Swap Book Hedging using Stochastic Optimisation with Realistic Risk Factors

---

Master Thesis 30 ECTS
The Department of Management and Engineering (IEI)
ISRN: LIU-IEI-TEK-A–21/04219—SE

Emil Mårtensson & Rickard Nordin
January 2022

Supervisor:
Jörgen Blomvall, Linköping University

Examiner:
Jonas Ekblom, Linköping University

External Supervisor:
Johan Hagenbjörk, Swedbank

Abstract

Market makers such as large banks are exposed to market risk in fixed income by acting as a counterparty for customers that enter swap contracts. This master thesis addresses the problem of creating a cost-effective hedge for a realistic swap book of a market maker in a multiple yield curve setting. The proposed hedge model is the two-stage stochastic optimisation problem created by Blomvall and Hagenbjörk (2020).

Systematic term structure innovations (components) are estimated using six different component models including principal component analysis (PCA), independent component analysis (ICA) and rotations of principal components. The component models are evaluated with a statistical test that uses daily swap rate observations from the European swap market. The statistical test shows that for both FRA and IRS contracts, a rotation of regular principal components is capable of a more accurate description of swap rate innovations than regular PCA.

The hedging model is applied to an FRA and an IRS swap book separately, with daily rebalancing, over the period 2013-06-21 to 2021-05-11. The model produces a highly effective hedge for the tested component methods. However, replacing the PCA components with improved components does not improve the hedge.

The study is conducted in collaboration with two other master theses, each done at separate banks. This thesis is done in collaboration with Swedbank and the simulated swap book is based on the exposure of a typical swap book at Swedbank, which is why the European swap market is studied.

Acknowledgements

We would like to sincerely thank our supervisor at the Department of Management and Engineering at Linköping University, Jörgen Blomvall. Without his help, the progress of the thesis would have been much slower and would not have yielded such interesting results. He has continuously contributed with his knowledge and important insights which we are very grateful for. Mr. Blomvall also provided valuable input in the underlying mathematical expressions and how to correctly formulate them for readers to better understand.

We also want to express our gratitude to our external supervisor Johan Hagenbjörk at Swedbank who provided us with vital information regarding the European interest rate market. He gave us the information needed for the thesis but also explained how trading within the bank works and why managing risk is so important.

Lastly, we want to thank our examiner Jonas Ekblom for thoroughly reading our report and giving us constructive feedback throughout the whole of the thesis. We also have to mention the other groups since this thesis was made in collaboration with two other theses. We have provided each other with sections of code, help during times of troubleshooting and support throughout the length of the project. To all of the above mentioned, thank you.

Contents

List of Figures

List of Tables

Nomenclature

Abbreviations

cdf	Cumulative Density Function
EONIA	Euro Overnight Index Average
EURIBOR	Euro Interbank Offered Rate
EWMA	Exponentially Weighted Moving Average
FRA	Forward Rate Agreement
GARCH	Generalised Autoregressive Conditional Heteroscedasticity
HB	Handelsbanken
IBOR	Interbank Offered Rate
ICA	Independent Component Analysis
i.i.d	Independent and Identically Distributed
IRS	Interest Rate Swap
MLE	Maximum likelihood Estimation
NPV	Net Present Value
OIS	Overnight Index Swap
ON	Overnight
pdf	Probability Density Function
PCA	Principal Component Analysis
RIC	Refinitiv Instrument Codes
SEB	Skandinaviska Enskilda Banken
Sw	Swedbank

Variables

α	A penalty parameter that punishes decisions that lead to transaction costs in the stochastic programming hedge model
W	Simulated value of swap book
h	Current value of swap book
V_P	NPV of payer swap contracts
V_R	NPV of receiver swap contracts
Δx_P	Number of bought contract of each payer swap type
Δx_R	Number of bought contracts of each receiver swap type
c_P	Proportional transaction fee for bought payer swaps
c_R	Proportional transaction fee for bought receiver swaps
y_{gauss}	Gaussian random variable
$\tilde{\xi}_t$	All risk factors associated with contract price
ξ_t	Significant risk factors
$\xi_{t,0}$	Significant risk factors from risk free forward rate curve
$\xi_{t,\pi}$	Significant risk factors from tenor specific forward rate curve
$P_{t,i}(\tilde{\xi}_t)$	Price of financial security i as a function of all risk factors
$\Delta P_{t,i}(\Delta\tilde{\xi})$	Change in price of financial security as a function of changes in all risk factors
$D_{t,i}$	Cash flow payment from financial security i at time t
$\epsilon^I_{t,i}$	Price error of financial security from exclusion of insignificant risk factors
$\theta_{t-1,i}\Delta t$	Price change due to shortened discounting period (including cash flows) between $t-1$ to t known as the rollover effect
$\Delta P^A_{t,i}$	Taylor approximation of price change as a function of changes in significant risk factors
$g_{t,i}$	Gradient of the price of asset i w.r.t. changes to significant risk factors
$H_{t-1,i}$	Hessian of the price asset i w.r.t. changes to significant risk factors
$\epsilon^A_{t,i}$	Pricing error due to taylor approximation and discrete time increments
$\Delta\epsilon^P_{t,i}$	Pricing error due to noise in market data or pricing model
$\bar{P_{t,i}}$	Observed market price of asset i
V_t	Value of portfolio of financial securities
$h_{t,i}$	Holdings of asset i at time t
$\mathcal{I}$	The set of all financial securities in a portfolio
$\mathcal{E}$	The set of all currencies
$h_{t,e}$	Holdings of currency e
$f_{t,e}$	Currency exchange rate from base currency e to the currency in which the portfolio performance is measured
$D_{t,i}$	Cash flow from the financial security i at time t
$e(i)$	Currency associated with the financial security i
$\Delta f_{t,e(i)}$	Change in exchange rate, with base currency $e(i)$ and the currency in which the portfolio performance is measure as term currency, from time $t-1$ to t
$\Delta P^L_{t,i}$	Change in the price of financial security i, measured in the currency of which the portfolio performance is measured
ΔV^h_t	Change in portfolio value due to asset prices
$\Delta h^B_{t,i}$	Number of financial securities i bought at time t
$\Delta h^S_{t,i}$	Number of financial securities i sold at time t
$r_{e,t-1}$	The simple interest of currency e between time $t-1$ and t

Variables

$\Delta t_{e,t-1}$	Time period between $t-1$ and t, measured with the day counting convention of the market associated with currency e
$s^S_{t,i}$	Spread that the seller has to cross when they sell a financial security i at time t
$s^B_{t,i}$	Spread that the buyer has to cross when they buy a financial security i at time t
$\Delta h^B_{t,e_2,e}$	Amount of currency e deposited from account associated with currency e_2
$f_{t,e_1,e}$	Currency exchange rate with e as base currency and e_2 as term currency, at time t
$s^B_{t,e_1,e}$	Spread that the buyer has to pay for each unit of the base currency e when they buy the currency e_1
$\Delta h_{t,e,e_1}$	Number of units of (term) currency e_1 bought with base currency e
I_n	Identity matrix with n rows and columns
κ	Upper limit of the cost function of the interior value problem
η	Lower limit of the cost function of the interior value problem
$s_i(t)$	Source signal
B	Mixing matrix for source signals
D	Unmixing matrix for source signals (inverse of B)
A	Integrating matrix
$f_0(t)$	The continuous risk free forward rate curve
$f_{t,0}$	The discrete risk free forward rate curve
$\pi_\tau(t)$	The continuous tenor specific forward curve
$\pi_{t,\tau}$	The discrete tenor specific forward curve
π_t	The discrete forward tenor specific forward curve for tenor 6M
f_t	Single interest rate curve
$f_\tau(t)$	The continuous combined forward rate
$\Psi_\tau(t)$	Integral of $f_\tau(t)$ times the time t.
$\Psi_0(t)$	Integral of $f_0(t)$ times the time t.
$\mathcal{T}$	The set of tenors studied
τ	A specific tenor
E	An arbitrary matrix with decomposed vectors
E_π	An matrix with decomposed vectors based on tenor specific forward curve
E_0	An matrix with decomposed vectors based on risk free forward rate curve
r_{float}	Floating rate in FRA/IRS
r_{fix}	Fix rate in FRA/IRS
P_{IRS}	Price for an IRS contract using theoretical valuation
P_{FRA}	Price for an FRA contract using theoretical valuation
$\mathbb{Z}^+$	Integer numbers between one and infinity
F	Arbitrary multivariate distribution
Φ	Cdf of the standard normal distribution
θ	Parameters of a pdf/cdf
u_i	Uniformly distributed random number
μ	Expected value
σ	Standard deviation
x_i	Samples from a distribution

Variables

p	Poisson process
S	Number of scenarios in hedging model
H	General notation for hessian
∇	General notation for gradient
$r_{t,0}^{T_j}$	The spot rate of the discrete risk free forward rate curve evaluated at T_j
$r_{t,\pi}^{T_j}$	The spot rate of the discrete tenor specific forward rate curve evaluated at T_j
$r_{t,\tau}^{T_j}$	The sum of $r_{t,0}^{T_j}$ and $r_{t,\pi}^{T_j}$
λ_i	An eigenvalue, a lower i means a more significant eigenvalue
λ	A matrix with zeros and eigenvalues on the diagonal (largest first)

Functions

$E[\cdot]$	Expected value of a random variable
$f_U(\cdot)$	Pdf of the uniform distribution
$F_U(\cdot)$	Cdf of the uniform distribution
$f_N(\cdot)$	Pdf of the normal distribution
$F_N(\cdot)$	Cdf of the normal distribution
$f_{\text{stud}}(\cdot)$	Pdf of the Student's t-distribution
$\Psi_\tau(\cdot)$	Zero coupon rate multiplied by a time period
$c_{\text{FRA}}(\cdot)$	Cash flow size of an FRA
$h(\cdot)$	Regularisation function for the risk free forward yield curve
$h_\tau(\cdot)$	Regularisation function for a tenor specific forward curve
$g_b(\cdot)$	Function that converts forward rates to market prices
$\text{kurt}[\cdot]$	Kurtosis of a centered random variable
$H(\cdot)$	Differential entropy of a random variable
$J(\cdot)$	Negentropy of a random variable
$G_1(\cdot)$	Function used for approximating negentropy of a random variable
$G_2(\cdot)$	Function used for approximating negentropy of a random variable
$g_1(\cdot)$	Derivative of G_1
$g_2(\cdot)$	Derivative of G_2
$g_1'(\cdot)$	Derivative of g_1
$g_2'(\cdot)$	Derivative of g_2
$f(\cdot)$	Oblique rotation criterion of an arbitrary matrix
$Q(\cdot)$	Oblique rotation criterion of an oblique rotation matrix
$\rho(\cdot)$	Rescales the columns of a matrix to unit length
$dg(\cdot)$	Zeroes all non-diagonal elements of a matrix
$\langle\cdot\rangle_F$	Frobenius inner product of a two matrices
$F_i(\cdot)$	Marginal cdf of a continuous random variable
$C(\cdot)$	Copula of random variables
$F(\cdot)$	Multivariate cdf of random variables
$C_N(\cdot)$	Gaussian copula
$C_{\text{stud}}(\cdot)$	Student's t-copula
$P_{t,i}(\cdot)$	Theoretical price of an asset
$\Delta P_{t,i}(\cdot)$	Theoretical price change of an asset between two time points

Chapter 1

Introduction

Hedging risk on the interest rate market is an important and complex challenge for banks with large interest rate derivative portfolios. The hedging problem on interest rate markets consists of two major problems to create an effective hedge: accurate yield curve estimation and accurate risk factor modelling. In this thesis, the generalised multiple yield curve estimation framework developed by Blomvall and Ndengo (2013) is used to create an improved risk factor model and consequently a more cost-effective hedge.

Banks act as market markers and offer interest rate swap contracts, and in return gain money from customers crossing the bid-ask spread. The bank acts as a counterparty of each customer trade and the portfolio of these swap contracts is called a swap book. Since the bank is the counterparty, this swap book introduces interest rate risk that the bank has to manage. If the risk exposure is not properly managed, the bank could lose large amounts of money when the underlying interest rates change over time. By hedging the swap book, the bank makes the profits less volatile, but it also introduces transaction costs.

Creating hedges with swaps is a common way to hedge against yield changes which can be noted in Swedbank's annual report from 2019 where the outstanding notional amount of interest rate swaps as hedge accounted for 1 billion SEK (Swedbank 2019). To determine the combination of swaps needed to create a hedge, several methods could be used. Traditional methods for hedging on interest markets include bumping, boxes and waves presented by Hagan and West (2006), in which the sensitivity of the portfolio value to simple yield curve changes is measured and hedged. However, these methods result in expensive hedges because they do not take systematic risk or transaction costs into account, which appears indirectly when trading swaps and crossing bid-ask spreads.

Instead, a stochastic programming framework could be used to solve the hedging problem in this setting by generating scenarios of future yield curves to make prominent decisions today with regard to both volatility and transaction costs. To generate yield curves, the systematic risk factors on the interest rate market must be determined to be able to model their changes. This can be done using several techniques. Litterman and Scheinkman (1991) proposes principal component analysis, a model that is commonly used for dimension reduction as it maximises the variance explained by the components. Nonetheless, there is no guarantee that this is the best risk factor model.

The risk factors have been observed to be related to changes in the forward rate yield curve, as

seen in research made by Nelson and Siegel (1987). The forward rate yield curve is unobservable and requires modelling via an inverse problem. A desirable property of the yield curve model is that small changes in the input should result in small changes in the output. A generalised framework for multiple yield curve estimation developed by Blomvall and Ndengo (2013) has been shown to have the desired properties by having noise reducing effects toward input deviations in comparison to other methods. The noise could originate from several factors such as incorrect pricing, non-synchronised data, different contributors or misprints. The resulting forward yield curves have also been shown to be decomposable by data-driven dimension reduction techniques that result in systematic forward yield curve movements that have been termed systematic risk factors by other authors. The area of feature extraction on the yield curves from the generalised yield curve estimation framework has not been properly explored. Because of the paramount importance risk factors have on hedging decisions, different methods of feature extraction can have a significant impact on the performance of the hedged swap book.

Blomvall and Hagenbjörk (2020) present a stochastic programming optimisation problem that weighs portfolio variance against transaction costs in a single interest rate curve setting. Their proposed model can be extended to multiple yield curves, but that requires accurate measurements of tenor premiums (compensation for being exposed to longer tenors and thereby higher credit risk) and their corresponding risk factors.

1.1 Background

The Swedish bank Swedbank is primarily active in the segments of loans, payments and savings. Swedbank also trade securities on a daily basis as a market maker. To reduce their exposure to interest rate risk, they enter interest rate swap contracts with other banks. Interest rate risk in this context infers a change in value of the swap book of the bank due to yield changes. The risk appetite within the bank is set by the board. All parties which are allowed to expose the company to market risk have a risk limit issued by the CEO (Swedbank 2019). Risk management is vital for traders and failure to comply to the risk limit results in immediate dismissal (Hull 2018).

The importance of risk management is clear but how it should be accomplished is not. The owner of a swap book must weigh transaction costs against variance when performing a hedge to avoid unnecessary costs. By creating a more cost-effective hedge based on Blomvall and Hagenbjörk (2020) in relation to traditional methods like boxes and waves by Hagan and West (2006) the bank is able to spend less money to set up and maintain it. This is the reason for this study: to examine if this can be utilised based on a typical swap book at Swedbank. To create an effective hedge, representative and realistic risk factors must be used to generate scenarios in the stochastic programming problem. Estimating these factors is no simple task and there exists no self-evident answer to the authors' knowledge which is why this area will be explored.

1.2 Purpose

The purpose of this thesis is to create a realistic risk factor model for interest rate market risk to improve the stochastic optimisation based hedging framework developed by Blomvall and Hagenbjörk (2020) and expand it to a multiple yield curve framework.

1.3 Delimitations

The delimitations that have been made to limit the study are:

- The hedge and swap book can only consist of (EURIBOR) IRS and FRA contracts with tenor 6 months.
- The time to maturity of contracts in the swap book and the hedge is limited to 10 years.

Chapter 2

Methodology

In this chapter, an overview of the methodology proposed to fulfil the purpose of the study is presented. The layout of the approach is presented in figure 2.1 below.

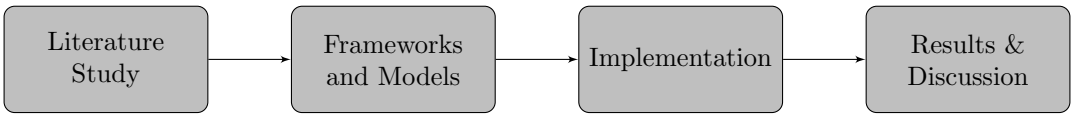


Figure 2.1: Overview of the methodology in this thesis.

Initially a literature study is performed to orient the authors on available models and the current state of research. This results in possible frameworks to fulfil the purpose of the study. An overview of theories is presented in section 2.2. Based on the potential approaches, a method is developed to produce the result of the thesis within the available time frame. The chosen method and models are then implemented and are described further in section 2.3.

This master's thesis is done in collaboration with two other theses with the same end purpose; creating a hedge for a swap book, each with different specialisations. For an overview of the joint methodology see figure 2.2, segments marked Swedbank are the specialisations of this thesis and are studied in detail and later implemented.

2.1 Literature Study

The literature study is made to orient the authors on available models and give them an idea of what has been done in the area previously to be able to answer the purpose. The basis of this inquiry is foremost made out of books and publications within finance, machine learning, data dimension reduction and signal processing. This enables the authors to suggest prominent methods to answer the purpose.

The databases that mainly are used, are Scopus, DiVA, and Libris. This selection was made to find high-quality sources. Recent sources with many citations were prioritised in the literature study.

2.2 Frameworks and Models

This thesis is part of a collaboration with two other theses (at Handelsbanken and Skandinaviska Enskilda Banken) that both have hedging of a swap book as purpose. Each thesis has an individual research area that is analysed in order to improve the hedge as a whole. An overview of the joint methodology is displayed in figure 2.2.

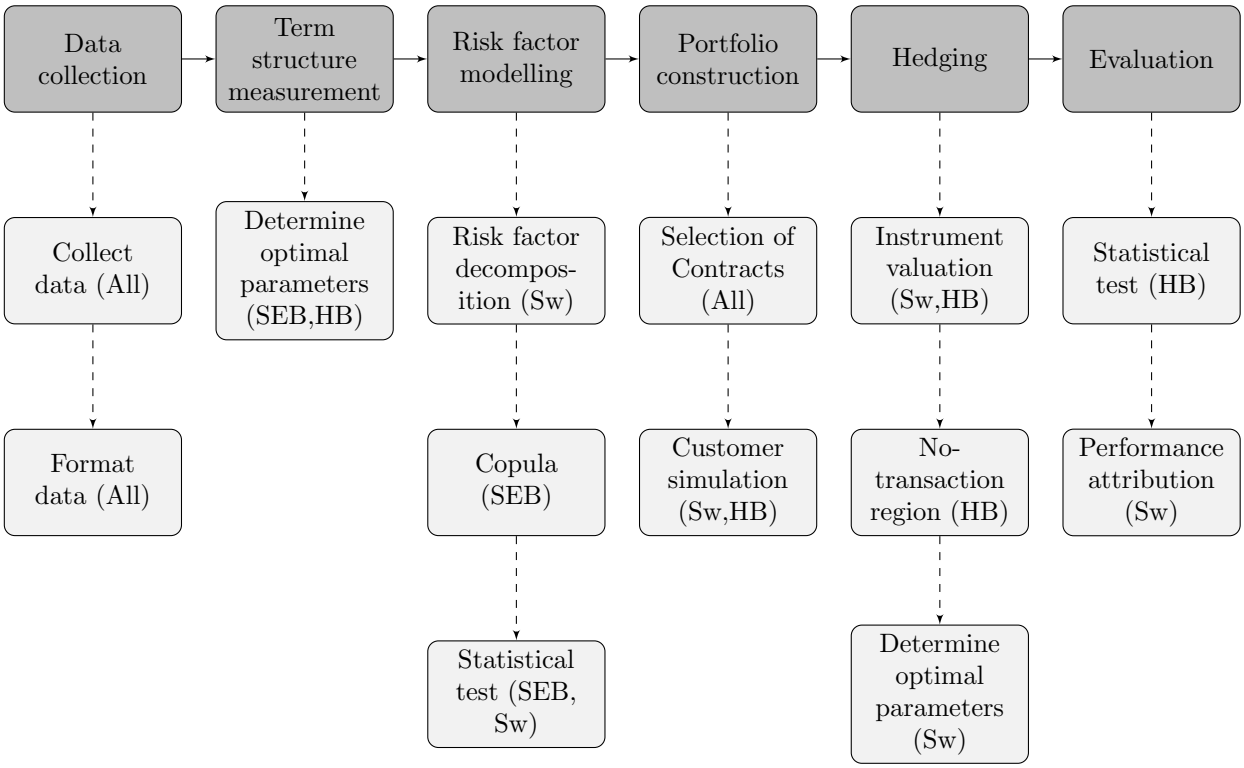


Figure 2.2: Overview of the framework in this thesis in relation to the collaborative theses.

In this thesis, all of these areas are described. However, more focus will be on the areas marked with Swedbank (Sw). A brief overview of the models that could be used to answer the purpose of this thesis is presented in this section.

2.2.1 Data Collection

The data is divided into an in-sample period and an out-of-sample period to avoid overfitting of model parameters and create a realistic scenario for the bank. The information is primarily collected from Refinitiv Eikon. Floating rates, prices on contracts and their specific details such as maturity dates will be gathered. These values are required for yield curve estimation according to the multiple yield framework proposed by Blomvall and Ndengo (2013). The values are also needed for portfolio valuation and performance attribution of the hedge model.

2.2.2 Term Structure Measurement

Yield curves are required for the pricing of contracts since all contracts that contain some form of discounting or IBOR rates depend on yield curves. Before the financial crisis of 2007, swap contracts of different tenors were priced consistently with one single yield curve. However, during and after the crisis, inconsistencies between the pricing of swaps of different tenors were noticed, as observed by Bianchetti (2011). This implies that there exist distinct term structures for each tenor. Because of this discrepancy, a multiple yield curve framework is needed. A multiple yield curve framework separates yield curves into individual curves that describe the credit risk premium associated with each tenor.

Estimation of a single yield curve is an inverse problem, which means that there are infinitely many different yield curves that can reprice contracts exactly. Many models are exact models, where the repricing of contracts must be exact. In a research paper made by Blomvall and Ndengo (n.d.), it is established that noise is present in the quoted contract prices. In the article, an inexact single yield curve estimation model is presented that is more stable over time. It also performs better than other, commonly used single yield curve models in terms of creating realistic curves. This makes the model useful for dimension reduction purposes since the data has a more systematic behaviour.

Multiple yield curve estimation models are typically extended from a single yield curve framework. In Blomvall and Ndengo (2013), the previously mentioned single yield curve model is extended to a multiple yield curve model that has the same stable properties as their single yield curve model. The multiple yield curve model is thus a suitable model to use for the estimation of risk factors in a multiple yield curve setting. The model has a regularisation parameter for curve smoothness. This parameter must be set in order to balance the trade-off between pricing errors and curve smoothness. The trade-off is studied in detail by de Flon and Landgren (2021). The estimated yield curves are then used for the estimation of risk factors.

2.2.3 Risk Factor Modelling

After the yield curves have been estimated, risk factors can be extracted. Risk factors on the interest rate market are systematic changes in yield curves. Since yield curves are continuous, they are an example of infinite-dimensional data. These continuous signals can be approximated as discrete, with a daily discretisation of the duration. But the discretised yield curves are still high-dimensional and must be further decomposed by dimension reduction techniques. Analysis of high-dimensional data is a common theme in a variety of topics, such as analysis of brain signals, image analysis or quantitative studies in psychology.

There are many different dimension reduction techniques that rely on different assumptions about the underlying data and result in different extracted components, and an extensive review of the active research field of dimension reduction has been provided by Burges (2009). Some dimension reduction models are also misused in different ways, resulting in suboptimal or even unrealistic components. Common pitfalls have been explained in a very pedagogic paper authored by Preacher and MacCallum (2003). In the problem of risk factor estimation on the interest rate market, the true number of risk factors is unknown. It is also unknown if these true risk factors are independent or not, if they have a nonlinear relationship and if they are stationary. For this reason, it is important to choose a model that makes as few of these assumptions as possible.

Two common techniques are principal component analysis (PCA) and independent component analysis (ICA). PCA extracts components that maximise the variance of the data when it has been projected onto the components (or equivalently, minimises the squared residual error of the projected data). The extracted components are linearly uncorrelated for the in-sample data. ICA is similar to PCA but extracts components that are maximally independent instead. This is achieved by maximising the non-Gaussianity of the projected data. Non-Gaussianity is commonly measured as the absolute value of excess kurtosis of the projected data. Both PCA and ICA and their extensions are discussed in detail by Hyvärinen, Karhunen and Oja (2001). Both of these models assume that the risk factors are uncorrelated or independent, which is an assumption that could result in unrealistic risk factors if the true risk factors are correlated.

Another promising dimension reduction technique involves the usage of neural networks, more specifically a feedforward network. A feedforward network is a nonlinear model that consists of layers of neurons (weights). Each output variable of a layer is a nonlinear function, known as the activation function, applied to a linear combination of the input variables. The number of layers and neurons in each of them are set as parameters, while the weights are determined by the optimisation of a cost function. If there are more input variables than output variables, the model is forced to perform a dimension reduction of the data. Dimension reducing neural networks are called autoencoders. This is a very computationally demanding method, as the optimisation problem in this setting has multiple local optima and a very large number of variables. For this reason, dimension reduction of yield curve innovations using autoencoders is not pursued further in this thesis. A more thorough explanation of neural networks can be found in Bishop (2006) or Goodfellow, Bengio and Courville (2016). A practical example of an autoencoder is described by Hinton and Salakhutdinov (2006).

All curve estimation models result in noisy yield curves because they solve an inverse problem. A bad yield curve (unrealistic outlier) affects dimension reduction models that are based on higher-order moments greatly. In an article about performance attribution by Blomvall and Hagenbjörk (2019a), in which risk factors are extracted, there is a problem with noisy yield curves that is handled by a data processing step before the risk factors are estimated. Thus a desirable property of the risk factor estimation model is robustness against noise.

The risk factors are stochastic variables describing the observed magnitudes of the systematic yield curve innovations. Volatility in the financial market is almost always heteroscedastic. This is also assumed to be the case for the extracted risk factors. There are multiple models that estimate the time-varying volatility, for example exponentially weighted moving average (EWMA) or generalised autoregressive conditional heteroscedasticity (GARCH). The mentioned models depend on parameters that require estimation. This is typically done by maximum likelihood estimation.

Marginal distribution functions for random variables give information about the multivariate distribution of the random variables. A copula is a multivariate cumulative distribution function that describes the dependence between random variables. Univariate marginal distributions together with a copula can be used to model any multivariate joint probability distributions. This implies that marginal univariate functions paired with a copula is an equivalent description of the multivariate cumulative distribution function for the risk factors and can thus be used for simulation, as described by Sklar (1959). Examples of copulas are the Gaussian copula and Student's t-copula. Both have parameters that can be estimated with maximum likelihood estimation.

After the risk factors have been estimated with the different risk factor models and the copulas have

been fitted for each model, the performance of the estimated risk factors are evaluated. This is done according to the statistical test proposed by de Flon and Landgren (2021), in which the estimated ability of the risk factors to explain changes in observed swap rates on the European interest rate market is measured. The model or models that perform best in the test are used in the hedge model to simulate future values of the swap book. The simulated swap book values are used to estimate the variance of the swap book. The chosen risk factor model will also be used in the performance attribution framework, using PCA as a reference component model for comparison.

2.2.4 Portfolio Construction

The portfolio is constructed in collaboration with Swedbank to create a portfolio that replicates the exposure of a typical swap book at Swedbank. This resulted in a portfolio consisting of only IRS and FRA contracts on the European market with a time to maturity of up to 10 years. Even with these restrictions, the studied case is still very realistic. FRA and IRS contracts made up around 90% of the European OTC swap market according to Bank for International Settlements (2020), and the trading activity of swaps with longer time to maturity than 10 years is significantly lower (Swedbank 2021).

In order to create a realistic hedge (from Swedbank's perspective), which changes every day, customer trades are simulated. This means determining the number of trades for each tenor on both paying and receiving sides as well as the nominal value of the trades each day. The simulation of customer trades is described in the theory section, 3.1.5. Parameter values for the customer simulation are determined in collaboration with Swedbank and presented in the method, 4.5.

2.2.5 Hedging

Hedging allows an investor to reduce or eliminate unwanted exposure to risk factors. To do so, the investor must pay a transaction fee or cross a bid-ask spread. The investor is thus forced to make a trade-off between transaction costs and future portfolio variance. The estimated risk exposure depends on the risk factors, thus different risk factors result in different optimal hedging decisions.

The institutional risk management process is summarised by Hull (2018). It is traditionally done by measuring exposure to bumps in the yield curve but also by viewing simple risk factors such as duration, modified duration or bucket deltas (boxes). To estimate the variance of the portfolio, it is possible to use the Monte Carlo simulation of the risk factors and measure a sample variance of the swap book values over the scenarios. This requires the valuation of the contracts with regard to risk factors.

Valuation of IRS and FRA contracts using a multiple yield curve framework has been described in detail by Blomvall and Ndengo (2013). First, the future IBOR rate is estimated using the credit risk premium curve. The expected IBOR rate(s) are then used in a discounted cash flow model that determines the expected NPV of the swap contract.

The hedging model that will be used is the model proposed by Blomvall and Hagenbjörk (2020). The model is based on a single yield curve framework and requires the multivariate distribution of the risk factors and their loading to be known, for simulation of portfolio values. By using valuation models based on multiple yield curves and simulating the risk factors for each yield curve, the

hedging model can be extended to the multiple yield curve setting.

The hedge model is a two-stage stochastic programming model, which makes the model faster than tree-based optimisation models. The latter would take an unreasonably long time to run because every decision requires a new Monte Carlo estimation of the swap book variance. As shown in the results by Blomvall and Hagenbjörk (2020), the performance is still better than other common hedging methods.

The hedging model results in the purchase of new contracts to hedge the original swap book. The next day, the hedge is created to hedge the contracts from the day before (both original swap book and the hedge that day) and the customer trades entered today which results in a swap book that increases in size every day until contracts start to expire.

2.2.6 Evaluation

The total value of the portfolio (including transaction costs) is compared to the portfolio volatility. This results in an efficient frontier which, depending on the risk appetite of the bank, tells the bank how optimal parameters should be set in the hedging model to achieve a cost-effective hedge. The hedge model is also tested against the traditional method of boxes described in Hagan and West (2006), in terms of portfolio value and volatility. A more robust hedge parameter analysis and the comparison between the two hedge models is described and presented by Bohlin and Harling (2021).

Performance attribution of the swap book and the hedge is performed over time to attribute portfolio gains to specific factors of the portfolio. This is vital to do in order to ensure that the portfolio gains from the Blomvall and Hagenbjörk (2020) model originates from lower transaction costs and not from exposing the portfolio to more risk. The performance attribution is based on the residual free framework that is described in Blomvall and Hagenbjörk (2019a). In the performance attribution, the change in price between two days will be expressed as a second-order Taylor expansion with regards to changes in risk factors. This is the foundation for the performance attribution and further progression is made to only consider the most significant risk factors. In the attribution effects regarding risk factor exclusion, carry effect and error from Taylor expansion will be computed. Based on these results the complete swap book can be broken down into values from risk factors, error terms and carry effect. Due to the portfolio consisting of both the original swap book and the hedge, some effects from the hedge must be added. These effects are described in detail by Blomvall and Hagenbjörk (2020) and include the transaction costs for the bank creating a hedge and the earnings from customers crossing the bid-ask spread.

A hedge model is better if the hedged portfolio has lower transaction costs for each portfolio volatility level. It is possible that a hedging model only outperforms others in a limited region. In that case, the risk aversion of the investor determines what model they would prefer. The evaluation of the performance of the stochastic programming hedge model against a more simple hedging method is described by Bohlin and Harling (2021). To evaluate if a good representation of the systematic risk factors has been found, the statistical test developed by de Flon and Landgren (2021) and the risk factor error in the performance attribution is used to compare the different risk factor models. The statistical test measures the ability of the risk factor model to explain changes in observed swap rates.

2.3 Implementation

The chosen method and models are mainly implemented in Matlab. For a deeper understanding of the implementation phase, see figure 2.2. The areas marked with 'Sw' (Swedbank) & 'All' are areas that are described and implemented in this thesis. These are the areas that will be implemented in this thesis and those marked with SEB and HB will not be studied in detail as a part of this master's thesis.

The optimisation problem used to create the hedge has to be solved and for this AMPL is used as it is a versatile solver which is easy to use and can compute large-scale complex problems. Due to the low amount of data available on OIS contracts, the same in-sample period is used for parameter estimation of the yield curve model and risk factor estimation models.

Chapter 3

Theory

This chapter presents the theoretical framework needed to fulfil the purpose of improving the stochastic programming model. Portfolio construction, hedge formulation and performance attribution are also described in detail in this section. Multiple theories on data decomposition that are applicable to yield curve data are also presented.

3.1 Basic Statistics

This section contains a brief introduction to the basic statistical tools that are used in the other theoretical sections of this thesis. The mathematical formulas for the distributions and the Poisson process in this section are described by Ross (2010) if not otherwise stated.

3.1.1 Covariance Matrix

The covariance of two random variables X and Y, with mean values μ_X and μ_Y respectively, is defined as

$$\mathrm{Cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)]. \tag{3.1}$$

If $Z \in \mathbb{R}^{n\times 1}$ is a n-dimensional random vector $Z = (Z_1, Z_2, \ldots, Z_n)^T$, the covariance matrix $C_Z \in \mathbb{R}^{n\times n}$ of Z is

$$C_Z = E[(Z - \mu_Z)(Z - \mu_Z)^T], \tag{3.2}$$

where μ_Z is the mean vector of the random vector Z.

3.1.2 Uniform Distribution

A univariate distribution of a random variable describes the probability of outcomes of a random variable. An example of this is the uniform distribution. The uniform distribution is often used to generate random numbers between 0 and 1. The probability density function (pdf) of a uniformly distributed random variable X with outcomes $x \in [a, b]$ is described in Blom et al. (2005):

$$f_U(x) = \begin{cases} \frac{1}{b-a}, & \text{if } a < x < b \\ 0, & \text{otherwise.} \end{cases} \tag{3.3}$$

They also present the cdf of the uniform distribution as

$$F_U(x) = \begin{cases} 0, & \text{if } x < a \\ \frac{x-a}{b-a}, & \text{if } a \leq x \leq b \\ 1, & \text{if } x > b. \end{cases} \tag{3.4}$$

Uniformly distributed random variables are denoted u_i where i is the variable index if multiple random variables are generated. If a and b do not correspond to zero and one respectively, an interval is presented along with u_i and this interval corresponds to the values a and b, where the lower bound is a and upper bound b.

3.1.3 Normal Distribution

Another commonly used distribution is the normal distribution, of which the pdf is

$$f_N(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}. \tag{3.5}$$

The cdf of the normal distribution is defined as

$$F_N(x) = \int_{-\infty}^{x} f_N(x)dx. \tag{3.6}$$

For future reference, $F_N(x)$ and $f_N(x)$ denotes the distribution with μ=0 and σ=1. If other parameter settings are used these are specified like this, $F_N(x, \mu, \sigma)$ and $f_N(x, \mu, \sigma)$. The parameters μ and σ are mean value and standard deviation, respectively.

3.1.4 Student's t-Distribution

Financial data modelling often requires random variables where extreme outcomes are more common than those of normal distributed random variables. This makes Student's t-distribution a good candidate. This distribution is a generalisation of the normal distribution, where the shape of the tails of the distribution is controlled with an additional parameter ν. Student's t-distribution (pdf) is defined by Yanev and Ahsanullah (2012) as

$$f_{\text{stud}}(x) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{\pi\nu}} \Big(1 + \frac{x^2}{\nu}\Big)^{-(\nu+1)/2} \tag{3.7}$$

with the degrees of freedom, $\nu \in \mathbb{Z}^+$. The function Γ is the gamma function. As ν approaches infinity, the distribution turns into a normal distribution.

3.1.5 Poisson Process

The Poisson process describes the accumulation of a Poisson distributed random variable. The Poisson process, $p(t), t \geq 0$ with rate λ, $\lambda \geq 0$ satisfies three conditions:

1. $p(0) = 0$

2. The increments are independent
3. The number of events in any interval of length t is Poisson distributed with mean λt $\forall s, t \geq 0$ $P(p(t+s) - p(s) = n) = e^{-\lambda t} \frac{(\lambda t)^n}{n!}$.

The expected value of this process is $E[p(t)] = \lambda t$.

3.2 Contract Valuation

The theory for the contracts that are valuated in this study is presented here, along with an introduction to the notations used.

Using multiple yield curves, an accurate valuation of contracts with future interest payments can be performed. This is done by isolating two yield curve components $f_0(t)$ and $\pi_\tau(t)$. The risk-free yield curve component $f_0(t)$ is used as a discounting tool for tenor free cash flows, while the tenor specific forward curve $\pi_\tau(t)$ is the term premium for fixating the interest rate over a tenor, τ. The spot rates for a future time point, T is denoted $\psi_0(T)$ and $\psi_\tau(T)$. These spot rates are

$$\psi_0(T) = \int_0^T f_0(t)dt \tag{3.8}$$

and

$$\psi_\tau(T) = \int_0^T f_\tau(t)dt = \int_0^T (f_0(t) + \pi_\tau(t))dt. \tag{3.9}$$

Discretised forward rate curves are often studied as the curves can seldom be described analytically. To integrate a discretised yield curve, an integrating matrix can be used. Blomvall and Hagenbjörk (2019a) used the square matrix A, defined as

$$A_{i,j} = \begin{cases} \frac{1}{i}, j \leq i \\ 0, \text{otherwise} \end{cases} \tag{3.10}$$

where i, j are indices for rows and columns respectively in A.

3.2.1 Forward Rate Agreements

A forward rate agreement (FRA) is an interest rate derivative where two parties exchange a future floating rate for a fixed rate. This contract (where the fixed rate is paid and the floating received) can be valuated as a future cash flow, determined by formula (3.11) according to Hull (2015).

$$c_{FRA}(T_1, T_2) = N \frac{(r_{float} - r_{fix})\Delta T}{1 + r_{float}\Delta T} \tag{3.11}$$

Where T_1 and T_2 are floating rate fixing date and the maturity date of the contract respectively. N denotes the nominal value of the contract and r_{fix} is the fixed rate in the contract and is determined on the day that the contract is entered. The value of r_{fix} is set by the contract seller (the bank). It is usually set to the expected value of the forward floating rate over the active time of the contract

and a spread premium that the customer has to cross. r_{float} is the floating rate from the market between T_1 and T_2. The rates are here given on simple form. ΔT is the time difference between T_2 and T_1 with the appropriate day count convention.

In the multiple yield curve framework developed by Blomvall and Ndengo (2013) the floating rate is approximated as

$$r_{float} = E_t^{T_2}[L_\tau(T_1, T_2)] = \frac{e^{\psi_\tau(T_2)-\psi_\tau(T_1)} - 1}{\Delta t}. \tag{3.12}$$

Where τ denotes the tenor of the contract and the expected value is given under the forward neutral measure. $L_\tau(T_1, T_2)$ is the floating forward rate between T_1 and T_2 for tenor τ. t denotes the time point that floating rate is valuated on. This expression inserted in (3.11) gives

$$c_{FRA}(\tau, T_1, T_2) = N\frac{(E_t^{T_2}[L_\tau(T_1, T_2)] - r_{fix})\Delta T}{1 + E_t^{T_2}[L_\tau(T_1, T_2)]\Delta T}. \tag{3.13}$$

Up to time point T_1 the expected value in equation (3.12) is used as an estimate of the floating rate. If valuation is being done on or after T_1 the expected value is not needed because the floating rate, r_{float} is known on the financial market. This simplified expression is presented in (3.14).

$$c_{FRA}(\tau, T_1, T_2) = N\frac{(r_{\tau,float} - r_{fix})\Delta T}{1 + r_{\tau,float}\Delta T} \tag{3.14}$$

Where $r_{\tau,float}$ is the correct floating rate with a certain tenor for the particular contract studied. At time point T_2 when the contract expires, the difference between the floating leg and the fixed leg is paid by one of the parties.

3.2.2 Interest Rate Swaps

An IRS contract is similar to an FRA contract but instead of one payment in one period, this swap contract has several payments resulting in multiple cash flows. The value of an IRS contract is the difference in NPV between the floating payments, P_{float} and fixed payments, P_{fix} according to Hull (2015). The value of a swap, in the position of a fixed-rate payer, can be regarded as a short position in a fixed-rate bond and a long position in a floating-rate bond. This results in the expression

$$P_{IRS} = P_{float} - P_{fix}. \tag{3.15}$$

The fixed leg P_{fix} is viewed as a bond with coupons of y except that the nominal value is not exchanged. The fixed rate of the interest rate swap in the multiple yield curve framework by Blomvall and Ndengo (2013) is calculated as

$$y = \frac{\sum_{(i,j)\in\mathcal{I}} \Delta\hat{t}_i E_t^{T_j}[L_\tau(T_i, T_j)] \exp\left\{\Big(\psi_0(T_0) - \psi_0(T_j)\Big)\right\}}{\sum_{j\in\mathcal{J}} \Delta\hat{t}_j \exp\left\{\psi_0(T_0) - \psi_0(T_j)\right\}}. \tag{3.16}$$

The day counting convention used in ψ is assumed to be the same as the convention used by the floating (IBOR) rate. The fixed payments are made at time points T_j, $j \in \mathcal{J}$ and floating payments

at time points T_i, $i \in \mathcal{I}$.

Before the floating rate has been fixed for the corresponding floating payment this cash flow is stochastic. To calculate the expected value of the IBOR rate, (3.12) is used. This combined with (3.16) makes it possible to calculate the expected theoretical swap rate for any IRS using only the risk-free forward rate and tenor specific premium observed at the trade date of the contract. The expression for the fixed rate in (3.16) together with the estimate of the floating rate in (3.12) is all that is needed to value the IRS according to (3.15) as they determine the value of the two bonds along with contract-specific data found on financial markets.

3.2.3 Overnight Index Swaps

An overnight index swap (OIS) is a swap contract where a floating rate is exchanged for a fixed rate. An OIS is very similar to the IRS, as they both have recurring fixed and floating payments. The floating rate is the daily compounded overnight (ON) rate fixed at the end of each business day. On the European market, this rate is called the Euro Overnight Index Average (EONIA). Because the overnight rate spans such a short time interval, the credit and liquidity risks are neglected and thus the contract can be considered risk free. Due to this phenomenon, the OIS is often considered as the best proxy for the risk-free rate.

In the multiple yield curve framework developed by Blomvall and Ndengo (2013) the OIS can be modelled as an IRS but only considering the risk free forward rate curve. The fixed rate in the OIS contract is set to make the floating respectively fixed payments cancel out on the entering date which results in a contract worth nothing. The corresponding equation for the fixed rate, y, behind this reasoning is presented here as

$$y = \frac{d_{t,T_0} - d_{t,T_N}}{\sum_{i=1}^{N} \Delta_i d_{t,T_i}} \tag{3.17}$$

where d denotes the discount factors based on the risk-free interest rate. The two indices on d represent the two dates that the discounting is made between. The discount factor $d_{t,T}$ is defined as

$$d_{t,T} = e^{-\psi_0(T)T}, \tag{3.18}$$

where the spot rate $\psi_0(T)$ has been observed at time t.

The second index indicates from what time point the discounting is made and the first index infers to what time point the discounting is done. Δ_i represents the year fraction (with the appropriate day count convention) for the the fixed payments between time points $[T_{i-1}, T_i]$, $i = 1, ..., N$. N is the total number of fixed payments in the contract.

Using the notations defined in section 3.2 the expression can be rewritten to

$$y = \frac{e^{-\psi_0(T_0)} - e^{-\psi_0(T_N)}}{\sum_{i=1}^{N} \Delta_i e^{-\psi_0(T_i)}} \tag{3.19}$$

which is the fixed rate of the contract expressed in the current interest rate curves.

3.2.4 Market Conventions

For the FRA, IRS and OIS contracts described above, the value of the contracts is not quoted on financial markets. Instead, fixed rates are quoted, leaving the discounting calculations to the investor.

In Refinitiv Eikon, these quotes include bid and ask rates which are set by the market maker. The bid rate is lower than the corresponding ask rate because the market maker wants to earn money from customers entering positions. If an investor buys one of the above-mentioned swap contracts in the position of a fixed rate payer the higher rate (ask) will be used as the fixed rate of the contract. If an investor buys the contract in the position of a floating rate payer (fixed rate receiver) instead, the lower rate (bid) will be received.

As the theoretical interest rate (which results in a NPV equal to zero) might not lie exactly between the bid and ask rate, the sign of the contract value can not be changed in order to determine the value of the other direction. It has to be calculated explicitly via the valuations formulas with the appropriate rate (bid or ask) depending on the direction of the contract. An investor can always buy any swaps in both directions, but they must always pay a small premium included in the fixed rate of the contract which is called perceived transaction cost in this thesis.

3.2.5 Day Count Conventions

Financial markets use different day count conventions to measure time periods that are used for contract valuation. Swap contracts are separated into two legs: The floating leg and the fixed leg. Each leg has a different day count convention and payment frequencies. The following information has been gathered from Swedbank (2021) and is valid on the European interest rate market.

EURIBOR FRA contracts have two settlement days. The fixed leg follows the 30/360 day count convention and has one payment. The floating leg of a 6M EURIBOR FRA has one payment and follows the 30/360 day count convention.

An EURIBOR IRS has two settlement days. The fixed leg follows the 30E/360 day count convention and has annual payments. The floating leg of a 6M EURIBOR IRS has semi-annual payments and follows the ACT/360 day count convention.

EONIA OIS contracts have zero settlement days. The fixed leg follows the ACT/360 day count convention and has annual payments. The floating leg of an EONIA OIS has annual payments and follows the ACT/360 day count convention.

The EURIBOR rate is a spot rate that is valid two days after it has been fixed. The standard tenor on the European market is six months. The RIC for the 6M rate is EURIBOR6MD=.

3.3 Yield Curve Estimation

In this section, the theory behind the method used to generate multiple yield curves on an interest rate market is presented. The terms forward rate curves, term structures, interest rate curves and yield curves are used interchangeably when referring to the multiple forward rate curves derived in this section.

3.3.1 The Generalised Optimisation Framework

Blomvall and Ndengo (2013) describe how to estimate multiple yield curves via an optimisation problem that allows for deviations from market yields and penalises the non-smoothness of the curves. The optimisation model used to generate yield curves is presented in equation (3.20) below.

One risk-free curve, f_0, is produced and multiple tenor specific forward curves, π_τ, for different tenors τ.

$$\begin{aligned} \min_{f_\tau, z_e, z_b, \pi_\tau} \quad & h(f_0) + \frac{1}{2} z_e^T Z_e z_e + \frac{1}{2} z_b^T Z_b z_b + \sum_{\tau \in \mathcal{T}} h_\tau(\pi_\tau) \\ \text{subject to} \quad & g_e(f) + F_e z_e = x_e \\ & x_l \leq g_b(f) + F_b z_b \leq x_u \\ & f_\tau = f_0 + \pi_\tau; \tau \in \mathcal{T} \\ & \pi_{\tau_i} \geq \pi_{\tau_{i+1}} \\ & f_0 \geq f_l \end{aligned} \tag{3.20}$$

where

- $g_e(f)$, $g_b(f)$ are functions which convert forward rates to market prices.
- $f = \{f_\tau\}_{\tau \in \mathcal{T}_E}$ where $\mathcal{T}_E = \{0\} \cup \mathcal{T}$.
- $\mathcal{T} = \{\tau_1, \tau_2, ..., \tau_M\}$ denotes the set of tenors which forward rate curves are generated for.
- The variables z_e and z_b are deviations from unique and bid/ask market prices respectively.
- The functions h and h_τ describes regularisations of the curves by penalising non-smoothness in terms of first- and second-order derivatives.
- x_l and x_u are lower and upper bounds for the market prices for contracts with both a bid and ask price.
- x_e denotes the vector with unique (only bid or ask price is quoted) market prices.
- F_e and F_b are diagonal matrices and determine what instruments are allowed to deviate from market prices.
- Z_e and Z_b are diagonal matrices which determine how much an instrument should be punished in the cost function for deviating from the market price.
- f_l is the lower bound for the risk free forward rate.
- f_0 is the risk free forward rate curve.
- π_τ is the tenor specific forward curve for tenor τ which exists as a compensation for being exposed to contracts with longer maturities.

The optimisation problem is based on the discretisation $T = T_1, T_2, ..., T_n$ of the curves and uses the regularisations $h(f_0)$ and $h_\tau(\pi_\tau)$, defined as

$$\begin{aligned} h(f_0) = & \frac{1}{2} \sum_{t=0}^{n-2} a'_{t,0} \left(\frac{f_{t+1,0} - f_{t,0}}{\Delta T_t} \right)^2 \Delta T_t \\ & + \frac{1}{2} \sum_{t=1}^{n-2} a''_{t,0} \left(\frac{2}{\Delta T_{t-1} + \Delta T_t} \left(\frac{f_{t+1,0} - f_{t,0}}{\Delta T_t} - \frac{f_{t,0} - f_{t-1,0}}{\Delta T_{t-1}} \right) \right)^2 \frac{\Delta T_{t-1} + \Delta T_t}{2} \end{aligned} \tag{3.21}$$

and

$$h_\tau(\pi_\tau) = \frac{1}{2}\sum_{t=0}^{n-2} a'_{t,\tau}\left(\frac{\pi_{t+1,\tau}-\pi_{t,\tau}}{\Delta T_t}\right)^2 \Delta T_t + \frac{1}{2}\sum_{t=1}^{n-2} a''_{t,\tau}\left(\frac{2}{\Delta T_{t-1}+\Delta T_t}\left(\frac{\pi_{t+1,\tau}-\pi_{t,\tau}}{\Delta T_t} - \frac{\pi_{t,\tau}-\pi_{t-1,\tau}}{\Delta T_{t-1}}\right)\right)^2 \frac{\Delta T_{t-1}+\Delta T_t}{2}. \tag{3.22}$$

The functions $a'_{t,\tau}, a''_{t,\tau}, a'_{t,0}, a''_{t,0}$ are penalty functions which are set to make the resulting curves to smooth. The time increment is defined as $\Delta T_t = T_{t+1} - T_t, t = 1, 2, ..., n-1$.

3.4 Risk Factor Decomposition Models

Yield curves are examples of infinitely long and infinitesimal data that change over time. To analyse yield curves, assumptions need to be made. The regularised framework for estimating yield curves build discretised yield curves that are only defined up to the maturity of the longest included contract. This makes the yield curves suitable for data-driven dimension reduction techniques. The estimated forward yield curves f_t follow a systematic behaviour. This suggests that the dimension of f_t can be reduced significantly since the full representation uses $\frac{T}{\Delta T}$ variables, where T is the longest maturity of the yield curve and ΔT the length of the discrete time steps of the curve.

Since the price of interest derivatives is heavily impacted by the changes in the yield curves $\Delta f_t = f_t - f_{t-1}$, describing and understanding these changes are of paramount importance. Blomvall and Hagenbjörk (2019a) separates the forward curve into systematic components that describe individual movements on the forward curve. These components are defined as the columns of a component matrix E. All components are thought of as constant over time, meaning that changes in the yield curve only comes from linear combinations of the columns in E. The forward curve f_t is thus described as

$$f_t = E\xi_t, \tag{3.23}$$

where ξ_t is known as the component scores. The scores are called risk factors if the decomposition is applied to yield curves or yield curve innovations. If the components are orthogonal, then the risk factors can be calculated as the projection

$$\xi_t = E^T f_t \tag{3.24}$$

regardless of the number of components used.

If the component matrix E is not orthogonal while also not containing every component (as is the case in dimension reduction), this way of calculating the risk factors does not work. Instead, minimising the squared error from the component representation gives an estimate of ξ_t. The minimisation problem is

$$\xi_t = \arg\min_{\xi_t} ||E\xi_t - f_t||_2. \tag{3.25}$$

This is known as the least-squares problem and has the solution

$$\xi_t = (E^T E)^{-1} E^T f_t. \tag{3.26}$$

Blomvall and Hagenbjörk (2019a) also stated that the yield curve contains noise and that systematic changes to the yield curve can be found as a matrix $\tilde{E}$ of reduced dimension. This means that problem of finding the component representation E can be modelled as a source separation problem. In the article, principal component analysis (PCA) was applied to the yield curve innovations $\Delta f_t = f_t - f_{t-1}$, but there is no guarantee that this is the most effective method of systematic risk extraction.

3.4.1 Source Separation

A common way of reducing the dimension of signals is by finding source signals. Assume that observed signals $(x_1(t), x_2(t), \ldots, x_n(t))$ at time t are a linear combination of source signals $(s_1(t), s_2(t), \ldots, s_n(t))$ described by

$$\begin{pmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_n(t) \end{pmatrix} = B \begin{pmatrix} s_1(t) \\ s_2(t) \\ \vdots \\ s_n(t) \end{pmatrix}. \tag{3.27}$$

where $B \in \mathbb{R}^{n \times n}$ is unknown. An alternative formulation of the problem is

$$\begin{pmatrix} s_1(t) \\ s_2(t) \\ \vdots \\ s_n(t) \end{pmatrix} = E \begin{pmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_n(t) \end{pmatrix}. \tag{3.28}$$

With the simple relation $E = B^{-1}$. If $x_i(t)$, $i = 1, 2, \ldots, n$ are random, then $s_i(t)$, $i = 1, 2, \ldots, n$ are also random. The matrix E contains the components associated with each source $s_i(t)$. It is common for source separation models to analyse properties of the source signals $s_i(t)$ to find a suitable value of E. This usually much easier than attempting to find B and taking its inverse, which could also introduce numerical problems. Principal component analysis (PCA) and independent component analysis (ICA) are examples of source separation models.

3.4.2 Principal Component Analysis

Principal Component Analysis (PCA) is a separation technique commonly used to reduce the dimensionality of data. Given a dataset of dimension $\mathbb{R}^N$ this method finds a low dimensional subspace of $\mathbb{R}^m$, where m is usually significantly lower than N. The subspace is created by minimising the mean squared error of the data after projection onto the principal components. For the case of PCA, this is equivalent to maximising the variance of the projected data. The principal components in regular PCA are the eigenvectors of the covariance matrix of the multivariate distribution of the data. The eigenvalues describe the amount of variance that each component is able to explain in the data,

$$v_m = \frac{\sum_{i=1}^{m} \lambda_i}{\sum_{i=1}^{N} \lambda_i}. \tag{3.29}$$

PCA does not make any assumption about the distribution of the components, but it does assume that the covariance matrix is constant.

3.4.3 PCA as an Optimisation Problem

In an article by Golub (2006), the problem of finding eigenvectors of a symmetric matrix is described by two continuous optimisation problems. The first problem is the extreme eigenvalue problem, which is the problem of finding an eigenvalue λ_1 and the corresponding eigenvector x_1 and the interior eigenvalue problem. The second problem is the interior eigenvalue problem, which is the problem of finding an eigenvalue λ_i on a predefined interval $[a, b]$ and the corresponding eigenvector x_i.

The covariance matrix is a symmetric matrix, and the eigenvectors are the components sought after in PCA. Using these optimisation problem formulations applied to a covariance matrix, a regular PCA can be extended with a regularisation to promote certain component properties.

In the article, Golub and Liao suggests the following formulation of the extreme eigenvalue problem, where the goal is to find the eigenvectors of A:

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & x^T A x - c x^T x \\ \text{s.t} \quad & x^T x \leq 1. \end{aligned} \tag{3.30}$$

The constant c must satisfy the condition

$$c \geq \lambda_n + 1. \tag{3.31}$$

If the largest eigenvalue is unknown, c can be set to

$$c = \|A\|_1 \tag{3.32}$$

It has been proven that a local optimum of (3.30) is also a global optimum. The global optimum of this optimisation problem is the eigenvector corresponding to the smallest eigenvalue of the matrix A.

3.4.4 Independent Component Analysis

Independent Component Analysis (ICA) is a signal separation technique. Similar to PCA, it finds linear combinations of the random variables $x_i(t)$ expressed as component variables or source signals $s_i(t)$. ICA attempts to maximise independence between the components variables, but independence of random variables is impossible to measure in practice. Instead, non-Gaussianity is used as a proxy for independence.

The central limit theorem states that linear combinations of independent variables are Gaussian in limit. This implies that if a component were to contain multiple independent source signals $s_i(t)$, it would be more Gaussian than an isolated independent source signal. Hyvärinen, Karhunen and Oja (2001) suggests using kurtosis or negentropy to find non-Gaussian components.

Excess kurtosis of a random variable Y with zero mean is

$$\text{kurt}[Y] = E[Y^4] - 3(E[Y^2])^2. \tag{3.33}$$

Kurtosis is very easy to estimate for a data set and is thus a simple way to find independent components. The measure is however not robust, because outliers of the data have a significant effect on the measure. If the data set is suspected to contain outliers, other measures of non-Gaussianity are of interest.

A more robust measure of non-Gaussianity is entropy. Entropy is a basic concept in information theory that measures the average information level of a discrete random variable. Differential entropy is the equivalent measure for continuous random variables. The differential entropy H of a continuous random variable y is defined as

$$H(y) = -\int p_y(z)\log(z)dz \tag{3.34}$$

Gaussian variables have the largest entropy of all random variables of equal variance. By defining negentropy as the difference of (differential) entropy between a Gaussian variable y_{gauss} and the random variable of interest y:

$$J(y) = H(y_{\text{gauss}}) - H(y), \tag{3.35}$$

Maximally non-Gaussian components can be found by maximising $J(y)$. A problem with using entropy is that it requires knowledge of the full probability distribution of the components. Thankfully, there are ways of approximating negentropy. The model described above is a basic ICA model that assumes that the variances of the independent component variables are constant. It is also important to note that entropy can only be approximately measured in practice.

3.4.5 FastICA Algorithm using Negentropy and Symmetric Orthogonalisation

Hyvärinen, Karhunen and Oja (2001) observed that the negentropy is approximately proportional to the squared difference between the expected value of $G(Y)$ and $G(X)$, where Y is a potential source signal and X is a Gaussian random variable and G is (almost) any non-quadratic function:

$$J(Y) \propto \left(E[G(Y)] - E[G(X)]\right)^2 \tag{3.36}$$

For approximating negentropy, G can be chosen as either

$$G_1(Y) = \frac{1}{\alpha_1}\log\cosh\left(\alpha_1 Y\right), \qquad 1 \leq \alpha_1 \leq 2 \tag{3.37}$$

$$G_2(Y) = -\exp\left\{-\frac{Y^2}{2}\right\}. \tag{3.38}$$

Both $G_1(Y)$ and $G_2(Y)$ are univariate functions that yield a simple, fast and robust estimator for the negentropy $J(Y)$. α_1 is a parameter often set to 1, according to Hyvärinen, Karhunen and Oja (ibid.).

The authors mentioned above have also proposed a fast fixed-point algorithm to estimate a single independent component, by maximisation of negentropy. The first step is to perform a PCA on the measured signal data x to find a representation where each variable is uncorrelated. These variables are further scaled to unit length. This transformed data z has the covariance matrix $C = I$. The algorithm is described as

1. Center the data x to make its mean zero.

2. Perform PCA and scale the data to give z, such that $C_z = I$.
3. Choose an initial (e.g., random) vector w of unit norm.
4. Let $w \leftarrow E[zg(w^T z)] - E[g'(w^T z)]w$.
5. Let $w \leftarrow \frac{w}{||w||}$.
6. If not converged, go back to step 4.

where $g(y)$ is the derivative of the function $G(y)$ in (3.36), chosen as either

$$g_1(y) = \tanh(\alpha_1 y) \qquad 1 \leq \alpha_1 \leq 2 \tag{3.39}$$

$$g_2(y) = y \exp\{-\frac{y^2}{2}\}, \tag{3.40}$$

with the derivatives

$$g_1'(y) = \alpha_1 \left(1 - \tanh^2(\alpha_1 y)\right), \qquad 1 \leq \alpha_1 \leq 2 \tag{3.41}$$

$$g_2'(y) = \left(1 - y^2\right) \exp\{-\frac{y^2}{2}\}. \tag{3.42}$$

The convergence of a solution $w^{(k)}$ from iteration k is measured as the dot product of the solution and the solution from a previous iteration $k-1$:

$$||w^{(k)} \cdot w^{(k-1)}| - 1| < \epsilon \tag{3.43}$$

where ϵ is the tolerance level for a single component extraction. Note that independent component analysis can't determine the sign of w, which is why the absolute value of the dot product is used in (3.43).

Hyvärinen, Karhunen and Oja (2001) describe an algorithm for estimating multiple components with symmetric orthogonalisation as the following:

1. Center the data x to make its mean zero.
2. Perform PCA and scale the data to give z, such that $C_z = I$.
3. Choose m, the number of independent components to estimate.
4. Choose initial values for each component w_i, $i = 1, \ldots, m$, each of unit norm. Orthogonalise the matrix W as in step 6 below.
5. For every $i = 1, \ldots, m$, let $w_i \leftarrow E[zg(w_i^T z)] - E[g'(w_i^T z)]w_i$, where g is defined as in (3.39) or (3.40).
6. Do a symmetric orthogonalisation of the matrix $W = (w_1, \ldots, w_m)^T$ by

$$W \leftarrow (WW^T)^{-\frac{1}{2}} W \tag{3.44}$$

7. If not converged, go back to step 5.

When the matrix W is no longer substantially changing, the algorithm has converged. Assuming that the algorithm has performed k iterations, the convergence is measured as the sum of the inner products of each component $w_i^{(k)}$ the component of the same index from the previous iteration, $w_i^{(k-1)}$. Using ϵ as tolerance, the algorithm has converged if

$$\left| \sum_{i=1}^{m} \left| w_i^{(k)} \cdot w_i^{(k-1)} \right| - m \right| < \epsilon. \tag{3.45}$$

3.5 Determining the Number of Risk Factors

Estimating the number of significant components when using PCA is very difficult. In an article written by Camacho and Ferrer (2014), the research area of PCA has been divided into three main purposes: accurate data compression, interpretation of latent variables and distribution estimation of latent variables and residuals. We choose to focus on accurate data compression. For the purpose of accurate data compression, Jolliffe (2010) describes three distinct model categories for determining the number of components that a PCA should use: ad-hoc rules, statistical tests and computational criteria. The ad-hoc rules are the most commonly used methods, as they are easy to implement, but they rely on subjective assumptions. Examples of ad-hoc involve total explained variance, scree charts and the Kaiser criterion. There are attempts at estimating the number of components using statistical tests on the eigenvalues of the covariance matrix, but there is no statistical test that has been shown to yield satisfying results (to the authors' knowledge). Computational criteria such as cross-validation are commonly used for machine learning successfully, but it is unfortunately computationally heavy.

3.5.1 Total Variance Explained

In PCA, the components are eigenvectors of the covariance or correlation matrix. The variance of the components are exactly their corresponding eigenvalue λ_i which means that the proportion of variance v_m explained by m components for a data set with N variables is described by

$$v_m = \frac{\sum_{i=1}^{m} \lambda_i}{\sum_{i=1}^{N} \lambda_i}. \tag{3.46}$$

By setting a threshold c for how much variance of the data must be explained, the dimension of the data is reduced to the lowest value of m that satisfies

$$v_m \geq c, \quad 0 < c \leq 1. \tag{3.47}$$

There is no good way of determining the optimal value of c. Jolliffe (ibid.) suggests that values between 0.7 and 0.9 are common but iterates that exact levels depend on the specific data set.

3.5.2 Scree Graph

Another simple method of determining the number of significant components is by plotting the eigenvalues λ_m against the number of components m. This type of graph is called a scree graph. The method is described in Cattell (1966) and is highly subjective in its methodology. The scree

test was originally created for determining the number of true factors for factor analysis models, but it has been adopted for component analysis models such as PCA as well. It is a graphical method that relies on the user looking for a shape in the eigenvalues that resembles a flat line. The first eigenvalue of this flat line is the last eigenvalue that should be included. Depending on the data set, some scree graphs could give the researcher difficulty finding a distinct flat line.

3.5.3 Kaiser Criterion

The last simple ad-hoc method is described by Kaiser (1960). Kaiser's empirical study resulted in a suggestion that factor analysis models should use the same number of factors as there are eigenvalues above one in the sample correlation. Thus, performing a PCA on the correlation matrix of the data, the Kaiser criterion is

$$\lambda_i > 1, \qquad i = 1, 2, \ldots, m. \tag{3.48}$$

This condition resulted in a number of factors that is consistent with many factor-analytic studies in psychology and fulfils algebraic lower bounds for a number of factors used in a factor analysis model. Unfortunately, little emphasis is put on the statistical significance of the result.

3.5.4 Cross-validation

In machine learning, many models have numerous variables that can lead to overfitting. To find the best balance between the number of parameters and model accuracy, cross-validation has been a very successful tool.

Let θ be the model parameter that requires estimation. Using a goal function $f(\theta)$, usually sum of squared prediction errors. Jolliffe (2010) describes the process as dividing the in-sample data set into k subgroups. For each subgroup, the goal function value is measured after it has been calibrated on the other $k-1$ subgroups. This is repeated for a range of values of θ. The optimal θ is found by comparing the function values for the different parameter values and choosing the value of θ that gives the optimal function value.

3.6 Rotations

Even if some data is approximated well by few components, the components may still be difficult to model. By rotating the components, the same data can be explained while the rotated components are also more easily understood. It could also improve the accuracy of simulations if the component scores of the rotated loadings have an improved fit to a known distribution.

Rotations are separated into orthogonal and oblique rotations. The former makes the resulting basis orthogonal while the latter does not. In an article by Jennrich (2002), a general method for oblique rotations is described. In the article, an oblique rotation criterion $Q(\Lambda)$ is minimised with a gradient projection algorithm. Λ is a (rotated) loading matrix where each loading has unit length. The oblique rotation is described as

$$\Lambda(R) = A(R^T)^{-1} \tag{3.49}$$

in which R is an arbitrary non-singular matrix with columns of unit length and A is the initial loading matrix before rotation. The optimisation problem described as a function of the matrix R is

$$\underset{T \in \mathcal{M}}{\arg\min}\, f(R) = Q(\Lambda(R)). \tag{3.50}$$

With $\mathcal{M}$ being the manifold (set) of all $k \times m$ matrices with columns of unit length, Jennrich (2002) also shows that a projection of any matrix $X \in \mathbb{R}^{k \times m}$ with nonzero columns has the following projection onto $\mathcal{M}$

$$\rho(X) = X \left(\mathrm{dg}\left(X^T X\right)\right)^{-\frac{1}{2}} \tag{3.51}$$

where $\mathrm{dg}(X^T X)$ is a matrix containing only the diagonal elements of $X^T X$. Jennrich further proposes an algorithm for solving (3.50) numerically as a gradient projection algorithm:

1. Choose step length $\alpha > 0$ and an initial rotation matrix $R \in \mathcal{M}$.
2. Compute the gradient $G = \frac{\partial f(R)}{\partial R}$.
3. Replace R by $\rho(R - \alpha G)$.
4. If $s > ||G - R \cdot \mathrm{diag}(R^T G)||$, stop. Otherwise, go to Step 2.

The algorithm is a simple gradient descent, which means the algorithm is only capable of finding local optima. The step length is usually set as $\alpha = 1$ and the stopping criterion s to 10^{-6}. Assuming that algorithm has iterated k times, the convergence is measured as the entrywise norm of the difference in the R matrix, $|| R^{(k)} - R^{(k-1)} ||_2$. G is the gradient of the function f with respect to the matrix R, defined as

$$G = \begin{pmatrix} \frac{\partial f}{\partial R_{11}} & \frac{\partial f}{\partial R_{12}} & \cdots & \frac{\partial f}{\partial R_{1m}} \\ \frac{\partial f}{\partial R_{21}} & \frac{\partial f}{\partial R_{22}} & \cdots & \frac{\partial f}{\partial R_{2m}} \\ \vdots & \vdots & & \vdots \\ \frac{\partial f}{\partial R_{k1}} & \frac{\partial f}{\partial R_{k2}} & \cdots & \frac{\partial f}{\partial R_{km}} \end{pmatrix}. \tag{3.52}$$

Jennrich (ibid.) also defines the oblique rotation criterion for the oblimin family of rotations including quartimin, bi-quartimin and covarimin using a parameter $\gamma = 0, \frac{1}{2}$ and 1, respectively. The rotation criteria for the oblimin family are

$$Q(\Lambda) = \frac{1}{4} \left\langle \Lambda^2 \,,\, (I - \gamma C)\Lambda^2 N \right\rangle_F, \tag{3.53}$$

where Λ^2 is the matrix Λ squared element-wise. N is a square matrix with zeros on the diagonal and ones elsewhere. C is a square matrix with each element equal to $\frac{1}{p}$, where p is the number of loadings. The Frobenius norm of two matrices $A \in \mathcal{R}^{m \times n}$ and $B \in \mathcal{R}^{m \times n}$ of equal size is defined as

$$\langle A \,,\, B \rangle_F = \sum_{i=1}^{m} \sum_{j=1}^{n} A_{ij} B_{ij}. \tag{3.54}$$

The number of local optima for the optimisation problem (3.50) depend on the used rotation criterion. Therefore, special care must be taken when choosing this criterion to ensure that the problem does not become too difficult to solve.

3.7 Risk Factor Modelling

Modelling of risk factors is discussed in this section. Several distributions are presented with the purpose of being used in the hedging model to simulate possible value changes in the swap book portfolio. A multivariate distribution for the risk factors can be modelled using a copula in conjunction with marginal distributions of the individual risk factors.

3.7.1 Copula

A copula is a mathematical tool that can be used to capture the dependency between different risk factors. To correctly model the co-dependence of the risk factors, marginal distributions were estimated for each risk factor and a copula was fitted with MLE as described by Blomvall and Hagenbjörk (2019b). This can be applied to all studied risk factors.

Sklar (1959) describes how a multivariate distribution can be separated in a copula och univariate distributions. The theorem states that for continuous marginal distributions, $F_i(x_i)$, $i = 1, 2, ..., n$ for values

$$x_i = F_i^{-1}(u_i) \tag{3.55}$$

where u_i is a uniformly distributed random number on [0,1], a copula C, defined as

$$C(F_1(x_1), F_2(x_2), ..., F_n(x_n)) \tag{3.56}$$

is a multivariate distribution with marginal distributions $F_i(x_i)$.

The theorem also holds in the opposite direction. If F is a multivariate distribution with the marginal distributions $F_i(x_i)$ there exists a copula C for which

$$F(x_1, x_2, ..., x_n) = C(F_1(x_1), F_2(x_2), ..., F_n(x_n)). \tag{3.57}$$

There exists several copulas that could be used to model the risk factors. Only the Gaussian and Student's t-copula are presented here as these are typically used in financial applications. They are also presented by Blomvall and Hagenbjörk (2019b) who found them useful in the application of risk factor modelling on interest rate markets.

The Gaussian copula as presented by Cherubini et al. (2012) is introduced as

$$C_N(u_1, u_2, ..., u_n) = \Phi_n(\Phi^{-1}(u_1), \Phi^{-1}(u_2), ..., \Phi^{-1}(u_n); \boldsymbol{\rho}) \tag{3.58}$$

where Φ denotes the univariate standard normal distribution (cdf) and $\boldsymbol{\rho}$ is the n-dimensional correlation matrix for the values $\Phi^{-1}(u_i)$ for $i = 1, 2, ..., n$.

They also introduce the Student's t-copula in a similar manner. They express it as

$$C_{stud}(u_1, u_2, ..., u_n) = T_n(T_\nu^{-1}(u_1), T_\nu^{-1}(u_2), ..., T_\nu^{-1}(u_n); \boldsymbol{\rho}, \nu) \tag{3.59}$$

where T_ν denotes the univariate Student's t distribution with ν degrees of freedom and $\boldsymbol{\rho}$ is the corresponding correlation matrix for $T_\nu^{-1}(u_i)$.

3.7.2 Maximum Likelihood Estimation

Maximum likelihood estimation is presented by Blom et al. (2005) and can be described as a method to estimate unknown parameters for a certain distribution. For a pdf, f with parameters denoted θ and the sample values x_i, $i = 1, 2, ..., n$ (assuming the samples independent) of the random variables $X_1, X_2, ..., X_n$, maximum likelihood is the largest product of the likelihood value of each observation,

$$\max_{\theta} \prod_{i=1}^{n} f(x_i). \tag{3.60}$$

This means maximising the probability of observing the series x_i if assumed that these samples are generated from the pdf, f. The logarithm of equation (3.60) is often presented instead and as the logarithmic function is injective and strictly growing it results in the equivalent optimal parameters, θ. The resulting expression yields an expression based on a sum, which is often easier to interpret and also numerically stable,

$$\max_{\theta} \sum_{i=1}^{n} \ln(f(x_i)). \tag{3.61}$$

3.8 Monte Carlo Simulation

Wang. (2012) describes Monte Carlo simulation as determining an approximation of the expected value of a function. Consider a random variable X. The expected value of a function $h(X)$ is

$$\mu = E[h(X)]. \tag{3.62}$$

A Monte Carlo simulation means generating independent and identically distributed (i.i.d) samples $x_1, x_2, ..., x_n$ of X. An estimate of the expected value can be calculated as

$$\hat{\mu} = \frac{1}{n}[h(x_1) + h(x_2) + ... + h(x_n)] \tag{3.63}$$

where n is the sample size. The value estimate of the expected value, $\hat{\mu}$ is a random variable whose value depends on the values of the samples.

The basis for Monte Carlo simulation is the strong law of large numbers which states that

$$\hat{\mu} = \frac{1}{n}[h(X_1) + h(X_2) + ... + h(X_n)] \to E[h(X)] = \mu \text{ when } n \to \infty. \tag{3.64}$$

Monte Carlo simulation is a powerful tool in complex quantitative models. However it also has shortcomings. For example, the convergence rate towards the real expected value is rather slow and requires many samples. Monte Carlo simulation is a central part of the hedging model presented in section 3.9.1.

The variance in the estimate results in uncertainty and inaccurate results. There are ways to reduce the variance of Monte Carlo simulations that reduce the number of samples needed to achieve a good estimate of the expected value. These methods are called variance reducing methods.

A popular method is Latin hypercube sampling, which Huynh, Lai and Soumare (2011) present as a variance reduction method which is based around dividing samples in sections. Instead of generating uniformly distributed random variables on the interval $u_i \in [0, 1]$, the random variables are generated on K subintervals of [0,1] instead. This ensures an even spread in the random variables and thus eliminates the risk of getting a sample with clustering. The observed random variables are then inverted using the inversion principle to generate a sample from the desired distribution. The advantage of using Latin hypercube sampling is that a lower number of samples are required for convergence of the estimated value $\hat{\mu}$. Latin hypercube sampling is a very simple variance reduction technique due to not demanding any advanced mathematics or implementation. It is therefore a suitable approach for this study.

3.9 Hedge Model

In this section, the theoretical framework of swap book hedging is presented. This is a central part in the thesis and almost all previous theory is combined and used in this section.

3.9.1 The Stochastic Programming Model

Blomvall and Hagenbjörk (2020) have produced a framework for hedging interest rate risk based on stochastic programming. They formulate a two-stage stochastic programming model which considers both portfolio variance and transaction costs. The foundation for the framework is accurate measurements of single yield curves and their corresponding risk factors. With accurate modelling of the risk factors, they can be simulated in the hedge to make precise evaluations of the perceived transaction costs.

The stochastic programming problem is formulated as

$$\begin{aligned} \min_{\Delta x_P, \Delta x_R} \quad & \frac{1}{S-1}\sum_{i=1}^{S}(W_i - \frac{1}{S}\sum_{j=1}^{S} W_j)^2 + \alpha(c_P^T \Delta x_P + c_R^T \Delta x_R) \\ \text{subject to} \quad & W_i = h_i + V_{P,i}^T \Delta x_P + V_{R,i}^T \Delta x_R,\ i = 1, ..., S \\ & \Delta x_P, \Delta x_R \geq 0. \end{aligned} \tag{3.65}$$

The notations used are

- Δx_P and Δx_R describes the trading decisions determined by the hedge, in terms of the notional amount.
- S are the number of simulated risk factor scenarios.
- W_i is the value of the hedged portfolio in scenario i.
- α is a penalty parameter that punishes decisions that lead to transaction costs. Larger values of α make the hedge less likely to rebalance.
- c_P and c_R are the transaction costs that appear when entering the contracts paying and receiving fixed rate, respectively.
- h_i is the value of the swap book before hedging.

- $V_{P,i}$ and $V_{R,i}$ are the NPVs of the interest paying and receiving fixed rate, respectively.

The transaction costs depend on the term structure estimation model, because the transaction costs on swap contracts appear indirectly when crossing the bid-ask spread. If the term structure model does not reprice the instruments used to generate the curve exactly, a skew in transaction costs will exist for both directions of the contract (both bid and ask). The perceived transaction costs (which are seen by the hedge model) are determined by valuing the theoretical NPV based on the curves generated from a term structure model. The transaction costs that occur when entering a swap contract are

$$c_{t,i,P} = \max\{0, -P_{i,t}(f_t, r_{fix,ask})\} \tag{3.66a}$$
$$c_{t,i,R} = \max\{0, -P_{i,t}(f_t, r_{fix,bid})\}. \tag{3.66b}$$

Here $c_{t,i,P}$ and $c_{t,i,R}$ denotes perceived transaction costs for paying and receiving the fixed rate respectively. i denotes what instrument is studied and t at what time the transaction costs are calculated. $P_{i,t}(f_t, r_{fix,ask})$ denotes the NPV (price) for this instrument i at time t given the yield curve f_t as time t and using the ask yield $r_{fix,ask}$ in the valuation. $P_{i,t}(f_t, r_{fix,bid})$ is the same expression except that instead the bid yield is used in the valuation.

In the simulation made by Blomvall and Hagenbjörk (2020), they simulated the hedge using 5000 scenarios of daily discretised term structures. They used alternating α values when solving the optimisation problem and present several charts with different results depending on the chosen α parameter. The α values they tested span between 10^{-4} and 10^2. The simulations in the hedge model are based on a model for simulating risk factor innovations $\Delta\xi$ between succeeding days.

In their paper, they used only six risk factors to generate scenarios for their single forward rate curves. These factors were retrieved using PCA on innovations in forward rate curves in their in-sample period, spanning from 1996-2001. Then these components were used and kept constant in their out-of-sample period 2002-2018 to create scenarios for the forward rate curves the succeeding days. The scenario generating method used by them were Monte Carlo with antithetic sampling of uniformly distributed random numbers.

When generating the simulated curves, noise was added in order to avoid linear dependence between the six risk factors which otherwise would be exploited by the hedging model.

$$v_{\text{unexp}} = \text{diag}(C) - \text{diag}(E^T \lambda E) \tag{3.67}$$

The variance is denoted $\text{diag}(v_{\text{unexp}})$ with C as the covariance matrix for daily innovations in the forward rate curves. $E^T \lambda E$ is the approximated covariance matrix using only six risk factors, based on E presented in (3.24). Samples from the noise distribution are drawn from the normal distribution $F_N^{-1}(u_i, 0, \sqrt{\text{diag}(v_{\text{unexp}})})$ using the inversion principle.

However, in the single yield curve setting presented in the paper, this proved insufficient and additional noise was added in order to achieve a more stable hedge over time. Therefore a constant variance term v is added to shift the variance to a higher level which results in the noise distribution $F_N^{-1}(u_i, 0, \sqrt{\text{diag}(v_{\text{unexp}} + v)})$. The noise level must be carefully determined because too much noise has the opposite effect and makes the hedge unstable. In the paper by Blomvall and Hagenbjörk (ibid.), they found that $v = 0.01\%$ works good and results in the lowest portfolio volatility in their in-sample period. They also scaled the amount of noise by the fraction between the current variance

level (from the variance model used) and the in-sample term structure variance, as

$$\frac{\sum_{i=1}^{k} \sigma_{i,t}^2}{\sum_{i=1}^{k} \lambda_i}. \tag{3.68}$$

The volatility model used in the paper is a variant of a GARCH-model.

3.9.2 Performance Attribution

Let $\tilde{\xi}_t$ be a vector of M financial risk factors at time t and assume that the theoretical price of a financial asset $P_{t,i}(\tilde{\xi}_t)$ only depends on the risk factors $\tilde{\xi}_t$. A change in the risk factors leads to the following change in the theoretical price of the asset:

$$\Delta P_{t,i}(\Delta\tilde{\xi}_t) \equiv P_{t,i}(\tilde{\xi}_{t-1} + \Delta\tilde{\xi}_t) + D_{t,i} - P_{t-1,i}(\tilde{\xi}_{t-1}). \tag{3.69}$$

$D_{t,i}$ is the dividend or cash flow from the financial asset i paid at time t. By discarding insignificant risk factors and keeping the significant risk factors ξ_t, as discussed in 3.4, the dimensionality of the price change is reduced. The dimension reduction also introduces an error $\epsilon_{t,i}^I$,

$$\epsilon_{t,i}^I \equiv P_{t,i}(\tilde{\xi}_t) - P_{t,i}(\tilde{\xi}_{t-1} + [\Delta\xi_t \ ; \ 0]). \tag{3.70}$$

The rollover effect on the asset price that occurs after a time Δt has passed is defined as

$$\theta_{t-1,i}\Delta t \equiv P_{t,i}(\tilde{\xi}_{t-1}) + D_{t,i} - P_{t-1,i}(\tilde{\xi}_{t-1}), \tag{3.71}$$

which describes the effect of the asset price sensitivity to time. Small changes in the theoretical price are approximated with a second-order taylor approximation when the rollover effect is included, described by

$$\Delta P_{t,i}^A(\Delta\xi_t) \equiv \theta_{t-1,i}\Delta t + g_{t-1,i}^T\Delta\xi_t + \frac{1}{2}\Delta\xi_t^T H_{t-1,i}\Delta\xi_t. \tag{3.72}$$

The Taylor approximation introduces an error $\epsilon_{t,i}^A$ that comes from omitting higher-order derivatives as well as the time not being infinitesimal in a discretised yield curve (which is assumed by using the gradient and hessian of the theoretical price). The error is calculated as

$$\Delta\epsilon_{t,i}^A = P_{t,i}(\tilde{\xi}_{t-1} + [\Delta\xi_t \ ; \ 0]) - P_{t,i}(\tilde{\xi}_{t-1}) - \Delta P_{t,i}^A(\Delta\xi_t). \tag{3.73}$$

The market price does not match with the theoretical price if the bid or ask prices are used. The theoretical pricing error $\Delta\epsilon_{t,i}^P$ is

$$\epsilon_{t,i}^P = \bar{P}_{t,i} - P_{t,i}(\tilde{\xi}_t) \tag{3.74}$$

and

$$\Delta\epsilon_{t,i}^P = \epsilon_{t,i}^P - \epsilon_{t-1,i}^P. \tag{3.75}$$

The approximated change in the theoretical market price in combination with the error terms defined above creates a residual free model for price changes in financial assets:

$$\Delta\bar{P}_{t,i} = \theta_{t-1,i}\Delta t + g_{t-1,i}^T\Delta\xi_t + \frac{1}{2}\Delta\xi_t^T H_{t,i}\Delta\xi_t + \epsilon_{t,i}^I + \epsilon_{t,i}^A + \Delta\epsilon_{t,i}^P. \tag{3.76}$$

The performance attribution model above can be further generalised to a model for portfolios consisting of international financial assets. Consider the set of different financial assets, $\mathcal{I}$, that is included

in a portfolio of assets. Let $\mathcal{E}$ be the set of exchange rates and $e(i)$ the index of an exchange rate associated with asset i and its exchange rate by $f_{t,e(i)}$ (where the term currency is the currency in which the portfolio performance is being measured in). The number of each asset in the portfolio is denoted $h_{t,i}$ and holdings in each currency at time t are denoted $h_{t,e}$ The portfolio value in local currency is described as

$$V_t \equiv \sum_{i \in \mathcal{I}} h_{t,i} \bar{P}_{t,i} + \sum_{e \in \mathcal{E}} h_{t,e} f_{t,e}. \tag{3.77}$$

The change in the quoted price (in the quoted currency) is

$$\Delta \bar{P}_{t,i} \equiv \bar{P}_{t,i} + D_{t,i} - \bar{P}_{t-1,i}, \tag{3.78}$$

and the change of the price described in local currency is

$$\Delta P^L_{t,i} = \Delta \bar{P}_{t,i} f_{t-1,e(i)} + \bar{P}_{t-1,i} \Delta f_{t,e(i)} + \Delta \bar{P}_{t,i} \Delta f_{t,e(i)}, \tag{3.79}$$

where

$$\Delta f_{t,e(i)} = f_{t,e(i)} - f_{t-1,e(i)}. \tag{3.80}$$

The change in portfolio value without adjusting for currency holdings is calculated as

$$\Delta V^h_t = \sum_{i \in \mathcal{I}} h_{t-1,i} \Delta P^L_{t,i}. \tag{3.81}$$

Changes in holdings of asset i at time t depend on buy transactions, $\Delta h^B_{t,i}$, and the sell transactions $\Delta h^S_{t,i}$ of each asset i, according to

$$h_{t,i} \equiv h_{t-1,i} + \Delta h^B_{t,i} - \Delta h^S_{t,i}. \tag{3.82}$$

The transaction sizes are required to calculate the effects of crossing the bid-ask spread. The holdings of a currency depend on the previous (simple) interest rate $r_{e,t-1}$ and the corresponding time period $\Delta t_{e,t-1}$, measured with the appropriate day count convention. The currency holdings are also affected by buying and selling assets, including the spreads $s^B_{t,i}$ and $s^S_{t,i}$ respectively. Finally, the currency holding depends on currency transactions, where

$$\begin{aligned} h_{t,e} = {} & h_{t-1,e}(1 + r_{e,t-1} \Delta t_{e,t-1}) \\ & + \sum_{i:e(i)=e} h_{t-1,i} D_{t,i} \\ & + \sum_{i:e(i)=e} \left(\bar{P}_{t,i} - s^S_{t,i}\right) \Delta h^S_{t,i} \\ & - \sum_{i:e(i)=e} \left(\bar{P}_{t,i} + s^B_{t,i}\right) \Delta h^B_{t,i} \\ & + \sum_{e_2 \in \mathcal{E}} \Delta h^B_{t,e_2,e} \\ & - \sum_{e_1 \in \mathcal{E}} \left(f_{t,e_1,e} + s^B_{t,e_1,e}\right) \Delta h^B_{t,e,e_1}. \end{aligned} \tag{3.83}$$

A change in portfolio value can thus be fully described as

$$\Delta V_t = \Delta V^h_t + \sum_{e \in \mathcal{E}} \left(\Delta h_{t,e} f_{t-1,e} + h_{t-1,e} \Delta f_{t,e} + \Delta h_{t,e} \Delta f_{t,e}\right). \tag{3.84}$$

Chapter 4

Method

This section describes the implementation of the methodology used and the application of the theories presented. The meaning of this section is to connect the theories with the methodology and purpose of this thesis, which is to find a realistic risk factor model in order to make better decisions in the hedging model. Since this thesis is a collaboration with two other theses, some parts of the implementation are prepared by those groups. In these cases, the reader will be referred to the work of the authors who implemented the solution.

The chapter follows a similar arrangement as the one presented in section 2.2. To start with the contracts used to generate the interest rate curves are presented and then data processing is discussed. The data processing is necessary to eliminate unrealistic curves, which could affect the subsequent component estimation model. After these steps the risk factor decomposition is addressed and how the risk factors will be modelled. Then the portfolio construction will be discussed to replicate a typical swap book at Swedbank. Together with this section, the valuation of IRS and FRA contracts expressed in risk factors will be presented. Finally, the hedging model and evaluation methods will be discussed.

4.1 Data Collection

The first step to conduct this study is to gather information regarding the contracts used to estimate the term structures. This will be accomplished with Refinitiv Eikon, using closing prices for daily swap rates. These swap rates along with contract-specific data are used for the yield curve estimation.

4.1.1 Contracts to Generate Yield Curves

The contracts used to estimate the interest rate curves ($f_0(t)$ and $\pi_\tau(t)$) are presented in table 4.1. In the table, the tenor of the contract is presented along with the Refinitiv instrument codes (RIC). As only the standard tenor on the European market is studied according to section 1.3 the index τ on π can be dropped. Each risk-free forward rate curve $f_0(t)$ is created based on the OIS contracts in the second column. Each tenor specific forward curve $\pi(t)$ are estimated using the IRS and FRA contracts in the third and fourth column, respectively. The fixed rates of the contracts are gathered every day for the in-sample & out-of-sample period. The periods are presented in table 4.2. The length of the in-sample and out-of-sample periods are limited by the period of available data for OIS contracts because those were invented after the IRS and FRA contracts. This results in two interest

Table 4.1: RIC of contracts used to generate the risk-free and the 6M tenor specific forward rate curves on the European market.

Tenor	OIS	IRS	6M FRA
ON	EONIA=		
SW	EUREONSW=		
2W	EUREON2W=		
3W	EUREON3W=		
1M	EUREON1M=		EUR1X7F=
2M	EUREON2M=		EUR2X8F=
3M	EUREON3M=		EUR3X9F=
4M	EUREON4M=		EUR4X10F=
5M	EUREON5M=		EUR5X11F=
6M	EUREON6M=		EUR6X12F=
7M	EUREON7M=		
8M	EUREON8M=		
9M	EUREON9M=		EUR9X15F=
10M	EUREON10M=		
11M	EUREON11M=		
1Y	EUREON1Y=	EURAB6E1Y=	EUR12X18F=
15M	EUREON15M=		
18M	EUREON18M=	EURAB6E18M=	EUR18X24F=
21M	EUREON21M=		
2Y	EUREON2Y=	EURAB6E2Y=	
3Y	EUREON3Y=	EURAB6E3Y=	
4Y	EUREON4Y=	EURAB6E4Y=	
5Y	EUREON5Y=	EURAB6E5Y=	
6Y	EUREON6Y=	EURAB6E6Y=	
7Y	EUREON7Y=	EURAB6E7Y=	
8Y	EUREON8Y=	EURAB6E8Y=	
9Y	EUREON9Y=	EURAB6E9Y=	
10Y	EUREON10Y=	EURAB6E10Y=	

Table 4.2: In-sample and out-of-sample periods to generate yield curves and risk factors on.

Type of period	Dates
In-sample period	2005-08-15 to 2013-06-20
Out-of-sample period	2013-06-21 to 2021-05-11

rate curves for each day of the sample periods. All of the IRS and FRA contracts presented has 6M EURIBOR as the floating rate, which is the standard tenor for European interest rate derivatives. These floating rates are collected for each day as well, to ensure that the cash flows for the floating leg in both FRA and IRS contracts can be calculated after the floating rate has been observed. The

RIC code for the 6M floating EURIBOR rate is EURIBOR6MD=.

4.1.2 Swap Rate Data Clean Up

To fully ensure that the swap rate data does not contain unnecessary outliers, a cleanup process described by de Flon and Landgren (2021) is applied. This results in the exclusion of days where unrealistic swap rates have been observed, which is important because the other subsequent steps are greatly affected by outliers.

4.2 Yield Curve Estimation

In this section, the method that is used to estimate forward rate curves based on the collected contract data is presented. First, the bid and ask swap prices are averaged into a mid-price (swap rate). These mid rates along with the contract-specific data gathered in 4.1.1 are used in the optimisation problem (3.20) to generate forward yield curves. In order to achieve a realistic term structure, the smoothness and pricing error of the curve must be weighed carefully. An examination of different parameter values in the optimisation problem is performed by de Flon and Landgren (ibid.). The parameters are estimated in the in-sample period and tested in the out-of-sample period presented in table 4.2. The test is based on controlling the valuation using the forward rate curve generated against the actual quoted fix rates of the contracts. The goodness of the fit is also estimated using methods described by de Flon and Landgren (ibid.).

4.3 Risk Factor Decomposition of a Single Yield Curve

In this section the different risk factor decomposition models that have been found are discussed. The risk decomposition models separate observed yield curve innovations into observed risk factors that may be correlated. This decomposition method requires estimation of the number of risk factors and their distributions, which will differ depending on the type of yield curve analysed. This means that the described method in this section must be done for each type of yield curve individually.
Rather than using a component analysis model on the observed forward curves, component analysis is performed on innovations of observed forward curves. Innovations of a forward rate curve f_t is defined as

$$\Delta f_t = f_t - f_{t-1}. \tag{4.1}$$

Mathematically, simulating Δf_t is equivalent to simulating f_t and f_{t-1}. The only difference is that the statistical properties of the term structure innovations make them easier to model. The works of Litterman and Scheinkman (1991) and Blomvall and Hagenbjörk (2019a) have both successfully applied PCA on innovations in the term structure. For this reason, the proposed component analysis models in this thesis will be applied to the term structure innovations Δf_t as well. The subsequent sections assume that the yield curve data is in the form of daily yield curve innovations.

Six different component models are used in the study. PCA, regularised PCA towards a kernel, ICA with symmetric orthogonalisation, ICA using the deflation method, quartimin rotation of principal components and a rotation of principal components towards a kernel. The details of how the components are estimated are described in 3.

4.3.1 Rotations and Significant Number of Components

The hedge model described in 3.9 requires accurate simulations of risk factors, thus a component representation is sought after such that the corresponding risk factor has a statistical distribution that is easy to estimate. PCA yields components that are orthogonal and rotated in such a way that the variance of each risk factor is maximised, but that is not necessarily the most useful component description. The ICA model and the rotation framework in section 3.6 describe rotations of principal components, which could potentially improve statistical modelling of the risk factors. In this study, the rotations are performed on principal components to retain the highest possible variance. After dimension reduction, the rotated significant principal components are able to explain exactly the same amount of variability of the data. This is because a vector basis is invariant to rotations, which means that components may be rotated for the purpose of improving the subsequently estimated multivariate distribution of the risk factors.

There is no doubt that rotating every component from the initial dimension reducing PCA is useful, but there is no definitive way of determining what the rotated components should look like. In this thesis, the only required property of the components is that simulated innovations in the term structure should be as accurate as possible. But before the rotation is applied, the number of components must be determined due to the large dimension of the data. This is problematic, because different types of rotations may optimally require a different number of components. This makes cross-validation a suitable method for finding the number of components. Due to the limited time frame of this project, the number of components is limited to six (as used by Blomvall and Hagenbjörk (2020)) to decrease the workload.

4.4 Risk Factor Modelling

This section describes how the risk factors are modelled with a multivariate distribution through marginal distributions of each risk factor combined with a copula. This is an important step as the hedging model requires accurate simulations of future risk factors in order to make good hedging decisions. These simulations are made, based on samples drawn from the distributions described here.

4.4.1 Univariate Distribution

The marginal distributions of the risk factors are assumed to be Student's t-distributed with zero mean and time-varying volatility. Since volatility is commonly non-stationary in financial markets, heteroscedastic models such as GARCH will be used. For a detailed explanation of the volatility models that are examined, see de Flon and Landgren (2021). The parameters are determined using MLE, which is described in section 3.7.2, as MLE is a simple yet effective method to find parameter values for each individual risk factor model.

4.4.2 Copula

In this section an overview of the copula models used are presented along with how the parameters for the copula are estimated, for a thorough explanation see de Flon and Landgren (ibid.).

The copulas used are the Gaussian and Student's t copulas presented in section 3.7.1. The copula is, as stated earlier, important to model the co-dependence between risk factors. The copula

models are tested in the same manner as the univariate distributions (in section 4.4.1). The Student's t copula should represent the co-dependence better than a gaussian copula because the multivariate Student's t-distribution can account for fat tails but this is tested by de Flon and Landgren (2021) before deciding which copula to use in the hedge model. The copula is estimated using the maximum-likelihood estimator presented in section 3.7.2.

The copula (and the underlying marginal distributions) together with the risk factor decomposition and yield curve estimation are evaluated using the test described by de Flon and Landgren (ibid.). In this test, the simulated term structures are compared with actual swap rates of the contracts used in the curve generation observed on the financial market. This is a critical step in order to ensure that the risk factor modelling and all previous steps are sufficiently good to model the actual swap market. Otherwise, the simulations in the stochastic programming hedge model would not give good anticipation of future portfolio value and lead to faulty hedge decisions. For a further explanation of this test and details regarding the simulation, see de Flon and Landgren (ibid.).

4.5 Portfolio Construction

To simulate the trading in the portfolios, the distributions and the process defined in section 3.1.5 are used. Each portfolio can be thought of as two sub-portfolios (from the perspective of the bank): The swap book generated from customer positions and those entered through the hedge. The one which originates from customer positions and the hedge are in practice added together but here considered to be separated to give the theoretical reasoning behind the trading in the portfolio. Trading will be done once a day as well as hedging to mimic the situation used by Blomvall and Hagenbjörk (2020) as it is a realistic scenario that also yielded satisfying results for them. Another factor restricting how often it is possible to hedge is the run time of the algorithm. In order to avoid high demands on computational efficiency, hedging once a day seems sensible.

The customer portfolios are hedged using the stochastic programming model presented in 3.9.1. Each day, the time to maturity of the held contracts decrease. To handle this phenomenon an approach with "buckets" will be used where the buckets are the different maturities presented in column one in table 4.3. For example, a contract with tenor 2Y will after a day have a tenor of one year and 364 days. The weight of this contract will be divided by the two tenors closest to this one using linear interpolation, in this case, 2Y and 18M. The day count convention differs for the fixed and floating legs of the contracts. As a simple approximation, 30E/360 is assumed and thus the weights will be 1/180 on tenor 18M and 179/180 on tenor 2Y. The weights are also separated into FRA and IRS contracts.

In order to keep the tenor exposure of the swap book relatively constant and achieve a realistic portfolio over time, customer trading will be simulated. The trading will be made in the portfolio originating from the customer positions (as the hedge portfolio is determined by the bank and not random in that way). The skewness in paying or receiving fixed rate in the customer portfolio has also been discussed. The customers prefer to enter contracts where they pay the fixed rate and receive the floating rate with a ratio of about 3:1.

The nominal value for each contract entered by customers will be simulated by generating a number using a modified log-normal distribution, which can be derived from the normal distribution. The expression

$$e^{F_N^{-1}(u_i)} \tag{4.2}$$

returns samples from the log-normal distribution. By modifying it to

$$100 + \exp\{2 + F_N^{-1}(u_i, 3, 0.5)\} \tag{4.3}$$

the samples drawn are more representative of actual notional values on financial markets. They are floored and then also scaled by a factor of €10^6 in order to make them as realistic as possible. To get an understanding of what the generated notional values look like, a histogram of the unscaled generated samples are presented in figure 4.1. To determine if the customers enter the contract by

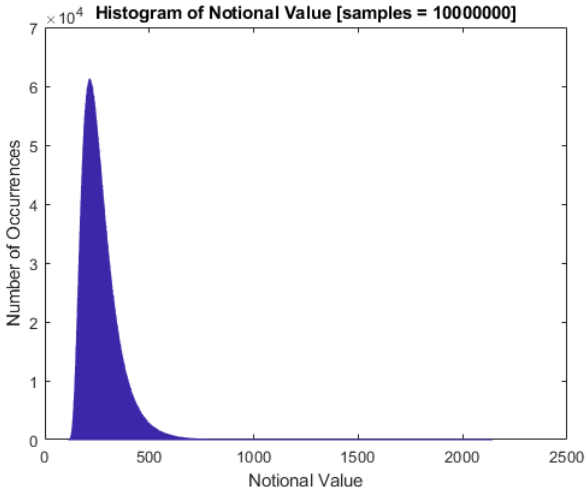


Figure 4.1: A histogram of the notional values generated by the modified log-normal distribution.

paying a fixed rate or floating rate (which has to be known to calculate the transaction costs in the hedge), the uniform distribution F_U is used. For each generated nominal value, another value u_j is also generated from the uniform distribution on the interval [0,1]. The following limits are used:

$$\begin{cases} \text{the customer receives fixed rate} & \text{if } u_j \leq 0.25 \\ \text{the customer pays fixed rate} & \text{if } u_j > 0.25. \end{cases} \tag{4.4}$$

This means that the customer portfolios will be skewed in a way such that the total nominal amount of the contracts of which the customers pay fixed rate is three times as large as the total nominal amount of contracts in the other direction.

A Poisson process is used to model how many customer trades are made each day. The swap book is treated as stationary when the hedging starts. This is the case that is most realistic for the bank because they are already in a state of stationarity at the start of the in-sample period. For the first ten years of the in-sample data, a stationarity state is impossible to simulate fully. This is because the contract with the longest maturity is an IRS with ten years to maturity and when time has passed such that the contract has transformed into an IRS with a time to maturity of a single day, the swap rate of the contract is almost 10 years old. The effect of using buckets is analytically calculated in this case, but instead of using old data, the current swap rate (first day in the out-of-sample period) is used for the initial portfolio.

To calculate the number of contracts in each bucket, taking the trickle down effect into account for all IRS contracts, assume that $T_{\text{mat}} = (T_{\text{mat,1}}, ..., T_{\text{mat,n}})$ is a vector containing the time to maturities for each contract at the creation date, in ascending order, and $\lambda = (\lambda_1, ..., \lambda_n)$ is a vector with the trading intensity of each maturity (average number of new contracts per day) , in the same contract order as T_{mat}. For IRS contracts, the number of contracts in the bucket with longest maturity (10Y) at stationarity is

$$\begin{aligned} N_n &= \sum_{t=1}^{T_n - T_{n-1}} \frac{t}{T_n - T_{n-1}} \lambda_n \\ &= \frac{1}{T_n - T_{n-1}} \cdot \frac{(T_n - T_{n-1})(T_n - T_{n-1} + 1)}{2} \lambda_n \\ &= \frac{T_n - T_{n-1} + 1}{2} \lambda_n. \end{aligned} \tag{4.5}$$

The shorter buckets are affected by other contracts with longer time to maturity trickling downwards as well. The number of contracts in bucket i, where $i = 1, 2, ..., n-1$, is

$$\begin{aligned} N_i = &\sum_{t=1}^{T_i - T_{i-1}} \frac{t}{T_i - T_{i-1}} (\lambda_i + \lambda_{i+1} + ... + \lambda_n) \\ &+ \sum_{t=1}^{T_{i+1} - T_i - 1} \frac{t}{T_{i+1} - T_i} (\lambda_{i+1} + \lambda_{i+2} + ... + \lambda_n) \\ &= \frac{T_i - T_{i-1} + 1}{2} (\lambda_i + \lambda_{i+1} + ... + \lambda_n) + \frac{T_i - T_{i-1} - 1}{2} (\lambda_{i+1} + \lambda_{i+2} + ... + \lambda_n) \\ &= \frac{T_i - T_{i-1} + 1}{2} \lambda_i + (T_i - T_{i-1}) (\lambda_{i+1} + \lambda_{i+2} + ... + \lambda_n) \end{aligned} \tag{4.6}$$

The intensity of the Poisson processes are predetermined using the λ values presented in table 4.3. These have been chosen to represent a scenario at Swedbank. The weights for longer maturities are lower as these accumulate overtime on to the lower maturities. Contracts with a shorter time to maturity are generally also traded more frequently.
The initial holdings of the swap book are determined with equation (4.5) and (4.6) for IRS and FRA contracts separately (the calculations are identical for FRA contracts). The number of contracts in each direction is set as their expected values, meaning 25% of the initial contracts pay fixed payments to the customers and 75% pay floating payments to the customers. The number of contracts in the initial FRA and IRS swap books, along with the percentage of the total notional value, is displayed in 4.4 and 4.5. All contracts are assumed to follow the 30E/360 day count convention, which introduces a small but negligible error in the holdings of the initial portfolio. Due to the interpolation of the time to maturity of contracts seen in (4.6) and (4.5), the difference in time to maturity between contracts has a great effect on the number of contracts in the initial portfolios.

Table 4.3: RIC and λ values of contracts traded in a portfolio similar to a typical swap book at Swedbank.

Tenor	IRS	6M FRA	λ values of contracts
1M		EUR1X7F=	2
2M		EUR2X8F=	2
3M		EUR3X9F=	2
4M		EUR4X10F=	2
5M		EUR5X11F=	2
6M		EUR6X12F=	2
9M		EUR9X15F=	2
12M		EUR12X18F=	2
18M		EUR18X24F=	2
1Y	EURAB6E1Y=		2
18M	EURAB6E18M=		2
2Y	EURAB6E2Y=		2
3Y	EURAB6E3Y=		2
4Y	EURAB6E4Y=		2
5Y	EURAB6E5Y=		10
6Y	EURAB6E6Y=		2
7Y	EURAB6E7Y=		2
8Y	EURAB6E8Y=		2
9Y	EURAB6E9Y=		2
10Y	EURAB6E10Y=		10

Table 4.4: Number of contracts held in the initial FRA swap book and contract weights.

Tenor	6M FRA	# Contracts	Weight
1M	EUR1X7F=	511	15.3%
2M	EUR2X8F=	451	13.5%
3M	EUR3X9F=	391	11.7%
4M	EUR4X10F=	331	9.9%
5M	EUR5X11F=	271	8.1%
6M	EUR6X12F=	391	11.7%
9M	EUR9X15F=	451	13.5%
12M	EUR12X18F=	361	10.8%
18M	EUR18X24F=	181	5.4%

4.6 Contract Valuation

In this section the valuations for the studied swap book consisting of IRS and FRA contracts are further described in a multiple yield curve setting. The valuations presented in the theoretical framework are expanded and expressed as changes of the risk factors from section 4.3. The gradients and hessians of the theoretical contract prices are also presented here, because these are later used in

Table 4.5: Number of contracts held in the initial IRS swap book and contract weights.

Tenor	IRS	# Contracts	Weight
1Y	EURAB6E1Y=	10081	13.1%
18M	EURAB6E18M=	6301	8.2%
2Y	EURAB6E2Y=	8821	11.5%
3Y	EURAB6E3Y=	11161	14.5%
4Y	EURAB6E4Y=	10441	13.6%
5Y	EURAB6E5Y=	8285	10.8%
6Y	EURAB6E6Y=	6121	7.9%
7Y	EURAB6E7Y=	5401	7.0%
8Y	EURAB6E8Y=	4681	6.1%
9Y	EURAB6E9Y=	3961	5.1%
10Y	EURAB6E10Y=	1805	2.3%

the performance attribution in section 4.8. In this study, only FRA and IRS contracts with a tenor of 6M are studied. The floating rate of those contracts is 6M EURIBOR. This means that the index τ corresponds only to the 6M tenor in this case.

The optimisation problem described in 3.3.1 requires a discrete representation of the forward rate curves. Using the notations in 3.3.1, the integral in equation 3.9 is approximated by integration of the discrete forward curves, using the integrating matrix A defined in equation (3.10), as

$$\begin{aligned}\psi_\tau(\hat{T}) &= \int_0^{\hat{T}} f_\tau(t)dt = \int_0^{\hat{T}} (f_0(t) + \pi_\tau(t))dt \\ &\approx r_{t,\tau}^{\hat{T}}\hat{T} = (r_{t,0}^{\hat{T}} + r_{t,\pi}^{\hat{T}})\hat{T}.\end{aligned} \tag{4.7}$$

where $r_{t,0} = Af_{t,0}$ and $r_{t,\pi} = Af_{t,\pi}$. Assuming that the forward curve discretised over n time points, $r_{t,\tau}$ can be interpreted as the tenor specific spot rate. The tenor specific spot rate is separated further into the risk free spot rate $r_{t,0}$ and the tenor specific spot rate $r_{t,\pi}$. There are multiple indices used in the section and to illustrate the meaning of them, consider the risk free rate $r_{t,0}^{\hat{T}}$. This spot rate corresponds to the discrete approximation of $\psi_0(\hat{T})$, estimated using the forward curve that was observed at the time point t. $\psi_0(\hat{T})$ can be used to discount a risk free cash flow from the date $\hat{T}$ to the spot date t.

4.6.1 Forward Rate Agreements

An FRA contract can be valuated as a future cash flow, determined by formula (3.13) and presented here again.

$$c_{FRA}(\tau, T_1, T_2) = N\frac{(E_t^{T_2}[L_\tau(T_1, T_2)] - r_{fix})\Delta T}{1 + E_t^{T_2}[L_\tau(T_1, T_2)]\Delta T} \tag{4.8}$$

The expression above can be simplified if the cash flow is considered at time point T_2 instead of T_1. This results in the contract being discounted from T_2 rather than T_1. This approximation makes the

gradient of the theoretical price much easier to calculate. After moving the cash flow to the time point T_2, the following expression is obtained:

$$c_{FRA}(\tau, T_1, T_2) = N(E_t^{T_2}[L_\tau(T_1, T_2)] - r_{fix})\Delta T. \tag{4.9}$$

The goal is to describe the price in terms of the discrete representation of continuous forward rates, $f_{t,0}$ and $f_{t,\tau}$. The cash flow is rewritten as an expression based on $\psi_\tau(\hat{T})$ from (3.9) using the approximation of the expected value, (3.12),

$$c_{FRA}(\tau, T_1, T_2) = N(e^{(\psi_\tau(T_2) - \psi_\tau(T_1))} - 1 - r_{fix}\Delta T) \tag{4.10}$$

which leads to

$$c_{FRA}(\tau, T_1, T_2) = N(\exp\left\{\int_0^{T_2} f_\tau(t)dt - \int_0^{T_1} f_\tau(t)dt\right\} - 1 - r_{fix}\Delta T). \tag{4.11}$$

When the expected cash flow c_{FRA} has been estimated, the expected price of the contract can be derived by discounting c_{FRA} with the risk-free rate that is implied by $f_0(t)$ at the day of the estimation.

$$P_{FRA}(\tau, T_1, T_2) = e^{-\int_{T_0}^{T_2} f_0(t)dt} c_{FRA}(\tau, T_1, T_2). \tag{4.12}$$

The cash flow is discounted to time point T_0 which normally is the settlement date of the contract. However, when using the interest rate curves derived in section 3.3.1 the curves are created from today (and not from the settlement date) and can therefore be discounted to zero instead of T_0.

The cash flow should be received at time point T_2 when disregarding from the denominator as presented in (4.9). When time point T_1 is reached however the cash flow at T_2 is already known as the floating rate has been fixed and then the cash flow is calculated as

$$c_{FRA}(\tau, T_1, T_2) = N\frac{(r_{\tau,float} - r_{fix})\Delta T}{1 + r_{\tau,float}\Delta T} \tag{4.13}$$

and received at T_1 as presented in equation (3.14).

A discrete price expression has to be produced because the interest rate curves are expressed on discrete form. This discrete form requires a discretised time vector $T = (T_0, ..., T_n)$ and the discretised spot rate curves $\left(r_{t,0}^0, r_{t,0}^1, ..., r_{t,0}^n,\right)$ and $(r_{t,\pi}^0, r_{t,\pi}^1, ..., r_{t,\pi}^n)$. Because $r_{t,0} = AE_0\xi_{t,0}$, the spot rate $r_{t,0}^{T_j}$ can be extracted for time T_j via the integrating matrix A. By defining $a_{j,0}^T$ as the j-th row in the matrix AE_0. A change in the risk free spot rate is

$$\Delta r_{t,0}^{T_j} = a_{j,0}^T \Delta\xi_{t,0} \tag{4.14}$$

This also holds for the tenor specific spot rate if $\xi_{t,0}$ and E_0 is replaced by $\xi_{t,\pi}$ and E_π, respectively. Given these definitions, the cash flow in equation (4.11) is approximated (due to the discretisation) as

$$\begin{aligned} c_{FRA}(t, \Delta\xi_{t,\tau}) = N \cdot \Bigg\{ & \exp\{(r_{t-1,0}^{T_2} + a_{T_2,0}^T \Delta\xi_{t,0} + r_{t-1,\pi}^{T_2} + a_{T_2,\pi}^T \Delta\xi_{t,\pi})T_2 \\ & -(r_{t-1,0}^{T_1} + a_{T_1,0}^T \Delta\xi_{t,0} + r_{t-1,\pi}^{T_1} + a_{T_1,\pi}^T \Delta\xi_{t,\pi})T_1\} - (1 + r_{fix}\Delta T)\Bigg\}. \end{aligned} \tag{4.15}$$

By discounting this expression, as done in equation (4.12), it results in the fully discrete price expression of an FRA contract,

$$\begin{aligned} P_{FRA} = \exp(-(T_2(r^{T_2}_{t-1,0} + a^T_{T_2,0}\Delta\xi_{t,0}) - T_0(r^{T_0}_{t-1,0} + a^T_{T_0,0}\Delta\xi_{t,0}))) \cdot N \cdot \Big\{ \exp\{(r^{T_2}_{t-1,0} + \\ a^T_{T_2,0}\Delta\xi_{t,0} + r^{T_2}_{t-1,\pi} + a^T_{T_2,\pi}\Delta\xi_{t,\pi})T_2 \\ -(r^{T_1}_{t-1,0} + a^T_{T_1,0}\Delta\xi_{t,0} + r^{T_1}_{t-1,\pi} + a^T_{T_1,\pi}\Delta\xi_{t,\pi})T_1\} - (1 + r_{fix}\Delta T)\Big\}. \end{aligned} \tag{4.16}$$

The following new variables are introduced to simplify the price expression:

$$\begin{aligned} \Delta\xi_t &= \begin{pmatrix} \Delta\xi_{t,0} \\ \Delta\xi_{t,\pi} \end{pmatrix} \\ A_1 &= \begin{pmatrix} -(T_2 a_{T_2,0} - T_0 a_{T_0,0}) + T_2 a_{T_2,0} - T_1 a_{T_1,0} \\ T_2 a_{T_2,\pi} - T_1 a_{T_1,\pi} \end{pmatrix} = \begin{pmatrix} T_0 a_{T_0,0} - T_1 a_{T_1,0} \\ T_2 a_{T_2,\pi} - T_1 a_{T_1,\pi} \end{pmatrix} \\ A_2 &= \begin{pmatrix} -(T_2 a_{T_2,0} - T_0 a_{T_0,0}) \\ 0 \end{pmatrix} \\ K_1 &= -\left(T_2 r^{T_2}_{t-1,0} - T_0 r^{T_0}_{t-1,0}\right) + T_2 r^{T_2}_{t-1,\tau} - T_1 r^{T_1}_{t-1,\tau} = T_0 r^{T_0}_{t-1,0} + T_2 r^{T_2}_{t-1,\pi} - T_1 r^{T_1}_{t-1,\tau} \\ K_2 &= -\left(T_2 r^{T_2}_{t-1,0} - T_0 r^{T_0}_{t-1,0}\right) \end{aligned} \tag{4.17}$$

These variables lead to the price expression

$$P_{FRA} = N\big(e^{K_1 + A_1^T \Delta\xi_t} - (1 + r_{fix}\Delta T)e^{K_2 + A_2^T \Delta\xi_t}\big). \tag{4.18}$$

The gradient and hessian of (4.18) with respect to $\Delta\xi_t$ is

$$\nabla_{\Delta\xi_t} P_{FRA} = N\left(A_1 e^{K_1 + A_1^T \Delta\xi_t} - A_2\left(1 + r_{fix}\Delta T\right) e^{K_2 + A_2^T \Delta\xi_t}\right) \tag{4.19a}$$

$$H_{\Delta\xi_t} P_{FRA} = N\left(A_1 A_1^T e^{K_1 + A_1^T \Delta\xi_t} - A_2 A_2^T\left(1 + r_{fix}\Delta T\right) e^{K_2 + A_2^T \Delta\xi_t}\right) \tag{4.19b}$$

(4.19a) and (4.19b) with $\Delta\xi_t = 0$ gives

$$\begin{aligned} \nabla_{\Delta\xi_t} P_{FRA}\Big|_{\Delta\xi_t = 0} &= N\left(A_1 e^{K_1} - A_2\left(1 + r_{fix}\Delta T\right) e^{K_2}\right) \\ H_{\Delta\xi_t} P_{FRA}\Big|_{\Delta\xi_t = 0} &= N\left(A_1 A_1^T e^{K_1} - A_2 A_2^T\left(1 + r_{fix}\Delta T\right) e^{K_2}\right). \end{aligned} \tag{4.20}$$

The gradient and hessian expressions are valid to $T_1 - 1$ and become zero after this time point. The presented gradient and hessian are used in 4.8 to approximate price changes in FRA contracts.

4.6.2 Interest Rate Swaps

The value of an IRS contract is found by separating the contract into a fixed leg and a floating leg,

$$P_{IRS} = P_{float} - P_{fix}. \tag{4.21}$$

The fixed leg has regular payments that are calculated using the swap rate of the IRS. The size of these payments depends on the time intervals of the fixed payments that the swap rate is accumulated over. For each fixed cash flow that is paid at time T_j, the notation Δt_j is used to describe the time interval that the swap rate has been accumulated over. The index $j \in \mathcal{J}$ is used to distinguish each fixed payment of an IRS.

The floating leg also has regular payments that depend on time intervals and payment dates. Every IBOR rate has a determined start point and endpoint for which the IBOR rate is accumulated over. T_i is the starting point and T_j is the stopping points of the time intervals of an IBOR rate $L(T_i, T_j)$. This means that the IBOR rate for each day is associated with an index pair (i, j) that determine the start and end of the period. The size of a floating cash flow is found by multiplying the IBOR rate $L(T_i, T_j)$ with the time period $\hat{t_i}$ for which the rate is active over (using the IRS specific day counting convention). For an IRS, the timing of each floating does not necessarily coincide with the stopping point of the active IBOR rate. The time of payment for the floating leg is denoted T_k. Each floating payment from an IRS has an associated vector (i, j, k) that determines the start, stop and pay date, respectively. The set $\mathcal{F}_{obs}$ is defined as the set containing the index vectors for floating payments with a determined IBOR rate. For payments where IBOR has not yet been observed, the same notation is used, but the set is defined as $\mathcal{F}_{exp}$. This separates the value of the floating leg into two separate parts, P^{obs}_{float} and P^{exp}_{float}. The value of the IRS is thus

$$P_{IRS} = P^{obs}_{float} + P^{exp}_{float} - P_{fix}, \tag{4.22}$$

where

$$\begin{aligned} P^{obs}_{float} &= \sum_{(i,j,k)\in\mathcal{F}_{obs}} \Delta\hat{t}_i L(T_i, T_j) d(T_k) \\ P^{exp}_{float} &= \sum_{(i,j,k)\in\mathcal{F}_{exp}} \Delta\hat{t}_i E^{T_i}_t \left[L(T_i, T_j)\right] d(T_k) \\ P_{fix} &= \sum_{j\in\mathcal{J}} y\Delta t_j d(T_j). \end{aligned} \tag{4.23}$$

The spot rate curves implicitly estimated from the optimisation framework described in 3.3.1 are used to estimate the IBOR rates and the risk-free discount factors. Since these spot rates are continuous and IBOR rates are simple, the expected IBOR is calculated as

$$E^{T_i}_t \left[L(T_i, T_j)\right] = \frac{\exp\left\{r^{T_j}_{t,\tau} T_j - r^{T_i}_{t,\tau} T_i\right\} - 1}{\Delta t_i}, \tag{4.24}$$

where Δt_i is the time period for which the IBOR rate is active, measured in the day counting convention that has been used in the estimation of the spot rate curves (ACT/365). Using the discrete spot curves, P^{obs}_{float}, P^{exp}_{float} and P_{fix} are calculated as

$$P^{obs}_{float} = \sum_{(i,j,k)\in\mathcal{F}_{obs}} \Delta\hat{t}_i L\left(T_i, T_j\right) \exp\left\{-r^{T_k}_{t,\tau} T_k\right\} \tag{4.25}$$

$$P^{exp}_{float} = \sum_{(i,j,k)\in\mathcal{F}_{exp}} \Delta\hat{t}_i \frac{\exp\left\{r^{T_j}_{t,\tau} T_j - r^{T_i}_{t,\tau} T_i\right\} - 1}{\Delta t_i} \exp\left\{-r^{T_k}_{t,0} T_k\right\} \tag{4.26}$$

$$P_{fix} = \sum_{j \in \mathcal{J}} y \Delta t_j \exp\big\{ - r_{t,0}^{T_j} T_j \big\}. \tag{4.27}$$

For performance attribution, the gradient and hessian of the price of the IRS is needed w.r.t the risk factor innovation $\Delta\xi_t$. The gradient and hessian is separated in the same way as the price in (4.22):

$$\nabla_{\Delta\xi_t} P_{IRS} = \nabla_{\Delta\xi_t} P_{float}^{obs} + \nabla_{\Delta\xi_t} P_{float}^{exp} - \nabla_{\Delta\xi_t} P_{fix} \tag{4.28}$$

$$H_{IRS} = H_{float}^{obs} + H_{float}^{exp} - H_{fix}. \tag{4.29}$$

The gradients of accrued interest for the different spot rate curves $r_{t,0}, r_{t,\pi}$ and $r_{t,\tau}$ from time $T = 0$ to $T = T_k$ w.r.t. risk factor innovations $\Delta\xi_t$ are

$$\begin{aligned}
\nabla_{\Delta\xi_t} \exp\left\{ r_{t,0}^{T_k} T_k \right\} &= a_{k,0} T_k \exp\left\{ r_{t,0}^{T_k} T_k \right\} \\
\nabla_{\Delta\xi_t} \exp\left\{ r_{t,\pi}^{T_k} T_k \right\} &= a_{k,\pi} T_k \exp\left\{ r_{t,\pi}^{T_k} T_k \right\} \\
\nabla_{\Delta\xi_t} \exp\left\{ r_{t,\tau}^{T_k} T_k \right\} &= a_{k,\tau} T_k \exp\left\{ r_{t,\tau}^{T_k} T_k \right\},
\end{aligned} \tag{4.30}$$

where $a_{k,0}^T$, $a_{k,\pi}^T$ and $a_{k,\tau}^T$ are the k-th rows of the matrices $(AE_0, \mathbf{0})$, $(\mathbf{0}, AE_\pi)$ and (AE_0, AE_π), respectively. The gradients of P_{float}^{exp}, P_{float}^{obs} and P_{fix} can be found by applying (4.30) on (4.25), (4.26) and (4.27). This gives the gradients

$$\nabla_{\Delta\xi_t} P_{float}^{obs} = \sum_{(i,j,k) \in \mathcal{F}_{\text{obs}}} -a_{k,0} T_k \Delta\hat{t}_i L\left(T_i, T_j\right) \exp\left\{ -r_{t,0}^{T_k} T_k \right\}, \tag{4.31}$$

$$\begin{aligned}
\nabla_{\Delta\xi_t} P_{float}^{exp} = \sum_{(i,j,k) \in \mathcal{F}_{exp}} \frac{\Delta\hat{t}_i}{\Delta t_i} \Big(& A_{exp}(i,j,k) \exp\left\{ r_{t,\tau}^{T_j} T_j - r_{t,\tau}^{T_i} T_i - r_{t,0}^{T_k} T_k \right\} \\
& - (a_{k,0} T_k) \exp\big\{ - r_{t,0}^{T_k} T_k \big\} \Big),
\end{aligned} \tag{4.32}$$

and

$$\nabla_{\Delta\xi_t} P_{fix} = \sum_{j \in \mathcal{J}} a_{j,0} T_j y \Delta t_j \exp\big\{ - r_{t,0}^{T_j} T_j \big\}, \tag{4.33}$$

where the function $A_{exp}(i,j,k)$ is defined as

$$A_{exp}(i,j,k) = a_{j,\tau} T_j - a_{i,\tau} T_i - a_{k,0} T_k, \quad (i,j,k) \in \mathcal{F}_{exp} \tag{4.34}$$

to make the gradient expression for the expected IBOR payments more compact.

The hessian of (4.25), (4.26) and (4.27) are found by applying (4.30) on (4.31), (4.32) and (4.33) in the same way as before. The hessians of P^{exp}_{float}, P^{obs}_{float} and P_{fix} are

$$H^{obs}_{float} = \sum_{(i,j,k)\in\mathcal{F}_{\text{obs}}} -a_{k,0}T_k\Delta\hat{t}_i L\left(T_i,T_j\right)\exp\left\{-r^{T_k}_{t,0}T_k\right\}, \tag{4.35}$$

$$\begin{aligned} H^{exp}_{float} = \sum_{(i,j,k)\in\mathcal{F}_{exp}} \frac{\Delta\hat{t}_i}{\Delta t_i}\Big(&A_{exp}(i,j,k)A_{exp}(i,j,k)^T\exp\left\{r^{T_j}_{t,\tau}T_j - r^{T_i}_{t,\tau}T_i - r^{T_k}_{t,0}T_k\right\} \\ &-\left(a_{k,0}T_k\right)\left(a_{k,0}T_k\right)^T\exp\left\{-r^{T_k}_{t,0}T_k\right\}\Big) \end{aligned} \tag{4.36}$$

and

$$H_{fix} = \sum_{j\in\mathcal{J}}\left(a_{j,0}T_j\right)\left(a_{j,0}T_j\right)^T y\Delta t_j\exp\left\{-r^{T_j}_{t,0}T_j\right\}. \tag{4.37}$$

4.6.3 Implementation phase

The contract valuation is a time-consuming algorithm. To increase the speed of the process, the valuation step is simplified by grouping the cash flows by the day that they have been paid/received. To determine the NPV of the cash flows for a certain day, the sum of every cash flow of that day is discounted once, instead of discounting every single cash flow individually. Since there is a large number of contracts in the simulated swap book of the bank, the valuation algorithm becomes much faster.

4.7 Hedging

The stochastic programming model presented in section 3.9.1 will be used with some modifications. The presented expression is valid for single yield curves and must be extended to use the multiple yield curves produced by the term structure estimation model presented in section 3.3.1.

In this study, only FRA and IRS contracts on the European market are considered which means that these are the only possible interest rate contracts to be used in the hedge. The transaction costs for these two types of contracts are calculated as presented in equation 3.66 with the modification that the single interest rate curve is substituted for the two curves $f_{t,0}$ and π_t used in this application. The price expression used to estimate the perceived transaction costs can be derived from the discrete definitions in sections 4.6.1 and 4.6.2.

In the hedge, the number of simulations S must be determined as well as the penalising constant α. The theoretical reasoning behind the values on these parameters originates from an alpha frontier presented in Blomvall and Hagenbjörk (2020). In order to avoid having to spend an unnecessary amount of time optimising the value of α, which is not the purpose of this thesis, it is set to a range similar to what is used by Blomvall and Hagenbjörk (ibid.). The risk factor simulation is not affected by α, which means the chosen value is not critical for conclusions about the risk factor model. The number of simulations S is set to the same value (5000) as used by Blomvall and Hagenbjörk (ibid.),

see section 3.9.1. In their paper, they utilised antithetic sampling to generate the scenarios for the forward rate curves but here Latin hypercube sampling is used instead, described in detail by de Flon and Landgren (2021).

The no-transaction region is an intuitive way of visualising the trade-off between transaction costs and rebalancing. The region defines a section where the perceived profits of rebalancing is too low in relation to the costs of rebalancing. The hedge model trades such that the hedged swap book reaches the edge of the no-transaction region, where further rebalancing is no longer profitable. A detailed description is given by Bohlin and Harling (2021), they also discuss the problem of setting the parameter α.

In order to determine if this stochastic programming based hedge model outperforms a traditional hedge, the hedging method of boxes introduced by Hagan and West (2006) will be implemented as a reference model for comparison. The implementation and comparison of the methods are described in Bohlin and Harling (2021) for a portfolio of foreign exchange swaps and is not discussed further here. The two different hedging models will be presented in a chart that weighs transactions costs against swap book volatility for different parameter settings.

During the implementation phase, the hedge based on stochastic programming is implemented in AMPL and solved using CPLEX. To make the runs in the hedge comparable the randomised customer trades are saved and used every time the when hedge is run. This decreases the degrees of freedom for the hedge and makes the results more comparable between runs. Noise will also be added in a similar way as made by Blomvall and Hagenbjörk (2020) and presented in 3.9.1 but might also be scaled further to achieve stability in a swap book consisting of both IRS and FRA contracts in a multiple yield curve setting on the European interest rate market.

In order to evaluate the risk factor decomposition model used, the best model according to the statistical test is implemented in the stochastic hedging model. The results from this hedge will be compared to the hedge achieved via PCA components which is used as a benchmark as it was previously used in the hedging model. This is done to decide if the components only give a better representation of changes in swap rates (as the result from test states) or also improves the hedge which is desirable. The comparison is based on the statement "A hedge model is better if the hedged portfolio has lower transaction costs for each portfolio volatility level" from section 2.2.6 and is evaluated through a graph comparing hedged portfolio volatility and the corresponding transaction cost for the different decomposition models tested.

4.8 Performance Attribution

The performance attribution framework described in 3.9.2 is applied to the hedged swap book. The (rotated) significant risk factors ξ_t are used with the performance attribution framework together with the contract-specific calculations introduced in 3.2.1 and 3.2.2 to derive the sources of the changes in portfolio value. The performance attribution is done on the out-of-sample period of the data for both hedge models. The error term $\Delta\epsilon_t^P$ that defines the error between the theoretical price and the market price of the portfolio will include the spread that the customers and banks have to cross in their transactions. This error term is separated into the pricing error $\Delta\epsilon_{t,i}^{P_c}$ that comes from customers crossing the spread to the bank and the pricing error $\Delta\epsilon_{t,i}^{P_h}$ that comes from the hedging transactions of the bank. $\Delta\epsilon_{t,i}^{P_h}$ is calculated the same way as (3.66), while $\Delta\epsilon_{t,i}^{P_c}$ swaps places of

the bid and ask swap rates (since the customers are paying the bank to cross the spread). The separation into customer and hedge pricing errors replaces the theoretical price error in equation (3.74) with

$$\Delta\epsilon^{P}_{t,i} = \Delta\epsilon^{P_c}_{t,i} + \Delta\epsilon^{P_h}_{t,i}. \tag{4.38}$$

This results in the residual free performance attribution equation for individual contracts:

$$\Delta\bar{P}_{t,i} = \theta_{t-1,i}\Delta t + g^{T}_{t-1,i}\Delta\xi_t + \frac{1}{2}\Delta\xi^{T}_t H_{t,i}\Delta\xi_t + \epsilon^{I}_{t,i} + \epsilon^{A}_{t,i} + \Delta\epsilon^{P_c}_{t,i} + \Delta\epsilon^{P_h}_{t,i}. \tag{4.39}$$

This equation and the equivalent aggregation to portfolio level will be used to separate the performance into its (aggregated) error terms of the portfolio. This is the performance attribution of a hedged swap book. The performance attribution assumes the perspective of a European investor and the swap book value is measured in €. This makes the performance of the swap book immune to currency exchange rates.

Chapter 5

Results

In this section, the results from the hedge and decomposition models in the thesis are presented along with brief commentary. To make the comparison of the components generated by the decomposition models easy, they have also been included in appendix A without any surrounding text.

Under section 4.3.1 a method was described in order to determine the number of components to extract but this step was not implemented due to lack of time. Instead, six components were retrieved in all decomposition models as this led to good results when used by Blomvall and Hagenbjörk (2019a) in their performance attribution.

5.1 Risk Factor Decomposition

Several variants of PCA has been implemented and tested and the results for each one is presented under their respective section. First the standard PCA is discussed and then PCA viewed as an optimisation problem using the self-developed kernel penalisation to achieve realistic components. The last two methods are based on rotations of the standard PCA components, first, the results from the quartimin rotation are presented and then the self-developed rotation using the same kernel used in PCA as an optimisation problem. Lastly, the components generated from the two methods of ICA is presented.

5.1.1 Ordinary PCA

The components generated from ordinary PCA is presented in figure 5.1 and 5.2 for the risk-free curve and the tenor premium respectively. In these graphs, it can be noted that the first three components of the risk free yield curve are the characteristic shift, twist and butterfly components.

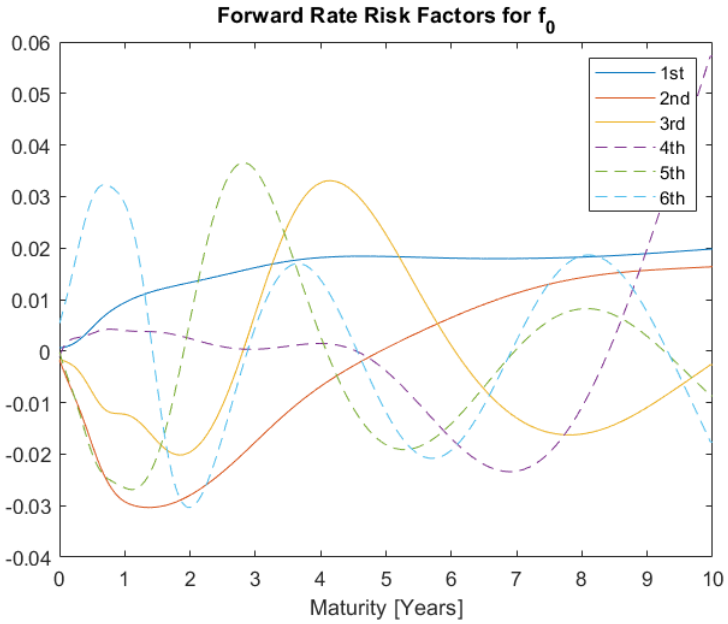


Figure 5.1: Components for innovations of f_0 after PCA

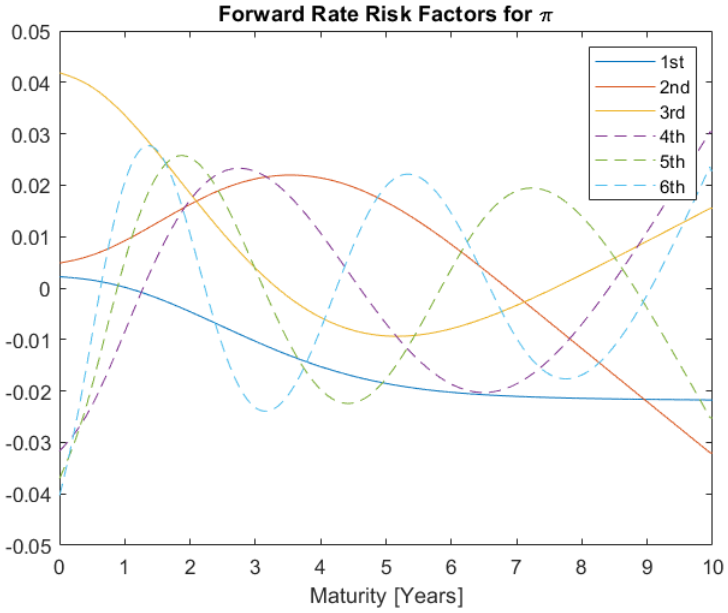


Figure 5.2: Components for innovations of π_τ after PCA.

5.1.2 PCA as an Optimisation Problem

The systematic forward curve innovations are separated into two different types of innovations. The first type of innovations are changes in the forward rate curve for longer maturities. These changes are thought to originate from changes in market expectations of future forward rates. The second type of innovations are forward curve changes for shorter maturities. These innovations are thought to be affected mainly by changes in the risk-free rate set by central banks.

To separate long term market expectations and short term financial policy decisions, each component is allowed to describe changes in either the shorter or longer maturities, but never both. For this reason, a regularisation is introduced to the extreme eigenvalue problem in the form of a weight vector h with the same length as the component x. The weight matrix H is defined as the matrix containing h on the diagonal.

$$H = \begin{pmatrix} h_1 & & \\ & \ddots & \\ & & h_n \end{pmatrix} \tag{5.1}$$

The chosen regularisation is the inner product of the weight vector h and the squared elements of the component x. The problem of finding the first component is thus described as

$$\begin{aligned} \max_{x \in \mathbb{R}^n} \quad & x^T A x - x^T H x \\ \text{s.t} \quad & x^T x = 1. \end{aligned} \tag{5.2}$$

To find another component, an additional condition must be added to prevent the same component from being found again. By collecting previously found components in a component matrix E, the similarity between the new component x and the previous components are penalised. The inner product of the columns E_i and x is forced to be less than a tolerance level $\epsilon \geq 0$. If $\epsilon = 0$, the optimal component x will be orthogonal to the components in E. The optimisation problem with the added conditions from the component matrix is

$$\begin{aligned} \max_{x \in \mathbb{R}^n} \quad & x^T A x - x^T H x \\ \text{s.t} \quad & x^T x = 1 \\ & -\epsilon \leq E_i x \leq \epsilon, \quad i = 1, ..., p, \end{aligned} \tag{5.3}$$

where p is the number of components in the component matrix.
The kernel functions h_p that was used to find the first three components ($p = 0, 1, 2$) are

$$h_p(t) = \begin{cases} \beta_p \left(\frac{(T-t)}{T} \right)^{\alpha} & t < T \\ 0 & t \geq T \end{cases} \tag{5.4}$$

A total of six components were sought after. The last three components were found with the following kernel function, to isolate movements on the short end:

$$h_p(t) = \begin{cases} \beta_p \left(\frac{t}{T} \right)^{\alpha} & t < T \\ \beta_p & t \geq T \end{cases} \tag{5.5}$$

Finding a suitable level of β_p without a reference is difficult. For this reason, β_p was calculated using the corresponding eigenvalue λ_p of the covariance matrix scaled with a multiplier m_p. The parameters for each kernel is presented in table 5.1 and 5.2 below. The regularised PCA was performed on the in-sample innovations of the observations of the risk-free forward curve f_0 and the tenor specific forward premiums π_τ separately. The components are shown in figure 5.3 and 5.4.

For both the risk-free rate and the tenor premium, the first three components only affect the longer maturities. Components 4-6 only affect the shorter maturities. This is considered to be a realistic separation of risk factors, as the movements for shorter maturities originates from monetary policies while the movements for longer maturities are affected by market sentiment.

Table 5.1: Parameter settings for regularised PCA on innovations of the risk-free forward rate f_0.

Component	p	α	T	λ_p	m_p	β_p	ϵ
1	0	8	5	$8.3 \cdot 10^{-4}$	10	$8.3 \cdot 10^{-3}$	
2	1	8	5	$1.8 \cdot 10^{-4}$	10	$1.8 \cdot 10^{-3}$	0.9
3	2	8	5	$8.5 \cdot 10^{-5}$	10	$8.5 \cdot 10^{-4}$	0.9
4	3	8	5	$6.7 \cdot 10^{-5}$	10^3	$6.7 \cdot 10^{-2}$	0.9
5	4	8	5	$4.3 \cdot 10^{-5}$	10^3	$4.3 \cdot 10^{-2}$	0.9
6	5	8	5	$2.1 \cdot 10^{-5}$	10^3	$2.1 \cdot 10^{-2}$	0.9

Table 5.2: Parameter settings for regularised PCA on innovations of the tenor specific tenor premium π_τ.

Component	p	α	T	λ_p	m_p	β_p	ϵ
1	0	8	5	$2.2 \cdot 10^{-4}$	1	$2.2 \cdot 10^{-4}$	
2	1	8	5	$5.7 \cdot 10^{-5}$	1	$5.7 \cdot 10^{-5}$	0.9
3	2	8	5	$1.9 \cdot 10^{-5}$	1	$1.9 \cdot 10^{-5}$	0.9
4	3	8	5	$2.0 \cdot 10^{-6}$	$1 \cdot 10^4$	$2.0 \cdot 10^{-3}$	0.9
5	4	8	5	$2.7 \cdot 10^{-7}$	$1 \cdot 10^4$	$2.7 \cdot 10^{-3}$	0.9
6	5	8	5	$2.5 \cdot 10^{-8}$	$5 \cdot 10^4$	$1.3 \cdot 10^{-3}$	0.9

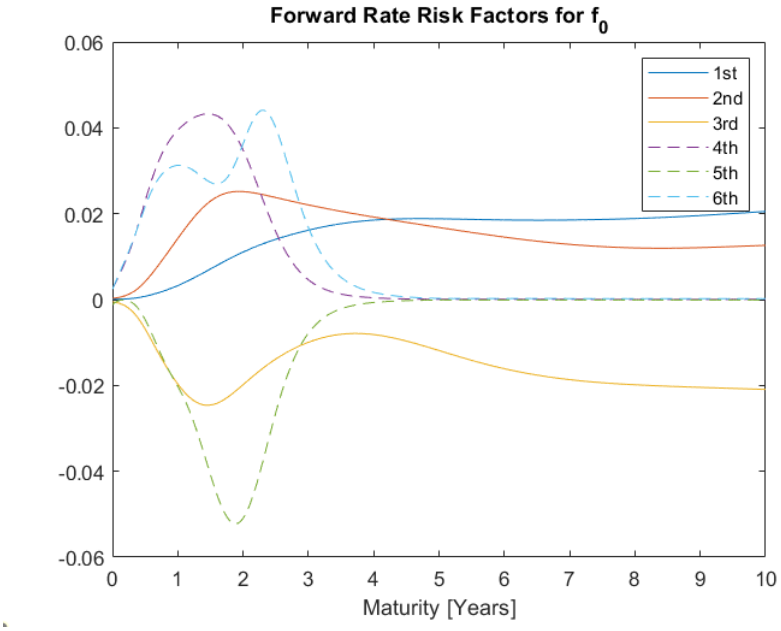


Figure 5.3: Components for innovations of f_0 after regularised PCA.

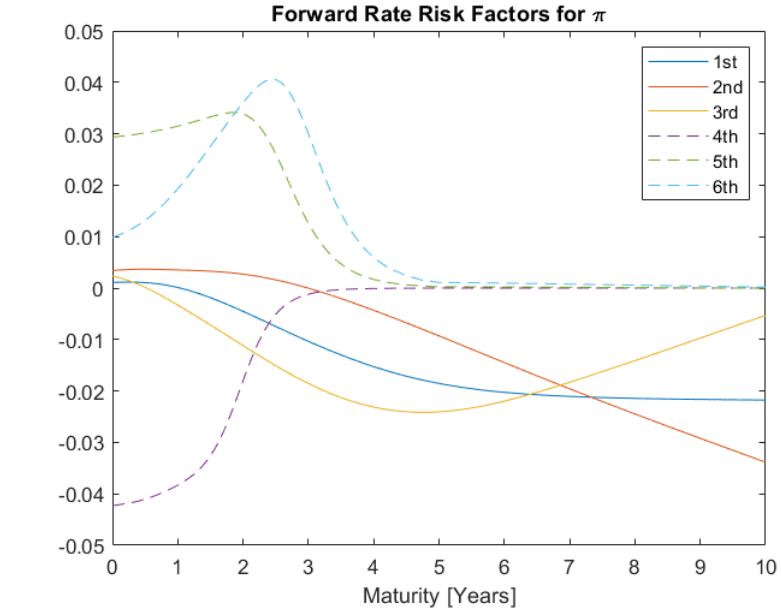


Figure 5.4: Components for innovations of π_τ after regularised PCA.

5.1.3 Rotated PCA Components using Quartimin

Using the rotational framework (in 3.6), a quartimin rotation was applied to the components generated by the PCA. The parameter values used to generate the components are presented in table 5.3 and the initial rotation matrix R is set to the identity matrix. The resulting components are shown

Table 5.3: Parameter settings for the quartimin rotation defined in section 3.6.

α	s
1	10^{-6}

in figure 5.5 and 5.6. Each component has a distinct section of the forward curves that are explained, similar to the case of the regularised PCA. Therefore it should be a good alternative to use, however the same interpretability of the short and long end, as described in 5.1.2 is not achieved.

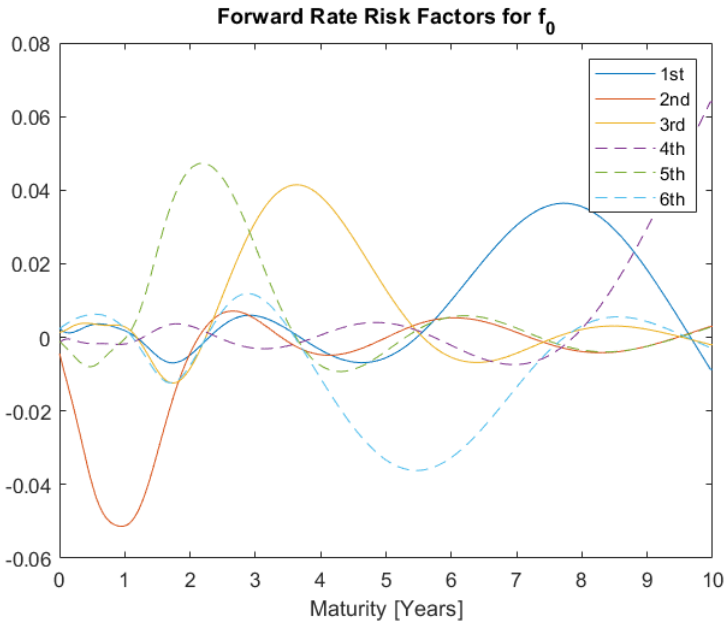


Figure 5.5: Components for innovations of f_0 after a quartimin rotation of PCA components.

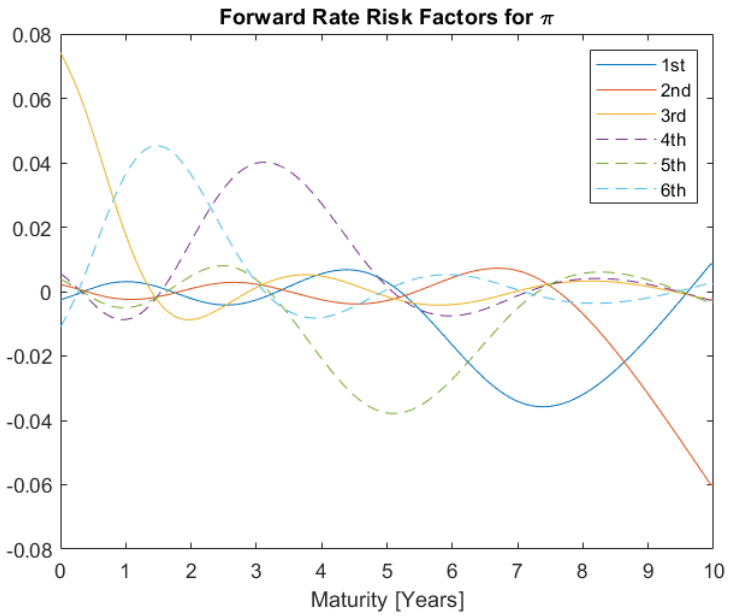


Figure 5.6: Components for innovations of π_τ after a quartimin rotation of PCA components.

5.1.4 Principal Components Rotated Towards a Kernel

The kernels h_i that are used in regularised PCA are applied as a cost function with the generalised rotation framework described in 3.6. The same parameter settings were used here as for the quartimin rotation and is presented in table 5.5. The identity matrix is used as the initial rotation matrix. The cost function $Q(\Lambda)$ is defined as

$$Q(\Lambda) = \sum_{i=1}^{p} = \Lambda_i^T \operatorname{diag}(h_i)\Lambda_i \tag{5.6}$$

with the gradient

$$\nabla Q(\Lambda) = \sum_{i=1}^{p} 2h_i^T \Lambda_i. \tag{5.7}$$

The components for the innovations of the risk-free forward rates and the tenor premiums are shown in figure 5.7 and 5.8, respectively. The regularisation of the PCA optimisation problem yields components that explain less variance than components from a regular PCA does. By rotating the PCA components, the new components exhibit no loss in explained variance, thus making them more accurate than the regularised PCA components. The rotated components also explain more independent movements than regular PCA components since the shape discussed in 5.1.2 is promoted.

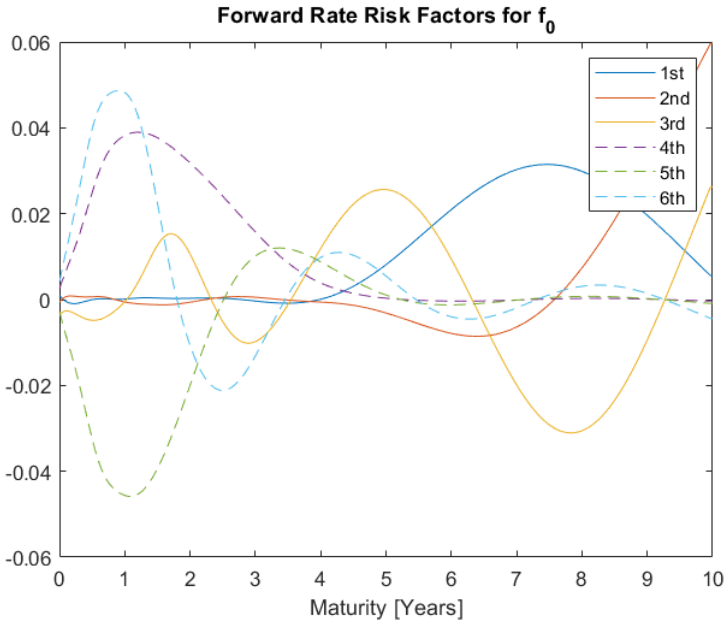


Figure 5.7: Components for innovations of f_0 after rotation of PCA components using a kernel.

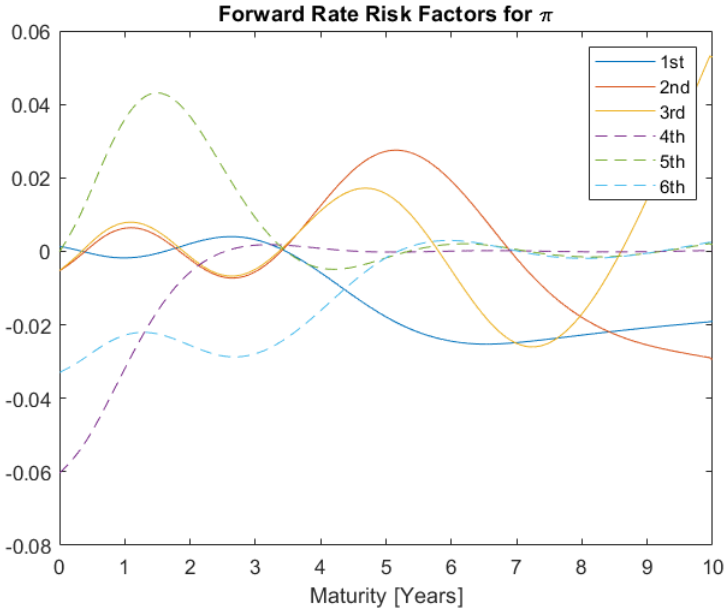


Figure 5.8: Components for innovations of π_τ after rotation of PCA components using a kernel.

5.1.5 Independent Component Analysis

For the ICA model, a third-party Matlab package is used. This package is called fastICA and was created by Hugo Gävert, Jarmo Hurri, Jaakko Särelä and Aapo Hyvärinen. Hyvärinen is a professor of computer science (Machine Learning) at the University of Helsinki.

By using the fastICA package, the method described in 3.4.5 is applied, with the exception that independent components that do not improve the determinant of the reconstructed covariance matrix over a predefined threshold are removed. The two different orthogonalisation techniques were investigated separately, as there is no way of determining which one is more appropriate before applying the statistical test. The components for the innovations of the risk-free forward rate curves and the tenor specific forward premium are shown in the figures below. Figure 5.9 and 5.10 show the components that were found using the deflation method. Figure 5.11 and 5.12 show the components that were found with the symmetric orthogonalisation. The 6th component for π_τ has been removed for both methods due to the previously mentioned condition on the reconstructed covariance matrix.

The components do not resemble the previously mentioned systematic market behaviour, but instead have been optimised to have non-Gaussian scores.

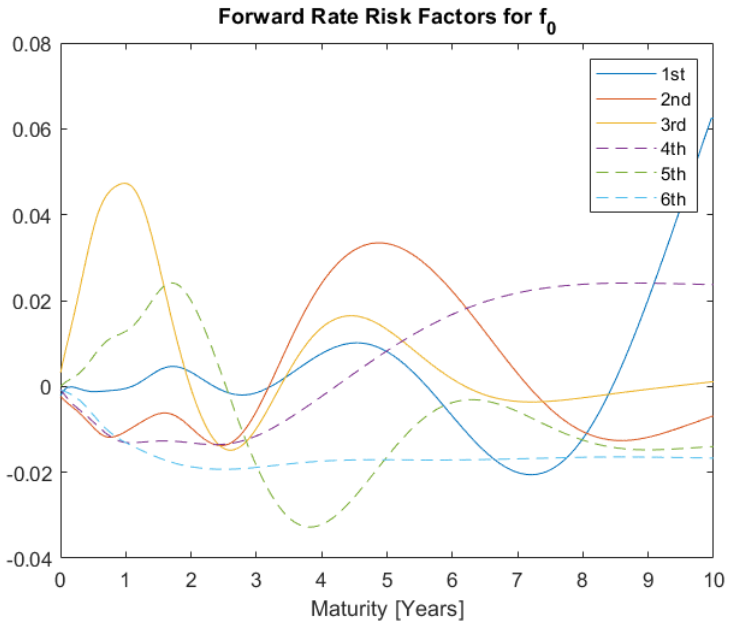


Figure 5.9: Components for innovations of f_0 after ICA using the deflation algorithm.

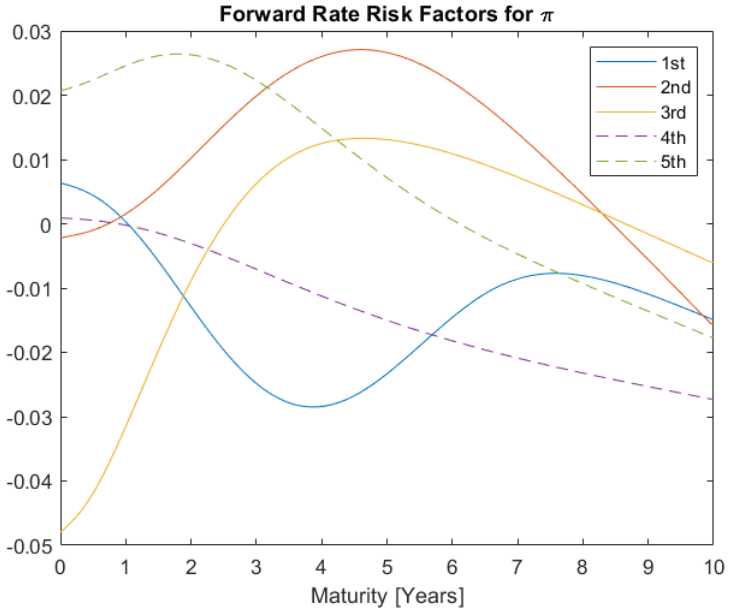


Figure 5.10: Components for innovations of π_τ after ICA using the deflation algorithm.

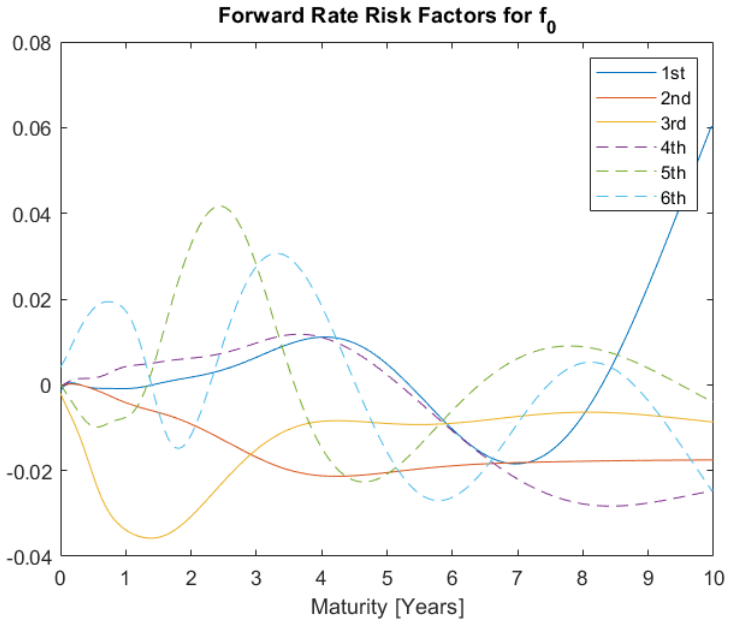


Figure 5.11: Components for innovations of f_0 after ICA using symmetric orthogonalisation.

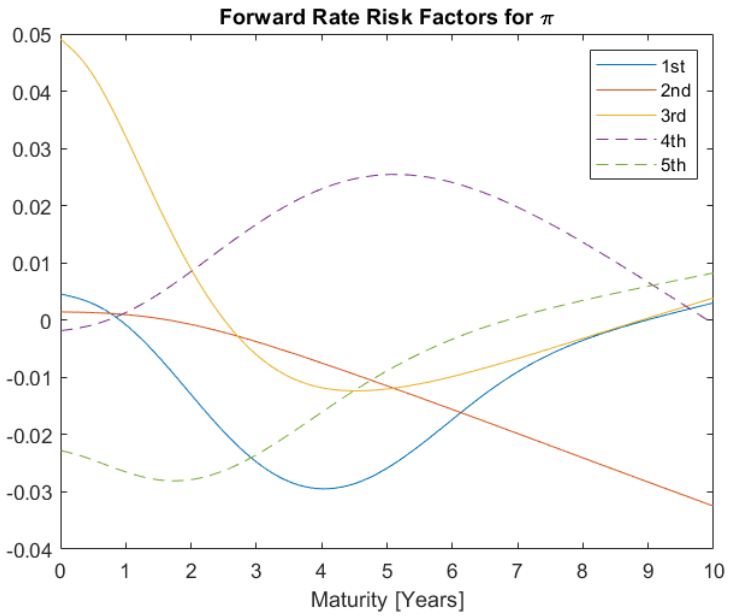


Figure 5.12: Components for innovations of π_τ after ICA using symmetric orthogonalisation.

5.2 Statistical Test

The results from the statistical test (likelihood ratio test) performed by de Flon and Landgren (2021) is given here and for a detailed description, see their thesis. In tables, 5.4 and 5.5, the p-values are presented. Each row shows the probability of the model of first column being better than the other models. The test is relative and only evaluates how good the models are at explaining changes in swap rates relative to one another. During the evaluation, however, a naive model was used which assumed the change in interest rate curve between days was zero. This model was compared to all other models in order to make sure that no implementation error was made, and all decomposition models outperformed the naive model as they should. Both Gaussian and Student's t-distributions were tested as the univariate distribution and copula respectively. Using the Student's t-distribution for both copula and univariate distribution resulted in higher likelihood values and therefore they were used in the statistical test presented in table 5.4 and 5.5 and also in the hedge.

In summary, rotated PCA components using quartimin is the best model to describe changes in swap rates for the FRA contracts studied. For the IRS swap rates, the rotation of PCA components using the rotational kernel is a favourable choice. These will therefore be implemented in the hedging model.

Table 5.4: p-values for models of FRA swap rates.

	Reg. PCA	ICA (defl)	ICA (symm)	PCA	Rot. kernel	Rot. quartimin
Reg. PCA	NaN	0	0	0	0	0
ICA (defl)	1	NaN	1	1	0	0
ICA (symm)	1	0	NaN	0	0	0
PCA	1	0	1	NaN	0	0
Rot. kernel	1	1	1	1	NaN	0
Rot. quartimin	1	1	1	1	1	NaN

Table 5.5: p-values for models of IRS swap rates.

	Reg. PCA	ICA (defl)	ICA (symm)	PCA	Rot. kernel	Rot. quartimin
Reg. PCA	NaN	0	0	0	0	0
ICA (defl)	1	NaN	1	1	0	0
ICA (symm)	1	0	NaN	0	0	0
PCA	1	0	1	NaN	0	0
Rot. kernel	1	1	1	1	NaN	1
Rot. quartimin	1	1	1	1	0	NaN

5.3 Hedging

The original idea was to create a hedge for the portfolio consisting of both IRS and FRA contracts. This was not possible due to the hedging model matching cash flows for the longer FRA contracts and shorter IRS contracts. This lead to the hedge buying excessive amounts of contracts to reach the lowest possible volatility tomorrow but instead resulted in enormous value spikes today. Over time this lead to a hedge that diverged and therefore the hedge was studied for FRA and IRS contracts separately.

Based on the results in 5.2, the Student's t-distribution was used as the univariate distribution of the risk factors and also for the copula generating the term structure scenarios.

5.3.1 Hedging an FRA Portfolio

The results from the implemented hedge, (3.65) for only FRA contracts are presented in figure 5.13 and 5.14. The first figure is a plot of the portfolio value of the unhedged portfolio generated from customer transactions and the hedged portfolio respectively. The components used in this hedge were generated using standard PCA and used as a benchmark to show that the hedge implementation functions correctly. From the graph it can be noted that a very effective hedge has been implemented, there is however more to be desired in terms of return if this was to be used as a real hedge at the bank due to the high transaction costs generated by the hedge, presented in figure 5.14. The results are sufficient to conclude that an effective hedge model has been developed and can cope with FRA contracts in a multiple yield curve setting. The noise used in the implementation was the same as described by Blomvall and Hagenbjörk (2020).

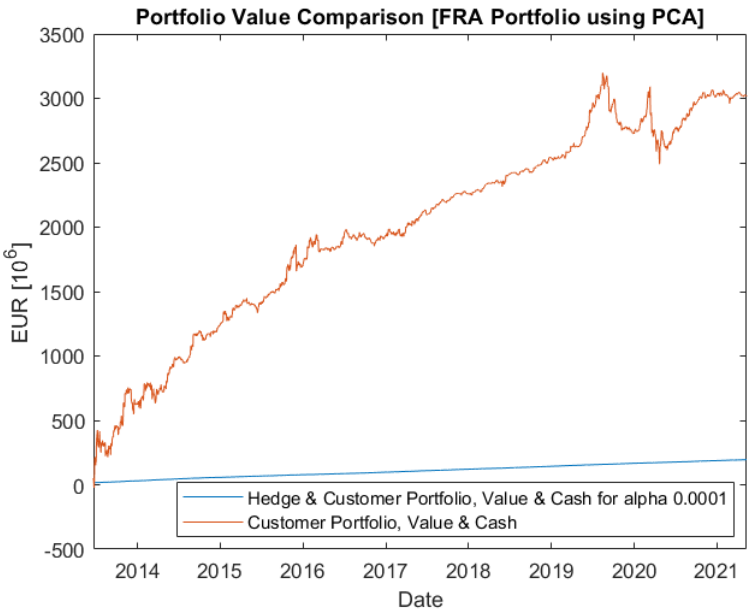


Figure 5.13: Hedge over the out-of-sample period, 2013-06-21 to 2021-05-11, for a portfolio consisting of FRA contracts using components generated with PCA.

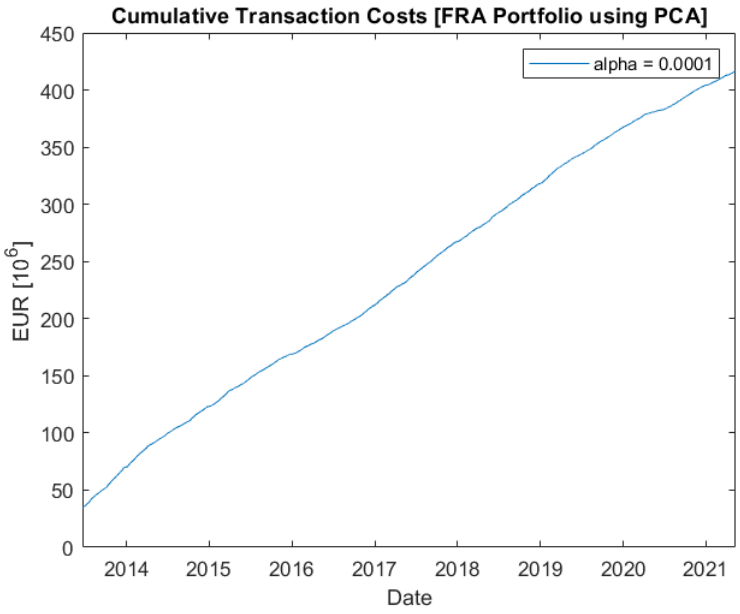


Figure 5.14: Cumulative transaction costs over the out-of-sample period, 2013-06-21 to 2021-05-11, for a portfolio consisting of FRA contracts using components generated with PCA.

In order to decrease the volatility of the portfolio, the hedge model trades contracts with other banks, which means crossing the bid-ask spread in a non-beneficial way. This gives rise to perceived transaction costs (discussed in section 3.9.1) and as noted in figure 5.14 these are high because low portfolio volatility is prioritised over low transaction costs when using such a low value of α.

In figure 5.15 an underwater plot is presented, showing the drawdown of the portfolio value. It signals how fast the hedge recovers to its highest peak in portfolio value.

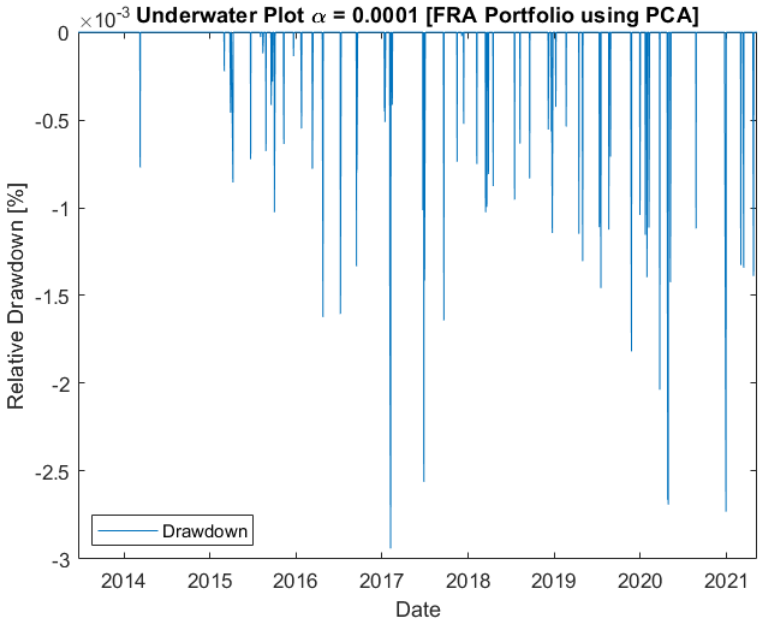


Figure 5.15: Relative drawdown for $\alpha = 0.0001$ over the out-of-sample period plus one day in order to only study the hedged portfolio, 2013-06-24 to 2021-05-11, for a portfolio consisting of FRA contracts.

In figure 5.16 the excess value for the hedged portfolio for different α values is presented. The excess value is computed against the hedged portfolio with $\alpha = 0.0001$ and this can be used from the bank's perspective to select an appropriate α value depending on the risk appetite.

From the results in 5.2 it was concluded that the best decomposition model from the ones tested was quartimin. Therefore this model was implemented and tested in the hedging model (with all other affecting parameters kept constant like customer trades, simulated risk factor scenarios and noise levels) and is presented in 5.17. In this graph, it can be noted that quartimin does not perform better than PCA.

During testing of the hedge it was noted that, for low values of α along with a decrease in noise added to the simulated yield curves, the results indicated that quartimin resulted in both lower

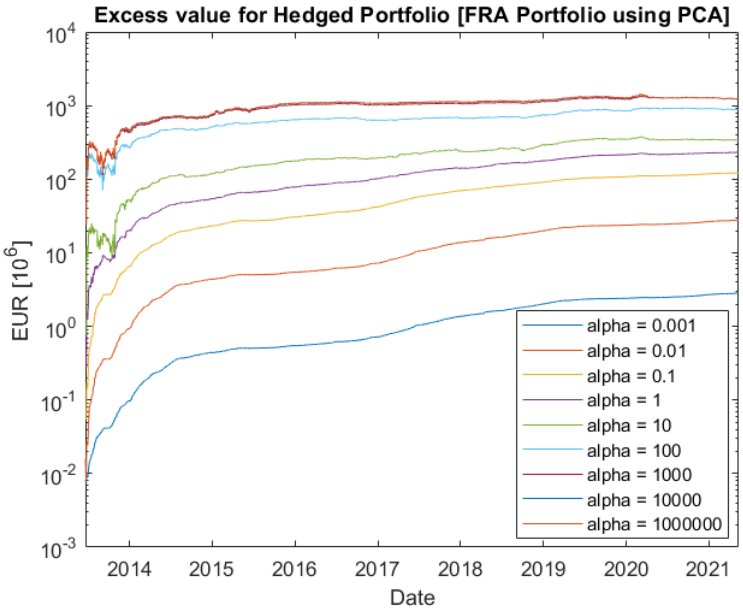


Figure 5.16: Excess value over the out-of-sample period, 2013-06-21 to 2021-05-11, for a portfolio consisting of FRA contracts using components generated with PCA with $\alpha = 0.0001$ as a benchmark.

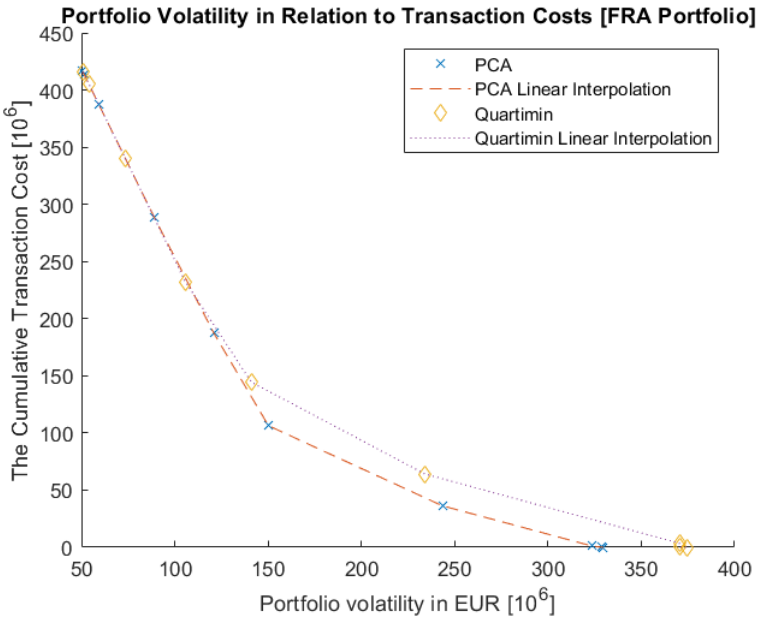


Figure 5.17: Comparison between total transaction cost and portfolio volatility for the FRA portfolio using PCA components and rotated PCA components using quartimin. α-values used in the plot: 0.0001, 0.001, 0.01, 0.1, 1, 10, 100, 1000, 10000, 1000000.

transaction costs and lower portfolio volatility but the hedge was not stable. This was not studied further due to lack of time but this would be interesting to test further as the noise has a significant impact on the results for low values of α.

5.3.2 Hedging an IRS Portfolio

This section contains the same graphs, in the same order as in 5.3.1 but for a portfolio consisting of only IRS contracts. It can be noted that the portfolio value and cumulative transaction costs in 5.18 and 5.19 is larger than the corresponding values for the FRA case. The effect is derived from the initial contracts in table 4.5 being bigger than 4.4 and that it occurs more trading in IRS contracts than FRA contracts, as seen in table 4.3.

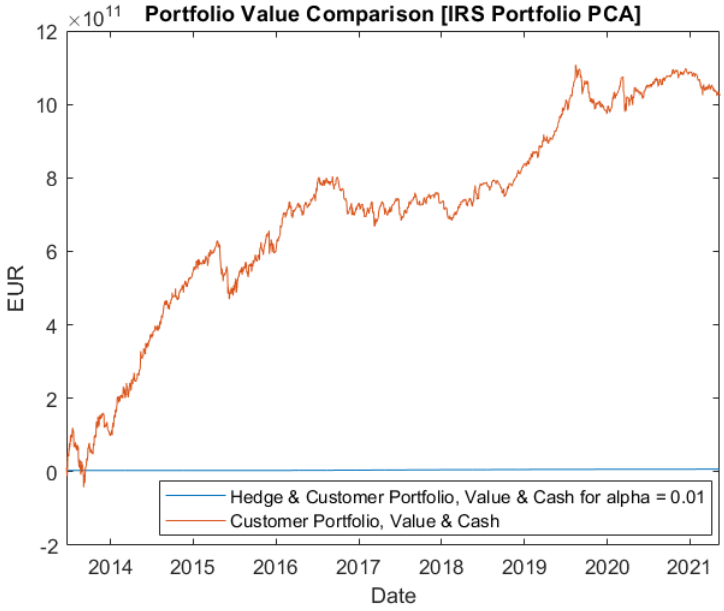


Figure 5.18: Hedge over the out-of-sample period, 2013-06-21 to 2021-05-11, for a portfolio consisting of IRS contracts using components generated with PCA.

In figure 5.20 an underwater plot is presented, showing the drawdown of the portfolio value. It can be noted that this portfolio recovers from spikes in portfolio value in a similar way to the FRA portfolio.

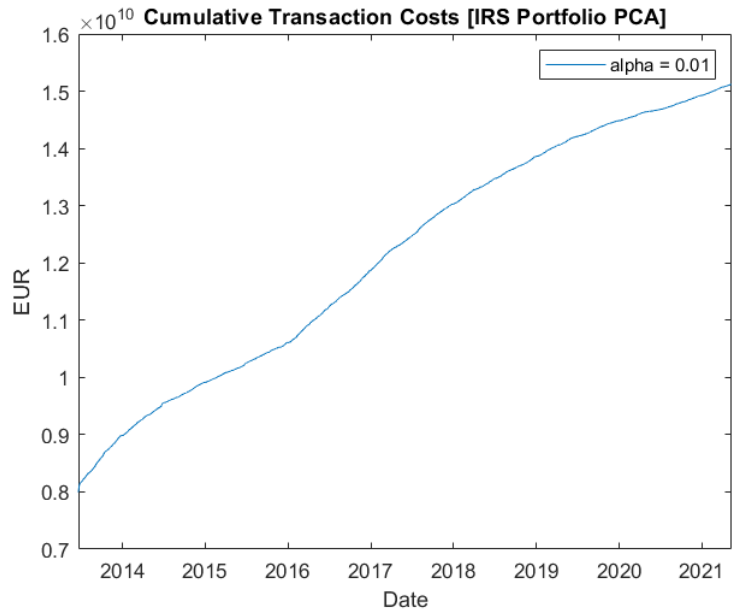


Figure 5.19: Cumulative transaction costs over the out-of-sample period, 2013-06-21 to 2021-05-11, for a portfolio consisting of IRS contracts using components generated with PCA.

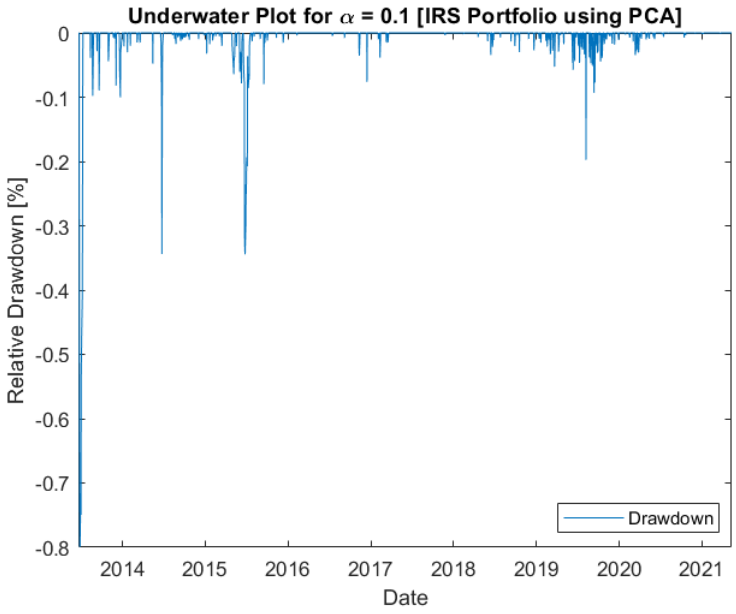


Figure 5.20: Relative drawdown for $\alpha = 0.1$ over the out-of-sample period plus one business day in order to only study the hedged portfolio, 2013-06-24 to 2021-05-11, for a portfolio consisting of IRS contracts.

In figure 5.21 the excess value for the hedged portfolio for different α values is presented. The excess value is computed against the hedged portfolio with $\alpha = 0.01$ and this can be used from the bank's perspective to select an appropriate α value depending on the risk appetite. The plot looks different for α values of 100 and 1000 compared to figure 5.16 for the FRA portfolio. This happens because the hedge (for $\alpha = 100$ and $\alpha = 1000$) is more volatile than the hedge for $\alpha = 0.01$ but still has a low portfolio value around zero. It results in the portfolio value being lower than the value of the lower alpha portfolios during certain periods due to the bigger fluctuations and this leads to a lower excess value.

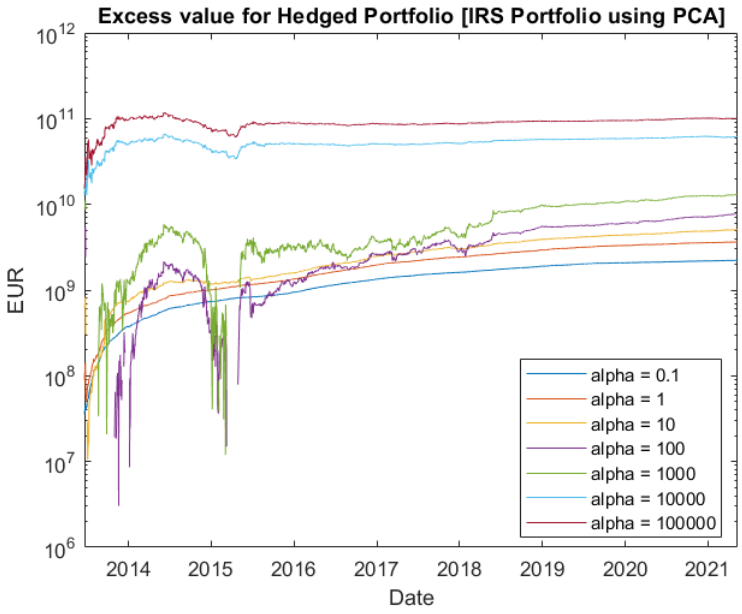


Figure 5.21: Excess value over the out-of-sample period, 2013-06-21 to 2021-05-11, for a portfolio consisting of IRS contracts using components generated with PCA with $\alpha = 0.01$ as a benchmark.

For the runs in 5.22 the noise, customer trades and simulated risk factor scenarios was the same for the different α-levels. Each α-level was only run once for each decomposition model used. In the graph, it can be noted that there is no significant improvement using the model that was shown to explain swap rates better.

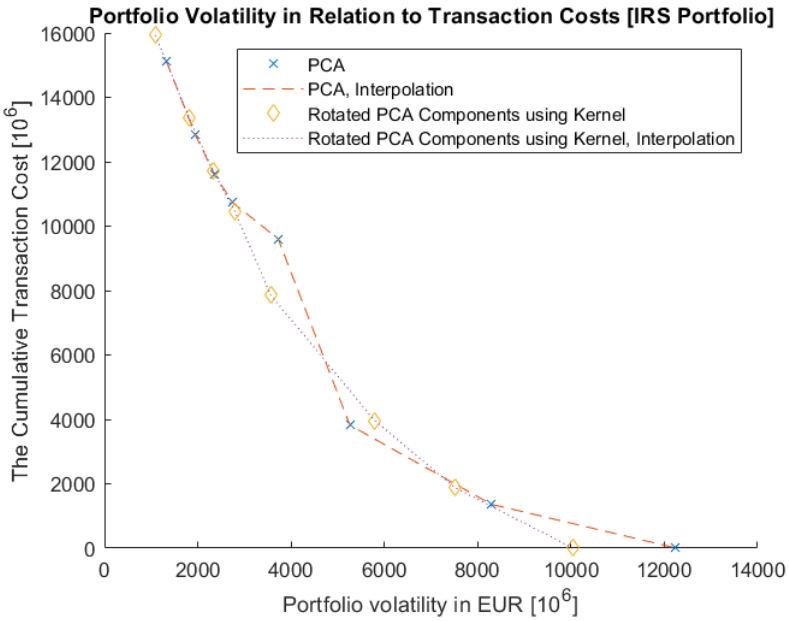


Figure 5.22: Comparison between total transaction cost and portfolio volatility for the IRS portfolio using PCA components and rotated PCA components using a Kernel. α-values used in the plot: 0.01, 0.1, 1, 10, 100, 1000, 10000, 100000.

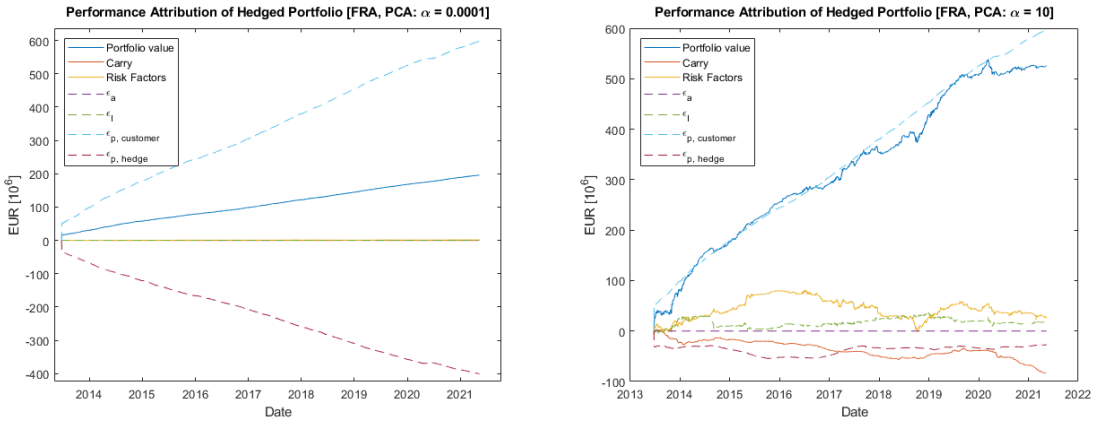
5.4 Performance Attribution

In this section, the performance attributions for the different types of hedges are presented. The performance attribution separates the portfolio returns into different sources. The change in portfolio value that comes from risk factor innovations is called 'Risk Factors' in the figures. The risk factors do not capture all market risk, which can be seen by the error term ϵ_I. The tested component types are all able to produce very stable hedges.

In the plots presented in this section, it is interesting to study $\epsilon_{p,customer}$ and $\epsilon_{p,hedge}$ as these variables describe the amount of portfolio value that is added from customers and the hedge algorithm crossing the spread. In the perfect hedges 5.23a, 5.24a, 5.25a and 5.26a the risk factor exposure is almost zero as it does not contribute to the value of the portfolios. The only factors affecting the value of the portfolios are the customer and hedge positions and in these runs, the value from the customer positions is bigger than the hedge positions which results in a positive portfolio value for the bank.

In figures 5.23b, 5.24b, 5.25b and 5.26b the hedge is not perfect as the risk factor exposure contributes to the total portfolio value. Worth noting is that different α values are used for the FRA and IRS portfolios as the value of the portfolios are in different order of magnitude and this must be taken into account when choosing the α parameter for the portfolio. $\alpha = 10$ would have resulted in a perfect hedge (due to the high portfolio value) for the IRS portfolio and therefore a higher α was used in the illustrations. As a result of using higher values of α, the portfolio value is more volatile due to exposing the portfolio to more risk but instead $\epsilon_{p,hedge}$ is lower and this illustrates the trade-off that must be decided on when hedging and the flexibility of the model.

5.4.1 Performance Attribution of FRA Portfolio: PCA



(a) $\alpha = 0.0001$ (b) $\alpha = 10$

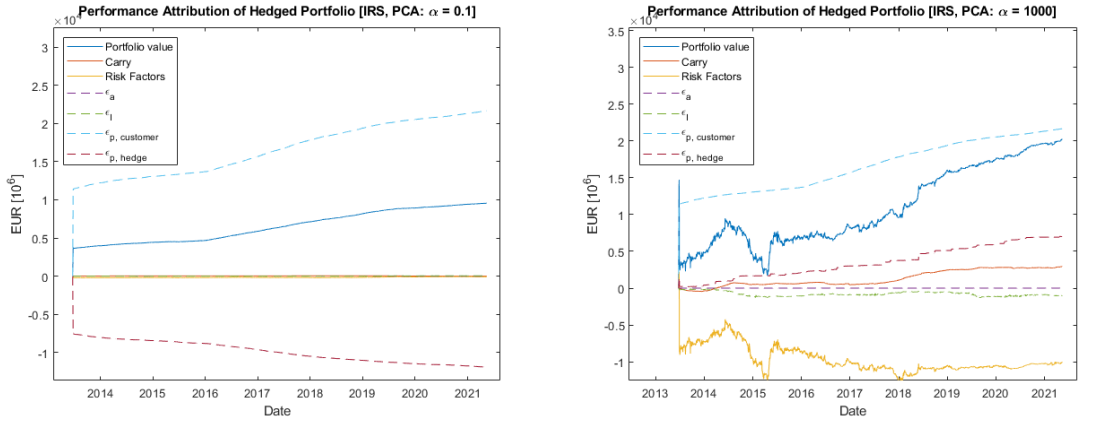
Figure 5.23: Performance attribution of the hedged FRA portfolio using principal components for the complete out-of-sample period.

5.4.2 Performance Attribution of FRA Portfolio: Quartimin Rotation

(a) $\alpha = 0.0001$ (b) $\alpha = 10$

Figure 5.24: Performance attribution of the hedged FRA swap book using a quartimin rotation of principal components for the complete out-of-sample period.

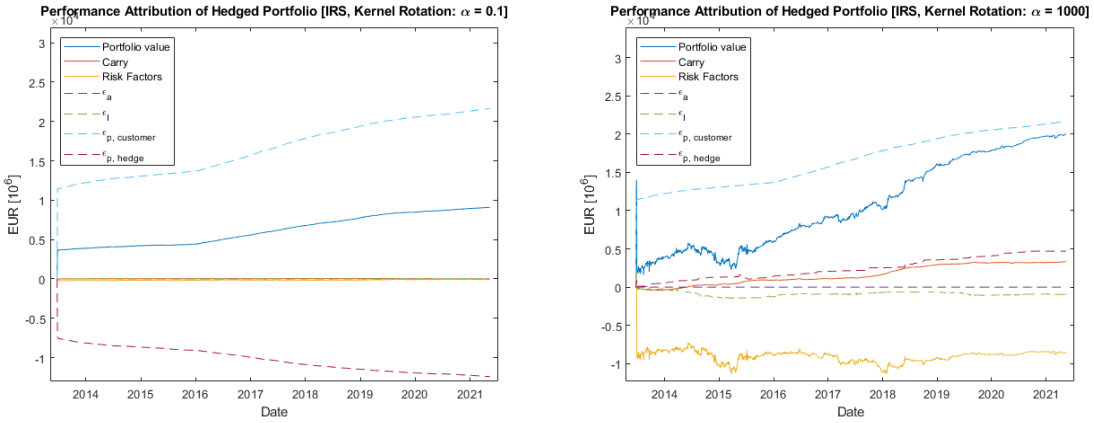
5.4.3 Performance Attribution of IRS Portfolio: PCA



(a) $\alpha = 0.1$ (b) $\alpha = 1000$

Figure 5.25: Performance attribution of the hedged IRS swap book using principal components for the complete out-of-sample period.

5.4.4 Performance Attribution of IRS Portfolio: Kernel Rotation



(a) $\alpha = 0.1$ (b) $\alpha = 1000$

Figure 5.26: Performance attribution of the hedged IRS portfolio using a rotation of principal components toward a kernel function for the complete out-of-sample period.

Chapter 6

Discussion

In this section, a discussion is given regarding the results and the challenges that have been encountered in the study, along with a conclusion, ethical aspects of the project and suggestions for further research.

6.1 Customer Simulation

In this thesis, emphasis is put on simulating a realistic initial swap book portfolio. Using the initial IRS swap book as an example, the swap book only consists of eleven IRS contracts with a total of eleven fixed future payments. This is a trivial scenario to hedge since every contract has matching cash flow dates. This would not be the case if the bank actually had been accumulating contracts over a period of ten years before the date of hedging.

6.2 Risk Factors

In this study, several realistic risk factor models were found that explain changes in swap rates better than PCA. The statistical test of the component models shows that a simple rotation of principal components increases the quality of the estimated probability density function of swap rates for FRA and IRS contracts. Hedging with the new components does not improve the hedge and this has two possible causes. The first is that the statistical test does not prioritise properties of the risk factor model that are important to the hedge model. The second possible cause is that the noise that is added to the simulated curves in the hedge model nullifies the effect of the improved risk factor simulations.

The simulation of risk factors rely on the accuracy of the estimated yield curves, the explained variance of the components and how well the distributions fit the risk factors. Principal components retain the highest amount of variance in the data, but the corresponding risk factors may fit poorly to an analytical distribution. By rotating the principal components, the explained variance remains the same, and the rotated risk factors could become easier to fit an analytical distribution. This would result in a more realistic simulation of yield curves.

The component kernel is used to force components to separate market risk into shorter and longer maturities. The statistical test confirmed that this economic reasoning is better than pure data-

driven models for the IRS portfolio. This also provides a possible explanation for why the FRA portfolio did not benefit from this separation: The FRA portfolio is only exposed to the first two years of maturities on the yield curve and unaffected by components with emphasis on longer maturities.

6.3 Hedge Model

During the implementation phase, we realised the difficulty of setting good parameters and creating a realistic simulation in order to achieve a well functioning and realistic hedge as there are many exceptions that have to be dealt with. One of these is the combination of α and noise in the hedging model, if not combined in the right way it leads to either an unstable hedge or a too stable hedge. Different level of noise combined with α values was tested with a background in the noise level presented in section 3.9.1 in order to achieve a stable hedge over the complete out-of-sample period. A high level of noise resulted in an uncertain hedge which is not sure which decision to take as the outcome is too random and too low noise led to an overconfident hedge that took big positions and diverged over time. A higher noise level means that the hedge model becomes less confident and can not eliminate all portfolio volatility because of the decreased accuracy of the simulations. To not spend an excessive amount of time trimming in the parameters, which is not the focus of this thesis (see Bohlin and Harling (2021) for a detailed description of how to derive good parameter values), the noise level was set to the same level as used by Blomvall and Hagenbjörk (2020) which was sufficient in this setting.

Another issue with the current hedging algorithm is it does not realise when the current yield curves are of poor quality (from missing swap rates). This leads to poorly estimated transaction costs and subsequently worse decision making. For a full description of the development process of interest rate curves and what criteria are used for constructing the interest rate curves used in this thesis, see de Flon and Landgren (2021).

The hedge model has been tested on the combined FRA and IRS swap books, thus allowing synergies between the contracts to be utilised. However, on days where some FRA and IRS contracts use the same (future) IBOR rate and cash flow date, the hedge buys an excessive amount of contracts, regardless of α. This behaviour means that a small error in the yield curve that is used to calculate the transaction cost of the contracts could have catastrophic effects. This problem may be resolved by forbidding the hedge to buy contracts that have matching the IBOR and cash flow dates.

6.4 Conclusion

In conclusion, we have managed to create a hedging model which can hedge a swap book containing different swap contracts with the same tenor. The hedging model developed by Blomvall and Hagenbjörk (2020) has been expanded to a multiple yield curve framework and can hedge portfolios consisting of IRS or FRA contracts. More realistic decomposition models have been found, that explains changes in swap rates on the European interest rate market better. However, the hedging did not show any significant improvement after the inclusion of the realistic risk factors.

6.5 Ethical Aspects

Market makers have a very important function in financial markets, and the financial markets are deeply connected in a complex global network. This makes the global financial system very fragile to the mistakes of these market makers, especially in the field of risk management. The framework proposed in this thesis relies on a simplified description of reality in the form of assumptions and approximations, just like any scientific model. Practitioners of risk management have an obligation to make sure they clearly understand the practical risks of the proposed framework and the underlying assumptions and approximations before they rely on it.

The parameters used in the project, for example, the portfolio simulation, has been developed in collaboration with the three banks but are still very arbitrary and do not reflect reality in an optimal way. When evaluating the results and implementing the hedge model, the sensitivity to the chosen parameter values was relatively high. Therefore the results are sensitive to the implementation made by the authors and must be refined before they can truly be trusted.

Some extreme observations have been removed from the in-sample data set, which could make the model less realistic. The approach that has been taken by the authors is that as few points should be removed as possible and that observations should only be removed if they are obvious measurement errors.

Lastly, changes of extreme magnitude in market data happen rarely. For this reason, it is important to note that the estimated distributions may fit historical data well while also being a poor representation of reality in extreme scenarios such as future financial crises. The statistical toolbox can not provide more information than the data itself, and managers of risk must be vigilant for scenarios that are rare or unobserved.

6.6 Further Research

The component framework can be improved by reducing the sensitivity towards noise and allowing component structures to change over time. The quality of the noise-reduction may be increased by using exploratory factor analysis, which could potentially result in components of higher quality. Using time-invariant components could increase the accuracy of the simulation of term structures, but measuring market risk with this type of component description becomes more abstract. As for the number of components to retain, cross-validation remains a promising method.

The statistical test suggests that the rotations improve estimations of the multivariate probability density, but a deeper examination of this topic would provide a decisive conclusion to the value of rotations in this context.

Simulation of the initial swap book could be made more realistic. In this study, the initial portfolio was required to consist of existing contracts. By relaxing this constraint and allowing the initial portfolio to be a basket of (non-contract-specific) IBOR rate exposures of varying sizes, the initial swap book comes closer to its intended scenario. This also makes the hedging of the initial swap book non-trivial, because the majority of the cash flows can not be perfectly matched with available contracts.

The statistical test has a very high certainty, but the unclear empirical relation between the test and hedge performance warrants further investigation to determine the cause.

As for the hedge model, the decision making may be robustified by introducing uncertainty or noise to the calculation of transaction costs. Finally, if the hedge model is improved to run stable for the combined swap book of FRA and IRS contracts, synergies between the contract types can be utilised to achieve an even more cost-efficient hedge.

Bibliography

Bank for International Settlements (2020). *OTC, interest rate derivatives*. Accessed: 2021-06-14. URL: https://stats.bis.org/statx/srs/table/d7.

Bianchetti, C. (2011). 'Interest Rates After The Credit Crunch: Multiple-Curve Vanilla Derivatives and SABR.' In: URL: http://arxiv.org/abs/1103.2567.

Bishop, C. M. (2006). *Pattern recognition and machine learning*. Information science and statistics. Springer. ISBN: 0387310738.

Blom, G., Enger, J., Englund, G., Grandell, J. and Holst, L. (2005). *Sannolikhetsteori och statistikteori med tillämpningar*. Studentlitteratur.

Blomvall, J. and Hagenbjörk, J. (2019a). 'A generic framework for monetary performance attribution'. In: *Journal of Banking and Finance* 105. cited By 0, pp. 121–133. DOI: 10.1016/j.jbankfin.2019.05.021. URL: https://www.sciencedirect.com/science/article/abs/pii/S0378426619301268.

Blomvall, J. and Hagenbjörk, J. (2019b). 'Simulation and evaluation of the distribution of interest rate risk.' In: *Computational Management Science* 16.1-2, pp. 297–327. ISSN: 1619-697X. URL: https://link.springer.com/article/10.1007%2Fs10287-018-0319-8.

Blomvall, J. and Hagenbjörk, J. (2020). 'Hedging of interest rate risk using a stochastic programming approach'. Working paper.

Blomvall, J. and Ndengo, M. (2013). 'Multiple Yield Curves Estimation Using A Generalized Optimization Framework'. In: *Term structure estimation based on a generalized optimization framework*.

Blomvall, J. and Ndengo, M. (n.d.). 'Accurate measurements of yield curves and their systematic risk factors'. Working paper.

Bohlin, E. and Harling, J. (2021). 'Hedging of a foreign exchange swapbook using Stochastic programming'. MA thesis. Linköping University, Production Economics, p. 76.

Burges, C. (2009). 'Dimension reduction: A guided tour.' In: *Foundations and Trends in Machine Learning* 2.4, pp. 275–365. ISSN: 19358237. DOI: 10.1561/2200000002. URL: https://www.nowpublishers.com/article/Details/MAL-002.

Camacho, J. and Ferrer, A. (2014). 'Cross-validation in PCA models with the element-wise k-fold (ekf) algorithm: Practical aspects.' In: *Chemometrics and Intelligent Laboratory Systems* 131, pp. 37–50. ISSN: 0169-7439. DOI: 10.1016/j.chemolab.2013.12.003. URL: https://www.sciencedirect.com/science/article/abs/pii/S0169743913002335.

Cattell, R. B. (1966). 'The Scree Test For The Number Of Factors'. In: *Multivariate Behavioral Research* 1.2. PMID: 26828106, pp. 245–276. DOI: 10.1207/s15327906mbr0102_10. eprint: https://doi.org/10.1207/s15327906mbr0102_10. URL: https://doi.org/10.1207/s15327906mbr0102_10.

Cherubini, U., Gobbi, F., Mulinacci, S. and Romagnoli, S. (2012). *Dynamic Copula Methods in Finance*. cited By 61. DOI: 10.1002/9781118467404.

de Flon, M. and Landgren, H. (2021). 'Hedging of an interest rate swap portfolio using stochastic programming focusing on risk factor simulation'. In:

Golub, L. (2006). 'Continuous methods for extreme and interior eigenvalue problems.' In: *Linear Algebra and Its Applications* 415.1, pp. 31–51. ISSN: 0024-3795. URL: https://www.sciencedirect.com/science/article/pii/S0024379505000315.

Goodfellow, I., Bengio, Y. and Courville, A. (2016). *Deep Learning*. http://www.deeplearningbook.org. MIT Press.

Hagan, P. and West, G. (2006). 'Interpolation methods for curve construction'. In: *Applied Mathematical Finance* 13.2. cited By 60, pp. 89–129. DOI: 10.1080/13504860500396032. URL: https://www.tandfonline.com/doi/abs/10.1080/13504860500396032.

Hinton, G. E. and Salakhutdinov, R. R. (2006). 'Reducing the Dimensionality of Data with Neural Networks.' In: *Science* 313.5786, pp. 504–507. ISSN: 00368075. URL: https://www.jstor.org/stable/3846811.

Hull, J. (2015). *Options, futures, and other derivatives.* Pearson. ISBN: 9780133456318.

Hull, J. (2018). *Risk management and financial institutions.* Wiley Finance Series. Wiley. ISBN: 9781119448112.

Huynh, H. T., Lai, V. S. and Soumare, I. (2011). *Stochastic Simulation and Applications in Finance with MATLAB Programs [Elektronisk resurs]*. Hoboken: John Wiley & Sons. ISBN: 9780470722138.

Hyvärinen, A., Karhunen, J. and Oja, E. (2001). *Independent component analysis.* Adaptive and learning systems for signal processing, communications and control. Wiley. ISBN: 047140540X.

Jennrich, R. I. (2002). 'A Simple General Method for Oblique Rotation.' In: *PSYCHOMETRIKA* 1, p. 7. ISSN: 0033-3123. URL: https://link.springer.com/article/10.1007/bf02294706.

Jolliffe, I. T. (2010). *Principal component analysis.* Springer series in statistics. Springer. ISBN: 9781441929990.

Kaiser, H. F. (1960). 'The Application of Electronic Computers to Factor Analysis'. In: *Educational and Psychological Measurement* XX.1, pp. 141–151.

Litterman, R. B. and Scheinkman, J. (1991). 'Common Factors Affecting Bond Returns'. In: *The Journal of Fixed Income* 1.1, pp. 54–61. ISSN: 1059-8596. DOI: 10.3905/jfi.1991.692347. eprint: https://jfi.pm-research.com/content/1/1/54.full.pdf. URL: https://jfi.pm-research.com/content/1/1/54.

Nelson, C. R. and Siegel, A. F. (1987). 'Parsimonious Modeling of Yield Curves.' In: *The Journal of Business* 60.4, pp. 473–489. ISSN: 00219398. URL: https://www.jstor.org/stable/2352957.

Preacher, K. J. and MacCallum, R. C. (2003). 'Repairing Tom Swift's Electric Factor Analysis Machine.' In: *Understanding Statistics* 2.1, p. 13. ISSN: 1534844X. URL: https://www.tandfonline.com/doi/abs/10.1207/S15328031US0201_02.

Ross, S. M. (2010). *Introduction to probability models.* Academic Press. ISBN: 9780123756862.

Sklar, A. (1959). 'Fonctions de r´epartition 'a n dimensions et leurs marges'. In: *Publications de l'Institut de Statistique de l'Universit´e de Paris*, 8:229–231.

Swedbank (2019). *2019 Annual and Sustainability Report.* Accessed: 2021-02-08. URL: https://www.swedbank.com/investor-relations/reports-and-presentations/annual-reports.html.

Swedbank (2021). *Personal Interview.*

Wang., H. (2012). *Monte Carlo Simulation with Applications to Finance [Elektronic resource].* Chapman and Hall.

Yanev, G. and Ahsanullah, M. (2012). 'Characterizations of Student's t-distribution via regressions of order statistics'. In: *Statistics* 46.4. cited By 5, pp. 429–435. DOI: 10.1080/02331888.2010.535904. URL: https://www.tandfonline.com/doi/abs/10.1080/02331888.2010.535904?journalCode=gsta20.

Appendix A

Graphs of Decomposition Models

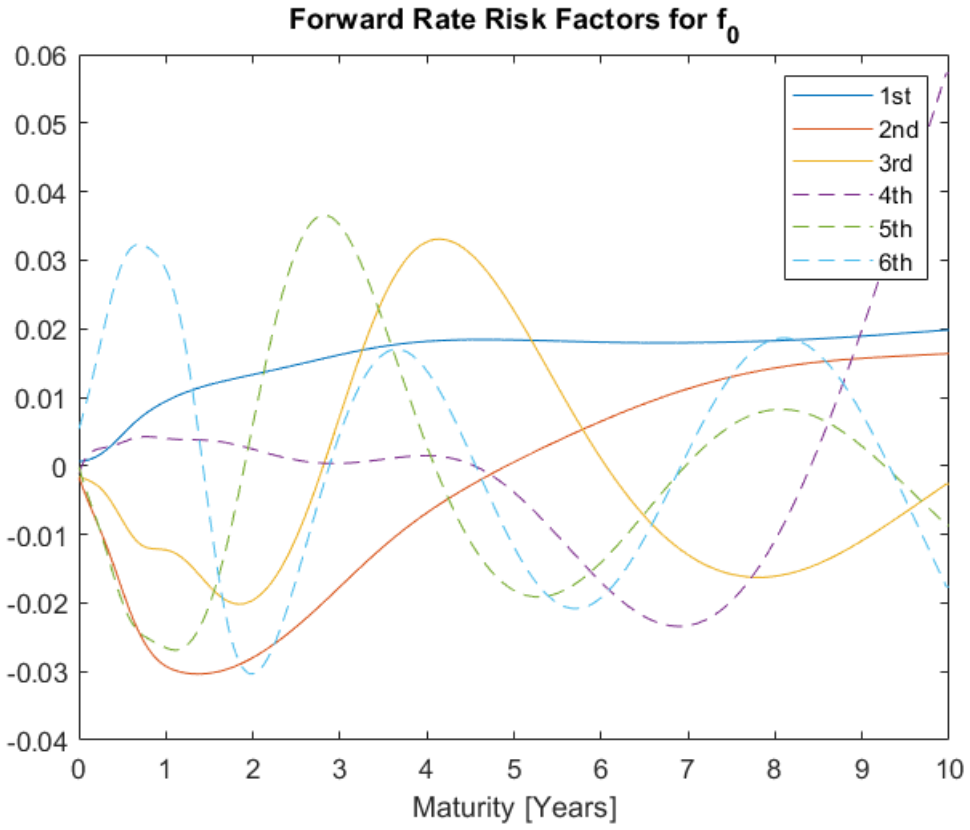


Figure A.1: Components from PCA, based on changes in the risk free forward rate curves.

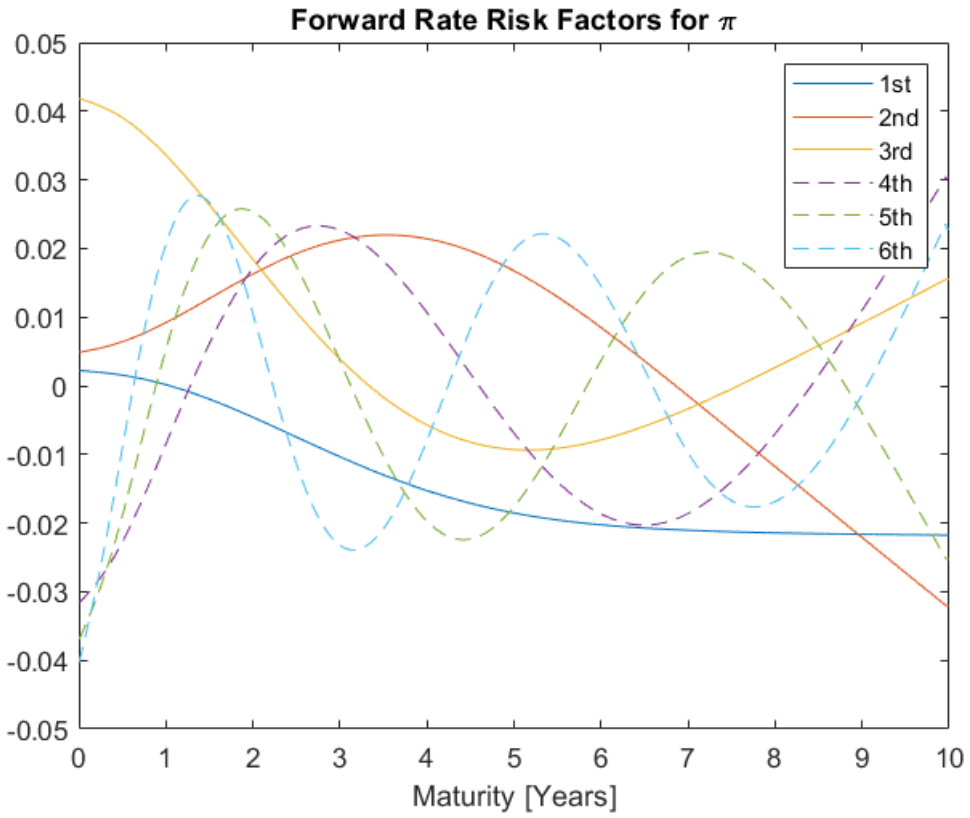


Figure A.2: Components from PCA, based on changes in the tenor premium forward rate curves.

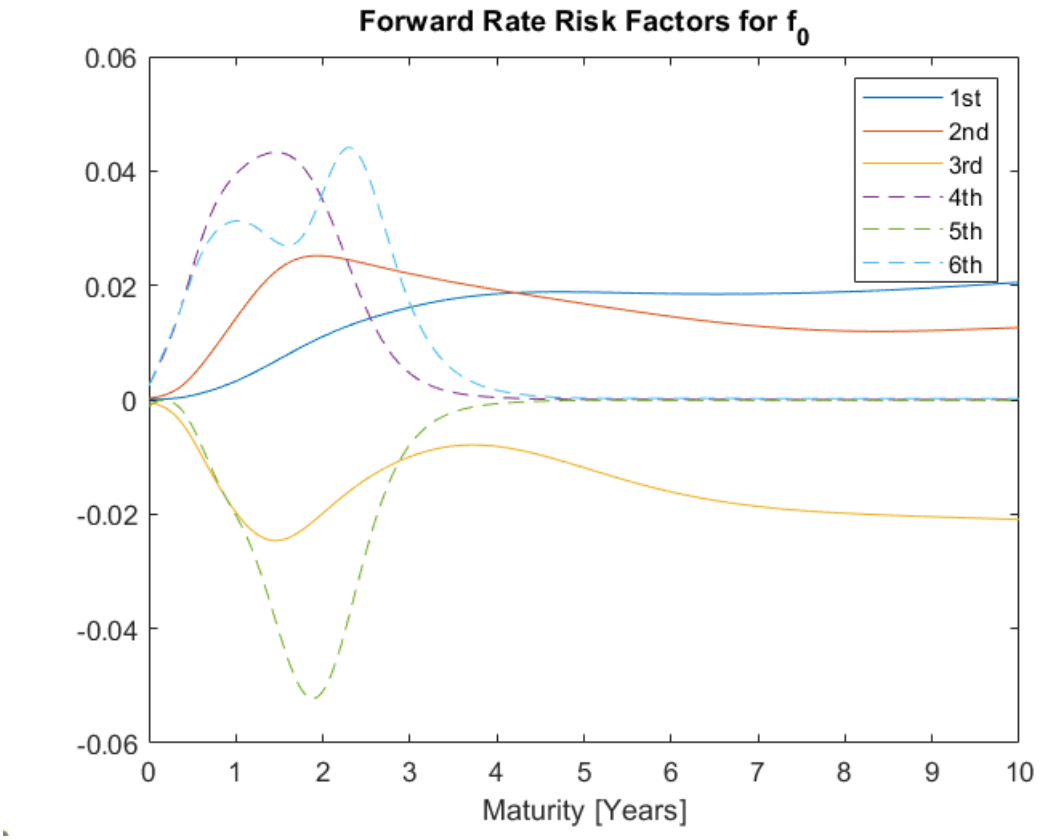


Figure A.3: Components from PCA as an optimisation problem, based on changes in the risk free forward rate curves.

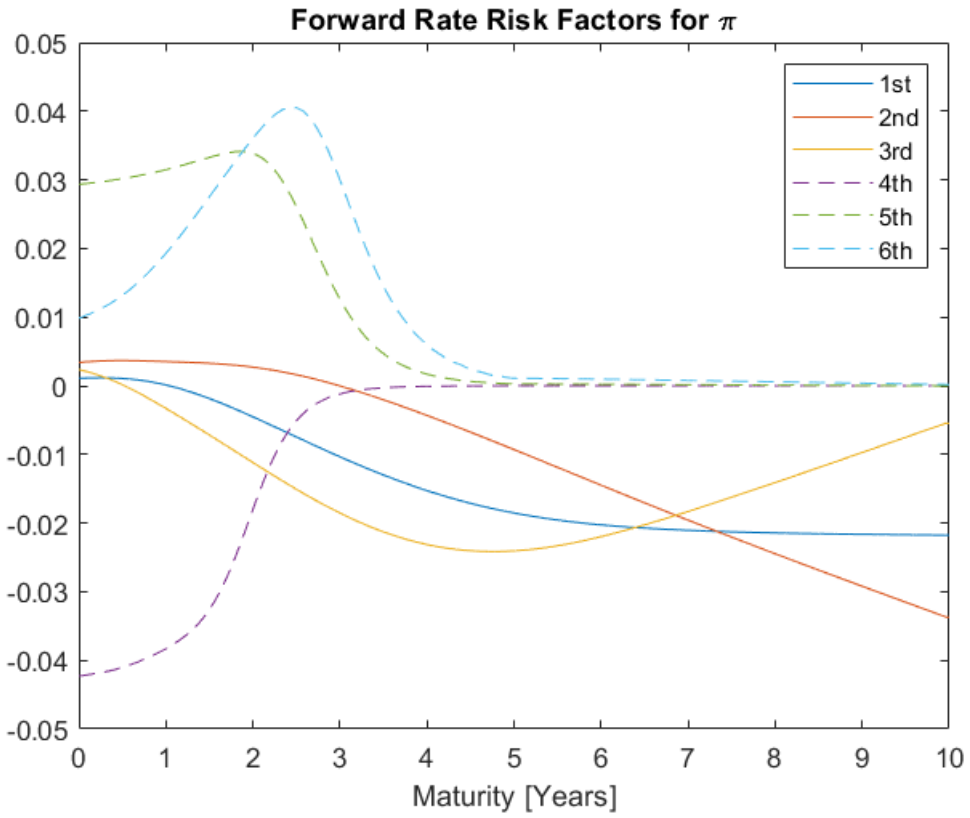


Figure A.4: Components from PCA as an optimisation problem, based on changes in the tenor premium forward rate curves.

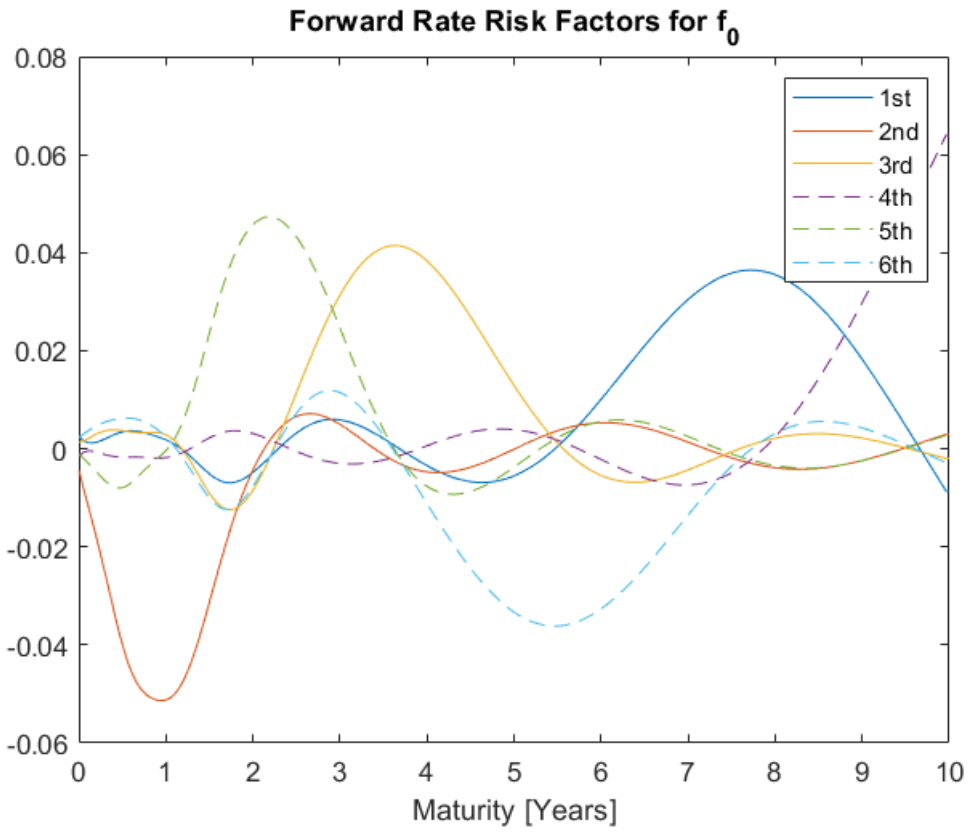


Figure A.5: Components from rotated PCA components using quartimin, based on changes in the risk free forward rate curves.

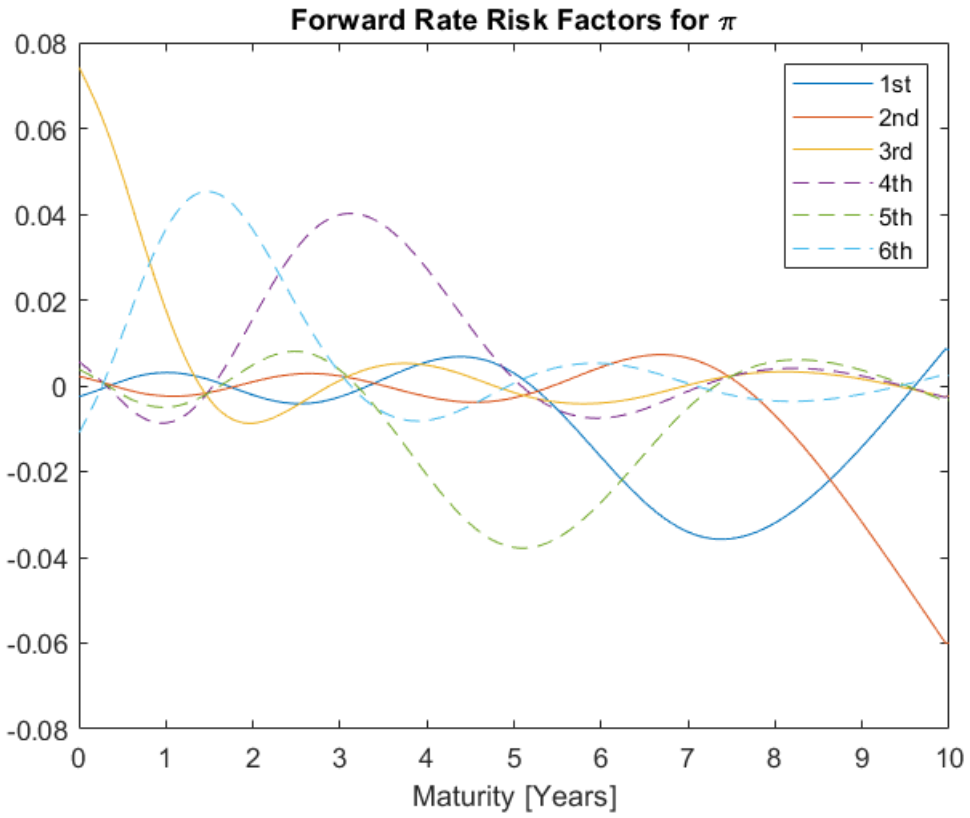


Figure A.6: Components from rotated PCA components using quartimin, based on changes in the tenor premium forward rate curves.

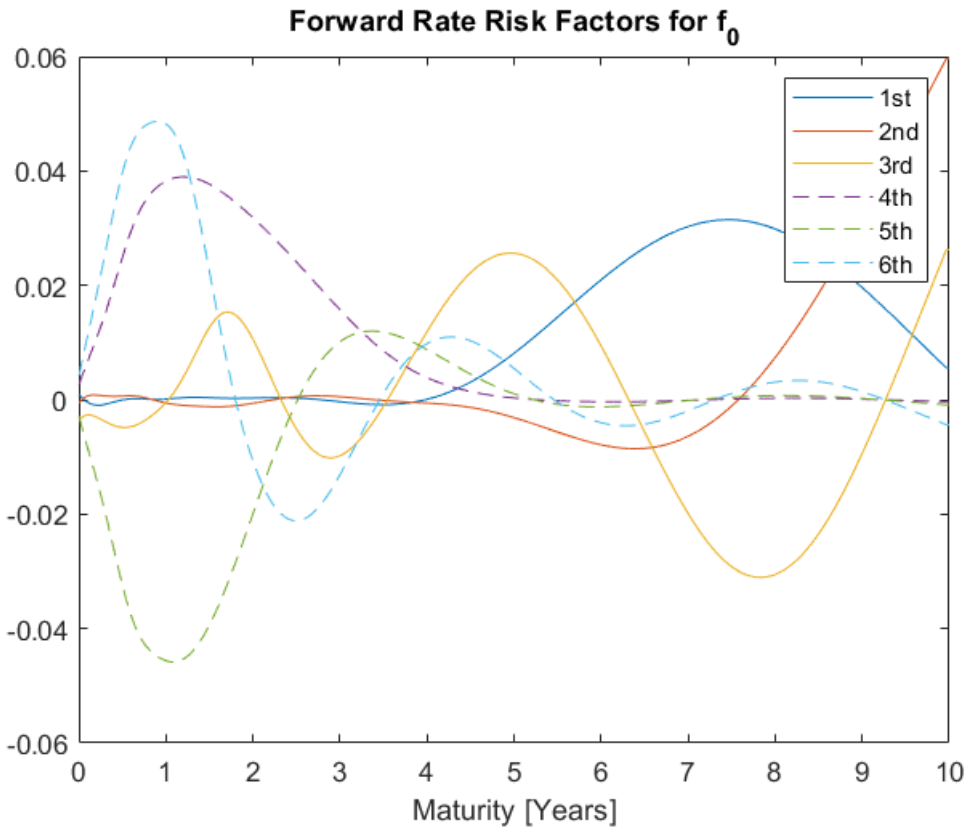


Figure A.7: Components from rotated PCA components using a Kernel, based on changes in the risk free forward rate curves.

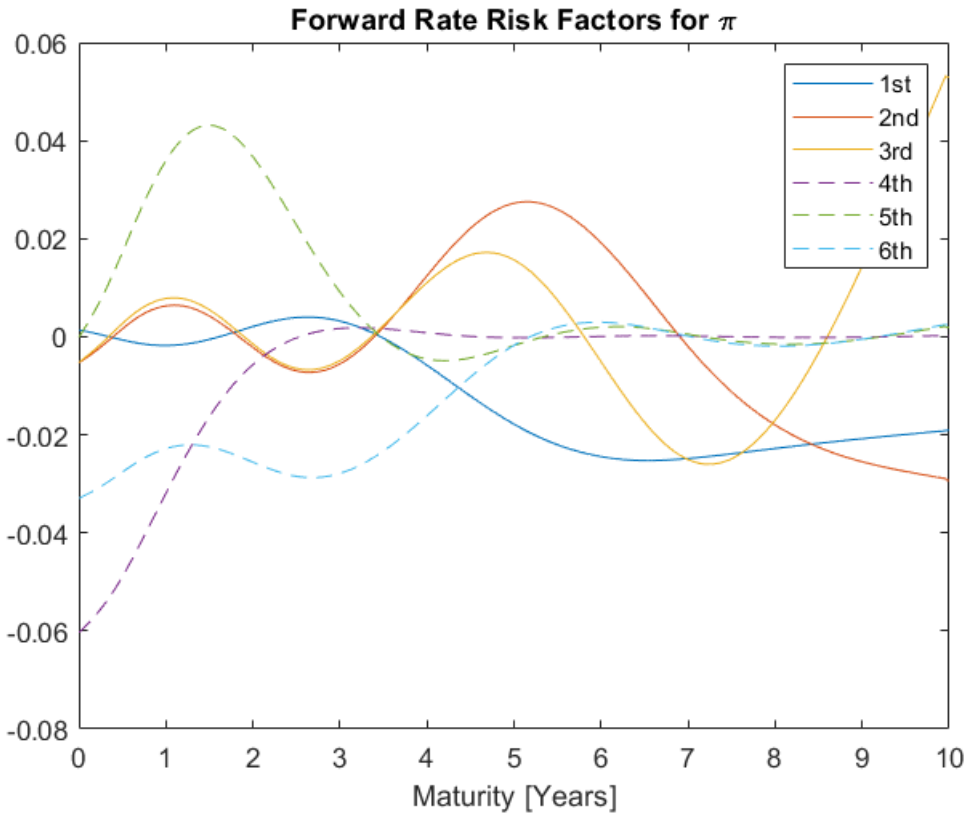


Figure A.8: Components from rotated PCA components using a Kernel, based on changes in the tenor premium forward rate curves.

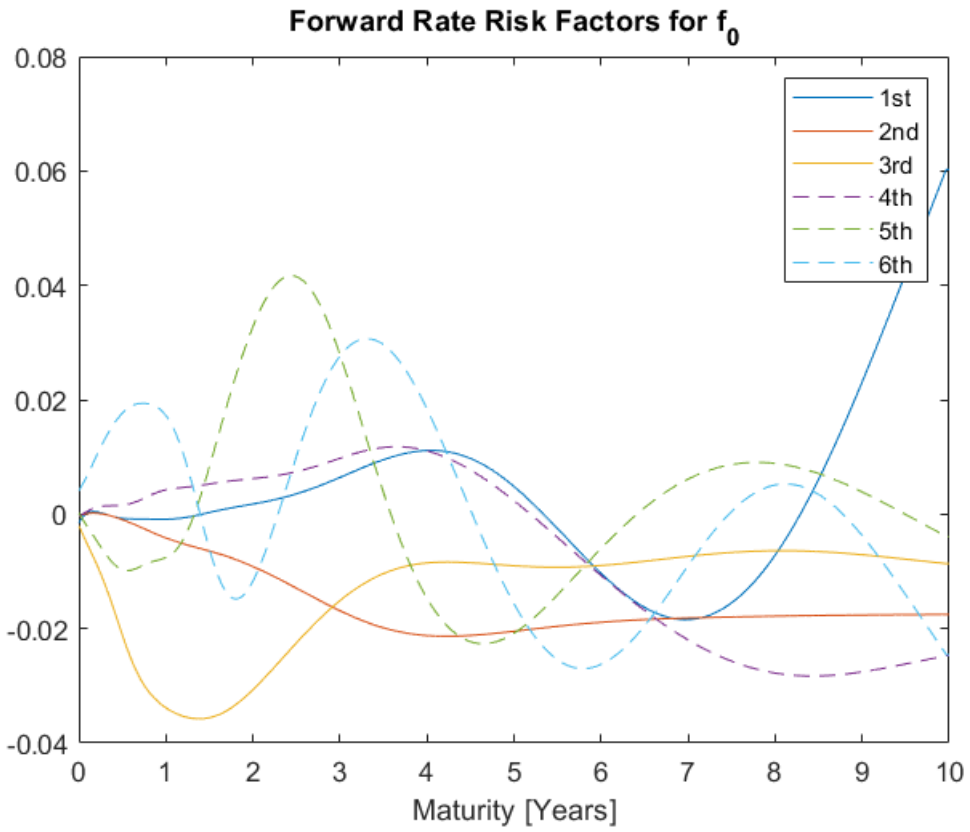


Figure A.9: Components from ICA symmetric, based on changes in the risk free forward rate curves.

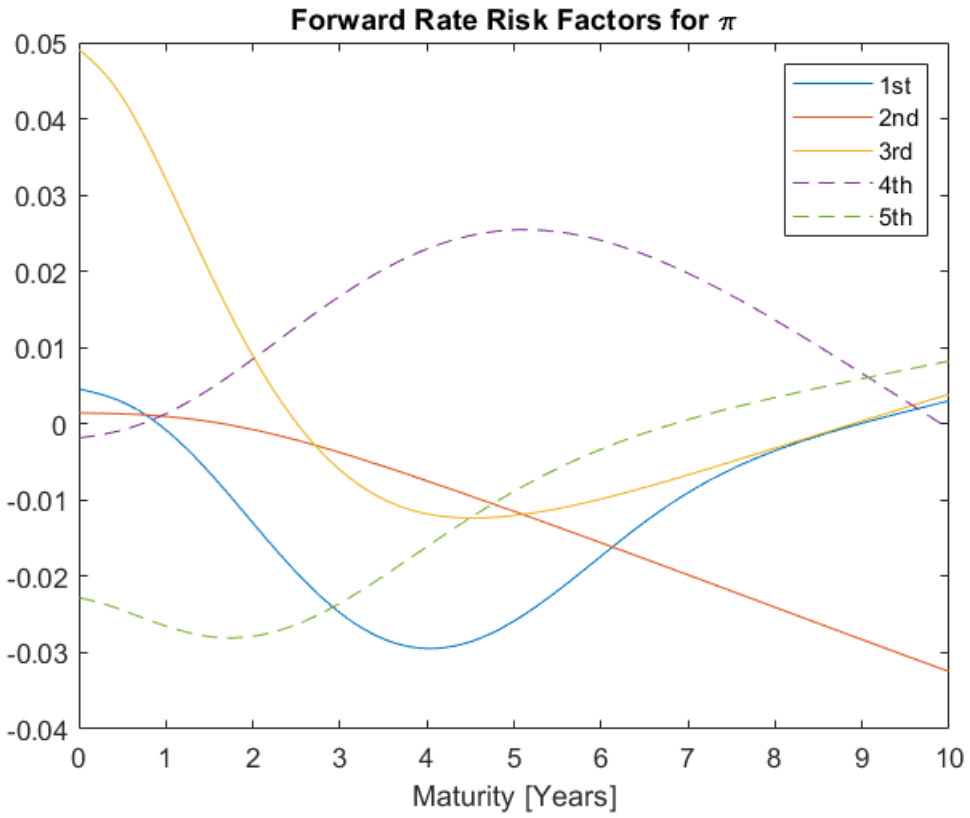


Figure A.10: Components from ICA symmetric, based on changes in the tenor premium forward rate curves.

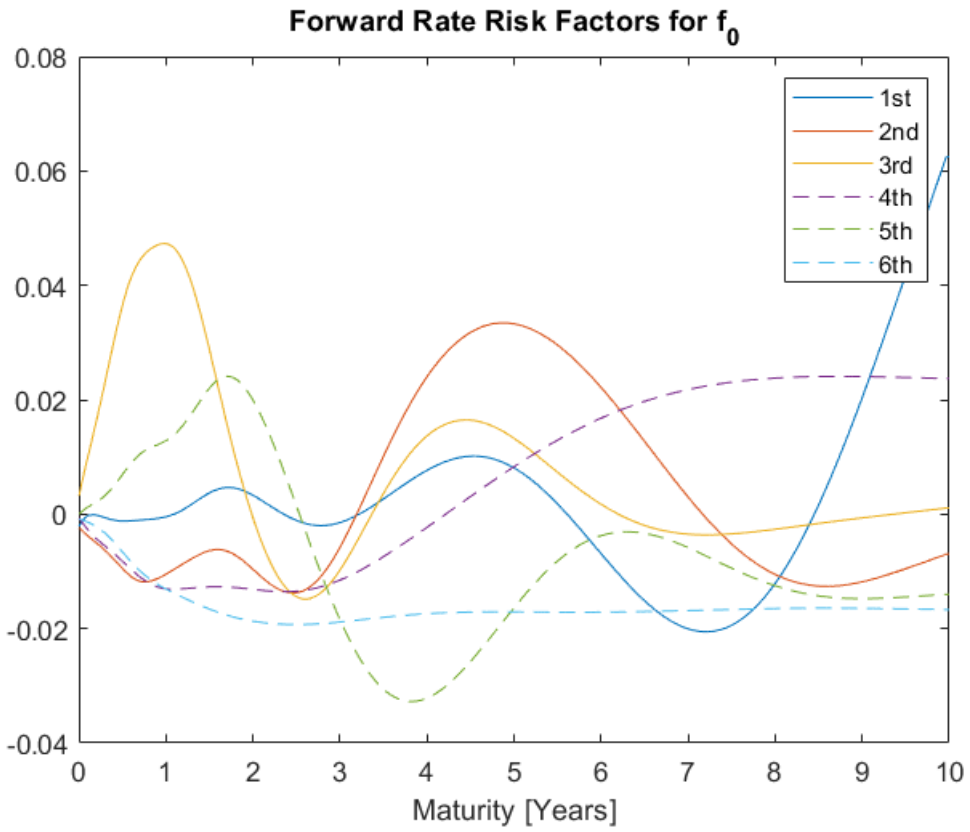


Figure A.11: Components from ICA deflation, based on changes in the risk free forward rate curves.

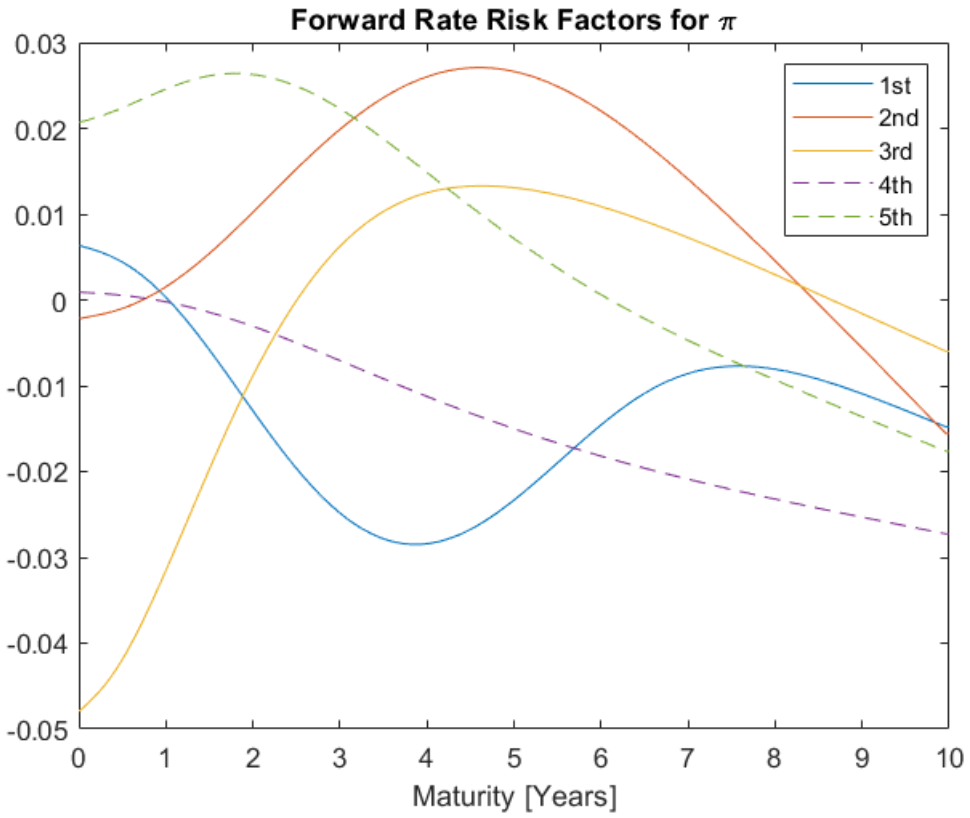


Figure A.12: Components from ICA deflation, based on changes in the tenor premium forward rate curves.