Karlsruher Institut für Technologie

Institut für Finanzwirtschaft, Banken und Versicherungen

Lehrstuhl Financial Engineering und Derivate
Prof. Dr. Marliese Uhrig-Homburg

Master Thesis

Filtered historical Simulation for Portfolios: Model Selection and Calibration

Author: Andreas Höcherl
Im Cäcilienbusch 3
53340 Meckenheim
E-Mail: a.hoecherl@googlemail.com

Karlsruhe, den 08. September 2015

Contents

List of figures **iii**

List of tables **iv**

1 Introduction and Motivation **1**

Glossary **1**

2 Theory and Literature Review **3**

2.1 Estimating Value at Risk based on Risk factors 3

2.2 Popular Value at Risk Models 4

2.2.1 Filtered historical simulation 6

2.3 Volatility Modeling 8

2.3.1 GARCH Models 9

2.3.2 Popular GARCH extensions 10

2.3.3 Exponentially weighted moving average as a special case of GARCH 12

2.3.4 Distribution and Autocorrelation tests 13

2.3.5 Modeling of the mean return 14

2.4 Maximum Likelihood Estimation 15

2.4.1 The Bayesian Information Criterion 18

2.4.2 Robust confidence intervals 19

2.5 Additional components of a Filtered historical VaR model 20

2.5.1 Correlation break adjustment 21

2.5.2 Liquidity adjustment 22

2.5.3 Measures to prevent procyclicality 23

2.6 Backtesting 26

2.6.1 Measurement of Backtesting Performance 27

2.6.2 Unconditional tests of backtesting performance 29

2.6.3 Conditional tests of backtesting performance 30

2.7 Risk factor based pricing of a Forward Rate Agreement 31

2.7.1 Using interest rate curves as risk factors 33

2.8 VaR estimation and backtesting of a Portfolio of Forward Rate Agreements - an example 35

2.8.1 Identifying and gathering relevant data 35

2.8.2 Choosing and calibrating a volatility model . . . 36
2.8.3 Calculating the current value at risk . . . 37
2.8.4 Backtesting the calibrated model . . . 38

3 Empirical Analysis **40**
3.1 Data used . . . 40
3.2 Individual Modelselection . . . 41
3.3 Simultaneous Modelselection . . . 48
3.3.1 Results of simultaneous EWMA-Modelselection . . . 50
3.4 The use of multiple volatility models across one discount curve . . . 55
3.5 Calibration results for an entire discount curve . . . 60
3.6 Backtests . . . 65
3.6.1 Backtests with a constant discount factor . . . 66
3.6.2 Backtests with calibrated volatility models . . . 68

4 Conclusion **76**

A Backtesting Performance **77**
A.1 Lag 1 based independence and conditional coverage . . . 77
A.2 Duration based independence and conditional coverage . . . 78
A.3 Autoregression based independence and conditional coverage . . . 80

B Empirical Appendix **81**
B.1 Choosing the length of the calibration window . . . 81
B.2 Seperating the impact of different discount curves in VaR estimation of a FRA . . . 84

Bibliography **88**

List of Figures

1 Interaction of filtered and stress value at risk in the calculation of Margin Requirement . . . 25
2 Dependency between tenor, degree of freedom and log likelihood . . . 47
3 Pseudocode Calibration of Portfolio . . . 51
4 LambdaForTenor . . . 54
5 The issue with using different volatility models for the same discount curve . . . 58
6 Time dependency of the use of different decay factors . . . 59
7 Impact of the use of a common decay factor on autocorrelation and distributional property . . . 61
8 Relation between Curve Specific λ and λ calibrated across multiple curves 63
9 Backtesting with different constant λ factors . . . 67
10 Exemplary backtest with three different volatility scenarios . . . 72
11 Backtest of short end CHF FRA . . . 74
12 Stability of Lambda over time for calibration periods of differing length and position . . . 82
13 Backtesting a FRA with disabled OIS discount risk factors . . . 85
14 Backtesting a FRA with disabled forward risk factors . . . 86

List of Tables

2 Information needed for a Lag 1 independence test 30

3 Time series used for analysis 42

4 Count of risk factors, for which a given model performs best according to BIC. Analysis made for different calibration periods. 45

5 Amount of risk factors, for which *EWMAT* or *GARCH11T* works best out of the two. 48

6 Comparison of some approaches to splitting up the risk factors into subsets. 55

7 Parametrizations of EWMA models with and without splitting risk factors of a common discount curve. The calibration period is June 2007 to June 2010. The λ calibrated for scenario one applies to all € risk factors even though it has been calibrated based only on those with a tenor above one year. 57

8 Teststatistics of backtests of € products 69

9 Teststatistics of backtests of $ and CHF products 71

10 Backtests of € products with only one risk factor 87

Glossary

ARCH	Autoregressive Conditional Heteroscedasticity.
BIC	Bayesian Information Criterium.
CCP	Central Counterparty.
DoF	Degree of Freedom.
EONIA	Euro OverNight Index Average.
EURIBOR	Euro InterBank Offered Rate.
EWMA	Exponentionally weighted moving average.
FHS	Filtered Historical Simulation.
FRA	Forward Rate Agreement.
FX	Foreign Exchange.
GARCH	Generalized Autoregressive Conditional Heteroscedasticity.
HS	Historical Simulation.
i.i.d.	independent and identically distributed.
IRS	Interest Rate Swap.
MLE	Maximum Likelihood Estimation.
OIS	overnight index swap.
pdf	Probability Density Function.
PnL	Profit and Loss.
VaR	Value at Risk.

1 Introduction and Motivation

This thesis has been written in cooperation with the Eurex Clearing AG in Frankfurt.

Eurex Clearing is a Central Counterparty and as such clears trades of derivatives between its customers, which are usually major financial institutions such as universal banks or large funds. Since Eurex inherits positions in case of a default it charges its customers so called margins to cover possible losses. These charges are based on value at risk estimates which are calculated using filtered historical simulation. Filtered historical simulation does estimate value at risk based on historic returns while correcting for higher or lower volatility in the markets. Therefore, volatility has to be modeled for margin calculation. Eurex has implemented a portfolio based margining approach, charging one single margin for the entire portfolio of a customer. This margin does take risk mitigating effects such as hedges into account.

Eurex offers a broad range of different products which can have indices, single stocks, bond prices, exchange rates or other products as underlying. Nevertheless all products use a common volatility method and even a common parametrization for products with similar underlyings. This is necessary to reduce complexity and operational risk. It is also suspected that the use of different volatility models for products which are underlying the same portfolio might have unwanted effects on the margin charged for the portfolio. However, the German regulator Bafin scrutinized the use of a common model and only few parametrizations during a review forcing Eurex to investigate the impact of this more thoroughly.

Value at risk is currently the most important financial risk measure. Despite its well-known shortcomings, which could be overcome by using a coherent risk measure such as the expected shortfall, value at risk is still the benchmark risk measure for regulators and no major financial institution can take the liberty not to calculate it.

In academia empirical studies regarding volatility modeling in value at risk estimation can be found to some extent, however the portfolios studied often only have one underlying to which the volatility model is fitted exactly. In the case of Eurex however value at risk is calculated for an entire portfolio, based on a range of different underlyings and it is unclear if it would be better to use a common volatility model for all underlyings or to use a customized one for each underlying.

This thesis therefore examines the differences in value at risk modeling with a common volatility model for the entire portfolio, individual volatility models for each underlying and increments in between those extremes. Portfolios are limited to interest rate products such as forward rate agreements and a portfolio can contain products based on different currencies.

Different volatility models will be taken to the test on a number of different underlyings. At first it will be examined, which volatility model perform best and how far parametrizations differ between the underlyings. In a seconds step it will be examined, which underlyings can be grouped together, while still maintaining good Value at risk estimates and it will be examined if a too granular differentiation of volatility models might actually do more harm than good. Volatility models are not only chosen based on maximum likelihood estimation but extensive backtesting will be done, to assess the quality of value at risk estimates.

The goal of this thesis is to study empirically how many different volatility models are necessary to calculate the value at risk of a portfolio.

2 Theory and Literature Review

2.1 Estimating Value at Risk based on Risk factors

The introduction simplified the process of Value at Risk (VaR) estimation by implying that the VaR of a derivative can be calculated based on the current price of its underlying. This simplification has been made, since the definition of an underlying is well-known while the more appropriate concept of a risk factor is not.

To familiarize the reader with the concept of risk factors the VaR estimation of a derivative using historical simulation will be described up to a certain detail.[1] Historical simulation is used in this example, since it is the simplest method, but the concept of risk factors applies to any of the VaR models described in section 2.2.

A stock option has only one underlying - the stock it is based upon. However, for pricing a stock option other information than the price of its underlying is needed. The Black-Scholes-Merton model needs the volatility of the underlying, maturity, risk free rate and strike in addition to the current price of the underlying to price a vanilla put or call option[2]. All of the stochastic inputs are called risk factors. Therefore risk factors of a vanilla put are the current price of the underlying, the volatility of the underlying and the risk free rate. The strike is constant and the maturity can be determined deterministically for any point in the future. The maturity and strike therefore aren't risk factors.

For VaR estimation with historical simulation the VaR is a quantile of the historic returns. For example based on a 500 day history of Adidas returns the 99% VaR of the Adidas stock is the 5$^{\text{th}}$ largest loss of the Adidas stock in the last 500 days. Estimating the VaR of an option on Adidas is not as straight forward. To calculate the VaR of a call option with strike 70, expiring in 13 days in the same way we would need a 500 day history of this exact product. While a call with a strike of 70 might be continuously traded it's probably impossible to find the Profit and Loss (PnL) for one expiring in exactly 13 days for each day in the 500 day history. Call options on Adidas emitted by Deutsche Bank for example expires on the third Wednesday each month[3]. Assuming the third Wednesday is the 14$^{\text{th}}$ then the PnL of an option with an expiry of 13 days can only be measured on the 1$^{\text{st}}$.

[1] In the context of VaR estimation the term risk factor is used in the same fashion e.g. by Butler and Schachter (1997) or Pritsker (2006)

[2] It is assumed, that the stock doesn't pay dividend. Compare Hull (2006, Chapter 14)

[3] Compare boerse.de (2015)

Therefore a PnL time series of this exact call is not available and the VaR of an option is not calculated based on the PnL time series of the option but the time series of its risk factors. With this approach a synthetic time series of end of day prices of the desired option can be calculated. For each of the 500 days the option price is calculated using the Black-Scholes Merton model with the desired strike and maturity - in our example a constant 70 and 13 days - and the observed underlying prices, implied volatilities and risk free rates of the respective day.

A pricing algorithm, such as the Black-Scholes-Merton model is used to transform the risk factor PnLs of the last 500 days into 500 PnLs of the option - a historical PnL distribution of the wanted product. The 1% quantile of this distribution then is the 99% VaR of an Adidas call with strike 70, which expires in 13 days.

All VaR models mentioned within the course of this thesis work with this process. They create a time series or distribution of the risk factor returns and then use pricing algorithms to transform the risk factor returns into a PnL distribution of a derivative. However, the VaR models differ in their way of creating risk factor returns.

Estimating the VaR of a portfolio is as simple as summing up the prices of each product across one scenario. In case of the exemplary historical simulation each one of the past 500 days is a scenario. Assuming that the portfolio would contain additionally a short Adidas-Future and that the described Call-Option would be deep in the money, the VaR would be considerably lower, than that of the Call option alone, since the portfolio would be approximately delta hedged.

2.2 Popular Value at Risk Models

The three common VaR methodologies are Monte Carlo Simulation as a sophisticated parametric approach, Delta-Gamma approach as a less sophisticated parametric approach and the already briefly introduced Historical Simulation as a non-parameteric approach.[4] Finally, filtered historical simulation is an extension of historical simulation, which corrects for volatility clustering.

For a Monte Carlo Simulation return distributions and correlations of different risk factors have to be estimated. Based on these estimations multidimensional copula of the joint risk factor returns can be created from which samples for the simulation can be drawn. An example would be parametrizing two student-t distributions for the returns of the DAX and a government bond as well as their correlation. With this information a

[4] Compare Jorion (1997)

two dimensional student-t copula can be created. Given a scenario of joint movements of these risk factors the prices of derivatives based on these elemental instruments can also be calculated with pricing algorithms. By calculating and adding up scenario prices of all instruments in a portfolio the scenario price of the portfolio can be calculated. If one now reiterates the draw and calculation process many times a return distribution is created. Quantiles of this distribution - which can be interpreted as its VaR - and tail properties like the expected tail loss can be estimated very precisely, if enough computational power is available. The time horizon for Monte Carlo simulation is also perfectly flexible. If instead of a one day VaR a one year VaR is required this can be created by simply chaining 250 daily scenarios and relabeling these as one yearly scenario. Given a strong computer and a correct model yearly VaR can be calculated this way without using the square-root-of-time-rule.

The square-root-of-time-rule is a common approach to translate high frequency VaRs to a lower frequency.[5] By simply multiplying a one day VaR by $\sqrt{T}$ a T-day VaR is estimated. However this approach assumes a normal return distribution - an assumption, which is usually not fulfilled. Therefore estimating T-day returns by chaining one day returns, as is possible with a Monte Carlo Simulation even for large T is the superior approach in comparison with the square-root-of-time-rule.[6]

A simpler parametric approach is the delta-gamma approach, that was popularized by J.P. Morgans RiskMetrics in the 90s.[7]. It assumes that risk factors follow a conditionally normal multivariate distribution with the mean percentage change of each variable being zero. Under the assumption that products have a linear dependency with their underlying - which is correct for a future but not for a call - normally distributed underlying returns result in normally distributed portfolio returns. This methodology is capable of estimating the VaR of portfolios of linear products with acceptable precision based only on the covariance matrix of its risk factors. To compensate for some of the models inadequacy regarding non-linear models squared risk factor returns can be added as an additional risk factor.[8] Nevertheless the delta-gamma approach has been largely replaced by the better working Monte Carlo or filtered historical simulation.[9]

Being a non-parametric approach Historical Simulation (HS), which was used ex-

[5]Compare e.g. Hull and White (1998)
[6]Danielsson and Zigrand (2006)
[7]Compare Morgan (1996)
[8]Compare Hull and White (1998) and Jorion (1997)
[9]Compare Gurrola-Perez and Murphy (2015)[Chapter 2] - in this paper as well as many other places the Delta-Gamma approach is referred to as RiskMetrics.

emplary in section 2.1, on the other hand does not require an assumption of return distributions or correlations. Instead, every single day of a time series of historic returns serves as a scenario. Given for example a 750 day history and a 99% quantile for the VaR, the VaR will be the 7^{th} largest loss in the last three years.[10] The largest issue of historical simulation becomes directly apparent here. Even with a three year history only about ten days with large losses are relevant for calculating the VaR. Therefore the estimation can be very instable if a large loss enters or exits the 750 day time window. Additionally the statistical significance might be questionable given the low number of relevant samples. This issue is even worse, when regarding a VaR of an even higher quantile, such as 99.9% or measures that focus on the tail beyond a certain quantile like the expected shortfall. When calculating VaR for a longer time horizon it is almost mandatory to make use of the imperfect square-root-of-time procedure, since the sample size decreases drastically if a longer time window has to be regarded.

Another issue is the implicit assumption of homoscedasticity of historical simulation. Historical Simulation assumes that the past is a good indicator for tomorrows return and that the market conditions of the past are the same as today. However, it has been shown, that volatility clusters in financial time series and therefore the volatility on the day of estimating VaR might be higher or lower than on any given day in the last 750 days.[11] Filtered historical simulation is an approach based on historical simulation which mitigates this shortcoming and incorporates volatility clustering into the model.

2.2.1 Filtered historical simulation

At the end of the 90s Giovanni Barone-Adesi and Kostas Giannopoulos came up with the idea of combining the techniques of historical simulation and volatility modeling to create the so called filtered historical simulation.[12] Hull and White (1998) published a very similar idea at the same time.

A few years later Barone-Adesi and Giannopoulos were able to show, that their model outperformed more traditional VaR models on productive portfolios of commercial banks.[13] Since then Filtered Historical Simulation (FHS) developed into the most popular tool for VaR estimation in practice.[14]

[10]Since $750 * 1\% = 7.5$ it's actually the mean between the 7^{th} and 8^{th} largest loss.
[11]Compare e.g. Bollerslev et al. (1992)
[12]Compare Barone-Adesi et al. (1999) also published as a shorter version Barone-Adesi et al. (1998)
[13]Compare Barone-Adesi et al. (2002)
[14]Compare Gurrola-Perez and Murphy (2015)

As described in the previous section a weakness of historical simulation is its assumption of constant volatility. FHS tries to remedy this weakness of HS while retaining its strengths. Just like HS no return distributions need to be assumed and no variance covariance matrix is required. Instead a quite accurate VaR of even complex , non-linear products can be calculated with only a few assumptions.[15] The fact that only few assumptions need to be made makes it relatively simple to defend the approach when regulators are reviewing the risk methodology, since every additional assumptions needs to be backed up with solid empirical and theoretical evidence. Nevertheless, regulators have published papers pointing out potential weaknesses of FHS in the more recent past.[16]

The general idea of FHS is to divide realized returns by their estimated standard deviation to yield a time series with constant standard deviation and afterwards scale this filtered time series up with the currently estimated standard deviation.
Assuming that

- r_j is the return time series of risk factor j with length T, which shows volatility clustering - it therefore is not stationary
- σ_j is the corresponding time series of standard deviations for each day

a stationary time series r_j^* with a constant σ_j=1 can be created with

$$
\begin{aligned}
r_{t,j}^* &= \frac{r_{t,j}}{\sigma_{t,j}} \text{ for } t = 1, \ldots, T. \\
r_j^* &= \begin{pmatrix} r_{1,j}^* \\ \vdots \\ r_{T,j}^* \end{pmatrix}
\end{aligned}
\tag{1}
$$

Given a desired standard deviation σ^* a scenario time series r_k^{scn} recreating the price movements of r_j under the constant volatility σ^* can be created as

$$r_j^{scn} = \sigma^* \cdot h_j^*.$$

Using this alternative time series r_j^{scn} as the return time series of risk factor j instead of r_j is the sole difference between HS and FHS. Afterwards VaR is estimated

[15] Given that a long time series of all risk factors are available, which can proof difficult, when calculating VaR of a new product.

[16] Compare Gurrola-Perez and Murphy (2015), Murphy et al. (2014b) and especially Pritsker (2006).

in exactly the same way as it is done for HS which was described in section 2.1 and section 2.2. In vectorized form, the time series r_j^{scn} can be created as[17]

$$r_j \,./\, \sigma_j \cdot \sigma_{N,j} = r_j^{scn} \ , \tag{2}$$

where N is the final day of the time series, for which we want to estimate the VaR. The biggest challenge of FHS is therefore the correct estimation of the time series of standard deviations σ_j, which will be described in detail in section 2.3.

2.3 Volatility Modeling

Modeling returns is one of the primary goals of financial time series research and one of its generally accepted results is, that the standard deviation of any assumed distribution has to vary in time to yield reasonable results.[18] While section 2.2.1 did establish, how this insight may be incorporated in the estimation of VaR this section closer investigates different means of modeling volatility.

The fundamental difference between modeling returns and volatility is that return realizations can be measured directly, while realizations of volatility may only be measured indirectly under certain model assumptions. The daily PnL of a stock is a realization of its return distribution, the volatility of this stock however is not easily measurable for the same day. Nevertheless, the volatility of a stock is a critical information not only for FHS-VaR models but especially for the pricing of derivatives based on the stock. For pricing purposes implied volatilities are used. Given a pricing algorithm which uses the volatility of the underlying as an input $P(\sigma_U, \theta)$, where θ is the set of other inputs and the current market price p of a derivative, it is possible to numerically determine a standard deviation ${\sigma_U}^*$ such that

$$P({\sigma_U}^*, \theta) = p.$$

An issue with these implied volatilities is, that the prices of different derivatives based on the same underlying may imply different volatilities of the underlying due to inaccurate assumptions of the pricing algorithms. The underlying, however can obviously only have one volatility at any given time. This drawback is accepted by banks, since their primary priority is to accurately calculate market prices rather than

[17] ./ is used as the symbol for element-wise division.
[18] Compare e.g. Bollerslev et al. (1992)

making sure that model assumptions are fulfilled [19].

For application on FHS-VaR models, only one volatility per risk factor is needed and it is important that model assumptions hold even in extreme market conditions. Therefore implied volatilities are not used to scale historic returns up and down but volatilities are modeled instead[20].

2.3.1 GARCH Models

The most common way to calculate realized volatility is by calculating the sample variance of a random variable X as

$$\sigma^2 = \frac{1}{T}\sum_{t=1}^{T}(x_t - \mu)^2, \tag{3}$$

where x_t are realizations of X and μ is the sample mean.

This assumes, that all realizations x_t origin from i.i.d. distributions with the same variance. However, there might be evidence, that different x_t in fact origin from different distributions Y_t instead of a common distribution X. We therefore have a time series of realizations $[x_1, \dots x_T]$, which origin from a series of distributions $[Y_1, \dots Y_T]$ each of which has its own standard deviation $[\sigma_1, \dots \sigma_T]$. To estimate a specific σ_k the only directly connected information is the return x_k. However when $[x_1, \dots, x_T]$ is a time series of returns it is reasonable to assume, that previous returns $[x_1, \dots, x_{k-1}]$ might carry valuable information for estimating σ_k. It is also reasonable to assume, that more recent returns like x_{k-1} could carry more information to estimate σ_k than returns long past, such as x_1. Based on these basic thoughts Engle (1982) developed so called Autoregressive Conditional Heteroscedasticity (ARCH) models and received the Nobel memorial prize in economic sciences for his work in 2003.

With the additional assumption, that $[Y_1, \dots Y_T]$ arise from the same distributional family and their parametrization only differs in the standard deviation, a time series

[19] For this reason banks create volatility surfaces for each underlying. These surfaces contain the implied volatilities as a function of the time to expiry and moneyness of the derivative which implied them. Storing these surfaces for a number of underlyings and at high frequency is expensive and extremely data intensive but without alternative to calculate accurate prices.

[20] In fact, volatility surfaces are needed nonetheless. While risk factors are scaled up and down using modeled volatilities, these risk factor scenarios are afterwards used as input for pricing algorithms to calculate the scenario value of the portfolio in question. These pricing algorithms often use volatility surfaces to price more complex derivatives. Backtesting might therefore require historical volatility surfaces to be available.

following an ARCH model is defined as

$$
\begin{aligned}
x_t &= \sigma_t \epsilon_t \\
\sigma_t^2 &= \omega + \sum_{p=1}^{P} \alpha_p x_{t-p}^2 \ ,
\end{aligned}
\tag{4}
$$

where ϵ is a series of independent and identically distributed (i.i.d.) random variables[21] with $E(\epsilon_t) = 0$ and $\sigma(\epsilon_t) = 1$ for all t and $[\omega, \alpha_1 \dots \alpha_P]$ are non-negative parameters. The value P indicates how far the ARCH model looks backwards for volatility estimation and since an individual parameter a_p is needed for each lag P tends to be relatively low in practice.

Bollerslev (1986) enhanced the work of Engle with his Generalized Autoregressive Conditional Heteroscedasticity (GARCH) models. He added a recursive component to the model and therefore estimates σ not only based on past returns but also on previously estimated σ. A GARCH model is defined as

$$
\begin{aligned}
x_t &= \sigma_t \epsilon_t \\
\sigma_t^2 &= \omega + \sum_{p=1}^{P} \alpha_p x_{t-p}^2 + \sum_{q=1}^{Q} \beta_q \sigma_{t-q}^2 \ .
\end{aligned}
\tag{5}
$$

A significant advantage of a GARCH model is its simpler parametrization. A GARCH(1,1)[22] model for example requires only three parameters ω, α_1 and β_1 but due to it recursive character with

$$
\begin{aligned}
\sigma_t &= f(x_{t-1}, \sigma_{t-1}) \\
\sigma_{t-1} &= f(x_{t-2}, \sigma_{t-2}) \\
&\vdots
\end{aligned}
$$

still contains information from the entire time series.

2.3.2 Popular GARCH extensions

Over the course of the last few years several extensions of GARCH models have been developed, which gained popularity for modeling financial time series volatility.

[21] These are often assumed to be normally distributed, but don't have to be. The distributional properties of ϵ are further investigated in section 2.3.2.

[22] GARCH(1,1) refers to a GARCH model with $P = Q = 1$ and is by far the most common GARCH model.

E-GARCH models as introduced in Nelson (1991) with their logarithmic structure and GJR-GARCH models as introduced in Glosten et al. (1993) with their leverage parameter allow for the consideration of the leverage effect in volatility estimation.

GJR-GARCH models assume the volatility process

$$\sigma_t^2 = \omega + \sum_{i=1}^{P} \beta_i \sigma_{t-i}^2 + \sum_{j=i}^{Q} \alpha_j \epsilon_{t-j}^2 + \sum_{j=1}^{Q} \zeta_j I[\epsilon_{t-j} < 0] \epsilon_{t-j}^2$$
$$\text{where} \quad I[\epsilon_{t-j} < 0] = \begin{cases} 1 & \text{if } \epsilon_{t-j} < 0 \\ 0, & \text{otherwise} \end{cases} \tag{6}$$

and E-GARCH models assume

$$\log \sigma_t^2 = \omega + \sum_{i=1}^{P} \beta_i \log \sigma_{t-i}^2 + \sum_{j=i}^{Q} \alpha_j \left[\frac{|\epsilon_{t-j}|}{\sigma_{t-j}} - \sqrt{\frac{2}{\pi}} \right] + \sum_{j=1}^{Q} \zeta_j \left(\frac{\epsilon_{t-j}}{\sigma_{t-j}} \right) \tag{7}$$

if innovations are assumed to have a Gaussian distribution.

The so called leverage effect, which these models regard, bases on the empirical observation, that a negative PnL tends to increase volatility more than a positive PnL of the same magnitude. This effect was assumed to descent from the fact that a loss of one euro is larger in relation to the remaining market capitalization than a profit of one euro and was therefore called leverage effect. However it has been empirically found, that the leverage parameter in GJR-GARCH models is too significant to be attributed solely to the shift in market capitalization. The full reason for the good performance of E-GARCH and GJR-GARCH models might therefore be more complex, than originally assumed. For practitioners however it seems very intuitive that markets would react stronger to downward movements than to upward movements and GJR-GARCH models therefore gain popularity in practice.

Nevertheless GJR-GARCH models are a rare exception in this regard. Dozens of extensions of GARCH models have been proposed over the course of the last three decades and most of them faded away barely noticed, while practitioners stuck to their known and tested GARCH(1,1) models. With large empirical effort Hansen and Lunde (2005) took some of these extensions to the test and concluded, that despite the large effort put into developing more sophisticated GARCH models almost none perform significantly better than a simple GARCH(1,1) model. Hansen and Lunde found, that no GARCH extension was able to consistently outperform a GARCH(1,1) model for estimating volatility of an Foreign Exchange (FX) rate, while extensions accounting for a leverage effect such as E-GARCH or GJR-GARCH were able to outperform a GARCH(1,1) model in modeling volatilities of stock returns. This result seems logical, since FX rates do not have a market capitalization and a drop in the EUR-USD FX rate corresponds to a rise in the USD-EUR FX rate, which makes it tough to argue, why the market should react stronger towards negative price movements than to positive price movements.

2.3.3 Exponentially weighted moving average as a special case of GARCH

Exponentionally weighted moving average (EWMA) models, were originally named Integrated GARCH (IGARCH) and have been first presented by Engle and Bollerslev (1986). Theoretical results for IGARCH models are of large practical value, since EWMA models are frequently used by banks.[23] Their popularity has two reasons: On the one hand an EWMA can be defined with as few as a single parameter. On the other hand their popularity is largely owed to the historic development of VaR. JP Morgan introduced the VaR parametric first and sold their approach as the RiskMetrics framework to all major banks in the 90s.[24] Since RiskMetrics used EWMA for volatility estimation EWMA quickly became the market standard.

EWMA is a special case of a GARCH model, as presented with the additional constraint of $\alpha_i + \beta_i = 1$ and therefore a parametrization of a GARCH model might very well result in an IGARCH model. Usually - when compared to eq. (5) - calibrations of EWMA models result in $\omega = 0$, since the volatility would constantly rise otherwise and they usually use $P = Q = 1$. Since $\alpha + \beta = 1$ they are usually reformulated as $\alpha = 1 - \lambda$ and $\beta = \lambda$. This results in the following definition of an EWMA model:

$$\begin{aligned} x_t &= \sigma_t \epsilon_t \\ \sigma_t^2 &= (1-\lambda)x_{t-1}^2 + \lambda\sigma_{t-1}^2 \ . \end{aligned} \tag{8}$$

λ is often called *decay factor* since it determines the speed at which the impact of previously estimated σ decays. In academics IGARCH models tend to prove especially useful in use for high frequency data.[25]

[23]Compare Pafka and Kondor (2001)
[24]Compare Morgan (1996)
[25]Compare Andersen and Bollerslev (1996) and Beltratti and Morana (1999)

2.3.4 Distribution and Autocorrelation tests

The issue with HS is that it does not adjust for volatility clustering. This means that there are market regimes in which volatility is higher and such in which volatility is lower. This suggests in turn that high returns - in absolute terms - should be followed by other high absolute returns and vice versa. This property is measured by examining the autocorrelation of absolute or squared returns of the time series.

As described in section 2.2.1 the idea of FHS is to standardize historic returns to unit standard deviation. When $\mu = 0$ is assumed realized volatilities are measured as squared returns. Unlike their unfiltered counterparts the time series of filtered realized volatilities should show no or close to no autocorrelation. Since the entire idea of FHS is to correct for volatility clustering when calculating VaR it is absolutely crucial, that the squared filtered returns show significantly less auto correlation than unfiltered ones. If the filtered time series still shows significant autocorrelation it has to be concluded, that the process used for modeling volatility is inadequate.

One of the most popular tests for autocorrelation is the Ljung-Box Q-test and has been introduced in Ljung and Box (1978). It is tested, if the hypothesis that a time series does not show autocorrelation has to be rejected. It is portmanteau test, meaning that a Ljung-Box Q-test for lag 10 does not only test for independence with regard to the 10^{th} lag but for all lags ≤ 10. The test uses the sample autocorrelations up to the specified lag. The autocorrelation for lag k of a time series x is defined as[26]

$$\begin{aligned} ac_k &= \frac{c_k}{c_0} \\ \text{where} \qquad c_k &= \frac{1}{T-1} \sum_{t=1}^{T-k} \left(x_t - \bar{x}\right)\left(x_{t+k} - \bar{x}\right) \end{aligned} \tag{9}$$

where c_0 is the sample variance as it has been defined in eq. (3). Based on this autocorrelation the test statistic of a Ljung-Box Q-test up to a maximum lag of L is

$$Q = T(T+2) \sum_{k=1}^{L} \left(\frac{ac_k^2}{T-k} \right)$$

and is approximately X^2 distributed with L degrees of freedom under the null hypothesis[27]. If the test statistic exceeds a quantile of the X_L^2 distribution the null hypothesis has to be rejected on the level of this quantile.

[26]Compare Box et al. (2011)
[27]Compare Ljung and Box (1978)

The model assumptions of a GARCH model as introduced in eq. (5) can be tested by means of the error term ϵ. During parametrization of a GARCH model ϵ is either assumed to be normally or student-t distributed and parameters are optimized to result in error terms which are as close as possible to the desired distribution. Nevertheless 'as close as possible' might still be way off the actual mark simply because the assumed volatility process is incorrect for the time series it is used upon. Therefore distributional tests have to be made to ensure, that the resulting ϵ resemble the desired distribution. Testing, if an empirical sample could origin from a target distribution is a quite common task and therefore will not presented in much detail.

Within the scope of this thesis it is most common to test, if a sample could origin from a specified student-t distribution. For this reason well known tests like the t-test[28] or the Jarque-Bera test[29] are not suited, since they can only test against a normal distribution. This thesis will exclusively make use of the one sample Kolmogorow-Smirnow test, whose test statistic is simply the largest distance between the theoretical, desired and the empirical, measured cumulative density function. For large sample sizes n the hypothesis H_0 that the sample has been created by a stochastic process of the target distribution has to be rejected with significance level α if this distance exceeds[30]

$$d_\alpha = \frac{\sqrt{ln\left(\frac{2}{\alpha}\right)}}{\sqrt{2n}}.$$

2.3.5 Modeling of the mean return

The GARCH model family introduced in section 2.3.1 does exclusively model volatility. One of the preceding assumptions is, that x in eq. (5) has a constant mean of zero. Therefore is might be necessary to subtract the mean of the original time series to yield x. In financial time series the mean is commonly assumed to be constantly zero, a constant μ calculated as its sample mean

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i$$

or as a mean time series of its own. A common assumption for a mean time series is an autoregressive moving average or (ARMA) model.[31] The stochastic process of an

[28]Compare Student (1908)
[29]Compare Jarque and Bera (1980)
[30]Compare Massey Jr (1951)
[31]CompareGreene (2008)[Start of Chapter 21]

ARMA(P,Q) model is defined as

$$\mu_t = c + \sum_{i=1}^{P} \alpha_i x_{t-i} + \sum_{i=1}^{Q} \beta_i \epsilon_{t-i} + \epsilon_t$$

and ϵ is usually assumed to be normally distributed.

Over the course of this thesis it will be tested, which of the three approaches fits the data best.

2.4 Maximum Likelihood Estimation

For calibration of any of the models presented in section 2.3 maximum likelihood estimation should be used. Maximum Likelihood Estimation (MLE) is one of the most fundamental techniques in econometrics.[32] With MLE it is for example possible to parametrize a given distribution to fit as closely as possible to a data sample or parametrizing a model such that the model output is as close as possible to a desired distribution. It is even possible to simultaneously parametrize a model and the distributions its results are supposed to be fit to, to find the most probable combination of model parametrization and target distribution. An example for the latter approach is the parametrization of a GARCH(1,1) model with a student-t target distribution as it is presented in eq. (23). During MLE of this model the degree of freedom ν and the GARCH parameters α and β are optimized in parallel to yield the most probable combination of model and target distribution of the family of student-t distributions.

Based on the likelihood value, which is the target value during MLE it is also possible, to compare for example different distributional families with information criteria or to determine confidence bounds for the parametrization found with a MLE.

Despite of how powerful procedures based on likelihoods are the concept of likelihood is rather simple. Given a realization x and a Probability Density Function (pdf) f_ν under which the likelihood L should be determined it follows, that

$$L(x, f_\nu) = f_\nu(x),$$

where ν is the set of parameters used for the distribution f.

[32]This section is mainly based on Greene (2008, Chapter 14) and Millar (2011).

Assuming that X is a set of realizations $[x_1 \dots x_n]$ the likelihood of this set is[33]

$$\begin{aligned} L(X, f_\theta) &= \prod_{i=1}^{n} f_\nu(x_i) \\ \Leftrightarrow \quad log\left(L\left(X, f_\nu\right)\right) &= \sum_{i=1}^{n} log\left(f_\theta\left(x_i\right)\right) \\ \Leftrightarrow \quad l\left(X, f_\nu\right) &= \sum_{i=1}^{n} log\left(f_\theta\left(x_i\right)\right), \end{aligned} \tag{10}$$

where l is to so called log likelihood. In practice log likelihoods are used almost exclusively, since likelihoods can become extremely small resulting in numerical issues and due to the fact, that addition is computationally more efficient than multiplication. Since the logarithm is a monotone transformation any parametrization which maximizes l also maximizes L. It has to be noted, that this assumes that all x_i are i.i.d.. This is, why MLE can be performed on the filtered error terms of a GARCH model, which do have constant variance unlike the return data itself.

A likelihood value by itself has no significance. This is a fundamental difference to a probability which can be interpreted without comparing it to another probability. Given a probability of 0.95 it can be said, that this probability is high. In contrast a likelihood of 0.95 can neither be labeled as high nor low - it does only gain significance when other likelihoods have been measured. If the majority of likelihoods are lower than 0.95 it's relatively high, otherwise relatively low. It has to be noted, that only comparing likelihoods calculated from the same sample size produces meaningful results.

Just like probability, likelihood is floored at zero. However, unlike probability it is not ceiled. A simple example of this property is the likelihood of drawing five zeros from a normal distribution with mean zero and standard deviation 0.1. According to the pdf it follows that $\phi_{(0,0.1)}(0) = 3.98$ which results in a likelihood greater than one of $3.98^5 = 998.65$ and a non-negative log likelihood of $5 * log(3.98) = 6.90$. This phenomenon is common for time series which have very small absolute standard deviation, such as the daily movements of tenor points on a yield curve used in this thesis.[34]

Based on the definition of l given in section 2.4 the problem to solve to fit a distri-

[33] Compare MathWorks (2015)

[34] Be aware, that likelihood functions in software packages oftentimes return negative log likelihood for use in a minimizer.

bution to a set of observations X is

$$\begin{aligned} & \underset{\nu}{\text{maximize}} & & l(X, f_\nu) \\ \Leftrightarrow \quad & \underset{\nu}{\text{minimize}} & - \quad & l(X, f_\nu). \end{aligned}$$

Assuming, that the goal is rather to optimize to parametrization θ of a model which produces the results X_θ with regards to a given and fixed target distribution f_ν the problem transforms to

$$\underset{\theta}{\text{minimize}} \quad -l(X_\theta, f_\nu).$$

In the final case where for example the model parameters as well as the distribution parameters ν are parametrized the problem transforms to

$$\underset{\theta,\nu}{\text{minimize}} \quad -l(X_\theta, f_\nu).$$

The latter kind of problem would need to be solved in calibration of a GARCH model with student-t distributed error terms ϵ.

As an example assume that we want to fit a normal distribution ϕ_ν to the data $X = [-0.5, 0.5, 1.5]$ with the unknown distribution parameter $\nu = [\mu, \sigma]$. The occurrence of these three draws is most probable, if the underlying normal distribution has a mean of 0.5 and a standard deviation of 0.816, since

$$\begin{aligned} & \underset{\nu}{\text{minimize}} & & -\phi_\nu(-0.5) * \phi_\nu(0.5) * \phi_\nu(1.5) \\ \Rightarrow & & & \nu = [0.5, 0.816]. \end{aligned}$$

This result has been calculated using a numerical optimizer. The same is true for any other MLE presented in this thesis. This optimization with all the issues that come with such as the existence of multiple local minima could be studied in great detail and the theory behind it presented but this is beyond the scope of this thesis.[35]

If we now assume for the sake of simplicity, that the set of parameters which we are optimizing is θ, no matter if the parameters are actually used as part of the model or the target distribution the MLE is the $\hat{\theta}$ which solves the likelihood equation

$$\frac{\partial l(X_\theta, f_\theta)}{\partial \theta} = 0.$$

The reason, why the $\hat{\theta}$ solving this equation is of such interest are the asymptotic

[35] Optimizations in this thesis are done using the *fmincon* and *fminbnd* MATLAB functions.

properties of a maximum likelihood estimator. As the sample size grows to infinity the maximum likelihood estimator $\hat{\theta}$ shows the following properties:

1. $\hat{\theta}$ will converge towards the true parametrization θ_0 with probability.

$$\hat{\theta} \underset{plim}{\rightarrow} \theta_0$$

2. Possible $\hat{\theta}$ which could occur based on different samples will be normally distributed around θ_0 with the variance being inversely proportional to the sample size.[36]

3. A maximum likelihood estimator is asymptotically efficient. No other method of estimating $\hat{\theta}$ could yield a result with lower variance.

4. If the maximum likelihood function is continuously differentiable, $\hat{\theta}$ is unique.

The proofs of these can be found in Greene (2008, pp. 545) where the regularity conditions which have to be fulfilled to ensure the existence of the described properties are also mentioned. The existence of the regularity conditions will not be formally tested in this thesis but all plotted likelihood functions appeared smooth and should therefore probably offer the required continuous derivability up to the third order.

2.4.1 The Bayesian Information Criterion

The previous section has shown how likelihood may be used to compare parametrizations of a given model. However, it is not very practical in comparing parametrizations of different models, since it doesn't penalize the use of additional variables. For this reason the likelihood of an optimally parametrized GARCH(3,3) model will always be higher than that of an optimally parametrized GARCH(1,1) model, since the latter is a special case of the former[37]. A model selection based on likelihood alone would therefore lead to over fitted models using an excessive amount of variables.

Given the likelihoods of their optimal parametrizations, two distinct models can be compared with the Bayesian Information Criterium (BIC) as

$$BIC = -2 \ln L_j + K \ln(n) \tag{11}$$

[36]How exactly the variance behaves is dependent on the sample distribution, however for usual distributions such as the normal or student-t distribution the property $\sigma \propto \frac{1}{\sqrt{n}}$ holds.

[37]A GARCH(1,1) model might be interpreted as a GARCH(3,3) model with $\alpha_2 = \alpha_3 = \beta_2 = \beta_3 = 0$.

where L_j and K_j are the likelihood and parameter count of the j^{th} model and n is the sample size.

The BIC was derived in Schwarz et al. (1978). The BIC is an approximation of the computationally expensive Bayes Factor

$$\text{Bayes Factor} = \frac{p(Data|Model_2)}{p(Data|Model_1)}.$$

This approximation requires of a large sample, nested models[38] and multivariate normal prior expectations of model parameters. Since the result of a parameter estimation with MLE is asymptotically normal, as mentioned in section 2.4, all models that will be compared within the course of this thesis are special cases of ARMA-GARCH models, and sample sizes are large, all requirements for using the BIC are fulfilled.

The BIC does penalize the use of additional variables, and a more complex model will have a worse BIC, than a simpler model, if the additional variables do not improve the likelihood sufficiently. If one model has a higher BIC than another, the former should be favoured over the latter.

2.4.2 Robust confidence intervals

The estimation of confidence intervals of the MLE parameter estimates is obviously a crucial necessity for using these estimates in practice. Estimating a parameter is not worth much without knowing how close to the real value the estimate is likely to be. Due to the approximate normality of the parameter estimates, it is relatively easy to establish parameter confidence bounds, once the standard deviation is known. One possibility is using the so called Wald test[39] and is pretty much as simple as defining $(\hat{\theta} - 1.96\sigma_\theta, \hat{\theta} - 1.96\sigma_\theta)$ as the 95% confidence bound for the true parametrization θ_0 given the maximum likelihood estimate $\hat{\theta}$.

However, estimating σ_θ is a bit tricky. If the model structure represents reality perfectly[40] it is sufficient to estimate Σ_θ[41] based on the first derivatives of the likelihood

[38] Models are nested, if one model is a special case of the other. A GARCH(1,1) model is nested in a GARCH(3,3) model.

[39] Compare Wald (1943)

[40] In the case of a GARCH(1,1) model this would mean, that filtrating a return time series with an optimally calibrated GARCH(1,1) model would result exactly in $\epsilon \sim \phi(0,1)$ and ϵ is i.i.d..

[41] Σ being the covariance matrix

function at the maximum likelihood estimator $\hat{\theta}$:

$$\begin{aligned} \hat{\Sigma}(\hat{\theta}) &= \left[\sum_{i=1}^{N} \hat{g}_i \hat{g}_i{}'\right]^{-1} \\ \text{where} \qquad \hat{g}_i &= \frac{\partial \ln f(x_i, \hat{\theta})}{\partial \hat{\theta}} \end{aligned} \tag{12}$$

and where$x$ is the sample and f is the assumed pdf of the sample.[42] This estimator is known as the BHHH estimator.[43]

If the model is misspecified though - which should be the case for any volatility model, since it usually won't reconstruct the characteristics of an empirical time series flawlessly - the BHHH estimator is not best suited. Assume, that $j(x, \theta)$ is the assumed pdf which is different from the true pdf $f(x)$, since the model is misspecified. Then $\hat{\theta}$ is the maximum likelihood estimator from trying to shapeshift j as much as possible to resemble the sample x, even though it would be correct to do the same with f instead of j . Redefine

$$\hat{g}_i = \frac{\partial \ln j(x_i, \hat{\theta})}{\partial \hat{\theta}},$$

which is only a change in theory, since f is unknown, if the model is misspecified, then the covariance matrix

$$\hat{\Sigma}_{sand}(\hat{\delta}) = \frac{1}{N}[\bar{H}(\hat{\theta})]^{-1}\frac{1}{N}\hat{\Sigma}(\hat{\theta})[\bar{H}(\hat{\theta})]^{-1} \tag{13}$$

where $\bar{H}$ is the empirical Hessian matrix, estimates the covariance matrix of θ better than eq. (12) would do, if the model is misspecified.[44] This estimator is known as the sandwich estimator or as HC covariance matrix estimator - HC standing for heteroscedasticity consistent, since the misspecification leads to heteroscedasticity in the time series, which would not exist if the model was correctly specified.

2.5 Additional components of a Filtered historical VaR model

The FHS VaR is only one component of a VaR calculation, albeit the most important one. To adjust for model errors and make the estimation less vulnerable against violated model assumptions additional components are added. Some of these component are even designed with a goal different from estimating a return quantile as close as possible

[42]Compare Greene (2008)[Section 14.4.6]
[43]Named after Berndt et al. (1974).
[44]Compare Greene (2008)[Sections 14.8.2 and 14.8.3] for further detail and partial proof.

such as the measures to prevent procyclicality, which will be presented in section 2.5.3.

2.5.1 Correlation break adjustment

In the section 2.2.1 it was established, how a scenario is derived in FHS. Historical risk factor returns are rescaled to current volatilities to adjust for the fact, that an assumption of constant volatilities is wrong. The same argument could be brought forward for correlations. FHS does make the assumption of constant correlations, which are supposed to hold in the future[45]. The correlations assumed in FHS are not the same as in the covariance matrix one would usually calculate from realized returns. Due to the volatility filtration the filtered scenario returns are assumed to have constant correlation instead of the original, unfiltered returns. As a matter of fact, this approach does make the assumption of constant correlations more likely. Longin and Solnik (1995) have shown, that correlations of returns filtered with a GARCH(1,1) process are closer to being constant than those of unfiltered returns. Nevertheless the same paper has shown that the assumption of constant correlation between returns filtered with a GARCH(1,1) model can still not be confirmed empirically. Research has tried to improve on this by designing more elaborate GARCH processes intended to produce filtered time series which not only exhibit constant volatility within one time series but also constant correlation across multiple filtered time series.[46]

The issue with these unstable correlations is especially pressing in the context of VaR estimation. It is accepted across the industry and has been shown empirically[47] that correlations increase during time of financial crisis. This means that in times when VaR models are needed most their assumption that current correlations are close to these measured in the past is most likely to be violated.

In practice models such as those used in Engle and Sheppard (2001) and Cappiello et al. (2006) might be too complicated to be applied across the vast amount of risk factors, with which a commercial bank is facing. Therefore makeshift solutions are in place to adjust for the described model error made in estimating VaR. At Eurex Clearing a so called correlation break adjustment is calculated for this purpose.[48] This adjustment compares the VaR calculated based on a short time series, which would be e.g. six months long with the VaR based on the usual three year period. If the

[45]Compare Pritsker (2006)[section 4.2]
[46]Compare Engle and Sheppard (2001) and Cappiello et al. (2006)
[47]Compare among others Sandoval and Franca (2012)
[48]Compare Eurex Clearing AG (2014)[Section 3.4.1]

VaR estimated based on the short time series differs from that of the long time series it is assumed, that the correlations within these six months differs from the long run average. The difference of VaR estimates based on a short and long time series is calculated for each day over time period of multiple years. The average magnitude of this difference is taken as an indication of how susceptible the portfolio in question is for unstable correlations and a correlation break adjustment, which scales with this indicator is added to the estimated VaR

Within the scope of this thesis it will be assumed, that the used volatility models do result in filtered time series with constant correlations. Therefore no correlation break adjustment is done. The impact of this simplification should not be too large, since this thesis focuses on the VaR estimation of interest rate based products. Phenomena like varying correlations and an asymmetry regarding positive and negative returns[49] are much more prevalent for equity products than for interest rate products[50]. Using simpler volatility models therefore shouldn't be as much of a compromise as it would be if equity derivatives were in question.

2.5.2 Liquidity adjustment

When calculating VaR it is always calculated based on a certain number of days. A VaR is related to a percentile and a time period and defined as the loss, which the portfolio will not surpass at the percentiles probability within the time period. The time period is of special importance for a Central Counterparty, since a Central Counterparty (CCP) takes on the position of a defaulted clearing member and then sells it again as fast as possible. The sole purpose of margins taken for a portfolio is to cover potential losses between the time of default until the CCP is able to sell the inherited portfolio. For this reason in the case of a CCP the time period for which the VaR is calculated has to match the time it takes to sell the portfolio.

Based on the asset class of a portfolio it is expected to take not more than n days to sell the portfolio after it is inherited. VaR estimation is based on this n since historical n-day returns are used.[51] In the case of Eurex clearing[52] the length n of the liquidation period is only a rough guess based on the types of assets within the portfolio. The specific structure of the portfolio is not taken into account but

[49] This problem will be discussed in section 2.3.2

[50] See Cappiello et al. (2006)

[51] Alternatively one day returns could be scaled up using the square-root-of-time-rule introduced in section 2.3

[52] Compare Eurex Clearing AG (2014)[Section 3.5]

only that it is e.g. an portfolio consisting of fixed income positions. Therefore a Liquidity adjustment component is added on top of the n-day FHS VaR. This liquidity adjustment of a Portfolio is based on the instruments in the portfolio and position size of these instruments. A minimal liquidity adjustment is charged for a well diversified portfolio consisting of highly liquid products, while a very high adjustment is charged for a portfolio consisting only of a few, large positions in rather exotic products.

To reduce complexity a liquidity adjustment is not considered within this thesis. Additionally, Value at Risk is calculated based on historical one day returns, s.t. $n = 1$. As VaR calculation is done for a CCP this implies that the CCP would be certain that he can sell any inherited portfolio in only one day and therefore does not need to charge a liquidity adjustment.

Results gained in this thesis are not necessarily transferable to cases, where longer liquidation periods with $n > 1$ are used. The distributional properties of e.g. four day returns might differ considerably from the one day returns studied within this thesis.

2.5.3 Measures to prevent procyclicality

The issue of procyclicality has steadily risen in significance for regulators in the last couple of years[53]. Regulators fear, that financial strains, which behave in a procyclical manner might severely worsen or even trigger the next financial crisis. Financial strains behave procyclical if they increase in times of volatile market and especially in times of a crisis. If a bank is customer of a CCP during a crisis and the margin model of the CCP behaves largely procyclical rapidly rising margin requirements from the CCP towards the bank might force the bank into a default, which the bank might not have suffered if the margin model of the CCP was less procyclical. Rapidly rising margin requirements during a crisis could even bring a healthy bank into troubles since liquidity dries up and assets might suddenly become non-eligible as collateral. Regulators start preferring margin models for a CCP which have rather stable margin requirements even through a crisis.

A very similar development has been observed in credit rating in recent years. So called through-the-cycle models have gained favor with regulators. Internal rating models following the through-the-cycle approach base ratings on an entire economic cycle as opposed to point in time models which only base a rating decision on the current economic situation. This leads to more stable ratings mitigating procyclical

[53] Compare Murphy et al. (2014a) and Gurrola-Perez and Murphy (2015)

effects in the calculation of counterparty risk. Additionally regulators found empiric evidence, which suggests, that credit risk management is too liberal in times of financial upswing and too restrictive in times of financial downturn.[54]

However, the primary goal of a margin model has to be to cover possible losses in case a portfolio is inherited. Margin stability can only be a secondary goal if possible losses are adequately covered. The three most important measures which can prevent procyclicality within a FHS-VaR Modell are Stress VaR, Volatility flooring and the filtration itself.

Stress VaR acts as a floor to the filtered VaR estimation. Stress VaR is the VaR of a classical, non-filtered historical simulation using returns of high stress periods. Since Stress VaR is usually much higher than filtered VaR it has to be scaled down with a parameter $\omega < 1$ because it would otherwise dominate the VaR estimation.

Compared to filtered VaR stress VaR tends to be much more stable over time for two reasons. On one hand stress VaR does not depend on currently estimated volatility and therefore does neither rise nor fall when market volatility increases or decreases. On the other hand the stress scenarios themselves tend to be fairly consistent. If the financial crisis of 2008 is used as the stress period, the stress period will not change until a larger crisis occurs or the financial crisis of 2008 is not regarded relevant anymore. By contrast filtered VaR fluctuates in conformity with currently estimated volatility and every day the oldest scenario is replaced with the daily returns that have just been realized.

The synthetic time series in Figure 1 show this behaviour well. An ω of 0.9 has been chosen. Since the margin requirement is calculated as $max(Stress\ VaR, Filtered\ VaR)$ using a stress component in the VaR-Modell limits margin volatility by flooring it. Extending the VaR-model with a stress component is therefore a very feasible way to decrease procyclicality while preserving a sufficient reactivity of the model.

The second measure to limit procyclicality is to use a volatility floor for the assumed volatility in the day of VaR estimation. This floors the volatility used for VaR calculation on a certain quantile of the actually estimated volatility. If the estimated volatility would e.g. be below the 25% quantile of the empirical distribution of estimated volatilities, the 25% quantile would be used instead. This leads to higher estimated margins during a low volatility regimen and therefore to higher stability in the margin requirement and less procyclicality.

[54] Compare Jesus and Gabriel (2006)

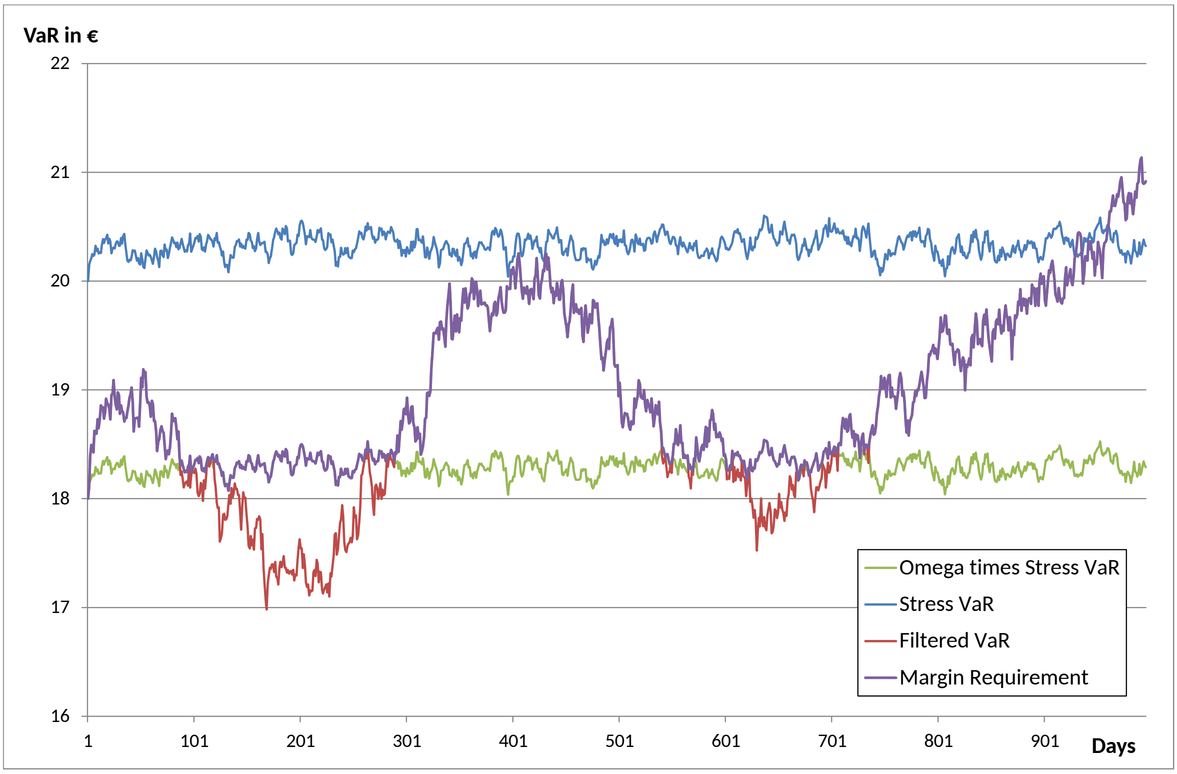


Figure 1: Synthetic example to show the interaction of Stress VaR and Filtered VaR in the calculation of the Margin Requirement of an Portfolio: The Margin Requirement is the maximum of the filtered VaR and a calibrated parameter ω times the Stress VaR. The composition of the Portfolio changes over the course of time.

The third component that heavily influences procyclicality is the filtration itself. If a simple EWMA model is considered a higher decay factor will result in more stable variance estimation and therefore more stable margins. This becomes apparent if the extreme case of $\lambda = 1$ is considered. In this case the filtered historical simulation becomes historical simulation with the assumption of homoscedasticity and Variance is estimated as the classical sample variance with a mean of zero

$$\frac{1}{N}\sum_{i=1}^{N} x_i^2$$

As described in section 2.3.3 any $\lambda < 1$ can be connected with a lookback period of returns which impact the estimated variance. The lower λ the shorter the lookback period and the more volatile will the variance estimation be. In accordance with this consideration Gurrola-Perez and Murphy (2015) found that filtered historical simulation based on an EWMA process exhibits less procyclicality with rising λ. However, this thesis will later show, that the lower reactivity of EWMA models with a $\lambda >= 0.99$ is connected with bad results in backtesting. Therefore fighting, procyclicality with measures like stress VaR or volatility flooring seems more reasonable than lowering the reactivity of an otherwise well working model just to lower procyclicality.

2.6 Backtesting

A VaR model is a complex interaction of assumptions and simplifications. Using Monte Carlo simulation one assumes to know return distributions and correlations of risk factors. When using filtered historical simulation one assumes that volatility-filtered history repeats itself. Additionally any VaR model uses pricing algorithms to calculate the VaR of portfolios. If an derivative is e.g. priced with a simple Black & Scholes model among others the assumption of normally distributed returns and homoscedasticity of the underlying is made. The assumption of normal distribution of the loss distribution is made when using the square-root-of-time-rule and modelling volatility is also based on assumptions.

While every single one of these assumptions can and should be tested by itself using historical data, the performance of the entire VaR model should be tested as well by backtesting. In backtesting, historical scenarios are applied on current portfolios to test, if the used VaR model would have properly worked. For every historic scenario the VaR of the portfolio that would have been estimated can be compared with the

PnL of this portfolio would have realized on this past day. For some historic scenarios the loss will be greater than the estimated VaR, which is a so called exception[55]. Some of these exceptions are to be expected, since the VaR only covers losses up to a certain quantile - typically 0.95 or 0.99. It would therefore be expected that the VaR is exceeded in five or respectively one out of a hundred days. Different tests of the backtesting performance of VaR models have been proposed in recent years, which will be introduced in sections 2.6.1 and 2.6.3.

It is important that no information younger than the point of observation is used for the VaR estimation. If for example the 4^{th} of April 2011 is the historical scenario the VaR model used has to be calibrated based on information available earlier than the 4^{th} of April 2011.

A possible issue of backtesting is, that VaR models might be overfitted to historical data. One run of backtesting does not use any ex-post information. However if the first backtest did not run well and the VaR model is afterwards iteratively altered to produce better backtesting results it becomes questionable how meaningful these backtests actually are. A VaR model that has been developed solely based on parameter MLE without ever running a backtest and then performs well in the first backtest it is run through is certainly more desirable than a model which has been parametrized with the backtesting result as the target variable. The latter model is more likely to be overfitted to the backtesting period and might not perform well in the future.

2.6.1 Measurement of Backtesting Performance

The backtesting of a VaR model may test if the VaR model fulfills three important properties a VaR should have[56].

1. The count of realized exceptions should be close enough to the expectation.

2. Exceptions should occur independent of each other.

3. The performance of the VaR model should not be dependent on the chosen quantile α.

Focusing on the first two properties first, tests for the first property are called tests for unconditional coverage, while tests for the second property are tests for independence. Tests for conditional coverage are simultaneously verifying both properties.

[55] Often also called exceedance of failure
[56] Compare Campbell (2005)

While tests of unconditional coverage are very widespread and also used by regulators the independence property is at least of equal importance. The claim of a VaR model is to produce a loss value which will only be surpassed with a probability of α. Tests for unconditional coverage can only assess this claim in average, while tests for independence might find an indication that the VaR model underestimates risk during certain periods.

A VaR model that passes a test for unconditional coverage might very well underestimate risk for some days. A common phenomenon could be, that a model claiming to calculate the VaR at $\alpha = 0.01$ might in reality only calculate the VaR at $\bar{\alpha} = 0.2$ within one week after an exception. Over the course of a long testing period the overall exception count might not reveal this issue. Therefore a model with this property might very well pass a test for unconditional coverage, while it should fail a test for independence. Since exceptions usually occur in stressed market conditions the VaR model in this case underestimates risk when it is needed most.

Tests for conditional coverage seem to be more powerful than tests for independence since they cover independence and unconditional coverage. However, a model might pass the test for conditional coverage and unconditional coverage, while it fails a test for independence. When independence and unconditional coverage are assessed in a single test statistic, good results regarding the unconditional coverage might blur bad results regarding the independence. A test for conditional coverage can therefore not fully replace two independent tests for independence and unconditional coverage.

Focusing on the third property, VaR models based on Monte Carlo or historical simulation claim to be able to forecast the entire PnL distribution for the time horizon of the model. However, regarded is usually only on a single percentile of this distribution while the rest of it is ignored. The conclusiveness of the tests previously introduced suffers severely from the small sample sizes, which are common in backtesting. If only one quantile is considered backtesting produces only one dataset per historic scenarios. If multiple quantiles are regarded one hit function as defined in eq. (14) may be built per quantile increasing the size of the dataset. Among others Berkowitz (2001) developed performance measures for VaR models based on multiple quantiles. Since sample size in this thesis is unusually large such tests have been passed upon and will therefore not be examined closer in the sections to follow.

2.6.2 Unconditional tests of backtesting performance

The fundamental function used to examine backtesting performance is the hit function, which indicates if a loss breaches the VaR or not. It is defined as

$$\begin{aligned} I_{t+1}(\alpha) &= \begin{cases} 1 & \text{if } x_{t,t+1} \leq -VaR_t \\ 0 & \text{otherwise} \end{cases} \\ I_{sum}(\alpha) &= \sum_{t=2}^{T} I_t(\alpha) \end{aligned} \tag{14}$$

where α is the VaR quantile and x is the return time series. If $I_t = 1$ this t is called an exceedance, exception or failure. Kupiec (1995) developed a test statistic comparing the expected proportion of failure α with the observed proportion of failure $\hat{\alpha} = \frac{1}{T} I_{sum}(\alpha)$

$$Q = 2 * log\left(\left(\frac{1-\hat{\alpha}}{1-\alpha}\right)^{T-I(\alpha)} \left(\frac{\hat{\alpha}}{\alpha}\right)^{I(\alpha)}\right). \tag{15}$$

If this statistic is below a threshold ϵ the null hypothesis of appropriate unconditional coverage at the α quantile is rejected with a confidence level of $1 - \epsilon$. This test statistic is two-sided. This means that both models producing too few exceptions and models producing too many exceptions are rejected. However regulators won't criticize a model, since it produces too few exceptions. Measures to prevent procyclicality as described in 2.5.3 will oftentimes force productive VaR models to yield not a single exception during a low variance regime. Therefore most operating VaR models would be rejected by a two sided Kupiec test but regulators do not penalize this.

The Basel committee dictates $\alpha = 0.01$ and a backtesting period of 250 days. Just like a Kupiec test their traffic light approach is based on unconditional coverage.[57] Models are labeled *green* if they produce less than 5 exceptions, *red* if they produce more than 10 exceptions and *yellow* otherwise. If one focuses on the distinction between *green* and *yellow* the traffic light system favors rather conservative models: To reject a model yielding five exceptions over a course of 250 days and $\alpha = 0.01$ the Kupiec test would require $\epsilon \geq 0.17$. The null hypothesis $\alpha = 0.01$ can only be rejected on an 83% confidence level - the model will be penalized by the regulator nevertheless. This might be connected to the fact that the Kupiec test is prone to produce type II errors when used on a small sample such as 250. For example the probability to reject

[57] Compare on Banking Supervision (1996)

Table 2: Information needed for a Lag 1 independence test

	$I_{t-1} = 0$	$I_{t-1} = 1$	
$I_t = 0$	n_{00}	n_{01}	$n_{00} + n_{01}$
$I_t = 1$	n_{10}	n_{11}	$n_{10} + n_{11}$
	$n_{00} + n_{10}$	$n_{01} + N_{11}$	N

the hypothesis $\alpha = 0.01$ when the actual value is $\bar{\alpha} = 0.03$ when using a confidence level of $1 - \epsilon = 0.95$ is only 62.5%. The reason for this is, that based on the binomial distribution the probability to receive $I_{sum}(0.03) \in \{0, ..., 6\}$ is

$$P\left(I_{sum}(0.03) \leq 6\right) = \sum_{k=0}^{6} \binom{250}{k} 0.03^k * (1 - 0.03)^{250-k} = 37.5\%. \tag{16}$$

2.6.3 Conditional tests of backtesting performance

Kupiec already criticized his own unconditional coverage test when publishing it for yielding non-significant results on small sample sizes [58]. Christoffersen (1998) developed the first independence and conditional coverage tests for VaR models which found great appreciation by theoreticians and practitioners and are today widespread in risk management departments around the globe. Christoffersens independence test examines, if the probability for an exception differs conditional on the occurrence of an exception the day before. The null hypothesis of the independence test is

$$H_{0,ind}: \; P(I_t = 1 | I_{t-1} = 0) \; = \; P(I_t = 1 | I_{t-1} = 1). \tag{17}$$

His test for conditional coverage includes the claimed VaR α, such that the null hypothesis becomes

$$H_{0,cc}: \; P(I_t = 1 | I_{t-1} = 0) \; = \; P(I_t = 1 | I_{t-1} = 1) \; = \; \alpha. \tag{18}$$

To perform the test for independence as proposed in Christoffersen (1998) one simply has to transform the hit function I as shown in table 2. Once all n have been determined the test statistics for the independence and conditional coverage test can be calculated and the null hypotheses from equations (17) and (18) can be rejected or accepted. Further details on the test including the teststatic can be found in appendix A.1.

[58] Compare Kupiec (1995)

The obvious flaw of this approach is the fact that only exceptions that follow directly on each other are identified as clustering. If an exception would occur every second day for three weeks straight this would clearly be a clustering of exceptions but the just presented test could not identify it.

Christoffersen was able to improve upon his own work a few years later. Christoffersen and Pelletier (2004) developed an independence test based on the duration between two exceptions. If the occurrence of exceptions is interpreted as a Markov chain the realized gaps between two exceptions should not deviate too far from the expectation of $\frac{1}{\alpha}$. This test is presented in further detail in appendix A.2 including the used test statistics.

The duration based approach might be unsuited for practical application. As was explained in section 2.5.3, commercial banks will usually overestimate their risk during low volatility regimes. A duration based test would penalize these long intervals between exceptions. A test penalizing too many unexpectedly short intervals between exceptions as they could but should not occur during a high volatility regime, while not penalizing too long intervals would be desirable but has not been proposed yet.

An alternative approach to assess backtesting results has been developed by Engle and Manganelli (2004). They test, if information available at time t can be used to anticipate if an exception occurs in $t+1$ more accurately than the null $P(I_{t+1} = 1) = \alpha$. If an alternative that works significantly better than the null could be found the VaR(α) model that yielded the backtesting results has to rejected, since the occurrence of exceptions is not i.i.d. for each day. This test is presented in a bit more detail in appendix A.3

Berkowitz et al. (2011) applied all of the four backtesting performance measures introduced above on the HS VaR figures calculated for different business lines of a commercial bank. They find most praise for the CAViaR approach by Engle and Manganelli (2004), label the duration based approach from Christoffersen and Pelletier (2004) as good, while massively criticizing the use of unconditional performance measures such as the massively popular Kupiec test introduced in Kupiec (1995).

2.7 Risk factor based pricing of a Forward Rate Agreement

As introduced in section 2.1 risk factors are the most granular parts of a VaR model and the VaR of any more complex products is calculated based on pricing algorithms using the risk factors and deterministic components. This thesis focuses on the VaR estimation of forward rate agreements.

A Forward Rate Agreement (FRA) is one of the most basic interest rate products. It is an over-the-counter agreement between two parties, regarding the difference between an agreed upon interest rate and a realized interest rate of a reference index at a future point in time. One party is called the borrower while the other is the lender. The contract has three dates of interest: t_0, the date of initializing the agreement, t_1 the date at which the realized reference interest rate is compared with the agreed upon interest rate and t_2, since the reference interest rate will be the interest rate to borrow the principal for the duration $t_2 - t_1$. Assume that

- R_K is the rate of interest agreed to in the FRA based on quarterly compounding[59]
- R_{t,t_1,t_2} is the forward interest rate in t for a period from t_1 to t_2 based on quarterly compounding
- L is the principal amount
- R_2 is the continuously compounded riskless zero rate for maturity t_2 at the time of valuation

and R_K and R_{t,t_1,t_2} will be bound to an agreed upon reference index. If $R_K > R_{t_1,t_1,t_2}$ the lender pays the interest rate difference times the discounted principal to the borrower in t_1 and vice versa, s.t.[60]

$$Cashflow_{t1,Lender} = -(R_K - R_{t_1,t_1,t_2})\frac{t_2 - t_1}{360} * L.$$

FRAs usually have a value of zero when they are arranged. Such a FRA is labeled as *fair* and has the advantage that no payments have to be exchanged between the two parties at the time of entering the FRA. A FRA has a value of zero in t, if $R_K = R_{t,t_1,t_2}$[61]. Therefore R_K will be set as $R_K := R_{t_0,t_1,t_2}$. For this reason the only cash flow will be in t_1 while there is no cash flow in t_0 and t_2. The fact, that the principal is not actually lend from the lender to the borrower between t_1 and t_2 has the advantage, that the lender does not suffer a large counterparty credit risk in this time period. In fact both parties can only have a relatively minor credit risk between t_0 and

[59]The compounding type is dependent on the specific FRA. A 6/9 FRA with $t_1 = t_0 + 6$ months and $t_2 = t_0 + 9$ months would for example need this compounding type since $t_2 - t_1$ is a quarter year. This notation is the market practice and simplifies the formula for valuation.

[60]Compare Hull (2006, Chapter 4.7)

[61]For an explanation compare Hull (2006, Chapter 4.6 and 4.7).

t_1, which is usually mitigated by requesting margins from the counterparty or clearing the FRA at a CCP. Due to this design a FRA is a useful tool for trading interest rate market risk without the large associated credit risk that a forward on a zero coupon bond would for example have.

If for a point in time t, s.t. $t_0 < t < t_1$ the forward interest rate $R_{t,t_1,t_2} > R_{t_0,t_1,t_2}$ the FRA initiated in t_0 has positive value for the borrower, since he managed to lock in a lower interest rate in t_0 than he could in t. The value in t of a fair FRA initiated in t_0 for the borrower is[62]

$$V_{FRA} = L(R_{t,t_1,t_2} - R_K)(t_2 - t_1)e^{R_2 t_2} \tag{19}$$

The most heavily traded interest rate product are interest rate swaps followed by the just introduced forward rate agreements[63]. Conveniently enough an Interest Rate Swap (IRS) can be characterized as a portfolio of FRAs. This thesis focuses exclusively on the VaR estimation of FRAs, since their structure is simpler than that of an IRS and therefore uses less risk factors for pricing. Since this thesis examines the influences of differences in volatility modeling of different risk factors it is beneficial to use products which are priced using as few risk factors as possible. This way the precision of the VaR estimation of the product can be directly contributed to the quality of volatility modeling of a very small set of risk factors. If instead more complex products would be used for VaR estimation the use of many risk factors in pricing would make it difficult to assign good or bad results to the risk factors, whose volatility modeling caused them.

In practice it will be of higher importance to accurately estimate the VaR of an IRS than it is to accurately estimate the VaR of a FRA since exposure in the former will usually be larger. However, since an IRS can be characterized as a portfolio of FRAs properly estimated FRA VaR will result in properly estimated VaR of an IRS. The replication of an IRS with a portfolio of FRAs is e.g. explained in Hull (2006, Pages 162-164).

2.7.1 Using interest rate curves as risk factors

As is apparent from eq. (19) the pricing formula of an IRS uses two risk factors. The forward interest rate R_{t,t_1,t_2} and the risk free rate used for discounting the final payment R_2. Time behaves deterministic and R_K is fixed while the FRA exists and therefore

[62]Derived in further detail in Hull (2006, Chapter 4.7)
[63]Compare Gyntelberg and Upper (2014)

are not risk factors.[64] R_2 is the continuously compounded risk free zero rate. While the existence of truly risk free interest rates has been a subject of discussion in the last years, overnight index swap rates such as the EONIA in case of the Euro will be used as risk free rate in this thesis. R_2 can be transformed into a continuously compounded discount factor $d_{OIS,2}$ as $d_{OIS,2} = e^{-R_2 t_2}$. It is desirable to transfer all risk factors into a representation based on continuously compounded discount factors to store as few curves as necessary. Any type of compounding as well as a forward curve or a yield curve can be transformed into a continuously compounded discount curve.[65] Transforming the other risk factor R_{t,t_1,t_2} is a bit more challenging. First the quarterly compounded interest rate has to be transformed into a continuously compounded one year rate by $C_{t,t_1,t_2} = ln\left((1 + R_{t,t_1,t_2})^4\right)$. Afterward it has to be transformed from a forward curve to the discount curve. While C_{t,t_1,t_2} is a single point on the three month continuously compounded forward curve at t_1 its representation using the discount curve requires the discount factors at t_1 and t_2:

$$C_{t,t_1,t_2} = \frac{-ln(d_{3M,1}) - ln(d_{3M,2})}{t_2 - t_1}$$

The additional subscript $3M$ is necessary, since the discount curve, which can be derived from the three month forward curve is not the same discount curve as the one with the overnight index swap (OIS) subscript derived from the overnight index swap rate. In the case of the euro the discount curve for $3M$ is based on the three month Euro InterBank Offered Rate (EURIBOR), while the *OIS* curve is based on the Euro OverNight Index Average (EONIA).[66] Without going into too much detail at this point both of these indexes are based on daily surveys at large commercial banks requesting the interest they get for unsecured interbank funding with either a one day scope in case of EONIA or a three month scope in the case of the three month EURIBOR. Since even large banks can default a commitment to lend money to another large bank for three months is tied to a larger credit risk than lending it for only one day. Therefore yields of the three month EURIBOR are higher than those for the EONIA and therefore the $3M$ discount curve differs from the OIS discount curve.[67] The two discount curves can be

[64] Compare section 2.1

[65] A good explanation of the relation between different compounding methods and interest rate curves can be found e.g. in Hull (2006)[Chapter 4]

[66] In practice these discount curves are bootstrapped from forward rates implied by liquid interest rate products like swaps. Compare Hagan and West (2006)

[67] This is commonly known as the *OIS-Libor Spread* and in more detail discussed for example in Thornton (2009)

differentiated by their so called reset frequency. The OIS curve has a reset frequency of one day while the $3M$ curve has a reset frequency of three months. Since a set of the most common reset frequencies can be used as the underlying of an IRS or FRA it is necessary to not only store one discount curve per currency but a discount curve for each of the most frequently used reset frequencies of a currency.

2.8 VaR estimation and backtesting of a Portfolio of Forward Rate Agreements - an example

To wrap up the theoretical section and lead over to the empirical section we will visualize in an example how all the pieces introduced up to now can work together to actually estimate a VaR and test if the estimation is any good. The goal is to estimate the VaR of a portfolio of two FRAs a 6/9 € FRA as the borrower and a 3/9 £ FRA as the borrower. First we will need to identify the risk factors of the portfolio and get a time series of each of them, secondly we need to calibrate volatility models for the risk factors, in a third step we will need to calculate the VaR of our portfolio and finally perform a backtest to see if our calibrated model would have performed well, if used in the past. In our example we will use the same filtered historical simulation approach to estimate the VaR that will be used throughout the entire empirical section.

2.8.1 Identifying and gathering relevant data

Based on section 2.7 we are able to identify the risk factors of the two products in our Portfolio. We are assuming, that the two FRAs have just been arranged and therefore - in accordance with the notation of section 2.7 - the point of observation t is exactly $t = t_0$ and not $t_0 < t < t_1$. For the 6/9 € FRA our risk factors are therefore the six month tenor point on the three months forward curve and the nine months tenor point on the EONIA curve. If our point of observation t would be after t_0 our risk factors would be the $t_1 - t$ tenor point of the forward curve and the $t_2 - t$ tenor point on the EONIA curve. After transforming the tenor point on the forward curve to the discount curve our three risk factors are the six and nine month tenor point on the three month EURIBOR discount curve and the nine month tenor point on the EONIA discount curve. We will name these $d_{€,3M,6M}$, $d_{€,3M,9M}$ and $d_{€,1D,9M}$. The first spot of the subscript is the currency, the second the reset frequency and the third the tenor point. Doing the same for the 3/9 £ FRA we end up with the risk factors $d_{£,6M,3M}$, $d_{£,6M,9M}$ and $d_{£,1D,9M}$. The reset frequency used for the 3/9 £ FRA is 6M, since $t_2 - t_1$

is six months unlike the 6/9 € FRA where it is three months.

For each of these six risk factors we now need a history of PnLs. For model calibration and calculating the current VaR a history of about three years is usual[68] to do backtesting a longer history of for example ten years is necessary. Assuming that daily stored discount curves of the four necessary curves are available the time series of e.g. the risk factor $d_{€,3M,6M}$ is created from the daily movements of the six month tenor point of the three month EURIBOR curve. Just as one would do with daily settlement prices of a stock the log returns of this series of discount factors is calculated. Even though discount factors aren't technically prices the process is the same for any risk factor no matter if it happens to be e.g. an underlying price or a discount factor. This is also, where the liquidation period, which was introduced in section 2.5.2 comes into play. If we are assuming a one day liquidation period - as we do throughout this thesis - we only need a distance of one day in between the settlement prices with which we calculate one return. If the liquidation period was instead e.g. four days the settlement prices used to calculate one return would need to be four days apart.

2.8.2 Choosing and calibrating a volatility model

After the relevant time series has been created we need to investigate the properties of these six return time series. First we need to test, if the time series do exhibit volatility clustering, which would make it necessary to use FHS instead of the simpler HS. For this we could for example use the Ljung-Box Q-test introduced in section 2.3.4. Making the likely assumption, that the test detects volatility clustering in all time series we need to use FHS and identify a volatility model, which captures the properties of the time series well.

The step of identifying and calibrating the proper volatility model is the most important part of the process in the context of this thesis. This thesis investigates the impact that alterations in this step have on the estimated VaRs.

At this point we do have six different time series, which need modeling of their volatility. The first step is to find out which type of volatility model is best suited for each time series. Different GARCH models as they are introduced in section 2.3.1 and section 2.3.2 are at choice. For each combination of model type and time series an optimal parametrization can be found using maximum likelihood estimation as it is presented in section 2.4. After the best parametrization for each model for a specific

[68] In the empirical section different lengths of the calibration windows will be compared

time series has been found the parametrized models can be compared with each other using the Bayesian information criterion, that has been introduced in section 2.4.1. Based on the BIC the best working model for each time series can be identified. In the context of this example we will assume, that a conventional GARCH(1,1) model performs best for each of the six risk factor time series.

While a GARCH(1,1) model is the best fit for modeling the volatility for all the six time series the parametrization that results from the MLE is still unique for each time series. By estimating confidence intervals of the parametrization results - as introduced in section 2.4.2 - it might be possible to identify subsets of risk factors for which these confidence intervals overlap. An overlap of the confidence intervals of the parametrizations of two time series indicates, that it is feasible to use the same parametrization for both volatility models instead of two individual ones. In the context of this example it is assumed that one overlap has been found between the three € risk factors $d_{€,3M,6M}$ $d_{€,3M,9M}$ and $d_{€,OIS,9M}$ while another overlap has been found for the three £ risk factors $d_{£,6M,3M}$, $d_{£,6M,9M}$ and $d_{£,1D,9M}$. This implies, that the six unique GARCH(1.1) parametrizations we started with can be replaced by only two - one covering the € risk factors, the other covering the £ risk factors - without loss of statistical significance.

The next step is to find these two parametrizations. Using MLE again we can find the parametrization with the highest likelihood to capture the properties of all three €risk factors and repeat this for the three £risk factors respectively. The procedure for this step is sketched in fig. 3.

This concludes the selection and calibration of volatility models for the risk factors. Using the chosen models we are able to create six additional time series of estimated volatilities using eq. (5). Each of these volatility time series spans again three years and belongs to the return time series from which is has been created.

2.8.3 Calculating the current value at risk

With the return time series and volatility time series for all risk factors and the FRA pricing equation 19 everything that's necessary for calculating the FHS-VaR of the portfolio is available. As explained in section 2.2.1 all return series need to be transformed on unit standard deviation and multiplied with the currently estimated standard deviation to obtain the scenario returns:[69]

$$\begin{pmatrix} r_{t_{-750}} \\ \vdots \\ r_{t_{-1}} \end{pmatrix} \odot \begin{pmatrix} \sigma^{-1}_{t_{-750}} \\ \vdots \\ \sigma^{-1}_{t_{-1}} \end{pmatrix} \cdot \sigma_{t_0} = \begin{pmatrix} r_{scn_1} \\ \vdots \\ r_{scn_{750}} \end{pmatrix}$$

Please note that each risk factor uses its specific volatility time series and a specific σ_0 to scale the normalized returns on the currently estimated market volatility of this specific risk factor.

Afterwards the scenario returns of the risk factors need to be converted into scenario returns of the portfolio. Using all riskfactor returns of scenario one - the rescaled returns of 750 days ago - the PnL of the two FRAs can be calculated using the pricing formula. Adding up these two PnLs results in the portfolio PnL of scenario one. Repeating this for all other 749 scenarios yields 750 portfolio PnLs with which a return distribution can be created. Assuming that it is the goal to estimate a 99% VaR the 7$^{\text{th}}$ largest portfolio loss is the estimated VaR in t_0.

2.8.4 Backtesting the calibrated model

After a model has been chosen and calibrated - as was just done in section 2.8.2 - it needs to be tested. While the model calibrated with MLE has good properties in estimating volatilities the VaRs calculated by it might still improperly estimate the loss which will only be exceeded in one out of one hundred cases. Instead of using the calibrated model productively it's therefore sensible to test how the model would have fared in the past and compare the VaR predictions the model would have made in the past with the realized returns. This is done with backtesting as it was introduced in section 2.6. All that needs to be done for backtesting is to shift the point of observation from t_0 into the past and repeat the steps described in section 2.8.2 and section 2.8.3 after the volatility model has been selected. Given a new point of observation, e.g. t_{-500} the volatility model needs to be calibrated using only data available before t_{-500} and the VaR of the portfolio needs to be calculated using the returns $(r_{t_{-1250}} \cdots r_{t_{-501}})$. The VaR that would have been estimated with the new model in t_{-500} can then be compared with the return the portfolio would have realized between t_{-500} and t_{-499}. After repeating this for as many past points of observation as possible the introduced in section 2.6.1, section 2.6.2, section 2.6.3 and A can be used to determine if the model

[69] $\odot$ is the operator for the Hadamard product, which is element-wise multiplication.

would have worked well in the past.

In the case of a portfolio of FRAs the products need to be 'rolled' for backtesting. Since a FRA matures we need to create a portfolio for each past point of observation that has exactly the same maturities as the portfolio original in t_0. In case of the 6/9 € FRA we would need to use a FRA with reference dates around t_{-380} and t_{-320}, assuming that a month has 20 banking days.

3 Empirical Analysis

After the theoretical foundation has been laid and the process of estimating the VaR of a FRA has been established in section 2 an empirical study can be made. The goal of the empirical section is to analyze, how many different volatility models are necessary to produce decent VaR estimates. The tried volatility models were introduced in section 2.3 and performance of the VaR estimates will be measured by the tests introduced in section 2.6.

The empirical section is based on multiple hundred risk factors and at the end of the empirical section an answer will be found just how many different volatility models are necessary to cover all of these risk factors, while still yielding good VaR estimates.

3.1 Data used

Since we want to calculate the VaR of FRA Portfolios we need to identify the necessary risk factors first. As shown in section 2.7 risk factors of a FRA with dates t_0, t_1 and t_2, as they were defined previously, are the t_2 tenor point on the riskless continuously compounded discount curve and the t_1 and t_2 tenor point on the continuously compounded discount curve with reset frequency $t_2 - t_1$.

As FRAs mature $t_2 - t_0$ can pretty much assume any value between one day and 50 years, while $t_2 - t_1$ is limited to only a few common reset frequencies per currency. Within the scope of this thesis, up to five different reset frequencies, each for five different currencies are considered. Overall, this results in 23 different discount curves, which are subject to study. Table 3 on page 42 has one column for each of these discount curves. Each discount curve is represented by a set of its tenor points, which are the curve's risk factors. Since each curve in theory consists of an infinite number of tenor points, these could be translated to an infinite number of risk factors. Instead, tenor points are chosen to represent the curve, which are close to the tenor points, from which the curve is originally created. The discount curves used were provided by productive systems already in place at Eurex Clearing and their creation was not part of the empirical work for this thesis.

Eurex creates discount curves based on the implied discount factors of a set of very liquid FRAs and IRSs for each curve. Using these discount factors as pillars the rest of the discount curve is bootstrapped afterwards.[70] The risk factors chosen to represent

[70] A good survey on yield curve interpolation can be found in Hagan and West (2006). This paper was also the point of orientation for the methodology put in place at Eurex.

a discount curve are oriented towards these pillar discount factors to keep the impact of the bootstrap as low as possible. However, it has to be considered, that these pillars will move slightly on a day to day basis, as the liquid products, which implied them, mature. The yield points chosen as risk factors on the other hand have to be locked at constant maturity to be able to create a return time series based on them.

The tenor points chosen as representative risk factors for each discount curve can be looked up in table 3. Overall the time series of 517 different risk factors are subject to study.

3.2 Individual Modelselection

Given the daily discount curves and after we have identified relevant tenor points a risk factor return time series can be created. Given a time series of discount factors $d_{Ccy,Rfq,Tenor}$ where the tenor point discount curve can be identified by its currency Ccy, its reset frequency Rfq and it's tenor $Tenor$ a log return for time t is calculated as

$$r_{Ccy,Rfq,Tenor}(t) = \ln\left[\frac{d_{Ccy,Rfq,Tenor}(t)}{d_{Ccy,Rfq,Tenor}(t-1)}\right] \tag{20}$$

resulting in the return time series $r_{Ccy,Rfq,Tenor}$.

In a first step we fitted different volatility models to the return series of each tenor point individually. Starting with the assumption of an ARMA(1,1) process for the estimation of a mean time series and a GARCH(1,1) process for the estimation of a volatility time series gradually simpler models are tested.[71] The following eight models were fitted to the time series:

ARMA(1,1)-GARCH(1,1) model with student-T distributed innovations labelled as '*ARMAGARCH11T*'

$$\begin{aligned} r_t &= \mu + \gamma * r_{t-1} + \delta * \epsilon_{t-1} + \epsilon_t \\ \epsilon_t &= z_t * \sigma_t \\ z_t &\sim T_\nu(0,1) \\ \sigma_t^2 &= \omega + \alpha * \epsilon_{t-1}^2 + \beta * \sigma_{t-1}^2 \end{aligned} \tag{21}$$

GARCH(1,1) with a constant mean possibly unequal to zero and student-t distributed

[71] GARCH models with a lag structure $P = Q > 1$ have also been briefly tested but quickly discarded due to poor results. The use of lags beyond the first for financial time series is also very rare in the literature. Compare e.g. Hansen and Lunde (2005)

Table 3: Time series used for analysis

Currency	EUR					CHF					JPY					GBP					USD				
Resetfrequency	OIS	1M	3M	6M	12M	OIS	1M	3M	6M	12M	OIS	1M	3M	6M	12M	OIS	1M	3M	6M	12M	OIS	1M	3M	6M	12M
Tenorpoint																									
1 Day	✓					✓					✓					✓					✓				
7 Days	✓										✓					✓					✓				
14 Days	✓										✓					✓					✓				
21 Days	✓										✓					✓					✓				
1 Month	✓		✓*	✓*		✓		✓	✓		✓		✓	✓		✓		✓	✓		✓		✓*	✓	
2 Months	✓	✓	✓*	✓*		✓	✓	✓	✓		✓		✓	✓		✓	✓	✓	✓		✓	✓	✓*	✓	
3 Months	✓	✓	✓*	✓*		✓	✓	✓	✓		✓	✓	✓	✓		✓	✓	✓	✓		✓	✓	✓*	✓	
4 Months	✓	✓	✓*	✓*		✓	✓	✓	✓		✓		✓	✓		✓	✓	✓	✓		✓	✓	✓*	✓	
5 Months	✓	✓	✓*	✓*		✓	✓	✓	✓		✓		✓	✓		✓	✓	✓	✓		✓	✓	✓*	✓	
6 Months	✓	✓	✓*			✓	✓	✓			✓	✓	✓			✓	✓	✓			✓	✓	✓*		
7 Months	✓	✓	✓*			✓	✓	✓			✓		✓			✓	✓	✓			✓	✓	✓*		
8 Months	✓	✓	✓*			✓	✓	✓			✓		✓			✓	✓	✓			✓	✓	✓*		
9 Months	✓	✓				✓	✓				✓	✓				✓	✓				✓	✓			
10 Months	✓	✓				✓	✓				✓					✓	✓				✓	✓			
11 Months	✓	✓				✓	✓				✓					✓	✓				✓	✓			
1 Year	✓	✓	✓*	✓*		✓	✓				✓	✓	✓	✓		✓	✓	✓	✓		✓	✓	✓*	✓	
2 Years	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	
3 Years	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	
4 Years	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	
5 Years	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	
6 Years	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓		✓	✓	✓	✓	✓	✓		✓		
7 Years	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	
8 Years	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓		✓	✓	✓	✓	✓	✓		✓		
9 Years	✓	✓	✓	✓	✓		✓	✓	✓		✓	✓	✓	✓		✓	✓	✓	✓	✓	✓		✓		
10 Years	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	
11 Years											✓					✓			✓				✓		
12 Years	✓	✓	✓	✓	✓		✓	✓	✓		✓	✓	✓	✓		✓	✓	✓	✓	✓	✓		✓		
15 Years	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓		✓	✓	✓	✓	✓	✓		✓		
20 Years	✓	✓	✓	✓	✓		✓	✓	✓		✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	
25 Years	✓	✓	✓	✓	✓		✓	✓	✓		✓	✓	✓	✓		✓	✓	✓	✓	✓	✓		✓		
30 Years	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	
40 Years		✓	✓	✓	✓											✓	✓	✓	✓	✓	✓		✓	✓	
50 Years	✓		✓	✓	✓											✓	✓	✓		✓	✓		✓	✓	

All tenorpoints on the discountcurves for which enough data is available to use them as a returnseries without risking problems caused by insufficient data. Euro Tenorpoints marked with an asterix are just in use after April 2012 since data is insufficient beforehand. US Dollar 3M Tenorpoints marked with an asterix are just in use after April 2014 for the same reason.

innovations labeled as '*CGARCH11T*'

$$\begin{aligned} r_t &= \mu + \epsilon_t \\ \epsilon_t &= z_t * \sigma_t \\ z_t &\sim T_\nu(0,1) \\ \sigma_t^2 &= \omega + \alpha * \epsilon_{t-1}^2 + \beta * \sigma_{t-1}^2 \end{aligned} \tag{22}$$

GARCH(1,1) with the assumption, that the mean of the daily continuous returns is zero and student-t distributed innovations labeled '*GARCH11T*'

$$\begin{aligned} r_t &= \epsilon_t \\ \epsilon_t &= z_t * \sigma_t \\ z_t &\sim T_\nu(0,1) \\ \sigma_t^2 &= \omega + \alpha * \epsilon_{t-1}^2 + \beta * \sigma_{t-1}^2 \end{aligned} \tag{23}$$

Exponentially weighted moving average with student-t distributed innovations labeled '*EWMAT*'

$$\begin{aligned} r_t &= \epsilon_t \\ \epsilon_t &= z_t * \sigma_t \\ \sigma_t^2 &= (1-\lambda) * \epsilon_{t-1}^2 + \lambda * \sigma_{t-1}^2 \end{aligned} \tag{24}$$

with the same distributional assumption

$$z_t \sim T_\nu(0,1) \tag{25}$$

Additionally the same four models were considered with normally distributed innovations, such that

$$\begin{gathered} z_t \sim \phi(0,1) \\ \text{instead of} \\ z_t \sim T_\nu(0,1) \end{gathered} \tag{26}$$

which are labeled without the trailing '*T*'.

From top to bottom, each model in eq. (21) to (24) uses one or two fewer variables, than the previous model. As explained in section 2.4.1 each model is therefore nested within the previous ones. Additionally, any model assuming $z_t \sim \phi(0,1)$ is nested within the same model with the assumption $z_t \sim T_\nu(0,1)$, since the normal distribution is a special case of the student-t distribution. Since models are nested, we can use the Bayesian Information Criterion to compare their performance. Each of the eight models was individually calibrated for the time series of every single risk factor adding up to about 4000 combinations. Calibrations were done for different time windows and lengths: One time window spanning the entire eight years of available data, as well as three different three year subwindows.[72] This adds up to about 16,000 calibrations.

To gain an overview over these calibrations the model, which results in the highest *BIC*[73] for each risk factor is identified and the count of risk factors which are best calibrated with a given model are displayed in table 4. As an explanation of the table: The first entry means, that 64 risk factors with a tenor of 6 months or less are best calibrated with the *'ARMAGARCH11T'* model of eq. (21) based on an eight year time series reaching from January 2006 to January 2014.

The first observation to strike the eye in table 4 is, that no model with the assumption of normally distributed z performed best for any risk factor. The assumption of student-t distributed z improves the likelihood significantly. As the second observation, the *EWMAT* model seems to always be covering the majority of risk factors. The third observation is, that *ARMAGARCH11T* seems to perform well for low tenor risk factors and *GARCH11T* models seem to perform well for high tenor risk factors. Finally, the *CGARCH11T* model which makes the assumption that it would be best to shift the return series with a constant μ before estimating volatilities performs worse and will therefore be disregarded from this point on.[74]

Since the tenor of the risk factor has a noticeable impact on the chosen model a further study of calibration results in relation to the tenor of the corresponding risk factor suggests itself. Figure 2 plots the average log likelihood value and the average calibrated Degree of Freedom (DoF) of the best performing models against the tenors of the risk factors. As can be clearly seen, likelihoods and DoF rise steadily and significantly in unison with the tenor of the corresponding risk factors. Calling to mind section 2.4 log likelihood is $\prod (T_\nu(z))$ where z is the time series of errors created - in this case - by the best working volatility model for the given risk factor. The likelihoods of high and low tenor time series can be compared directly with each other, since all time series do have the same sample size. The relatively high likelihoods produced by the high tenor risk factors, suggest that their z match the target distribution T_ν closely.

The z calculated based on low tenor risk factors on the other hand probably match

[72]The three year length is chosen based on the result in appendix B.1
[73]Compare eq. (11) on page 18
[74]Additional tests did not find any other notable patterns regarding best working models, when grouping risk factors by their currencies or reset frequencies.

	ARMAGARCH11T	CGARCH11T	EWMAT	GARCH11T	Overall
$\leq$ 6 months	64	9	28	6	107
7 to 12 months	7	2	52	6	67
$\geq$ 2 Years	14	3	226	74	317
Overall	85	14	306	86	491

(a) 8 Years: 2006-2014

	ARMAGARCH11T	CGARCH11T	EWMAT	GARCH11T	Overall
$\leq$ 6 months	54	4	43	6	107
7 to 12 months	4		60	3	67
$\geq$ 2 Years	4		271	42	317
Overall	62	4	374	51	491

(b) 3 Years: 2008-2011

	ARMAGARCH11T	CGARCH11T	EWMAT	GARCH11T	Overall
$\leq$ 6 months	46	8	36	17	107
7 to 12 months	13	2	36	16	67
$\geq$ 2 Years	4	1	242	70	317
Overall	63	11	314	103	491

(c) 3 Years: 2010-2013

	ARMAGARCH11T	CGARCH11T	EWMAT	GARCH11T	Overall
$\leq$ 6 months	44	12	39	12	107
7 to 12 months	8	2	48	9	67
$\geq$ 2 Years	5		275	37	317
Overall	57	14	362	58	491

(d) 3 Years: 2011-2014

Table 4: Count of risk factors, for which a given model performs best according to BIC. Analysis made for different calibration periods.

their desired distribution very poorly. This is also implied by the low calibrated ν for these risk factors. Keeping in mind that the minimal degree of freedom of a student-t distribution is two and that kurtosis rises as ν falls it seems, as if the calibration tries to shape the target distribution as fat tailed as possible to fit it to a time series, which in reality can't be matched well by any student-t distribution.

At this point some interesting properties of the student-t distribution should be mentioned. While a DoF between one and two is possible, the variance of such a distribution is not defined. Since the error term in any GARCH calibration needs to have a mean of zero and a variance of one, a student-t distribution with a degree of freedom below two is out of the question. Additionally the skewness of a student-t distribution is not defined if the DoF is below three and the kurtosis is not defined, if the DoF is below four.[75] This means e.g. that the sample kurtosis does not converge, when sampling from a student-t distribution with less than four DoF.[76]

As a next step, the good performance of *ARMAGARCH11T* models for low tenor risk factors is studied. Their good BICs could have one of two reasons: Either the process is really well depicted by an *ARMAGARCH11T* model, or - as fig. 2 suggests already - none of the models, including *ARMAGARCH11T* captures the movements of low tenor risk factors well. In the latter case *ARMAGARCH11T* models are just chosen, because they are the most flexible out of a set of bad performing models thanks to their relatively high parameter count.

To answer this question, the performance in autocorrelation and distributional tests between *ARMAGARCH11T* and *EWMAT* models for these low tenor risk factors was compared. The Kolmogorow-Smirnow test was used for the test of distributional property, while the Ljung-Box Q test was used, to test for autocorrelation. Both tests were introduced in section 2.3.4. The analysis was made for the time period 2010 to 2013 and the result was, that - with the exception of three out of 46 risk factors, for which *ARMAGARCH11T* performed best - test results are not better when the *ARMAGARCH11T* models are used. In fact, the null hypothesis had to be rejected in almost all Kolmogorow-Smirnow tests and tests for autocorrelation were also quite

[75]Compare Spanos (1999)[Chapter 10.6]

[76]Given a time series z of length n the sample kurtosis is calculated as

$$K = \frac{n-1}{(n-2)(n-3)}[(n+1)(\frac{m_4}{{m_2}^2} - 3) + 6]$$

where m_2 and m_4 are the second and forth moment of the time series respectively. Compare Joanes and Gill (1998)

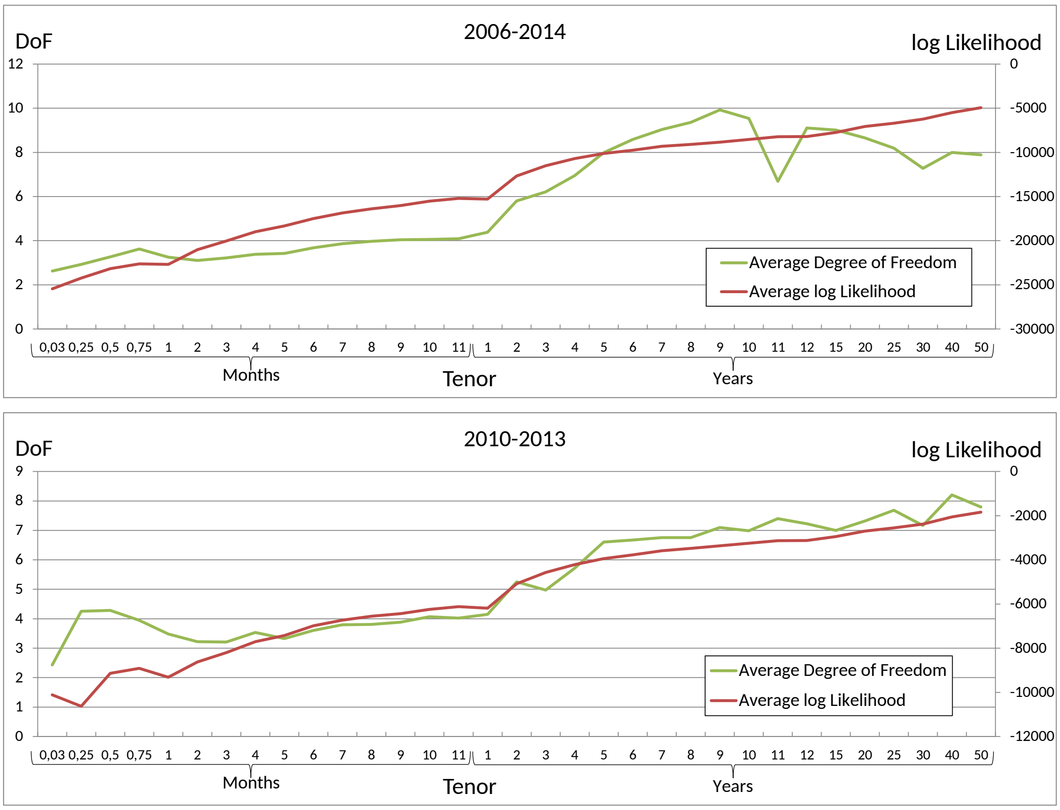


Figure 2: The figure shows for two different calibration periods, how the average optimal degree of freedom of the student-t error term and log likelihood increase as a function of the tenor of the used risk factors.

	EWMAT	GARCH11T	Overall
$\leq$ 1Year	60	7	67
7 to 12 months	90	17	107
$\geq$ 2Year	232	84	316
Overall	382	108	490

(a) 8 Years: 2006-2014

	EWMAT	GARCH11T	Overall
$\leq$ 1Year	47	20	67
7 to 12 months	62	45	107
$\geq$ 2Year	244	73	317
Overall	353	138	491

(b) 3 Years: 2010-2013

Table 5: Amount of risk factors, for which *EWMAT* or *GARCH11T* works best out of the two.

mixed. It seems as if the time series of many low tenor risk factors can't be depicted well by any of the tested models. For this reason the relatively good performance of *ARMAGARCH11T* models for low tenor risk factors will be disregarded at this point.

After narrowing the models down to *EWMAT* and *GARCH11T* the analysis of table 4 is repeated with only these two models. The result may be found in table 5. Apparently most of the low tenor risk factors that were previously best represented by *ARMAGARCH11T* models are covered better by *EWMAT* models than they are by *GARCH11T* models. Further analysis will focus primarily on the *EWMAT* model.

3.3 Simultaneous Modelselection

After determining, which model is best suited for each individual risk factor it is of interest to determine, which models are capable of representing a large portion of these time series. Modeling each time series individually would result in way too many models and parameters to handle operationally. In addition, the time series can't be treated as being independent from each other, since they represent points on a discount curve and therefore are highly dependent at least as long as they originate from the same curve. For this reason a way of simultaneously calibrating a model on multiple time series has been implemented for the two models, which have had the most success during individual calibration - EWMA and GARCH with student-t innovations, both of which assume a constant mean of zero. Since model approaches with normally distributed error terms had no success in section 3.2, these models will not be considered further and will from this point on EWMA will refer to an EWMA model with student-t distributed error terms. When calibrating a common model for multiple time series, the likelihood value can be used in a very similar way as it is used for individual model selection. Given a set of parameters, the log likelihood for each time series can be calculated by adding up the log likelihood of every single point in the time series. In the same way, the likelihood of a given set of parameters for multiple time series can be calculated by adding up the likelihoods of each time series. As an illustrative example the log likelihood of an EWMA model for the three independent time series a b and c with the parametrization $\lambda = 0.94$ and $\nu = 6$ is calculated. First, the three time series are taken and filtered with λ, which returns three time series of filtered innovations which are assumed to be student-t distributed with mean zero, standard deviation one and a DoF of six. The log likelihood of this assumption is afterwards calculated using the student-t log likelihood function with the three time series of filtered innovations as input. The log likelihood of a time series z being student-t distributed with mean zero, standard deviation one and degree of freedom ν is calculated as

$$l_\nu(z) = N log\left[\frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\pi(\nu-2)}\Gamma\left(\frac{\nu}{2}\right)}\right] - \frac{1}{2}\sum_{t=1}^{N} log\left(\sigma_t^2\right) - \frac{\nu+1}{2}\sum_{t=1}^{N} log\left[1 + \frac{z_t^2}{\sigma_t^2(\nu-2)}\right] \tag{27}$$

where Γ is the Gamma function.

Using this likelihood function the likelihood of the parametrization $[\lambda, \nu] = [0.94, 6]$ for the time series a, b and c can be inferred. By setting $r = a$ and $\lambda = 0.94$ in eq. (24) on page 43 an error time series z_a can be calculated. Setting $[\lambda, \nu] = [0.94, 6]$ and $z = z_a$ in eq. (27) one is able to calculate $l_{\nu=6}(z_a)$ and respectively $l_{\nu=6}(z_b)$ and $l_{\nu=6}(z_c)$. The likelihood of an EWMA model with $\lambda = 0.94$ and $T_{\nu=6}$-distributed innovations of the three time series $[a, b, c]$ is then

$$L_{[\nu=6,\lambda=0.94]}(z_a, z_b, z_c) = l_{\nu=6}(z_a) + l_{\nu=6}(z_b) + l_{\nu=6}(z_c). \tag{28}$$

To calibrate a common EWMA model for the three time series the first intuition would be to search the parametrization $[\nu, \lambda]$, which maximizes $L_{[\lambda,\nu]}(\cdot)$ in eq. (28). In the context of the introduced example this would mean finding a λ which results in three time series of innovations $[z_a, z_b, z_c]$, which are most likely under a distribution T_ν. However, in the opinion of the author, there is no reason, why the likelihood of $[z_a, z_b, z_c]$ should be measured with regards to the same distribution T_ν.[77] Since ν primarily influences kurtosis, it seems reasonable, that some risk factors might result in fatter tailed innovations z than others and the model should adjust to this by using a proper DoF for each risk factor. In the context of the example one therefore would

[77] Since literature on the calibration of a single model for multiple independent time series is very scarce, no previous work has been found regarding this particular issue.

use a separate $[\nu_a, \nu_b, \nu_c]$ for each time series and maximize $L_{[\lambda,\nu_a,\nu_b,\nu_c]}(\cdot)$ instead. Once the λ of an EWMA model is calibrated standard deviations are estimated recursively as

$$\sigma_t^2 = (1-\lambda)\epsilon_{t-1}^2 + \lambda\, \sigma_{t-1}^2.$$

The assumed DoF of the innovations only influences the calibration of λ but is not used after calibration for estimating volatilities. The use of specific DoF per risk factor during calibration would therefore still result in only one common EWMA model for all risk factors.

Based on this idea a process has been developed which finds the maximum likelihood estimate $[\hat{\lambda}, \hat{\nu_1}, \ldots, \hat{\nu_n}]$, which is the most likely combination of a common λ and individual ν for n different risk factors. Applied on the example one would search the parameters $[\hat{\lambda}, \hat{\nu_a}, \hat{\nu_b}, \hat{\nu_c}]$, which maximizes $L_{[\lambda,\nu_a,\nu_b,\nu_c)]}(z_a, z_b, z_c)$. The time series of innovations z_a, z_b and z_c are all inferred with eq. (24) on page 43 using the same λ. To describe this process in further detail, the pseudocode of the estimation of a common EWMA model with individual DoFs is shown in fig. 3.

A similar process has been implemented for a GARCH(1,1) process with student-t innovations. For calibration of a model for n different time series a common set of parameters $[\omega, \alpha, \beta]$ and individual DoFs $[\nu_1, \ldots, \nu_n]$ is calibrated. It turned out, that calibrations using a GARCH(1,1) model were very similar to those with an EWMA model, s.t. ω is close to zero, α close to $1-\lambda$ and β close to λ. After this initial result, the main focus was laid on volatility estimation with EWMA.

Based on the BIC the use of individual DoFs seems warranted, since it improved the BIC significantly for samples, for which the two approaches were compared. Using the example again, the likelihood $L_{\left[\tilde{\lambda},\tilde{\nu}\right]}$, where $\tilde{\lambda}$ is the MLE under the assumption of a common DoF would be much lower than $L_{\left[\hat{\lambda},\hat{\nu_a},\hat{\nu_b},\hat{\nu_c}\right]}(\cdot)$. Even though the BIC penalizes the use of four instead of only two variables in the latter case, the BIC would still be better, when individual DoFs are used.

3.3.1 Results of simultaneous EWMA-Modelselection

The BIC has been introduced in eq. (11) on page 18 as a function $BIC(L, k, n)$ of likelihood parameter count and sample size. In contrast to likelihoods, models resulting in lower BIC are preferred over such with higher ones. In the following we will omit the sample size n, since all comparable BIC need to have the same sample size anyway. Based on this information criterion it is possible to analyze, how many different models should be calibrated for a set of risk factors and which risk factors should be grouped together for calibration. Let C be the set of all risk factors and A and B subsets, s.t. $A \cap B = \emptyset$ and $A \cup B = C$. The process proposed in fig. 3 can calibrate an optimal

```
function [ lambda, llf] = PFEWMATCalib(PortfolioReturns)
    %PortfolioReturns is a Matrix of the returns of multiple risk factors. The goal is to
    calibrate a common optimal Lambda according to an EWMA model with student-t innovations.

    lowerbound=0;
    upperbound=1;

    [lambda,llf]=fminbnd(@(lambda) loglikPFEWMAT(lambda, PortfolioReturns), lowerbound,
        upperbound, options);
        %Optimization of Lambda with the target function loglikPFEWMAT
        %fminbnd is a proprietary matlab function for optimizing one, bounded parameter
end

function [ negPFllf ] = loglikPFEWMAT(lambda, PortfolioReturns)
    %Given a fixed lambda and a return matrix of positions in a portfolio loglikPFEWMAT
    returns the negative likelihood of this Lambda. To do this an optimal individual degree
    of freedom for every single risk factor conditional on the common lambda is determined.
    After the common Lambda and individual DoF are known log likelihood of each risk factor
    is calculated with the time series and these log likelihoods are summed up to yield
    maximal portfolio log likelihood for the given lambda.

    [timeserieslength, assetcount]=size(PortfolioReturns);

    for i=1:assetcount
        lowerbound=2;
        upperbound=200;

        [DoF,negllf]=fminbnd(@(DoF) loglikewmat(PortfolioReturns(:,i), lambda, DoF),
            lowerbound, upperbound, options);
            %Optimization of the Degree of Freedom of an individual risk factor with the
            target function loglikewmat

        negPFllf=negPFllf+negllf;
    end

end

function [llf] = loglikewmat(returns, lambda, DoF)

    %Calculates log likelihood of Parameterset [Lambda, DoF] for a given time series
    %The used EWMA process assumes the mean return to be zero. Starting value of the
    variance estimation process is the mean of squared returns of the entire time series.

    e(1)=sqrt(mean(returns.^2));
    h(1)=mean(returns.^2);

    for i=2:(length(returns)+1)
        e(i)=returns(i-1);
        h(i)=(1-lambda)*e(i-1)^2+lambda*h(i-1);
    end

   %Innovations are deducted by dividing squared returns e^2 through estimated variance h.
   Student-t log likelihood of these estimation is calculated as presented in eq. (25)
   LogLikelihoods  =  0.5 * (log(h)  +  (DoF + 1) * log(1 + (e.^2)./(h * (DoF - 2))));
   LogLikelihoods  =  LogLikelihoods - log(gamma((DoF + 1)/2) / (gamma(DoF/2) * sqrt(pi *
       (DoF - 2))));
   llf            =  sum(LogLikelihoods);

end
```

Figure 3: Pseudocode Calibration of Portfolio

EWMA model and the corresponding log likelihood value for each one of these sets. The count of risk factors in a set plus one equals the necessary parameter count for an EWMA model for the entire set - one parameter for the optimal DoF of each risk factor and an additional parameter for the decay factor λ of the EWMA model.

Unions of disjoint sets, which cover the same amount of risk factors, can be compared with each other using eq. (11). In the example $A \cup B$ is a union of disjoint sets, which covers the same amount of risk factors as the set C. Since the union $A \cup B$ has the same sample size as the set C they are comparable with the BIC. After an optimal EWMA model has been calibrated for each of the sets A, B and C the likelihood of these models L_A, L_B and L_C are known and the corresponding amount of parameters is $k_A + 1$, $k_B + 1$ and $k_C + 1$, where k is the number of risk factors in a set.

Therefore the BIC of the two sets A and B together is

$$BIC_{[A,B]} = BIC\left((L_A + L_B), (k_a + k_b + 2)\right)$$

and the BIC of the set C is

$$BIC_C = BIC\left(L_C, (k_C + 1)\right).$$

If $BIC_C > BIC_{[A,B]}$ it is statistically feasible to rather use two distinct EWMA models - one for the set A and one for the set B - than a single one for the entire set C. Assuming that $BIC_{[A,B]} < BIC_C$ one might still be able to find two sets D and E, s.t. $D \cap E = \emptyset$ and $D \cup E = C$ for which

$$BIC_{[D,E]} = BIC\left((L_D + L_E), (k_D + k_E + 2)\right) < BIC_{[A,B]}$$

or even three disjoint sets F, G and H, s.t. $F \cup G \cup H = C$, for which

$$BIC_{[F,G,H]} = BIC\left((L_F + L_G + L_H), (k_F + k_G + k_H + 3)\right) < BIC_{[A,B]}.$$

By trying out all possible unions of disjoint sets, which equal C one could find a combination of sets, which has the lowest BIC value and therefore offers the best tradeoff between a high log likelihood and parsimony.

With this approach one is able to search for a combination of disjoint sets, which has a relatively high sum of BIC. Since the set C, encompassing all risk factors, consists of 517 risk factors the number of different possibilities of dividing C into disjoint subsets is extremely large. To overcome the task nevertheless, a few rules had to be put in place for the creation of subsets.

1. C may at maximum be split into four disjoint subsets.

2. One discount curve may at most be divided into three interconnected sections along the tenor dimension.

3. If curves are divided into multiple sections all curves to be divided at the same tenor.

4. The analysis has only be done for one calibration period and assuming, that an EWMA model works well for all subsamples.

To make these rules more tangible a possible split would be:

1. All risk factors of the currencies € or £ and tenor < 12 months

2. All risk factors of currencies \$, ¥, CHF and tenor < 12 months

3. All risk factors with a reset frequency of one day, one month or three months and a tenor $>= 12$ months

4. All risk factors with a reset frequency of six months or twelve months and a tenor $>= 12$ months

and as one can see in table 3 on page 42 the first set would comprise 76 risk factors, the second 104, the third 231 and the fourth 106.

Performing this analysis on the January 2010 to January 2013 dataset unveiled, that to minimize the BIC it is most important to separate the risk factors with high tenors from those with low tenors. Apparently the low tenor risk factors behave similar across different discount curves, while low tenor and high tenor risk factors of the same curve behave very different. The highest BIC of a split into four subsets was observed for the following split:

1. All risk factors with a tenor ≤ 5 months

2. All risk factors with a tenor between 6 and 12 months

3. All risk factors with a tenor above 12 months and of the currencies \$ or £

4. All risk factors with a tenor above 12 months and of the currencies €, ¥ or CHF.

During the analysis an interesting relationship between the decay factor λ and the tenor of the risk factors it was calibrated of was observed: For risk factors up to one year λ rises steadily in unison with the tenor but then keeps a relatively steady level for risk factors above one year. This behavior is displayed in fig. 4. In conjunction with the previous result this indicates, that the tenor is the most important metric to

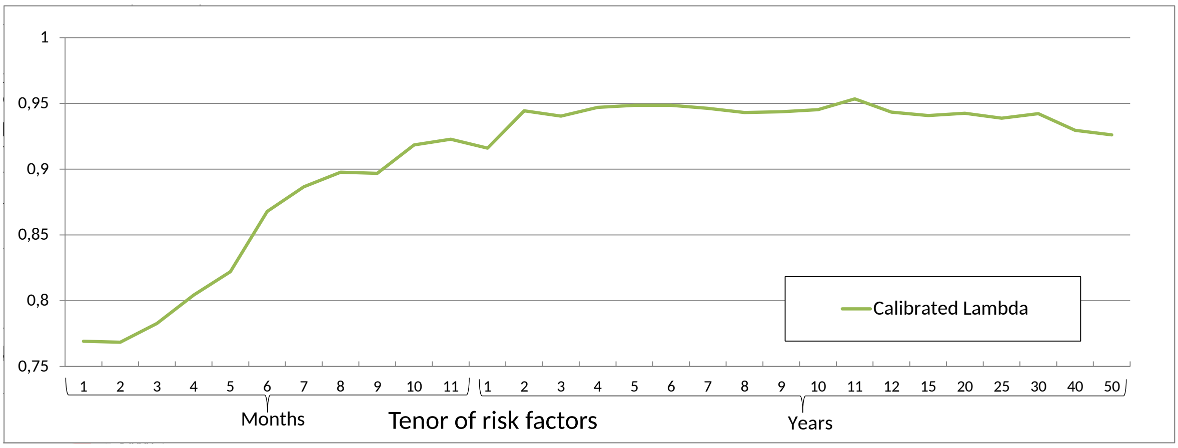


Figure 4: Value of the optimal decay factor λ for subsets, which consist of all risk factors with a common tenor value.

differentiate between risk factors with a tenor of one year or less. On the other hand, based on the identified optimal split, this doesn't seem to be the case for risk factors with a tenor above one year. For this reason, a quick comparison is made to test if risk factors with a tenor above one year are best differentiated based on the currency, reset frequency or tenor.

All risk factors have been split into subsets in three different ways: First all risk factors with a common tenor were grouped together, just like they were in fig. 4, secondly all risk factors with a common currency and a tenor above one year made up a subset and thirdly all risk factors with a common reset frequency and a tenor above one year made up a subset. These three approaches are compared in table 6 on the following page. As can be seen a split with regard to the tenor results in the highest likelihood for risk factors with a tenor of one year or below, while a split regarding the currency results in the highest likelihood for risk factors with a tenor above one year. Interestingly enough, a split with regard to the reset frequency has the best BIC value for low tenors, due to its use of fewer parameters.

The main result of this section is that - based on the likelihood value - it is most important to differentiate between risk factors of the short end and those of the long end of the discount curves. Additionally, evidence was found that this differentiation should be made at the one year tenor point. This would suggest to use different volatility models for risk factors with low tenors and such with a high tenor. However, it might have unwanted results to use different volatility models along a single discount curve, which leads over to the next section.

Split approach	Number of calibrated models	Δ LLF	Δ BIC
One subset per tenor	16	0	0
One subset per currency	5	-362,48	595,04
One subset per reset frequency	4	-35,06	-71,60
One set ≥5, one ≤ 12 months	2	-111,46	57,57

(a) Risk factors with a tenor of one year and below

Split approach	Number of calibrated models	Δ LLF	Δ BIC
One subset per tenor	16	0	0
One subset per currency	5	6,19	-161,31
One subset per reset frequency	5	1,62	-152,17

(b) Risk factors with a tenor above one year

Table 6: Comparison of different split approaches. Δ are calculated with the split into tenor subsets as reference. As can be seen, a division into single tenor groups is best with regard to likelihood for risk factors with a tenor of one year and below, while a division into reset frequency is best for these risk factors if decisions are based on the BIC. For risk factors above one year a division according to currencies is best suited with regards to both metrics.

3.4 The use of multiple volatility models across one discount curve

The use of different volatility models for risk factors of a common curve - as the previous section's results would suggest - might be problematic. It is the goal, to estimate VaR of FRAs and since a FRA matures, the risk factors, which are points on a discount curve, used to calculate its value move down along the discount curve. Since a discount curve is represented by multiple risk factors, the risk factors used for pricing a FRA do change as the product matures. The use of different volatility models along the tenor dimension of a discount curve would therefore mean, that as the product matures, the volatility model of its risk factors would change. This could result in significant VaR jumps, which would be purely model induced and without economical explanation.

As an example, we will assume, that volatilities of a discount curve are modeled with an EWMA model with $\lambda = 0.84$ for all risk factors with a tenor below one year and an EWMA model with $\lambda = 0.94$ for risk factors with a tenor of one year or above. Referring to table 3 on page 42 it is apparent, that the used volatility model switches, as the relevant risk factor transitions from the one year risk factor to the eleven months risk factor. To understand the connection between the different λ of the two EWMA models and the estimated VaR it is best to focus on the parameter σ^* in eq. (1) on page 7. σ^* is the volatility used for scaling of the filtered returns - in the context of estimating the current VaR of a product it is the currently estimated volatility of a risk factor. Given a return time series r of a risk factor and an estimated λ of the model the current volatility of the risk factor σ_T is calculated recursively as

$$\begin{aligned} {\sigma_T}^2 &= (1-\lambda) r_{T-1}^2 + \lambda\, \sigma_{T-1}^2 \\ \text{with start value} \quad {\sigma_1}^2 &= r_0^2 \end{aligned} \tag{29}$$

If this recursive structure is resolved it becomes apparent, that σ_T^2 is a weighted sum of past squared returns, where the weight of the $(T-i)^{\text{th}}$ squared return is

$$\frac{(1-\lambda)^i}{\sum_{j=1}^{j=T} (1-\lambda)^j} \ .$$

In the just introduced example λ would switch from 0.94 to 0.84 between the 11 and 12 months tenor risk factor. Therefore it is worth exploring the two risk factors r_{11M} and r_{12M}, where the subscript is the tenor in months, which are calculated as shown in eq. (20) on page 41. With r_{11M} and r_{12M} being the returns of two adjacent tenor points of the same discount curve they will not be exactly the same, but can be expected to be very similar. This thesis will not explore how a discount curve could or could not move on a day to day basis at this point, but it should still be fair to assume, that the movements of adjacent tenor points on the same discount curve are similar in magnitude and direction.[78] For this reason r_{11M} and r_{12M} should be quite similar and a common volatility model would predict very similar but not equal current volatilities $\sigma_{11M,T}$ and $\sigma_{12M,T}$ for them. However, if a different λ parametrization is used for the calculation of $\sigma_{11M,T}$ than for $\sigma_{12M,T}$ the volatilities can differ significantly despite the similarity between r_{11M} and r_{12M}. Due to its lower λ the calibration of $\sigma_{11M,T}$ will be more dependent on the more recent past than the calibration of $\sigma_{12M,T}$ - a model with a lower λ is more reactive than one with a higher λ. If the market currently cools down, more recent squared returns can be expected to be lower than such further in the past which would result in a considerably lower $\sigma_{11M,T}$ than $\sigma_{12M,T}$. On the other hand, if the market would have been homoscedastic for a while, $\sigma_{12M,T}$ and $\sigma_{11M,T}$ should be fairly similar, even if they are calculated based on different λ. Since filtered returns with a standard deviation of one are scaled up with σ_T a higher σ_T directly results in larger scenario returns of the risk factor and therefore probably a higher estimated VaR (the latter relation depends on the pricing formula of the specific product).

To test, how much impact this anticipated problem could have VaR calculated based

[78] If the reader is interested in possible discount curve movements Hagan and West (2006) published a survey on this subject.

	Currency	Tenor	Resetfrequency	Calibrated λ
Scenario 1	Euro	≤ 2 Months	1 Day	0.8093
	Euro	> 2 Months & ≤ 1 Year	1 Day	0.8953
	Euro	> 1 Year	1 Day	0.9592
	Euro	≤ 5 Months	1 Month	0.8699
	Euro	> 5 Months & < 1 Year	1 Month	0.9293
	Euro	≥ 1 Year	1 Month	0.9519
Scenario 2	Euro	> 1 Year	All	0.9519

Table 7: Parametrizations of EWMA models with and without splitting risk factors of a common discount curve. The calibration period is June 2007 to June 2010. The λ calibrated for scenario one applies to all € risk factors even though it has been calibrated based only on those with a tenor above one year.

on only one volatility model per discount curve are compared with such based on two volatility models per discount curve. VaR of € FRAs based on the one month reset frequency curve were tested. As explained in section 2.7, such a FRA has the three risk factors $d_{€,1M,t1}$, $d_{€,1M,t2}$ and $d_{€,1D,t2}$. We will calculate VaR under two different scenarios. In the first, six different volatility models are used - three EWMA models for the one month € discount curve and three for the OIS curve, while in the second, the same EWMA model with the same λ is used for all € risk factors. The optimal parametrizations for both scenarios are displayed in table 7.

Under the assumption of each scenario VaRs were calculated for each day of the year 2011 for FRAs with different t_2.[79] The chosen t_2 range from two months to 25 months in the future with a one month increment. When e.g. calculating VaRs for the 1st of March 2011, we would estimate the VaRs for 24 different products, the first being a FRA, which expires at the 1st of May 2011 and the last being a FRA expiring on the 1st of April 2013.

As would be expected the results were dependent on the day of observation. For some days the different λ did not make a difference, since market conditions had been homoscedastic for a sufficient period of time. For others, the difference in the estimated VaRs was very large between the two volatility scenarios and the VaR jumped noticeably when transitioning from one λ parameter to another. An example of the latter case would be the 14th of April 2011, for which the VaR for the 24 FRAs with different maturities are plotted in fig. 5 on the following page. The graph for scenario two is relatively smooth, while under scenario one the VaR estimation drops signifi-

[79]The choice of the year 2011 is arbitrary. These FRAs and all others in the empirical section are calculated with a nominal L of 2.5 million and an interest rate R_K of 2.6% continuously compounded. Since these two values are not risk factors as explained in section 2.7 they scale the VaR magnitude but don't influence the estimation process.

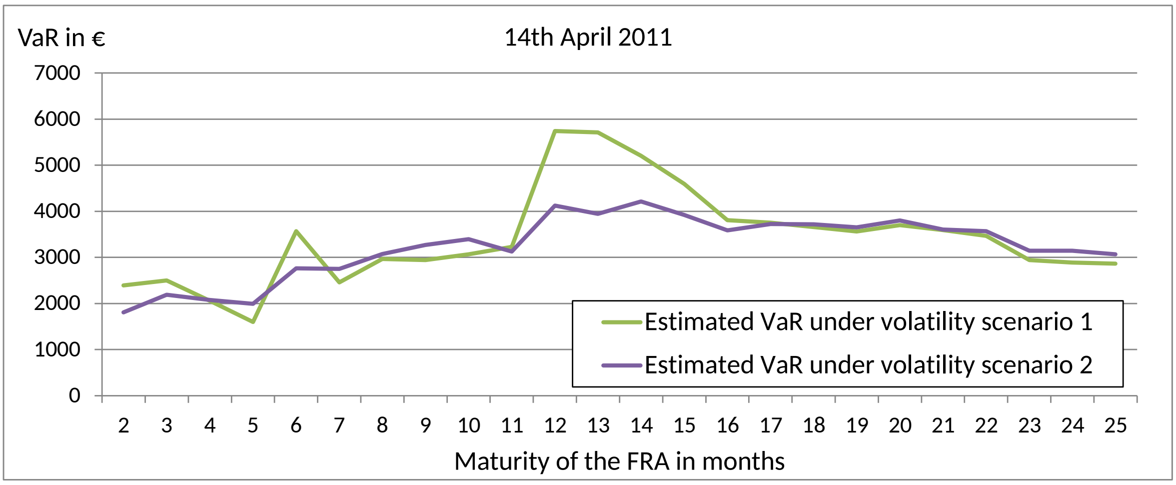


Figure 5: The graph shows the estimated VaR of FRAs with different maturities at the 14^{th} of April 2011. If VaR is estimated with volatility scenario two, jumps of the VaR are clearly visible at the junction of different EWMA calibrations of the 1 month € curve.

cantly from the maturity 12 month to the maturity 11 months FRA as well as from the maturity 6 month to the maturity 5 months VaR. It stands out, that these jumps happen to be exactly at the junction of different EWMA parametrizations of the one month € discount curve. While a deviation at the junctions was to be expected the magnitude certainly comes as a surprise - the VaR of the FRA with a twelve month maturity was estimated to be more than 40% higher under volatility scenario one than under volatility scenario two.

To show that the VaR impact of different λ depend largely on the point of observation, a comparison between some days in April 2011 and some in November 2011 is displayed in fig. 6 on the next page. In April, market conditions are currently changing, which produces large jumps at the junctions of the different λ used under scenario one. These jumps do not appear when using only one λ for the entire curve, as it is done in scenario two. At the start of November on the other hand VaR estimates are very similar under both volatility scenarios, indicating a recent homoscedastic behavior of the market, in which the currently estimated volatility is relatively insensitive against the chosen decay factor.

The behavior under scenario one is unacceptable for two reasons: First, a FRA will mature which would in our example mean, that the VaR of a FRA with a maturity between eleven and twelve months would drop rapidly just due to the switch of the volatility parametrization. Secondly and more significantly this could have hardly predictable effects on portfolios with approximate hedges: Assuming in our example

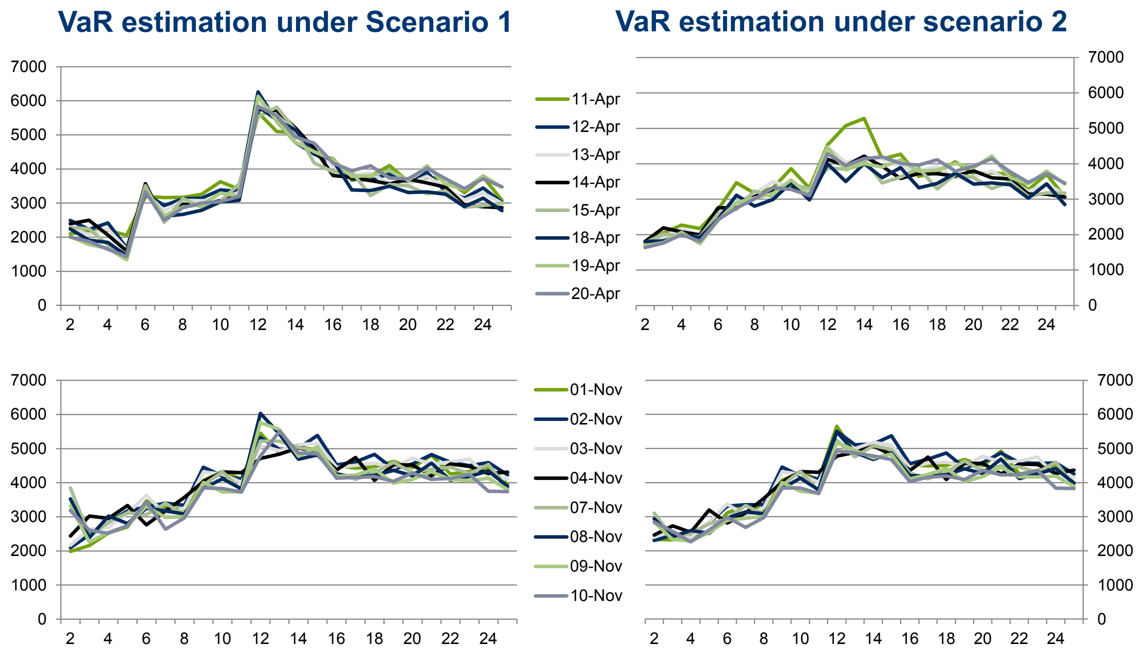


Figure 6: The graph shows the VaR of FRAs with different maturities at a few days in April and some in November. On the left volatility scenario 1 is applied, on the right scenario 2. While for the samples of April the junction of the different λ can be clearly made out as spikes the two scenarios produce very similar VaR in November.

that a customer would hedge his borrower position in a twelve month FRA with a lender position in an equivalent eleven month FRA the following issue would occur: Since the returns of the risk factors of the customers two positions would be very similar and the valuation functions differ only in sign, he would expect the returns of the two products to always roughly offset each other leading to a very small portfolio VaR. However, due to the different λ, the two filtered return time series could be scaled up with very different standard deviations, and therefore the offset between the two products would be much less exact leading to a significant overestimation of VaR.

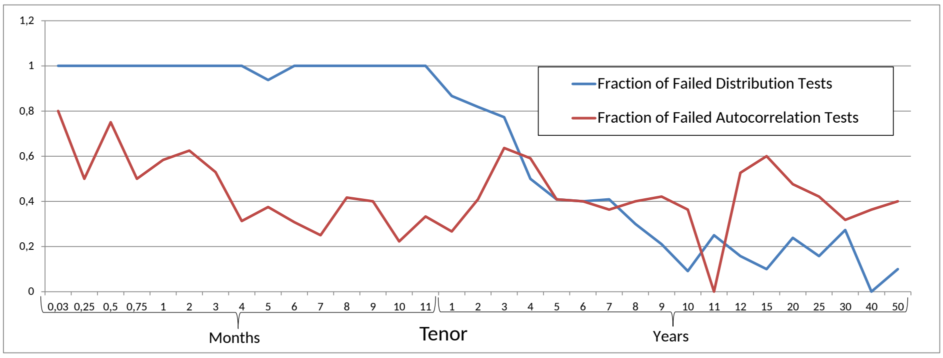
Due to this issue it is inadvisable to use different volatility models for risk factors of the same discount curve and this thesis will therefore focus the estimation of volatility models for entire discount curves from this point on. This does contradict the best working division into subsets presented at the end of section 3.3, however, the BIC based selection approach, which was motivated in section 3.3 can be used nevertheless, if an additional constraint is added, which states, that risk factors of a common discount curve cannot be split across multiple subsets.

Using a common model for a discount curve might however negatively impact the main target of volatility modeling - the removal of autocorrelation from the time series of squared returns. Additionally the model assumption of properly distributed error times might be more frequently violated, when a common model is used for the entire curve.

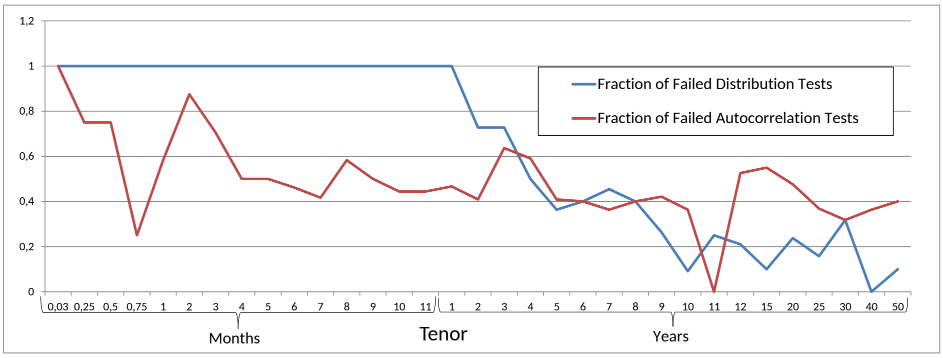
For this reason it has been tested, if volatility scenario one, two and a third scenario using individually calibrated EWMA models for each risk factors yield different results in tests for autocorrelation and distributional fit. The used Ljung-Box Q and Kolmogorow-Smirnow test were introduced in section 2.3.4. Results are displayed in fig. 7 on the following page. As can be seen, the use of volatility scenario two does not noticeably worsen the results of the autocorrelation and distributional tests. In fact, results of these tests are already quite bad, if a separate model is calibrated for each risk factor. Especially for the risk factors below one year it is striking, that almost no risk factor fulfills the distributional properties. With regard to the autocorrelation we need to mention, that even filtered time series, which fail the Ljung-Box Q test show significantly less autocorrelation than their unfiltered counterparts. This was confirmed for some sample risk factors, for which autocorrelation charts of the unfiltered and the filtered squared returns were compared.

3.5 Calibration results for an entire discount curve

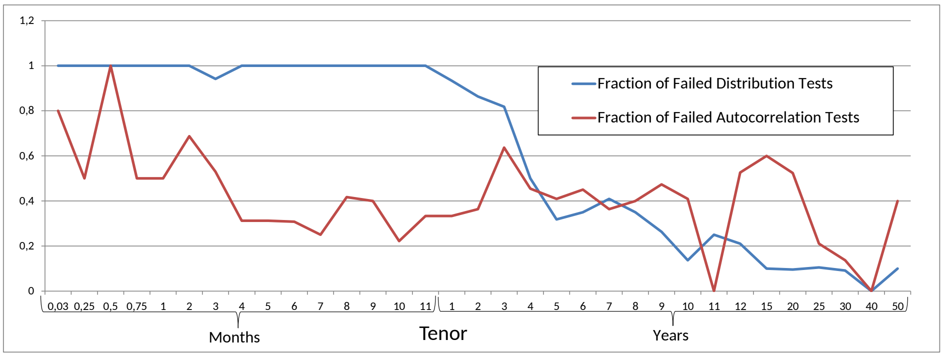
Due to the result of section 3.4 the next goal is to further investigate the calibration results of entire discount curves. Since no evidence up to now has been found that



(a) Three λ per curve - volatility scenario one



(b) One λ per currency - volatility scenario two



(c) Up to 33 λ per curve - one model per risk factor

Figure 7: Impact of the use of volatility scenarios one or two on autocorrelation of filtered squared returns and distributional property of the errorterm. Autocorrelation is tested for with the Ljung-Box Q test. The test for distribution is done with a Kolmogorow-Smirnow test against a Student-t distribution with the calibrated optimal DoF. All tests are rejected based on a 1% confidence interval. Displayed is the fraction of risk factors with a specified tenor for which the null assumption of a test had to be rejected.

would discourage from the use of EWMA models with t-distributed error terms, we will exclusively focus on these models due to their simple structure and high relevance in practice. The approach to calibration in this section is exactly the same as it was introduced in section 3.3, but risk factors of the same discount curve cannot be part of different subsets.

For the entire history available EWMA models were calibrated for each discount curve at a yearly frequency based on a three year calibration period. Additionally, common models for discount curves of the same currency as well as a common model for all discount curves were calibrated.

Due to the results in section 3.3 it became apparent, that risk factors with a tenor of one year or less behaved differently than such with a higher tenor. Calibrating a common model using all risk factors of a discount curve should therefore yield unstable parametrizations with relatively large confidence bounds, since the algorithm tries to fit the model to return time series, which behave very differently. Due to this reason and the fact, that section 3.2 has already shown, that it is even very hard to find well working models for *individual* risk factors on the short end, risk factors with a tenor of one year or below are omitted during calibration of a discount curve's parametrization. We have made the same adjustment in the previous section already, without mentioning it explicitly - volatility scenario two is calculated based on all € risk factors with a tenor above one year but still used for all € risk factors. Additionally confidence bounds were calculated as shown in section 2.4.2 for the individual discount curve parametrizations. The results can be seen in fig. 8 on the next page. The figure shows the individually calibrated λ of the different discount curves and their confidence intervals. Examining the figure, the calibrated models generally seem to be similar across the different reset frequencies of a common currency - especially for the € and the £ curves. Additionally the yellow line, which represents the λ calibrated for the entire currency, lies within the confidence intervals of the calibrations of the individual reset frequencies in most cases. Since 95% confidence intervals are displayed, this means, that the use of the λ calibrated for the entire currency instead of the λ calibrated for the individual reset frequency can't be rejected at the 95% confidence level. Taking the $ discount curves of the year 2013 as an example, one could just as well use $\lambda = 0.958$ for the four discount curves with different reset frequencies instead of bothering to use the individually calibrated λ - which would actually differ quite a bit - if one would take a 95% confidence level as the foundation of his decisions. As fig. 8 shows, this could be done for the clear majority of discount curves. Only a few curves are not well represented by the respective λ of the entire currency including the Swiss franc after it was bound to the Euro in September

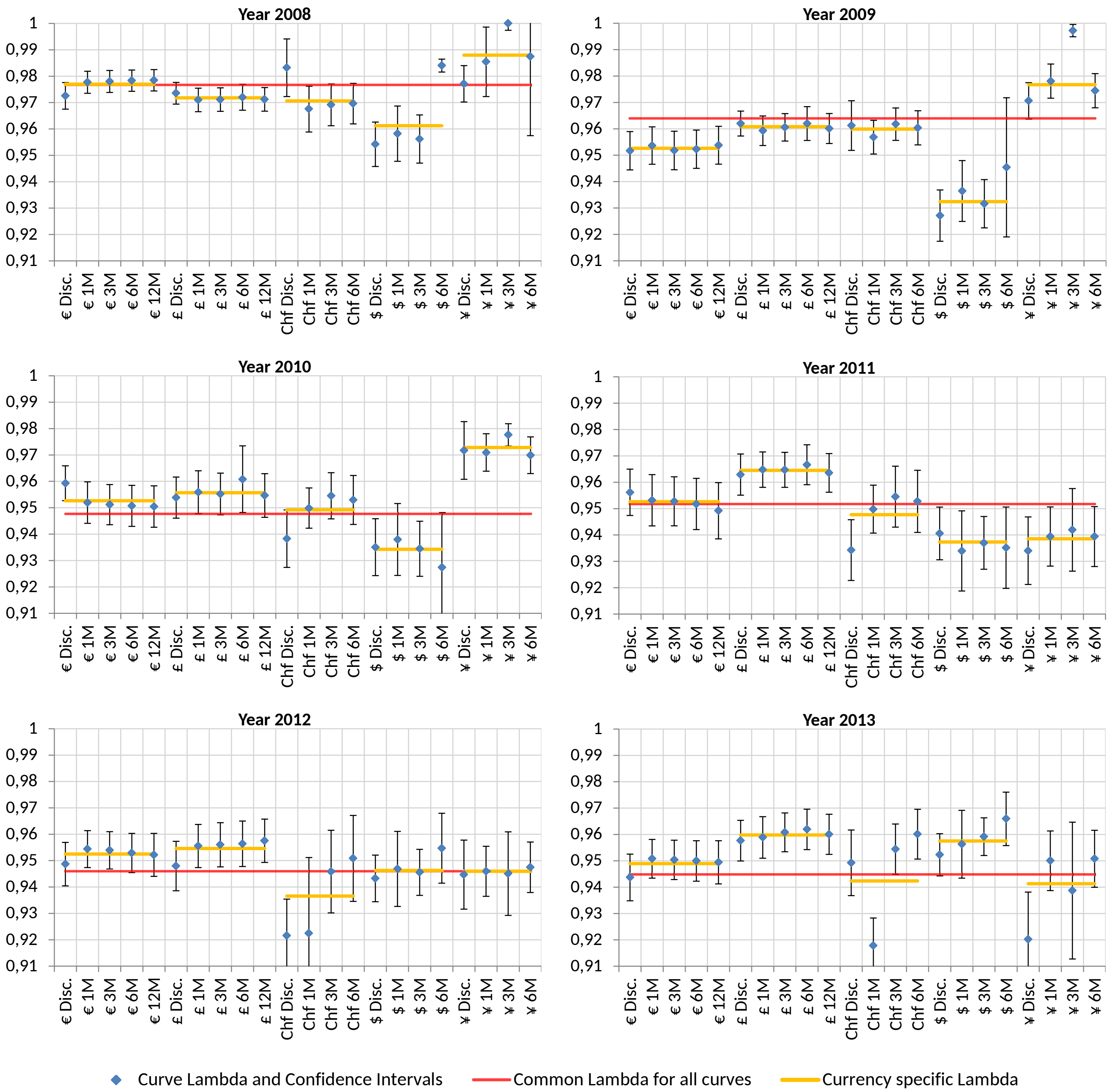


Figure 8: Relation between curve specific λ and λ calibrated across multiple curves: Displayed in blue are the calibrated λ of a curve and their respective 95%confidence Intervals. For comparison, λ calibrated across an entire currency are displayed in yellow and the λ calibrated simultaneously across all curves is displayed in red. Calibrations are based on all riskfactors with a tenor above one year over a three year period ending on the 31st of May of the respective year.

2011.[80]

Figure 8 also explains the result regarding the ideal way of dividing risk factors into subsets in section 3.3. The proposed result included a subset of all risk factors with a tenor above one year of the currencies \$ and £ and another subset of all risk factors of the currencies ¥, € or CHF. Since the used calibration period in section 3.3 reached from 2010 to 2013 we need to focus on the 2013 chart in fig. 8. Indeed λ calibrated for the £ and \$ discount factors are very similar at about 0.96, while those of the other three currencies hover around 0.945. However, fig. 8 also shows, that this relation does not stand the test of time. Across the six different calibration periods no clear relations between different currencies can be discovered. While discount curves of a common currency tend to produce similar result to each other for each calibration period, this does not seem to be the case between entire different currencies. The only exception might be the currencies € and £, for which it should be possible to find a common decay factor λ without violating too many confidence bounds of the individual discount curves. Overall, the confidence bounds calculated in fig. 8 advice against combining different currencies for λ calibration.

The simplest, and certainly most practical, approach to volatility modeling for the up to 517 risk factors would be to use the same volatility model for every single risk factor. If risk factors with a tenor equal or below one year are omitted, as they also were in the analysis before, the resulting λ for each calibration period is displayed as the red line in fig. 8. As can be seen, the common λ violates the confidence bound of the individual discount curves very frequently and therefore this extreme simplification can't be recommended based on the confidence bounds of the calibrations of individual discount curves.

However, the estimated confidence bounds could be narrower, than they need to be in practice: The aspiration of the VaR models, which were introduced in section 2.2 is to estimate the entire PnL distribution of a portfolio and afterwards a certain percentile of this distribution is returned as the VaR estimate[81]. During calibration λ is calibrated in such a way, that all returns - no matter if common or extreme - are well captured by the VaR model. However, only a few of these returns are relevant for the estimation of the e.g. 1% percentile of the most severe losses, which is the returned estimated

[80] The calibration results of the Swiss franc are very similar to the euro from 2008 to 2011 while margins of error significantly increase and λ of the individual reset frequencies differ from each other as soon as the time period after the central bank intervention is regarded in calibration. Calibration results of Yen time series and the six month USD reset frequency for the years 2008 to 2010 have to be regarded with caution due to issues with the underlying data. On the 6$^{\text{th}}$ September 2011 the Swiss national bank declared, that it will no longer accept an exchange rate CHF:EUR below 1.20€. Before the intervention the CHF noted at about 1 €. Compare Swiss National Bank (2011)

[81] This aspiration led to the VaR performance metric presented in Berkowitz (2001), which regards multiple different percentiles of the return distribution instead of only the desired VaR percentile

VaR. For this reason, a volatility model using a λ outside the confidence bound of a certain discount curve might still result in well estimated VaR percentiles of the return distribution. To properly estimate a certain percentile of the return distribution is a weaker requirement than estimating the entire return distribution and the confidence bounds displayed in fig. 8 are related to the latter requirement.[82]

3.6 Backtests

As introduced in section 2.6, the common way to assess the quality of a VaR estimation model, once it has been calibrated using MLE, is to run backtests with it. Different volatility models would have resulted in different VaR estimates in the past, which can be compared with the realized returns as explained in section 2.6. All backtests were performed over the longest possible time period, which the available data allowed. Backtests generally struggle for statistical significance due to the low expected number of exceptions. Since VaR in this thesis are estimated for the 0.99 percentile one would expect one exception over the curse of a hundred days. Since data was available starting 2005 and three preceding years are required for an initial calibration, backtests reach from mid-2008 to mid-2014, which results in about 1300 daily observations. However, for products based on low tenor risk factors backtests cannot be performed before 2009 due to insufficient data quality. Therefore backtests of products with a low tenor unfortunately can't include the financial crisis of 2008. In contrast to section 2.7 a different notation will be used for the FRAs. Since the backtested products tend to have a very high t_2 and a comparatively low $t_2 - t_1$ the usual notation of t_1/t_2 is replaced by $(t_2 - t_1)/t_2$ therefore a 117/120 FRA would be a 3M/10Y FRA instead, where $t_2 - t_1$ is three months and t_2 is ten years.

Backtesting is a computationally expensive procedure and estimating the VaR of 1300 days takes longer than a calibration using MLE. Therefore the backtest of a single product using a single volatility model can take the better part of a day. For this reason the range of backtested products (and their associated risk factors) and volatility models are a lot smaller than during MLE based model selection due to time constraints.

Overall 15 different products were backtested, which were based on three different discount curves across three different currencies. For each product a variety of calibration approaches were applied, all of which are EWMA models. Calibrations during

[82] The fact that the VaR models presented in this thesis try to estimate the entire return distribution, while the actual target value is just a single percentile of said distribution might very well be pointed out as a weakness of these models. Alternative approaches, which focus exclusively on the estimation of the target percentile by using extreme value theory, have been proposed. Compare e.g. McNeil and Frey (2000).

each backtest are renewed once a year in July and therefore the used λ has never been calibrated more than a year ago and no information unknown at the point of observation is used for λ calibration.

3.6.1 Backtests with a constant discount factor

To get an initial idea of backtesting and how VaR and the used volatility model interact a floating rate agreement on the one month € discount curve with a maturity of ten years is backtested. The backtesting period covers the time from July 2008 to June 2014. Based on the known movements of the discount curves the price movement of the described FRA maturing in 10 years is calculated and compared to the calculated Value at Risk. It is important to note, that the product does not mature during backtesting, it is instead rolled daily, such that the FRA in this example always matures in exactly 10 years at any point of observation. The value at risk in January 2009 is calculated based on the filtered risk factor return between January 2006 and January 2009. Volatilities used for filtering as displayed in eq. (2) on page 8 are estimated with an EWMA model, which is specified by its λ.

To start with, five different EWMA models with constant λ are specified. Theses λ are chosen naively to cover a range of reasonable values. A graphical representation of the result can be found in fig. 9 on the next page.[83] The decreasing reactivity with rising λ can be clearly seen. The model, which stands out the most, is the model with a λ of one. $\lambda = 1$ is actually an homoscedasticity assumption and therefore results in classical historical simulation as described in section 2.2.[84] One of the biggest problems with historical simulation can also be seen in this chart: In October 2011 the Value at Risk of the historical simulation drops significantly, while the VaR of filtered historical simulation models rises. The reason for this is, that the stress period of 2008 drops out of the three year time window. Since with a λ of one, returns in October 2008 have exactly the same impact on the estimated current volatility as a return yesterday the estimated current volatility drops as soon as the most stressful days of October 2008 drop out of the timeseries and with a drop in volatility the predicted Value at Risk drops as well. This also shows, that while the length of the time window used for volatility estimation is of great importance for historical simulation its importance for volatility estimation is comparatively smaller for filtered historical simulation. Spikes in the estimation of EWMA models with $\lambda < 1$ are primarily caused by an increase of the estimated current volatility caused by recent large losses or gains.

[83] VaR displayed are actually one week averages to improve readability. This does conceal a few spikes in VaR for models with high reactivity.

[84] This relation between FHS and HS is e.g. described in Gurrola-Perez and Murphy (2015)

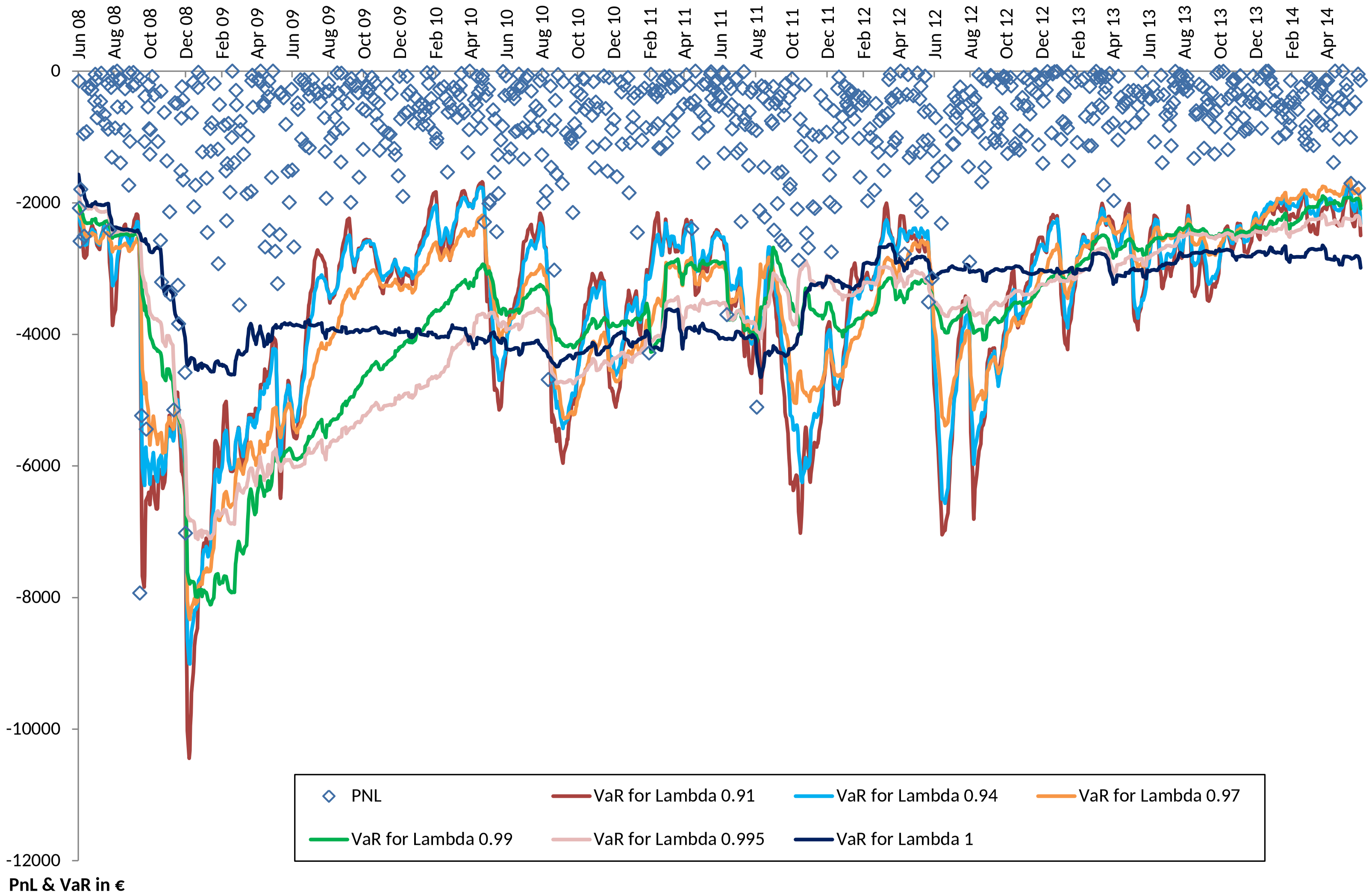


Figure 9: Backtesting results of different constant λ factors. It can be seen, that VaR rises quicker and drops quicker for lower λ reflecting the higher reactivity. Using a λ of one is the same as doing historical simulation without any volatility filtering. Since only the negative scale is shown only losses are displayed. Whenever a loss is larger than the estimated VaR a so called exceedance occurs.

While knowing the daily estimated VaR of each model is interesting and of high practical importance its performance has to be derived from the exceedances it creates. As described in further detail in section 2.6 already, an exception in backtesting occurs, if the tested portfolios hypothetical loss on a past date would have been higher than the estimated VaR of this day. Using die Kupiec proportion of failure and the Christoffersen test, the amount and frequency of these exceptions can be used to assess the quality of a VaR model. The amount of exceptions and results of these tests are, among others, presented in table 8 on the following page. Corresponding values to fig. 9 span from the 4^{th} to the 8^{th} row and the HS case is displayed in row 41. As can be seen, all models pass the Kupiec test, which - with a sample size of about 1300 and a confidence level of five percent - means, that they produce no less than six and no more than 22 exceptions over the entire backtesting period. The Christoffersen tests rejects independence and conditional coverage for the historical simulation with $\lambda = 1$, while all filtered historical simulations pass the Christoffersen tests. When applying the more sophisticated duration based test from Christoffersen & Pelletier and the CAViaR performance measure by Engle & Mangnelli FHS methods with very low reactivity due to the use of $\lambda = 0.99$ and $\lambda = 0.995$ fail a few metrics, which don't reject their more reactive counterparts.

The biggest result, that can be taken away from this initial test is, that FHS seems to work very well and that its backtesting results are not very sensitive to the parametrization of the EWMA volatility model. The range of 0.91 to 0.97, for which the 1M/10Y € FRA performed well in all performance metrics, is much wider than the calculated confidence bounds of the 1M € discount curve in fig. 8. Applied on € FRAs with shorter maturity constant $\lambda = [0.91, 0.94, 0.97]$ still performed well, although only the model with $\lambda = 0.94$ performed well for the FRA with a maturity of two months.

As expected historical simulation shows its inferiority in comparison to FHS as models applying $\lambda = 1$ fail the Kupiec test and most of the other test metrics. This result for HS is in strong contrast to reactive FHS models.

3.6.2 Backtests with calibrated volatility models

In the previous section the volatilities were estimated with EWMA models using arbitrarily chosen constant λ parameters. This section on the other hand will make use of the results of sections 3.3 to 3.5 and apply a range of EWMA calibration approaches during backtesting. Those models are

- M1: Calibration of three different λ per discount curve, the position of the junctions between different λ is optimized individually for each discount curve. Overall this results in 69 different models.

Table 8: Test statistics of backtests of € products. *UC* stands for unconditional coverage, *Ind* for independence and *CC* for conditional coverage. Tests are described in section 2.6.2, section 2.6.3 and appendix A. Shown is the value of the individual test statistic. Input for the tests was the Hit sequence under a given volatility model. Tests for which the null hypothesis had to be denied with a 95% confidence level are marked red. A denial of the null hypothesis indicates bad backtesting performance.

			Kupiec	Christtofersen		Engle / Manganelli - CAViaR									Christoffersen & Pelletier	
Row	Product	Lambda	UC	Ind	CC	DQ UC Lag 1	DQ UC Lag 2	DQ UC Lag 3	DQ Ind Lag 1	DQ Ind Lag 2	DQ Ind Lag 3	DQ CC Lag 1	DQ CC Lag 2	DQ CC Lag 3	Duration CC	Duration Ind
1		M 1	53,4%	63,8%	73,8%	54,2%	54,0%	57,0%	75,8%	66,2%	27,5%	79,4%	90,7%	15,1%	8,8%	**4,4%**
2		M 2	53,4%	63,8%	73,8%	54,2%	54,0%	57,0%	75,8%	66,2%	27,5%	79,4%	90,7%	15,1%	8,8%	**4,4%**
3		M 3	72,3%	61,2%	82,6%	72,5%	72,4%	74,3%	73,1%	62,4%	33,2%	88,7%	94,8%	20,9%	10,8%	**4,4%**
4	€ 1M/10Y	0,91	37,0%	66,4%	60,9%	38,5%	38,3%	38,1%	78,5%	69,8%	63,3%	66,5%	82,8%	91,5%	38,3%	46,0%
5		0,94	23,9%	69,0%	46,2%	26,0%	25,8%	25,6%	81,1%	73,4%	67,6%	52,0%	71,5%	84,2%	28,5%	54,2%
6		0,97	92,6%	58,6%	85,9%	92,6%	92,7%	93,3%	70,2%	58,7%	39,4%	92,6%	95,9%	26,5%	17,3%	6,6%
7		0,99	72,3%	11,5%	27,1%	74,4%	74,2%	57,5%	**1,9%**	16,8%	**1,1%**	5,9%	11,6%	**1,6%**	**1,5%**	**0,5%**
8		0,995	72,3%	11,5%	27,1%	74,4%	74,2%	57,5%	**1,9%**	16,8%	**1,1%**	5,9%	11,6%	**1,6%**	**1,5%**	**0,5%**
9		M 1	87,1%	56,2%	83,4%	86,8%	87,3%	87,0%	67,4%	23,1%	43,5%	90,3%	20,2%	31,3%	14,8%	5,1%
10		M 2	72,3%	61,2%	82,6%	72,5%	74,3%	74,3%	73,1%	15,3%	30,8%	88,7%	12,4%	21,0%	7,7%	**2,9%**
11		M 3	72,3%	61,2%	82,6%	72,5%	74,3%	74,3%	73,1%	15,3%	30,8%	88,7%	12,4%	21,0%	7,7%	**2,9%**
12	€ 1M/2Y	0,91	53,4%	63,8%	73,8%	54,2%	57,1%	57,0%	75,8%	12,1%	25,3%	79,4%	8,4%	15,1%	14,7%	8,1%
13		0,94	53,4%	63,8%	73,8%	54,2%	57,1%	57,0%	75,8%	12,1%	25,3%	79,4%	8,4%	15,1%	5,6%	**2,6%**
14		0,97	87,1%	56,2%	83,4%	86,8%	87,3%	87,0%	67,4%	23,1%	43,5%	90,3%	20,2%	31,3%	38,3%	16,7%
15		0,99	**3,8%**	77,2%	11,1%	5,8%	5,8%	5,7%	88,2%	83,3%	79,6%	16,7%	30,8%	46,1%	**0,3%**	**1,5%**
16		0,995	**3,8%**	77,2%	11,1%	5,8%	9,7%	9,7%	88,2%	**2,1%**	5,8%	16,7%	**0,2%**	**0,4%**	**0,2%**	**0,9%**
17		M 1	81,7%	10,9%	26,9%	83,1%	83,0%	83,1%	**1,4%**	14,8%	30,4%	**4,7%**	9,5%	16,9%	18,8%	7,9%
18		M 2	96,8%	12,9%	31,6%	97,0%	97,0%	96,7%	**2,0%**	**0,1%**	**2,1%**	6,8%	**1,8%**	**3,2%**	**4,1%**	**1,2%**
19		M 3	96,8%	12,9%	31,6%	97,0%	97,0%	96,7%	**2,0%**	**0,1%**	**2,1%**	6,8%	**1,8%**	**3,2%**	**4,1%**	**1,2%**
20	€ 1M/8M	0,91	75,9%	15,2%	34,1%	77,0%	76,4%	79,1%	**2,9%**	22,6%	**0,0%**	8,8%	16,0%	**0,0%**	**2,6%**	**0,7%**
21		0,94	96,8%	12,9%	31,6%	97,0%	96,7%	96,5%	**2,0%**	18,4%	36,6%	6,8%	13,0%	21,8%	8,4%	**2,7%**
22		0,97	81,7%	62,0%	86,1%	81,7%	84,5%	84,6%	73,3%	**0,0%**	**0,4%**	91,9%	**0,0%**	**0,0%**	13,3%	5,2%
23		0,99	8,6%	76,0%	21,9%	11,3%	16,4%	16,3%	86,9%	**2,2%**	6,1%	28,3%	**0,3%**	**0,7%**	**0,3%**	**0,5%**
24		0,995	**0,2%**	87,9%	**0,7%**	**0,8%**	**4,9%**	5,0%	95,4%	**0,0%**	**0,2%**	**2,9%**	**0,0%**	**0,0%**	**0,0%**	**0,6%**
25		M 1	81,7%	62,0%	86,1%	81,7%	81,7%	81,7%	73,3%	62,7%	55,0%	91,9%	96,2%	98,2%	67,3%	46,5%
26		M 2	57,1%	54,1%	70,6%	55,7%	57,6%	57,0%	64,0%	25,7%	48,3%	75,8%	18,9%	29,1%	28,8%	12,2%
27		M 3	57,1%	54,1%	70,6%	55,7%	57,6%	57,0%	64,0%	25,7%	48,3%	75,8%	18,9%	29,1%	28,8%	12,2%
28	€ 1M/5M	0,91	41,1%	51,6%	57,8%	38,9%	38,2%	37,4%	60,9%	46,6%	36,9%	61,1%	73,9%	81,9%	39,9%	21,8%
29		0,94	75,9%	56,7%	81,0%	75,4%	76,5%	76,1%	67,2%	21,1%	40,8%	87,1%	16,9%	27,0%	19,0%	6,8%
30		0,97	81,7%	62,0%	86,1%	81,7%	83,1%	83,2%	73,3%	13,3%	27,7%	91,9%	10,2%	17,9%	23,9%	10,7%
31		0,99	**4,1%**	79,0%	11,9%	6,4%	11,1%	11,1%	89,3%	**1,2%**	**3,7%**	17,9%	**0,1%**	**0,2%**	**1,6%**	10,4%
32		0,995	**1,7%**	81,9%	5,6%	**3,4%**	15,7%	15,8%	91,5%	**0,0%**	**0,0%**	10,5%	**0,0%**	**0,0%**	**0,0%**	**0,3%**
33		M 1	11,9%	44,4%	22,2%	9,2%	8,7%	8,2%	51,3%	35,1%	25,0%	20,3%	30,3%	39,2%	40,4%	75,6%
34		M 2	11,9%	44,4%	22,2%	9,2%	10,4%	9,9%	51,3%	49,6%	82,9%	20,3%	12,4%	18,9%	40,9%	78,4%
35		M 3	18,8%	46,7%	32,3%	15,9%	17,5%	16,9%	54,5%	42,9%	73,7%	31,7%	15,6%	23,6%	42,9%	45,6%
36	€ 1M/2M	0,91	**1,3%**	35,8%	**3,0%**	**0,5%**	**0,5%**	**0,4%**	38,9%	21,9%	12,9%	**1,6%**	**2,9%**	**4,3%**	7,8%	77,8%
37		0,94	18,8%	46,7%	32,3%	15,9%	15,2%	14,6%	54,5%	38,8%	28,7%	31,7%	44,2%	54,4%	54,6%	78,5%
38		0,97	**0,7%**	47,1%	**2,0%**	**0,3%**	**0,4%**	**0,3%**	26,6%	12,0%	46,0%	**0,6%**	**0,9%**	**1,4%**	**1,9%**	17,1%
39		0,99	7,3%	**3,9%**	**2,4%**	7,6%	10,1%	9,1%	**0,0%**	**0,0%**	**0,0%**	**0,0%**	**0,0%**	**0,0%**	**0,0%**	**0,0%**
40		0,995	7,3%	**3,9%**	**2,4%**	7,6%	10,1%	9,1%	**0,0%**	**0,0%**	**0,0%**	**0,0%**	**0,0%**	**0,0%**	**0,0%**	**0,0%**
41	€ 1M/10Y	1 (HS)	51,0%	**1,6%**	**4,4%**	54,3%	53,2%	73,4%	**0,0%**	**1,1%**	**0,1%**	**0,0%**	**0,0%**	**0,0%**	**0,0%**	**0,0%**
42	€ 1M/2Y	1 (HS)	**0,2%**	85,7%	**0,8%**	**0,8%**	**3,5%**	**3,6%**	94,2%	**0,2%**	**0,8%**	**2,9%**	**0,0%**	**0,0%**	**0,0%**	**3,7%**
43	€ 1M/8M	1 (HS)	**0,0%**	90,9%	**0,2%**	**0,4%**	5,6%	5,8%	97,0%	**0,0%**	**0,0%**	**1,4%**	**0,0%**	**0,0%**	**0,0%**	**3,4%**
44	€ 1M/5M	1 (HS)	**0,0%**	90,9%	**0,2%**	**0,4%**	5,6%	5,8%	97,0%	**0,0%**	**0,0%**	**1,4%**	**0,0%**	**0,0%**	**0,0%**	5,5%
45	€ 1M/2M	1 (HS)	**1,7%**	81,9%	5,6%	**3,4%**	7,7%	7,7%	91,5%	**0,6%**	**2,0%**	10,5%	**0,0%**	**0,0%**	**0,1%**	**0,8%**

- M2: Calibration of one λ per currency, which, as shown in fig. 8 should be sufficient based on confidence bounds. This approach results in 5 different models.
- M3: Calibration of one λ for the risk factors on all discount curves, which results in only one model at any given time.

While the risk factors of a FRA are always spread across two different discount curves, the one on the OIS curve has relatively little impact on the VaR estimation as is shown in appendix B.2. For this reason we will exclusively focus on the two risk factors on the discount curve with a reset frequency of $t_2 - t_1$. Products based on three discount curves across three different currencies were chosen.

First the one month € discount curve, which appeared especially well behaved in fig. 8. Individual calibrations of this curve are close to the overall λ of the € currency and the result across all risk factors. Additionally, calibrations of this discount curve showed relatively narrow confidence bounds. For products based on this discount curve one would expect very similar results between the three different volatility scenarios M1 to M3, if $t_2 \geq 2$ years for the FRA in question.

As the second discount curve the CHF one month curve was chosen, since - in direct contrast to the first choice - it shows extremely broad confidence bounds, which rarely intersect with the calibrated overall and currency specific λ. Additionally the curve seems to change its character after the intervention of the swiss central bank in September 2011.

To cover yet another currency and also extend backtesting to another reset frequency the \$ three month discount curve is chosen as the third curve. Examining fig. 8 again, it can be seen, that confidence bounds of this curve only seldom intersect with the confidence bounds of the one month CHF discount curve. Additionally, calibrations of this curve has an opposing trend with calibrations of the one month CHF curve being high in early years and low in later years. Furthermore, products based on the one month CHF, which will be backtested, are focused on the short end of the curve, while those based on the three month \$ curve focus on the long end. This should make it as hard as possible to find a common model working equally well for products based on the one month CHF curve and on those based on the three month \$ curve. One would therefore expect the results between M1, M2 and M3 to be quite different for these two curves. If it would be possible to achieve good results with M3 for products based on the one month CHF curve as well as such based on the three month \$ curve this would suggest, that one common EWMA calibration could be sufficient to appropriately estimate the VaR of any FRA portfolio.

Results of the backtests using these three volatility scenarios are displayed in table 8 and 9. Volatility scenario M1 is not used upon the \$ products, since they all have

Table 9: Teststatistics of backtests of $ and CHF products. *UC* stands for unconditional coverage, *Ind* for independence and *CC* for conditional coverage. Tests are described in section 2.6.2, section 2.6.3 and appendix A.

		Kupiec	Christtofersen		Engle / Manganelli - CAViaR									Christoffersen & Pelletier	
Product	Lambda	UC	Ind	CC	DQ UC Lag 1	DQ UC Lag 2	DQ UC Lag 3	DQ Ind Lag 1	DQ Ind Lag 2	DQ Ind Lag 3	DQ CC Lag 1	DQ CC Lag 2	DQ CC Lag 3	Duration CC	Duration Ind
CHF 1M/10Y	M 1	72,3%	11,5%	27,1%	74,4%	74,2%	76,0%	**1,9%**	16,8%	**1,2%**	5,9%	11,6%	**2,0%**	71,0%	58,2%
	M 2	72,3%	11,5%	27,1%	74,4%	74,2%	76,0%	**1,9%**	16,8%	**1,2%**	5,9%	11,6%	**2,0%**	71,0%	58,2%
	M 3	72,3%	11,5%	27,1%	74,4%	74,2%	76,0%	**1,9%**	16,8%	**1,2%**	5,9%	11,6%	**2,0%**	71,0%	58,2%
	0,94	72,3%	11,5%	27,1%	74,4%	74,2%	76,0%	**1,9%**	16,8%	**1,2%**	5,9%	11,6%	**2,0%**	71,0%	58,2%
CHF 1M/2Y	M 1	37,0%	66,4%	60,9%	38,5%	38,3%	42,2%	78,5%	69,8%	22,3%	66,5%	82,8%	9,8%	49,8%	89,4%
	M 2	72,3%	61,2%	82,6%	72,5%	72,4%	74,3%	73,1%	62,4%	33,2%	88,7%	94,8%	20,9%	56,0%	37,8%
	M 3	92,6%	58,6%	85,9%	92,6%	92,7%	93,3%	70,2%	58,7%	39,4%	92,6%	95,9%	26,5%	49,1%	25,5%
	0,94	53,4%	63,8%	73,8%	54,2%	54,0%	57,0%	75,8%	66,2%	27,5%	79,4%	90,7%	15,1%	61,5%	67,5%
CHF 1M/1Y	M 1	92,6%	58,6%	85,9%	92,6%	92,7%	92,8%	70,2%	58,7%	50,3%	92,6%	95,9%	97,8%	87,6%	70,7%
	M 2	53,4%	63,8%	73,8%	54,2%	54,0%	53,8%	75,8%	66,2%	59,0%	79,4%	90,7%	95,7%	66,7%	90,8%
	M 3	37,0%	66,4%	60,9%	38,5%	38,3%	38,1%	78,5%	69,8%	63,3%	66,5%	82,8%	91,5%	50,3%	98,6%
	0,94	53,4%	63,8%	73,8%	54,2%	54,0%	53,8%	75,8%	66,2%	59,0%	79,4%	90,7%	95,7%	66,3%	87,2%
CHF 1M/8M	M 1	53,4%	63,8%	73,8%	54,2%	54,0%	53,8%	75,8%	66,2%	59,0%	79,4%	90,7%	95,7%	**0,3%**	**0,1%**
	M 2	53,4%	63,8%	73,8%	54,2%	54,0%	53,8%	75,8%	66,2%	59,0%	79,4%	90,7%	95,7%	11,7%	6,2%
	M 3	72,3%	61,2%	82,6%	72,5%	72,4%	72,4%	73,1%	62,4%	54,7%	88,7%	94,8%	97,5%	15,1%	6,6%
	0,94	37,0%	66,4%	60,9%	38,5%	38,3%	38,1%	78,5%	69,8%	63,3%	66,5%	82,8%	91,5%	25,7%	24,6%
CHF 1M/2M	M 1	**0,1%**	13,6%	**0,1%**	**0,0%**	**0,1%**	**0,1%**	**0,9%**	**0,0%**	**2,2%**	**0,0%**	**0,0%**	**0,0%**	**0,7%**	51,4%
	M 2	17,1%	29,0%	22,4%	16,3%	15,6%	17,0%	11,5%	51,4%	14,2%	10,2%	16,8%	10,4%	25,6%	22,9%
	M 3	17,1%	29,0%	22,4%	16,3%	15,6%	17,0%	11,5%	51,4%	14,2%	10,2%	16,8%	10,4%	25,6%	22,9%
	0,94	**2,3%**	42,3%	5,6%	**1,6%**	**1,4%**	**1,7%**	22,9%	79,3%	35,1%	**2,3%**	**4,0%**	**4,3%**	6,8%	26,5%
$ 3M/15Y	M 2	51,0%	51,4%	65,0%	49,3%	48,6%	50,4%	61,5%	47,4%	61,0%	70,1%	80,8%	34,6%	78,4%	57,4%
	M 3	87,1%	56,2%	83,4%	86,8%	86,5%	87,0%	67,4%	54,9%	46,2%	90,3%	94,3%	31,1%	79,2%	49,8%
$ 3M/10Y	M 2	87,1%	15,7%	36,2%	87,6%	87,3%	87,0%	**3,5%**	24,6%	46,0%	10,8%	19,2%	30,0%	53,3%	26,4%
	M 3	51,0%	20,5%	36,1%	51,8%	51,0%	50,4%	6,0%	34,1%	60,4%	13,6%	22,7%	33,7%	26,6%	11,5%
$ 3M/7Y	M 2	25,6%	46,8%	40,2%	22,8%	22,1%	21,4%	55,6%	40,1%	30,0%	41,4%	54,6%	64,4%	65,2%	81,5%
	M 3	11,0%	42,4%	20,2%	8,4%	9,4%	8,9%	49,6%	57,0%	91,7%	18,6%	13,7%	20,3%	29,4%	44,1%
$ 3M/5Y	M 2	25,6%	26,0%	27,8%	25,2%	24,4%	28,8%	9,4%	45,2%	**0,4%**	12,2%	20,0%	**0,0%**	48,4%	42,1%
	M 3	25,6%	26,0%	27,8%	25,2%	24,4%	26,2%	9,4%	45,2%	10,9%	12,2%	20,0%	10,7%	53,6%	50,4%
$ 3M/30M	M 2	51,0%	51,4%	65,0%	49,3%	48,6%	50,4%	61,5%	47,4%	61,0%	70,1%	80,8%	34,6%	80,6%	60,9%
	M 3	25,6%	46,8%	40,2%	22,8%	22,1%	21,4%	55,6%	40,1%	30,0%	41,4%	54,6%	64,4%	32,0%	22,4%

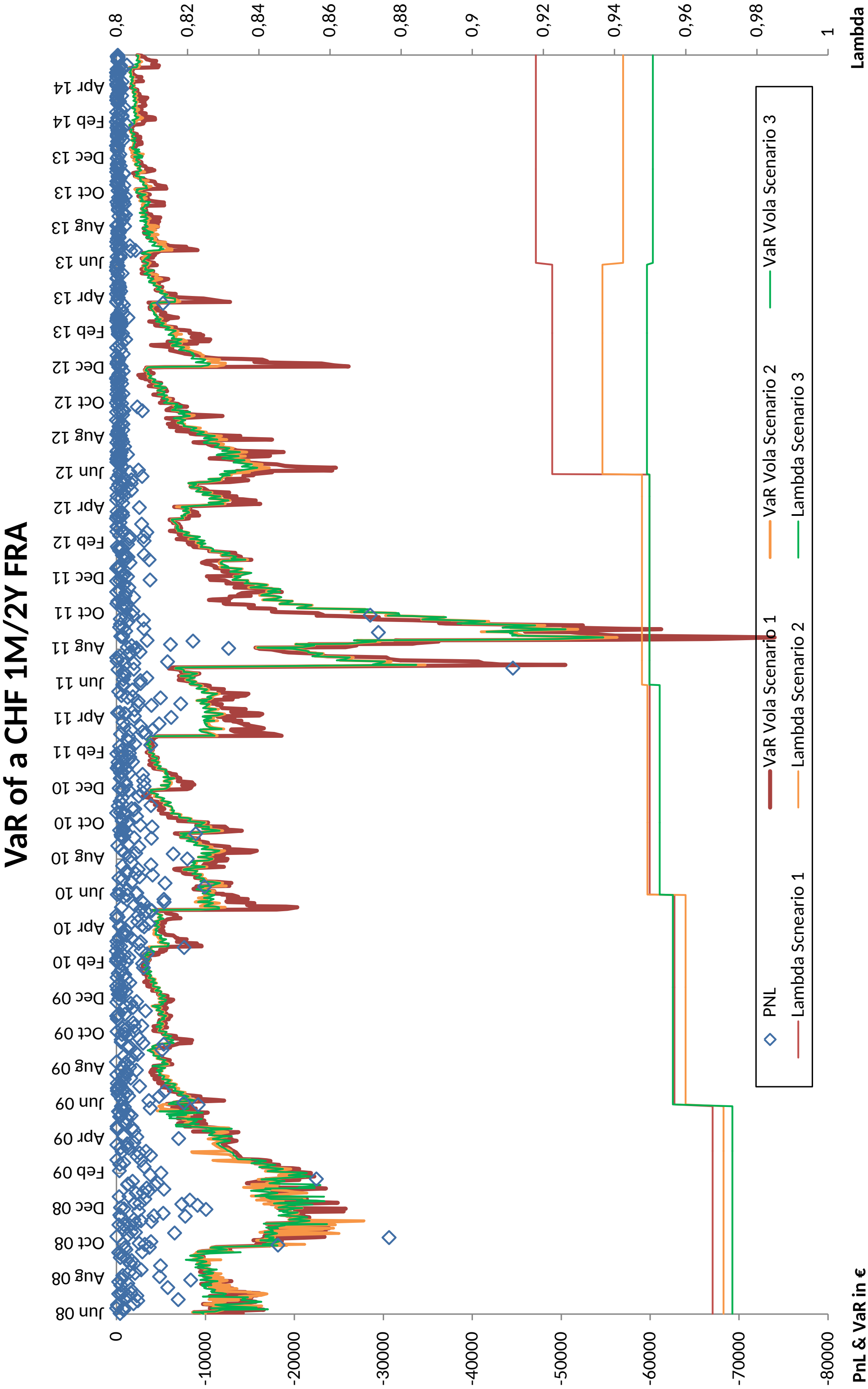


Figure 10: The λ used for the one month € discount curve for the respective volatility scenario is displayed as a horizontal jumping line at relates to the secondary axis. The utilized λ jumps once a year, since models are recalibrated at the first of July every year.

a relatively high maturity and therefore none of them should be impacted by the segmentation of the front end accomplished by scenario M1. Results are similar across the three volatility scenarios with all of them achieving very good backtesting results. The similarity of the results can oftentimes simply be explained by the calibrations of very similar λ using any of the three scenarios. Such an example is displayed in fig. 10. The calibrated λ are displayed in the figure and remain very similar to each other until the recalibration in 2012. Consequently, estimated VaR are very similar as well, which explains the similar performance of all models in table 9. In addition fig. 10 is also a very good example to show the power of FHS. As the central bank intervention suddenly hit the market in 2011 estimated VaR jumped massively to adjust for the shift in the market. Despite this unprecedented incident, FHS - which estimates VaR based purely on historical returns - was able to adjust well thanks to its volatility filtering and all volatility scenarios performed well under all backtesting performance metrics. Unfortunately the removal of the intervention in January 2015 was not part of the available dataset.

Another reason for the similar performance of all volatility scenarios is the fact that all performance metrics used in this thesis are based exclusively on the hit function as presented in eq. (14) on page 29, which can be the same even for somewhat different time series of VaR. An extreme example of this is the CHF 1M/10Y FRA for which all volatility models resulted in exactly the same hit function, even though λ estimated were relatively different with a difference of up to 0.043 between the λ under scenario M3 and the one under scenario M1 in 2013.[85]

As the only tested product the CHF 1M/2M FRA has shown considerable performance differences between the different volatility scenarios. The backtesting chart is displayed in fig. 11 on the following page and shows, that the relevant λ of M1 is extremely low. For this reason the VaR prediction based on scenario M1 is too reactive and drops too fast and frequently underestimates the volatility in the market, resulting in too many exceptions.

Overall four lessons can be taken away from backtesting with the different backtesting scenarios. First and foremost all volatility scenarios produce well performing VaR estimations. Secondly the exclusion of low tenor risk factors during calibration does not result in bad backtesting results for products based on the short end - in contrast, scenario M1 even performed worse than M2 and M3 for one product based on the short end of a discount curve. Thirdly, within the scope of the backtested products it is completely acceptable with regards to backtesting performance to use a common decay factor for all curves instead of an individual one for each currency. The

[85] The PnL and VaR time series are not shown.

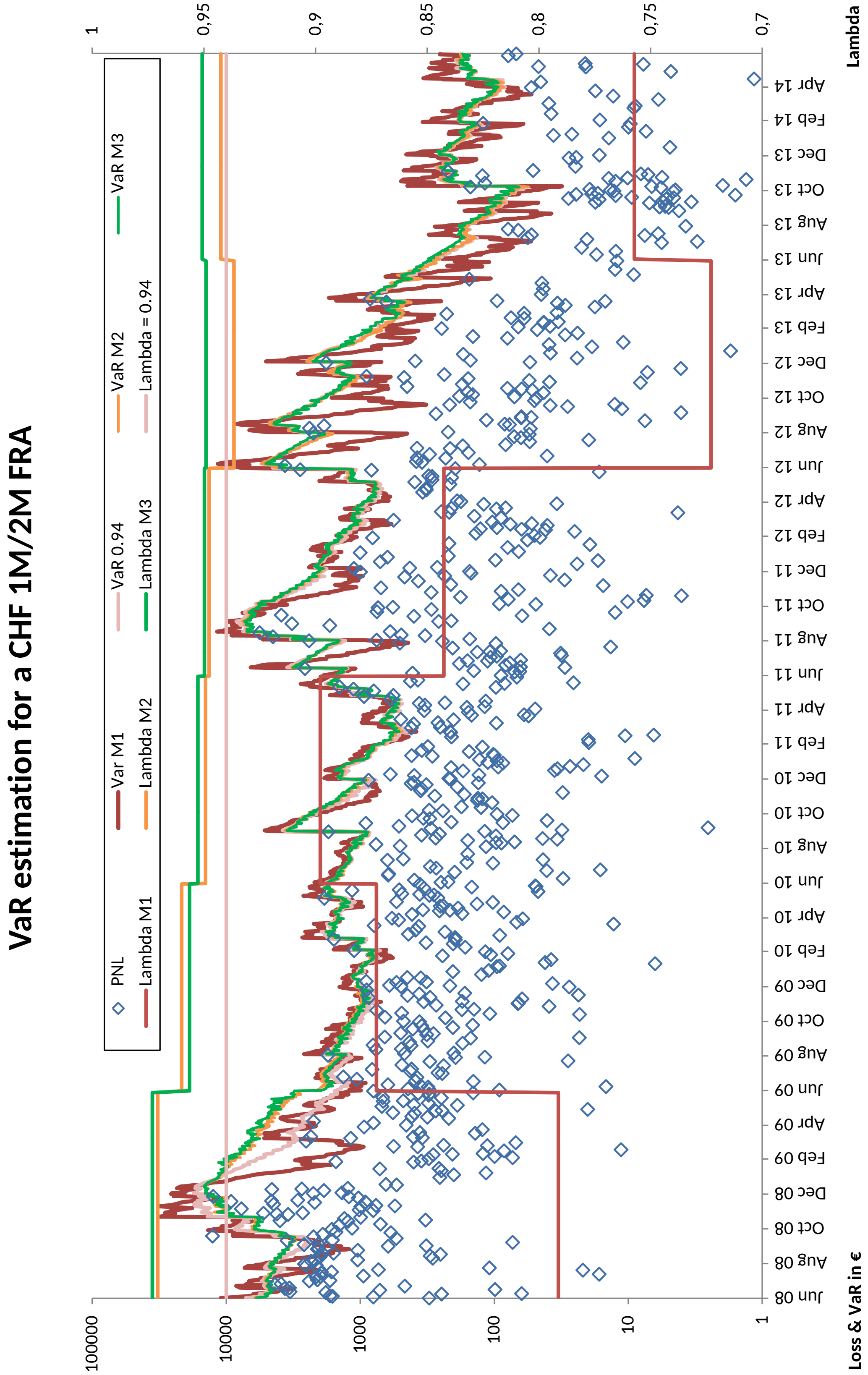


Figure 11: For this product PnLs were extremely fat tailed, requiring a logarithmic scale. For this reason losses have to be depicted as positive values. Note that the relevant λ estimated under scenario M1 is extremely low, dropping down to only 0.72 in July 2012.

backtested products were chosen to cover the extremes and therefore this result should be reproducible for other FRAs and their portfolios. The fourth result is, that not calibrating anything and just assuming an EWMA process with a constant $\lambda = 0.94$ as it was proposed by Morgan (1996)[86] almost 20 years ago also performs really well for the databasis of this thesis.

[86]For EWMA processes based on daily returns. Compare Morgan (1996)[Section 5.3.2]

4 Conclusion

The goal of this thesis was to gain insight into the role of volatility modeling during value at risk estimation with filtered historical simulation of a portfolio and how the volatility model should be chosen. The empirical study was conducted on forward rate agreements.

The thesis has shown yet again, that adjusting for volatility clustering is a very effective way to improve the estimation of VaR based on historic returns. While existing literature focuses on the calibration of a volatility model based on a single return time series, a process for calibrating a single volatility model based on multiple return time series has been developed. This process is based on the addition of likelihood values and therefore allows to compare different models with each other using the Bayesian information criterion.

Comparisons have shown, that a simple exponentially weighted moving average model is sufficient to estimate volatilities of movements of tenor points on discount curves. Additionally it has been shown, that using different volatility models for tenor points of the same discount curve can result in unwanted side effects and might compromise an approximate hedge. For this reason using a common EWMA volatility model for all tenor points on a single discount curve is a good approach.

Additionally backtests of a range of forward rate agreements have shown, that VaR performance using filtered historical simulation with an EWMA model is very insensitive towards the chosen parametrization of the EWMA model. A wide range of parametrizations of the decay factor λ produced good results over a six year time period. The most common backtesting performance metrics - the two sided Kupiec test and the Christoffersen test for conditional coverage - were very rarely by any λ between 0.91 and 0.97. Even more advanced backtesting performance metrics have shown to be insensitive towards the calibration of the volatility model. Models calibrated with the developed process performed well, but did not outperform alternative approaches. In fact, the calibration almost appeared pointless with regard to backtesting, as one might just as well use the decay factor $\lambda = 0.94$ proposed for EWMA processes by RiskMetrics in 1996 for all risk factors and yield the same good backtesting results as one would get by properly calibrating λ.

A Backtesting Performance

A.1 Lag 1 based independence and conditional coverage

To create his test statistic Christoffersen (1998) compares the likelihood of the most probable transition matrix with the supposed transition probabilities[87]. Assuming a working VaR(α) model the probability of the appearance of an exception the next day should always be α and respectively $1-\alpha$ for the opposite. An observed Hit function $\hat{I}$ as defined in eq. (14) can be interpreted as a Markov chain and based on table 2 on page 30 the observed transition probability matrix $\hat{\Pi}$ can be calculated as

$$\begin{aligned}
\hat{\pi}_{01} &= \frac{\hat{n}_{01}}{\hat{n}_{00}+\hat{n}_{01}} \\
\hat{\pi}_{11} &= \frac{\hat{n}_{11}}{\hat{n}_{10}+\hat{n}_{11}} \\
\hat{\pi}_{00} &= 1-\hat{\pi}_{01} \\
\hat{\pi}_{10} &= 1-\hat{\pi}_{11} \\
\hat{\Pi} &= \begin{bmatrix} \hat{\pi}_{00} & \hat{\pi}_{01} \\ \hat{\pi}_{10} & \hat{\pi}_{11} \end{bmatrix}.
\end{aligned}$$

The likelihood of the realization of any observed Hit function I yielding n_{00}, n_{01}, n_{10} and n_{11} under an arbitrary transition probability matrix Π is

$$L\left(\Pi \mid I\right) = {\pi_{00}}^{n_{00}} * {\pi_{01}}^{n_{01}} * {\pi_{10}}^{n_{10}} * {\pi_{11}}^{n_{11}}$$

and $\hat{\Pi}$ maximizes this likelihood function for $\hat{I}$.

Under the assumption of independence it should be true that $\pi_{01} = \pi_{11}$. Therefore, by replacing $\hat{\pi}_{01}$ and $\hat{\pi}_{11}$ with a variable π_2 and optimizing π_2 such that

$$\begin{aligned}
&\underset{\pi_2}{\text{maximize}} && L\left(\Pi_2 \mid \hat{I}\right) = (1-\pi_2)^{\hat{n}_{00}} * {\pi_2}^{\hat{n}_{01}} * (1-\pi_2)^{\hat{n}_{10}} * {\pi_2}^{\hat{n}_{11}} \\
&\text{where} && \Pi_2 = \begin{bmatrix} (1-\pi_2) & \pi_2 \\ (1-\pi_2) & \pi_2 \end{bmatrix}
\end{aligned}$$

the best possible transition matrix Π_2 under the assumption of independence can be found.

Since the occurence of an exception is the transition $0 \to 1$ or $1 \to 1$ of the Hit function and the probability of the occurence of an exception in a VaR(α) model should

[87] A good introduction to likelihood and likelihood ratio tests can be found in chapter 14 of Greene (2008). Alternatively it is also briefly covered in section 2.4.

always be α the implied transition matrix of an VaR(α) model is

$$\hat{\Pi}_\alpha = \begin{bmatrix} (1-\alpha) & \alpha \\ (1-\alpha) & \alpha \end{bmatrix}.$$

It follows that

$$L\left(\hat{\Pi} \mid \hat{I}\right) \geq L\left(\Pi_2 \mid \hat{I}\right) \geq L\left(\Pi_\alpha \mid \hat{I}\right)$$

and based on standard literature regarding likelihood tests of different parametrizations of binary markov chains[88] the test statistics are

$$LR_{ind} = -2log\left[\frac{L\left(\Pi_2 \mid \hat{I}\right)}{L\left(\hat{\Pi} \mid \hat{I}\right)}\right] \overset{asy}{\sim} X^2(1) \tag{30}$$

and

$$LR_{cc} = -2log\left[\frac{L\left(\Pi_\alpha \mid \hat{I}\right)}{L\left(\hat{\Pi} \mid \hat{I}\right)}\right] \overset{asy}{\sim} X^2(1). \tag{31}$$

A.2 Duration based independence and conditional coverage

Duration based tests for independence and conditional coverage where proposed by Christoffersen and Pelletier (2004). While the tests introduced in appendix A.1 on page 77 tested only for Lag 1 independence Christoffersen and Pelletier used the duration between two exceptions in order not to make the assumption that clustering of exceptions is limited up to a certain Lag. The downside of this is that the authors did have to make distributional assumptions regarding the distance of two exceptions, while the test for Lag 1 independence was free of distributional assumptions.

If exceptions occur at time t_i and t_{i+1} without any exception in between then let $D_i = t_{i+1} - t_i$ be the duration between the two exceptions. D is the set of all D_i produced in backtesting a VaR model. Under the assumption of independence the probability for an exception should be equal for each new day. It follows for the a D_i,

[88] Compare Hoel (1954)

that

$$\begin{aligned} Pr(D_i = 1) &= p \\ Pr(D_i = 2) &= (1-p)p \\ Pr(D_i = 3) &= (1-p)^2 p \\ &\cdots \\ Pr(D_i = d) &= (1-p)^{d-1} p \end{aligned}$$

$$\text{and} \qquad Pr(D_i = d \mid D_i \notin \{1, \cdots, (d-1)\}) = p \tag{32}$$

if the VaR model exhibits independence.

The only continuous distribution that has the no memory property of eq. (32) is the exponential. Therefore the null hypothesis is, that all D_i are distributed as

$$f_{exp}(D_i; p) = pe^{-pD_i}. \tag{33}$$

Since the exponential distribution is a special case of the Weibull distribution

$$\begin{aligned} f_W(D_i; a, b) &= a^b b D_i^{b-1} e^{-(aD_i)^b)} \\ f_W(D_i; p, 1) &= f_{exp}(D_i; p) \end{aligned}$$

the Weibull distribution with $b \neq 1$ can be assumed as the counter hypothesis[89]. When calibrating the Weibull distribution based on a VaR model with exception clustering it would result in a parametrization with $b < 1$ representing the unexpected large number of low and high D_i. Low D_i would occur in times of stress and high D_i would occur in times of low volatility. A proper VaR model exhibiting independence on the other hand should neither yield many low D_i nor many high D_i and therefore have a b close to one when calibrating a Weibull distribution based on D.

Similar to appendix A.1 on page 77 independence can be tested for by comparing the likelihood of realizing D with the best possible Weibull distribution with the likelihood of realizing D with the best possible exponential distribution. Conditional coverage of a VaR(α) model can be tested for by comparing the likelihood of realizing D with the best possible Weibull distribution with the likelihood of realizing D with an exponential distribution with parameter α.

Assuming that $[a^*,\ b^*]$ is the MLE parametrization for a Weibull distribution real-

[89] Alternatively Christoffersen and Pelletier also proposed a Gamma distribution of which the exponential distribution is also a special case as a counter hypothesis.

izing D and p^* is the MLE parametrization for an exponential distribution realizing D it follows that

$$L_{WB}\left(\,[a^*,\ b^*] \mid D\right) \geq L_{WB}\left(\,[p^*,\ 1] \mid D\right) \geq L_{WB}\left(\,[\alpha,\ 1] \mid D\right)$$

$$LR_{ind} = -2log\left[\frac{L_{WB}\left(\,[p^*,\ 1] \mid D\right)}{L_{WB}\left(\,[a^*,\ b^*] \mid D\right)}\right] \overset{\text{asy}}{\sim} X^2\left(1\right) \tag{34}$$

$$\text{and} \quad LR_{cc} = -2log\left[\frac{L_{WB}\left(\,[\alpha,\ 1] \mid D\right)}{L_{WB}\left(\,[a^*,\ b^*] \mid D\right)}\right] \overset{\text{asy}}{\sim} X^2\left(1\right). \tag{35}$$

where L_{WB} is the likelihood function of a Weibull distribution.

A.3 Autoregression based independence and conditional coverage

Engle and Manganelli (2004) proposed in their paper an alternative way of calculating and testing VaR. Their test methodology rests on the idea of identifying autoregression between hits and past hits or previous VaRs. They test if it is possible to predict a hit I_t with

$$I_t = \theta + \sum_{k=1}^{n} \beta_{1k} I_{t-k} + \sum_{k=1}^{n} VaR_{t-k+1} + u_t. \tag{36}$$

At time $t-1$ the information known are the VaR estimation for t and all previous VaR estimates as well as all hits up to I_{t-1}. Using a proper VaR model this information should be of no value for predicting more accurately if the I_t is more likely to be 0 or 1. Assuming the error term u follows a logistic distribution, since the target variable is binary, none of the β coefficients should be statistically significant. Just like in appendix A.1 and appendix A.2 a likelihood ratio test with the maximum likelihood estimated β and all $\beta = 0$ should lead to the conclusion that the assumption of $\beta = 0$ is not significantly worse. To test not only for independence but for conditional coverage of a VaR(α) model extend the assumption of $\beta = 0$ by the assumption

$$\frac{e^{\theta}}{(1+e^{\theta})} = \alpha. \tag{37}$$

B Empirical Appendix

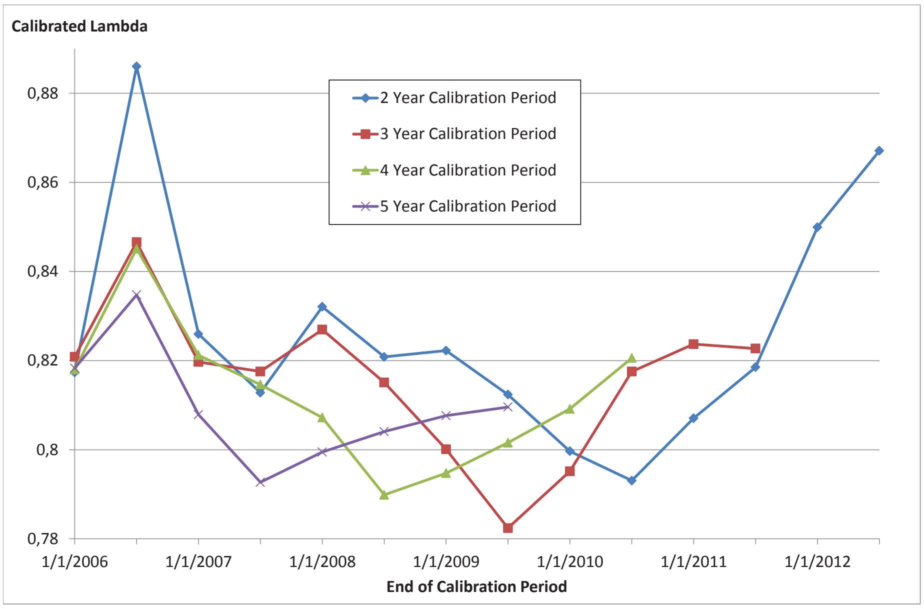
B.1 Choosing the length of the calibration window

All calibrations in the empirical section have been done based on a three year calibration period. It is tough to evaluate scientifically, weather the length of a calibration period is particularly good or bad but a choice has to be made nevertheless. As the length of the calibration window increases the calibration results would be expected to get more stable, since the chances to have 'bad luck' with the sample decreases. On the other hand long samples would decrease the period of time for which a backtest could be done, since no backtesting can be done for the first initial calibration period.

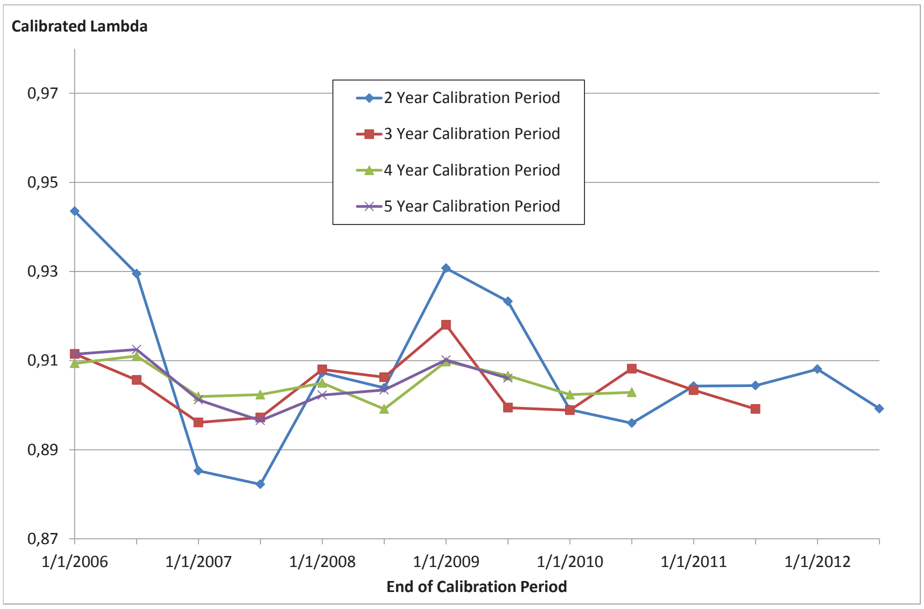
Using different calibration period lengths EWMA model were calibrated oriented on the optimal division into risk factor subsets identified in section 3.3 on page 48. The results are displayed from fig. 12a on the following page to fig. 12d on page 83.

The figures show, that λ calibrated based on a three year calibration period show similar variance as λ calibrated based on longer calibration periods. Calibrations based on the two year period on the other side tend to jump around a lot more and have significantly higher variance than calibrations based on a three year period. Calibrations based on a one year history have been made but were so unstable that they couldn't be properly displayed in the λ range of the figures.

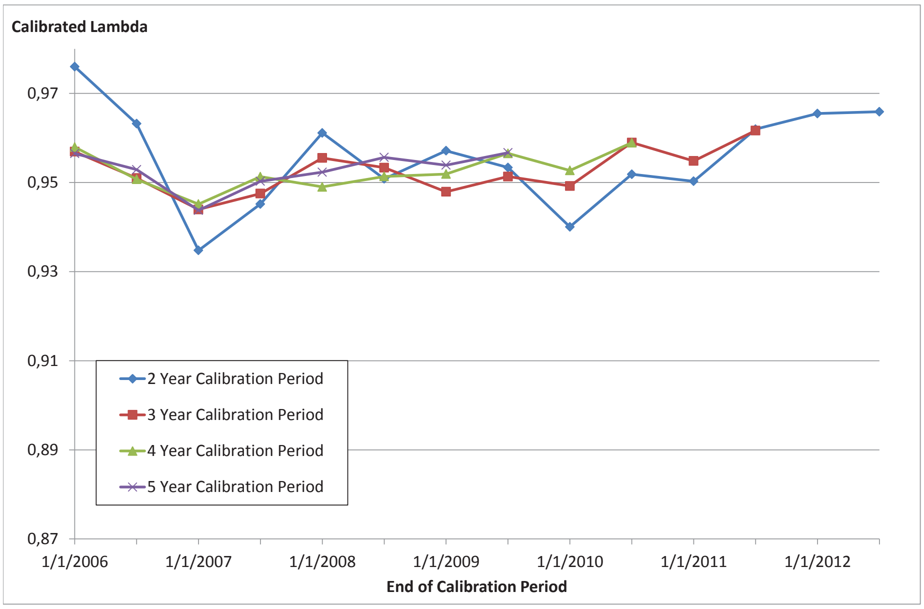
Figure 12: Stability of Lambda over time for calibration periods of differing length and position



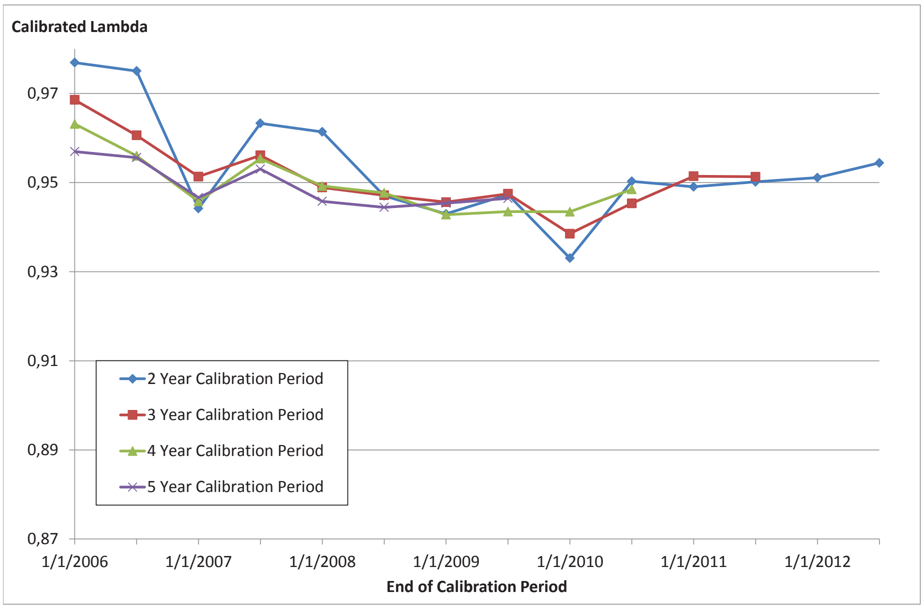
(a) Calibration for all Risk factors with a Tenor below 6 Months



(b) Calibration for all Risk factors with a Tenor between 6 and 12 Months



(c) Calibration for all Risk factors in the GBP and USD currency with a Tenor above 1 Year



(d) Calibration for all Risk factors in the EUR, CHF and JPY currency with a Tenor above 1 Year

B.2 Seperating the impact of different discount curves in VaR estimation of a FRA

As explained in section 2.7 on page 31 a FRA has three risk factors on two different discount curves. By fixing either R_{t,t_1,t_2} or R_2 at any value in eq. (19) on page 33 risk factors can be deactivated. If R_{t,t_1,t_2} is replaced by a constant the two risk factors on the discount curve with a $t_2 - t_1$ reset frequency do not influence the FRA value anymore and therefore aren't risk factors of the FRA anymore. The same is true for the risk factor at t_2 on the OIS curve if R_2 is replaced by a constant. However the main risk driver is R_{t,t_1,t_2} as can be seen when comparing the VaR values in figure fig. 13 on the following page with those in fig. 14 on page 86 Backtesting performance metrics for a range of products based on the one month € discount curve can be found in appendix B.2 on page 87. As can be seen all volatility scenarios still work well, when the two discount curves are regarded independently. Based on these results backtesting performance without partial risk factor deactivation is contributed to a good modeling of the volatility of the risk factor on the forward curve. Nevertheless, results should also hold for alternative products, which are based mainly on the OIS discount curve as appendix B.2 on page 87 shows, that the volatility scenarios also result in good backtesting performance metrics for products, which exclusively use OIS risk factors.

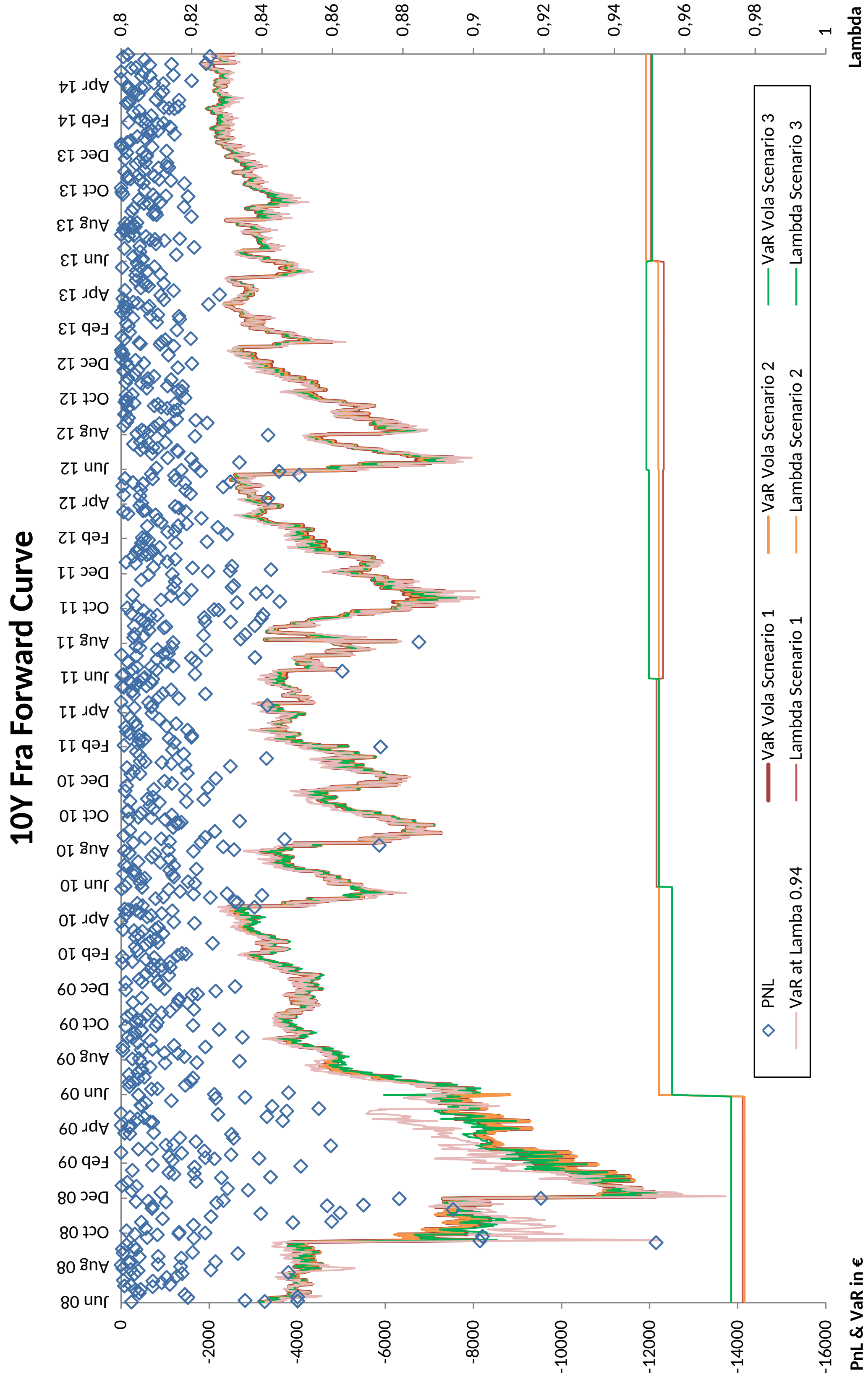


Figure 13: Backtesting results under the different volatility scenarios using a valuation formula for the FRA for which the only risk factor is on the forward curve.

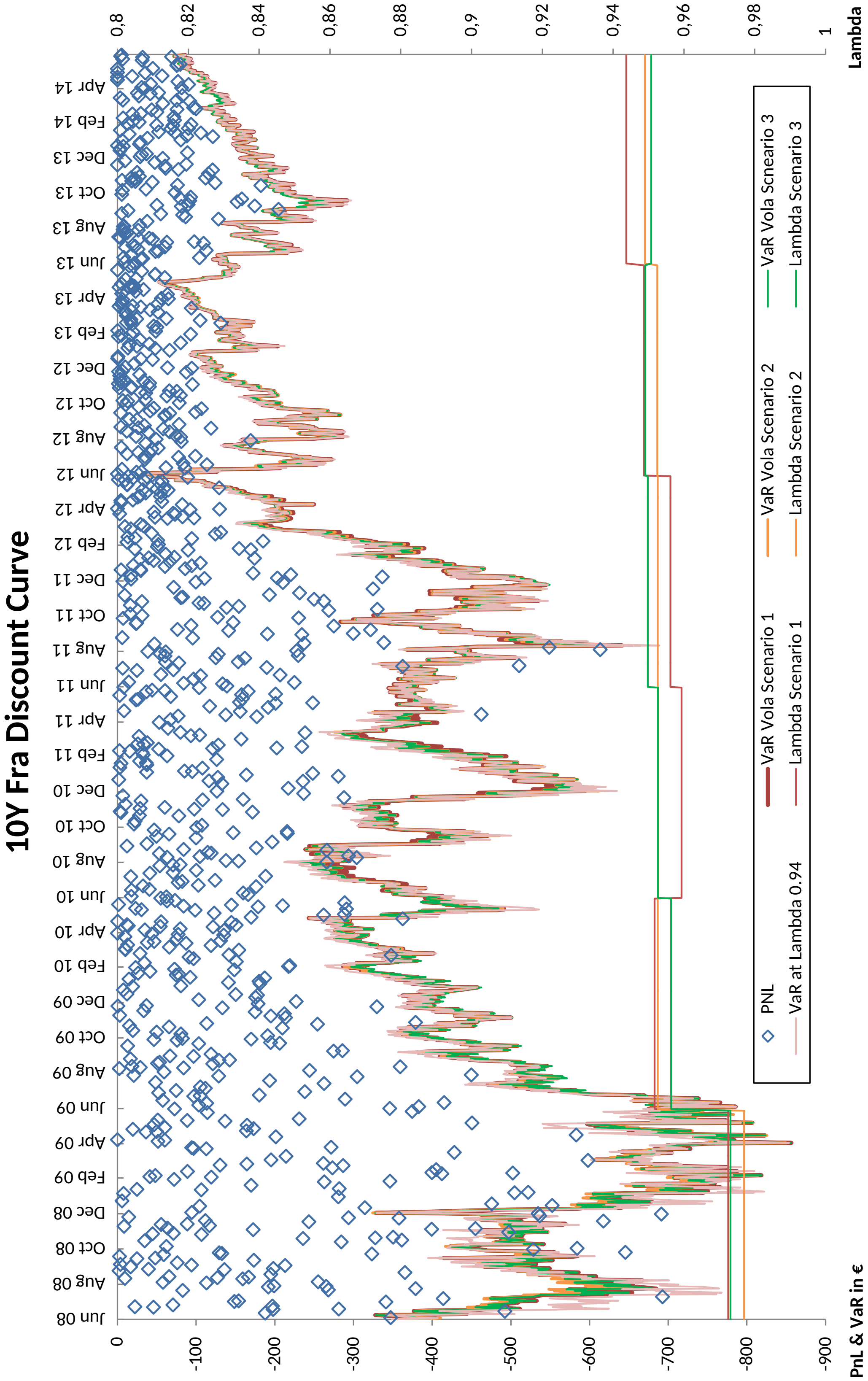


Figure 14: Backtesting results under the different volatility scenarios using a valuation formula for the FRA for which the only risk factor is on the OIS discount curve.

Table 10: Backtests of € products with only one risk factor

			Kupiec	Christtofersen		Engle / Manganelli - CAViaR									Christoffersen & Pelletier	
Product	Revaluation	Lambda	UC	Ind	CC	DQ UC Lag 1	DQ UC Lag 2	DQ UC Lag 3	DQ Ind Lag 1	DQ Ind Lag 2	DQ Ind Lag 3	DQ CC Lag 1	DQ CC Lag 2	DQ CC Lag 3	Duration CC	Duration Ind
€ 1M/10Y	Forward Only	**M 1**	53,4%	63,8%	73,8%	54,2%	54,0%	57,0%	75,8%	66,2%	27,5%	79,4%	90,7%	15,1%	8,8%	**4,4%**
		M 2	53,4%	63,8%	73,8%	54,2%	54,0%	57,0%	75,8%	66,2%	27,5%	79,4%	90,7%	15,1%	8,8%	**4,4%**
		M 3	72,3%	61,2%	82,6%	72,5%	72,4%	74,3%	73,1%	62,4%	33,2%	88,7%	94,8%	20,9%	10,8%	**4,4%**
		0,94	14,2%	71,7%	31,9%	16,7%	16,5%	16,3%	83,6%	76,8%	71,7%	38,1%	57,9%	73,4%	21,2%	95,7%
€ 1M/2Y	Forward Only	**M 1**	87,1%	56,2%	83,4%	86,8%	87,3%	87,0%	67,4%	23,1%	43,5%	90,3%	20,2%	31,3%	14,8%	5,1%
		M 2	72,3%	61,2%	82,6%	72,5%	74,3%	74,3%	73,1%	15,3%	30,8%	88,7%	12,4%	21,0%	7,7%	**2,9%**
		M 3	72,3%	61,2%	82,6%	72,5%	74,3%	74,3%	73,1%	15,3%	30,8%	88,7%	12,4%	21,0%	7,7%	**2,9%**
		0,94	53,4%	63,8%	73,8%	54,2%	57,1%	57,0%	75,8%	12,1%	25,3%	79,4%	8,4%	15,1%	5,6%	**2,6%**
€ 1M/8M	Forward Only	**M 1**	81,7%	10,9%	26,9%	83,1%	83,0%	83,1%	**1,4%**	14,8%	30,4%	**4,7%**	9,5%	16,9%	18,8%	7,9%
		M 2	96,8%	12,9%	31,6%	97,0%	97,0%	96,7%	**2,0%**	**0,1%**	**2,1%**	6,8%	**1,8%**	**3,2%**	**4,1%**	**1,2%**
		M 3	96,8%	12,9%	31,6%	97,0%	97,0%	96,7%	**2,0%**	**0,1%**	**2,1%**	6,8%	**1,8%**	**3,2%**	**4,1%**	**1,2%**
		0,94	96,8%	12,9%	31,6%	97,0%	96,7%	96,5%	**2,0%**	18,4%	36,6%	6,8%	13,0%	21,8%	8,4%	**2,7%**
€ 1M/5M	Forward Only	**M 1**	81,7%	62,0%	86,1%	81,7%	81,7%	81,7%	73,3%	62,7%	55,0%	91,9%	96,2%	98,2%	67,3%	46,5%
		M 2	57,1%	54,1%	70,6%	55,7%	57,6%	57,0%	64,0%	25,7%	48,3%	75,8%	18,9%	29,1%	28,8%	12,2%
		M 3	57,1%	54,1%	70,6%	55,7%	57,6%	57,0%	64,0%	25,7%	48,3%	75,8%	18,9%	29,1%	28,8%	12,2%
		0,94	75,9%	56,7%	81,0%	75,4%	76,5%	76,1%	67,2%	21,1%	40,8%	87,1%	16,9%	27,0%	19,0%	6,8%
€ 1M/2M	Forward Only	**M 1**	**4,3%**	39,9%	9,0%	**2,5%**	**3,0%**	**2,8%**	45,0%	64,3%	98,0%	6,7%	6,1%	9,6%	19,6%	78,8%
		M 2	7,3%	42,1%	14,5%	5,0%	7,2%	6,8%	48,2%	**3,2%**	14,3%	12,1%	**0,0%**	**0,1%**	22,9%	46,0%
		M 3	11,9%	44,4%	22,2%	9,2%	12,4%	11,8%	51,3%	**2,2%**	10,5%	20,3%	**0,0%**	**0,1%**	20,9%	23,3%
		0,94	11,9%	44,4%	22,2%	9,2%	10,4%	9,9%	51,3%	49,6%	82,9%	20,3%	12,4%	18,9%	41,1%	80,0%
€ 1M/2Y	Discount Only	**M 1**	17,1%	29,0%	22,4%	16,3%	15,6%	15,0%	11,5%	51,4%	84,7%	10,2%	16,8%	24,8%	18,5%	14,8%
		M 2	17,1%	29,0%	22,4%	16,3%	15,6%	15,0%	11,5%	51,4%	84,7%	10,2%	16,8%	24,8%	18,5%	14,8%
		M 3	25,6%	46,8%	40,2%	22,8%	22,1%	21,4%	55,6%	40,1%	30,0%	41,4%	54,6%	64,4%	44,4%	36,5%
		0,94	17,1%	29,0%	22,4%	16,3%	15,6%	15,0%	11,5%	51,4%	84,7%	10,2%	16,8%	24,8%	18,5%	14,8%
€ 1M/8M	Discount Only	**M 1**	96,8%	12,9%	31,6%	97,0%	96,7%	96,5%	**2,0%**	18,4%	36,6%	6,8%	13,0%	21,8%	42,1%	19,6%
		M 2	60,9%	9,0%	20,8%	64,4%	64,1%	64,0%	**0,9%**	11,6%	24,7%	**2,9%**	6,3%	11,8%	30,3%	18,5%
		M 3	60,9%	9,0%	20,8%	64,4%	64,1%	64,0%	**0,9%**	11,6%	24,7%	**2,9%**	6,3%	11,8%	30,3%	18,5%
		0,94	96,8%	12,9%	31,6%	97,0%	96,7%	96,5%	**2,0%**	18,4%	36,6%	6,8%	13,0%	21,8%	42,1%	19,6%
€ 1M/5M	Discount Only	**M 1**	81,7%	10,9%	26,9%	83,1%	83,0%	83,1%	**1,4%**	14,8%	30,4%	**4,7%**	9,5%	16,9%	82,4%	72,1%
		M 2	96,8%	59,3%	86,6%	96,7%	96,5%	96,2%	70,2%	58,7%	50,3%	92,9%	96,1%	97,8%	90,6%	70,5%
		M 3	81,7%	62,0%	86,1%	81,7%	81,7%	81,7%	73,3%	62,7%	55,0%	91,9%	96,2%	98,2%	67,6%	47,0%
		0,94	81,7%	62,0%	86,1%	81,7%	81,7%	81,7%	73,3%	62,7%	55,0%	91,9%	96,2%	98,2%	84,1%	76,6%
€ 1M/2M	Discount Only	**M 1**	81,7%	60,6%	85,2%	81,9%	81,7%	61,5%	72,3%	16,5%	36,1%	91,5%	13,0%	18,6%	49,0%	24,1%
		M 2	96,8%	14,0%	33,6%	98,2%	96,4%	96,3%	**2,7%**	17,5%	36,2%	8,6%	13,0%	21,7%	32,6%	13,4%
		M 3	96,8%	14,0%	33,6%	98,2%	96,4%	96,3%	**2,7%**	17,5%	36,2%	8,6%	13,0%	21,7%	32,6%	13,4%
		0,94	81,7%	11,8%	28,8%	81,6%	83,3%	83,2%	**1,9%**	13,9%	29,9%	6,2%	9,5%	16,8%	50,6%	25,3%

References

Andersen, T. G. and Bollerslev, T. (1996). Heterogeneous information arrivals and return volatility dynamics: uncovering the long-run in high frequency returns. Technical report, National Bureau of Economic Research.

Barone-Adesi, G., Giannopoulos, K., and Bourgoin, F. (1998). Don't look back. *Risk*, 11(8):100–104.

Barone-Adesi, G., Giannopoulos, K., and Vosper, L. (1999). Var without correlations for portfolios of derivative securities. *Journal of Futures Markets*, 19(5):583–602.

Barone-Adesi, G., Giannopoulos, K., and Vosper, L. (2002). Backtesting derivative portfolios with filtered historical simulation (fhs). *European Financial Management*, 8(1):31–58.

Basle Committee on Banking Supervision (1996). bcbs22 - Supervisory Framework for the use of "Backtesting" in conjunctiuon with the internal models approach to market risk capital requirements. Technical report, Basle Comitte on Banking Supervision.

Beltratti, A. and Morana, C. (1999). Computing value at risk with high frequency data. *Journal of empirical finance*, 6(5):431–455.

Berkowitz, J. (2001). Testing density forecasts, with applications to risk management. *Journal of Business & Economic Statistics*, 19(4):465–474.

Berkowitz, J., Christoffersen, P., and Pelletier, D. (2011). Evaluating value-at-risk models with desk-level data. *Management Science*, 57(12):2213–2227.

Berndt, E. R., Hall, B. H., Hall, R. E., and Hausman, J. A. (1974). Estimation and inference in nonlinear structural models. In *Annals of Economic and Social Measurement, Volume 3, number 4*, pages 653–665. NBER.

boerse.de (2015). Wkn: A1ewww.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 31(3):307–327.

Bollerslev, T., Chou, R. Y., and Kroner, K. F. (1992). Arch modeling in finance: A review of the theory and empirical evidence. *Journal of econometrics*, 52(1):5–59.

Box, G. E., Jenkins, G. M., and Reinsel, G. C. (2011). *Time series analysis: forecasting and control*, volume 734. John Wiley & Sons.

Burns, P. (2002). Robustness of the ljung-box test and its rank equivalent. *Available at SSRN 443560*.

Butler, J. and Schachter, B. (1997). Estimating value-at-risk with a precision measure by combining kernel estimation with historical simulation. *Review of Derivatives Research*, 1:371–390.

Campbell, S. D. (2005). *A review of backtesting and backtesting procedures*. Divisions of Research & Statistics and Monetary Affairs, Federal Reserve Board.

Cappiello, L., Engle, R. F., and Sheppard, K. (2006). Asymmetric dynamics in the correlations of global equity and bond returns. *Journal of Financial econometrics*, 4(4):537–572.

Christoffersen, P. and Pelletier, D. (2004). Backtesting value-at-risk: A duration-based approach. *Journal of Financial Econometrics*, 2(1):84–108.

Christoffersen, P. F. (1998). Evaluating interval forecasts. *International economic review*, pages 841–862.

Danielsson, J. and Zigrand, J.-P. (2006). On time-scaling of risk and the square-root-of-time rule. *Journal of Banking & Finance*, 30(10):2701–2713.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica: Journal of the Econometric Society*, pages 987–1007.

Engle, R. F. and Bollerslev, T. (1986). Modelling the persistence of conditional variances. *Econometric reviews*, 5(1):1–50.

Engle, R. F. and Manganelli, S. (2004). Caviar: Conditional autoregressive value at risk by regression quantiles. *Journal of Business & Economic Statistics*, 22(4):367–381.

Engle, R. F. and Sheppard, K. (2001). Theoretical and empirical properties of dynamic conditional correlation multivariate garch. Technical report, National Bureau of Economic Research.

Eurex Clearing AG (2014). Eurex clearing prisma - user guide: Methdology description release 2.0. Only available for Eurex Clearing Members.

Glosten, L. R., Jagannathan, R., and Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *The journal of finance*, 48(5):1779–1801.

Greene, W. H. (2008). *Econometric analysis*. Granite Hill Publishers.

Gurrola-Perez, P. and Murphy, D. (2015). Working Paper No 525 - Filtered historical simulation Value-at-Risk models and their competitors. Technical report, Bank of England.

Gyntelberg, J. and Upper, C. (2014). The otc interest rate derivatives market in 2013. `http://www.bis.org/publ/qtrpdf/r_qt1312h.htm`. Accessed: 2015-07-31.

Hagan, P. S. and West, G. (2006). Interpolation methods for curve construction. *Applied Mathematical Finance*, 13(2):89–129.

Hansen, P. R. and Lunde, A. (2005). A forecast comparison of volatility models: does anything beat a garch (1, 1)? *Journal of applied econometrics*, 20(7):873–889.

Hoel, P. G. (1954). A test for markov chains. *Biometrika*, 41:430–433.

Hull, J. and White, A. (1998). Incorporating volatility updating into the historical simulation method for value-at-risk. *Journal of Risk*, 1(1):5–19.

Hull, J. C. (2006). *Options, futures, and other derivatives*. Pearson Education India.

Jarque, C. M. and Bera, A. K. (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics letters*, 6(3):255–259.

Jesus, S. and Gabriel, J. (2006). Credit cycles, credit risk, and prudential regulation. *International Journal of Central Banking*.

Joanes, D. and Gill, C. (1998). Comparing measures of sample skewness and kurtosis. *The statistician*, pages 183–189.

Jorion, P. (1997). *Value at risk: the new benchmark for controlling market risk*. Irwin Professional Pub.

Kupiec, P. H. (1995). Techniques for verifying the accuracy of risk measurement models. *THE J. OF DERIVATIVES*, 3(2).

Ljung, G. M. and Box, G. E. (1978). On a measure of lack of fit in time series models. *Biometrika*, 65(2):297–303.

Longin, F. and Solnik, B. (1995). Is the correlation in international equity returns constant: 1960–1990? *Journal of international money and finance*, 14(1):3–26.

Massey Jr, F. J. (1951). The kolmogorov-smirnov test for goodness of fit. *Journal of the American statistical Association*, 46(253):68–78.

MathWorks (2015). Maximum likelihood estimation for conditional variance models.

McNeil, A. J. and Frey, R. (2000). Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. *Journal of empirical finance*, 7(3):271–300.

Millar, R. B. (2011). *Maximum likelihood estimation and inference: with examples in R, SAS and ADMB*, volume 111. John Wiley & Sons.

Morgan, J. (1996). *Riskmetrics: technical document*. Morgan Guaranty Trust Company of New York.

Murphy, D., Vasios, M., and Vause, N. (2014a). An investigation into the procyclicality of risk-based initial margin models. Technical report, Bank of England.

Murphy, D., Vasios, M., and Vause, N. (2014b). An investigation into the procyclicality of risk-based initial margin models. *Bank of England Financial Stability Paper*, 29.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica: Journal of the Econometric Society*, pages 347–370.

on Banking Supervision, B. C. (1996). *Supervisory Framework for the Use of" backtesting" in Conjunction with the Internal Models Approach to Market Risk Capital Requirements*. Bank for International Settlements.

Pafka, S. and Kondor, I. (2001). Evaluating the riskmetrics methodology in measuring volatility and value-at-risk in financial markets. *Physica A: Statistical Mechanics and its Applications*, 299(1):305–310.

Pritsker, M. (2006). The hidden dangers of historical simulation. *Journal of Banking & Finance*, 30(2):561–582.

Sandoval, L. and Franca, I. D. P. (2012). Correlation of financial markets in times of crisis. *Physica A: Statistical Mechanics and its Applications*, 391(1):187–208.

Schwarz, G. et al. (1978). Estimating the dimension of a model. *The annals of statistics*, 6(2):461–464.

Spanos, A. (1999). *Probability theory and statistical inference: econometric modeling with observational data*. Cambridge University Press.

Student (1908). The probable error of a mean. *Biometrika*, pages 1–25.

Swiss National Bank (2011). Press release regarding market intervention of the snb. `http://www.snb.ch/de/mmr/reference/pre_20110906/source/pre_20110906.de.pdf`. Accessed: 2015-08-25.

Thornton, D. L. (2009). What the libor-ois spread says. *Economic Synopses*.

Wald, A. (1943). Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Transactions of the American Mathematical society*, 54(3):426–482.

Erklärung

Ich erkläre hiermit ehrenwörtlich, dass ich die vorliegende Masterarbeit

Filtered historical Simulation for Portfolios: Model Selection and Calibration

selbständig angefertigt habe. Die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sind als solche kenntlich gemacht.

Karlsruhe, den 8. September 2015

....

(Andreas Höcherl)