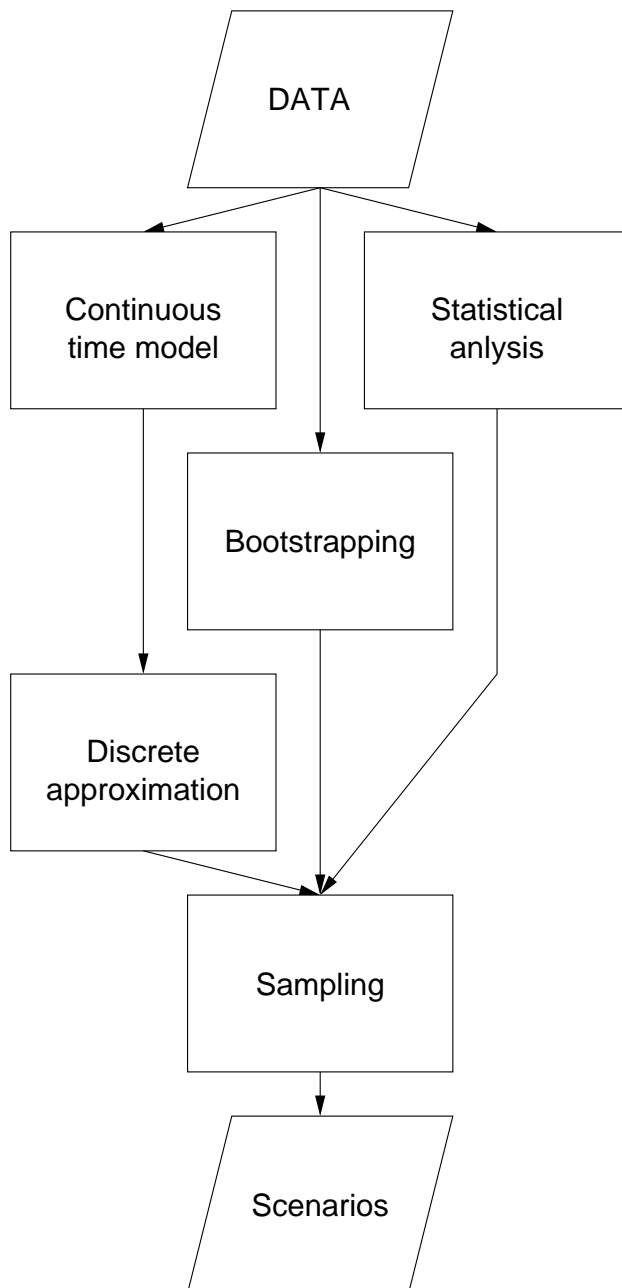


Figure 1.3: A digram showing several possibilities of generating scenarios, adapted from *Practical Financial Optimization* (2005).



the quality of scenario generation between different types of applications. But how to generate scenarios? Figure 5.1, shows three conventional ways of generating scenarios. The simplest of the methods shown is *bootstrapping*, which is the procedure of sampling observed data and use it as an direct input in to SP optimization. However a scenario generated with the bootstrapping method has the serious shortcoming that it can only reflect observations which have occurred before and is unable to come up with situations witch have not occurred, it lacks creativity, similar to learning something by rote.

To make up for the shortcomings of the bootstrapping method one can try to recognize the characteristics of the system instead of just mimicking past behavior. To do that some *statistical analysis* can be used to recognize the properties of the the underlying process. Those properties can then be used to generate scenarios having the same properties. The most common form of such statistical analysis is *moment matching*, where statistical moments of the underlying process, are found and then used to construct scenarios with matching moments, usually along with matching the correlation matrix. However generating scenarios using moment matching has some potential hazards as pointed out by Hocreiter & Pflug (2007). The hazards lie in the fact that different distributions can have the same moments, meaning that a scenario could be made out of completely different distributions than truly describe the underlying system which is being modeled. And as stated in their paper:

*“although moment matching performs better than crude random sampling and adjusted random sampling . . . it is obviously awkward to use this methodology in terms of reliability and credibility of the approximations”.*

An improvement to the moment matching is to develop a model of the underlying stochastic process, and then make a discrete approximation of that to sample scenarios from. Doing that the user can be sure that he is sampling from a process known to describe the system being modeled. That should address the reliability and credibility issues of moment matching.

## 1.2 Available Data

The historical data for the term structures of Danish interest rates for zero-coupon bonds was available. The data set covers the period from the 4. of January 1995 to the 8. of October 2007, issued with weekly intervals counting 659 issuing dates at all. Each issuing date contains the spot rates for maturities

up to thirty years in quarterly steps.

## 1.3 Outline of the Thesis

### 1.3.1 Layout of thesis

The rest of the thesis is organized out as follows.

#### **Chapter 2: Factor Analysis**

This chapter begins by covering the term structure of interest rate. Next a method for performing a factor analysis on the term structure is formulated and implemented, from which we find the factors which can be used to represent the term structure.

#### **Chapter 3: Normality of Interest Rates**

In this chapter the normality of the interest rates is tested, and the hypothesis that a log-normal distribution describes the data better is checked. The main result is the the log-normality assumption does not result in any benefits for the purpose of scenario generating. Therefore we use the data as it is.

#### **Chapter 4: Vector Autoregression**

In this chapter a VAR model is formulated for the purpose of modeling the term structure. It is investigated which order is suitable for the VAR model of the interest rates, which turns out to be order one. The stability of the model is also tested with positive results. A way to proxy for the factors with the rate data is derived and finally proxies for interest rate variability are derived.

#### **Chapter 5: Scenario Tree Construction**

In this chapter the construction of scenarios and scenario trees are covered in more depth than done in the introduction. The previous results are used as an input to a scenario generation system by Rasmussen & Poulsen (2007) to generate scenarios and look into how different approaches for the generation affect key issues such as existence of arbitrage and affects the number of scenarios has.

**Chapter 6: Conclusion**

Final overview of the results of this work along with elaborations of possible future work are given in this chapter.



## CHAPTER 2

# Factor Analysis

---

The first step in generating interest rate scenarios is to find some factors which describe the term structure of the rates and can serve as an input into an interest rate model. In this section a factor analysis is used to find the factors of use in the factor model of interest rates we wish to construct. The factor analysis is performed with data of Danish zero-coupon bonds, described in section [1.2](#).

The rest of the chapter is laid out as follows:

- In section [2.1](#) an overview over the term structure of interest rates is given.
- In section [2.2](#) an overview of the factor analysis, along with a formulation of it for the term structure of interest rates is given.
- In section [2.3](#) a factor analysis is performed on Danish yield curve data and the results analyzed.
- Finally section [2.4](#) concludes the chapter.

## 2.1 The Term Structure of Interest Rates

A *security* is a fungible financial instrument which represents a value. Securities are issued by some entity, such as a government or corporation, and they can be sub categorized as *debts*, such as bonds, or *equity*, such as common stock. Of particular interest to us is the term *fixed income securities* which refers to a specific kind of a financial instrument that yields a fixed income at a given time in the future, termed *maturity*. An example of fixed income instruments are bonds, where the issuer of the bond owes the holder a debt and is obliged to repay the face value of the bond, *the principal*, at the maturity possibly along with interests payments or *coupons* at specific dates prior to the maturity.

A fixed income securities which delivers no coupons is termed a *zero-coupon bond* (ZCB). Put differently a ZCB only delivers a single payment (the premium) when the bond reaches maturity. In an analytical sense, ZCB's are good to work with as they are the simplest type of bonds, but can however be used as building blocks for other types of fixed income securities. That is because it is possible to match other types of fixed income securities with a portfolio of ZCB's having different maturities which premiums are matched to the cash flow of the original ZCB's.

Changes on the term structure have direct opposite effects on the price of bonds. If the rates rise the prices of bonds fall and vice versa. The price of a fixed income security is the securities present value which is controlled by the interest rate termed as the *spot rate*. The concept "spot", used in financial sense, generally means buying or selling something upon immediate delivery and the concept applies in the same way for securities, meaning that the spot rate is simply the price of a security bought "on the spot". It is therefore easy to see why the price bond that pays fixed 5% interest is higher when the spot rate is 4% than when it is 6%. Formal definitions of spot rate and the term structure taken from *Practical Financial Optimization* (2005) are:

### DEFINITION 2.1 Spot Rate

The spot rate is the basic rate of interest charged for the risk free asset (cash) held during a period from time  $t = 0$  until some time  $t = \tau$ . We can think of the spot rate as the return on one unit of the risk free asset during the holding period  $\tau$  and denote it by  $r_{f\tau}$ .  $\diamond$

Next we define the *term structure of interest rates* which simply put is the relationship between interest rates and their time to maturity.

### DEFINITION 2.2 Term Structure of Interest Rates

Is the vector of spot rates for all holding periods  $t = 1, 2, \dots, T$ , denoted by  $(r_t)_{t=1}^T$ .  $\diamond$

If the term structure of interest rates is plotted the result is the the so called yield curve. An example of how yield curves look like can be seen in figure 2.1 which contains two instances of yield curves for Danish ZCB's from at two different historic time periods.

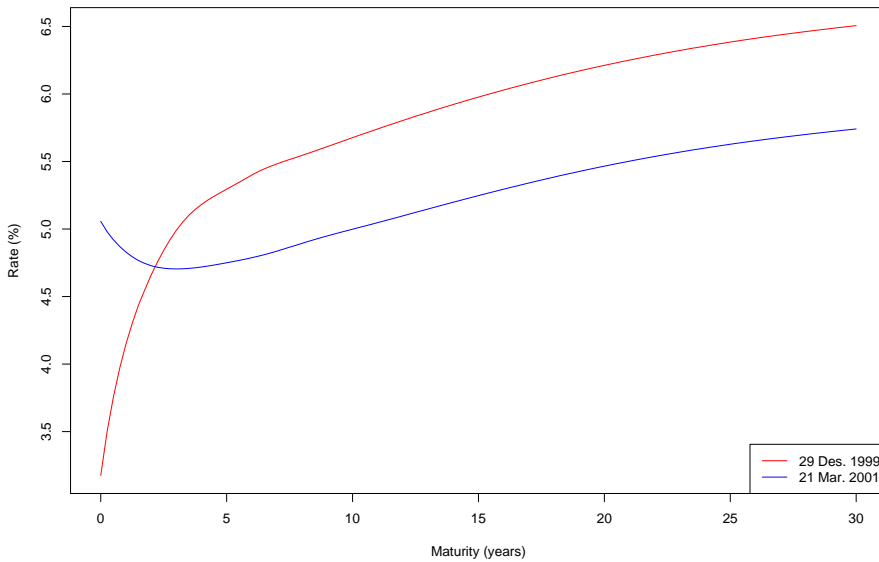


Figure 2.1: Yield curves for Danish zero-coupon bonds. The red curve is a normal shaped yield curve and the blue curve shows a yield curve where the short rate yield is inverted.

Yield curves can have various characteristics depending on economic circumstances at a given point in time. An upward sloping curve with increasing but marginally diminishing increases in the level of rates, for increasing maturities, is commonly referred to as a *normal shaped* yield curve. An example of such a curve is the red curve in figure 2.1. The reason for this naming is due to the fact that this is the shape of a yield curve considered to be normal for economically balanced conditions. Furthermore this shape has been the far most common for the past decades<sup>1</sup>.

<sup>1</sup>The normal shape has in fact been dominant in capitalized markets since the great depression.

Other types of yield curves include a *flat* yield curve where the yields are constant for all maturities. A *humped* shaped yield curve has short and long term yields of equal magnitude, different from the medium term yields which are consequently either higher or lower. An *inverted* yield curve is converted invert normal shaped curve, i.e. a downward sloping yield curve with decreasing but marginally diminishing decreases in yields.

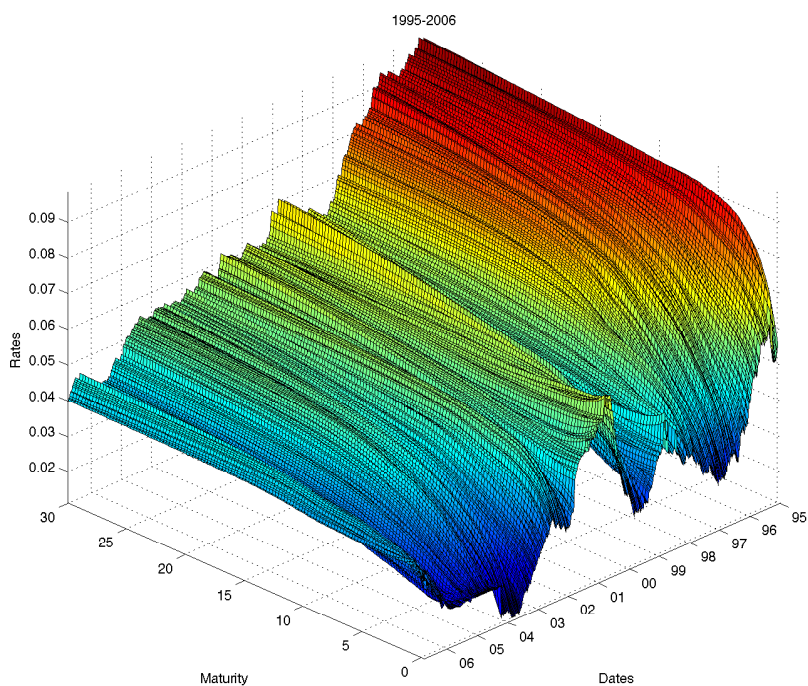


Figure 2.2: Historical data of Danish (zero-coupon) yield curves for the period 1995–2006.

Figure 2.2 shows a surface plot of Danish yield curves issued for the years 1995–2006. The plot simultaneously shows the yields plotted against time to maturity, and the yield of a given maturity plotted against issuing dates. From the figure, it can be observed that the yield curves are mostly normal shaped, with the exception of two short periods around the years 1999 and 2001.



## 2.2 Factor Analysis of the Term Structure

Now that we have described the term structure we turn our focus on how to model it. A simple procedure for modeling the term structure is the so called *parallel shift approach*, see e.g. *Options, Futures, and Other Derivatives* (2006). The parallel shifts approach is based on calculating the magnitude of a parallel shift of the yield curve caused by the change of the rate. This procedure however has the drawback that it does not account for non-parallel shifts of the yield curve, and

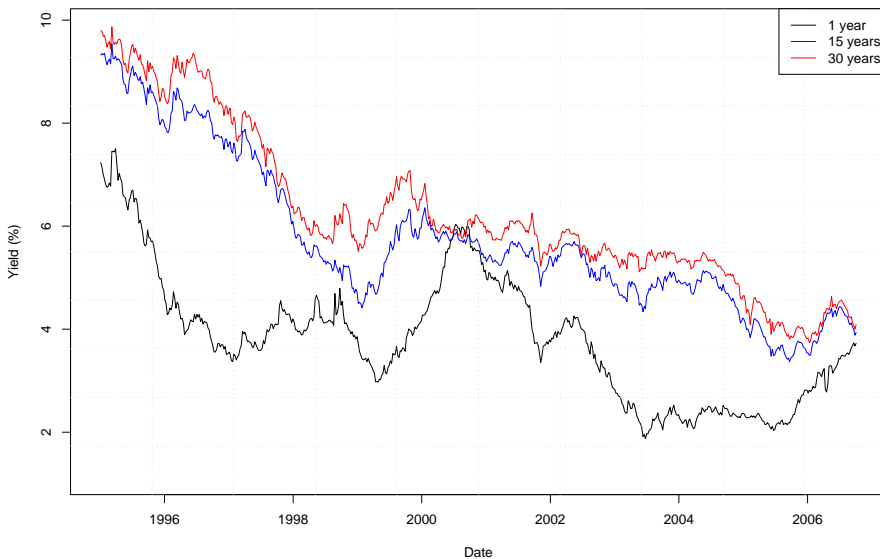


Figure 2.3: Short, medium and long term yields plotted for the same period as before.

as can be observed from figure 2.2 the parallel shift assumption simply does not hold. This can be further observed in figure 2.3, which gives cross-sections of the data shown in the preceding figure, for short medium and long term rates, from the figure it is evident that the yields are not perfectly correlated especially not the short and long term yields. Therefore we conclude that the yield curves evolve in a more complicated manner and a non parallel approach is needed.

A number of procedures are available to improve the parallel shift approach, such as dividing the curve into a number of sub periods, or so called buckets, and calculate the impact of shifting the rates in each bucket by one basis point

while keeping the rest of the initial term structure unchanged. Although the bucked approach leads to an improvement to the parallel shift approach it is still merely a patch on the parallel approach and still relies on the same assumption.

One commonly used method of modeling the term structure of interest rates, which does not rely on the parallel assumption, is to use Monte Carlo simulation to model the curve, based on some key rates used to describe the yield curve. According to the literature using a Monte Carlo simulation one can achieve better results than with the parallel assumption approach. However it has the disadvantages of high computational cost involving a huge number of trials, especially when working with multi currency portfolios, pointed out by Jamshidian & Zhu (1997), being . Furthermore the coverage of all “extreme” cases of the yield curve evolution is not guaranteed and the selection of the key interest rate is trivial often relying on arbitrary selected choices, making the quality of the simulation heavily dependent on those choices.

If historical data of the term structure is available another alternative is to investigate the internal relationship of the term structure. Such a method is called *factor analysis* which in general aims at describing the variability of a set of observed variables with a smaller set of unobserved variables, called *factors* or *principal components*. The factor analysis takes changes in the shape of the term structure in to account, allowing the parallel assumption approach to be relaxed.

Factor analysis has previously been applied in analysis of the term structure with great success, Litterman & Scheinkman (1991) find that the term structure of interest rates can be largely explained by a small number of factors. Performing factor analysis on data for US treasury bonds they find that about 95% of the the variation of the yield curve movements can be explained by just three factors which they name: *level*, *slope* and *curvature*. Level accounts for parallel shifts in the yield curve, affecting all the maturities with the same magnitude, slope describes changes in the slope of the yield curve and the curvature factor, accounts for change in the yield curve curvature.

Further applications of factor analysis on the term structure includes an analysis made on Italian treasury bonds by Bertocchi, Giacometti & Zenios (2000), considering yields with maturities up to 7 years, in that analysis the three most significant factors explained approximately 99% of the yield curve movement. Dahl (1996) found out that three factors were able to explained about 99.6% of the term structure variation of Danish ZCB's. Dahl's work on factor analysis is especially interesting in context to the work being done here because he performed his analysis on Danish ZCB's, analogous to the data used here, but from the 1980s. Therefore it is of interest to compare his results to the results which will be recited in this work.

### 2.2.1 Formulation of factor analysis for the term structure

Considering the success achieved in the past of applying factor analysis to model the term structure of interest rates and the analytical benefits the use of it brings, it was decided to apply factor analysis on the data. The analytical benefits weighting the most here are the relaxation of the parallel assumption of the yield curve and the low number of factors needed to describe it historically reported. But the small number of parameters is essential for using the results as a base for a factor model of the term structure.

The aim of factor analysis is, as said before, to account for the variance of observed data in terms of much smaller number of variables or factors. To perform the factor analysis i.e. to recognize the factors we apply a related method called *principal component analysis* (PCA). The PCA is simply a way to re-express a set of variables, possibly resulting in more convenient representation.

Ind. Sampl. [ $I$ ]	Variables [ $V$ ]			
	$V_1$	$V_2$	$\dots$	$V_p$
$I_1$	$x_{11}$	$x_{12}$	$\dots$	$x_{1p}$
$I_2$	$x_{21}$	$x_{22}$	$\dots$	$x_{2p}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$I_n$	$x_{n1}$	$x_{n2}$	$\dots$	$x_{np}$

Table 2.1:  $p$  variables observed on a sample of  $n$  individual samples.

PCA is essentially a orthogonal linear transformation of  $n$  individuals sets of  $p$  observed variables;  $x_{ij}$ ,  $i = 1, 2, \dots, n$  and  $j = 1, 2, \dots, p$ , such as shown in table 2.1, into an equal number of new sets of variables;  $y_{ij} = y_1, y_2, \dots, y_p$  along with coefficients  $a_{ij}$ , where  $i$  and  $j$  are indexes for  $n$  and  $p$  respectively. Along with obliging the properties listed in table 2.2.1. In our chase the historical yield curves are the  $n$  individual sets, containing  $p$  variables of different maturities each.

The last property in table 2.2.1 states that the new combinations  $y_i$  express the variances in a decreasing order so consequently the PCA can be used to recognize the most significant factors i.e. the factors describing the highest ratios of the variance. The method is perfectly general and the only assumption necessary to make is that the variables which the PCA is applied on are relevant to the analysis being conducted. Furthermore it should be noticed that the PCA uses no underlying model and henceforth it is not possible to test any hypothesis about the outcome.

According to Jamshidian & Zhu (1997), the PCA can either be applied to the

- Each  $y$  is a linear combination of the  $x$ 's i.e.  $y_i = a_{i1}x_1 + a_{i2}x_2 + \dots + a_{ip}x_p$ .
- The sum of the squares of the coefficients  $a_{ij}$  is unity.
- Of all possible linear combinations uncorrelated with  $y_1$ ,  $y_2$  has the greatest variance. Similarly  $y_3$  has the greatest variance of all linear combinations of  $x_i$  uncorrelated with  $y_1$  and  $y_2$ , etc.

Table 2.2: Properties of the PCA  $y$  is a new set of reduced  $x$ 's.

covariance matrix or the correlation matrix of a data set of rates. For clarity we give definitions of the covariance and correlation matrices, taken from *Applied Statistics and Probability for Engineers, third edition* (2003):

#### DEFINITION 2.3 Covariance Matrix

The **Covariance Matrix** is a square matrix that contains the variances and covariances among a set of random variables. The main diagonal elements of the matrix are the variances of the random variables and the off diagonal elements are the covariances between elements  $i$  and  $j$ . If the random variables are standardized to have unit variances, the covariance matrix becomes the correlation matrix.  $\diamond$

#### DEFINITION 2.4 Correlation Matrix

The **Correlation Matrix** is a square matrix containing the correlations among a set of random variables. The main diagonal elements of the matrix are unity and the off diagonal elements are the correlations between elements  $i$  and  $j$ .  $\diamond$

As stated in definition 2.3, the correlation matrix is the covariance matrix of the standardized random vector and it should therefore be adequate to use either of them to perform the PCA. Furthermore according to Jamshidian & Zhu (1997) the variance of all key interest rates are of the same order of magnitude so results from applying PCA on either should become very similar.

A general description and bibliography references of factor analysis and principal component analysis can for example be found in *Encyclopedia of Statistical Sciences* (1988). But our interest here lies in performing factor analysis on the term structure of interest rates and therefore we give formulation of the PCA based on such formulation from *Practical Financial Optimization* (2005), the formulation uses the covariance matrix.

Let  $R$  be a random variable return of a portfolio.

$$R(x, \tilde{r}) = \sum_{t=1}^T x_t \tilde{r}_t$$

where  $x_t$  represents the portfolio holdings in the  $t$ th spot rate, as given in definition 2.1, such that  $\sum_{t=1}^T x_t = 1$  and  $\tilde{r}_t$  is a random value return of that asset for the  $t$ th rate, with the expected value  $\bar{r}_t$  and the variance  $\sigma_t^2$ . The covariance between the returns of two assets  $t$  and  $t'$  in the portfolio is given by

$$\Sigma_{tt'}^2 = \mathbb{E}[(\tilde{r}_t - \bar{r}_t)(\tilde{r}_{t'} - \bar{r}_{t'})].$$

Let  $\mathbf{Q}$  denote the portfolios matrix of variance also known as the variance-covariance matrix or simply covariance matrix. The covariance matrix has the property of being real, symmetric and positive semidefinite and it can be shown that the portfolio variance can be written in a matrix format as

$$\Sigma^2(x) = x^\top \mathbf{Q} x. \quad (2.1)$$

Now the objective is to approximate the variance of the portfolio, without significance loss of variability. We will do that by surrogating the variance matrix  $\mathbf{Q}$  with a matrix  $\hat{\mathbf{Q}}$  of reduced dimensions. To do that we replace the original variable  $R$  with the *principal component*

$$\tilde{f}_j = \sum_{t=1}^T \beta_{jt} \tilde{r}_t$$

which is equivalent to create a new composite asset  $j$  as a portfolio  $\beta_{jt}$  in the  $t^{\text{th}}$  rate.  $j$ .

The variance-covariance matrix of the principal component  $\tilde{f}_j$ , written in vector form is

$$\Sigma_j^2 \doteq \Sigma^2(\beta_j) = \beta_j^\top \mathbf{Q} \beta_j.$$

Now if no priory structure is imposed on the data used, the PCA seeks to transforms the variables in to a set of new variables so that the properties in table 2.2.1 are fulfilled. To maximize the sample variance,  $\sigma_j^2 = \beta_j^\top \mathbf{Q} \beta_j$ , according to construction of orthogonally, we maximize the expression

$$\sigma_j^{2'} = \beta_j^\top \mathbf{Q} \beta_j - \lambda(\beta_j^\top \beta_j - 1).$$

It can be shown that the  $T$  equations in  $T$  unknowns  $\beta_1, \beta_2, \dots, \beta_T$  have consistent solution if and only if  $|\mathbf{Q} - \lambda \mathbf{I}| = 0$ . These condition leads to an equation of

degree  $T$  in  $\lambda$  with  $T$  solutions  $\lambda_1, \lambda_1, \dots, \lambda_T$ , named the *eigenvalues* of the covariance matrix  $\mathbf{Q}$ . Furthermore a substitution of each of all of the  $T$  eigenvalues  $\lambda_1, \lambda_1, \dots, \lambda_T$  in the equation

$$(\mathbf{Q} - \lambda_j \mathbf{I})\beta_j = 0$$

gives the corresponding solutions of  $\beta_j$ , which are uniquely defined if all the  $\lambda$ 's are distinct, called the *eigenvectors* of  $\mathbf{Q}$ .

Lets consider a portfolio consisting of a holding  $\beta_1$ , the portfolio has a variance  $\lambda_1$ , which accounts for the ratio  $\lambda_1/\sigma^2(x)$  of the total variance of the original portfolio. If we then collect the  $k$  largest eigenvalues in a vector  $\mathbf{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_k)$  and let the matrix  $\mathbf{B} = (\beta_1, \beta_2, \dots, \beta_k)$  denote the matrix of the corresponding  $k$  eigenvectors<sup>2</sup>. Then the covariance matrix of the portfolio can be approximated with  $\hat{\mathbf{Q}} = \mathbf{B}\mathbf{\Lambda}\mathbf{B}$  and henceforth an approximation of the variance-covariance matrix in equation 2.1 becomes:

$$\hat{\Sigma}^2(x) = x^\top \hat{\mathbf{Q}}x, \quad (2.2)$$

since the factors are orthogonal.

### The effects of factors on the term structure

Lets now look at what effects change of the  $j$ th principal component has on the value of return  $\tilde{r}$ . If  $\tilde{f} = (\tilde{f}_1, \tilde{f}_2 \dots, \tilde{f}_k)$  denotes a vector of  $k$  independent principal components and  $\mathbf{B}$  denotes matrix the  $k$  corresponding eigenvectors, then we have  $\tilde{f} = \mathbf{B}^\top \tilde{r}$ , and since  $\mathbf{B}\mathbf{B}^\top = \mathbf{I}$ , by construction, we have  $\tilde{r} = \mathbf{B}\tilde{f}$  and the  $T$  random rates are expressed as linear combinations of the  $k$  factors. Therefore a unit change in the  $j$ th factor will cause a change equal to the level of the  $\beta_{jt}$  to rate  $r_t$  and the changes of all factors have a cumulative effect on the rates.

Now assume that  $r_t$  changes by an amount  $\beta_{jt}$  from its current value,  $r_t^0$  and becomes  $r_t^0 + \beta_{jt}$ . Hence the  $j$ th principal component becomes

$$\begin{aligned} f_j &\rightarrow \sum_{t=1}^T \beta_{jt}(r_t^0 + \beta_{jt}) \\ &= \sum_{t=1}^T \beta_{jt}r_t^0 + \sum_{t=1}^T \beta_{jt}\beta_{jt} \\ &= f_j^0 + 1. \end{aligned}$$

---

<sup>2</sup>Note that since the matrix  $B$  is an product of an orthogonal linear transformation it is a *orthogonal* matrix, i.e. square matrix whose transpose is its inverse.

Where the last equality follows from the normalization of the eigenvectors achieved with the orthogonal transformation. What this means is that a unit change of the  $j$ th factor causes a change  $\beta_{jt}$  for each spot rate  $t$ . Since the factors are independent of each other we may therefore express the total change of the random variable spot rates,  $r_t$ , by

$$\Delta r_t = \sum_{j=1}^k \beta_{jt} \Delta f_j, \quad (2.3)$$

where  $k$  is the number of factors, identified by the eigenvector analysis, used to approximate the variance of the portfolio.

To summarize the results derived in this section we now give a definition of the principal components of the term structure of interest rates and a definition of factor loading which the coefficient  $\beta_{jt}$  will be called from now, taken from *Practical Financial Optimization* (2005).

**DEFINITION 2.5 Principal components of the term structure.**

Let  $\tilde{r} = (\tilde{r}_t)_{t=1}^\top$  be the random variable spot rates and  $Q$  be the  $T \times T$  covariance matrix. An eigenvector of  $Q$  is a vector  $\beta_j = (\beta_{jt})_{t=1}^\top$  such that  $Q\beta_j = \lambda_j\beta_j$  for some constant  $\lambda_j$  called eigenvalue of  $Q$ . The random variable  $\tilde{f}_j = \sum_{t=1}^T \beta_{jt}\tilde{r}_t$  is a principle component of the term structure. The first principal component is the one that corresponds to the largest eigenvalue, the second to the second largest, and so on.  $\diamond$

**DEFINITION 2.6 Factor loadings.**

The coefficients  $\beta_{jt}$  are called factor loadings, and they measure the sensitivity of the  $t$ -maturity rate  $t$  to changes of the  $j$ th factor.  $\diamond$

## 2.3 Application of Factor Analysis

A principal component analysis, as formulated in sections 2.2.1 and 2.2.1, was implemented on the data set described in section 1.2 in order to recognize the key factors of the Danish term structure. More precisely it was performed for yearly maturity steps dated from the 4. of January 1995 to the 4. of October 2006, all in all thirty maturities in 614 issue dates i.e.  $n = 614$  sets of  $p = 30$  observed variables.

In appendix A.2 the results of the factor analysis performed on data from 1995–2006, beginning from 1995 and adding one year at time are displayed. From those figures it can be seen that the shape of the factors becomes stable when

data from 4-5 years are included. Therefore it is concluded that the factors found from data groups containing more than five years of data give a stable estimation. The results displayed below are found from factor analysis performed on the years 1995-2006.

Table 2.3 shows the standard deviation, the proportion of the variance and the cumulative proportion of the seven most significant principal components found for the period. The first three components, or factors, explain 99.9% of the total variation and where as the first factor accounts by far for the most of the variation or 94.9%.

	PC1	PC2	PC3	PC4	PC5	PC6	PC7
Std.	5.335	1.1902	0.30696	0.15000	0.05260	0.02704	0.01863
Pr. of Var.	0.949	0.0472	0.00314	0.00075	0.00009	0.00002	0.00001
Cum. Prop.	0.949	0.9960	0.99912	0.99987	0.99996	0.99998	0.99999

Table 2.3: The seven most significance components found applying PCA on Danish ZCB from 1995–2006. Std. is the standard deviation, Pr.of Var. is the proportion of the total variance and Cum. Pr. is the cumulative proportion of the variance.

Figure 2.4 shows the three factor loadings corresponding to the three largest principal components in table 2.3 (the loadings are listed in appendix A.1). The loadings we recognize as the shift, steepness and convexity factors identified by Litterman & Scheinkman (1991).

From looking at figure 2.4 it can be observed that the the first factor, forms almost a horizontal line over the whole time period, excluding approximately the first five to six years. This corresponds to a change of slope for the first five years and a parallel shift for the rest of the maturity horizon. Although the slope in the first five to six years of the first factor is a deviation from what was observed in the other experiments mentioned in the introduction of section 2.2, the horizontal line is dominant for the rest of the term structure and hence the factor is recognized as the level factor.

The second factor, the slope, which corresponds to a change of the slope for the whole term structure accounts for 4.72% of the total variation. It can be seen from the plot that the slope is decreasing as a function of maturity which fits the description of a normal yield curve. This is in accordance to the fact that the yield curve the period investigated was for most parts a normal yield cure with marginally diminishing yields. It is also worth mentioning that the slope for the first ten years is much steeper.

The third factor, can be interpreted as the curvature factor since positive changes in it cause a decrease in yield for bonds with short and long maturities but cause an increase in yield for medium length maturities.



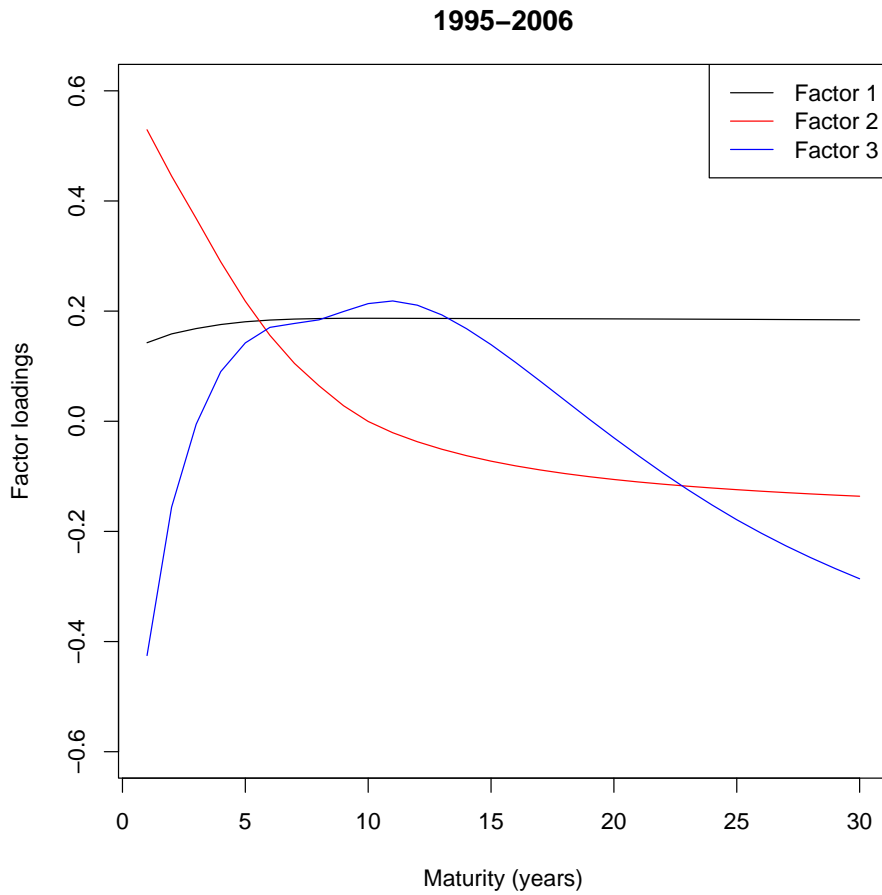


Figure 2.4: The first three factor loadings of the Danish yield curves, the values of the factor loadings can be seen in appendix [A.1](#).

In reference to equation 2.2 the three factors level, slope and curvature should be sufficient to form an estimation variance-covariance matrix  $\hat{Q}$  since they can explain the variance of the term structure up to 99.9%.

Although the first two factors are sufficient, from a statistical point of view, to describe the term structure accurately the third factor, which describes the curvature, is beneficial to include in a model since changes in the curvature of the term structure do occur. Therefore a model which does not take this change of term into account has a potential weakness of not capturing possible movements of the yield curve. Because of this we will use three factors throughout the report.

### Example of the effects of factors on rates

Equation 2.3 describes the relationship a change of the factors has on the level of rates, redisplay here for convenience

$$\Delta r_t = \sum_{j=1}^k \beta_{jt} \Delta f_j.$$

As an example lets see what effect a unit change ( $\Delta f_1 = 1$ ) of the level factor ( $j = 1$ ) has on the ten year rate ( $t = 10$ ).

$j$	1	2	3
$\beta_{j,10}$	0.1870124	-0.0003624621	0.213623944

Table 2.4: The values of  $\beta_{j,10}$  for the first three factors, taken from appendix A.1.

From table 2.4 we have  $\beta_{jt} = \beta_{1,10} = 0.1869201$  so a unit change in factor 1 causes 0.1869201 change in the ten year rate, which means that if the ten year rate is 5% a unit change in the level factor causes it to become 5.1869%.

In the same manner a unit change of three most significance factors ( $\Delta f_j = 1$ ) for  $j = (1, 2, 3)$ , again for ten years means:

$$\Delta r_{10} = \sum_{j=1}^3 \beta_{j,10} \Delta f_j = (0.1870 - 0.0004 + 0.2136) \cdot 1 = 0.4002$$

meaning that a 5% ten year rates would become 5.4002% if a unit change occurred for all the factors.

### 2.3.1 Comparison with results from H. Dahl

As mentioned earlier it is of interest to make a comparison with the results on Dahl's research of the factor analysis conducted on Danish bonds from the 1980's. The main facts from his analysis are that the most significant factor explains about 86% of the historical variation, the second most significant factor explains about 11% and the third most significant factor, which affects the term structure of maturities up to ten years, explains about 3%. All in all these three factors explain 99,6% of the term structure variance. Furthermore a fourth factor was able to explain what Dahl refers to as a twist of the term structure up to maturities of four years, explaining about 0.3% of the total variation of that time interval.

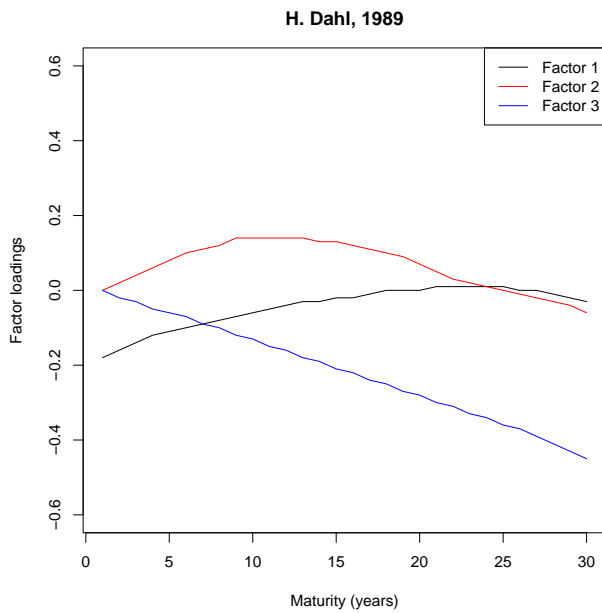
Figure 2.5a, shows the first three factors found by Dahl and figure 2.5b are factors from figure 2.4 redrawn for ease of comparison. It is visible that there have been some changes in the composition of these three factors. The factor 1 which is sloped in the older analysis has become level, apart from the first 5 years as previously mentioned. The proportion of variance explained by the first factor has also increased from 86% up to approximately 95%, which means that parallel shifts weigh more in the shape of the term structure. The main observation is that the shape of the first factor now looks more similar to results of factor analysis conducted on larger markets such as USA and Italy (Bertocchi et al. (2000) and D'Ecclesia & Zenios (1994)), which typically have a flat level curve over the whole maturities. The slope and curvature factors are also shaped differently in our analysis compared to Dahl's. Both in degree and level of explanation.

The difference in the shape of the factors must be explained by different economic circumstances present in Denmark for the past couple of decades. Dahl's work (including the data used) is from the eighties which was a turbulent time in Danish monetary policies but for the past years the situation has been fairly stable and has further begun closely to follow the trend of big markets such as the European and American respectively.

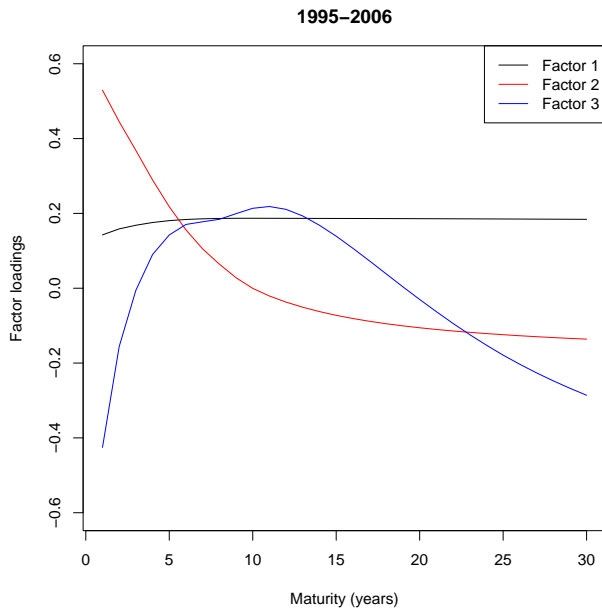
## 2.4 Conclusion

It could be concluded from figure 2.2 of the interest rates, that the assumption of a parallel shift of the term structure does not hold. There is in particular little correlation between short and long term yields so this assumption is especially dodgy to make when modeling long maturities.

The factor analysis gave the expected results, we were able to account for up



(a) Factors found by in the 1980's



(b) Factors found for the 1995-2006 data.

Figure 2.5: The three most significance factors found compared to the factors found in the 1980's

to an astonishing 99% of the variation with three factors, for the case studied here. Furthermore we found that the second factor counted for some 5% in the 1995–2006 period which indicates the magnitude of error associating with the parallel assumption.

It was furthermore found that the factor loadings of the Danish ZCB's, for the period considered, differ in one significance aspect from what has been observed from other markets, namely the slope evident in the first few years of the first factor, the level factor, is not observable in the level factor in other market areas that we know of. The Danish factors for the contemporary rates nevertheless behave in manner more similar to other market than it did in the eighties.



# Normality of Interest Rates

---

In the interest rate literature there are two main schools of research, one group assuming that interest rates follow a normal distribution and another which is more inclined to believe that interest rates are log-normally distributed. Therefore it is of interest to investigate firstly whether interest rates follow a normal distribution and secondly if the rates follow the log-normal distribution better. In this section those hypothesis are tested on the interest rate data we use. The main result is that there are no clear indicators that the rates are more log-normal distributed.

The rest of the chapter is laid out as follows:

- In section [3.1](#) an introduction to the procedures used for the analysis is given.
- In section [3.2](#) an analysis of the normality of interest rates is constructed.
- In section [3.3](#) the analysis in the previous section is repeated, but taking the log-normal of the interest rates.
- Finally, section [3.4](#) concludes the chapter.

### 3.1 Introduction

To conduct the investigation we choose different time horizons for the rates, namely the rates for one, five, fifteen and thirty years. Those maturities are chosen to cover the short, medium and long term yields. From looking at figure 2.2, in chapter 2, it is evident that the shape of the yield curve varies within the period shown. The rates are for example noticeably higher for the first years of the period, ranging from the beginning of 1995 to around 1998–1999, than for the last years of the period, from around 1998–1999 up to October 2006. That is especially evident for the medium to long term rates. Apart from that the period around the millennium behaves differently. That period shows behavior of a flat and inverted yield curve. Therefore it is also of interest to investigate the normality within some sub-periods of the time interval. We use two approaches to estimate the normality, namely visual inspection and goodness-of-fit tests.

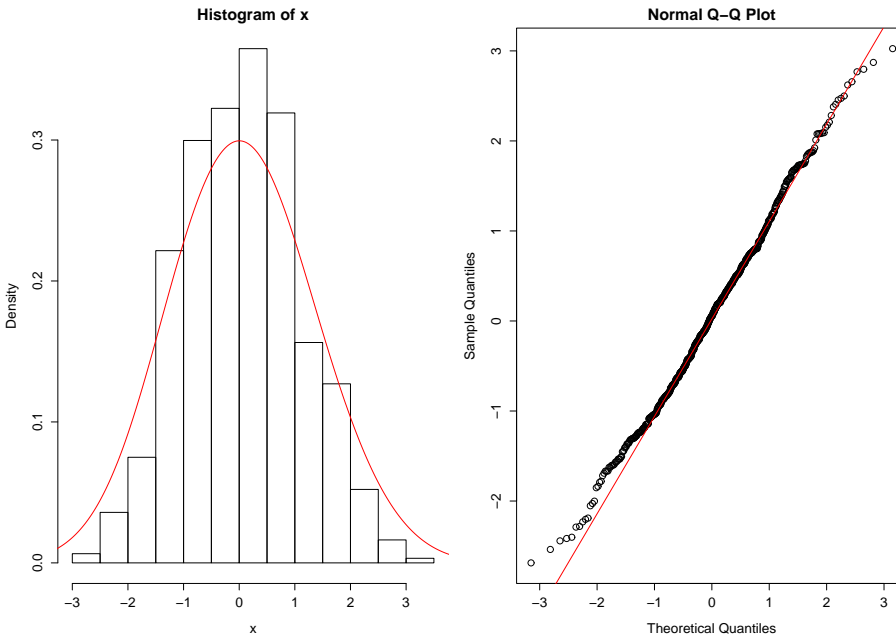


Figure 3.1: Histogram (left) and Q-Q plot (right) made for data from a random sample.

The visual inspection is conducted by plotting histograms of the rates along with smoothed curves, which are computed via kernel density estimation<sup>1</sup> of the data using a Gaussian (normal) kernel. Those normal plots can indicate

<sup>1</sup>A kernel is a weighting function used in non-parameter estimation techniques, used here to estimate the density function of the random variable.



if the data looks like it arrives from a normal population. However making a normal plot is not enough since other distributions exist which have similar shaped curves. Therefore Quantile to Quantile plots ( $Q-Q$  plots) of the data are also drawn. In a  $Q-Q$  plot the sample quantiles are plotted against the theoretical quantiles for the expected distribution, therefore a sample arriving from the expected distribution results in the data points being distributed along a straight line. Figure 3.1 shows an example of histogram along with its smoothed line and a  $Q-Q$  plot made from a random generated sample of 614 numbers with mean zero and standard deviation of one, i.e. sampled from standard normal distribution. Notice that the shape of the smoothed curve of the histogram in the figure is often said to be bell shaped.

The normality or goodness-of-fit tests which were applied on the data were the *Jarque-Bera* and *Shapiro-Wilk* tests. These tests are explained in the following two subsections.

### 3.1.1 The Jarque-Bera test for normality

The Jarque-Bera test is a goodness-of-fit test of departure from normality. It can therefore be used to test the hypothesis that a random sample  $X_i = (X_1, \dots, X_n)$  comes from a normally distributed population. The test is based on the sample kurtosis and skewness which are the third and fourth standardized central moments (mean and variance being the first and second ones). The skewness is a measure of the asymmetry of a probability distribution while the kurtosis is a measure of how much of the variance is due to infrequent extreme events. A sample drawn from a normal distribution has an expected skewness of zero and kurtosis of three, but in order to make the kurtosis equal to zero it is a common practice to subtract three from it. If that is done one can test the null hypothesis that a data comes from a normal distribution based on the joint hypothesis that the skewness and kurtosis are zero. One such test is the Jarque-Bera test (Jarque & Bera (1987)), which has the test statistic

$$JB = \frac{n}{6} \left( S^2 + \frac{(K - 3)^2}{4} \right), \quad (3.1)$$

where  $n$  is the number of observations.  $S$  is the sample skewness defined as

$$S = \frac{\mu_3}{\sigma^3} = \frac{\mu_3}{(\sigma^2)^{3/2}} = \frac{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^3}{\left( \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \right)^{3/2}}$$

where  $\mu_2$  is the second central moment or the variance,  $\mu_3$  is third central moment or the skewness,  $\sigma$  is the standard deviation and  $\bar{X}$  is the sample

mean.  $K$  is the sample kurtosis defined as

$$K = \frac{\mu_4}{\sigma^4} = \frac{\mu_4}{(\sigma^2)^2} = \frac{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^4}{\left(\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2\right)^2}.$$

where  $\mu_4$  is the fourth central moment or the kurtosis. In the test test statistic  $JB$  three is subtracted from the kurtosis to make the it equal to zero. The test statistic has an asymptotic  $\chi^2$  distribution with two degrees of freedom and the test has been reported to perform well for samples of small and large sizes.

### 3.1.2 The Shapiro-Wilk test for normality

The Shapiro-Wilk test is a another goodness-of-fit test which can be used for testing departure from normality. It is a so called omnibus test in which the explained variance in a set of data is significantly greater than the unexplained variance, overall and is regarded as one of the most powerful omnibus test procedures for testing univariate normality. The test statistic of the Shapiro-Wilk test,  $W$  is based on the method of generalized least-squares regression of standardized<sup>2</sup> ordered sample values. We will cover the method of least-squares in section 4.6.1, but the Shapiro-Wilk test can be computed in the following way, adapted from *Encyclopedia of Statistical Sciences* (1988).

Let  $\mathbf{M}' = (M_1, \dots, M_n)$  denote the ordered expected values of a standard normal order statistics for a sample of size  $n$  and let  $\mathbf{V}$  be the corresponding  $n \times n$  covariance matrix. Now suppose that  $X_i = (X_1, \dots, X_n)$  is the random sample to be tested ordered  $X'_1 < \dots < X'_n$ . Then the test statistic is defined as

$$W = \frac{(\sum_{i=1}^n w_i X'_i)^2}{\sum_{i=1}^n (X_i - \bar{X})^2}$$

where

$$\begin{aligned} \mathbf{w}' &= (w_1, \dots, w_n) \\ &= \frac{\mathbf{M}'\mathbf{V}^{-1}}{[(\mathbf{M}'\mathbf{V}^{-1})(\mathbf{V}^{-1}\mathbf{M})]^{1/2}} \end{aligned}$$

and  $\bar{X}$  is the sample mean. The test statistic  $W$  is a measure of the straightness of the normal probability plot and small values of  $W$  indicate departure from normality.

---

<sup>2</sup>The procedure of representing the distance of a normal random variable from its mean in terms of standard deviations.

In the literature the Shapiro-Wilk test is regarded as a very sensitive omnibus test and has shown to be a very good test against either skewed or short or very long-tailed populations. The Shapiro-Wilk test has also been shown to be usable for samples of size  $3 \leq n \leq 2000$  which is well within the scope considered here<sup>3</sup>.

### 3.1.3 Interpretation of the normality tests

The most convenient way of analyzing the tests results is by looking at the  $P$ -value of the test statistic. That is mainly due to two reasons, the former being that the  $P$ -value statistic is comparable between tests and the latter being that stating the  $P$  value gives more information than only stating whether or not certain hypothesis is or is not rejected at a specified level of significance.

The level of significance  $\alpha$  is the probability that a true hypothesis gets rejected and the  $P$ -value is the smallest level of significance that would reject the hypothesis. Or in other words, one would reject a hypothesis if the  $P$ -value is smaller than or equal to the chosen significance level. For example a  $P$ -value of 0.05 would lead to rejection at any level of significance  $\alpha \geq P\text{-value} = 0.05$ . Therefore the null hypothesis would be rejected if level of significance is chosen to be 0.1, but would be accepted if the chosen level were 0.001. Common choices of levels of significance are  $\alpha = 0.05$  for 5% and  $\alpha = 0.01$  for 1%. A  $P$ -value of 0.05 is a typical threshold used in industry to evaluate the null hypothesis.

A more abstract explanation of  $P$ -value is that a  $P$ -value laying close to zero signals that a null hypothesis is false, and typically that a difference from the expected distribution is likely to exist. Large  $P$ -value, closer to 1 imply that there is little or no detectable difference for the sample size used. Tables 3.1

<i>JB</i>	<i>P</i> -value	<i>W</i>	<i>P</i> -value
4.1247	0.1272	0.9956	0.08038

Table 3.1: Example of test  $JB$  test results for a data sampled from standard normal distribution. Table 3.2: Example of test  $W$  test results for a data sampled from standard normal distribution.

and 3.2 show test results for the Jarque-Bera and Shapiro-Wilk test on the same sample data as was used in figure 3.1. The  $P$ -value of 0.1272 for the  $JB$  test states that there is  $1 - 0.1272 = 0.87387.3\%$  change. Both the tests pass the sample as normally distributed for a significance level of 0.05.

<sup>3</sup>The R function used here to calculate the test gives the allowed sample size  $3 \leq n \leq 5000$ .

## 3.2 Normality Inspection

Now we look at the results of the histograms, the  $Q-Q$  plots and the goodness-of-fit tests applied to the data. The tests are both made for the whole data set ranging from 1995 to 2006 and for subsets of the period, because as mentioned in the beginning of the chapter the shape of the yield curves varies between sub periods of the whole set and therefore it is of interest to look at subsets spanning smaller time frames.

### 3.2.1 Normality Test on the Whole Data Set, 1995–2006

First we look at the whole data period from 1995 to 2006. Figure 3.2 shows the histograms for the selected maturity dates. From these histograms it is evident that the rates, in general, can hardly be regarded as a sample coming from a normally distributed population. The one and five year rates show a high level of skewness and have thick tails. The fifteen and thirty year rates have two humps which normally distributed data does not have. As for the two humps there is a period between 1995 and 1998 where the rates, especially for medium and long maturities, are noticeable higher. This period might be the cause for the hump in the curves for the fifteen and thirty year rates in the histograms of the data. Therefore it will be interesting to look at subsets of the data which excludes the 1995-1998 period. Of the different sets of maturities the one and five year rates look a little more likely to be regarded normally distributed.

Figure 3.3 displays the  $Q-Q$  plots for the selected interest rates of the data set. The  $Q-Q$  plots confirm what can be seen from the histograms, showing a one and five year maturity which is close to the line on some range, but far from it for the other ones, especially for the fifteen and thirty year rate in the higher values of the quantiles, which explains the double hump.

Tables 3.3 and 3.4 show the outcome from Jarque-Bera and Shapiro-Wilk tests performed on the data set. The  $P$ -values of the test statistics, both for the Jarque-Bera and the Shapiro-Wilk test, confirm the observations from the figures. The  $P$ -values are too low for the data to pass as a sample arriving from a normality distributed population.

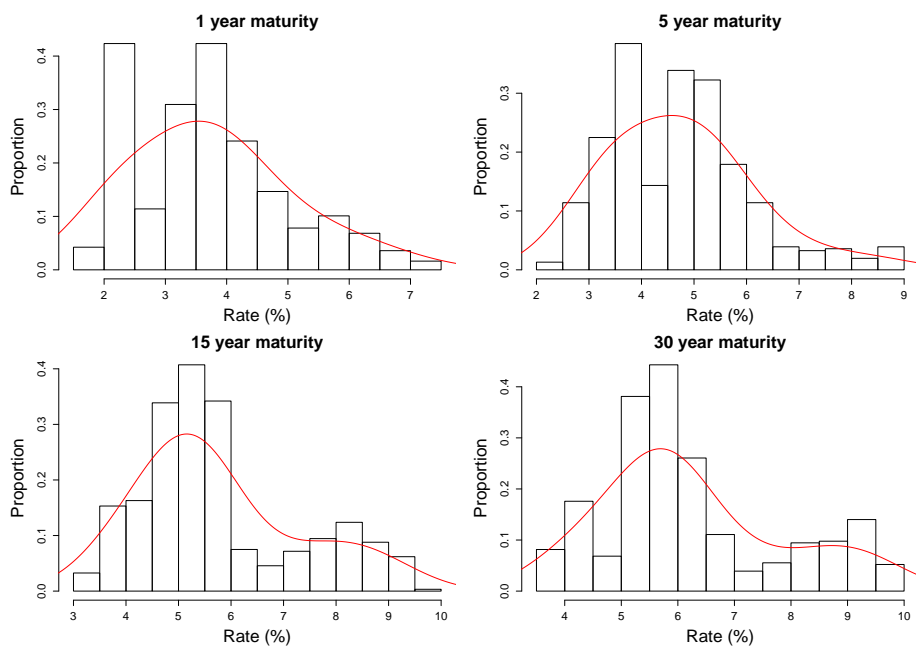


Figure 3.2: Histograms of selected interest rates from 1995-2006.

maturity	$JB$	$P$ -value	maturity	$W$	$P$ -value
1	40.2376	1.830e-09	1	0.9481	7.424e-14
5	73.557	2.2e-16	5	0.9509	2.062e-13
15	66.5859	3.442e-15	15	0.9131	< 2.2e-16
30	54.8556	1.225e-12	30	0.9117	< 2.2e-16

Table 3.3: Results of Jarque-Bera test of Table 3.4: Results of Shapiro-Wilk test for interest rates between 1995-2006.

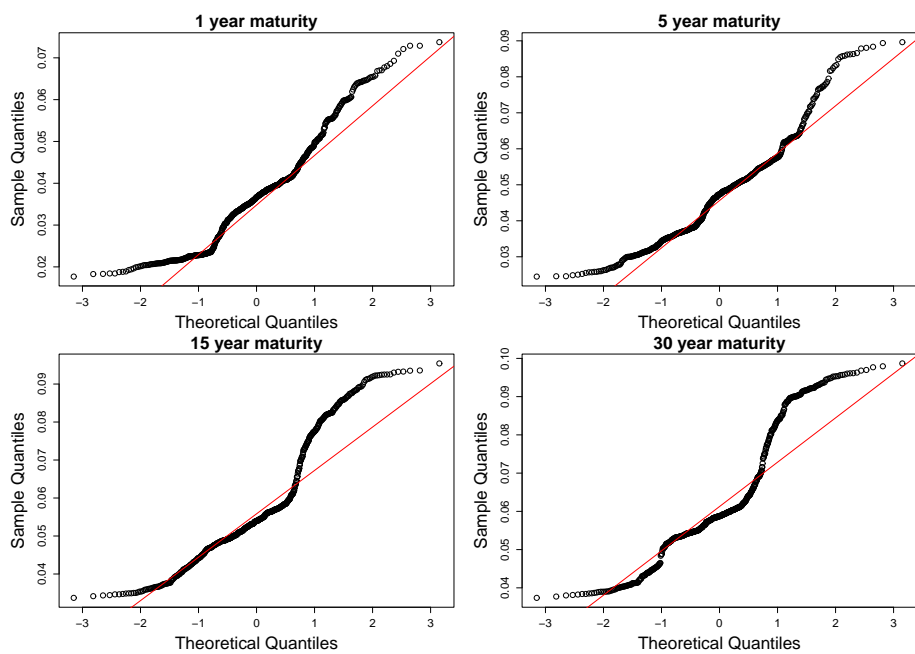


Figure 3.3:  $Q$ - $Q$  plot of selected interest rates from 1995-2006.

### 3.2.2 Normality Test on Data Ranging from 2001–2006

Now we look at the first subset of the data for the years from 2001 to 2006. The period is chosen to start from 2001 because of the unusual behavior of the yield curve around the millennium mentioned before.

The same procedure as before is performed for the selected sample resulting in figures 3.4 and 3.5 showing the histograms and the  $Q-Q$  plots respectively. There is some difference evident in these histograms compared to the histograms for the 1995-2006 period. The smoothed curve in the histograms is flatter and the data seems to be less skewed especially for the 5 year rates. Furthermore the double hump in the longer maturities in the 1995-2006 data is no longer visible, which can indicate that the oldest part of the data is the cause of it.

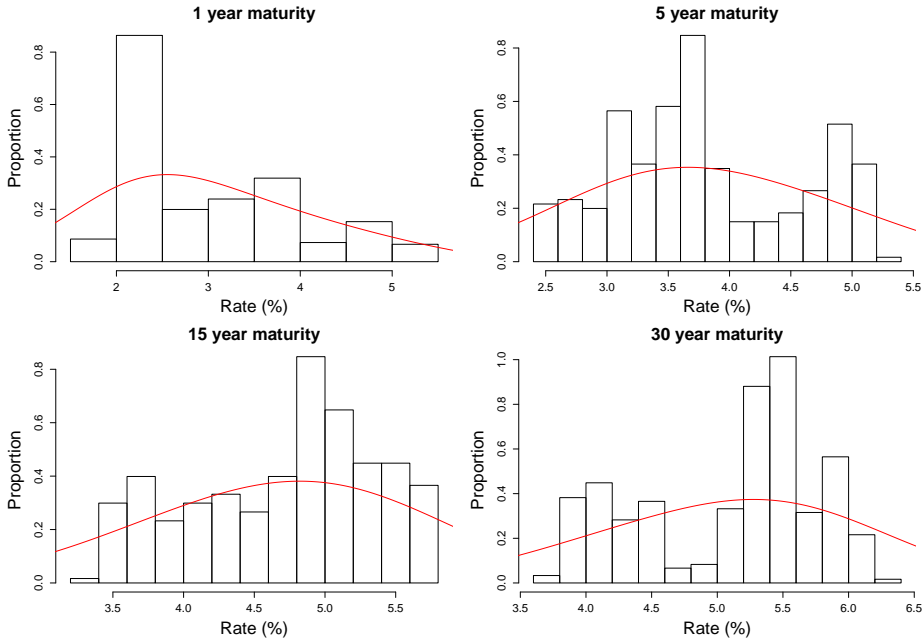


Figure 3.4: Histograms of selected interest rates from 2001-2006.

The  $Q-Q$  plots tell a similar story as the histograms. the fit looks significantly better for the fifteen and the thirty year rate, but there is no evident difference for the one and five year rate compared to the 1995-2006 data.

Tables 3.5 and 3.6 show the  $JB$  and  $W$  test statistics and the corresponding  $P$ -values. The  $P$ -values, although showing improvement for the 15 and 30 year rate, are too low for all of the maturities in both of these tests. The exception

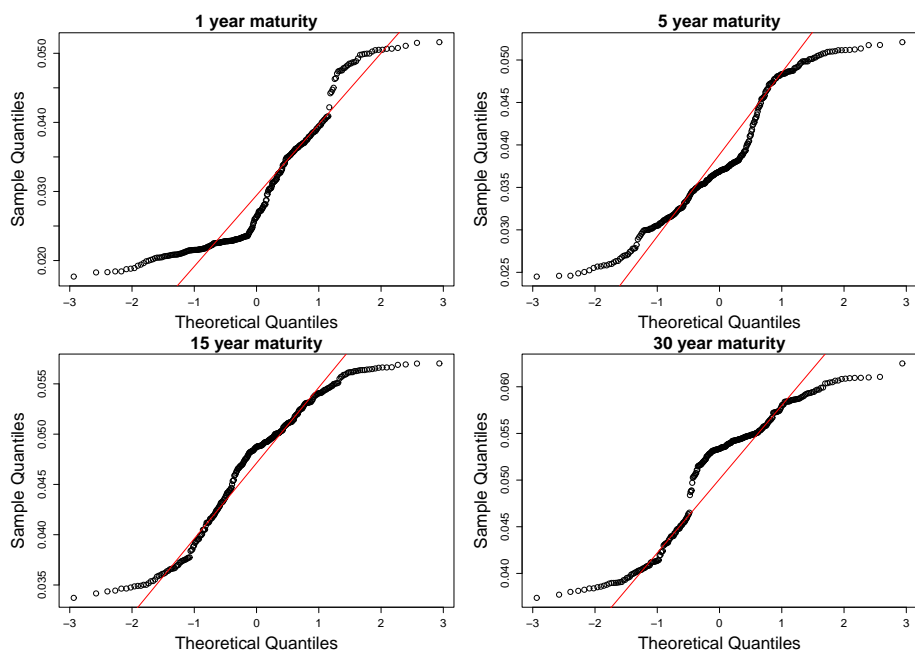


Figure 3.5:  $Q-Q$  plot of selected interest rates from 2001-2006.



is the 5 year rate  $JB$  statistic which has higher value than all the others, but still relatively low and is furthermore not backed up by the  $W$  statistic.

years	$JB$	$P$ -value	years	$W$	$P$ -value
1	36.5114	1.179e-08	1	0.8755	6.621e-15
5	15.8083	0.0003692	5	0.9451	3.695e-09
15	19.2273	6.681e-05	15	0.9447	3.356e-09
30	26.4184	1.834e-06	30	0.9087	1.546e-12

Table 3.5: Results of Jarque-Bera test of Table 3.6: Results of Shapiro-Wilk test of interest rates between 2001-2006. interest rates between 2001-2006.

From the results of the figures and the normality tests it is though evident that this time frame can not be passed as being normally distributed in general.

### 3.2.3 Normality Test on Data Ranging from 1995–1998

Now we take a look at the earliest period of the data, we choose to take the data ranging from 1995 up to 1999. The histograms shown in figure 3.6 show some level of normality for the five year maturities but apart from that there is not much sign of normality. The one and five year data are quite skewed and the two humps for the medium and long term is showing in the histograms of the whole data set (1995-2006) is visible again.

The  $Q$ - $Q$  plots that are displayed in figure 3.7 display more lack of fit than before, especially for the longer maturities where unusually high rates, compared to the rest of the data period, are visible.

The test statistics for the 1995 to 1998 in tables 3.7 and 3.8 echo what can be seen in the histograms before. they are in general too low to indicate normality.

years	$JB$	$P$ -value	years	$W$	$P$ -value
1	54.0437	1.839e-12	1	0.8159	4.55e-15
5	20.6486	3.283e-05	5	0.9344	3.898e-08
15	16.3801	0.0002774	15	0.9282	1.205e-08
30	20.6316	3.311e-05	30	0.8974	7.665e-11

Table 3.7: Results of Jarque-Bera test of Table 3.8: Results of Shapiro-Wilk test of interest rates between 1995-1998. interest rates between 1995-1998.

The tests were also run for 1999-2001 and showed similar results as before although the Jarque-Bera test gave slightly better results than before (more normal).

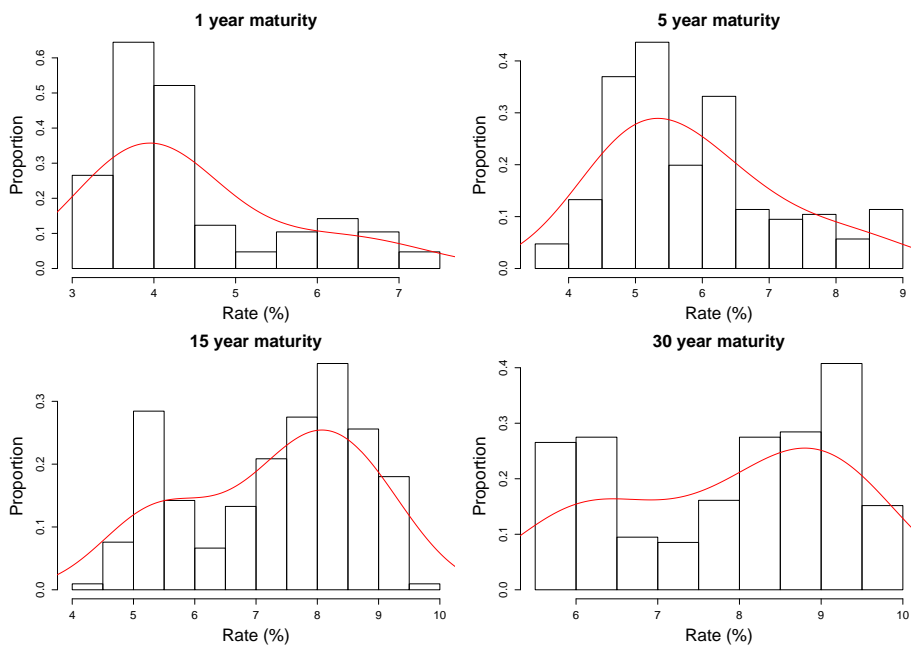


Figure 3.6: Histograms of selected interest rates from 1995-1998.

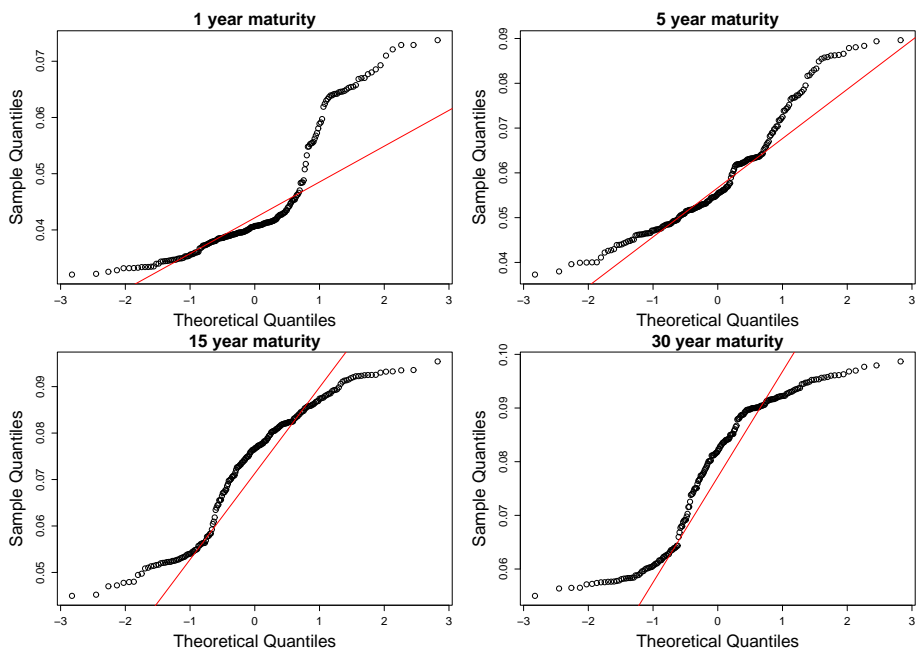


Figure 3.7:  $Q-Q$  plot of selected interest rates from 1995-1998.

### 3.2.4 Normality Test on Data Ranging from 2005–2006

We saw in the last subsection, that the time frames selected did not contain normally distributed data. It is therefore decided to reduce the time frame tested further. Looking only at the last few weeks. That resulted in more normally distributed rates than before. Adding one year at a time gave similar graphical results up to the fourth year added. But after that the histograms resulted in a less normally distributed data. As an example the 2005-2006 data set is displayed here in histograms and  $Q-Q$ -plots in figures 3.8 and 3.9, respectively.

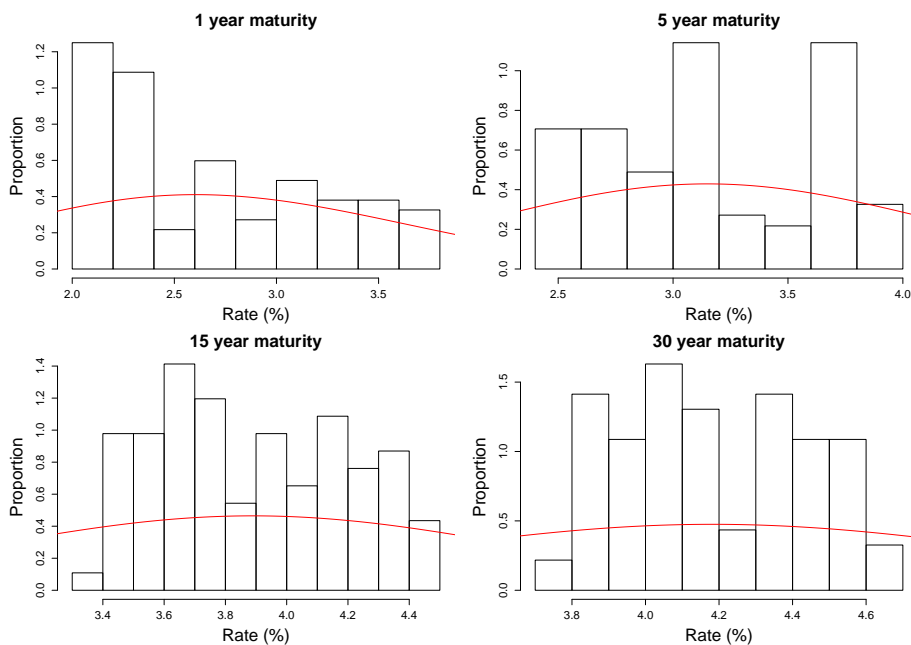


Figure 3.8: Histograms of selected interest rates from 2005-2006.

Figures 3.8 and 3.9 show the same characteristics as the 2001 to 2006 subset, the smoothed curve is flat but not very skewed nor with high kurtosis. The  $Q-Q$  plots seems to give better results though. The quantiles lie closer to the theoretical line, which indicates normality. That applies especially for the fifteen and thirty year rates.

The goodness-of-fit test results in tables 3.9 and 3.10 are the best ones up to now with the highest  $P$ -values overall, especially for the medium and long term rates and most of the  $P$ -values for the  $JB$  test statistic indicate normality (the

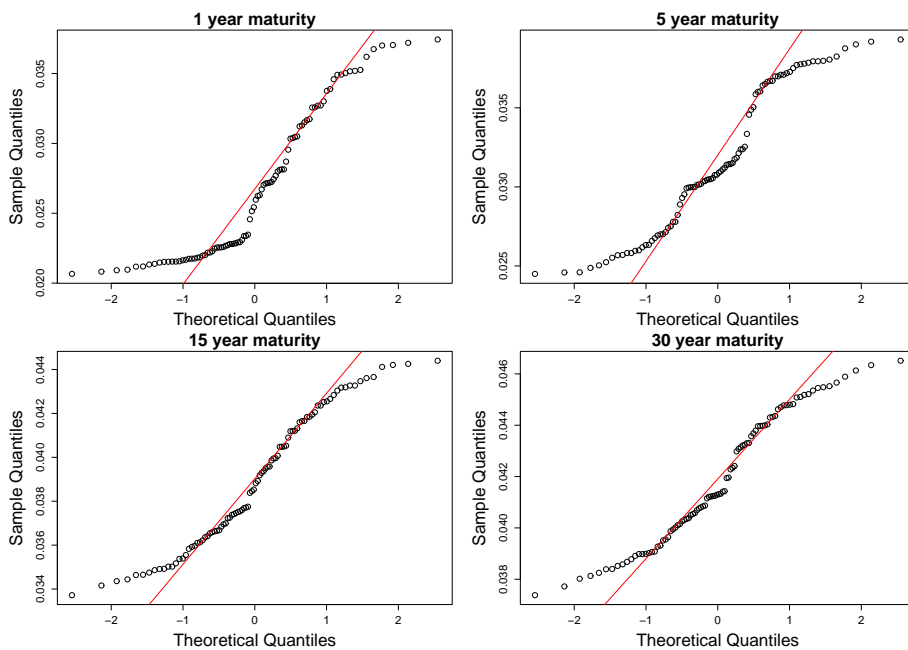


Figure 3.9: Q-Q plot of selected interest rates from 2005-2006.

years	$JB$	$P$ -value	years	$W$	$P$ -value
1	9.4569	0.00884	1	0.8768	3.352e-07
5	6.5902	0.03706	5	0.9255	5.705e-05
15	6.2529	0.04387	15	0.947	0.0009486
30	6.1375	0.04648	30	0.9488	0.001217

Table 3.9: Results of Jarque-Bera test of interest rates between 2005-2006. Table 3.10: Results of Shapiro-Wilk test of interest rates between 2005-2006.

5, 15 and 30 year rate) and the fifteen and the thirty year rates for the  $W$  test indicate the same. It is also noticeable that the five year rate, which did best in the 2001 to 2006 data, does not seem to be doing noticeably better in the in this subset of the data.

### 3.3 Normality Versus Log-normal

In this section we take the natural logarithm of our interest rate data and repeat our experiment in order to investigate whether the interest rates are more log-normal than normal?

#### 3.3.1 Normality test for data ranging from 1995–2006, $\ln$ taken

Considering the whole data set, figures 3.10 and 3.11 show the histograms and the  $Q$ - $Q$  plots respectively. The bell shape of the normality smoothing in the histograms is less biased than before, but the medium and long maturities appear to have too thick tails although they do not have the double hump seen for the data without the  $\ln$  taken.

The  $Q$ - $Q$  plot in figure 3.11 shows better tracing of the line than for the non  $\ln$  data, especially for the one and five year data, indicating better fits.

Tables 3.11 and 3.12 display the results from the  $JB$  and  $W$  tests. The  $JB$  test results indicate that the data could be considered normal but surprisingly the  $W$  test results are far off from being normal. But in general, taking the logarithm of the interest rates results in more normally distributed data for the whole data set.

years	$JB$	$P$ -value	years	$W$ -value	$P$ -value
1	15.0697	0.0005341	1	0.9668	1.433e-10
5	4.1357	0.1265	5	0.9859	1.229e-05
15	18.8934	7.895e-05	15	0.9577	2.74e-12
30	13.2719	0.001312	30	0.9493	1.131e-13

Table 3.11: Results of Jarque-Bera test of  $\ln$  of interest rates 1995-2006.

Table 3.12: Results of Shapiro-Wilk test of  $\ln$  of interest rates 1995-2006.

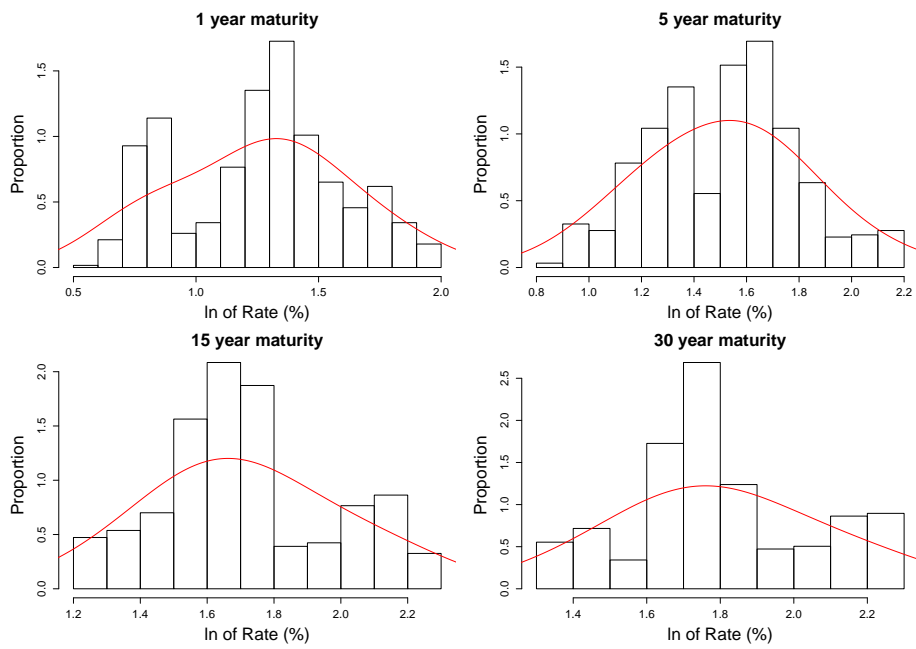


Figure 3.10: Histograms of the logarithm of the rates from 1995-2006.

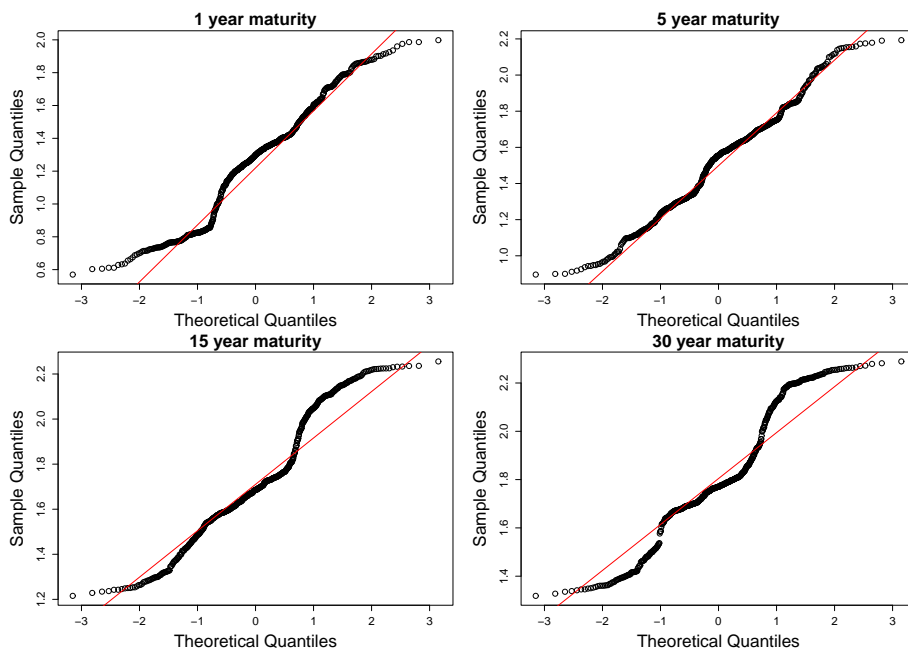


Figure 3.11:  $Q-Q$  plot of the logarithm of the rates from 1995-2006.



### 3.3.2 Normality Test for Data Ranging from 2001–2006, $\ln$ Taken

The same is done as before, taking the logarithm of the data for the years from 2001 to 2006. Figure 3.12 shows the histograms for the logarithms of the rates respectively. Furthermore the curves show some indication of bell shape, although being flat. The high kurtosis in the one and thirty year maturities does not seem to affect the shape of the curve.

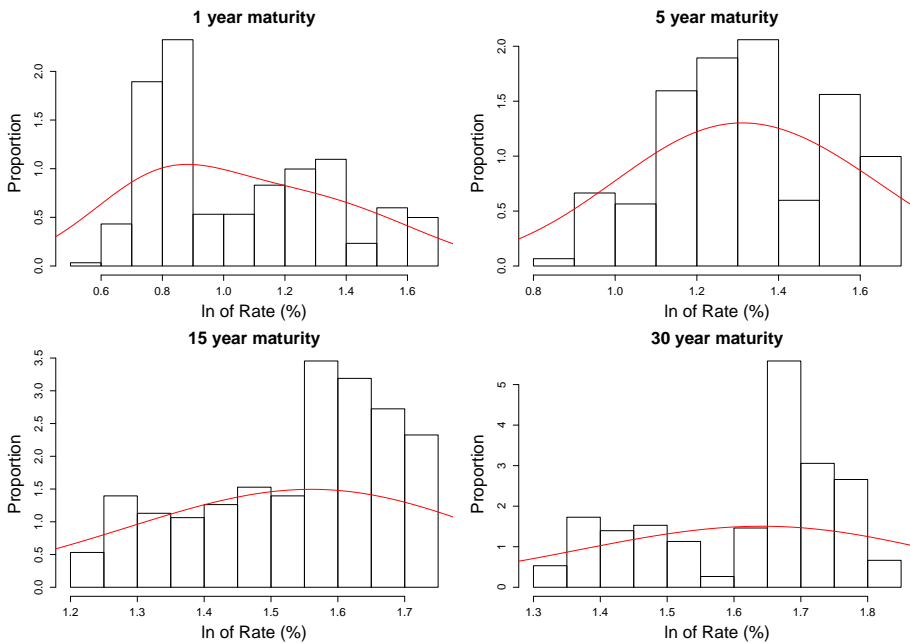


Figure 3.12: Histograms of the logarithm of the rates from 2001-2006.

The corresponding  $Q-Q$ -plots shown in figure 3.13 do not indicate anything special, not showing especially better or worse behavior of normality than has been seen before.

The result from the normality test in tables 3.13 and 3.14 do not give clear evidence of “more normality” compared to tables 3.5 and 3.6 for the same set without the logarithm taken. The  $P$ -values are little better for the 1 and 5 year rates with logarithm and a little worse for the 15 and 30 years.

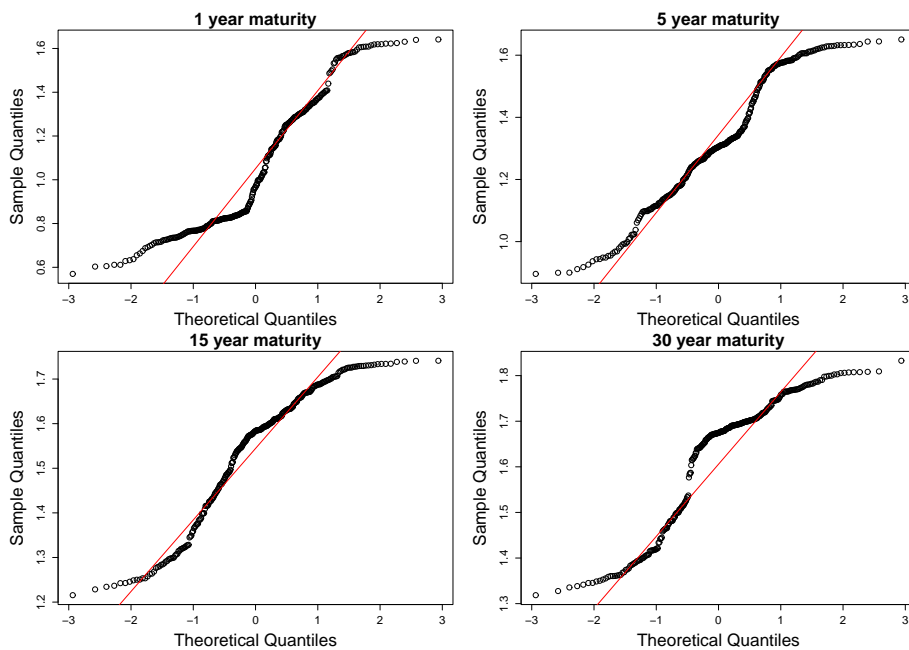


Figure 3.13: Q-Q plot of the logarithm of the rates from 2001-2006.

years	$JB$	$P$ -value	years	$W$	$P$ -value
1	25.566	2.808e-06	1	0.9111	2.408e-12
5	9.3409	0.009368	5	0.9588	1.618e-07
15	25.2801	3.240e-06	15	0.9273	5.981e-11
30	32.1768	1.030e-07	30	0.8898	6.041e-14

Table 3.13: Results of Jarque-Bera test of  $\ln$  of interest rates 2001-2006.Table 3.14: Results of Shapiro-Wilk test of  $\ln$  of interest rates 2001-2006.

### 3.3.3 Normality Test on data Ranging from 1995–1998, $\ln$ Taken

The histograms in figure 3.14 are highly skewed but the shape of the smoothed curve, looks more normal than the data before taking the logarithm. the five year data has the “best” distribution.

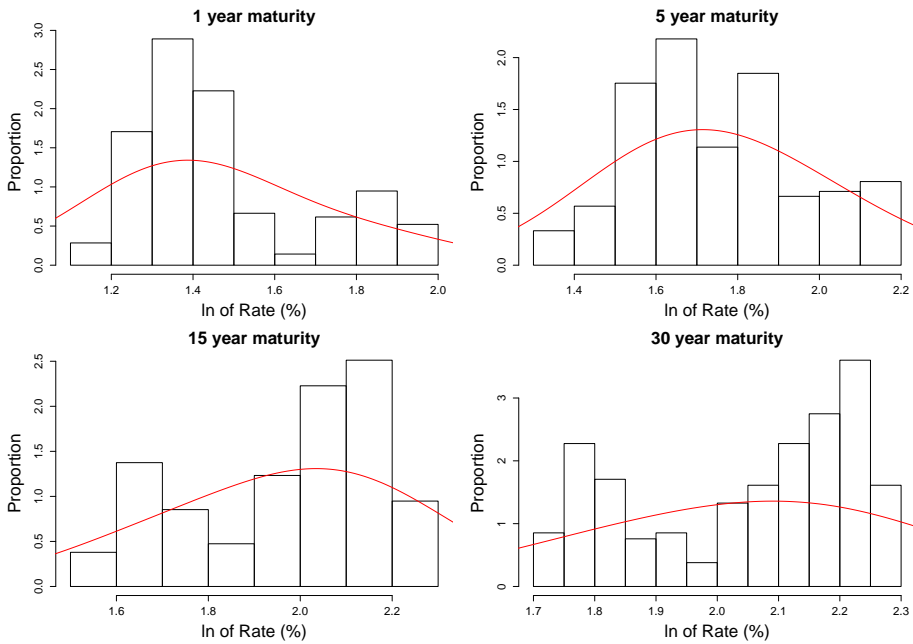


Figure 3.14: Histograms of the logarithm of the rates from 1995-1998.

The  $Q-Q$  plot in figure 3.15 does not indicate a good fit, except for the 5 year data.

The test results, as before, indicate better results with the logarithm than without taking the logarithm for this time interval, showing in higher  $P$ -values, but not high enough for the set to pass as a sample coming from a normal distribution.

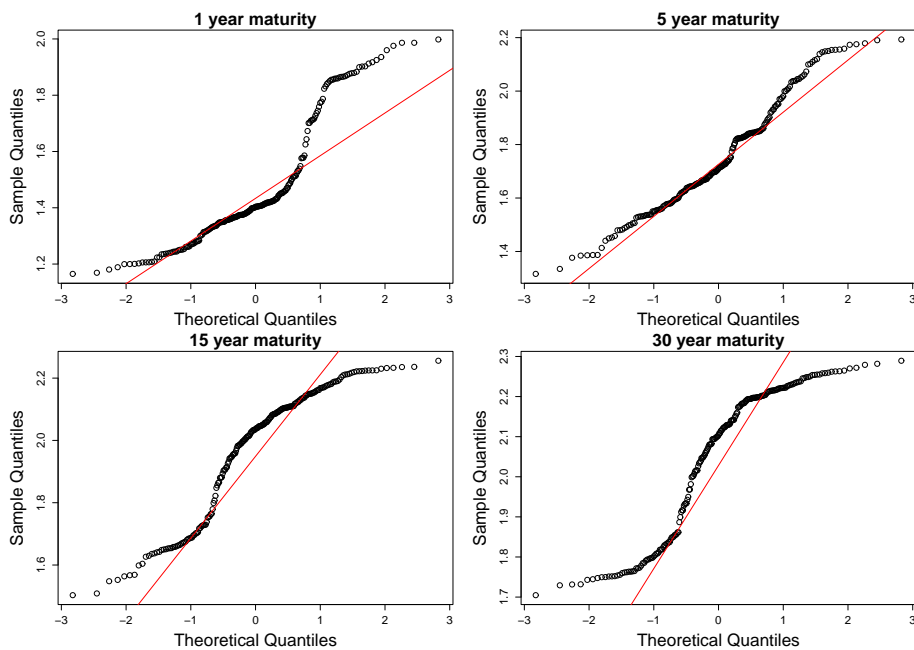


Figure 3.15: Q-Q plot of the logarithm of the rates from 1995-1998.

years	$JB$	$P$ -value	years	$W$	$P$ -value
1	33.5257	5.248e-08	1	0.8689	1.607e-12
5	7.8675	0.01957	5	0.9694	0.0001539
15	20.8803	2.923e-05	15	0.9051	2.447e-10
30	22.5967	1.239e-05	30	0.8821	8.947e-12

Table 3.15: Results of Jarque-Bera test of  $\ln$  of interest rates 1995-1998.Table 3.16: Results of Shapiro-Wilk test of  $\ln$  of interest rates 1995-1998.

### 3.3.4 Normality Test of Data Ranging from 2005– 2006, $\ln$ taken

The histograms in figure 3.16 generally show a flat curve especially for the 15 and 30 year rates that might be caused by fewer rate points used than before. Compared to the same data without the logarithm taken, the curves are similarly flat.

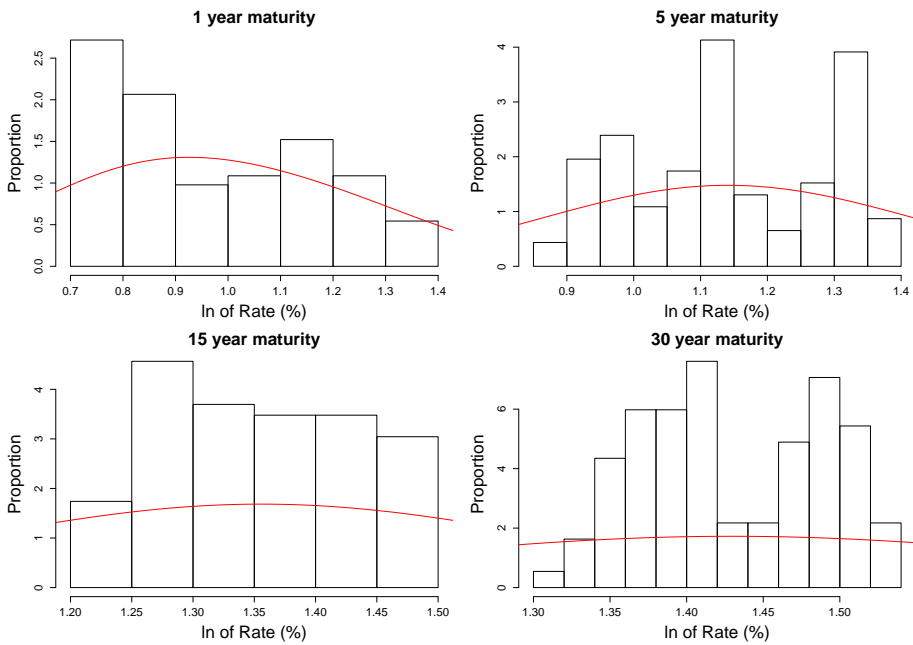


Figure 3.16: Histograms of the logarithm of the rates from 2005-2006.

The  $Q-Q$  plots on the other hand show a fairly good fit to the line compared to other sets and are in fact identical to the  $Q-Q$  plots from the figure 3.9 where the  $\ln$  is not taken.

The results from the goodness of fit tests shown in tables 3.17 and 3.18 indicate the best fit up to now, slightly better than for the same period without the logarithm taken. According to the tests the data would pass as normally distributed.

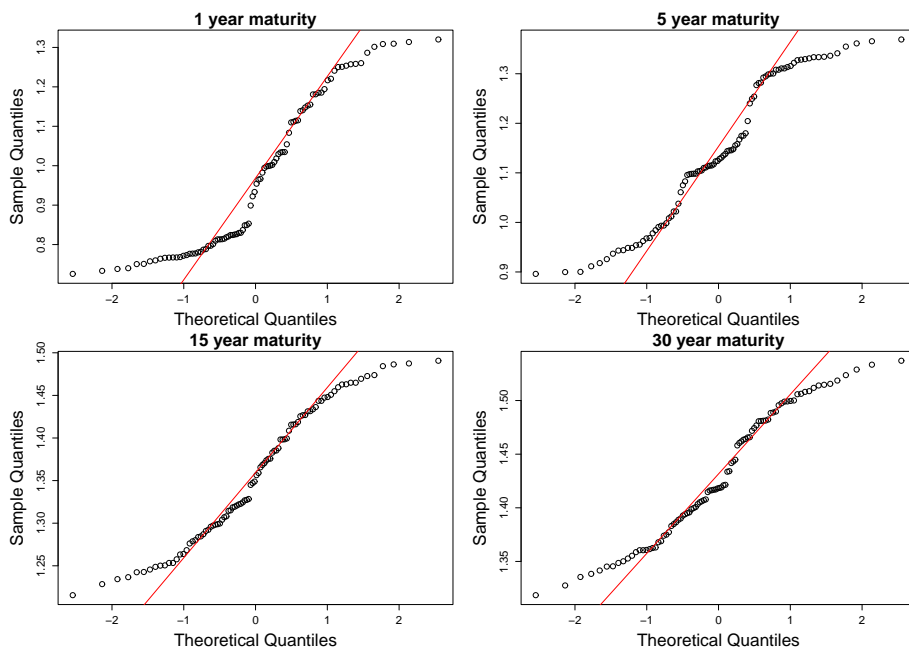


Figure 3.17:  $Q-Q$  plot of the logarithm of the rates from 2005-2006.

years	$JB$	$P$ -value	years	$W$	$P$ -value
1	9.0052	0.01108	1	0.8883	9.99e-07
5	5.9785	0.05032	5	0.9311	0.0001138
15	6.0941	0.0475	15	0.9493	0.001306
30	5.9534	0.05096	30	0.9507	0.001611

Table 3.17: Results of Jarque-Bera test of  $\ln$  of interest rates 2005-2006.

Table 3.18: Results of Shapiro-Wilk test of  $\ln$  of interest rates 2005-2006.

---

## 3.4 Conclusion

Taking the logarithm of the rates resulted in slightly more normally distributed results, but it did not bring any significance gain. It is therefore not quite evident that the interest rates observed are log-normal.

Although the interest rates do not pass as normally distributed according to the tests and visual observations, in the time frame analyzed, the histogram curves of the data had certain characteristics of a normal curve and looked more normal than not. Furthermore from looking at smaller sub periods it was discovered that some of the periods were less normally distributed and some of them more, were the newest data measured was more normal and that should be kept in mind when constructing a model. This fact is possibly due to a stable interest rate level for the last few years.

Taking that into account it is decided not abandon the assumption that the rate data arises from normal distribution, although it is possible that some other type of distribution might be able to describe the rates better, but that has to be up to future work.





# Vector Autoregression

---

In chapter 2 we concluded that three factors were sufficient to describe the term structure of interest rates. In this chapter a model of the term structure which uses those three factors as an input is formulated. The type of model we use is a *vector autoregression* (VAR) model which is a simple but powerful time series model which has proven to be useful for describing the behavior of econometric and financial time series. It is shown that a VAR model of first order is suitable for the data set and based on that we construct a VAR(1) process capable of describing the terms structure. Finally a least square estimation for the parameters in the model is formulated.

The rest of the chapter is laid out as follows:

- In section 4.1 an overview of the necessary time series concepts to formulate a vector autoregression model is given.
- In section 4.2 a formulation of the VAR process is made.
- In section 4.3 we study how the rates are used to proxy the rates.
- In section 4.4 an analysis of the order and stability of the vector autoregression model is conducted.
- In section 4.5 a formulation of a VAR(1) process suitable for generation of the term structure of interest rates is given.

- In section 4.6.1 we formulate a way to estimate the parameters in the VAR(1) process.
- Finally in section 4.7 concludes the chapter.

## 4.1 Stationary, Invertability and White Noise

It should be noted that the presentation of the subject covered in this chapter is primarily based on books by J.D. Hamilton, *Time Series Analysis* (1994) and by H. Madsen (also named) *Time Series Analysis* (2001), these books provide a general overview of time series analysis.

As mentioned the VAR process is a time series process and in order to formulate it it is suitable to introduce some basic time series analysis concepts. Since the VAR process is a linear model of a stochastic process in discrete time space  $t \geq 0$ , we will restrict our coverage of time series analysis subjects to what is needed to cover such a process. In general, for discrete time a time series  $\{x_t; t = 1, \pm 1, \dots\}$  is a realization of a stochastic process  $\{X_t; t = 1, \pm 1, \dots\}$ .

Lets begin by introducing the concept of a *system*. In the scope used here a system maps an input  $x(t)$  in continuous time and  $x_t$  discrete time respectively to an output  $y(t)$  or  $y_t$ . This mapping can be described with an operator  $\mathcal{F}$  often called filter. If the operator is linear the system correspondingly is a linear system and can be defined as:

### DEFINITION 4.1 Linear system

Given two inputs  $x_1$  and  $x_2$  and two scalar values  $\lambda_1$  and  $\lambda_2$  then a system is said to be *linear* if it fulfills

$$\mathcal{F}[\lambda_1 x_1(t) + \lambda_2 x_2(t)] = \lambda_1 \mathcal{F}[x_1(t)] + \lambda_2 \mathcal{F}[x_2(t)].$$

◇

If the system behavior is invariant to changes in time, it becomes time invariant and is described as:

### DEFINITION 4.2 Time invariant system

A system is said to be *time invariant* if

$$y(t) = \mathcal{F}[x(t)] \rightarrow y(t - \tau) = \mathcal{F}[x(t - \tau)]$$

◇

And the system is said to be stable if:

**DEFINITION 4.3 Stable system**

A system is said to be *stable* if any constrained input implies a constrained output.  $\diamond$

For modeling purposes all the properties above are desirable to make a robust model.

If a process is linear, this linear process  $\{Y_t\}$ , can be interpreted as a output from an linear system where the input is so called white noise, such as shown in figure 4.1. Following is a definition of white noise and Gaussian (normal) white noise in discrete time.



Figure 4.1: A digram showing the connections between white noise and a linear process.

**DEFINITION 4.4 White noise**

A sequence  $\epsilon_t$  of mutual uncorrelated identically distributed stochastic variables<sup>1</sup> with mean zero and a constant variance  $\sigma^2$  is called a *white noise process*. This implies

$$\begin{aligned}\mathbb{E}(\epsilon_t) &= 0 \\ \mathbb{E}(\epsilon_t^2) &= \sigma^2 \\ \mathbb{E}(\epsilon_t, \epsilon_{t+k}) &= 0 \text{ for } t \neq 0\end{aligned}\tag{4.1}$$

where  $\mathbb{E}$  is the expected value.  $\diamond$

**DEFINITION 4.5 Gaussian white noise**

If equations 4.1 hold along with the condition

$$\epsilon_t \sim N(0, \sigma^2),$$

that is the process has a probability density function of the normal or Gaussian distribution, the process is said to be a *Gaussian white noise process*<sup>2</sup>.  $\diamond$

<sup>1</sup> In the case of discrete time, a stochastic process amounts to a sequence of random variables i.e. time series.

<sup>2</sup>Note that a Gaussian noise is not necessary a white noise.

Since we will be working with vectors it is furthermore beneficial to give a vector generalization of white noise

$$\begin{aligned}\mathbb{E}(\epsilon_t) &= \mathbf{0} \\ \mathbb{E}(\epsilon_t^2) &= \mathbf{\Omega} \\ \mathbb{E}(\epsilon_t, \epsilon_{t+k}) &= \mathbf{0} \text{ for } t \neq 0\end{aligned}\tag{4.2}$$

where  $\mathbf{\Omega}$  is an  $(n \times n)$  symmetric positive definite matrix of variances.

A stochastic process which is time invariant is said to be stationary. Following is a definition of a stationary processes taken from *Probability and Random Processes* (2001) and note that these definitions are not restrict to discrete time.

**DEFINITION 4.6 Strong stationarity**

The stochastic process  $Y = \{Y(t) : t \geq 0\}$ , taking values in  $\mathbb{R}$ , is called *strongly stationary* if the families

$$\{Y(t_1), Y(t_2), \dots, Y(t_n)\} \text{ and } \{Y(t_1 + h), Y(t_2 + h), \dots, Y(t_n + h)\}$$

have the same joint distribution for all  $t_1, t_2, \dots, t_n$  and  $h > 0$ .  $\diamond$

If a process is strongly stationary it implies that its probability distribution is the same for all time steps  $t$ . This condition can be reduced to include only the first  $k$  moments:

**DEFINITION 4.7 Weak stationarity**

A process  $Y = \{Y(t) : t \geq 0\}$  is said to be *weakly stationary of order  $k$*  if all the first  $k$  moments are invariant to changes in time.  $\diamond$

By tradition a weakly stationary process of order 2 is called weakly stationary and from now on the term weakly stationary process will refer to that i.e. the mean value and the variance are constant over time.

Definition 4.7 implies that for a weakly stationary process both the mean value and the variance of the process are constant over time. A Weakly stationary process of order 2 is also sometimes refereed to as a covariance-stationary process.

## 4.2 The Vector Autoregression Process

Now that we have introduced the concepts needed to model a linear process in discrete time we can go on and formulate the vector autoregression process.

The VAR process is a vector generalization of the well known univariate *autoregression* (AR) process. Whereas the AR process describes interactions between time series variables  $y_t = (y_1, y_2, \dots, y_T)$  measured over the sample period  $t = (0, 1, \dots, T)$ , the VAR process captures the dynamic interactions between *sets* of such time series variables collected into an  $(n \times 1)$  vector  $\mathbf{Y}_t = (y_{1t}, y_{2t}, \dots, y_{nt})^\top$  measured over the sample period  $t = (0, 1, \dots, T)$ .

The VAR process is a vastly used tool in financial and econometric analysis. An example of its usage could be a model which captures the interactions between a country's gross domestic production  $y_p$  and the level of unemployment  $y_u$ , collected in the vector  $\mathbf{Y}_t = (y_p, y_u)^\top$ .

### 4.2.1 The Autoregression process

Since the building block of the VAR process is the AR process we begin by formulating it and then use the AR to formulate the VAR process.

An autoregression process of order  $p$  is an  $\text{AR}(p)$  process which satisfies the difference equation

$$Y_t = c + A_1 Y_{t-1} + A_2 Y_{t-2} + \dots + A_p Y_{t-p} + \epsilon_t, \quad (4.3)$$

where  $Y_t$  is a time series of random variables,  $c$  is a constant,  $A_j$  is a vector of autoregressive coefficients or simply parameters for  $j = (1, 2, \dots, p)$  and  $\epsilon_t$  is a Gaussian white noise sequence. The order of the model,  $p$  is often referred to as *lag* and simply tells how many past time steps are considered in the model.

As an example of an AR process we give a first order autoregression process  $\text{AR}(1)$ . A first order process uses one past observation,  $Y_{t-1}$  (lag one) along with an autoregressive coefficient  $A$  to express the current state.

$$Y_t = c + AY_{t-1} + \epsilon_t$$

The constant  $c$  and the white noise  $\epsilon_t$  are the same as in equation 4.3 before.

It can be shown see e.g. A.4 that if the roots of the AR process

$$1 - A_1 z - A_2 z^2 - \dots - A_p z^p = 0 \quad (4.4)$$

lie outside of the unit circle the process is weakly stationary.

## 4.2.2 The Vector Autoregression process

Now that we have formulated the  $AR(p)$  process for a single time series  $Y_t$  we can generalize it for a vector of time series  $\mathbf{Y}_t$ . The  $AR(p)$  process generalized to a  $p$ th order vector autoregression  $VAR(p)$  process becomes

$$\mathbf{Y}_t = \mathbf{c} + \mathbf{A}_1 \mathbf{Y}_{t-1} + \mathbf{A}_2 \mathbf{Y}_{t-2} + \cdots + \mathbf{A}_p \mathbf{Y}_{t-p} + \boldsymbol{\epsilon}_t \quad (4.5)$$

where  $\mathbf{Y}_t$  is an  $(n \times 1)$  vector of time series of random variables,  $\mathbf{c}$  is an  $(n \times 1)$  vector of constants,  $\mathbf{A}_j$  is an  $(n \times n)$  matrix of autoregressive coefficients for  $j = (1, 2, \dots, p)$  and  $\boldsymbol{\epsilon}_t$  is a vector generalization of Gaussian white noise. Since we intend to formulate a three factor VAR process, we give an example of such a process.

$$\begin{aligned} y_{1,t} &= c_1 + a_{11}y_{1,t-1} + a_{12}y_{2,t-1} + a_{13}y_{3,t-1} + \epsilon_{1,t} \\ y_{2,t} &= c_2 + a_{21}y_{1,t-1} + a_{22}y_{2,t-1} + a_{23}y_{3,t-1} + \epsilon_{2,t} \\ y_{3,t} &= c_3 + a_{31}y_{1,t-1} + a_{32}y_{2,t-1} + a_{33}y_{3,t-1} + \epsilon_{3,t} \end{aligned}$$

### Mean adjusted form

If a process fulfills the conditions of being stationary we can take the expectation of both sides of equation 4.5 to calculate the unconditional mean  $\boldsymbol{\mu}$  of the process

$$\boldsymbol{\mu} = \mathbf{c} + \mathbf{A}_1 \boldsymbol{\mu} + \mathbf{A}_2 \boldsymbol{\mu} + \cdots + \mathbf{A}_p \boldsymbol{\mu}$$

or equally

$$\boldsymbol{\mu} = (\mathbf{I} - \mathbf{A}_1 - \mathbf{A}_2 - \cdots - \mathbf{A}_p)^{-1} \mathbf{c}$$

which means that equation 4.5 can be written in terms of deviations from the mean, sometimes called the mean adjusted form.

$$(\mathbf{Y}_t - \boldsymbol{\mu}) = \mathbf{A}_1(\mathbf{Y}_{t-1} - \boldsymbol{\mu}) + \mathbf{A}_2(\mathbf{Y}_{t-2} - \boldsymbol{\mu}) + \cdots + \mathbf{A}_p(\mathbf{Y}_{t-p} - \boldsymbol{\mu}) + \boldsymbol{\epsilon}_t \quad (4.6)$$

## 4.2.3 Stationary of a VAR model

The next step is to analyze the stationarity of the  $VAR(p)$  process. For that purpose it is helpful to use the so called *companion form* of the VAR process, which is to rewrite the mean adjusted form in equation 4.6 in terms of deviation from a first order VAR process. The  $VAR(p)$  rewritten as a  $np$ -dimensional  $VAR(1)$  process is

$$\boldsymbol{\eta}_t = \mathbf{F} \boldsymbol{\eta}_{t-1} + \mathbf{v}_t$$

where

$$\boldsymbol{\eta}_t = \begin{bmatrix} \mathbf{y}_t - \boldsymbol{\mu} \\ \mathbf{y}_{t-1} - \boldsymbol{\mu} \\ \vdots \\ \mathbf{y}_{t-p+1} - \boldsymbol{\mu} \end{bmatrix},$$

$$\mathbf{F} = \begin{bmatrix} \mathbf{A}_1 & \mathbf{A}_2 & \cdots & \mathbf{A}_{p-1} & \mathbf{A}_p \\ \mathbf{I} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & \cdots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{I} & \mathbf{0} \end{bmatrix}$$

and

$$\mathbf{v}_t = \begin{bmatrix} \boldsymbol{\epsilon}_t \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix}$$

where  $\mathbf{v}_t$  is a vector generalization of white noise such that

$$\begin{aligned} \mathbb{E}(\mathbf{v}_t) &= \mathbf{0} \\ \mathbb{E}(\mathbf{v}_t^2) &= \mathbf{Q} \\ \mathbb{E}(\mathbf{v}_t, \mathbf{v}_{t+k}) &= \mathbf{0} \text{ for } t \neq 0 \end{aligned}$$

and

$$\mathbf{Q} = \begin{bmatrix} \boldsymbol{\Omega} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \end{bmatrix}.$$

The matrix  $\mathbf{F}$  is called the *companion matrix* of a VAR( $p$ ) process, it can be shown (see e.g. *Time Series Analysis* (1994)) that in order for the process to be weakly stationary all the *eigenvalues* of the companion matrix need to lie inside the unit circle. The eigenvalues of  $\mathbf{F}$  are all the  $\lambda$ 's that satisfy

$$|\mathbf{I}_n \lambda^p - \mathbf{A}_1 \lambda^{p-1} - \mathbf{A}_2 \lambda^{p-2} - \cdots - \mathbf{A}_p| = 0 \quad (4.7)$$

So a VAR( $p$ ) process is weakly stationary if  $|\lambda_i| \leq 1$  for all the eigenvalues  $\lambda_i$ ,  $i = (1, 2, \dots, p)$  i.e. the eigenvalues which lie inside the unit circle. Or in reference to equation 4.4 the VAR( $p$ ) is stationary if all values of  $z$  satisfying

$$|\mathbf{I}_n - \mathbf{A}_1 z - \mathbf{A}_2 z^2 - \cdots - \mathbf{A}_p z^p| = 0$$

lie outside the unit circle.

### 4.3 Choosing the Factors in the VAR Model

The main results from the factor analysis in chapter 2 was that three factors were to be used to construct the model of the term structure. But how are the factors recognized in the VAR model? Two methods were considered in this report. The former method is a naive approach which was used in the earlier stages in the construction and analysis of the VAR model. The latter was adopted later on.

The former method is based on taking three positions of the yield curve, a short, medium and long term maturity. The short term rate is chosen as a proxy for the level factor, the slope is chosen as the difference between two yields, a medium maturity yield minus the short maturity yield. And finally the curvature is chosen as two times the medium rate minus the long and short rate. If we note short, medium and long maturity as  $y_s$ ,  $y_m$  and  $y_l$  respectively the factors can be denoted in the following way

$$\begin{aligned} \text{level} &= y_s \\ \text{slope} &= y_l - y_s \\ \text{curvature} &= 2 \cdot y_m - (y_s + y_l) \end{aligned}$$

where we choose the short rate to be the one year rate, the medium to be the five year and the long to be the 30 year rate.

Later on in the work another approach to choose the factors was adopted, suggested by Christiansen & Lund (2007). The main difference is that this approach chooses the curvature of the yield curve differently, namely by using the mechanism of the so called butterfly spread.

A butterfly spread is a portfolio which consists of a long position in an intermediate maturity bond (the body of the butterfly) and two short positions of bonds whose maturities straddle the first bond (the wings of the butterfly). Figure 4.2 shows a diagram of how a butterfly spread looks for a concave (normal) yield curve and the spread  $c$ , is given as

$$c = y_m - (w_1 y_s + (1 - w_2) y_l)$$

where the weights  $w_1$  and  $w_2$  are chosen such that  $w_1 y_s = w_2 y_l$ . An example of how the weights are chosen if the maturities are 1, 5 and 30

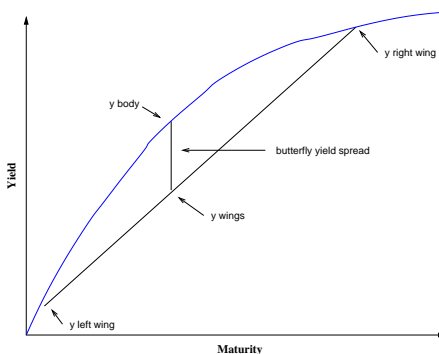


Figure 4.2: A butterfly yield spread.



years would be  $w_1 = (5 - 1)/(30 - 1) = 4/29$  and weight 2 would become  $w_2 = (30 - 5)/(30 - 1) = 25/29$ . The spread shown in the figure is positive and the more concave the yield curve becomes the more positive the spread gets and vice versa. This applies for both normal and inverted yield curves. Equivalently, a negative butterfly spread indicates a convex yield curve.

By the latter method the level is chosen in the same way as before, by taking the short rate as a proxy, but the slope is determined differently compared to the former method. The slope in the latter method is chosen to be the difference between the long and short rate in stead of the difference between the medium and short rate before. That is done in order to keep the correlation between the slope and the approximation of the curvature at a reasonable level, according to Christiansen & Lund (2007). Using the same notation as for the former method

$$\begin{aligned} \text{level} &= y_s \\ \text{slope} &= y_l - y_s \\ \text{curvature} &= y_m - (w_1 y_s + (w_2) y_l). \end{aligned}$$

## 4.4 Analyzing the Order and Stability of a VAR Model

To construct a accurate VAR model it is necessary to estimate a suitable model order and it is necessary for the model of the chosen order to be stationary. For estimation of the order of a VAR model we apply a model order selection criteria, and once we have established a suitable order we check the stability.

### 4.4.1 Suitable order of VAR model of interest rates

When choosing the appropriate model order one wants to balance between using as much available information as possible and the simplicity of the model constructed. To compensate between those aspects we refer to *the law of parsimony*, also known as Ockham's razor, which is commonly stated as

*Of two equivalent theories or explanations, all other things being equal, the simpler one is to be preferred.*

In a statistical sense the law of parsimony is interpreted as a simpler model that describes the data accurately enough is preferred over a more complicated one,

which leaves little of the variability unexplained. Put differently our objective is to choose the lowest order of VAR model which describes the data sufficiently. To achieve that we apply so called *information criteria* which is a statistical technique for estimating the goodness-of-fit of a model, versus its complexity. The general approach of a model selection criteria for a model such as VAR( $p$ ) is to fit models of orders  $i = (0, \dots, p_{max})$ , where  $p_{max}$  is some chosen upper limit of order, and select the value of  $p$  which scores highest in the information criteria. The general procedure of information criteria was derived by Akaike (1947), and the *Akaike information criterion* (AIC) is defined as

$$AIC = -2 \log(\text{maximum likelihood}) + 2(N)$$

where  $N$  is the number of independently adjusted parameters within the model. The first term in the AIC measures the goodness of fit of the model against the data whereas the second term is a penalty function which punishes the candidate model in accordance to the numbers of parameters used.

Under the assumption that the VAR is filtered from a multivariate Gaussian white noise, the likelihood of the noise can be estimated with maximum likelihood estimation of the covariance matrix,  $\tilde{\Sigma}(p) = N^{-1} \sum_{t=1}^T \hat{\epsilon}_t \hat{\epsilon}_t'$ . The AIC criterion for a VAR( $p$ ) model can therefore be formulated as

$$AIC(i) = \ln(|\tilde{\Sigma}_i|) + \frac{2Ni}{T} \quad (4.8)$$

where  $T$  is the number of observations used, or 614. Then for a given time series one chooses the order  $p$  such that the chosen order is  $AIC(p) = \min_{0 \leq i \leq p_{max}} AIC(i)$ , where  $i = (0, \dots, p_{max})$  is predefined.

Alternative criteria, which are in essence just variations of the AIC, have been developed. The two most common of them are the *Bayesian information criterion* (BIC) and *Hannan & Quinn* (HQ) criterion which are based upon taking the log normal of the sample size in various degrees. Their formulas are

$$BIC(i) = \ln(|\tilde{\Sigma}(i)|) + \frac{Ni \ln(T)}{T} \quad (4.9)$$

and

$$HQ(i) = \ln(|\tilde{\Sigma}(i)|) + \frac{Ni \ln(\ln(T))}{T}. \quad (4.10)$$

According to the literature the AIC criteria is considered to have the tendency to overestimate the order, especially when estimating a large number of parameters. The BIC and HQ on the other hand are considered to estimate the order more fairly if the true order  $p$  is less than or equal to the chosen limit  $p_{max}$ .

### A application of information criteria

The AIC, BIC and HQ information criteria were implemented on the same data set of interest rates as used in the factor analysis, ranging from 1995–2006. The criteria are applied on the VAR for orders  $i = (1, 2, \dots, 5)$  in equations

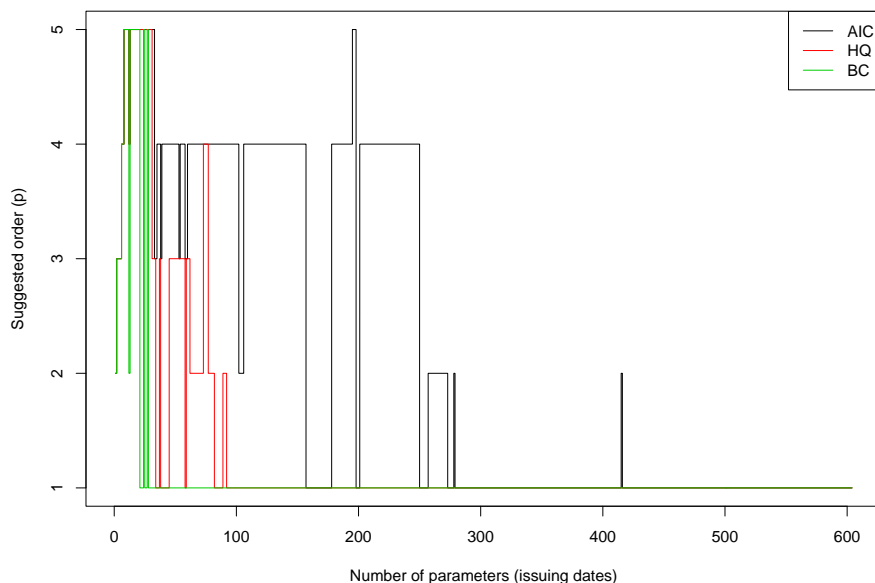


Figure 4.3: The number of lags suggested by the three information criteria considered, plotted against the number of historical interest rate data.

4.8–4.10 for increasing number of observations or issuing dates  $T$ , such that the most recent dates, from 2006, come first and older dates are added incrementally up to 1995.

In figure 4.3 the results of the applications of the information criteria are displayed in a graphical manner. Notice that the outcome of the criteria is a single value for each number of observations but are plotted as lines for easier analysis. From the figure it can be seen that the AIC lives up to its reputation of overestimating the model order compared to the other two. The HQ information criterion suggest a VAR(1) model for all subsets containing more than 91 observations whereas the BIC suggests VAR(1) for all sub sets containing more than 15 observations. Furthermore it can be seen from figure 4.3 that all the criteria suggest order one for a model containing more than 278 observations, with one exception for 415 observations, from the AIC, which as stated before

has the tendency to overestimate. Based on that, and in coordination to the law of parsimony, it can be concluded that a first order VAR model is suitable for modeling the term structure of the Danish interest rate data from the era considered here if we use more than ca. 90 observations.

#### 4.4.2 Stability of VAR model

Now that we have argued that a first order VAR process is suitable the next thing we need to do is check the stationarity of the VAR(1) model. To perform that stability test we use the condition of stationarity given in equation 4.7 stating that if the roots of all eigenvalues of the poles lie within the unit circle, then the process is stationary.

To investigate the stability versus number of observations we construct models containing 10 to 614 observations. The lower limit is chosen to be 10, because a certain number of degrees of freedom are needed for estimating the parameters, which should be at minimum higher than the lag of the model. The upper limit of the observations is the number of observation available in the data set.

Figure 4.4 shows calculations of the module for number of points from 10 to 604 (1996–2006). The data points are added incrementally to the model in the same fashion as used in the information criteria before. Each color in the figure shows the values of the three most significance eigenvalues, which are the decisive ones. It can be seen, as expected, that the first module is dominant. Furthermore a VAR(1) model constructed from less than ca. 60 data points (57 to be exact) tends to be unstable and surprisingly a VAR(1) model constructed from more data than 500 data points tends to be unstable, so the benefit of increasing the number of observed points used for the modeling seems to reach a limit in this case. Processes made from the most recent observations containing between 60 and 500 observation therefore seem to result in a good model in our case, with a critical area around the 300 points mark, but that would be considered an exception.

### 4.5 Construction of a VAR Model

To sum up the results of the previous subsection we have established, based on the sample used, that to construct a stationary VAR model it has to consist of ca. 60 to 500 past data observations. The most important constraint is the requirement of the model to be stationary, and it appears that one has to choose more than ca. 60 past observations and less than ca. 500 to achieve stationary

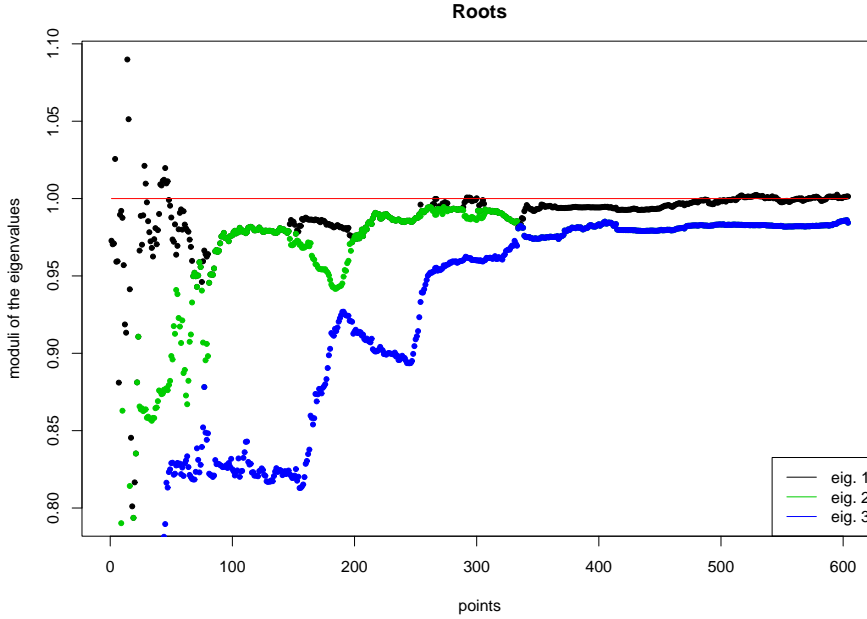


Figure 4.4: The module for the 3 eigenvalues plotted for 10 to 604 observed variables.

in the case considered here.

Furthermore if one wants to use a first order VAR, which is desirable, then it seems to be necessary to choose more than 90 points according to the BIC and HQ criteria, assuming that the AIC criterion overestimates the order. So if we take the results of the information criteria into account we should use somewhere in between 90 and 500 observation to construct a VAR(1) model.

Now that we have established a suitable model order and the number of observations necessary, we now give a relevant formulation of such a model with three factors. A VAR(1) model where  $i$  represents the factors  $i = 1, 2, 3$  in the mean adjusted form is

$$\mathbf{y}_{i,t+1} = \boldsymbol{\mu}_i + \mathbf{A}(\mathbf{y}_{i,t} - \boldsymbol{\mu}_i) + \boldsymbol{\epsilon}_{i,t+1}. \quad (4.11)$$

where  $\mathbf{y}_{i,t}$  is an  $(n \times 1)$  matrix of random time series variables,  $\boldsymbol{\mu}_i$  is an  $(n \times 1)$  vector of the mean of the sets,  $\boldsymbol{\epsilon}_{t+1} \sim N(0, \boldsymbol{\Omega})$  where  $\boldsymbol{\Omega}$  is an  $(n \times n)$  matrix of variances and  $\mathbf{A}$  is an  $(n \times n)$  matrix of factors where the constant  $\mathbf{c}$  in equation 4.5 is set to zero. Notice also that the time steps have been shifted in the equation compared to equation 4.5.

If we let the letters  $l, s$  and  $c$  abbreviate the level, slope and curvature, respec-

tively. Such a model would look like

$$\begin{aligned} y_{l,t+1} &= \mu_l + a_{ll}(y_{l,t} - \mu_l) + a_{ls}(y_{s,t} - \mu_s) + a_{lc}(y_{c,t} - \mu_c) + \epsilon_{l,t+1} \\ y_{s,t+1} &= \mu_s + a_{sl}(y_{l,t} - \mu_l) + a_{ss}(y_{s,t} - \mu_s) + a_{sc}(y_{c,t} - \mu_c) + \epsilon_{s,t+1} \\ y_{c,t+1} &= \mu_c + a_{cl}(y_{l,t} - \mu_l) + a_{cs}(y_{s,t} - \mu_s) + a_{cc}(y_{c,t} - \mu_c) + \epsilon_{c,t+1}. \end{aligned}$$

The VAR(1) model formulated so far can only be used for one-period predictions but we are interested in extending it so that it is capable to predict  $k$  periods ahead, so the model becomes

$$\mathbf{y}_{t+k} = \boldsymbol{\mu} + \mathbf{A}^k(\mathbf{y}_t - \boldsymbol{\mu}) + \boldsymbol{\epsilon}_{t+k} \quad (4.12)$$

where  $\boldsymbol{\epsilon}_{t+k} = N(0, \sum_{i=1}^k \mathbf{A}^{i-1} \boldsymbol{\Omega} (\mathbf{A}^{i-1})^\top)$ . The mean  $\boldsymbol{\mu}$  now becomes the longterm rate. It should also be noted that this the VAR(1) model is discrete in time but continuous in states and that the var model has a built in mean reversion, wich is a valuable propertie in rate modeling<sup>3</sup>

## 4.6 Estimation of the Parameters in a VAR Model

The unknown quantities in the VAR model i.e. the longterm rate  $\boldsymbol{\mu}$ , the matrix of autoregression coefficients  $\mathbf{A}$  and the matrix of variance  $\boldsymbol{\Omega}$ , from the vector of white noise  $\boldsymbol{\epsilon}$ , are parameters. In a statistical sense a parameter refers to quantities that define certain, relatively constant, characteristics of a model. When configuring a model over a specific period of time, such as we aim at doing with the VAR model here, independent variables vary over time, while the parameters are held constant. It is therefore neccesary to estimated the parameters to be able to use the model. The parameters are estimated from past observations and to do that we apply the well known method of *least squares*. Following is a general description of the method of least squares, heavily based on the book *Time Series Analysis* (2001).

If we have  $n$  observations of dependent and independent variables

$$(y_1, x_1), (y_2, x_2), \dots, (y_n, x_n),$$

where  $x$  is the input in the system and  $y$  is the output. Now we intend to estimate the parameters of the system  $\boldsymbol{\theta}$  with an estimator  $\hat{\boldsymbol{\theta}}$  such that  $f(x_t; \hat{\boldsymbol{\theta}})$  describes the observations as good as possible. To do that we apply some measure of closeness, such an estimation of closeness used here is the least square estimation.

---

<sup>3</sup>The tendency of a market variable (such as a interest rate) to revert back to some long-run average level (*Options, Futures, and Other Derivatives* 2006).

The LS measures the closeness by minimizing the *residual sum of squared errors* (*SSE*)  $[y_t - f(x_t; \boldsymbol{\theta})]^2$  of the observations.

The LS estimate can be performed via unweighted least squares estimation, which will be referred to as *ordinary least squares* (OLS) from now on, and *weighted least squares* (WLS). The term unweighted refers to that no consideration is taken to residuals which have larger variance than the average or are correlated i.e. the variance of the residuals is considered constant,  $\Sigma_{ii} = 1$ , and the residuals are mutually independent, where  $\Sigma$  is the correlation matrix. A general form of the OLS method is

$$\hat{\boldsymbol{\theta}} = \arg \min_{\boldsymbol{\theta}} S(\boldsymbol{\theta}) \quad (4.13)$$

where

$$S(\boldsymbol{\theta}) = \sum_{t=1}^n [y_t - f(x_t; \boldsymbol{\theta})]^2 = \sum_{t=1}^n \epsilon_t^2(\boldsymbol{\theta})$$

The assumptions above are considered justifiable for the interest rates.

It is not possible in general to find an explicit expression for the LS estimator and often numeric methods have to be used, however in the case of estimation for a general linear model, such as VAR is, an explicit expression can be found. The LS-estimator has several properties

- It is a linear function of the observations  $\mathbf{y}$
- It is central, i.e.  $\mathbb{E}[\hat{\boldsymbol{\theta}}] = \boldsymbol{\theta}$
- The variance of the estimator is

$$V[\hat{\boldsymbol{\theta}}] = \mathbb{E}[(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^\top] = \sigma^2(\mathbf{x}^\top \mathbf{x})^{-1}$$

- It has the smallest variance off all estimators which are linear functions of the observations

### 4.6.1 Formulation of a OLS-estimation for VAR(1) model

In this section a OLS estimation to estimate the parameters  $\boldsymbol{\mu}$ ,  $\mathbf{A}$  and  $\boldsymbol{\Omega}$  in a VAR(1) model in equations 4.11 and 4.12 is formulated. The model can be written on the form of OLS.

$$\mathbf{Y}_t = \boldsymbol{\alpha} + \boldsymbol{\beta}x_t + \boldsymbol{\epsilon}_t$$

for  $t = 1, \dots, n$  where  $\alpha$  and  $\beta$  denote  $\mu$  and  $A$  respectively, and  $\epsilon_t$  is a white noise in matrix form the model becomes

$$\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

The OLS-estimates are found by solving the *normal equation*.

$$\mathbf{x}^\top \mathbf{x} \begin{bmatrix} \hat{\alpha} \\ \hat{\beta} \end{bmatrix} = \mathbf{x}^\top \mathbf{Y}$$

where  $\hat{\alpha}$  and  $\hat{\beta}$  are the estimators for  $\alpha$  and  $\beta$ , now if  $\mathbf{x}^\top \mathbf{x}$  has full rank then the estimator is

$$\begin{bmatrix} \hat{\alpha} \\ \hat{\beta} \end{bmatrix} = (\mathbf{x}^\top \mathbf{x})^{-1} \mathbf{x}^\top \mathbf{Y}.$$

which minimizes the sum of least squares in the expression  $[\mathbf{Y} - \mathbf{x} \begin{bmatrix} \hat{\alpha} & \hat{\beta} \end{bmatrix}^\top]^2$ , according to equation 4.13. And the estimation for the estimator for the residuals becomes

$$\begin{aligned} res &= \mathbf{Y} - \mathbf{x} \begin{bmatrix} \hat{\alpha} \\ \hat{\beta} \end{bmatrix} \\ \hat{\Omega} &= res^\top \frac{res}{(n-1)} \end{aligned}$$

we now have a estimations for  $\mu$ ,  $A$  and  $\Omega$  which ca be used for constructing a VAR(1) model such as in 4.11 or for the longterm version in equation 4.12.

## 4.7 Conclusion

Analyzing the VAR process reveled that a process with lag 1 was suitable for modeling the rates, based on the results of information criteria. Investigating the stability of the VAR(1) process reviled that it was stable for the time frame of interest, but using all the data was not necessarily better, which is similar to what was concluded about the normal assumption in chapter 3. Finally a tailored VAR(1) model was construct witch can be used for generation of the term structure.



# Scenario Tree Generation

---

In last section we derived a three factor VAR(1) model which could be used for scenario generation. In this chapter the VAR(1) model will be used to sample values of the term structure in discrete states. Those samples are used by an existing scenario generation system which will be used to generate scenarios, analyze them and conduct some experiments on them. The chapter however begins covering scenarios and scenario trees in more detail.

The rest of the chapter is laid out as follows:

- In section [5.1](#) an overview over the representation of scenarios used here is given.
- In section [5.2](#) the issue of scenario tree quality is covered.
- In section [5.3](#) the model used for scenario generation is introduced.
- In section [5.4](#) scenarios are generated and analyzed.
- Finally in section [5.5](#) we conclude the chapter.

## 5.1 Scenarios and Scenario Trees

As mentioned in the introduction, scenarios are usually represented with a *scenario tree*. Scenario trees are a special case of *event trees*, which again are a sub-category of graph trees. In general drawing a graph tree is a graphical way of representing a hierarchical structure and in this section graph theories are used to formulate scenario trees. The next subsection covers the subject of scenario trees and using that formulation we can then define scenarios in more detail than done in chapter 1.

### 5.1.1 Scenario trees

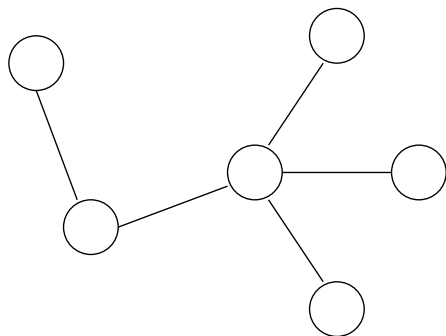


Figure 5.1: An example of a tree with six nodes and five edges.

To explain the construction of a scenario trees we begin by giving short overview of relevant graph theory concepts.

A graph is a set of objects referred to as vertices or nodes, connected by links, called arcs or edges. A *directed graph*, or digraph, is a graph with directed edges. If any two vertices in a graph are connected by *exactly one* edge then the graph fulfills the definition of being a (*graph*) *tree*. A useful proposition about graph tree is that the number of nodes  $v$  and edges  $e$  in a tree are related by  $e = v - 1$ . If the paths connecting the vertices are directed the tree becomes a *directed tree*. The definition of a directed tree

can be constrained further, such that all the edges are obligated to be directed towards, or from, a particular node. In that case the particular node is referred to as a *root* and a tree containing such a root becomes a *rooted tree*. If the vertices are given a unique label the tree is called a *labeled tree*. Note also that in subsequent parts of this thesis the word state is synonym for nodes or vertices in the above.

An event tree is, in accordance to the graph theory concepts listed above, a *labeled rooted directed tree*, with the root at the initial state. A formulation of an event tree which can be used to describe the scenarios which will be generated, taken from the book *Practical Financial Optimization* (2005) is.

**DEFINITION 5.1 Event tree.**

The event tree  $\mathcal{G} = (\Sigma, \mathcal{E})$  is made up of a set of nodes  $\Sigma$ , denoting time and state, and arcs  $\mathcal{E}$ , indicating links between states. At time  $t = (0, 1, \dots, T)$  the states are denoted by  $\Sigma_t = \{s_t^\nu | \nu = 1, 2, \dots, S_t\}$ , where  $S_t$  is the number of possible states at time  $t$ . An event tree has the following properties:

- (i)  $\Sigma_0 = \{s_0^0\}$  is a singleton<sup>1</sup>, and  $s_0^0$  is a unique state known as the *root node*, it has no predecessor.
- (ii) Every state  $s_t^{\nu(t)}$  has unique predecessor from the previous state  $\Sigma_{t-1}$  for all periods  $t = 1, 2, \dots, T$ .

◇

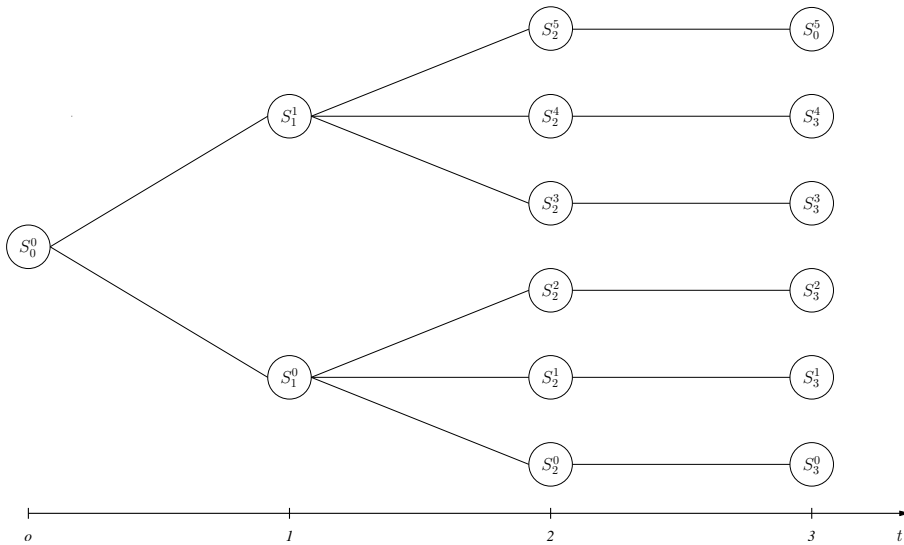


Figure 5.2: Digram of a tree structure showing a tree with three periods, labeled states.

The uniqueness of the predecessors implies that the graph  $\mathcal{G}$  has no cycles and is therefore a tree. An digram of such an event tree can be seen in figure 5.2, where each node, except for the nodes in the last period  $t = T$ , has child nodes, the number of child nodes does not have to be consistent within a level, each node or a state can have different number of child states. However we will only consider trees in which the nodes within a level all have the same number of child states, as shown in figure 5.2 and refer to such a tree a *scenario tree* in the

<sup>1</sup>A singleton is a set which contains exactly one element.

following.

To represent the structure of such a scenario tree in a convenient way a sequence of numbers will be used. The sequence starts at time  $t = 1$  (omitting the root state), and the sequence length denotes the number of periods considered. Each number of the sequence represents the number of states the period has. As an example of this, the tree in figure 5.2 would be denoted a 2-3-1 tree.

### 5.1.2 Scenarios

Now we can use definition 5.1 of the event tree to give a more concrete definition of scenarios than given in definition 1.1 from chapter 1.

**DEFINITION 5.2 Scenarios.**

A scenario is a path of the graph  $\mathcal{G} = (\Sigma, \mathcal{E})$  depicting an event tree, denoted by the sequence  $\{s_0^{\nu(t)}, s_1^{\nu(t)}, \dots, s_{\tau_l}^{\nu(t)}\}$  such that  $(s_t^{\nu(t)}, s_t^{\nu(t+1)}) \in \mathcal{E}$ , for all  $t = 0, 1, \dots, \tau_l, \tau_l < T$ , where  $\tau_l$  is the last period considered in scenario  $l$ , with the associated probability  $p^l \leq 0$ . Each scenario is indexed by  $l$  from sample set  $\Omega$ , and the probabilities satisfy  $\sum_{l \in \Omega} p^l = 1$ .  $\diamond$

## 5.2 The Quality of the Scenario Tree

Now that we have described the scenario trees, some insight into how to estimate the quality of them is given in this section.

A more complete theoretical formulation of the quality of a scenario tree can be found for example in Pflug (2001), Hocreiter & Pflug (2007) and Kaut & Wallace (2007). The outline of the theoretical error measurement of scenarios presented in those papers is that the error of approximating a stochastic process,  $\{\tilde{\xi}_t\}$  by a discretization  $\{\check{\xi}_t\}$  (the scenario tree) for a given stochastic problem is the difference between the optimal solution of the objective function, solved by using the stochastic process  $\{\tilde{\xi}_t\}$ , and the discretization of that process. However the issue of finding such an error is unsolved for practical problems, such as stated is section 1.1. And as the state of the art is today other methods, or procedures, to evaluate the quality of scenario trees generated have to be used.

Along with the discussion of the theoretical approach Kaut & Wallace (2007) mention some general guidelines for evaluating scenario trees. These guidelines include requirements, such as that a given scenario generation method should be able to generate trees which return stable results, meaning that if a large

number of scenarios are generated the trees should all behave similarly in terms of volatility and level of rates. Furthermore an event tree should not contain any biases compared to what is to be expected, should be evenly distributed over the periods, contain no negative rates and they should be plausible compared to the market situation they are generated for. The scenarios generated should also prevent the presence of arbitrage.

Zenios, in the book *Practical Financial Optimization* (2005) mentions three main criteria for estimating the quality of scenarios. The first one being *correctness*, meaning that the behavior of the scenarios should reflect academic theories such as mean reversion and they should be in accordance to the recent history. The second criterion is *accuracy* which involves that a scenario generation should represent the underlying process accurately, for example the moments and the correlation matrix should match. The third criteria Zenios talks about is *consistency* which is relevant when scenarios are generated for a number of financial instruments, meaning that the correlation between them should be accounted for.

Rasmussen & Poulsen (2007) suggest the following consistency criteria for an event tree of yield curves.

1. Some of the lower moments (for example the first two central moments and the correlation) of the factors of the underlying stochastic process of interest rates should be matched as closely as possible within any subtree of the event tree.
2. Some higher moments (for example up to the first four central moments and the correlation) should be matched for any given future period in the event tree conditioned on the initial state at the root node.
3. The actual levels of the generated scenarios should be realistic, for example the tree should not include any negative interest rates or many extreme scenarios.
4. The volatilities of the interest rates of different maturity should be consistent with the implied volatilities of the market benchmark.
5. There should be no arbitrage opportunities in any of the subtrees of the event tree.
6. Types of changes in the shape of the yield curve in future nodes of the event tree should reflect those observed historically from economic regimes which are assumed similar to the one the event tree is built for.
7. The model should be mean reversionary.

8. No volatility clumping; Volatility clumping refers to the case where a period of high volatility is followed by another period of high volatility. Empirical studies have shown that volatility clumping is not present in interest rates contrary to the equity market.
9. The normality assumption is acceptable for interest rates.

This listing more or less sums up the criteria mentioned before and it is therefore useful to use it as a guideline for the quality of generated scenarios.

## 5.3 A Scenario Generation Model

This section introduces a model by Rasmussen & Poulsen (2007) to generate scenario trees. This model is intended to fulfill criteria 1–9 listed in the last section. Note that the model formulated is single period, but can be extended to multi-period with minor changes. This model will be used for generating scenarios in the next section.

### 5.3.1 Formulation of yield curve scenario generation model

A definition of sets, parameters and variables:

SETS:

- $f$ : Set of factors (level slope and curvature),  $f'$  is alias for  $f$ .
- $i$ : Set of zero coupon bonds (ZCB's).
- $i'$ : A subset of the set  $i$  corresponding to the ZCB-rates which define the three factors,  $i'$  is chosen to be a set of 1, 5 and 30 year ZCB's.
- $j$ : Set of parameters of the Nelson-Siegel function; 0 to 3.
- $t$ : Set of time points.
- $s$ : Set of scenarios.

PARAMETERS:

- $Mean_f$ : The mean value factor  $f$ , arriving from the VAR(1) model.
- $Covar_{f,f'}$ : The covariance matrix of the error term, taken from the VAR(1) model.
- $Skewness_f$ : Skewness of factor  $f$ , which is assumed to be zero based on the normality assumption of the VAR(1) model.
- $\tau_i^t$ : Time to maturity for ZCB  $i$  at time  $t$ .

$PP_i^{\text{parent}}$ : Prices of the zero coupon bonds at the root. The prices are calculated using the initial rates  $PP_{ii''}^{\text{parent}} = e^{-r_i \tau_i^{\text{parent}}}$ .

$\psi^{\text{const}}$ : The martingale probability; assumed equal for all scenarios. It is found from the equation  $PP_{ii''}^{\text{parent}} = \sum_s \psi^{\text{Const}}$  where bond  $i''$  matures exactly at the children nodes of the tree with a price of 1.

VARIABLES:

$x_{f,s}$ : A future estimate of factor  $f$  in scenario  $s$  given by the VAR(1) model.

$\mathbb{E}(x)_f$ : The expected value of factor  $f$  over all scenarios.

$\sigma(x)_{f,f'}$ : The covariance matrix of factor  $f$  over all scenarios.

$\mathbb{E}3(x)_f$ : The skewness of factors across all scenarios.

$Y_{i',s}^{(\text{VAR}(1))}$ : The 3 yields comprising the 3 factors at scenario  $s$ .

$NSY_{i',s}$ : The 3 yields comprising the 3 factors at scenarios  $s$  given by the Nelson-Siegel function.

$\varphi_{s,j}$ : Parameter  $j$  of the Nelson-Siegel function at scenario  $s$ .

$R_{i,s}$ : The entire yield curve given by the Nelson-Siegel function at scenario  $s$ .

$CP_{i,s}$ : Price of bond  $i$  at scenario  $s$ .

The overall objective of the optimization model is to match the moments of the underlying stochastic process (the VAR(1) process) as closely as possible.

The objective function minimizes the sums of least squares corresponding to the overall objective of the model:

$$\begin{aligned} \text{Min } & \sum_f (\mathbb{E}(x)_f - \text{Mean}_f)^2 + \sum_f \sum_{f'} (\sigma(x)_{f,f'} - \text{Covar}_{f,f'})^2 + \\ & \sum_f (\mathbb{E}3(x)_f - \text{Skewness}_f)^2 + \sum_s \sum_{i'} (Y_{i',s}^{(\text{VAR}(1))} - NSY_{i',s})^2 \end{aligned} \quad (5.1)$$

The three moments used; expected value, variance and skewness as found by the optimization model are

$$\mathbb{E}(x)_f = \sum_s p_s x_{f,s} \quad \forall f \quad (5.2)$$

$$\sigma(x)_{f,f'} = \sum_s p_s (x_{f,s} - \mathbb{E}(x)_f)(x_{f',s} - \mathbb{E}(x)_{f'}) \quad \forall f, f' \quad (5.3)$$

$$\mathbb{E}3(x)_f = \frac{\sum_s (x_{f,s} - \mathbb{E}(x)_f)^3}{(\sum_s (x_{f,s} - \mathbb{E}(x)_f)^2)^{3/2}} \quad \forall f \quad (5.4)$$

In equations 5.5 the three yields corresponding to the three factors (underlying yields) used in the VAR(1) model are found by the Nelson-Siegel equation. Note

the the final term of the objective function requires that  $NSY_{i',s}$  should be as close as possible to the 3 yields found by the VAR(1) model. So Equation 5.6 calibrates the parameters of the Nelson-Siegel function in accordance to the objective function. These parameters are then used in equation 6 to decide the entire yield curve at each scenario.

$$NSY_{i',s} = \varphi_{s,0} + \varphi_{s,1}e^{-\varphi_{s,3}\tau_{i'}^{parent}} + \varphi_{s,2}\tau_{i'}^{parent}e^{-\varphi_{s,e}\tau_{i'}} \quad \forall i',s \quad (5.5)$$

$$R_{i,s} = \varphi_{s,0} + \varphi_{s,1}e^{-\varphi_{s,3}\tau_i} + \varphi_{s,2}\tau_i^{parent}e^{-\varphi_{s,3}\tau_i^{parent}} \quad \forall i,s \quad (5.6)$$

The VAR(1) model is defined in terms of factors not yields. Equations 5.7 to 5.9 reverse the factors, found by the VAR(1) model, to the yield for each scenario, note that the scales in equation 5.9 are the weights from section 4.3.

$$Y_{1,s}^{VAR(1)} = x_{1,s} \quad \forall s \quad (5.7)$$

$$Y_{30,s}^{VAR(1)} = x_{2,s} + Y_{1,s}^{VAR(1)} \quad \forall s \quad (5.8)$$

$$Y_{5,s}^{VAR(1)} = \frac{4}{29}Y_{30,s}^{VAR(1)} + \frac{25}{29}Y_{1,s}^{VAR(1)} + x_{3,s} \quad \forall s \quad (5.9)$$

Since the yield curve discretization is defined as an optimization model, it enables us to add constraints. One such constraint may be a lower bound in interest rates that is for example no negative rates

$$R_{i,s} \geq 0 \quad \forall i,s$$

Further constraints could be to eliminate arbitrage in interest rates. Equations 5.10 and 5.11 give more restraining condition than no-arbitrage condition, namely that martingale properties should be equal across all scenarios

$$CP_{i,s}^{child} = e^{-R_{i,s}\tau_i^{child}} \quad \forall i,s \quad (5.10)$$

$$PP_i^{parent} = \sum_s \psi^{Const} CP_{i,s}^{child} \quad \forall i \quad (5.11)$$

The model listed in equations 5.1 to 5.11 has the ‘‘shortcoming’’ of being non-linear and non-convex. Such a model therefore has several local minima. Solving such a model falls in the realm of global optimization. And since a general global solver is not available i.e. non existing, solving such a problem would require constructing a specialized solution for the problem at hand, Which is out of current scope of this work.

However as a workaround for solving the global optimization problem, Rasmussen & Poulsen (2007) suggest a approximate approach. In it the solving of the model is divided in to three parts which are solved one after another instead of solving the whole problem in one go, the parts are:

1. First, solve a model comprising the objective function less the 4th term with constraints 5.2–5.4. This model results in discretized factors match-



ing the first 3 moments of the underlying VAR(1) model one period ahead. We also add constraints 5.7 to 5.9 to guarantee no negative rates.

2. Then, solve a second model where the objective function is made of the fourth term and the only constraint is equation 5.5. Finding the parameters of the Nelson Siegel we now use simply equation 5.6 to find the entire yield curves for each scenario.
3. Finally, apply equations 5.10 and 5.11.

The two sub models are again non-linear and non-convex, but it is possible to find optimal solutions to these problems which is done using a standard non-linear solver.<sup>2</sup>

Furthermore, Rasmussen & Poulsen (2007) point out that apart from the no-arbitrage condition solving of the first two part of the approximation would correspond to solving the whole problem. And the solution found may be used as the initial solution of solving the entire problem. That however remains future work.

### 5.3.2 Smoothing the term structure

The model introduced in equations 5.1 to 5.11 uses a the *Nelson-Siegel* function to smooth the term structure of interest rates. It is however known from interest rate theories that the Nelson-Siegel function does not produce arbitrage free curves in any continuous mode. And therefore it not likely that the discretize models become arbitrage free, which is considered a shortcoming of the Nelson-Siegel function and therefore a different smoothing function, the *affine* function is also considered. The affine function is arbitrage free in the continuous setting, therefore it is possible that adding scenarios will also result in arbitrage free scenarios in discrete setting. Following are equations for both of the functions.

#### Nelson-Siegel smoothing

Nelson & Siegel (1987) proposed the following second order differential equation with real, equal roots to fit the yield curve

$$R(m) = \beta_0 + \beta_1 e^{(-m/\tau_1)} + \beta_2 [(m/\tau_2) e^{(-m/\tau)}]$$

---

<sup>2</sup>The solver system used is GAMS.

where  $r(m)$  is the instantaneous forward rate,  $\tau_1$  and  $\tau_2$  time constants associated with the equation and  $\beta_0$ ,  $\beta_1$  and  $\beta_2$  are determined by the initial conditions. In the case of modeling the term structure of interest rates it would be short medium and long maturity of rates used.

### Affine smoothing

Affine term-structure models are interest rate models where the continuously computed spot rate  $R(t, T)$ , at time  $t$  for maturity  $T$  is an affine function in the short rate  $r(t)$

$$R(t, T) = \alpha(t, T) + \beta(t, T)r(t)$$

where  $\alpha$  and  $\beta$  are deterministic function of time. Affine models are popular in the finance literature since they offer convenient forms for bond prices, yields and forward rates.

From generating multi-period scenarios Rasmussen & Poulsen observe ,among other things, that the affine model results in scenarios which follow the normal distribution more closely compared to Nelson Siegel. Furthermore they find that a VAR(1) model branched in the 4-4-4-4 fashion produces to large volatilities for all yields. They on the other hand conclude that a branch of the type 16-4-2-2 gives good approximation for the real world data of Danish ZCB rates from 1995–2006.

## 5.4 Test Case of Scenario Generation

As a test case we now use the model from section 5.3.1, with input parameters estimated with the OLS as described in 4.6.1. The model is used to generate scenarios from two initial dates; the eighth of August 2007 and the third of August 2007. The estimated parameters for the periods considered, can be seen in appendix A.5, but the result of generations are presented in this section.

### 5.4.1 Tree Plots made for 1 period

We begin with generating estimations for a one period tree with the agenda of investigating the effects of arbitrage. The trees are also generated using both types of smoothing of the yield curves considered (Nelson-Siegel and affine), all the combinations are plotted for 4, 8, 16, and 30 scenarios to check for the effects

of the number of scenarios used. The figures are displayed in appendix [A.7](#) and [A.6](#).

For all of the trees generated the same general behavior from arbitrage removal can be observed, removing it affects the structure of the one year rate but the effects, although visible, are much smaller for the five and thirty year rate.

As for the number of scenarios needed, for both of the generation dates and for both of the smoothing functions used, a general observations is that four scenarios are too few. The coverage of the spectrum of the rates is sparse. If eight scenarios are used it seems, in general, to result in evenly distributed scenarios, but the the coverage of the spectrum is sparse, as it is for the four scenarios. If the number of scenarios is further increased up to 16 the generation seems to better cover the spectrum in part of the range but the extreme chases lie to far from the rest for good results. However when using 32 scenarios we get acceptable coverage of the spectrum.

So using more scenarios results in better scenario structure, but one must keep in mind that it is on the cost of more computation time, which can start to count when generating scenarios for more periods and for optimizing a number of scenarios.

### 5.4.2 Simple arbitrage test

To try to quantify the level of arbitrage presented in the generation from the last section a simple arbitrage test was constructed. The test is set up as an optimization problem. The formulation for the simple arbitrage test is.

SETS:

$i$ : Set of zero coupon bonds (ZCB's).

$s$ : Set of scenarios.

PARAMETERS

$PPrice_i$ : Parent prices.

$CPrice_i$ : Child prices.

VARIABLES

$\psi_i$ : Constant, assumed equal for all scenarios.

$surplus$ : Positive variable.

$slack$ : Positive variable.

The objective function is to minimize the slack and surplus of the set of bonds  $i$

$$\text{Min } \sum_i (\text{slack}_i + \text{surplus}_i).$$

Subject to the constraints that the slack and surplus variables are added to the martingale constraint  $\psi$  from equation 5.11, and that the constant has to be larger than zero.

$$\begin{aligned} \text{s.t. } P\text{Price}_i &= \sum_s \psi(C\text{Price}_{i,s} + \text{slack}_i + \text{surplus}_i) && \forall i \\ \psi_s &\geq 0. && \forall s \end{aligned}$$

The results from this test are to be interpreted in the way that if we run the optimization on two sets of scenario tree, the one with smaller number has less arbitrage.

The bond prices in the simple arbitrage test are calculated in a following way, for discrete compounding:

A bond  $P_i$  that pays risk free amounts  $F_{ti}$  at  $t = 1, 2, \dots, T$  calculated as

$$P_i = \sum_{t=1}^T \frac{F_{ti}}{(1+r_t)^t}$$

where  $r_t$  is the rate at time  $t$ , now if we have ZCB's with maturity  $\tau$  the price simplifies to.

$$P_i^{\text{ZCB}} = \frac{F_\tau}{(1+r_\tau)^\tau}. \quad (5.12)$$

The hypothesis is that for affine smoothing the increased number of scenarios will results in less arbitrage. The Nelson-Siegel function however is not expected to behave in such manner. So of special interest is comparison between the smoothing methods used.

This test results are that displayed in tables 5.1 to 5.2.

Scen.	affine	Nelson-Siegel
4	0.794016709519	0.786190621762
8	0.688454007753	0.714680560675
16	0.589911137431	0.601367003083
32	0.913069293782	1.16986729789

Table 5.1: The results of arbitrage test for August 2007.

For the affine smoothing the arbitrage seems to decrease for four eight and sixteen scenarios in the 2007 data put surprisingly the number increases for

Scen.	affine	Nelson-Siegel
4	0.437493359765	0.380037577277
8	0.314511799745	0.164158712646
16	0.410136025530	0.393916431499
32	0.538091112163	0.467338413907

Table 5.2: The results of arbitrage test for August 2005.

the 32 scenarios, the same applies for the Nelson-Siegel smoothing. The results are much less indicating for the 2005 data. But in order to draw any solid conclusions about the affect the type of smoothing has on the level of arbitrage present in scenario tree more work needs to be conducted, especially if one wants to make assumptions for multi periods.

### 5.4.3 Multi-Period Trees

In order to try to cast some light on how a different number of scenarios affect the outcome of the scenarios trees for multi period chases we generate two chases of multi period tree structure. One with a moderate number of 256 scenarios, laid out in a 16-4-2-2 structure and another one containing a lot more or 2048 laid out in a 32-4-4-4 structure. The scenarios are generated from the same dates as in the one period chase (August 2005 and August 2007), with arbitrage removed, and it is chosen to limit the generation to affine smoothing because of the theoretical benefits associated with it and the observation by Rasmussen & Poulsen (2007) reflect that the affine smoothing follows the normal distribution more closely. The results are displayed in figures 5.3 to 5.6.

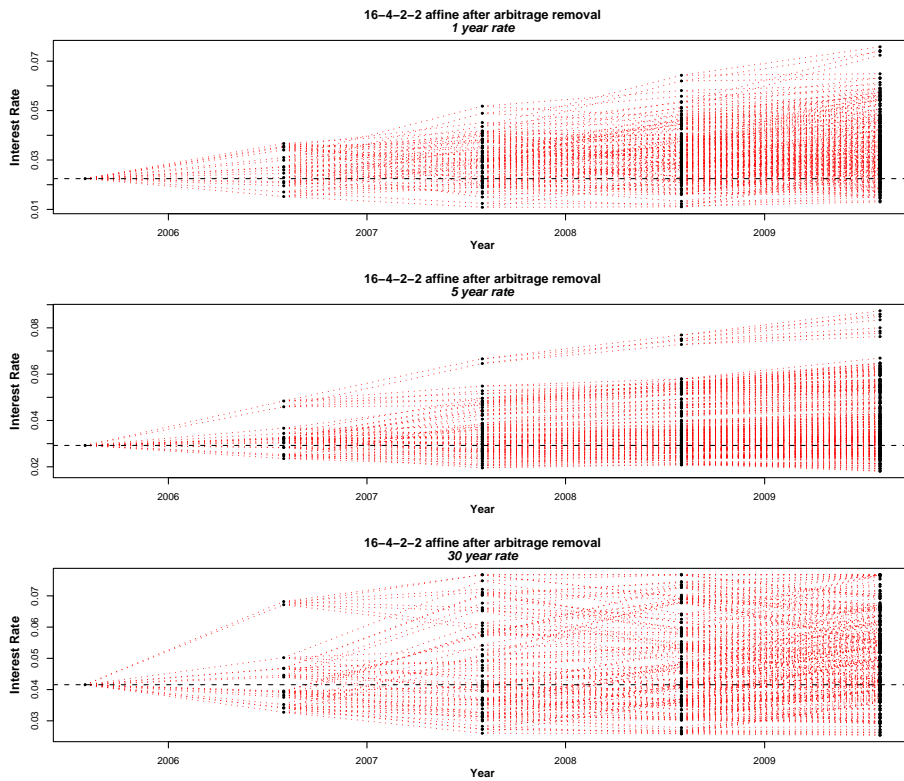


Figure 5.3: August 2005, four periods generated with the 16-4-2-2 structure.

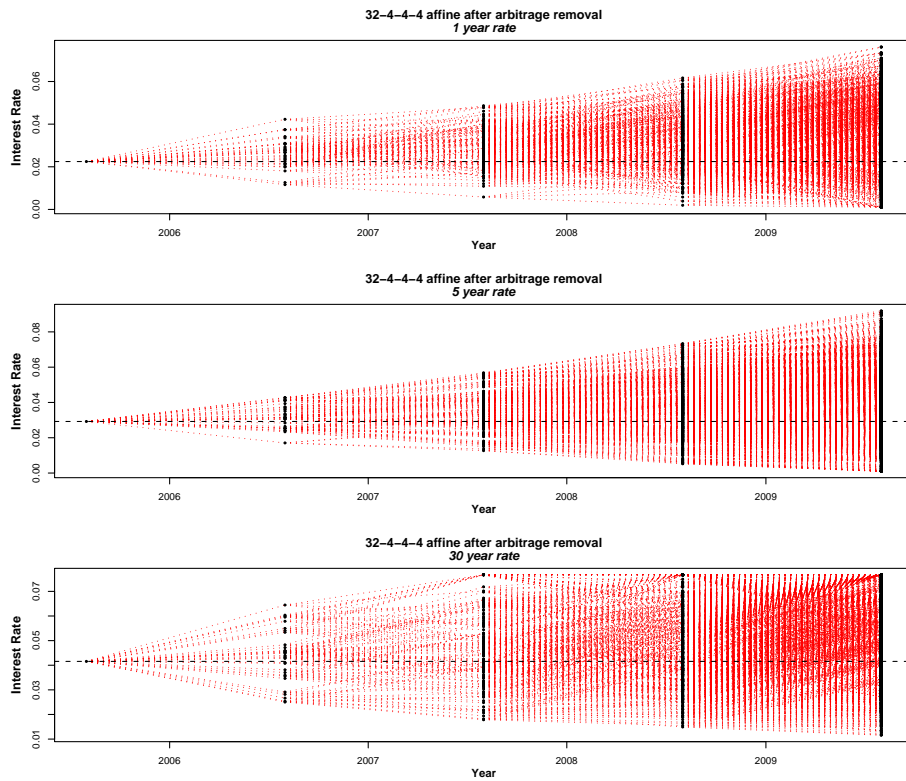


Figure 5.4: August 2005, four periods generated with the 32-4-4 structure.

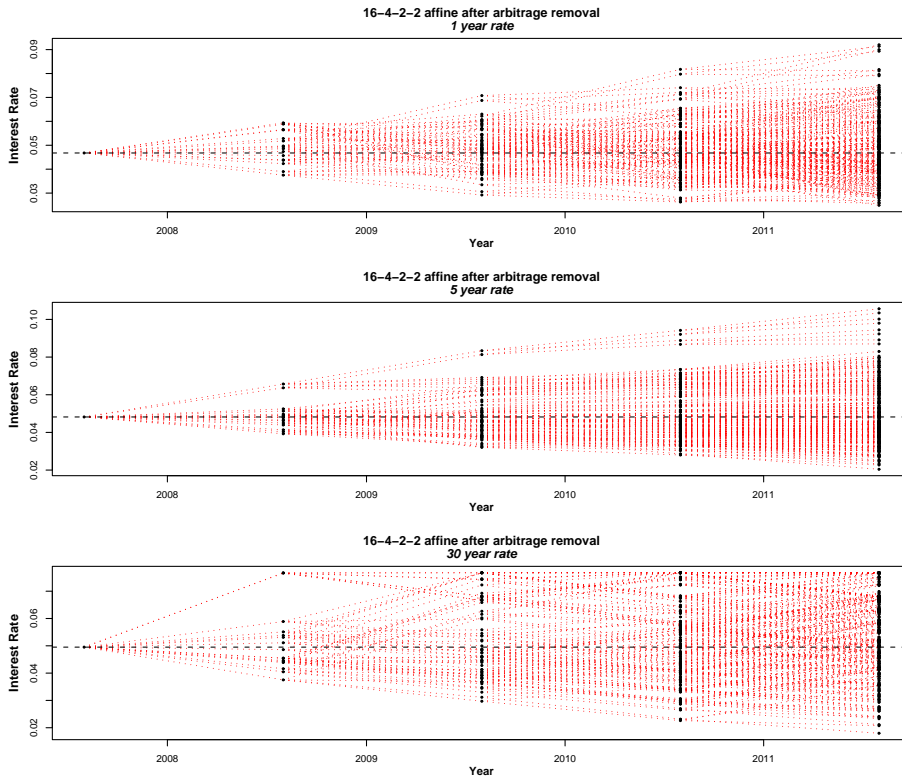


Figure 5.5: August 2007, four periods generated with the 16-4-2-2 structure.



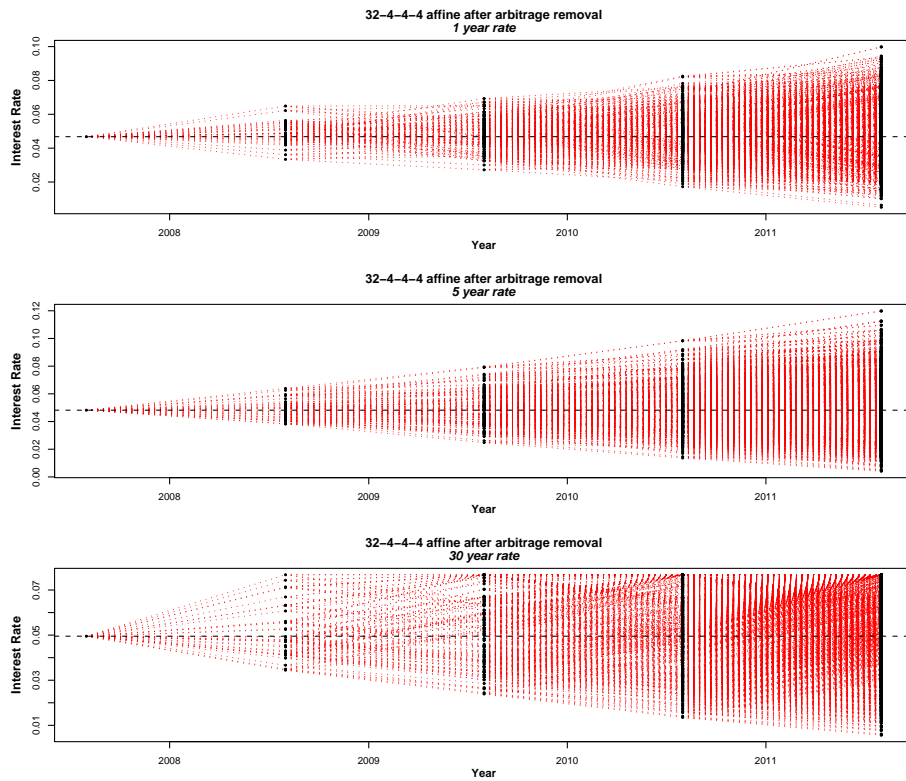


Figure 5.6: August 2007, four periods generated with the 32-4-4-4 structure.

When making comparison between the 16-4-2-2 and the 32-4-4-4 structure it can be seen that the range of rates is slightly larger on average for the 32-4-4-4 structure, but no general significance difference can be observed. However the 16-4-2-2 structure has sparse coverage of the range of the rates which is most clearly visible in the first period followed by sort of outlying scenarios for the higher end of rates generated. This is not wanted behavior and the coverage of the range is significantly better for 32-4-4-4 scenarios. In particular the 32-4-4-4 scenario tree looks better for the 2007 data than the 2005 data. An explanation of that might be that the 2007 tree is generated from of flatter initial yield curve. Although the conclusion which can be drawn is that an increased number of scenarios results in better trees, it should be noted that the 32-4-4-4 structure contains a vast number of scenarios which would result in long optimization solving time, so finding a good tradeoff between the coverage of the rates spectrum and solving time is important. A general notice is also that the thirty year rate seem to hit some sort of a roof which is not preferable.

#### 5.4.4 History Plush Trees

Finally to get some visualization for how the scenarios look compared to the historic behavior of the rates a series of plots, showing simultaneously the historical evolution of the rates and the scenarios generate is made. The scenarios shown are 16-4-2-2 smoothed with the affine function such as made in the section before. Figures 5.7–5.7 show those of plots for one five and thirty year rates as before. Note also that the rates in August 2007 fall in the range of two year prediction generated in August 2005.

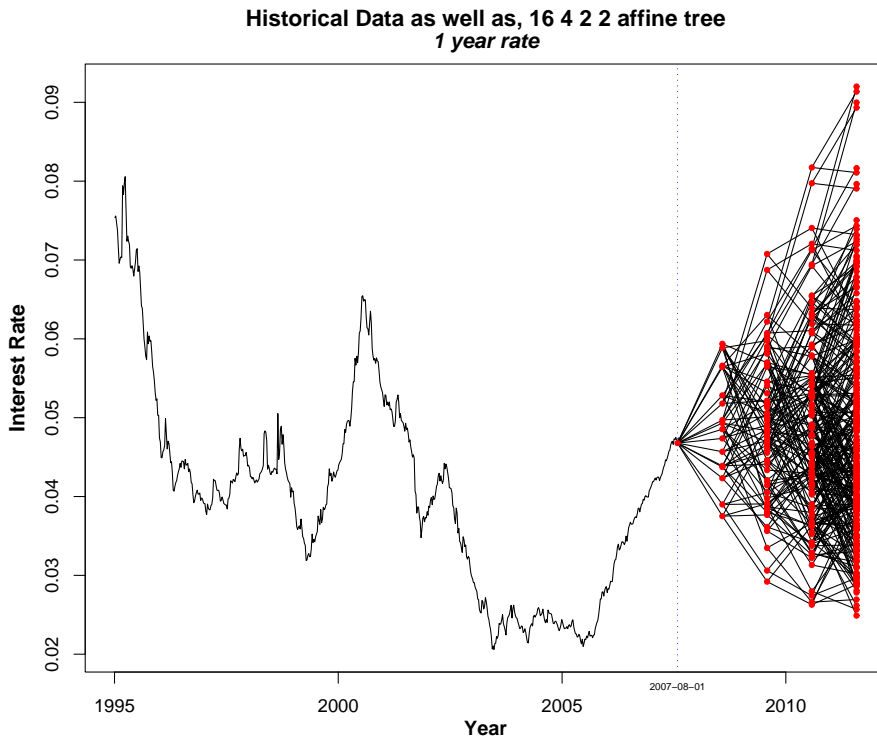


Figure 5.7: August 2007.

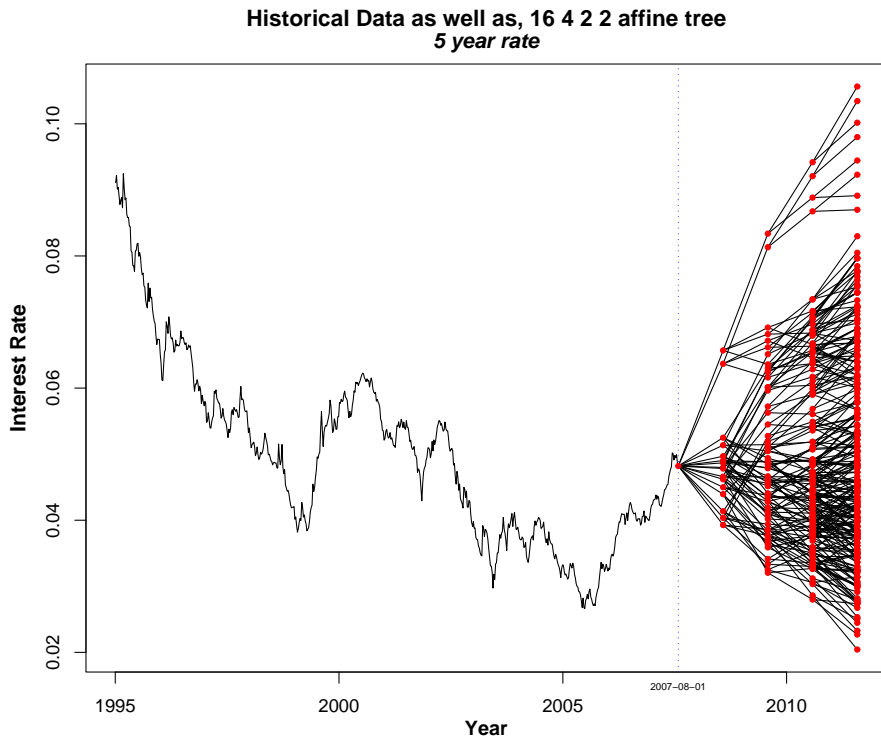


Figure 5.8: August 2007.

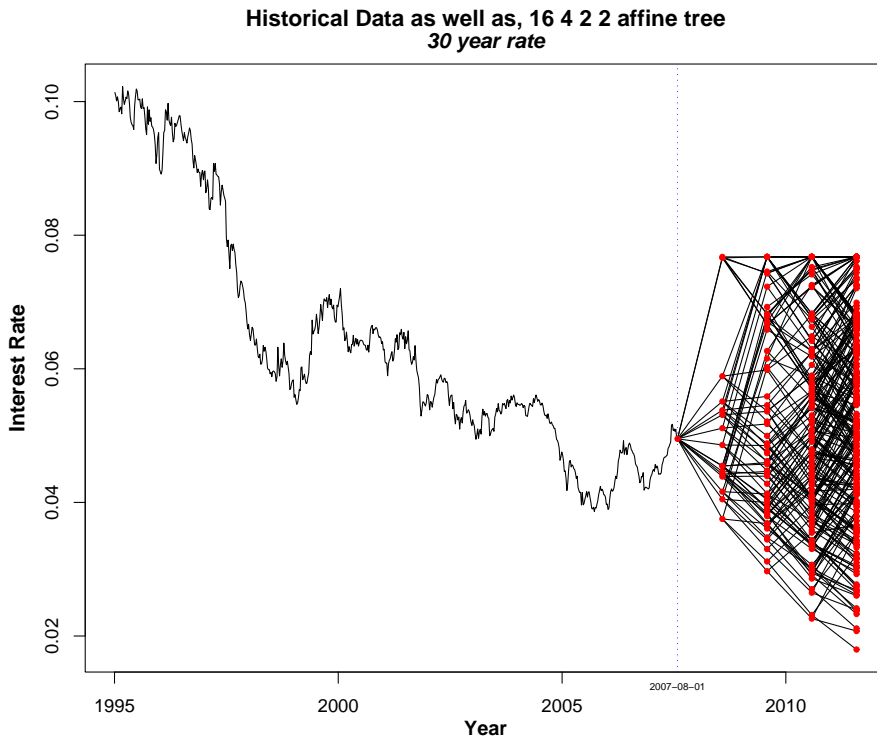


Figure 5.9: August 2007.

## 5.5 Conclusion

In this chapter a detailed formulation of scenarios and scenario trees was given. A model which used the three factor VAR(1) model, derived in chapter 4, to generate scenarios was introduced and used to make several types of scenarios. A comparison between the level of arbitrage resulting from two types of smoothing functions used in the model was made but it did not return decisive results. A comparison was also made of using a moderate number of scenarios vs. a high number of scenarios to construct a tree where it was clear that a high number of scenarios returns better trees. And finally it can be seen that the scenarios generated are on the same quantity as historical observations.

# Conclusion

---

## 6.1 Main findings

The main objective of the thesis was to construct scenario trees of interest rates. We begun by performing a factor analysis of the yield curve and found that the variability of the curve could be described adequately by only three factors. It was also found that the first factor of the Danish curve behaved differently, compared to results from other countries. It had a slope in the first few years instead of being level for all the maturities.

An analysis made on the data showed that log-normality did not describe the rates better for our case.

It was showed that a vector autoregression model could be used to model the underlying process of the yield curve using the factors found. And a way to estimate the parameters in the VAR model was formulated.

The vector autoregression model suitable had order one (VAR(1)) and it was shown that a stable VAR(1) model could be constructed from the term structure data.

A calibrated VAR(1) model was then successfully used to generate scenario trees, both in one period and in multi period settings.

From generating the scenarios, with and without arbitrage removal it was evident that there was arbitrage present in the generated trees and removing it would therefore be a vital part of a scenario generation and optimization process.

From generating large number of scenarios vs. generating a small number of scenarios it was clear result that the tree with a moderate number of scenarios did have some defects and making tree with high number reduced or eliminated those defects. So to get adequate results for optimization one needs to use enough scenarios.

## 6.2 Future work

### Possible improvements

There are possible potentials for improving the VAR model. One of them could be to take the difference of the interest rates and model them, meaning that one would model the change of the rates in stead of the level of rates at any given time, but whether that results in a better model is a question of further observation. It could also be worth investigating further how the length of the period of rates, used for the calibration of the model, would affect the outcome and furthermore how the tree construction affects the outcome. But it has to be kept in mind that interest rates do not fluctuate in nice periodic way according to the seasons, contrary to many other types of data, so comparing two seasons or adapting results from one period to another can become a dodgy business. However the the quality of optimization using the generated scenarios weights the most, so comparisons count the most. Finally it could be worth making comparisons to how scenarios generated by some other process perform compared to the VAR.

### Into the void

Apart from the possible fine tuning raised here above more interesting potential lies in conducting more work in estimation of the scenarios generated. To begin with, it might be well worth to use them to run optimization on portfolios and see how the results turn out.

Furthermore it is especially interesting to investigate in detail how the generated scenarios perform for multi period cases, such as shown in figure 6.1. Questions witch are worth asking are how the type of smoothing function affects the arbitrage? How tweaking of the underlying stochastic model affects



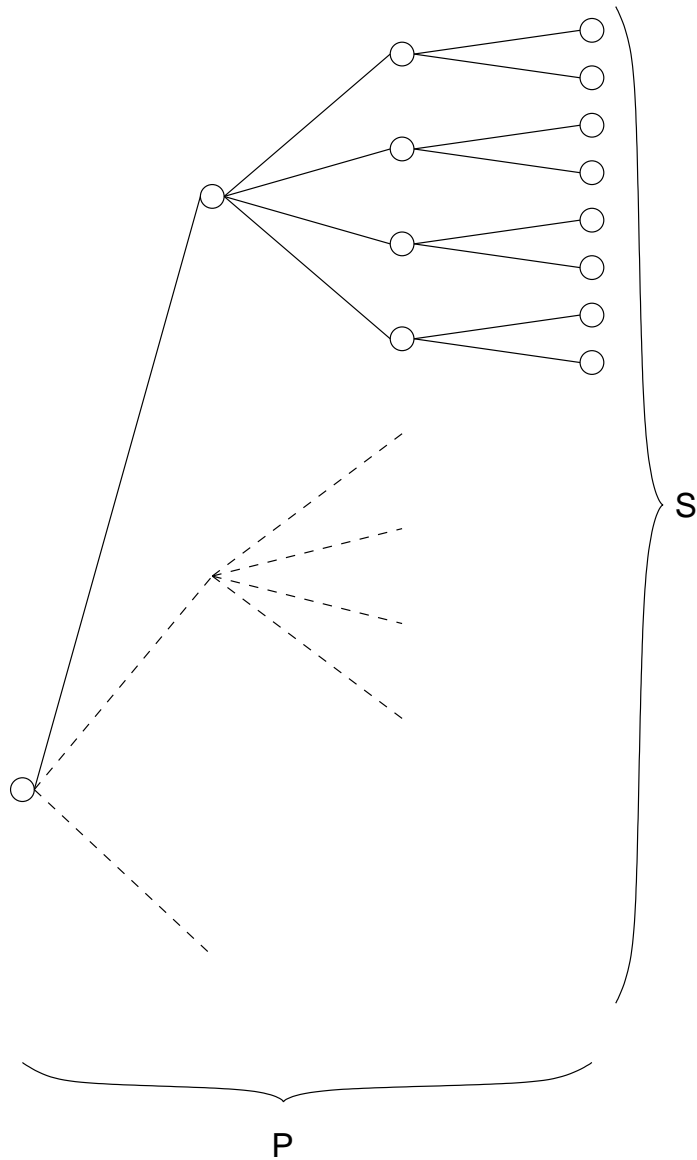


Figure 6.1: A digram of the a multi period problem, S referees to scenarios and P referees to the periods.

the term structure? And how different trees perform in optimization, where a related work could also be to find a non-general solution of the optimization problem in equations 5.1 to 5.11 which is now solved with approximations.

## APPENDIX A

# Further Results

---

### A.1 PCA, the Eigenvectors for 1995–2006

In the following the three largest eigenvalues and eigenvectors found with the PCA for the period 1995–2006, the first three are shown in figure [2.4](#).

```
> pca.prcomp.95.06$sdev[1:4]
[1] 5.3350439 1.1901862 0.3069631 0.1500022

> sum(pca.prcomp.95.06$sdev)
[1] 7.110224

> pca.prcomp.95.06$rotation[,1:4]
t      PC1      PC2      PC3      PC4
1  0.1426241  0.5293120636 -0.425361252  0.54467993
2  0.1586533  0.4452395851 -0.155949833  0.01666234
3  0.1683207  0.3682501490 -0.005428379 -0.21108618
4  0.1756762  0.2894477260  0.090169544 -0.28661033
5  0.1806667  0.2179523074  0.142316691 -0.26702293
6  0.1837765  0.1560215911  0.170477794 -0.23370559
7  0.1855677  0.1050615795  0.177594352 -0.19980076
```

---

8	0.1864985	0.0643999109	0.184151875	-0.14219372
9	0.1869447	0.0281257645	0.199329610	-0.04303345
10	0.1870124	-0.0003624621	0.213623944	0.05460835
11	0.1869201	-0.0208270441	0.218450756	0.12970364
12	0.1867961	-0.0371618002	0.210803692	0.17682810
13	0.1866792	-0.0508382372	0.193077045	0.20048241
14	0.1865757	-0.0624658518	0.168191779	0.20702130
15	0.1864796	-0.0724398125	0.139119554	0.20091853
16	0.1863868	-0.0809853185	0.106818152	0.18602378
17	0.1862889	-0.0884172528	0.072896359	0.16480060
18	0.1861802	-0.0949531114	0.038276516	0.13913007
19	0.1860594	-0.1006768578	0.003599528	0.11067244
20	0.1859248	-0.1057407095	-0.030034492	0.08070087
21	0.1857764	-0.1102745604	-0.062575202	0.04994467
22	0.1856168	-0.1142894334	-0.094114986	0.01923643
23	0.1854467	-0.1179231132	-0.123986709	-0.01092322
24	0.1852683	-0.1212237252	-0.152188240	-0.04014136
25	0.1850838	-0.1242364359	-0.178808673	-0.06818427
26	0.1848962	-0.1270351041	-0.203282675	-0.09468147
27	0.1847086	-0.1295760540	-0.226182026	-0.11948811
28	0.1845216	-0.1319290256	-0.247477891	-0.14270396
29	0.1843362	-0.1341224371	-0.267309912	-0.16442652
30	0.1841524	-0.1361673793	-0.286016107	-0.18485368

## A.2 PCA, Adding One Year at a Time

Figures A.1, A.2 and A.3 show the results of factor analysis starting at 1995 and adding one year at a time.

Note that the sign of the factor loadings is arbitrary when calculated with the R function used, see appendix B.2, that does however not matter when interpreting the factors for the purpose of the recognizing the factors, but it would be possible to find the correct signs.

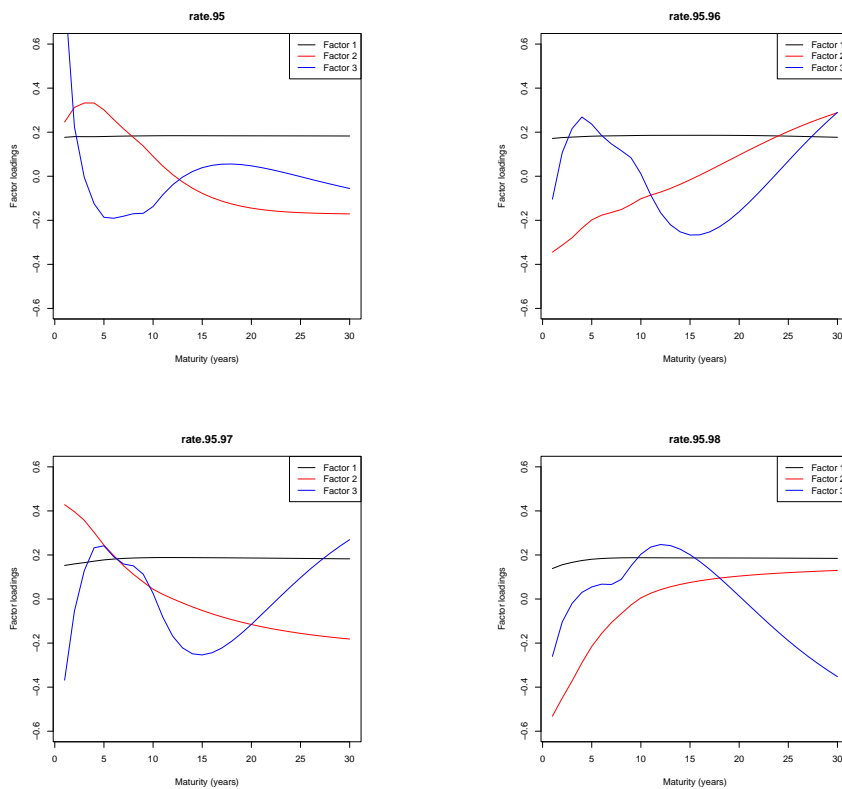


Figure A.1: Factor analysis, adding one year at time, 1995-1998.

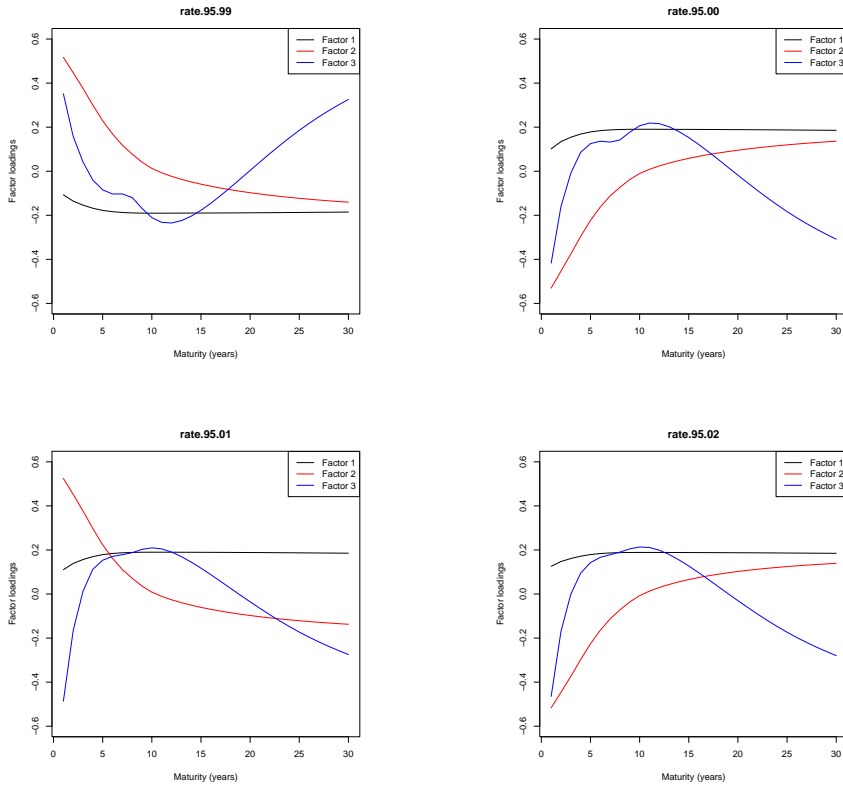


Figure A.2: Factor analysis, adding one year at time, 1995-2002.

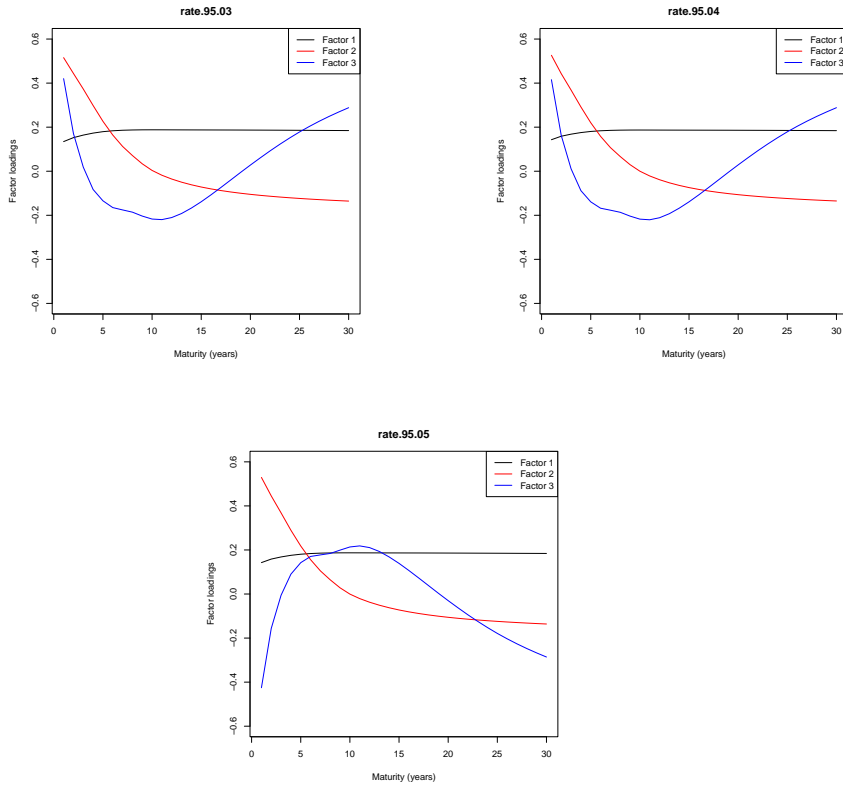


Figure A.3: Factor analysis, adding one year at time, 1995-2005.

### A.3 PCA, Performed Before and After the Changeover to Euro

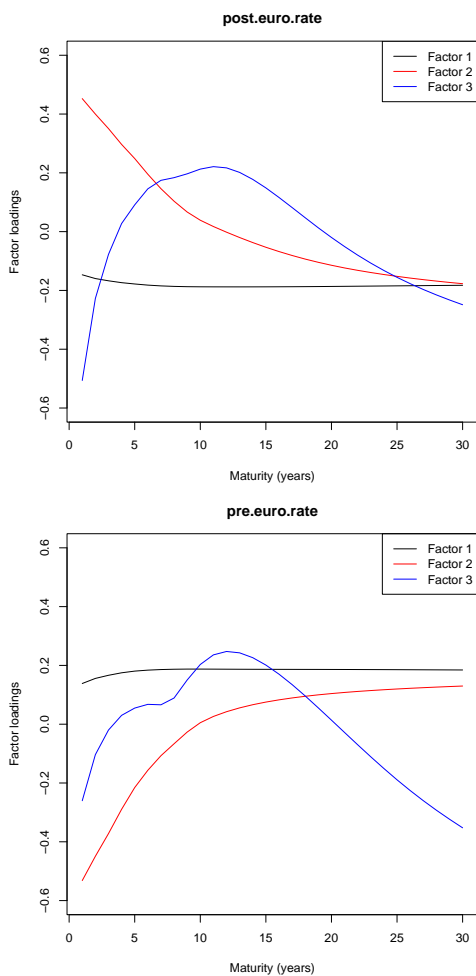


Figure A.4: The factor loading for before and after the introduction of the euro, 1. of January 1999. The Danish krona is pegged to the euro.



## A.4 Roots of $p$ th Order Difference Equations

Definitions of stability for the AR and VAR model, taken from the book *Time Series Analysis* (1994).

First let's define the lag operator; suppose that we start with sequence  $\{x_t\}_{t=-\infty}^{\infty}$  and generate new sequence  $\{y_t\}_{t=-\infty}^{\infty}$  such that  $y_t = x_{t-1}$ . This can be described as applying lag operator to  $\{x_t\}_{t=-\infty}^{\infty}$  which is represented with the symbol  $L$ :

$$Lx_t = x_{t-1}$$

Now if we apply the operator twice we get

$$L(Lx_t) = L(x_{t-1}) = x_{t-2}$$

Such lag operator is represented  $L^2$ , and in general for any integer  $k$

$$L^k x_t = x_{t-k}.$$

Now consider the  $p$ th order difference equation where  $w$  is variable, an example of such equation is the autoregression process given in equation 4.3

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + w_t$$

or in terms of lag operators

$$(1 - \phi_1 L + \phi_2 L^2 - \cdots - \phi_p L^p) y_t = w_t. \quad (\text{A.1})$$

If we factor the operator on the left side of A.1 we get

$$(1 - \phi_1 L + \phi_2 L^2 - \cdots) - \phi_p L^p = (1 - \lambda_1 L)(1 - \lambda_2 L) \cdots (1 - \lambda_p L).$$

This is the same as finding the values of  $(\lambda_1, \lambda_2, \dots, \lambda_p)$  such that the following polynomials are the same for all  $z$ :

$$(1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p) = (1 - \lambda_1 z)(1 - \lambda_2 z) \cdots (1 - \lambda_p z)$$

Now multiply both sides of the equation by  $z^{-p}$  and define  $\lambda \equiv z^{-1}$

$$(\lambda^p - \phi_1 \lambda^{p-1} - \phi_2 \lambda^{p-2} \cdots - \phi_p \lambda - \phi_p) = (\lambda - \lambda_1)(\lambda - \lambda_2) \cdots (\lambda - \lambda_p) \quad (\text{A.2})$$

Clearly, setting  $\lambda = \lambda_i$  for  $i = 1, 2, \dots, p$  causes the the right side of equation A.2 to equal zero. Thus the values  $(\lambda_1, \lambda_2, \dots, \lambda_p)$  must be the numbers that set the left site of equation A.2 to zero as well:

$$\lambda^p - \phi_1 \lambda^{p-1} - \phi_2 \lambda^{p-2} \cdots - \phi_p \lambda - \phi_p = 0 \quad (\text{A.3})$$

The difference equation A.2 is stable if the eigenvalues (the roots of A.3) lie inside the unit circle, or equivalently if the roots of

$$1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p = 0$$

lie outside the unit circle.

## A.5 The Parameters Estimated For Scenario construction

The parameters estimated via the ordinary least square estimation as formulated in section 4.6.1, for  $\boldsymbol{\mu}$ ,  $\mathbf{A}$  and  $\boldsymbol{\Omega}$  and used for tree construction are displayed here.

### A.5.1 August 2005

These are the parameters used for trees generated from 2005-08-03.

$$\boldsymbol{\mu} = \begin{bmatrix} 2.2260555 \\ 2.1573571 \\ 0.4980051 \end{bmatrix},$$

$$\mathbf{A} = \begin{bmatrix} 0.996886985 & -0.01491038 & 0.04553341 \\ -0.002193948 & 1.01263395 & -0.06281597 \\ -0.001929586 & 0.01634335 & 0.95151353 \end{bmatrix},$$

$$\boldsymbol{\Omega} = \begin{bmatrix} 0.009465929 & -0.006101418 & -0.001015083 \\ -0.006101418 & 0.013906296 & 0.003610049 \\ -0.001015083 & 0.003610049 & 0.004027632 \end{bmatrix}.$$

The parameters solved with the restriction that the coefficient matrix  $\mathbf{A}$  should be zero outside the diagonal used as input into the scenario.

$$\boldsymbol{\mu}^R = \begin{bmatrix} 2.1305334 \\ 1.7879579 \\ 0.3902231 \end{bmatrix},$$

$$\mathbf{A}^R = \begin{bmatrix} 0.9969801 & 0.0000000 & 0.0000000 \\ 0.0000000 & 0.9920899 & 0.0000000 \\ 0.0000000 & 0.0000000 & 0.9887856 \end{bmatrix},$$

$$\boldsymbol{\Omega}^R = \begin{bmatrix} 0.009586673 & -0.006268014 & -0.001110070 \\ -0.006268014 & 0.014141642 & 0.003751184 \\ -0.001110070 & 0.003751184 & 0.004140722 \end{bmatrix}.$$

The initial values for the term structure and the initial one five and thirty year rates are shown in tables A.1 and A.2 below.

Year	Init. Yield
1	2.24563463588144
5	2.9220541624358
30	4.1558371897765

Table A.1: The initial yields used in generation from 2005-08-03.

Year	Rate	Year	Rate
0	2.07552070845439	16	3.90933708615459
1	2.24563463588144	17	3.94966345971742
2	2.44694394887622	18	3.98466937618518
3	2.62082531962130	19	4.01525991031714
4	2.77685365664431	20	4.04231274106401
5	2.92205416243580	21	4.06652064416491
6	3.05939283570753	22	4.08788946640350
7	3.18739516721598	23	4.10625172326467
8	3.31092696624586	24	4.12144928320469
9	3.42075020847661	25	4.13333343560961
10	3.51913746141810	26	4.14190061805340
11	3.60507431053632	27	4.14773089857473
12	3.68055215085478	28	4.15154037523880
13	3.74839716219132	29	4.15402031366680
14	3.80915018081960	30	4.15583718977650
15	3.86275865264802		

Table A.2: The initial term structure used in generation from 2005-08-03 (yearly steps).

### A.5.2 August 2007

These are the parameters used for trees generated from 2007-08-1.

$$\boldsymbol{\mu} = \begin{bmatrix} 3.5851278 \\ 1.8714276 \\ 0.4047298 \end{bmatrix},$$

$$\mathbf{A} = \begin{bmatrix} 0.986428219 & -0.02131882 & 0.04933514 \\ 0.006998918 & 1.02092581 & -0.06908143 \\ -0.001699929 & 0.01888504 & 0.94683984 \end{bmatrix},$$

$$\boldsymbol{\Omega} = \begin{bmatrix} 0.012265425 & -0.007130835 & -0.001535627 \\ -0.007130835 & 0.014760175 & 0.004490134 \\ -0.001535627 & 0.004490134 & 0.004653980 \end{bmatrix}.$$

The parameters solved with the restriction that the coefficient matrix  $\mathbf{A}$  should be zero outside the diagonal used as input into the scenario.

$$\boldsymbol{\mu}^R = \begin{bmatrix} 3.0454845 \\ 2.7755317 \\ 0.5797868 \end{bmatrix},$$

$$\mathbf{A}^R = \begin{bmatrix} 0.9908742 & 0.0000000 & 0.0000000 \\ 0.0000000 & 0.9952236 & 0.0000000 \\ 0.0000000 & 0.0000000 & 0.9875937 \end{bmatrix},$$

$$\boldsymbol{\Omega}^R = \begin{bmatrix} 0.012437989 & -0.007352452 & -0.001684323 \\ -0.007352452 & 0.015072957 & 0.004656096 \\ -0.001684323 & 0.004656096 & 0.004831483 \end{bmatrix}.$$

The initial values for the term structure and the initial one five and thirty year rates are shown in tables [A.3](#) and [A.4](#) below. The initial term structure (in

<b>Year</b>	<b>Init. Yield</b>
1	4.67758712285553
5	4.82048201646060
30	4.95182577810421

Table A.3: The initial yields used in generation from 2007-08-1.

yearly steps) is displayed in table [A.4](#) below.

<b>Year</b>	<b>Rate</b>	<b>Year</b>	<b>Rate</b>
0	4.46265712494155	16	5.01467665942645
1	4.67758712285553	17	5.02148714812988
2	4.78427270635327	18	5.02517560051385
3	4.80580369089900	19	5.02630437300192
4	4.81161770772825	20	5.02541385480278
5	4.82048201646060	21	5.02295964720859
6	4.82740071431781	22	5.01911289998022
7	4.84363113854641	23	5.01396861194591
8	4.86507010601946	24	5.00761939525085
9	4.88337973756895	25	5.00015547105682
10	4.90652976823238	26	4.99168599803471
11	4.92985125096813	27	4.98240856527164
12	4.95155053353282	28	4.97253748101162
13	4.97173763255186	29	4.96227786999477
14	4.98960912990309	30	4.95182577810421
15	5.00416026118809		

Table A.4: The initial term structure used in generation from 2007-08-01 (yearly steps).

## A.6 Tree Plots, i Period, August 2005

### A.6.1 Nelson-Siegel smoothing

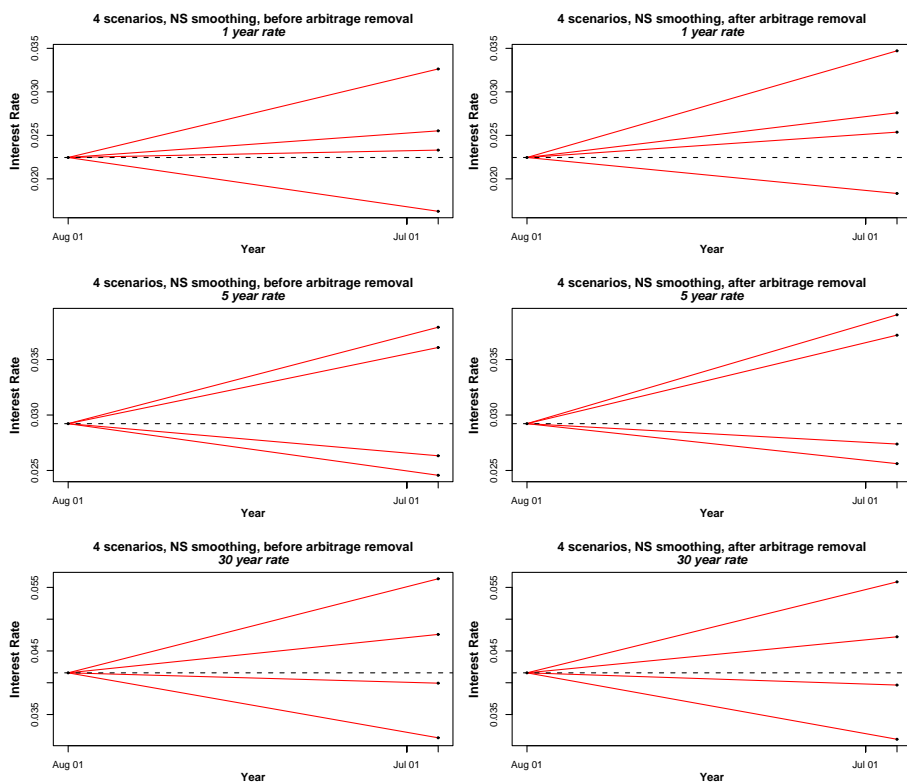


Figure A.5: August 2005.

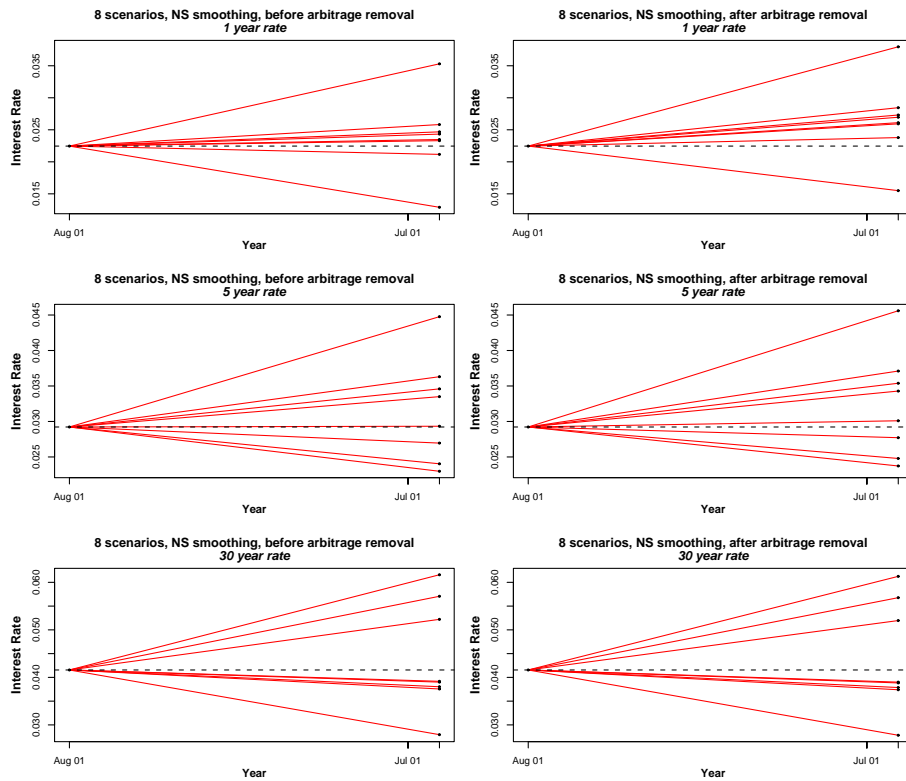


Figure A.6: August 2005.

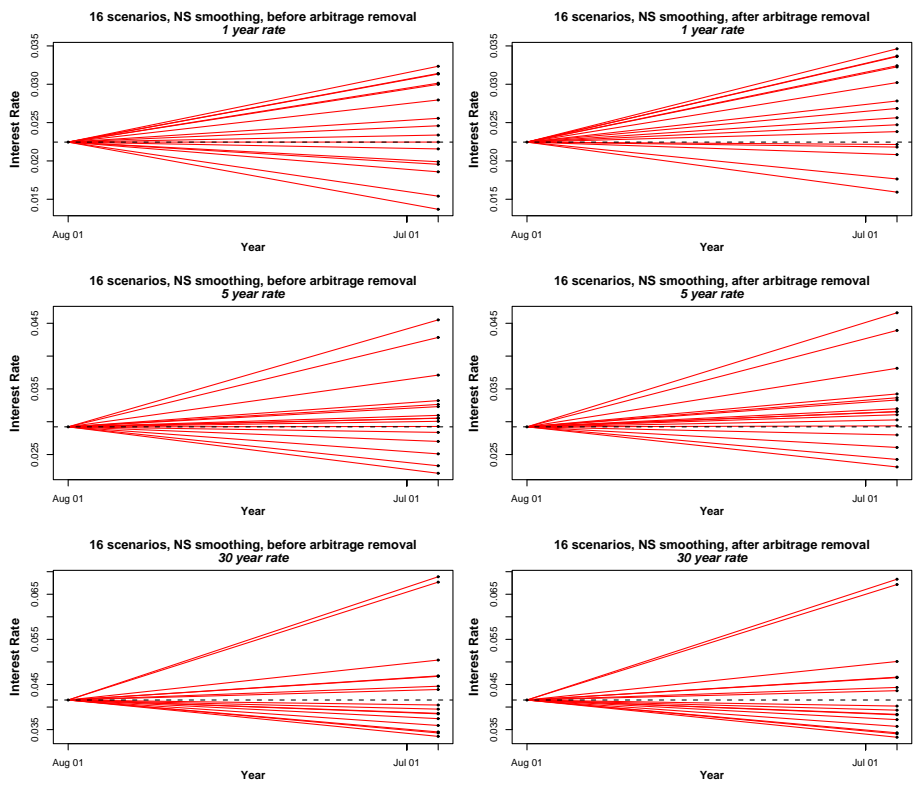


Figure A.7: August 2005.



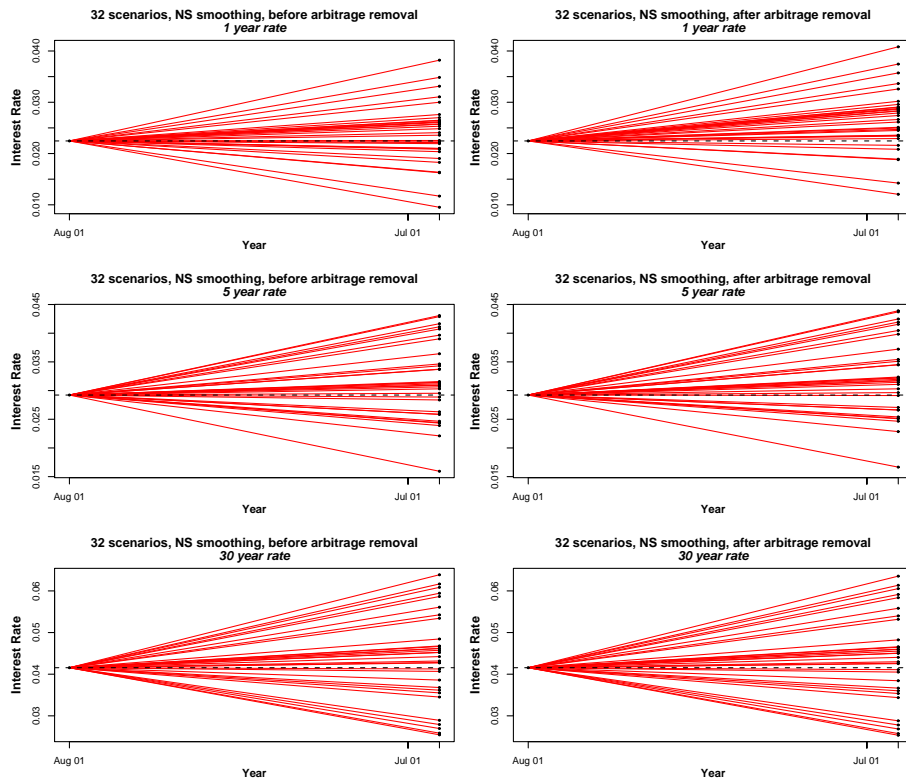


Figure A.8: August 2005.

### A.6.2 Affine smoothing

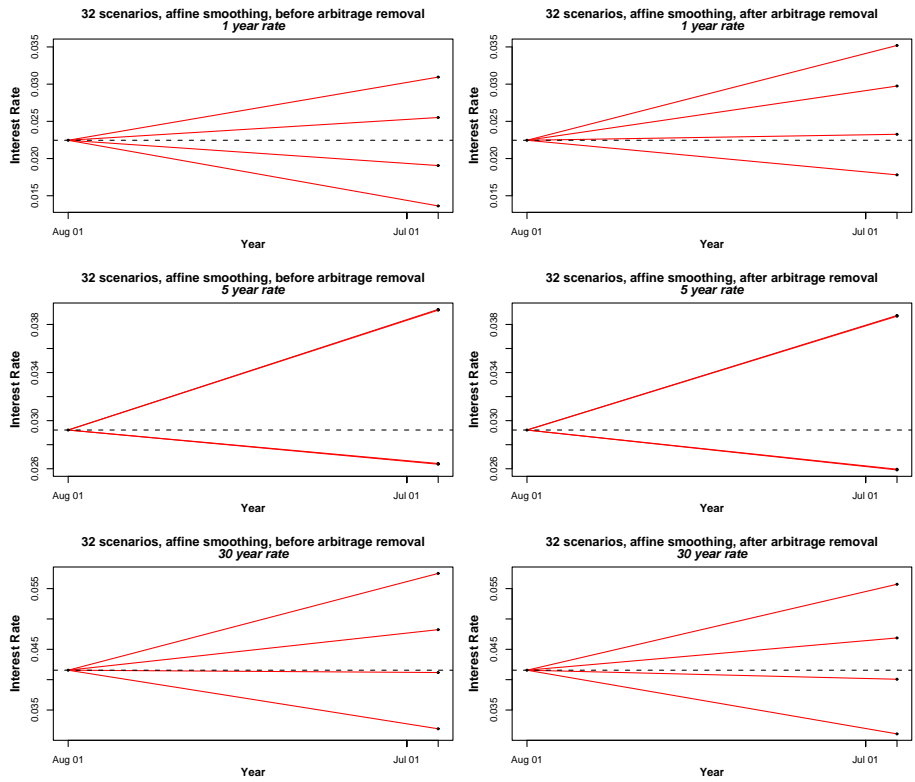


Figure A.9: August 2005.

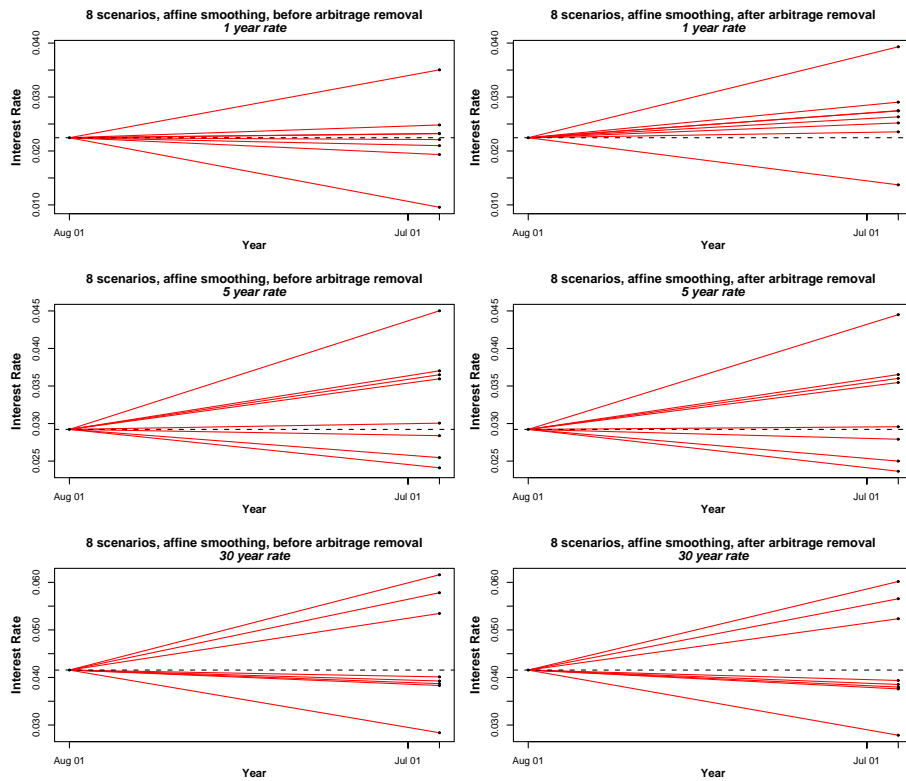


Figure A.10: August 2005.

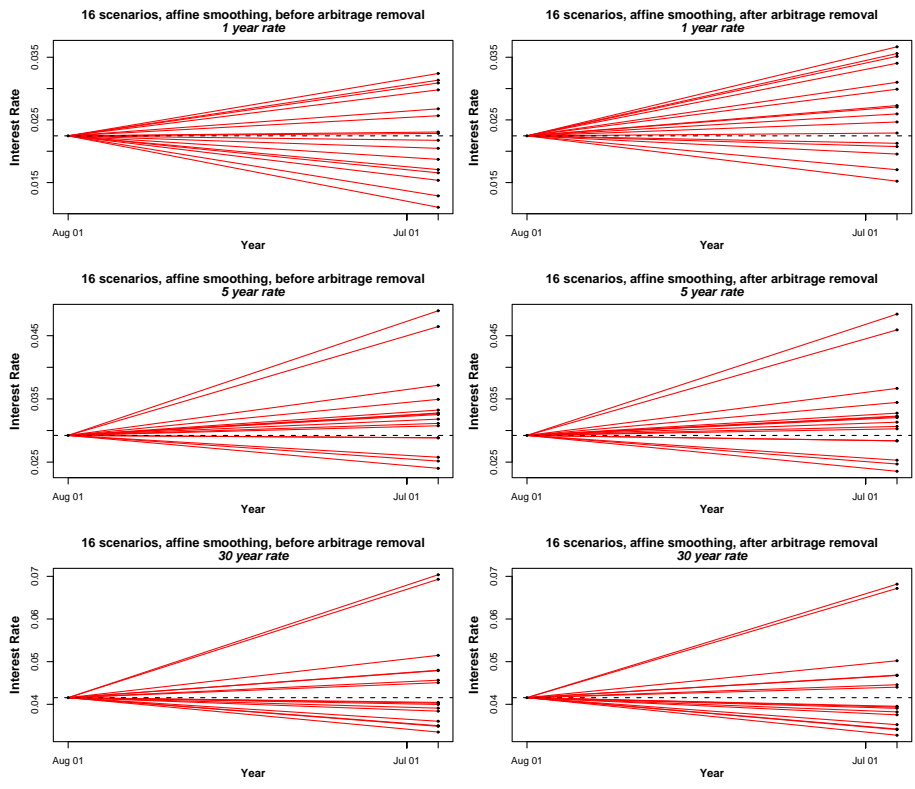


Figure A.11: August 2005.

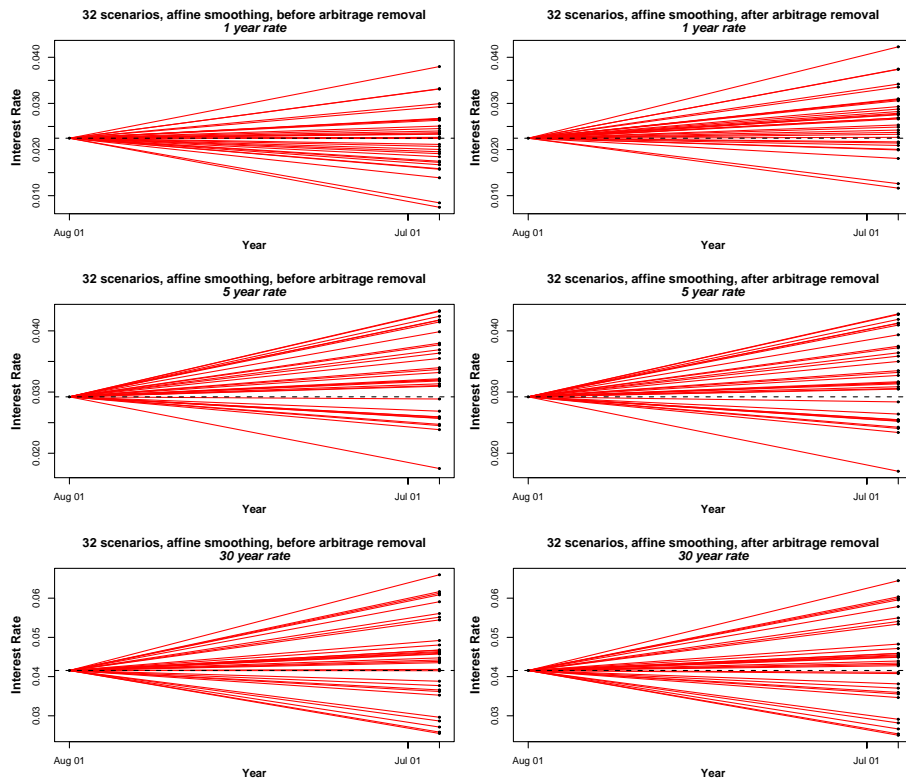


Figure A.12: August 2005.

## A.7 Tree Plots, 1 Period, August 2007

### A.7.1 Nelson-Siegel smoothing

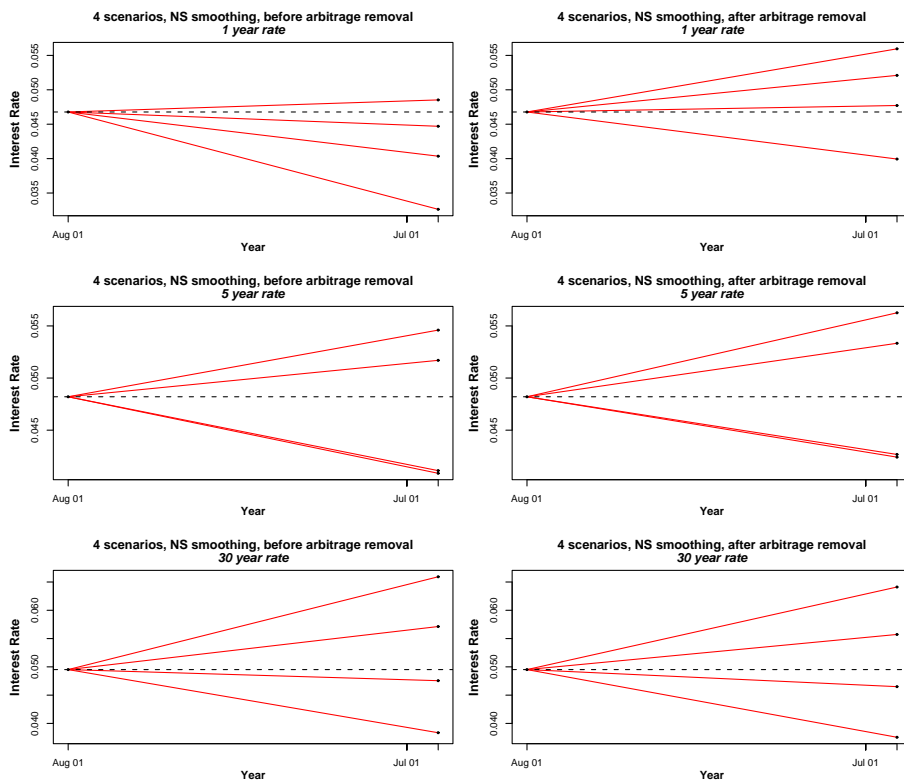


Figure A.13: August 2007.

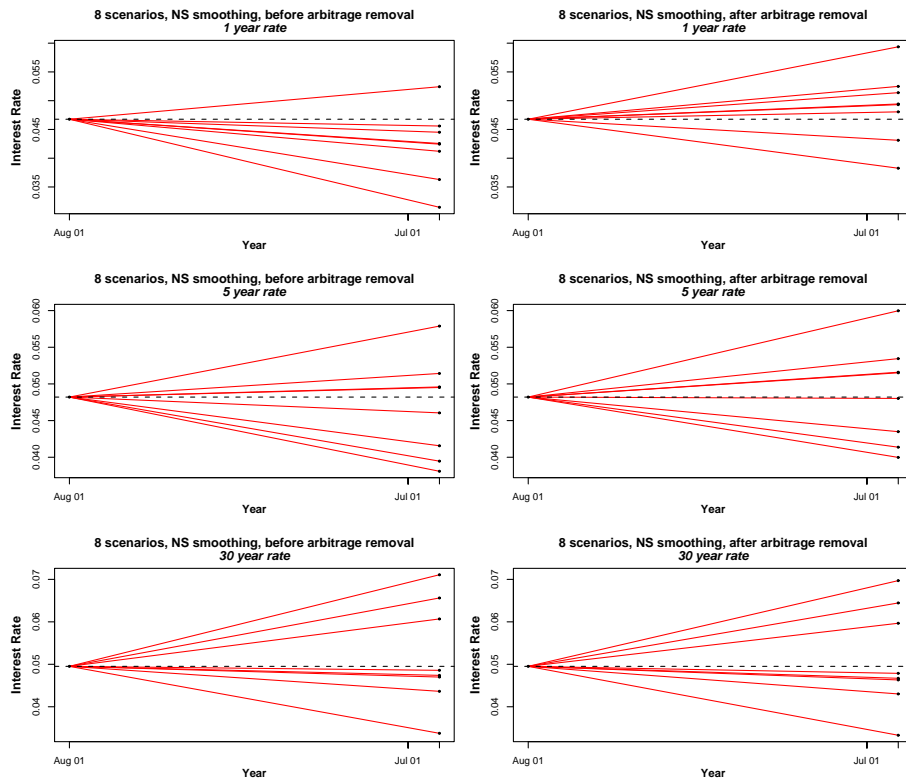


Figure A.14: August 2007.

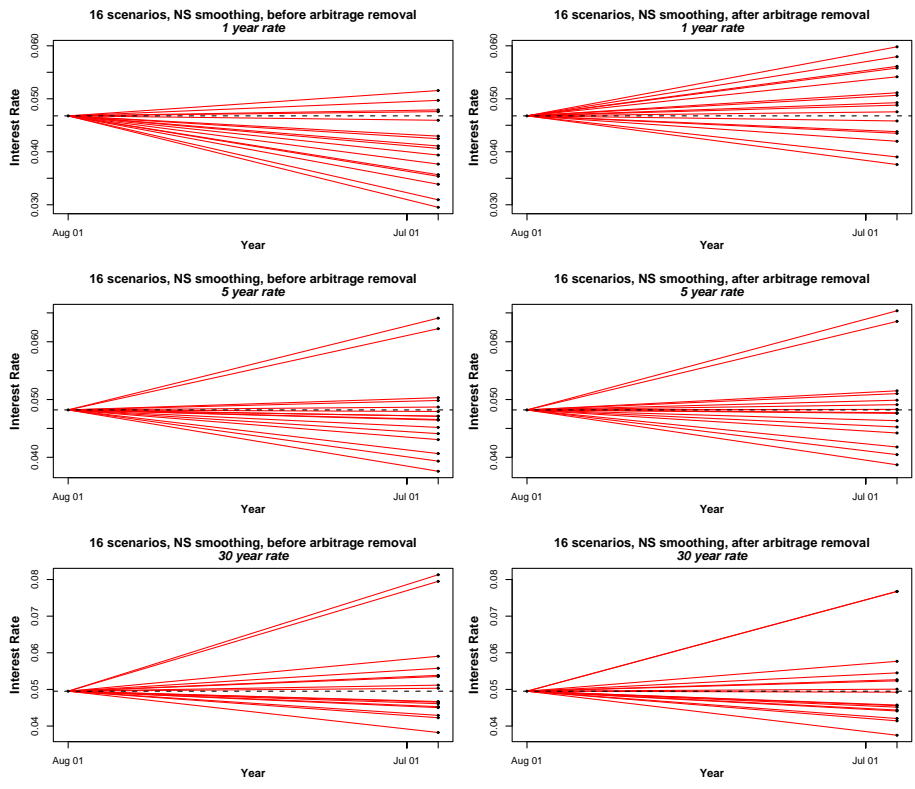


Figure A.15: August 2007.



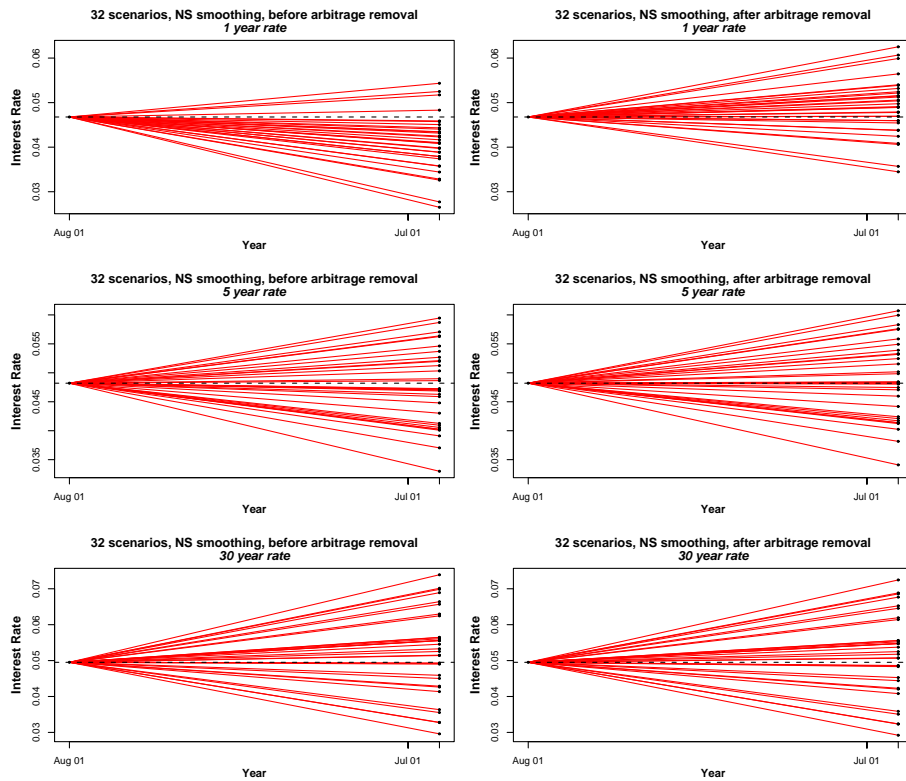


Figure A.16: August 2007.

### A.7.2 Affine smoothing

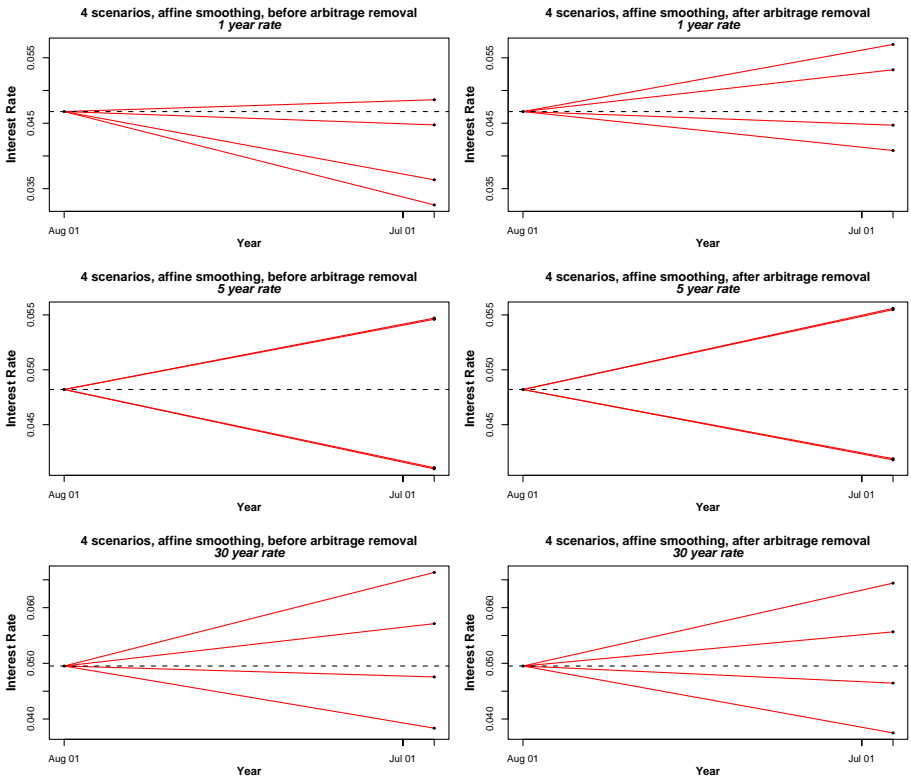


Figure A.17: August 2007.

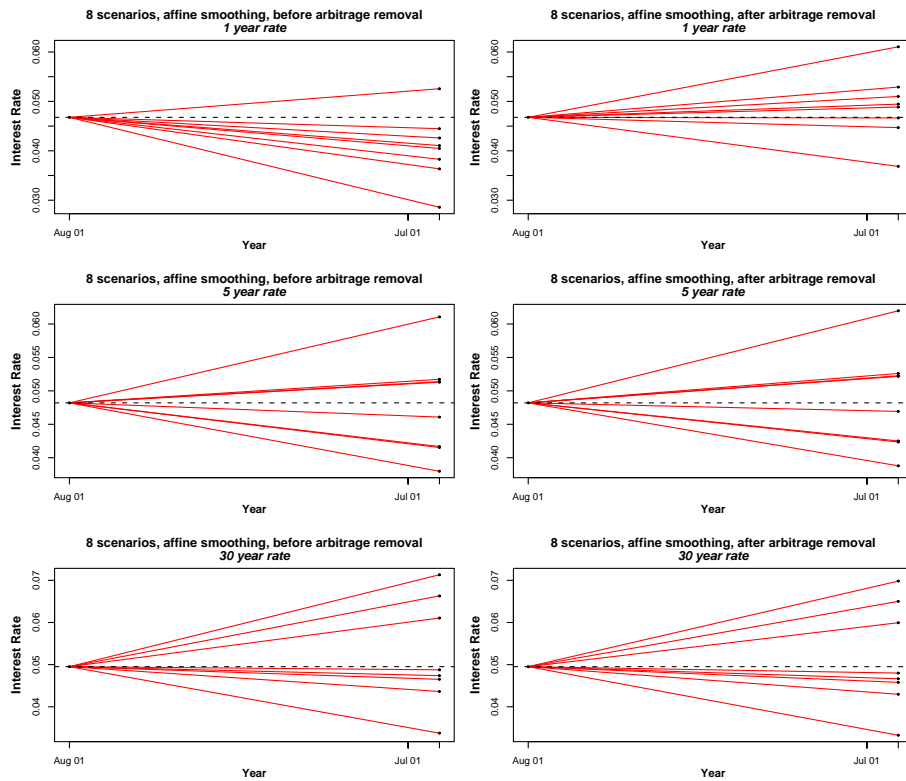


Figure A.18: August 2007.

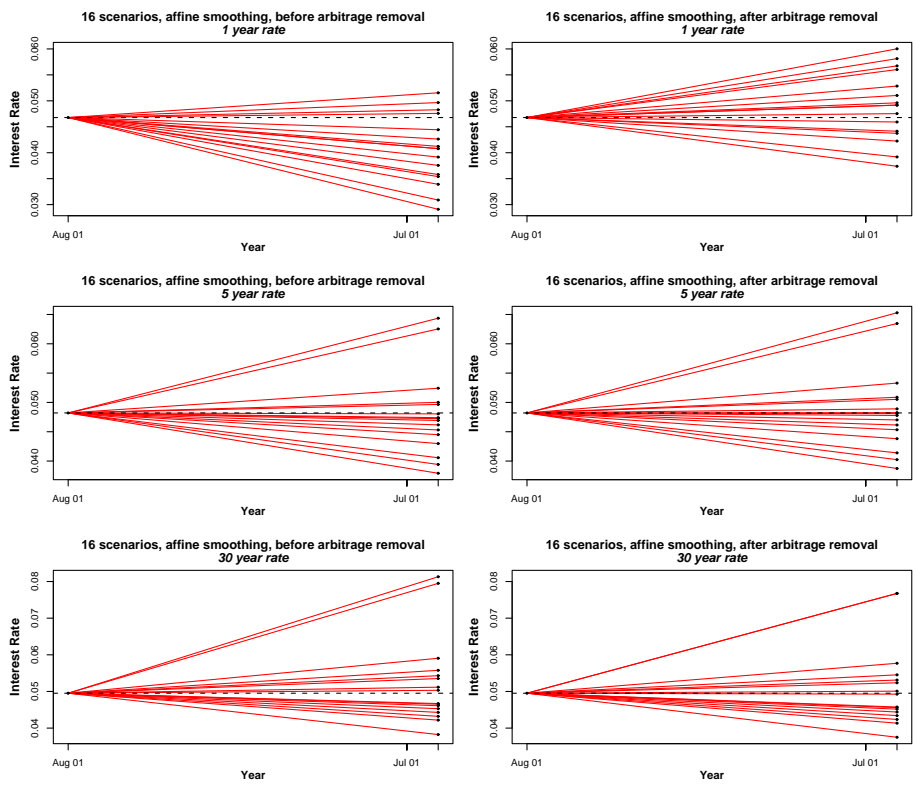


Figure A.19: August 2007.

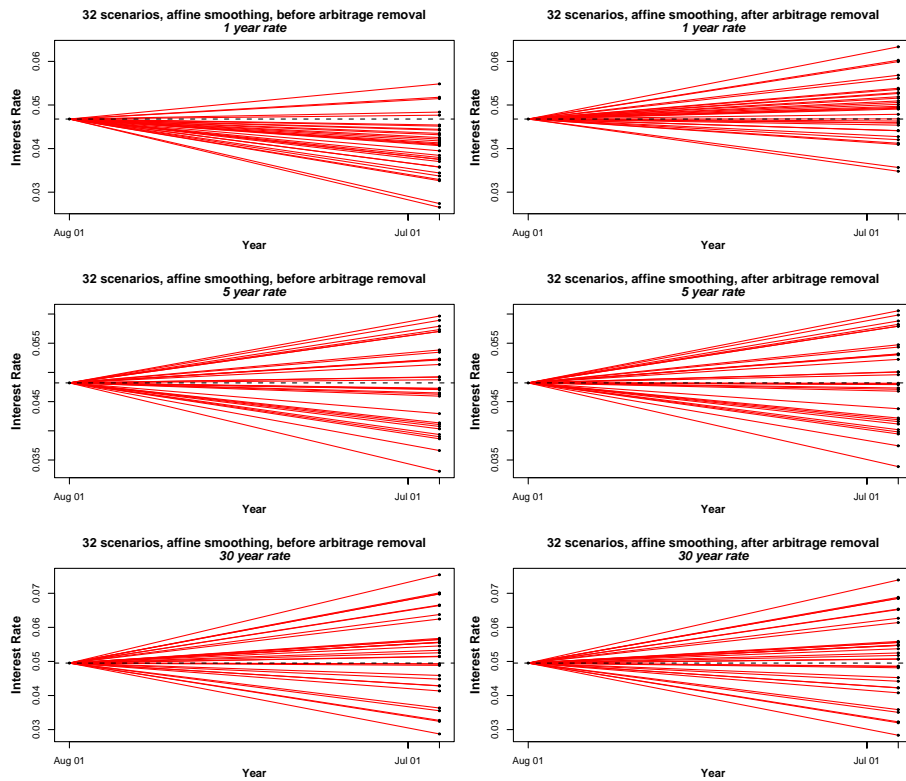


Figure A.20: August 2007.



## APPENDIX B

# Code

---

The maturity of the calculations in this thesis was done with the R language, R Development Core Team (2006).

### B.1 Data read In

Listing B.1: Script used to read in the data from data files in number of other scripts.

```
#####  
## Arnggrimur Einarsson, November 2007.  
## read.r  
## Script to read in and manipulate rate data.  
#####  
  
## 'rate.*' is the data containing only the rates.  
rate.06.01 <- read.table("rate06_01.csv", header = F)  
rate.01.97 <- read.table("rate01_97.csv", header = F)  
rate.97.94 <- read.table("rate97_94.csv", header = F)  
  
## Combine the rates in one large data frame.  
rate <- cbind(rate.06.01, rate.01.97, rate.97.94)  
remove(rate.06.01, rate.01.97, rate.97.94)
```

```

## 'dates.csv' scans in the dates.
dates = scan("dates.csv", what=character(), quiet=T)

## D casts the dates in the form yyyy-mm-dd.
D = as.Date(dates, "%m/%d/%Y")

## 'maturity' contains the time to maturity.
maturity <- read.table("loebetid.csv", header = F)

## Plotting the rates.
x11()
plot(D, as.vector(rate[1,]), mode="numeric", type="l",
      main="", ylab="Interst_rates")

## Print notification.
print('data_read_in_:')

```

Listing B.2: Script made in Matlab, used to plot the surface plot in figure 2.2.

```

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Arngrimur Einarsson, November 2007.
% rate_view.m
% A skript to view the Intetsts rates with surface plot.
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%% Importing the data.
% cd ~/Project/Code
int_06_01 = importdata('rentekurver_06_01.csv');
int_01_97 = importdata('rentekurver_01_97.csv');
int_97_94 = importdata('rentekurver_97_94.csv');

%% Extracting the interest rates.
Daint_06_01 = int_06_01.data(:,2:end);
Daint_01_97 = int_01_97.data(:,2:end);
Daint_97_94 = int_97_94.data(:,2:end);
int = [Daint_06_01, Daint_01_97, Daint_97_94];

%% Extracting the dates.
Date_06_01 = int_06_01.textdata(:,2:end);
Date_01_97 = int_01_97.textdata(:,2:end);
Date_97_94 = int_97_94.textdata(:,2:end);
dates = [Date_06_01, Date_01_97, Date_97_94]

%% Some calculation for plotting.
[m,n] = size(int);
dd = dates(1:77:n);
mi = min(min(int)); %mi=0.0114

```



```

ma = max(max(int)); %ma=0.0987

%% Plotting.
%mesh(int);
%ribbon(int);
surf(int);
xlim([0 n]), ylim([0 m]), zlim([mi ma]);
set(gca,'XTick',[40,92,144,197,249,301,354,405,457,510,562,
614]);
set(gca,'XTickLabel',{'06','05','04','03','02','01','00',
'99','98','97','96','95'});
set(gca,'YTick',(1:20:m))
set(gca,'YTickLabel',{'0','5','10','15','20','25','30'})
title('1995-2006')
xlabel('Dates'), ylabel('Maturity'), zlabel('Rates')
view([0 0 1])

%% Other calculation.
%Reduced matrix to find the eigenvectors and eigenvalues.
int_test = int_06_01.data(1:20:end,1:end);
int_test=int_test(2:end, 2:50:end);
[V,D] = eig(int_test);
C = cov(int_test);
RHO = corr(int_test);

```

Listing B.3: Script to plot the yield curves shown in chapter 2.

```

#####
## Arngrimur Einarrson.
## yield_plot.r
## Script to plot interest rate yields.
#####

## Read in the data.
source('read.r')

## Plotting yield curve for selectet samples.
x <- seq(0,30, length.out=length(rate[,1]))
opar <- par(cex=1.2,mar=c(4,4,1,1)+0.1)
#postscript("yieldCurve.ps")
## rate[,354]: "1999-12-29".
plot(x,(rate[,354])*100, type="l", col=2,
      xlab="Maturity_(years)", ylab="Rate_(%)")
## rate[,290]: "2001-03-21".
lines(x,(rate[,290])*100, col=4)
legend("bottomright", c("29_Des._1999","21_Mar._2001"),
      lty=1, col=c(2,4))
par(opar)

```

```

#dev.off()

## Plotting short vs. long term yield.
prate <- 100*rate
opar <- par(cex=1.2,mar=c(4,4,1,1)+0.1)
#postscript("long_short_yields.ps")
plot(D, prate[5,], type="l", ylim=range(prate), xlab="Date",
      ylab="Yield_(%)")
lines(D, prate[60,], col="blue")
lines(D, prate[121,], col="red")
grid(nx=10, ny=10)
legend("topright", c("1_yea
r", "15_years", "30_years"),
      lty=1, col=c("black", "blue", "red"), bg="white")
par(opar)
#dev.off()

```

## B.2 Principal Component Analysis

The R script `pca_script.r` in listing B.5 reads in the data, manipulates it and implements two functions `pca.basic` and `pca.pcomp` shown in listing B.4. The former function computes the eigenvalues and eigenvectors for a given data set and returns those values. The latter function is a call to the built in R method `prcomp`, for performing Principal component analysis. The two methods return numerically identical results apart from the fact that the eigenvectors can have opposite signs. But as stated in the documentation for `prcomp`, the sign of the factor loadings are arbitrary when calculating with it, that is it is possible that a factor causes a decrease in yield for bonds with short maturity and increase for those with a long maturity. This does not matter when results are used in factor analysis since we only care about the interaction between the factors. If it is of interest it would be possible to find the correct signs by finding the eigenvectors of the covariance matrix for the bond yield data.

Since the `prcomp` uses a *singular value decomposition* of the data matrix, which is the preferred method for numerical accuracy we use that method for the computations.

Listing B.4: The functions used for the principal component analysis.

```

#####
## Arngrimur Einarsson, November 2007.
## pca_func.r
## This script contains the functions pca.bas and pca.pcomp
## for computations of principal components of data sets.
#####

```

```

pca.basic <- function(rate, rate.name, plot, printplot){
  ## The pca.bas takes in data matrix (of interest rates
  ## and returns it's covariance matrix, eigenvalues and
  ## eigenvectors.
  t.rate <- t(rate)
  ## The cov() method calculates the covariance between columns,
  ## not the rows and therefore one needs to transform the data
  ## matrix.
  pca.cov <- cov(t.rate, y = NULL, use = "all.obs", method =
                 c("pearson"))#, "kendall", "spearman")

  ## The eigenvalues and eigenvectors calculated
  pca.bas <- eigen(pca.cov, only.values = F)

  a = 1.2
  b = 1.25
  if (plot==T){
    tmp <- par(cex=a, mar=c(4,4,3,2)+0.1)
    plot(pca.bas$vectors[,1], , ylim=c(-0.6,0.6), type="l",
         pch=1, col=1, main=rate.name, xlab="Maturity_(years)",
         ylab="Factor_loadings")
    #mtext("maturity for whole years", side = 3, cex=b)
    lines(pca.bas$vectors[,2], , type="l", pch=2, col=2)
    lines(pca.bas$vectors[,3], , type="l", pch=3, col=4)
    legend("topright", c("Factor_1", "Factor_2", "Factor_3"),
           lty=1, col=c(1,2,4))
    par(tmp)
  }

  if (printplot==T){
    postscript(paste(rate.name, ".ps", sep=""), width=8, height=8)
    tmp <- par(cex=a, mar=c(4,4,3,2)+0.1)
    plot(pca.bas$vectors[,1], , ylim=c(-0.6,0.6),
         type="l", pch=1, col=1, main=rate.name,
         xlab="Maturity_(years)", ylab="Factor_loadings")
    #mtext("maturity for whole years", side = 3)
    lines(pca.bas$vectors[,2], , type="l", pch=2, col=2)
    lines(pca.bas$vectors[,3], , type="l", pch=3, col=4)
    legend("topright", c("Factor_1", "Factor_2", "Factor_3"),
           lty=1, co=c(1,2,4))
    dev.off()
  }
  #return(pca.bas)
}

# <rate> is the data set, plot takes the arguments <T> for
# plotting and <TP> for prining the plot to a file.

```

```

pca.prcomp <- function(rate , rate.name, plot , printplot){
## The principal components calculated using pcomp() method
# in R.

# Transposed because the method calculates the covariance
# between the columns
t.rate <- t(rate)
pca.prc <- pcomp(t.rate , retx=T, centr=T, scale.=T, tol=0)

a = 1.2
b = 1.25
if (plot==T){
tmp <- par(cex=a, mar=c(4,4,3,2)+0.1)
plot(pca.prc$rotation[,1], , ylim=c(-0.6,0.6), type="l",
      pch=1, col=1, main=rate.name, xlab="Maturity_(years)",
      ylab="Factor_loadings")
#mtext("maturity for whole years", side = 3, cex=b)
lines(pca.prc$rotation[,2], , type="l", pch=2, col=2)
lines(pca.prc$rotation[,3], , type="l", pch=3, col=4)
legend("topright", c("Factor_1", "Factor_2", "Factor_3"),
      lty=1, co=c(1,2,4))
par(tmp)
}

if (printplot==T){
postscript(paste(rate.name, ".ps", sep=""), width=8, height=8)
tmp <- par(cex=a, mar=c(4,4,3,2)+0.1)
plot(pca.prc$rotation[,1], , ylim=c(-0.6,0.6), type="l",
      pch=1, col=1, main=rate.name, xlab="Maturity_(years)",
      ylab="Factor_loadings")
#mtext("maturity for whole years", side = 3, cex=b)
lines(pca.prc$rotation[,2], , type="l", pch=2, col=2)
lines(pca.prc$rotation[,3], , type="l", pch=3, col=4)
legend("topright", c("Factor_1", "Factor_2", "Factor_3"),
      lty=1, co=c(1,2,4))
dev.off()
}
#return(pca.prc)
}

```

Listing B.5: The script used to run the principal component analysis.

```

#####
## Arngrimur Einarsson , November 2007.
## pca_script.r
## Script to evaluate PCA with the functions in pca_func.r.
#####

```

```
## The functions used.
source('pca_func.r')
## The data read in.
source('read.r')

##-----
## Construction of test sets.
##-----
seq.1 <- seq(5,121,4) ## Every whole number of the maturity.
#seq.5 <- seq(5,121,20) ## Every fifth number of maturity.
## The rate with the whole maturity dates.
rate <- rate[seq.1,]

##-----
## Subsets of the maturity dates made.
##-----
## Single years.
test.set <- list(
rate.95_2006 = rate,
rate.06 = rate[, (1:40)],
rate.05 = rate[, (41:92)],
rate.04 = rate[, (93:144)],
rate.03 = rate[, (145:197)],
rate.02 = rate[, (198:249)],
rate.01 = rate[, (250:301)],
rate.00 = rate[, (302:353)],
rate.99 = rate[, (354:405)],
rate.98 = rate[, (406:457)],
rate.97 = rate[, (458:510)],
rate.96 = rate[, (511:567)],
rate.95 = rate[, (568:614)]
)

## Before and after euro (1. jan 1999 Denmark joins ERM II,
## but the DKK has pegged the D-mark since 1987.)
test.euro <- list(
pre.euro.rate = rate[, (354:614)],
post.euro.rate = rate[, (1:353)]
)

## Four year periods.
test.4.years <- list(
rate.95.98 = rate[, (406:614)],
rate.99.02 = rate[, (198:405)],
rate.03.06 = rate[, (1:197)]
)
```

```

## Adding year at a time.
test.add.year <- list(
rate.95 = rate[, (563:614)],
rate.95.96 = rate[, (511:614)],
rate.95.97 = rate[, (458:614)],
rate.95.98 = rate[, (406:614)],
rate.95.98 = rate[, (354:614)],
rate.95.99 = rate[, (302:614)],
rate.95.00 = rate[, (250:614)],
rate.95.01 = rate[, (198:614)],
rate.95.02 = rate[, (145:614)],
rate.95.03 = rate[, (93:614)],
rate.95.04 = rate[, (41:614)],
rate.95.05 = rate[, (1:614)]
)

##


---


## The principal components calculated and plottet.


---


##
## Example of function calls.
pca.basic.95.06 <- pca.basic(rate, "1995-2006", _basic", T, F)
pca.prcomp <- pca.prcomp(rate, "1995-2006", T, F)

# For loop to calculate and plot the data.
tmp <- length(names(test.add.year))
tmp2 = numeric(tmp)
for (i in (1:tmp)){
tmp2[i] = paste("pca.prc.", names(test.add.year[i]), "=",
               "pca.prcomp(test.add.year[[" , i, " ]]", ", ",
               "names(test.add.year[[" , i, " ]]", F, T)", sep="")
eval(parse(text=tmp2[i]))
}

# summary(pca.prc) ## Prints out numbers.

```

## B.3 Normality Inspection

Listing B.6: The following script is used to read in data, plot histograms,  $Q-Q$  plots and perform the Jarque-Bera and Shapiro-Wilk tests on data and subsets of data. The kernel estimation is made made with the rather “conservative” bandwidth of 0.8.

```

#####
## Arngrimur Einarsson, November 2007.
## rate_hist.r

```

```

## Script to analyze the normality of interest rates.
#####

## library path and extra library added to get the
## Jarque-Bera test.
.libPaths(c("~/Project/Code/Ae_R_lib", .Library))
library('tseries')

## Function used to plot the rates and smoothed line.
hist.rate <- function(rate, bw, name){
  ## Function used to plot histogram of the rates and
  ## smoothed line, takes in vector of rates (rate),
  ## parameter for bandwidth (bw) and name of the
  ## histogram (name).
  hist(rate, freq=F, main=NULL, xlab=NULL, ylab=NULL)
  title(main=name, xlab="ln_of_Rate_("%)", ylab="Proportion",
        cex.main=1.5, cex.lab=1.5)
  lines(density(as.matrix(rate), bw, kernel="gaussian"),
        col="red")
}

## The data read in and combined in a single matrix.
data95to97<-matrix(scan( 'NYK.95 to97.dat ' ),ncol=106,byrow=T)
data97to01<-matrix(scan( 'NYK.97 to01.dat ' ),ncol=254,byrow=T)
data01to06<-matrix(scan( 'NYK.01 to06.dat ' ),ncol=255,byrow=T)
data<-cbind(data01to06 , data97to01 , data95to97)

## The first column (maturity index) and first row (issued
## dates) taken out of the matrix.
data <- data[, -1];date <- data[1,];data <- data[-1,]

## A histogram and a QQ-plot made from a normal sample.
x <- rnorm(dim(data)[2])
##postscript("samp_norm.ps ")
opar <- par(mfrow=c(1,2),mex=0.8,mar=c(4,4,2,1)+0.1)
hist(x, freq=F); lines(density(x, bw=0.8,
        kernel="gaussian"), col="red")
qqnorm(x); qqline(x, col="red")
par(opar)
##dev.off()

par(cex.lab=2,mar=c(4,4,2,1)+0.1)
plot(sin , -pi , 2*pi)

x <- rnorm(dim(data)[2])
jarque.bera.test(x)
shapiro.test(x)

```

```
## Sub-sets constructed.

## Make set containing each year.
#rate.06 <- data[, (1:40)]
#rate.05 <- data[, (41:92)]
#rate.04 <- data[, (93:144)]
#rate.03 <- data[, (145:197)]
#rate.04 <- data[, (198:249)]
#rate.01 <- data[, (250:301)]
#rate.00 <- data[, (302:353)]
#rate.99 <- data[, (354:405)]
#rate.98 <- data[, (406:457)]
#rate.97 <- data[, (458:510)]
#rate.96 <- data[, (511:567)]
#rate.95 <- data[, (568:614)]

## Test set 1 2001–2006.
#data <- data[, 1:301]
## Test set 2 1995–1998
#data <- data[, 404:614]
## Test set 3 1999–2001
#data <- data[, (301:405)]
## Test set 4 2005–2006
#data <- data[, 1:92]

## Test set 5 2004–2006.
#data <- data[, 1:144]
## Test set 6 before millennium: just as skewed.
#data <- data[, 1:354]
## Test set 7 after millennium: nothing better than before.
#data <- data[, 354:614]
## Test set 8 1995–1999: no better
#data <- data[, 407:614]
## Test set 9 1999–2001
#data <- data[, (301:405)]

## The 1, 5, 20 and 30 year rate taken out of the data set.
rate1 <- data[4,]
rate5 <- data[20,]
rate15 <- data[60,]
rate30 <- data[120,]

## ln taken of the rates.
#rate1 <- log(rate1*100)
#rate5 <- log(rate5*100)
```



```
#rate15 <- log(rate15*100)
#rate30 <- log(rate30*100)

## Example of histogram plotting.
#postscript("hist_95_06.ps")
opar <- par(mfrow=c(2,2),mex=0.8,mar=c(4,4,2,1)+0.5)
hist.rate(rate1, 0.8, "1_year_maturity")
hist.rate(rate5, 0.8, "5_year_maturity")
hist.rate(rate15, 0.8, "15_year_maturity")
hist.rate(rate30, 0.8, "30_year_maturity")
par(opar)
#dev.off()

## Example of QQ-plotting.
#postscript("qqplot_95_06.ps")
opar <- par(mfrow=c(2,2),mex=0.8,mar=c(4,5,2,1)+0.1)
qqnorm(rate1, main="1_year_maturity", cex.main=1.5,
        cex.lab=1.5)
qqline(rate1, col="red")
qqnorm(rate5, main="5_year_maturity", cex.main=1.5,
        cex.lab=1.5)
qqline(rate5, col="red")
qqnorm(rate15, main="15_year_maturity", cex.main=1.5,
        cex.lab=1.5)
qqline(rate15, col="red")
qqnorm(rate30, main="30_year_maturity", cex.main=1.5,
        cex.lab=1.5)
qqline(rate30, col="red")
par(opar)
#dev.off()

## Jarque-Bera test for normality performed on the data.
jb1 <- jarque.bera.test(rate1)
jb5 <- jarque.bera.test(rate5)
jb15 <- jarque.bera.test(rate15)
jb30 <- jarque.bera.test(rate30)

## Shapiro-Wilk normality test performed on the data.
sw1 <- shapiro.test(rate1)
sw5 <- shapiro.test(rate5)
sw15 <- shapiro.test(rate15)
sw30 <- shapiro.test(rate30)
```

## B.4 Vector Auto Regression

### B.4.1 Calculation of the lag and stability of the VAR model

The roots and the information criteria are calculated with the function VARS package in R (Pfaff 2007).

Listing B.7: Functions calibration of the VAR model (contains also estimation of parameters).

```
#####
## Arnggrimur Einarsson, November 2007.
## var1_func.r
## Functions *** and *** used for computations of VAR(1)...
#####

var1.basic <- function(factors){
  ## The var1.basic function takes in an R data frame
  ## containig the a constant level, slope, and curvature
  ## factors.
  X <- as.matrix(factors[1:(ncol(rate)-1),])
  Y <- as.matrix(factors[2:(ncol(rate)),])

  tmp <- solve(t(X)%*%X) %*% t(X)
  d_hat <- matrix(0,nrow=3, ncol=4)
  for(i in 1:3){
    d_hat[i,] <- t(tmp%*%Y[, (i+1)])
  }

  alpha <- d_hat[1:3,1] # aka. mu
  beta <- d_hat[1:3,2:4] # aka. A
  res <- Y[,2:4] - X%*%t(d_hat)
  V <- t(res)%*%res/(ncol(rate)-1)
  invV <- solve(V)

  ans <- data.frame('alpha'=alpha, 'beta'=beta,
    'variance'=V)
  return(ans)
}

var1.lm <- function(factors){
  ## The var1.lm uses the R lm function...
  X <- as.matrix(factors[1:(ncol(rate)-1),])
  Y <- as.matrix(factors[2:(ncol(rate)),])
  lm1 <- lm(Y~X)
  return(lm1)
}
```

```

}

var.3.factors <- function(data){
  # Takes in the interest rate data and returns the factors
  # for the varp.3 model.
  level <- t(data[5,])*100
  slope <- t(data[21,]-data[5,])*100
  curv. <- t(2*data[21,]-data[5,]-data[121,])*100 # NB appr.
  factors <- data.frame(level,slope,curv.)
  names(factors)[1:3] <- c('level','slope','curv.')
  return(factors)
}

st.var1 <- function(data){
  ## Function used to estimate stability and information
  ## criteria for VAR(p), takes in the factors
  var1.3.ic <- VARselect(data, lag.max = 5, type = c("const"))
  var1.3 <- VAR(data, p=1, type="const")
  roots <- roots(var1.3)
  skila <- list(var1.3.ic$selection, roots)
  return(skila)
}

```

Listing B.8: Script used for estimating the order and stability (contains also estimation of parameters).

```

#####
## Arngrimur Einarsson, November 2007.
## var1.r
## Script to investigate the properties of the VAR(p) model.
#####

##.libPaths(), Adding VARS package to the libPaths
.libPaths(c("~/Project/Code/Ae_R_lib", .Library))
library("vars") #library("xtable")
## Running the functions used in the script.
source("var1_func.r")
## Data read in.
source("read.r")

## The rates plottet.
x11()
plot(D,as.vector(rate[1,], mode="numeric"), type="l",
      main="Base_rate", ylab="Interst_rates")
x11()
plot(D[1:52],as.vector(rate[1,1:52], mode="numeric"),
      type="l", main="Base_rate", ylab="Interst_rates")

```

```

###
### Factors put in data frame, for methods #1 & #2.
###
level <- t(rate[5,]*100)
slope <- t(rate[21,]-rate[5,])*100
### Naive method.
#curv. <- t(2*rate[21,]-rate[5,]-rate[121,])*100
### Alt. rates
curv. <- t(rate[21,]-(4/29)*rate[121,]-(25/29)*rate[5,])*100
cons. <- rep(1,(ncol(rate))); cons. <- as.data.frame(cons.)

### The factors combined in a data frame.
Z <- data.frame(cons., level, slope, curv.)
names(Z)[2:4] <- c('level', 'slope', 'curv.')
```

---

```

###
### Method #1 VAR1 hard coded.
###
### Using the function var1.basic
var1.1 <- var1.basic(Z)
```

---

```

###
### Method #2 the R lm() function used, returns the same :)
###
var1.2 <- var1.lm(Z)
```

---

```

###
### Method #3 the VARS package used.
###
### The function factors run to "extract the factors from the
### data set"
factors <- var.3.factors(rate)
```

---

```

### Loop to estimate the stability and compute the
### information criteria
num <- 604
ic <- vector("list", num)
root <- matrix(nrow=num, ncol=3)
for (r in 1:num) {
  st.param <- factors[1:(r+10),]
  tmp <- st.var1(st.param)
  ic[r] <- tmp[1]
  root[r,] <- as.matrix(t(tmp[[2]]))
}
```

---

```

### Plotting of the roots to see the moduli
```

```

#postscript("roots.ps")
opar <- par(cex=1.2,mar=c(4,4,3,2)+0.1)
plot(root[,1],type="p", pch=20, col=1, main="Roots",
      ylab="moduli_of_the_eigenvalues", xlab="points")
mtext("Roots", side = 3,cex=b)
points(root[,2],pch=20, col=3)
points(root[,3],pch=20, col=4)
l <- (rep(1,num))
lines(l, type="l", col=2)
legend("bottomright", c("eig._1", "eig._2", "eig._3"), pch=20,
      col=c(1,3,4))
par(opar)
#dev.off()

## Manipulating the IC information
rout <- file("ic.txt","w")
dump(ic, file = rout)
close(rout)
dump("ic", file = "ic.dump.txt")
## Edit the file with search/replace and shell programs
## including 'cut -b 4-29 bakk.txt > bakk.txt'.
ic.table <- read.table("ic.clean.txt", header = T)
## Plotting of the IC to see the sugj. order, not using the
## FPE criteria 'cose I don't know what it is (col=4).
#postscript("ic.ps")
opar <- par(cex=1.2,mar=c(4,4,3,2)+0.1)
plot(ic.table[,1], type="s", pch=20, col=1, main="",
      ylab="Suggested_order_(p)",
      xlab="Number_of_parameters_(issuing_dates)")
lines(ic.table[,2], type="s", col=2)
lines(ic.table[,3], type="s", col=3)
#lines(ic.table[,4], type="l", col=4)
legend("topright",c("AIC", "HQ", "BC"),
      lty=1, col=c(1,2,3))
par(opar)
#dev.off()

## Construction of a "solid" model!
var.3.fact <- factors[1:200,]
var.3.mod <- VAR(var.3.fact, p=1, type="const")
roots(var.3.mod)
#postscript(horizontal = FALSE, onefile = FALSE)
plot(var.3.mod)
#dev.off()

## Forecasting.
var.3.f <- predict(var.3.mod, nahead = 52, ci = 0.95)

```

```
names(var.3.f)
plot(var.3.f)
```

Listing B.9: The script used for estimating the parameters in the VAR(1) model.

```
#####
## Arngrimur Einarsson, November 2007.
## VARestimate3f.r
## Construction of VAR(1) model & scenario generation for 3
## factors.
#####
rm(list=ls())
## Adding path to additional R packages {.libPaths()}.
.libPaths(c("~/Project/Code/Ae_R_lib", .Library))
library("mvtnorm")

## The interest rate data read in.
data95to97<-matrix(scan('NYK.95to97.dat'),ncol=106,byrow=T)
data97to01<-matrix(scan('NYK.97to01.dat'),ncol=254,byrow=T)
data01to06<-matrix(scan('NYK.01to06.dat'),ncol=255,byrow=T)
## This data is reversed compared to the others.
data06to07<-matrix(scan('NYK.06to07.dat'),ncol=45,byrow=T)
data06to07 <- data06to07[,45:1]
data<-cbind(data06to07,data01to06,data97to01,data95to97)

NoDates<-ncol(data)-1
NoMats<-nrow(data)-1
Dates<-data[1,(NoDates+1):2]
Mats<-data[2:(NoMats+1),1]
Rates<-data[2:(NoMats+1),(NoDates+1):2]
Rates <- Rates[,2:491] # nr. 1 is August 8.
NoDates <- dim(Rates)[2]
D<-M<-Y<-DatesAsReal<-rep(0,NoDates)
#Rates <- log(Rates*100)

## Estimation of the VAR(1) model:  $(l,s,c)_{t+1} = a +$ 
##  $B*(l,s,c)_{t+eps_{t+1}}$ 
const<-rep(1,NoDates)

## Extract the selected factors from the interest rates,
## the simpeler method.
#level<-Rates[5,]
#slope<-Rates[21,]-Rates[5,]
#curvature<-2*Rates[21,]-Rates[5,]-Rates[121,]

## Extract the selected factors from the interest rates,
## Christiansen/Lund.
```

```

level<-Rates[5,]
slope<-Rates[121,]-Rates[5,]
curvature<-Rates[21,]-(25/29)*Rates[5,]-(4/29)*Rates[121,]

## Histogram of the factors made.
#hist(level, prob=T)
#lines(density(level, bw=2), col="red")
#x11();hist(slope, prob=T)
#lines(density(slope, bw=0.5), col="red")
#x11();hist(curvature, prob=T)
#lines(density(curvature, bw=0.5), col="red")

## The factors multiplied with 100 to work with percentage.
level<-level*100
slope<-slope*100
curvature<-curvature*100

## The factors combined in a single data frame.
Z<-cbind(const, level, slope, curvature)
X<-Z[1:(NoDates-1),]
Y<-Z[2:NoDates,]

## OLS estimation -----
Dummy <- solve(t(X)%*%X) %*% t(X)
est <- matrix(0, nrow=3, ncol=4)
for (i in 1:3) est[i,]<-t(Dummy%*%Y[, (i+1)]) # delta hat

a <- est[1:3,1]
B <- est[1:3,2:4]
longterm <- solve(diag(rep(1,3))-B,a) # mu

res <- Y[,2:4]-X%*%t(est) # residuals
V <- t(res)%*%res/(NoDates-1) # Omega hattur
invV <- solve(V)
## OLS ends -----

## The built in function lm() to fit linear models i.e the
## model in equation in chapter X. It confirms the
## calculations of the coefficients (a and B) and the
## residuals.

#lm1 <- lm(Y~X)

## Restricted model OLS estimation, off-diagonal elements
## equal to 0.
estR<-matrix(0, nrow=3, ncol=4)
resR<-Y[,2:4]

```

```

for (i in 1:3) {
  Dummy1<-solve(t(X[,c(1,(i+1))]) %*%X[,c(1,(i+1))]) %*%t
  (X[,c(1,(i+1))])

  Dummy2<-t(Dummy1%*%Y[, (i+1)])
  estR[i,1]<-Dummy2[1]
  estR[i,(i+1)]<-Dummy2[2]
  resR[i,i]<-Y[, (i+1)]-X[,c(1,(i+1))]%*%t(Dummy2)
}

VR<-t(resR)%*%resR/(NoDates-1)

aR<-estR[1:3,1]
BR<-estR[1:3,2:4]

longtermR<-solve(diag(rep(1,3))-BR,aR)
#yieldlongtermR<-solve(Trans,longtermR)

invVR<-solve(VR)
## restricted model ends _____

## Some code for Scenario generation _____
## initializing
Bprod <- diag(rep(1,3))
Bsum <- diag(rep(0,3))
Vsum <- V
initY <- Y[613,2:4]
#initStart <- longterm
mean <- matrix(0, ncol=1, nrow=3)
NoScen <- k <- 100 # Number of linear scenarios.
NoYears <- 5 # Number of years in time step.
NoWeeks <- 52 # Number of weeks in time step.

for (i in 1:NoWeeks){
  Bprod <- Bprod%*%B
  Bsum <- Bsum+Bprod

  if(i<NoWeeks) {Vsum <- Vsum + Bprod %*% V %*% t(Bprod)}
}

## First random numbers generated for the scenarios.
init_simpath <- rmvnorm(NoScen, mean, Vsum)
## Y_t = \mu + \beta Bprod*(Y_{t-1} - \mu)
Minit <- longterm+Bprod%*%(initY-longterm)
# Matrix with column for each scenario, containing the
# factors.

```



```

xV <- matrix(0, ncol=NoScen, nrow=3)
for (i in 1:k){
  xV[,i] <- Minit+init_simp[ i , ]
}

## generating the scenarios
scenarios <- vector("list", NoScen)
#scenarios2 <- vector("list", NoScen)
for (s in 1:NoScen){
  xV.sim <- matrix(0, ncol=NoYears+1, nrow=3)
  xV.sim[,1] <- longterm
  xV.sim[,2] <- xV[,s] # xV1[, (s+1)]?
  #xV2 <- matrix(0, ncol=NoYears+1, nrow=3)
  #xV2[,1] <- longterm
  #xV2[,2] <- Minit
  for(j in 2:NoYears){      #j is number of years
    sCounter <- 1
    #Bprod <- diag(rep(1,3)) #initialize
    #Bsum <- diag(rep(0,3)) #initialize
    #Vsum <- V              #initialize
    #for (i in 1:NoWeeks){
      #Bprod <- Bprod%*%B
      #Bsum <- Bsum+Bprod
      #if(i<NoWeeks) {Vsum <- Vsum + Bprod %*% V %*% t(Bprod)}
    #}
    if (j == 2){
      ## picks scenario Counter as a starting point
      start <- xV[,sCounter];
    }
    else
      {start <- tmp}
    simpath <- t(rmvnorm(1, mean, Vsum))
    M <- longterm + Bprod %*% (start-longterm)
    tmp <- M+simpath
    xV.sim[, (j+1)] <- tmp
    #xV2[, (j+1)] <- M
  }
  sCounter <- sCounter + 1
  scenarios [[s]] <- xV.sim
  #scenarios2 [[s]] <- xV2
}
print("done_:")
## scenario generation ends ## -----

## The 1 year rate calculatet, eps1 is average.
ratel <- matrix(0,ncol=NoYears+1, nrow=k)
for (i in 1:k){

```

```

    rate1[i,] <- scenarios [[i]][1,]
  }
  #eps1 <- scenarios2 [[k]][1,]
  ## The 30 year rate calculatet, eps30 is average.
  rate30 <- matrix(0,ncol=NoYears+1, nrow=k)
  for (i in 1:k){
    rate30[i,] <- scenarios [[i]][1,] + scenarios [[i]][2,]
  }
  #eps30 <- rep(0, NoYears+1)
  #eps30 <- scenarios2 [[k]][2,]+scenarios2 [[k]][1,]
  ## The 5 year rate calculatet, eps5 is average.
  rate5 <- matrix(0,ncol=NoYears+1, nrow=k)
  for(i in 1:k){
    rate5[i,] <- scenarios [[i]][3,] + (4/29) * rate30[i,] +
      (25/29) * scenarios [[i]][1,]
  }
  #eps5 <- rep(0, NoYears+1)
  #eps5 <-
  #scenarios2 [[k]][3,]+(4/29)*eps30+(25/29)*scenarios2 [[k]][1,]

plot.scenarios <- function(rates, name){
  plot(rates[k,], type="l", lty="dotted", col="red",
       ylim=c(0,10), xlim=c(1, NoYears+1), axes=FALSE,
       main=name, xlab="Years", ylab="rates_("%)", cex.main=1.5,
       cex.lab=1.5)
  axis(1, 1:6, c(0,1,2,3,4,5)); axis(2); box()
  for (i in 2:k){
    lines(rates[i,], lty="dotted", col="red")
  }
  #lines(eps)
}

#postscript("rScenarios.ps")
opar <- par(mfrow=c(3,1),mex=0.8,mar=c(4,4,2,1)+0.5)
plot.scenarios(rate1, "1_year_maturity")
plot.scenarios(rate5, "5_year_maturity")
plot.scenarios(rate30, "30_year_maturity")
par(opar)
#dev.off()

## Histogram of level
#postscript("hist_level.ps")
opar <- par(mfrow=c(3,2),mex=0.8,mar=c(4,4,2,1)+0.5)
for(j in 2:(NoYears+1)){
  h <- rep(0,k)
  for (i in 1:k){

```

```

    h[i] <- scenarios [[i]][1, j]
  }
  hist(h, br=22, prob=T, ylim=c(0,1),
       main=paste(j-1,"years"), xlab = "", cex.main=1.5,
       cex.lab=1.5)

  lines(density(h, bw=0.8), col="red")
}
par(opar)
#dev.off()
## Histogram of slope
#postscript("hist_slope.ps")
opar <- par(mfrow=c(3,2), mex=0.8, mar=c(4,4,2,1)+0.5)
for(j in 2:(NoYears+1)){
  h <- rep(0,k)
  for(i in 1:k){
    h[i] <- scenarios [[i]][2, j]
  }
  hist(h, br=22, prob=T, ylim=c(0,1), main =
       paste(j-1,"years"), xlab = "", cex.main=1.5,
       cex.lab=1.5)
  lines(density(h, bw=0.8), col="red")
}
par(opar)
#dev.off()
## Histogram of curvature
#postscript("hist_curv.ps")
opar <- par(mfrow=c(3,2), mex=0.8, mar=c(4,4,2,1)+0.5)
for(j in 2:(NoYears+1)){
  h <- rep(0,k)
  for(i in 1:k){
    h[i] <- scenarios [[i]][3, j]
  }
  hist(h, br = 22, prob=T, ylim=c(0,1),
       main = paste(j-1,"years"), xlab = "", cex.main=1.5,
       cex.lab=1.5)
  lines(density(h, bw=0.8), col="red")
}
par(opar)
#dev.off()

## ## Write rates to file
# rout <- file("rates.txt", "w")
## for(s in 2:2){
##   for(t in 1:10){
##     capture.output(rate1[s,t], file = rout, append = T)
##     capture.output(rate5[s,t], file = rout, append = T)

```

```

##      capture.output(rate30[s,t], file = rout, append = T)
##    }
## }
## close(rout)

```

## B.5 Simple Arbitrage test

Listing B.10: Calculation of the prices for the bonds used in listing B.11 below.

```

#####
## Arngrimur Einarsson, November 2007.
## CalcPrice.R
## Script to calculate the price of ZCBs.
#####

ZCBPrice <- function(principal,rate,time){
  ## Function to calculate the price of a ZCB.
  principal/((rate+1)^time)
}

data <- read.table("~/Project/Code/Gams/ArbitrageTest/
.....05_affine_Arb_1Per_4sce.dat")

PRate <- as.matrix(data[2,-c(1,2)])
CRate <- as.matrix(data[3:dim(data)[1],-c(1,2)])
time2mat <- c(30,29,28,27,26,25,24,23,22,21,20,15,10,5,1)

PPrice <- matrix(0, nrow=dim(PRate)[1], ncol=dim(PRate)[2])
CPrice <- matrix(0, nrow=dim(CRate)[1], ncol=dim(CRate)[2])

for(i in 1:length(PRate)){
  tmpRate <- as.numeric(PRate[i])
  PPrice[i] <- ZCBPrice(1,tmpRate,time2mat[i])
}

for(i in 1:dim(CRate)[2]){ #15
  for(j in 1:dim(CRate)[1]){ #8
    tmpRate <- as.numeric(CRate[j,i])
    CPrice[j,i] <- ZCBPrice(1,tmpRate,time2mat[i])
  }
}

write(PPrice, file="05_affine_4_PPrice",length(PPrice))
write(CPrice, file="05_affine_4_CPrice",dim(CPrice)[2])

```

Listing B.11: The script used for conducting the simple arbitrage test.

```
$eolcom #

set i 'Maturity' /zby1,zby5,zby30/;
set s 'states' /n1*n32/;

$include '05_NS_32_PPrice.dat';
$include '05_NS_32_CPrice.dat';

Variables
    psi(s)
    z;

Positive variables slack(i)
                    surplus(i);

Equations
    obj_func
    price(i)
    nominus(s);

obj_func.. z =e= sum(i, (slack(i)+surplus(i)));

price(i).. pprice(i) =e= sum(s,
    psi(s)*(cprice(s,i)+slack(i)-surplus(i))) ;

nominus(s).. psi(s) =g= 0.00001 ;

Model arb_level /all/ ;

solve arb_level using nlp minimizing z ;

.
```



# Bibliography

---

- Akaike, H. (1947), 'A new look at the statistical model identification', *IEEE Transactions on Automatic Control* **19**(6).
- Applied Statistics and Probability for Engineers, third edition* (2003), John Wiley & Sons, inc.
- Bertocchi, M., Giacometti, R. & Zenios, S. A. (2000), 'Risk factor analysis and portfolio immunization in the corporate bond market', *The Wharton Financial Institutions Center, Philadelphia USA* .
- Christiansen, C. & Lund, J. (2007), Revisiting the shape of the yield curve: The effect of interest rate volatility. Working paper.
- Dahl, H. (1996), A flexible approach to interest-rate risk management, in S. A. Zenios, ed., 'Financial Optimization', Cambridge University Press.
- D'Ecclesia, A. & Zenios, S. (1994), 'Risk factor analysis and portfolio immunization in the italian bond market', *The Journal of Fixed Income* .
- Encyclopedia of Statistical Sciences* (1988), John Wiley & Sons Inc.
- Hocreiter, R. & Pflug, G. C. (2007), 'Financial scenario generation for stochastic multi-stage decision processes as facility location problems', *Annual of Operations Research* **152**(1), 257–272.
- Jamshidian, F. & Zhu, Y. (1997), 'Scenario simulation: Theory and methodology', *Finance and Stochastic* pp. 43–67.
- Jarque, C. M. & Bera, A. K. (1987), 'A test for normality of observations and regression', *International Statistical Review* **55**(2), 163–172.

- Kall, P. & Wallace, S. (1994), 'Stochastic programming', *Wiley, Chester* .
- Kaut, M. & Wallace, S. W. (2007), 'Evaluation of scenario-generation methods for stochastic programming'.
- Litterman, R. & Scheinkman, J. (1991), 'Common factors affecting bond returns', *The Journal of Fixed income* .
- Nelson, C. & Siegel, A. (1987), 'Parsimonious modeling of yield curves', *The Journal of Business* **60**(4), 473–489.
- Options, Futures, and Other Derivatives* (2006), Prentice Hall.
- Pfaff, B. (2007), *vars: VAR Modeling*. R package version 0.6-5.  
**URL:** <http://www.pfaffikus.de>
- Pflug, G. C. (2001), 'Scenario tree generation for multiperiod financial optimization by optimal discretization', *Mathematical Programming* .
- Practical Financial Optimization* (2005), Unpublished.
- Probability and Random Processes* (2001), Oxford University Press.
- R Development Core Team (2006), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.  
**URL:** <http://www.R-project.org>
- Rasmussen, K. M. & Poulsen, R. (2007), Yield curve event tree construction for multi stage stochastic programming models. Working paper.
- Time Series Analysis* (1994), Princeton University Press.
- Time Series Analysis* (2001), DTU, IMM.