Applying Multivariate Expected Shortfall on High Frequency Foreign Exchange Data

SARA HOLMSÄTER

EMELIE MALMBERG
Applying Multivariate Expected Shortfall on High Frequency Foreign Exchange Data

SARA HOLMSÄTER
EMELIE MALMBERG

Master’s Thesis in Financial Mathematics (30 ECTS credits)
Master Programme in Industrial Engineering and Management (120 credits)
Royal Institute of Technology year 2016
Supervisor at SEB: Dr. Pär Hellström
Supervisor at KTH: Dr. Fredrik Armerin
Examiner: Prof. Dr. Boualem Djehiche

TRITA-MAT-E 2016:52
ISRN-KTH/MAT/E--16/52--SE
Abstract

This thesis aims at implementing and evaluating the performance of multivariate Expected Shortfall models on high frequency foreign exchange data. The implementation is conducted with a unique portfolio consisting of five foreign exchange rates; EUR/SEK, EUR/NOK, EUR/USD, USD/SEK and USD/NOK. High frequency is in this context defined as observations with time intervals from second by second up to minute by minute. The thesis consists of three main parts. In the first part, the exchange rates are modelled individually with time series models for returns and realized volatility. In the second part, the dependence between the exchange rates is modelled with copulas. In the third part, Expected Shortfall is calculated, the risk contribution of each exchange rate is derived and the models are backtested.

The results of the thesis indicate that three of the five final models can be rejected at a 5% significance level if the risk is measured by Expected Shortfall (ES_{0.05}). The two models that cannot be rejected are based on the Clayton and Student’s t copulas, the only two copulas with heavy left tails. The rejected models are based on the Gaussian, Gumbel-Hougaard and Frank copulas. The fact that some of the copula models are rejected emphasizes the importance of choosing an appropriate dependence structure. The risk contribution calculations show that the risk contributions are highest from EUR/NOK and USD/NOK, and that EUR/USD has the lowest risk contribution and even decreases the portfolio risk in some cases. Regarding the underlying models, it is concluded that for the data used in this thesis, the final combined time series and copula models perform quite well, given that the purpose is to measure the risk. However, the most important parts to capture seem to be the fluctuations in the volatilities as well as the tail dependencies between the exchange rates. Thus, the predictions of the return mean values play a less significant role, even though they still improve the results and are necessary in order to proceed with other parts of the modelling. As future research, we first and foremost recommend including the liquidity aspect in the models.

Keywords: Multivariate Expected Shortfall, Component Expected Shortfall, Backtesting, Time Series Modelling, ARIMA-RealGARCH, Copulas, Foreign Exchange Rates, High Frequency Data
Implementering av multidimensionell Expected Shortfall på högfrekvent växelkursdata

Sammanfattning


De slutgiltiga resultaten indikerar att tre av de fem föreslagna modellerna kan förkastas vid en signifikansnivå på 5% om risken mäts med Expected Shortfall (ES$_{0.05}$). De två modeller som inte kan förkastas är baserade på Clayton och Student’s t copulas, vilka särskiljer sig från övriga copulas genom att de har tjocka vänstersvansar. De modeller som förkastas är baserade på Gaussian, Gumbel-Hougaard och Frank copulas. Det faktum att några copula-modeller förkastas betonar vikten av att välja en lämplig beroendestruktur. Riskbidragsberäkningarna visar att EUR/NOK och USD/NOK bidrar mest till den totala risken i portföljen och att EUR/USD har det lägsta riskbidraget, där EUR/USD till och med minskar risken i vissa fall. Vad gäller underliggande modeller så visas det att för den tillgängliga datan i den här uppsatsen så fungerar tidsseriemodeller i kombination med copulas bra, givet att syftet är att måta risk. Dock tyder resultaten på att volatilitetsfluktuationer samt svansberoenden mellan växelkurserna är de mest väsentliga delarna att fånga. Väntevärdesprognoserna för avkastningarna har mindre inverkan på de slutgiltiga beräkningarna, även om de fortfarande förbättrar resultaten och i sig är nödvändiga för fortsatt modellering. För framtida studier rekommenderar vi först och främst att inkludera likviditetsaspekter i modellerna.
Acknowledgements

We would like to thank our supervisor Dr. Fredrik Armerin at the Department of Mathematics, The Royal Institute of Technology, for his valuable feedback and advice. We would also like to thank Dr. Pär Hellström at SEB for his relevant input and valuable expertise in the FX markets. Furthermore, we would like to thank friends and family for their extensive support throughout the progress of the thesis.
Contents

Abstract iii
Acknowledgements vii

1 Introduction 1
  1.1 Background . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 1
  1.2 Purpose . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 2
  1.3 Delimitations . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 3

2 Theory 5
  2.1 Expected Shortfall . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 5
    2.1.1 Definition . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 5
    2.1.2 Multivariate Expected Shortfall . . . . . . . . . . . . . . . . . . . . 6
    2.1.3 Decomposition of Risk Contributions . . . . . . . . . . . . . . . . 6
    2.1.4 Backtesting . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 7
  2.2 Time Series Modelling . . . . . . . . . . . . . . . . . . . . . . . . . . . . 11
    2.2.1 Predicting Returns using ARIMA Models . . . . . . . . . . . . . . 12
    2.2.2 Modelling Dynamic Volatility using GARCH Models . . . . . . . 19
    2.2.3 ARIMA-RealGARCH . . . . . . . . . . . . . . . . . . . . . . . . . . . . 24
    2.2.4 Standardized Residual Distributions . . . . . . . . . . . . . . . . . 25
    2.2.5 Model Selection and Validation . . . . . . . . . . . . . . . . . . . . . 27
  2.3 Dependence Structures . . . . . . . . . . . . . . . . . . . . . . . . . . . 29
    2.3.1 Pearson’s Linear Correlation . . . . . . . . . . . . . . . . . . . . . . 29
    2.3.2 Kendall’s Tau Rank Correlation . . . . . . . . . . . . . . . . . . . . 30
    2.3.3 Introduction to Copulas . . . . . . . . . . . . . . . . . . . . . . . . . 30
    2.3.4 Elliptical Copulas . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 31
    2.3.5 Archimedean Copulas . . . . . . . . . . . . . . . . . . . . . . . . . . . 33
    2.3.6 Parameter Estimation . . . . . . . . . . . . . . . . . . . . . . . . . . . 37
    2.3.7 Model Validation . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 39

3 Data 41
  3.1 Description . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 41
  3.2 Data Modifications . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 42
  3.3 Final Data . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 43

4 Results 45
  4.1 Time Series Models . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 45
    4.1.1 ARIMA-RealGARCH . . . . . . . . . . . . . . . . . . . . . . . . . . . 45
    4.1.2 Residual Diagnostics . . . . . . . . . . . . . . . . . . . . . . . . . . . 50
  4.2 Dependence Structures . . . . . . . . . . . . . . . . . . . . . . . . . . . . 52
    4.2.1 Correlation Measures . . . . . . . . . . . . . . . . . . . . . . . . . . . 52
    4.2.2 Graphical Analysis . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 53
    4.2.3 Copulas . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 55
4.3 Expected Shortfall .................................................. 59
  4.3.1 Convergence of Simulations ............................. 59
  4.3.2 Portfolio Expected Shortfall ......................... 60
  4.3.3 Component Expected Shortfall ..................... 63
  4.3.4 Backtesting ............................................. 64

5 Concluding Remarks ................................. 67
  5.1 Conclusion .................................................. 67
  5.2 Discussion ................................................ 68
  5.3 Future Research ....................................... 71

A Appendix ............................................. 73
  A.1 Descriptive Statistics ................................... 73
     A.1.1 EUR/SEK ................................................. 73
     A.1.2 EUR/NOK ................................................ 74
     A.1.3 EUR/USD ................................................. 74
     A.1.4 USD/NOK ................................................ 75
  A.2 Time Series Model Selection ............................ 76
     A.2.1 EUR/SEK ................................................. 76
     A.2.2 EUR/NOK ................................................ 76
     A.2.3 EUR/USD ................................................. 77
     A.2.4 USD/NOK ................................................ 77
  A.3 Time Series Model Diagnostics ....................... 78
     A.3.1 EUR/SEK ................................................. 79
     A.3.2 EUR/NOK ................................................ 80
     A.3.3 EUR/USD ................................................. 81
     A.3.4 USD/NOK ................................................ 82
  A.4 Time Series Forecasts ................................... 82
  A.5 Copulas .................................................... 83
  A.6 Expected Shortfall ....................................... 84
     A.6.1 Description of Implementation .................... 84
     A.6.2 Graphs .................................................. 85
List of Figures

2.1 Measured variance vs true variance ........................................ 21
2.2 Level curves and density plot of a Gaussian copula with marginal
distributions $\sim N(0, 1)$ and $\rho = 0.5$ ................................. 32
2.3 Level curves and density plot of a Student’s $t$ copula with marginal
distributions $\sim N(0, 1), \rho = 0.5$ and $\nu = 2$ ......................... 33
2.4 Level curves and density plot of a Clayton copula with marginal
distributions $\sim N(0, 1)$ and $\theta = 2$ ................................. 35
2.5 Level curves and density plot of a Gumbel-Hougaard copula with
marginal distributions $\sim N(0, 1)$ and $\theta = 2$ ..................... 36
2.6 Level curves and density plot of a Frank copula with marginal
distributions $\sim N(0, 1)$ and $\theta = 2$ ................................. 37

3.1 USD/SEK exchange rate, one month, 09.00-17.30 ........................ 43
3.2 USD/SEK, one month, 09.00-17.30 ........................................ 44
4.1 USD/SEK autocorrelation diagnostics ................................. 46
4.2 Actual log-returns vs forecasts ........................................ 47
4.3 Actual vs forecasted realized volatility ............................... 49
4.4 Actual vs forecasted realized volatility ............................... 50
4.5 USD/SEK diagnostics of standardized residuals ................... 51
4.6 EUR/SEK and EUR/USD against USD/SEK ......................... 54
4.7 EUR/NOK and EUR/USD against USD/NOK ......................... 54
4.8 Gaussian copula with marginal distributions ...................... 56
4.9 Student’s $t$ copula with marginal distributions .................. 57
4.10 Clayton copula with marginal distributions ...................... 57
4.11 Gumbel-Hougaard copula with marginal distributions .......... 58
4.12 Frank copula with marginal distributions ...................... 58
4.13 Residual portfolio outcomes - Gaussian copula .................. 59
4.14 VaR and ES of residual portfolio simulations .................. 60
4.15 Portfolio log-returns .................................................. 60
4.16 Value at Risk and Expected Shortfalls - all copulas .......... 61
4.17 Expected Shortfalls and Value at Risk - Clayton and Student’s $t$ 62
4.18 Calculated Expected Shortfalls and difference ................. 62
4.19 Expected Shortfall comparisons .................................. 63
4.20 Component Expected Shortfall ...................................... 63

A.1 EUR/SEK exchange rate, one month, 09.00-17.30 .................. 73
A.2 EUR/SEK, one month, 09.00-17.30 ................................. 73
A.3 EUR/NOK exchange rate, one month, 09.00-17.30 ................ 74
A.4 EUR/NOK, one month, 09.00-17.30 ................................. 74
A.5 EUR/USD exchange rate, one month, 09.00-17.30 ................ 74
A.6 EUR/USD, one month, 09.00-17.30 ................................. 75
A.7 USD/NOK exchange rate, one month, 09.00-17.30 ................ 75
<table>
<thead>
<tr>
<th>Section</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A.8</td>
<td>USD/NOK, one month, 09.00-17.30</td>
<td>75</td>
</tr>
<tr>
<td>A.9</td>
<td>EUR/SEK autocorrelation diagnostics</td>
<td>79</td>
</tr>
<tr>
<td>A.10</td>
<td>EUR/SEK diagnostics of standardized residuals</td>
<td>79</td>
</tr>
<tr>
<td>A.11</td>
<td>EUR/NOK autocorrelation diagnostics</td>
<td>80</td>
</tr>
<tr>
<td>A.12</td>
<td>EUR/NOK diagnostics of standardized residuals</td>
<td>80</td>
</tr>
<tr>
<td>A.13</td>
<td>EUR/USD autocorrelation diagnostics</td>
<td>81</td>
</tr>
<tr>
<td>A.14</td>
<td>EUR/USD diagnostics of standardized residuals</td>
<td>81</td>
</tr>
<tr>
<td>A.15</td>
<td>USD/NOK autocorrelation diagnostics</td>
<td>82</td>
</tr>
<tr>
<td>A.16</td>
<td>USD/NOK diagnostics of standardized residuals</td>
<td>82</td>
</tr>
<tr>
<td>A.17</td>
<td>Forecasted vs actual log-returns</td>
<td>83</td>
</tr>
<tr>
<td>A.18</td>
<td>USD/NOK forecasts</td>
<td>83</td>
</tr>
<tr>
<td>A.19</td>
<td>Elliptical copulas without EUR/USD</td>
<td>84</td>
</tr>
<tr>
<td>A.20</td>
<td>Archimedean copulas without EUR/USD</td>
<td>84</td>
</tr>
<tr>
<td>A.21</td>
<td>Gaussian copula</td>
<td>85</td>
</tr>
<tr>
<td>A.22</td>
<td>Frank copula</td>
<td>86</td>
</tr>
<tr>
<td>A.23</td>
<td>Gumbel copula</td>
<td>86</td>
</tr>
</tbody>
</table>
# List of Tables

3.1 Example of raw data .............................................. 41  
3.2 Example of adjusted data .......................................... 42  
4.1 USD/SEK - BIC matrix with RealGARCH(1,1) ................. 46  
4.2 Estimated best fit ARIMA-parameters ............................. 47  
4.3 Estimated best fit log RealGARCH-parameters ................. 48  
4.4 Estimated best fit log RealGARCH-parameters ................. 48  
4.5 Li-Mak test result \( p \)-values ................................. 49  
4.6 Jarque-Berra test for Normal Distribution ....................... 50  
4.7 Distributions of standardized residuals, \( z_t \) ............... 51  
4.8 Ljung-Box test for USD/SEK, joint \( H_0 \) ..................... 52  
4.9 Pearson’s Linear Correlation Matrix .............................. 53  
4.10 Kendall’s Tau Rank Correlation Matrix ......................... 53  
4.11 Copula parameter estimations .................................. 55  
4.12 Cramér-von-Mises statistics for copulas ....................... 56  
4.13 Component ES\(_{0.05}\), expressed as \( 10^{-3} \) .............. 64  
4.14 Number of VaR exceedances \( n_T \) and \( p \)-values from the backtests . 64  
A.1 EUR/SEK - BIC matrix with RealGARCH(1,1) ................. 76  
A.2 EUR/NOK - BIC matrix with RealGARCH(1,1) ................. 76  
A.3 EUR/USD - BIC matrix with RealGARCH(1,1) ................. 77  
A.4 USD/NOK - BIC matrix with RealGARCH(1,1) ................. 77  
A.5 \( p \)-values for the best fit ARIMA-parameters ................ 78  
A.6 \( p \)-values for the best fit RealGARCH-parameters ........... 78  
A.7 \( p \)-values for the best fit RealGARCH-parameters ........... 78  
A.8 Ljung-Box test for EUR/SEK, joint \( H_0 \) no autocorrelation 79  
A.9 Ljung-Box test for EUR/NOK, joint \( H_0 \) no autocorrelation 80  
A.10 Ljung-Box test for EUR/USD, joint \( H_0 \) no autocorrelation 81  
A.11 Ljung-Box test for USD/NOK, joint \( H_0 \) no autocorrelation 82  
A.12 Parameter estimations without EUR/USD ...................... 83
Chapter 1

Introduction

1.1 Background

During recent years Expected Shortfall (ES) has become an increasingly popular risk measure in the financial industry. In comparison to Value at Risk (VaR), which still is the most commonly used risk measure, ES has some noteworthy advantages. Two of the most important advantages are that ES better captures the risk arising from heavy tails (risk of extreme losses) and that it is a coherent measure of risk, meaning that diversification is rewarded (Yamai and Yoshiha 2005). In October 2013, the Basel Committee on Banking Supervision proposed a change to ES as the new industry standard for calculating losses and measuring risks. However, this proposal was criticized since practitioners pointed out that ES does not possess a mathematical property known as elicitability\(^1\), which made people believe that it could not be backtested. Following the ambiguity and uncertainty regarding backtesting, several articles were written about how to backtest ES, showing that it indeed is possible to backtest without the elicitability property (Acerbi and Szekely 2014). This gives further rise to the prospect that ES might soon be the most used risk measure in the industry.

Even though ES has been recognized as a well-performing risk measure, it is still to be tested in a wider context. As of today, it has mainly been implemented with longer time horizon data consisting of daily, weekly or yearly observations. Furthermore, most studies implement ES on portfolios consisting of stocks or stock indices. Hitherto, there does not seem to be any research on the performance of ES models on shorter time horizons and portfolios consisting of currencies. Moreover, even though ES is a relatively simple measure, there are many different ways of implementing it. Most of the current literature on both VaR and ES addresses only univariate approaches. Multivariate approaches, however, have the advantage that they from a portfolio perspective allow the models to capture both the characteristics of the underlying assets as well as the dependence structure between them. Unfortunately, the multivariate models are not yet as well-developed as their univariate counterparts, and it is not always straightforward to select an appropriate model and evaluate its performance.

\(^1\)Elicitability allows a measure to have a scoring function that makes comparison of different models possible in a natural way. However, even though it is relevant for model selection, it is not necessarily relevant for backtesting. For further elaboration on the topic, see e.g. (Sherif 2014).
In order to create an accurate and reliable ES model for measuring risk in financial data, there are some important aspects that need to be considered, both regarding the underlying assets and the dependence structure. Most importantly, the model needs to capture the riskiness of extreme portfolio outcomes in terms of probability and impact.

When modelling underlying assets, time dependencies need to be taken into account. For example, historic returns may entail information about future returns, which implies that there might be significant autocorrelation. Instead of assuming constant means, we may therefore implement time series models such as ARIMA and allow for time varying predictions of the mean values. Moreover, the volatilities of the underlying assets are most often time dependent as well (Cont 2005). This can for example be seen during periods of financial distress and concerns, when volatilities tend to be higher than during periods of stability and calm markets. This is referred to as volatility clustering, and to take the volatility clustering into account it is important to also model the volatilities using a dynamic time series model such as GARCH. The combination of both time dependent returns and volatility clustering can consequently be handled through ARIMA-GARCH-models, which capture both of these features.

When modelling the dependence structure, it is important to note that dependencies are not always linear. For example, when asset prices collapse, they tend to collapse together even though they might not be correlated when the markets are calm. In line with the purpose of risk modelling, it is crucial to capture these extreme outcomes. Modelling the dependence structure using copulas is therefore theoretically a suitable approach since copulas allow for non-linear tail dependencies, meaning that it is taken into consideration that assets might have stronger dependencies during periods of financial distress. However, neither the multivariate copula theory (for more than two assets) nor the implementation aspect is yet fully developed. Fortunately, five of the most popular and well-known copulas are extended to the multivariate case, namely the Gaussian, Student’s t, Clayton, Gumbel-Hougaard and Frank copulas. These copulas also cover all possible cases of no tail dependence, lower tail dependence, upper tail dependence and both lower and upper tail dependence.

By combining and implementing some of the latest contributions within the fields of ES, backtesting, copulas and time series analysis, we contribute to the existing knowledge with a thorough and complete analysis of each of the steps in modelling and computing multivariate ES. Furthermore, the implementation is conducted with an in this setting unique portfolio consisting of high frequency observations of foreign exchange rates.

1.2 Purpose

The purpose of this thesis is to evaluate the performance of different multivariate ES models on a portfolio consisting of currencies and intraday observations. The performance of the parametric models will be evaluated through backtesting of ES on historical returns. Thus, the main focus is to determine whether the ES models generate good results also for shorter time horizons and portfolios consisting of currencies. Throughout the thesis, good results is defined
as models that cannot be rejected on a 5% significance level. The five currency exchange rates that will be considered are: EUR/SEK, EUR/NOK, EUR/USD, USD/SEK and USD/NOK, with minute by minute observations of the returns for each of the exchange rates. Second by second observations will also be used in order to compute realized volatility for the minute by minute returns, which will be further introduced and discussed in section 2.2.2.

In order to obtain good results it is crucial to find suitable ways of modelling the underlying assets and the dependence between them. In this thesis, we will investigate whether time series models describe the individual currency fluctuations sufficiently and if the dependence structure can be satisfactorily described by copulas, with the purpose of predicting ES. When the performance of the models has been assessed, we will also derive the risk contributions of the underlying assets to determine how much each asset contributed to the total portfolio risk.

1.3 Delimitations

First and foremost, we will only look at the multivariate version of ES, and the parts included in the analysis are: time series models for returns, volatility models for time varying volatility and dependence through copulas. We will focus specifically on ARIMA-RealGARCH, where the actual volatility will be measured as the square root of the realized variance. The five different copulas that will be considered as dependence structures are: Gaussian, Student’s t, Clayton, Gumbel-Hougaard and Frank.

Another delimitation is that the thesis only will cover five exchange rates and observations minute by minute. This is primarily due to the fact that the data available for the thesis is of this nature, and it is not certain that the conclusions can be generalized to other currencies or time horizons without further research and validation. Besides, a lot of research has already been conducted for ES as a risk measure for longer time horizons and stock portfolios, which means that in order to contribute to the existing knowledge a more specific setting is needed.
Chapter 2

Theory

2.1 Expected Shortfall

2.1.1 Definition

Expected Shortfall (ES) is a natural extension of Value at Risk (VaR), which is one of the most commonly used risk measures today. In order to define ES we therefore start with a short recap of VaR.

VaR of a portfolio with payoff function $X$ at level $\alpha$, denoted $\text{VaR}_\alpha(X)$, is defined as the worst loss for a given confidence level $(1 - \alpha) \in (0, 1)$. Thus, for a confidence level of for example 99%, there is a 99% probability that at the end of the chosen risk horizon there will be no greater loss than the value of $\text{VaR}_{0.01}$.

If we now define $X$ as the net portfolio return$^1$ from time 0 to time 1 (defined as the end of the risk horizon), we can approximate the value of the loss at time 0 as $L = -X$ for short intraday time horizons$^2$. In statistical terms, $\text{VaR}_\alpha$ is the $(1 - \alpha)$-quantile of the loss $L$. If the distribution function $F_L$ is both continuous and strictly increasing we may express VaR as

$$\text{VaR}_\alpha(X) = F^{-1}_L(1 - \alpha). \quad (2.1)$$

Since VaR is only a quantile-value, it ignores the values beyond the chosen probability $\alpha$ (i.e. the highest values in the right tail of the loss function, or equivalently, the lowest values in the left tail of the payoff function), which makes it possible to hide risk in the tail. A natural extension of VaR is therefore to include these tail values and compute an average VaR below the level $\alpha$. This is referred to as ES, computed as

$$\text{ES}_\alpha(X) = \frac{1}{\alpha} \int_0^\alpha \text{VaR}_u(X)du. \quad (2.2)$$

$^1$In this thesis, we define $X$ as the net portfolio return and not as the net value of the portfolio. To obtain the actual size of the loss we simply multiply the obtained percentage by the size of the position.

$^2$For longer time horizons, $L = -X/R_0$, where $R_0$ is the risk-free return.
By combining formula 2.1 and 2.2 and using the defined loss function $F_L$, we see that ES is simply an average of quantile values of $L$

$$ES_\alpha(X) = \frac{1}{\alpha} \int_0^\alpha F_L^{-1}(1-u)du. \quad (2.3)$$

While VaR is translation invariant, monotone and positively homogeneous, it is not subadditive, meaning that diversification is not always rewarded. ES, on the other hand, inherits the translation invariance, monotonicity and positive homogeneity, but is also subadditive, which makes it a coherent measure of risk (Hult, Lindskog, Hammarlid, and Rehn 2012).

### 2.1.2 Multivariate Expected Shortfall

Following the subadditivity property of ES, we know that diversification is rewarded, meaning that $ES(Y_1 + Y_2) \leq ES(Y_1) + ES(Y_2)$. Thus, it is necessary to model the multivariate distribution of the asset returns, since the sum of expected shortfall for the underlying assets is only equal to the portfolio expected shortfall when the underlying random variables have perfect positive dependence (comonotonicity) (Yamai and Yoshina 2002).

With previous notation, we construct a portfolio net return $X$ at time 1 through a linear combination of individual net asset returns $Y_i$ and asset weights $w_i$, so that $X = w_1Y_1 + \cdots + w_nY_n$, or equivalently $X = w^T Y$ where $w = (w_1, \cdots, w_n)^T$ and $Y = (Y_1, \cdots, Y_n)^T$. With a given weight vector consisting of real numbers, the remaining task is to find a suitable multivariate distribution of the vector $Y$. The better the characteristics of the true multivariate distribution are captured by the estimated model of $Y$, the better the ES estimations will be.

### 2.1.3 Decomposition of Risk Contributions

When the portfolio ES has been computed, a natural next step is to determine where the greater parts of the portfolio risk originates from. One way to do this is through computing the individual assets’ risk contributions, referred to as component ES. With the previous definitions and positive homogeneity it holds that

$$ES_\alpha(X) = \sum_{i=1}^n \frac{\partial ES_\alpha(X)}{\partial w_i} \cdot w_i, \quad (2.4)$$

where the portfolio ES is decomposed into the portfolio weights times the partial derivatives of ES with respect to the portfolio weights. As proved by Tasche (Tasche 2000), if the underlying distributions are continuous, then the marginal risk contribution $\partial ES_\alpha(X)/\partial w_i$ for asset $i$ can be represented as the following conditional expectation
\[
\frac{\partial \text{ES}_\alpha(X)}{\partial w_i} = \mathbb{E}[ -Y_i \mid X \leq -\text{VaR}_\alpha(X)]. \tag{2.5}
\]

The component ES for asset \(i\) can then be expressed as \(^3\)

\[
c\text{ES}_i = \mathbb{E}[ -Y_i \mid X \leq -\text{VaR}_\alpha(X)] \cdot w_i. \tag{2.6}
\]

Note that the sum of all individual risk contributions (i.e. the sum of all component ES) is equal to the total portfolio ES, which can easily be seen with ES as \(\text{ES}_\alpha(X) = \mathbb{E}[ -X \mid X \leq -\text{VaR}_\alpha(X)]\).

Since it is difficult to compute the right hand side of the conditional expectation analytically (equation (2.5) and (2.6)), especially when the multivariate distribution of \(X\) is rather complex, we may instead approximate it through a procedure suggested by (Hallerbach 1999) and implemented by amongst others (Fan, Zeng, and Wong 2012). The marginal risk contribution for asset \(i\) is then estimated as

\[
\frac{\partial \text{ES}_\alpha(X)}{\partial w_i} = \mathbb{E}[ -Y_i \mid X \leq -\text{VaR}_\alpha(X)] \approx \frac{1}{T^*} \sum_{j=1}^{T^*} -Y_i^j, \tag{2.7}
\]

where \(Y_i^j\) is the return of asset \(i\), with \(1 \leq i \leq n\) and \(n\) as the total number of assets. Here, \(j\) is the observation number (either from historical observations or simulations) and we have that \(1 \leq j \leq T^*\), where \(T^*\) is the observation number of the observation where the portfolio loss is equal to \(\text{VaR}_\alpha(X)\). Thus, from a total sample of \(T\) observations, a data window where the portfolio losses are larger than or equal to the portfolio VaR is chosen.

From this we obtain the approximated marginal ES for each of the underlying assets for the worst portfolio outcomes. By multiplying the approximated marginal ES with the portfolio weights we obtain the approximated component ES, which can be interpreted as the risk contributions of the assets.

### 2.1.4 Backtesting

To evaluate the performance of an implemented ES model it is necessary to compare it to actual outcomes. As previously mentioned, there has been a long ongoing discussion about whether ES is backtestable or not, even though there is evidence of successful backtesting of ES as early as year 2000 (McNeil and Frey 2000). To overcome this misconception, Acerbi and Szekely (Acerbi

\(^3\)The marginal risk contribution is denoted as \(\mathbb{E}[X_i \mid X \geq \text{VaR}_p(X)]\) in e.g. (Fan, Zeng, and Wong 2012), where \(X_i\) represents the loss of asset \(i\). In order to be consistent with our notation we instead use the asset net return \(Y_i\), meaning that \(Y_i = -X_i\). Furthermore, our \(X\) represents the net portfolio return as opposed to the portfolio loss. However, the definition and meaning of \(\text{VaR}_p(X)\) is the same (even though \(X\) is different), which leads to our slightly modified formulas (2.5) and (2.6).
Chapter 2. Theory

and Szekely 2014) published a report about backtesting ES in December 2014, where they present three different alternatives of backtesting ES. All methods are non-parametric and free from distributional assumptions other than continuity (which is a necessary condition in banking regulations). In this thesis we have chosen to focus on and present the first and second backtesting method. The second method has the advantage that it tests ES directly, without requiring that VaR is tested first. Unfortunately, the acceptance rate in the second method tends to be too high. In order to perform a more conservative backtest, the first method is thus preferred (Wimmerstedt 2015). However, the first method is only valid given that the VaR is tested and accepted in advance.

One of the most common and straightforward ways of testing VaR is referred to as the unconditional coverage. In this test, the number of portfolio outcomes that exceed the estimations of VaR is calculated for a given sample of size \( N \). The test is then built on the number of exceedances \( n \), the sample size \( N \) and a given confidence level \( \alpha \). Assuming that all exceedances are independent and that the probability of exceedance is \( \alpha \), the distribution of \( n \) will be binomially distributed

\[
\begin{align*}
    f(n) &= \binom{N}{n} \alpha^n (1 - \alpha)^{N-n}, \quad (2.8)
\end{align*}
\]

that is, \( n \sim \text{Bin}(N, \alpha) \). Hence, the probability of an outcome larger than or equal to \( n \) can easily be calculated and used to assess whether the estimated VaR is plausible as the real VaR for the sample. If the probability is low, the hypothesis that the estimated VaR is the real VaR would be rejected. Note that this test only gives an assessment of whether VaR is underestimated or not, and may therefore suggest acceptance of a model where the estimated VaR is considerably larger than the real VaR (Campbell 2005).

To present the ES backtesting methods, we will now outline the underlying theory and tests as described in the report published by Acerbi and Szekely (Acerbi and Szekely 2014). Consider the pay-off function \( X_t \) for \( t = 1, \ldots, T \) with underlying true distribution \( F_t \). The random variables \( X = \{X_t\} \) are assumed to be independent but not necessarily identically distributed. \( X_t \) is then forecasted by a model predictive distribution \( P_t \), conditional to previous information. \( P_t \) is the estimated distribution which VaR\(_{\alpha,t}\) and ES\(_{\alpha,t}\) are computed from.

On this set-up we get the formula for ES\(_{\alpha,t}\),

\[
    \text{ES}_{\alpha,t} = \frac{1}{\alpha} \int_0^\alpha P_t^{-1}(1 - u) du, \quad (2.9)
\]

As can be seen, there are no restrictions regarding whether \( F_t \) and \( P_t \) vary with time. However, by assuming that the distributions are continuous and strictly increasing we can express ES as

\footnote{For a time series, there can be different VaR estimations for each point in time, as compared to only having one fixed VaR estimation over the entire time period.}
2.1. Expected Shortfall

\[ ES_{\alpha,t} = E[-X_t|X_t \leq -\text{VaR}_{\alpha,t}(X_t)], \quad \text{where} \quad \text{VaR}_{\alpha,t} = P_t^{-1}(1 - \alpha). \]  \hspace{1cm} (2.10)

**Test 1**

Test 1 is inspired by the conditional expectation 2.10 from which we can derive

\[ E \left[ \frac{X_t}{ES_{\alpha,t}} + 1 | X_t \leq -\text{VaR}_{\alpha,t} \right] = 0. \]  \hspace{1cm} (2.11)

To perform the first backtest we formulate the two hypotheses:

\[ H_0: \quad P_t^{[\alpha]} = F_t^{[\alpha]}, \quad \forall t \]

\[ H_1: \quad ES_{\alpha,t}^F \geq ES_{\alpha,t} \quad \forall t \text{ and } > \text{ for some } t \]

\[ \text{VaR}_{\alpha,t}^F = \text{VaR}_{\alpha,t} \quad \forall t \]

where \( P_t^{[\alpha]}(x) = \min(1, P_t(x)/\alpha) \) is the distribution tail for \( x < -\text{VaR}_{\alpha,t} \) and \( \text{VaR}_{\alpha,t}^F \) and \( ES_{\alpha,t}^F \) are the true risk measures of \( X \sim F \). Thus, we only test if the assumed predictive distribution underestimates the risk, and the hypothesis \( H_0 \) implies that the predictive distribution coincides with the true distribution in the \( \alpha \)-tail. We also see that \( \text{VaR}_{\alpha,t} \) is still correct under hypothesis \( H_1 \), since it is assumed that \( \text{VaR}_{\alpha,t} \) has already been tested and accepted.

The test statistic \( Z_1 \) for Test 1 is given by

\[ Z_1(X) = \frac{\sum_{t=1}^{T} X_t I_t}{n_T} + 1, \]  \hspace{1cm} (2.12)

where \( I_t = I(X_t < -\text{VaR}_{\alpha,t}) \) and \( n_T = \sum_{t=1}^{T} I_t \) has to be larger than zero.

The fact that Test 1 is completely insensitive to an excessive number of exceedances, \( n_T \), can be seen in the formula for the test statistic \( Z_1 \). Here, we see that the test statistic is in fact an average taken over the exceedances themselves. Therefore, it is crucial that the \( \text{VaR}_{\alpha,t} \) is not underestimated.

**Test 2**

Test 2 is based on the unconditional expectation
ES_{\alpha,t} = -E\left[\frac{X_tI_t}{\alpha}\right]. \quad (2.13)

To perform the second backtest we formulate the two hypotheses:

\begin{align*}
H_0 &: P_t^{[\alpha]} = F_t^{[\alpha]}, \quad \forall t \\
H_1 &: ES_{F,\alpha,t} \geq ES_{\alpha,t} \quad \forall t \quad \text{and} \quad \forall t > \text{for some } t \\
\text{VaR}_{F,\alpha,t} \geq \text{VaR}_{\alpha,t} \quad \forall t \quad \text{and} \quad \forall t > \text{for some } t
\end{align*}

The test statistic $Z_2$ is given by

$$Z_2(\tilde{X}) = \sum_{t=1}^{T} \frac{X_tI_t}{T \alpha ES_{\alpha,t}} + 1, \quad (2.14)$$

where $I_t = \mathbb{I}(X_t < -\text{VaR}_{\alpha,t})$ and ES_{\alpha,t} is defined as the unconditional expectation (2.13).

**Significance Test**

The significance test is performed in the same way for both Test 1 and Test 2, with test statistics $Z$ equal to $Z_1$ and $Z_2$, respectively. To start with, VaR_{\alpha,t} and the conditional ES_{\alpha,t} are calculated for Test 1 and the unconditional ES_{\alpha,t} is calculated for Test 2. Here it is important that VaR and ES are calculated using a sufficiently large set of simulations so that the outcomes converge. After this first step, the distribution $P_Z$ can be simulated under $H_0$. That is, test statistics are simulated for all $Z^i = Z(\{X_t\})$, where $i = 1, \ldots, M$. Each $i$ simulation corresponds to one simulated time series. One $\hat{Z}$ is thereafter calculated from the realizations, $\hat{Z} = Z(\{x_t\})$, where $\{x_t\}$ is the vector of actual observed time series values. The $p$-value, $p = P_Z(Z(\{x_t\}))$, and the steps to calculate the $p$-value can in other words be summarized as:

1. Test 1: Simulate VaR_{\alpha,t} and conditional ES_{\alpha,t}
   Test 2: Simulate unconditional ES_{\alpha,t}
2. Simulate independent $X^i_t \sim P_t, \forall t$ and $\forall i = 1, \ldots, M$
3. Compute $Z^i = Z(\{X^i_t\})$
4. Estimate $p = \sum_{i=1}^{M} (Z^i < Z(\{x_t\}))/M$

where $M$ has to be sufficiently large so that the outcomes converge. We define a significance level $\phi$ and the hypothesis is rejected if $p < \phi$ (Acerbi and Szekely 2014).
2.2 Time Series Modelling

When modelling financial data there are in general two important characteristics that need to be considered: autocorrelation in returns and time dependent volatility (Cont 2005). In this section, we present models that will be used to describe the log-returns of the exchange rates while accounting for these factors. We start by looking at ARIMA time series models that are based on autocorrelations. Thereafter, the standard deviation (volatility) of the series will be estimated for each point in time with GARCH models. After adjusting for this, the series should only contain random fluctuations, referred to as standardized residuals. In case of a correct model specification, the standardized residuals are covariance stationary and serially uncorrelated with mean value 0 and standard deviation 1. They are in general described by a white noise process WN(0, 1), but are most often modelled by an appropriate distribution such as Normal, Student’s t or Generalized Error Distribution (Diebold 2007).

Before we move over to the theory behind time series modelling, we start by introducing three main concepts that will be recurring throughout this section: covariance stationarity, autocorrelation and partial autocorrelation.

A time series is covariance stationary if its mean and covariance structure are stable over time, which means that the mean and covariance are independent of \( t \in (-\infty, \infty) \). That is, if we have a real-valued stochastic process \( \{X_t\} \), then the expected value of \( X_t \) is constant

\[
E[X_t] = \mu_X(t) = \{\text{independent of } t\} = \mu_X \quad (2.15)
\]

and the autocovariance, which is the covariance of a time series process with respect to itself at different points in time, is finite and depends only on the lag \( \tau \). The autocovariance of a covariance stationary time series is therefore given by

\[
\gamma_X(t, t \pm \tau) = \text{cov}(X_t, X_{t \pm \tau}) = \{\text{together with condition (2.15)}\}
\]

\[
= E[(X_t - \mu)(X_{t \pm \tau} - \mu)] = \{\text{independent of } t\}
\]

\[
= \gamma_X(\tau) < \infty. \quad (2.16)
\]

It can be shown that no autocovariance \( \gamma(\tau) \) of a time series can be larger in absolute value than \( \gamma(0) \) of the series, which automatically leads to that \( \gamma(\tau) \forall \tau < \infty \) if \( \gamma(0) < \infty \).

The autocorrelation is derived directly from the autocovariance, and is in general only meaningful if the time series is covariance stationary. For a covariance stationary time series, the autocorrelation function, \( \rho_X(\tau) \), is given by
\( \rho_X(\tau) = \frac{\text{cov}(X_t, X_{t+\tau})}{\sqrt{\text{cov}(X_t, X_t) \text{cov}(X_{t+\tau}, X_{t+\tau})}} \)

\[ \begin{align*}
\ &= \left\{ \text{using (2.15) and (2.16)} \right\} \\
\ &= \frac{\gamma_X(\tau)}{\gamma_X(0)}.
\end{align*} \]

(2.17)

The autocorrelation describes the linear dependence within the series for time lag \( \tau \). However, intermediate time lag effects indirectly affect the value of the autocorrelation for lag \( \tau \). To measure the direct dependence between \( X_t \) and \( X_{t+\tau} \), we may use the partial autocorrelation. For a covariance stationary time series, the partial autocorrelation \( \pi_X(j) = \pi_j \) can be found by performing a linear regression on \( X_t \) and its lagged values \( X_{t-1}, \ldots, X_{t-h} (h > 1) \), so that

\[ X_t = \pi_1 X_{t-1} + \cdots + \pi_h X_{t-h}. \] (2.18)

For the partial autocorrelation we have that \( \pi_X(0) = 1 \) (simply meaning that \( X_t = X_t \)), \( \pi_X(1) = \rho_X(1) \) (the partial autocorrelation is equal to the normal autocorrelation for the first lag) and \( \pi_X(h) = \pi_X(-h) \) (the partial autocorrelation is symmetric) (Diebold 2007).

### 2.2.1 Predicting Returns using ARIMA Models

The underlying components of time series models are usually divided into three main groups: trends, seasonals and cycicals. The trend describes the slow, long-run evolution of a time series and can be both linear and nonlinear. A common way of handling the trend in financial data is through differencing the series. That is, instead of modelling quoted prices, we model the returns or log-returns of the series. The seasonal component captures the repeating patterns that can be seen over a specific period, oftentimes consisting of monthly or quarterly effects. Seasonality not only appears in longer time frames, but can also be found in intraday data. For example, there might be significant seasonal effects for morning and afternoon data, both regarding volatility and returns (Zheng, Moulines, and Abergel 2012). Because of limited time and scope, the seasonal component will not be considered in this thesis. The cyclicity is any sort of dynamics not captured by trends or seasonals. Accordingly, there is a wider variety and complexity of cyclical patterns, but the purpose of the cyclical component is to describe different types of autocorrelation behavior in the series. A common way of modelling the cycle of a differentiated series is through ARIMA-models. Through applying ARIMA models we use the fact that historic returns and residuals may entail information about future returns, which makes it possible to forecast future returns using ARIMA models (Diebold 2007). The main reference for the following sections is the book Elements of Forecasting by Diebold (Diebold 2007).
2.2. Time Series Modelling

Moving Average Models

The $q$:th-order moving average process, MA$(q)$, is characterized by that the current value, $y_t$, of the observed time series can be expressed as a linear combination of current and lagged innovations $(\varepsilon_t, \varepsilon_{t-1}, \ldots, \varepsilon_{t-q})$. The process is on the form

$$y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}, \quad \varepsilon_t \sim WN(0, \sigma^2), \quad (2.19)$$

where WN stands for white noise, defined as a sequence of uncorrelated random variables with mean 0 and finite variance $\sigma^2$. If all the random variables are independent and come from the same probability distribution, then the white noise is independent identically distributed (iid) noise (Brockwell and Davis 2010).

An MA process is always covariance stationary. However, for a time series to be applicable for forecasting, it has to be possible to describe the process using an autoregressive representation. That is, the current value of the process has to be able to be described in terms of the current innovation plus lagged values of the series (as opposed to the current value in terms of current and lagged innovations). The autoregressive representation is achievable if the MA$(q)$-process is invertible, which is the case if the inverses of all of the $q$ roots are inside the unit cycle, and the process can then formally be written as

$$\frac{1}{\Theta(L)} y_t = \varepsilon_t, \quad (2.20)$$

with the numerator polynomial $\Theta(L)$ and the lag operator $L$ defined as

$$\Theta(L) = 1 + \sum_{i=1}^{q} \theta_i L^i, \quad L^i y_t = y_{t-i}. \quad (2.21)$$

Furthermore, the unconditional mean and variance of an MA$(q)$ process are

$$E[y_t] = E[\varepsilon_t] + \theta_1 E[\varepsilon_{t-1}] + \cdots + \theta_q E[\varepsilon_{t-q}] = 0 \quad (2.22)$$

and

$$\text{Var}(y_t) = \begin{cases} \text{using that cov}(\varepsilon_t, \varepsilon_s) = 0 \forall t \neq s \end{cases} \quad = \text{Var}(\varepsilon_t) + \theta_1^2 \text{Var}(\varepsilon_{t-1}) + \cdots + \theta_q^2 \text{Var}(\varepsilon_{t-q}) \quad = \sigma^2 + \theta_1^2 \sigma^2 + \cdots + \theta_q^2 \sigma^2 \quad = \sigma^2 \left(1 + \sum_{i=1}^{q} \theta_i^2\right), \quad (2.23)$$
respectively. As illustrated, the unconditional mean and variance are constant and independent of time. If we instead look at the conditional mean and variance, we see that the conditional mean evolves with the information set and that the larger the \( q \), the greater the potential for longer memory becomes.

We begin by defining the information set, which is the (in mathematical terms) filtration \( F_{t-1} \) given by

\[
F_{t-1} = \{ \varepsilon_{t-1}, y_{t-1}, \varepsilon_{t-2}, y_{t-2}, \ldots \},
\]

so that all information up until time \( t - 1 \) is known with the filtration \( F_{t-1} \).

Using this for the conditional mean and variance, with the innovation \( \varepsilon_t | F_{t-1} \sim WN(0, \sigma^2) \), we get

\[
E[ y_t | F_{t-1} ] = E[ \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} | F_{t-1} ]
= E[ \varepsilon_t | F_{t-1} ] + \theta_1 E[ \varepsilon_{t-1} | F_{t-1} ] + \cdots + \theta_q E[ \varepsilon_{t-q} | F_{t-1} ]
= \sum_{i=1}^q \theta_i \varepsilon_{t-i}
\]

and

\[
\text{Var}( y_t | F_{t-1} ) = \text{Var}( \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} | F_{t-1} )
= \left\{ \begin{array}{l} 
\text{using that } \text{cov}( \varepsilon_t, \varepsilon_s | F_{t-1} ) = 0 \ \forall \ t \neq s \\
\text{Var}( \varepsilon_t | F_{t-1} ) + \theta_1^2 \text{Var}( \varepsilon_{t-1} | F_{t-1} ) + \cdots + \theta_q^2 \text{Var}( \varepsilon_{t-q} | F_{t-1} ) \\
\text{using that } \text{Var}( \varepsilon_{t-1} | F_{t-1} ) = \cdots = \text{Var}( \varepsilon_{t-q} | F_{t-1} ) = 0 \\
\text{Var}( \varepsilon_t ) \\
\sigma^2.
\end{array} \right.
\]

For an MA(\( q \))-process, the autocorrelations are significant for all lags \( \leq q \), but then become zero beyond displacement \( q \). The partial autocorrelations, on the other hand, decay gradually in response to the infinite autoregressive representation of MA processes.

### Autoregressive Models

The autoregressive process is a mathematical model where the current value of a series, \( y_t \), is linearly related to its past values, \( \{ y_{t-1}, y_{t-2}, \ldots \} \), plus an innovation term, \( \varepsilon_t \). The \( p \)-th order autoregressive process, \( AR(p) \), can be written as

\[
y_t = \psi_1 y_{t-1} + \psi_2 y_{t-2} + \cdots + \psi_p y_{t-p} + \varepsilon_t, \quad \varepsilon_t \sim WN(0, \sigma^2).
\]
As for the moving average process, some conditions have to hold for the autoregressive process to be applicable for forecasting. It is already in autoregressive form (which had to be confirmed in the moving average case), however, it is not certain that an autoregressive process is covariance stationary. To examine whether an AR($p$)-process is covariance stationary or not, we first have to reformulate it on its lag operator form

\[
\Phi(L) y_t = (1 - \psi_1 L - \psi_2 L^2 - \cdots - \psi_p L^p) y_t = \varepsilon_t.
\] (2.28)

An AR($p$)-process is covariance stationary if and only if the inverses of all roots of the autoregressive lag operator polynomial $\Phi(L)$ are inside the unit circle. If this holds, the autoregressive process can be written on the convergent infinite moving average form

\[
y_t = \frac{1}{\Phi(L)} \varepsilon_t = \varepsilon_t + \sum_{i=1}^{\infty} \Psi_i \varepsilon_{t-i}.
\] (2.29)

The unconditional mean and variance for the process (when covariance stationary) are

\[
E[y_t] = E \left[ \varepsilon_t + \sum_{i=1}^{\infty} \Psi_i \varepsilon_{t-i} \right] = 0
\] (2.30)

and

\[
\text{Var} (y_t) = \text{Var} \left( \varepsilon_t + \sum_{i=1}^{\infty} \Psi_i \varepsilon_{t-i} \right)
\]
\[
= \sigma^2 + \sum_{i=1}^{\infty} (\Psi_i^2 \text{Var} (\varepsilon_{t-i}))
\]
\[
= \sigma^2 \left( 1 + \sum_{i=1}^{\infty} \Psi_i^2 \right).
\] (2.31)

If we instead look at the conditional mean and variance we have

\[
E[y_t | \mathcal{F}_{t-1}] = E[\psi_1 y_{t-1} + \psi_2 y_{t-2} + \cdots + \psi_p y_{t-p} + \varepsilon_t | \mathcal{F}_{t-1}]
\]
\[
= \psi_1 y_{t-1} + \psi_2 y_{t-2} + \cdots + \psi_p y_{t-p} + E[\varepsilon_t | \mathcal{F}_{t-1}]
\]
\[
= \sum_{i=1}^{p} \psi_i y_{t-i}
\] (2.32)
Var\( (y_t | \mathcal{F}_{t-1}) = E \left[ \left( y_t - E[y_t | \mathcal{F}_{t-1}] \right)^2 | \mathcal{F}_{t-1} \right] \)
\[ = E \left[ \left( y_t - E[\psi_1y_{t-1} + \cdots + \psi_py_{t-p} + \epsilon_t | \mathcal{F}_{t-1}] \right)^2 | \mathcal{F}_{t-1} \right] \]
\[ = E \left[ \left( \psi_1y_{t-1} + \cdots + \psi_py_{t-p} + \epsilon_t - (\psi_1y_{t-1} + \cdots + \psi_py_{t-p}) \right)^2 | \mathcal{F}_{t-1} \right] \]
\[ = E[\epsilon_t^2] \]
\[ = \sigma^2. \] (2.33)

Using these relations we can calculate the autocovariance, autocorrelation and partial autocorrelation, from which we can recognize AR\( (p) \)-processes. To find the autocovariance we use the recursive Yule-Walker equation. It is derived by multiplying the AR\( (p) \)-process at time \( t \) by \( y_{t-\tau} \), for \( \tau \geq 0 \), so that

\[ y_t y_{t-\tau} = \psi_1y_{t-1}y_{t-\tau} + \psi_2y_{t-2}y_{t-\tau} + \cdots + \psi_py_{t-p}y_{t-\tau} + \epsilon_t y_{t-\tau}. \] (2.34)

By taking expectations on both sides, using that the expected value of \( y_t \) is 0 (formula (2.30)) and that the expected value of the error term is 0, we get the Yule-Walker equation

\[ \gamma(\tau) = \psi_1 \gamma(\tau - 1) + \psi_2 \gamma(\tau - 2) + \cdots + \psi_p \gamma(\tau - p). \] (2.35)

By rewriting on matrix form, we get the relation

\[
\begin{pmatrix}
\gamma(1) \\
\gamma(2) \\
\vdots \\
\gamma(p)
\end{pmatrix} =
\begin{pmatrix}
\gamma(0) & \gamma(1) & \cdots & \gamma(p-1) \\
\gamma(1) & \gamma(0) & \cdots & \gamma(p-2) \\
\vdots & \vdots & \ddots & \vdots \\
\gamma(p-1) & \gamma(p-2) & \cdots & \gamma(0)
\end{pmatrix}
\begin{pmatrix}
\psi_1 \\
\psi_2 \\
\vdots \\
\psi_p
\end{pmatrix}. \] (2.36)

To get the autocorrelation we simply divide the autocovariance by \( \gamma(0) \). For the partial autocorrelation, \( \pi(k) \), it can be shown that

\[ \pi(k) = 0 \quad \forall \ k > p, \] (2.37)

which intuitively corresponds to the fact that AR\( (p) \)-processes are mathematical models where the current value of a series, \( y_t \), is linearly related to its past values, \( \{y_{t-1}, y_{t-2}, \ldots, y_{t-p}\} \). Hence, the partial autocorrelation will cut off after lag \( p \) and the autocorrelation will decay gradually with displacement.
2.2. Time Series Modelling

ARMA\((p, q)\)

The ARMA\((p, q)\)-model is, as the name suggests, a combination of the AR-model for lagged values of the series and the MA-model for a moving average of the innovations. It can be described as

\[
y_t = \psi_1 y_{t-1} + \cdots + \psi_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}, \quad \varepsilon_t \sim \text{WN}(0, \sigma^2),
\]

(2.38)

or on lag operator form

\[
\Phi(L)y_t = \Theta(L)\varepsilon_t,
\]

(2.39)

where

\[
\Phi(L) = 1 - \psi_1 L - \psi_2 L^2 - \cdots - \psi_p L^p
\]
\[
\Theta(L) = 1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q.
\]

(2.40)

If the inverses of all the roots of \(\Phi(L)\) and \(\Theta(L)\) are inside the unit circle, then the process is both covariance stationary and invertible. Hence, we can rewrite it both as a convergent infinite moving average process with the representation

\[
y_t = \frac{\Theta(L)}{\Phi(L)} \varepsilon_t,
\]

(2.41)

as well as a convergent infinite autoregressive process

\[
\frac{\Phi(L)}{\Theta(L)} y_t = \varepsilon_t.
\]

(2.42)

As for the AR and MA processes, the unconditional mean is constant and the conditional mean is time varying. However, in contrast to the AR and MA processes, where either the autocorrelation or partial autocorrelation cut off at a specific displacement, both the ARMA autocorrelation and partial autocorrelation decay gradually.

ARMA models can result from different scenarios, as for example when the innovation that drives the autoregressive process is itself a moving average process or when we have an aggregation of AR and MA processes. ARMA models often provide high accuracy without excessive usage of parameters. For example, it might take an AR(5) model to achieve the same approximation accuracy as could be obtained with an ARMA\((2, 1)\).
Chapter 2. Theory

**ARIMA** ($p, d, q$)

In case of a non-stationary series with a stochastic trend, it is in many cases possible to achieve stationarity by differencing the series. That is, subtracting $y_{t-1}$ from $y_t$ and then fitting a time series model to the new obtained series. Usually, the order of integration, $d$, needed to achieve stationarity (number of times that we take the difference) is restricted to order zero or one, and series of order $d > 2$ are almost nonexistent (Diebold 2007).

The lag operator representation of an ARIMA($p, d, q$) model is

$$\Phi(L)(1 - L)^d y_t = c + \Theta(L)\varepsilon_t,$$  

(2.43)

with $\Phi(L)$ and $\Theta(L)$ as in formula (2.40).

This representation is derived from the ARMA($p, q$) model when $p - d$ of the autoregressive roots are outside the unit circle and $d$ of the autoregressive roots are on the unit circle (hence, $y_t$ is not stationary since all autoregressive roots need to be outside the unit circle). In this case we can factorize the autoregressive lag operator polynomial as $(1 - L)^d$. If we look at the specific case of $d = 1$ we get the factorization

$$\Phi(L) = \Phi(L)^1(1 - L),$$  

(2.44)

where $\Phi(L)^1$ is of degree $p - 1$ and we get

$$\Phi(L)^1(1 - L)y_t = \Theta(L)\varepsilon_t \Rightarrow \Phi(L)^1\Delta y_t = \Theta(L)\varepsilon_t.$$  

(2.45)

By reformulating on the differentiated form, we create a covariance stationary and invertible ARMA($p-1, q$)-process for $\Delta y_t$, or in other words a ARIMA($p, 1, q$)-process with $c = 0$.

The main drawback of differentiating a series is the effect it has on the variance of the estimations. The variance of an ARIMA($p, 1, q$) grows unlimitedly as we forecast further into the future, which means that the uncertainty associated with forecasts also increases unlimetedly. The unlimited increase in variance translates into an unlimited increase in the spread of the density forecasts, as the length of the forecast horizon increases. However, in this thesis we only focus on a one-step-ahead forecast horizon, meaning that the increased uncertainty connected to longer forecast horizons is not affecting our results (Diebold 2007).

---

5Here, $c$ represents a drift term which will be assumed to be equal to zero in this thesis. For more elaboration on the subject see pages 288 to 295 in Elements of Forecasting by Diebold (Diebold 2007).
2.2.2 Modelling Dynamic Volatility using GARCH Models

Measuring Volatility

The most common way of estimating volatility, given that the volatility is defined as the standard deviation of returns, is probably as the sample standard deviation of logarithmic returns,

\[ s = \sqrt{\frac{\sum_{i=1}^{N} (r_i - \bar{r})^2}{(N - 1)}}, \]  

(2.46)

with \( r_i = \log(p_i/p_{i-1}) \), where \( \log \) is the natural logarithm, \( p_i \) is the price at time \( i \), \( N \in (2, T) \) is the number of observations in the measurement time frame and \( \bar{r} \) is the average return of the observations (Blom, Enger, Englund, Grandell, and Holst 2005). A possible disadvantage of using the sample standard deviation to compute the actual volatility is that historical returns from far away in time are included. Assuming that there is some time dependence in the volatility, the further away from the current observation we get, the more misleading the sample standard deviation may be.

Another common way of measuring the unobserved variance is as the squared return. Here, we simply assume that the variance each day is the squared return for that day. The main drawback of this approach is that it is extremely noisy. Thanks to the increased access to high frequency data, some new models have been proposed during the last decade to better capture the current volatility. Andersen et al. were the first ones to present a new and complementary volatility measure based on quadratic variation, termed realized volatility, which later proved very popular (Andersen, Bollerslev, Diebold, and Labys 2000). In practical terms, the realized variance is computed simply by summing intraperiod squared returns. For equally spaced minute by minute returns, the realized variance for each minute, \( RV_t \), can thus be computed through summing second by second squared logarithmic returns \( r_{i,t} \) for all seconds \( i \in (1, 60) \) and minutes \( t \)

\[ RV_t = \sum_{i=1}^{60} r_{i,t}^2. \]  

(2.47)

Even though the idea of realized variance is simple, the theory behind is deep. With logarithmic price \( Y(t) \), continuous drift \( \mu(t) \) of finite variation, \( \sigma(t) > 0 \) as the instantaneous volatility and \( W(t) \) as a standard Brownian motion, we need to assume that the logarithmic returns \( r_t = Y(t) - Y(t - 1) \) evolve as an Itô process,

\[ dY(t) = \mu(t)dt + \sigma(t)dW(t) \Leftrightarrow r_t = \int_{t-1}^{t} \mu(s)ds + \int_{t-1}^{t} \sigma(s)dW(s). \]  

(2.48)
Here, the population measure of the actual logarithmic return variance is defined as the integrated variance, $IV_t = \int_{t-1}^{t} \sigma^2(s) ds$. For Itô processes in particular, it holds that the quadratic variation $QV_t$ equals the integrated variance $IV_t$. The realized variance $RV_t$ is the empirical counterpart of the quadratic variation, and since $RV_t$ is a consistent estimator of $QV_t$ for all semimartingales, $IV_t$ is also consistently estimated by $RV_t$. The realized variance (squared volatility) measure can thus be made arbitrarily close to the integrated variance (squared volatility) over the interval of interest (Hansen and Lunde 2010).

One problem with this approach, though, is that the consistency of the estimator depends on the sample frequency. In practice, the sampling frequency is limited by the actual quotation or transaction frequency. Furthermore, for high frequency data there is also microstructure effects, referred to as noise, which will lower the accuracy of the measurements\(^6\). The benefit of the increased frequency will therefore be counterweighted by the increased noise effect, leading to a trade-off between these\(^7\). One of the main reasons that realized measures are still very valuable is that volatility oftentimes has dynamic properties, so that a more accurate measure of current volatility may be helpful for forecasting future volatility. Furthermore, in order to assess the performance of volatility models, it is of critical importance to have accurate measures of the actual volatilities. In fact, without the realized measures of volatility it is difficult to distinguish good volatility models from bad models (Hansen and Lunde 2010). An evaluation which is not based on a realized measure may actually be misleading and suggest an inferior model as the best with a probability that converges to one as the sample size increases (Hansen and Lunde 2005). In this thesis, we use the original definition of realized volatility to measure volatilities, since it is the most commonly used realized measure (Hansen and Lunde 2010), and the realized volatility per minute may simply be constructed by adding up squared logarithmic returns per second.

\(^6\)The noise effects can be due to a variety of reasons, including: bid-ask bounces, discreteness of price changes, differences in trade sizes, gradual response of prices to a block trade, strategic component of the order flow, etc. The reason why practitioners usually try to exclude this in the modelling is simply because the noise effect is believed to reflect momentary fluctuations rather than the underlying value (Ait-Sahalia and Yu 2009).

\(^7\)To get around this, some studies implement kernel-based estimators developed to reduce noise effects (Barndorff-Nielsen, Hansen, Lunde, and Shephard 2008). Moreover, instead of only looking at returns, there are volatility estimators that for example explore the price range or durations between specific levels of price changes since data is oftentimes not equally spaced (Pigorsch, Pigorsch, and Popov 2010). However, the data used in this thesis is limited to only consist of equally spaced observations of the best bid and offer, and hence these alternative volatility estimators will not be considered in this thesis.
2.2. Time Series Modelling

**Figure 2.1:** Measured variance vs true variance

The following simple example illustrates the significant improvement when changing measure from squared returns to realized variance. Assuming that the true distribution of the minute by minute returns \( r_m \) is \( N(0, 1) \) and that the returns are independent, the variance of the returns is \( \sigma^2_m = 1 \). In figure 2.1 (A) we see the squared returns from 1000 random outcomes of \( r_m \) plotted against the true variance. As can be seen, the measured variance is extremely noisy. In (B), we see the realized variance plotted against the true variance. Here, each of the 1000 minute by minute variance estimations are based on 60 underlying second by second return simulations, \( r_s \sim N(0, 1/\sqrt{60}) \), with variance \( \sigma^2_s = 1/60 \). The estimation of the realized variance in each minute is then computed as the sum of all 60 squared second by second returns.

**Forecasting Dynamic Volatility**

Many studies show that volatilities of market returns are time dependent, and therefore also possible to model and forecast with time series models. For example, during periods of financial distress and concerns, volatilities tend to be higher than during periods of stability and calm markets, which is referred to as volatility clustering (Cont 2005). Thanks to the dynamic nature of volatility, it is possible to model the current volatility as a function of past volatilities and residuals.

There are currently two types of models that are being discussed in the literature: reduced form volatility forecasts and model based forecasts. In the reduced form, realized measures are used to forecast future volatility with time series models (for example ARIMA). In the model based approach, the volatility forecasts are constructed from a model for returns that specifies the entire distribution (for example GARCH). Both of these approaches have gained popularity. However, in this thesis we choose to focus on GARCH. Even though there has been some doubts regarding the performance of GARCH out-of-sample, it has been shown that the poor performance was mainly due to the inaccurate measure of volatility, and that the models perform rather well when the realized volatility measure is used instead (Hansen and Lunde 2010). Furthermore, the GARCH models have some characteristics that are desirable for modelling financial data, which is discussed further in the following sections.
GARCH
When the variance (or volatility) of the error terms is changing over time we have a heteroscedastic time series, and the unconditional and conditional distribution of the error terms are given by

\[ \varepsilon_t \sim (0, \sigma^2) \quad \text{and} \quad \varepsilon_t | \mathcal{F}_{t-1} \sim (0, \sigma_t^2), \]  
(2.49)  

respectively. One model that has been designed to address and model the time-varying conditional volatility of the error terms is the GARCH \((p, q)\)-process which looks as follows

\[ y_t = \varepsilon_t, \quad \varepsilon_t | \mathcal{F}_{t-1} \sim (0, \sigma_t^2), \quad \varepsilon_t = \sigma_t z_t, \quad z_t \sim \text{iid}(0, 1) \]

\[ \sigma_t^2 = \omega + \alpha(L)\varepsilon_t^2 + \beta(L)\sigma_t^2, \quad \alpha(L) = \sum_{i=1}^{p} \alpha_i L_i, \quad \beta(L) = \sum_{j=1}^{q} \beta_j L_j \]

\[ \omega > 0, \quad \alpha_i \geq 0, \quad \beta_i \geq 0, \quad \sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j < 1. \]  
(2.50)  

The specified conditions assure that the process is covariance stationary and has positive conditional variance. The conditional variance, \( \sigma_t^2 \), is the key element of interest and is given by

\[ \text{Var}(\varepsilon_t | \mathcal{F}_{t-1}) = E \left[ (\varepsilon_t - E[\varepsilon_t | \mathcal{F}_{t-1}])^2 | \mathcal{F}_{t-1} \right] \]

\[ = \omega + \alpha(L)\varepsilon_t^2 + \beta(L)\sigma_t^2. \]  
(2.51)  

The unconditional variance is given by

\[ \text{Var}(\varepsilon_t) = E \left[ (\varepsilon_t - E[\varepsilon_t])^2 \right] \]

\[ = \frac{\omega}{1 - \sum_{i=1}^{p} \alpha_i - \sum_{j=1}^{q} \beta_j}. \]  
(2.52)  

As can be seen in formula (2.51), the conditional variance is time-varying and a serially correlated time series process. The GARCH process is in itself particularly interesting, not only since it allows for time-varying volatility, but also for some of its other characteristics. Even when modelled with normally distributed innovations, its implied unconditional distribution is symmetric and leptokurtic, which is typical for financial asset returns. It can also be shown

---

8Here, \( z_t \) is usually assumed to be normally distributed. However, the distribution can be elaborated with and changed to other distributions, such as Student’s or GED, in order to achieve a better representation of the underlying data set.
that even though the unconditional distribution of the GARCH process is lep-
tokurtic, the GARCH process converges towards normality under temporal ag-
grgregation\(^9\). Convergence to normality under temporal aggregation also holds
for financial asset returns, which together with the first characteristic shows the
suitability of GARCH processes for modelling financial asset volatility dynam-
ics (Diebold 2007).

### Realized GARCH

The standard GARCH model typically uses the squared return at time \( t \) as mea-
sure of the actual variance at time \( t \). As discussed in the beginning of the section,
using only one measurement point as estimate creates a scarce indication of the
real volatility. A model which instead utilizes the realized volatility measure is
the Realized GARCH model, RealGARCH. The RealGARCH\((p, q)\) is defined as
follows

\[
\varepsilon_t = \sigma_t z_t, \quad z_t \sim \text{iid}(0, 1) \quad (2.53)
\]

\[
\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i x_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2
\]

\[
x_t^2 = \xi + \delta \sigma_t^2 + \tau(z_t) + u_t, \quad u_t \sim \text{iid}(0, \lambda)
\]

\[
\tau(z_t) = \eta_1 z_t + \eta_2(z_t^2 - 1),
\]

where \( \varepsilon_t \) is the innovation term, \( \sigma_t^2 \) the conditional variance \( \sigma_t^2 = \text{Var}(\varepsilon_t|\mathcal{F}_{t-1}) \),
\( x_t^2 \) the realized variance and \( \tau(z) \) the leverage function (Hansen, Huang, and
Shek 2011). Here \( x_t^2 = \xi + \delta \sigma_t^2 + \tau(z_t) + u_t \) is referred to as the measurement
function. The measurement function relates the realized measure of volatility
to the latent volatility, which in this case means that it relates the realized vari-
ance to the conditional variance. In general, if \( x_t^2 \) is a consistent estimator of the
integrated variance (as in the case of realized variance) it makes sense to specify
\( x_t^2 \) as a function of the conditional variance and a random innovation. That is,
\( x_t^2 = \sigma_t^2 + w_t \), which is the simplest form of a measurement equation. However,
empirically it has been shown that the volatility tends to be asymmetrically
related to return shocks. Thus, the relation between volatility and return move-
ments may depend on the sign of the return movement. The leverage function is
therefore designed to capture this potential asymmetry and dependence struc-
ture between the returns and future volatility, which more commonly is referred
to as the leverage effect. To account for the potential asymmetry, the leverage
function is incorporated in the measurement function, as defined in equation
\( (2.53) \), where \( w_t \) can be seen as \( \tau(z_t) + u_t \).

To ensure that the conditional variance is positive, we may re-specify the Real-
GARCH on log-linear form

---

\(^9\)In this context, temporal aggregation means aggregation over time, where minute by minute
returns are turned into daily returns, daily returns turned into weekly returns and so forth.
\[ \varepsilon_t = \sigma_t z_t, \quad z_t \sim \text{iid}(0, 1) \] (2.54)

\[ \log \sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i \log \sigma_{t-i}^2 + \sum_{j=1}^{q} \beta_j \log \sigma_{t-j}^2 \]

\[ \log x_t^2 = \xi + \delta \log \sigma_t^2 + \tau(z_t) + u_t, \quad u_t \sim \text{iid}(0, \lambda) \]

\[ \tau(z_t) = \eta_1 z_t + \eta_2 (z_t^2 - 1), \]

which is the form being used for the modelling in this thesis. Thanks to the ensured positive conditional variance, the constraints on \( \omega, \alpha_i \) and \( \beta_j \) are no longer needed when RealGarch is defined on log-linear form. However, when a model is selected and parameters are estimated, we may still check that the persistence of the process lies between 0 and 1. The persistence \( P \) is given by (Hansen, Huang, and Shek 2011)

\[ P = \sum_{i=1}^{p} \beta_i + \delta \sum_{j=1}^{q} \alpha_i. \] (2.55)

Note that \( \delta \) is part of the persistence for the log-linear RealGARCH, which otherwise looks the same as the persistence for the normal GARCH process, formula (2.50).

For the linear RealGARCH, the unconditional (long-run) variance may be written as (Ghalanos 2015)

\[ E[\sigma_t^2] = \omega + \xi \sum_{i=1}^{q} \alpha_i. \] (2.56)

Thus, in the linear case it is straightforward to check that the unconditional variance is positive. For the log-linear case, we get that

\[ E[\log \sigma_t^2] = \frac{\omega + \xi \sum_{i=1}^{q} \alpha_i}{1 - P}. \] (2.57)

Here, we need to approximate the unconditional variance using for example a Taylor expansion, which will not be introduced in this thesis. However, it is notable that the right hand side of Equation (2.57) may very well take on negative values without the unconditional variance being negative.

### 2.2.3 ARIMA-RealGARCH

One way of combining time series models for returns with dynamic volatility is through ARIMA-RealGARCH models. By combining ARIMA and RealGARCH we end up with our final model specification: the ARIMA\((p, 1, q)\)-RealGARCH\((\alpha, \beta)\) model
2.2. Time Series Modelling

\[ y_t = \sum_{i=1}^{p} \psi_i y_{t-i} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} + \varepsilon_t, \quad \varepsilon_t = \sigma_t z_t, \quad (2.58) \]

where \( z_t \) and \( \sigma_t \) are defined in expression (2.54). In this thesis, we have restricted our GARCH-model to the dynamics of RealGARCH(1, 1).

To estimate the parameters of an ARIMA-RealGARCH-model, maximization of the log-likelihood function is a common approach. The selected parameter values are then the ones that maximize the likelihood of obtaining the data that was actually obtained. Since there is no closed-form expression for the ARIMA-RealGARCH log-likelihood estimator, the likelihood must be maximized numerically (Diebold 2007). However, this estimation procedure is often provided as a method incorporated in most statistical software, which is also the case in the R rugarch package used in this thesis (Ghalanos 2015).

2.2.4 Standardized Residual Distributions

As previously mentioned, the standardized residual \( z_t \) needs to be described by a white noise process \( WN(0, 1) \). This means that the standardized residual series should be uncorrelated, with mean value equal to 0 and variance equal to 1. Even though the exact distribution is not specified, it is in many cases reasonable to assume Gaussian white noise. To test the hypothesis that the residual series belongs to a Gaussian distribution, we can use the Jarque-Bera (JB) test. This tests the hypothesis that the skewness is 0 and that the kurtosis is equal to 3; the same as for a Normal distribution. The JB test statistic is computed as

\[ JB = \frac{T}{6} \left( \hat{S}^2 + \frac{1}{4}(\hat{K} - 3)^2 \right), \quad (2.59) \]

where \( \hat{S} \) is the sample skewness, \( \hat{K} \) is the sample kurtosis and \( T \) is the number of observations. Under the null hypothesis of independent normally distributed observations, the Jarque-Bera statistic is distributed in large samples as a \( \chi^2 \) random variable with \( \nu = 2 \) degrees of freedom (Diebold 2007).

Other popular distributions that have been used instead of the Gaussian are for example Student’s t and the Generalized Error Distribution (GED) (see e.g. (McNeil and Frey 2000)). These may exceed the performance of a Normal distribution in some cases, but before assuming that a more heavy-tailed distribution would be better, we have to examine the properties of the data.

In the case of a symmetric standardized error distribution with heavier tails than the Normal distribution (excess kurtosis), the standardized Student’s t distribution often performs better. When the degrees of freedom \( \nu > 2 \), a standardized Student’s t random variable \( z_T \) is given by
\[ z_T = \sqrt{\frac{\nu - 2}{\nu}} \cdot T_{\nu}, \]  
(2.60)

where \( T_{\nu} \) is a Student’s t distributed random variable and the degrees of freedom is selected to best fit the data (computed through maximum likelihood estimation in most statistical software). The GED can in some cases give an even better fit than the Student’s t distribution, partly thanks to its flexibility. It is defined by the density function (Ghalanos 2015)

\[ f(x|\alpha, \beta, \kappa) = \frac{\kappa e^{-\left|\frac{x - \alpha}{\beta}\right|^\kappa}}{2^{1+\kappa-1}\beta\Gamma(\kappa-1)}, \]  
(2.61)

where \( \alpha, \beta \) and \( \kappa \) represent the location, scale and shape parameters, respectively. It is a symmetric distribution and the location parameter equals the mode, median and mean of the distribution. The variance is given by

\[ \text{Var}(X) = \beta^2 2^{2/\kappa} \frac{\Gamma(3\kappa-1)}{\Gamma(\kappa-1)}. \]  
(2.62)

To be applicable for GARCH modelling, standardization is required also for GED. This is achievable by re-scaling, so that

\[ \text{Var}(X) = \beta^2 2^{2/\kappa} \frac{\Gamma(3\kappa-1)}{\Gamma(\kappa-1)} = 1, \]  
(2.63)

from which we get

\[ \beta = \sqrt{2^{-2/\kappa} \frac{\Gamma(\kappa-1)}{\Gamma(3\kappa-1)}}. \]  
(2.64)

The Student’s t distribution has heavier tails than the GED, but only in the regions far away from the mean, approaching \( \pm \infty \). In the more practical region of the distribution functions, the tail weights instead depend equally much on the distribution choice as on the parameter selection in GED. The distributions also differ when it comes to the shape around the mean, where the Student’s t distribution is smoother than the GED which instead has a sharp peak at the mean (Ruppert 2011).
2.2. Model Selection and Validation

A common way of deciding what number of lags \( p \) and \( q \) to include in an ARIMA\((p, 1, q)\)-model is through information criteria measures. The information criteria are used to assess the relative quality of the models through comparing their goodness of fit while accounting for the complexity. In general, the more parameters included the better the model will perform in-sample. However, too many included parameters leads to overfitting and bad results out-of-sample, which is why there is a need to penalize for the number of parameters included.

Two of the most commonly used information criterion tests are the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC), also known as Schwarz Information Criterion (SIC). The two criteria are very similar and select the same model most of the time, but there are cases of disagreement. The BIC penalizes the number of parameters more heavily than the AIC, which in some cases lead to a more parsimonious model selected by BIC while AIC sometimes selects models that are overparameterized. In fact, even as the sample gets large, AIC tends to select models that are too large. The benefit of BIC is thus that it is a consistent measure, while AIC is inconsistent\(^{10}\). However, the AIC has a property known as asymptotic efficiency, which BIC is lacking. This makes AIC a more efficient measure when it comes to selecting the best model out of an increasingly large sample of models (it selects the best model faster). Due to the parsimonious approach used in this thesis, we prefer to use BIC in cases of disagreement.

The formulas for AIC and BIC are

\[
\text{AIC} = e^{2k} \frac{\sum_{i=1}^{T} \varepsilon_i^2}{T} 
\]

\[
\text{BIC} = T^{\frac{k}{2}} \frac{\sum_{i=1}^{T} \varepsilon_i^2}{T},
\]

where \( k \) is the number of parameters, \( T \) is the number of total observations and \( \varepsilon_i \) is the error term. When comparing AIC and BIC model values, the best model is the one with the lowest AIC and BIC value (Diebold 2007).

The optimal model selected based on the information criteria is still only the best model among the given group of models. As a consequence, it is not certain that it is a well-performing model overall. One important aspect that we are considering in time series modelling is the autocorrelation. If there is still

\(^{10}\) A model selection criterion is consistent if the following properties are met (Diebold 2007)

(1) When the true model is among the models considered, the probability of selecting the true model approaches 1 as the sample size gets large

(2) When the true model is not among the models considered, the probability of selecting the best approximation approaches 1 as the sample size gets large.
significant autocorrelation present in the error terms, referred to as the residuals, then the model can be improved by taking this into account. It is therefore essential to measure and test for autocorrelation in the error terms. The autocorrelation function is useful to assess the autocorrelations one at the time for individual lags. However, the hypothesis that all autocorrelations are jointly 0 is not automatically tested.

One way to test the joint hypothesis that there is no serial correlation in the residuals is to use the null hypothesis that the residual series is a white noise process. The Ljung-Box Q-statistic is designed so that $Q_{LB}$ is approximately distributed as a $\chi^2_m$ random variable under the null hypothesis that $X$ is white noise, and it is computed as

$$Q_{LB} = T(T + 2) \sum_{\tau=1}^{m} \left( \frac{1}{T - \tau} \right) \hat{\rho}^2(\tau). \quad (2.67)$$

Here $m$ is the number of lags tested in the joint hypothesis and the sample autocorrelation $\hat{\rho}$ is given by

$$\hat{\rho}(\tau) = \frac{\sum_{t=\tau+1}^{T} ((y_t - \bar{y})(y_{t-\tau} - \bar{y}))}{\sum_{t=1}^{T} (y_t - \bar{y})^2}, \quad (2.68)$$

where $\bar{y}$ is the sample mean. As well as wanting to test whether the residuals are white noise or not, it is desirable to test for heteroscedasticity in the data. If the applied GARCH model is accurately defined, the heteroscedasticity should have been accounted for and the standardized residuals, $z_t$, should be homoscedastic. A diagnostic for this is the Li-Mak test. The Li-Mak test is built on the squared standardized residuals’ autocorrelation function

$$\hat{\rho}_{z^2}(\tau) = \frac{\sum_{t=\tau+1}^{T} ((z_t^2 - 1)(z_{t-\tau}^2 - 1))}{\sum_{t=1}^{T} (z_t^2 - 1)^2}. \quad (2.69)$$

Let $\hat{\rho}_{z^2}$ be the autocorrelation vector, $[\hat{\rho}_{z^2}(1), \hat{\rho}_{z^2}(2), \ldots, \hat{\rho}_{z^2}(m)]^T$, containing the autocorrelations from lag 1 to $m$. Then $\sqrt{T} \hat{\rho}_{z^2}$ has an asymptotic Normal distribution with mean zero and covariance matrix $\hat{V}$. The covariance matrix can be consistently estimated by $\hat{V}$,

$$\hat{V} = I_m - \frac{XG^{-1}X^T}{\frac{1}{T} \sum_{t=1}^{T} (z_t^2 - 1)^2}, \quad (2.70)$$

where $I_m$ is the $m \times m$ identity matrix and $G$ the consistent estimate of the asymptotic variance of $\sqrt{n}(\hat{\theta} - \theta)$, where $\theta$ is a vector of parameters capturing the conditional variances, $\sigma_t$, of the RealGARCH model (2.53) and $\hat{\theta}$ is the maximum likelihood estimator of $\theta$. Furthermore, $X = [X_1, X_2, \ldots, X_m]^T$ where,
2.3. Dependence Structures

\[ X_k = -\frac{1}{T} \sum_{i=k+1}^{T} \frac{\Delta \theta \sigma_t}{\sigma_t} (z_i^2 - 1) \]  

(2.71)

and \( \Delta \theta \sigma_t \) represents the derivative of \( \sigma_t \) with respect to \( \theta \). This yields the test statistic,

\[ Q_{LM}(m) = T \hat{\rho}^T \hat{V}^{-1} \hat{\rho}, \]

(2.72)

which is asymptotically distributed as \( \chi^2 \) with \( m \) degrees of freedom (Li and Mak 1994).

2.3 Dependence Structures

2.3.1 Pearson’s Linear Correlation

The easiest and most common way of measuring dependence in data is probably through pairwise linear correlations. In many cases, dependence structures are more complicated than this, but even for some of the more holistic dependence structures it is necessary to use the correlation matrix as input, which for example is the case with the Gaussian and Student’s t copulas. The linear correlation between the two random variables \( X_1 \) and \( X_2 \), calculated with the Pearson’s correlation coefficient measure, is

\[ \rho_{X_1,X_2} = \frac{\mathbb{E}[(X_1 - \mu_1)(X_2 - \mu_2)]}{\sigma_1 \sigma_2}, \]

(2.73)

where \( \mu_i = \mathbb{E}[X_i] \) and \( \sigma_i = \sqrt{\mathbb{E}[(X_i - \mu_i)^2]} \). For a sample we get

\[ r_{x_1,x_2} = \frac{\sum_{i=1}^{T} (x_{1,i} - \bar{x}_1)(x_{2,i} - \bar{x}_2)}{\sqrt{\sum_{i=1}^{T} (x_{1,i} - \bar{x}_1)^2} \sqrt{\sum_{i=1}^{T} (x_{2,i} - \bar{x}_2)^2}}, \]

(2.74)

where \( \bar{x}_i \) is the sample mean of variable \( i \). The sample correlation matrix for \( n \) variables is given by

\[ R = \begin{pmatrix}
    r_{x_1,x_1} & \cdots & r_{x_1,x_n} \\
    \vdots & \ddots & \vdots \\
    r_{x_n,x_1} & \cdots & r_{x_n,x_n}
\end{pmatrix}. \]

(2.75)
2.3.2 Kendall’s Tau Rank Correlation

For non-elliptical distributions, the linear correlation coefficients are inappropriate and sometimes even misleading. One of the best alternatives is to instead measure the non-linear correlation using Kendall’s Tau (Embrechts, Lindskog, and McNeil 2001). Kendall’s Tau is based on the concepts of concordance and discordance. A pair \((x_i, y_i)\) and \((x_j, y_j)\) is said to be concordant if \((x_i - x_j)(y_i - y_j) > 0\), which means that the direction of the change in the variable value is the same for both the \(x\) and the \(y\) variable between observation \(i\) and \(j\). The pair is said to be discordant for the opposite case, if \((x_i - x_j)(y_i - y_j) < 0\), meaning that one variable is increasing and the other one is decreasing from observation \(i\) to observation \(j\).

The sample estimation of Kendall’s Tau in the bivariate case is computed as

\[
\tau_2 = \frac{C - D}{C + D}, \tag{2.76}
\]

with \(C\) as the total number of concordant pairs and \(D\) as the total number of discordant pairs.

For the multivariate case there are two popular approaches of calculating the empirical Kendall’s Tau. The one which is in general preferable was defined by Kendall et al. in 1940 and is calculated as an average over all possible vector pairs \((X_r, X_s)\), where \(r \neq s\). The average \(\tau_n\) is often also referred to as the coefficient of agreement among the \(n \geq 2\) rankings (Genest, Neslehova, and Ghorbal 2011), and can be written as

\[
\hat{\tau}_n(X) = \frac{1}{n(n-1)} \sum_{r \neq s} \tau_2(X_r, X_s). \tag{2.77}
\]

The multivariate Kendall’s Tau can also be computed analytically in some cases, which is useful when estimating copula parameters for specific Archimedean copulas. Explicit formulas for the analytical Kendall’s Tau can be found in the subsection of each Archimedean copula, and the general formula can be found in Section 2.3.5.

2.3.3 Introduction to Copulas

An \(n\)-dimensional copula is a real function \(C\) in \(\mathbb{R}^n\) which couples multidimensional distribution functions to their marginal distributions. Hence, a copula explains the way in which random variables defined over a common probability space are connected or linked together (Sklar 1973).

Copulas are very powerful for modelling joint distributions, since they allow us to construct the multivariate distribution with only the marginal distributions and the real function \(C\). Furthermore, copulas do not require any specific assumptions regarding the marginal distributions, they allow us to decompose
any $n$-dimensional multivariate distribution into $n$ marginal distributions and a copula function.

**Copula Definition.** An $n$-dimensional copula (briefly, an $n$-copula) is a real valued function $C$ with domain $[0, 1]^n$ and range $[0, 1]$, which satisfies the following conditions:

1. $C(1, \ldots, 1, x_m, 1, \ldots, 1) = x_m$ for each $m \leq n$ and all $x_m \in [0, 1]$.
2. $C(x_1, \ldots, x_n) = 0$ if $x_m = 0$ for any $m \leq n$.
3. $C$ is $n$-increasing\(^{11}\), meaning that for each $n$-box $B = [a_1, b_1] \times \cdots \times [a_n, b_n]$ which is a subset of $[0, 1]^n$, we have that
   $$0 \leq \sum_{(c_1, \ldots, c_n) \in \times_{j=1}^n (a_j, b_j)} (-1)^{|\{j : c_j = a_j\}|} C(c_1, \ldots, c_n) \leq 1$$

The usefulness of the copula function is further clarified in Sklar’s Theorem:

**Sklar’s Theorem.** For $n \geq 2$, let $G$ be an $n$-dimensional distribution function with marginal distributions $F_1, \ldots, F_n$. Then there exists an $n$-copula $C$ such that

$$G(x_1, \ldots, x_n) = C(F_1(x_1), \ldots, F_n(x_n))$$

for all $n$-tuples $(x_1, \ldots, x_n)$ in $(-\infty, +\infty)^n$. Conversely, let $C$ be an $n$-copula and $(F_1, \ldots, F_n)$ an $n$-tuple of one-dimensional distribution functions. Let an $n$-place real valued function $G$ be defined via (2.78). Then $G$ is an $n$-dimensional distribution function with marginal distributions $F_1, \ldots, F_n$.

### 2.3.4 Elliptical Copulas

**The Gaussian Copula**

The Gaussian copula is constructed from a multivariate normal distribution. It is a symmetric dependence structure that does not take tail dependencies into account. The Gaussian copula function $C_{G_R}^{\Phi}$ is given by (Hult, Lindskog, Hammarlid, and Rehn 2012)

$$C_{G_R}^{\Phi}(u) = \Phi_R^n(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_n)),$$

where $R \in \mathbb{R}^{n \times n}$ is the linear correlation matrix, $\Phi^{-1}$ is the quantile function of a univariate standard normal distribution, and $\Phi_R^n$ is the multivariate cumulative normal distribution function with mean vector zero and covariance matrix equal to the correlation matrix $R$. Thus, the Gaussian copula function is given by

\(^{11}\)For a more general definition of an $n$-increasing function, see e.g. (Embrechts, Lindskog, and McNeil 2001).
The Student’s t Copula

The Student’s t copula is constructed from a multivariate Student’s t distribution. It also has a symmetric dependence structure, but unlike the Gaussian copula it takes both left and right tail dependencies into account. The Student’s t copula function $C_{\nu, R}^t(u)$ is given by (Hult, Lindskog, Hammarlid, and Rehn 2012)

$$C_{\nu, R}^t(u) = t_{\nu, R}^n(t_{\nu}^{-1}(u_1), \ldots, t_{\nu}^{-1}(u_n)), \quad (2.81)$$

where $R \in \mathbb{R}^{n \times n}$ is the linear correlation matrix, $t_{\nu}^{-1}$ is the quantile function of a univariate Student’s t distribution with $\nu > 0$ degrees of freedom and $\nu > 0$ is selected through for example maximum likelihood estimation (see Section 2.3.6 for details about the parameter estimation procedure). Here, $t_{\nu}^n$ is the multivariate cumulative Student’s t distribution function with mean vector zero and the positive-definite dispersion (or scatter) matrix equal to the correlation matrix $R$. The Student’s t copula function is given by (Demarta and McNeil 2004)

$$C_{\nu, R}^t(u) = \int_{t_{\nu}^{-1}(u_1)}^{-\infty} \cdots \int_{t_{\nu}^{-1}(u_n)}^{-\infty} \frac{\Gamma\left(\frac{\nu+n}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{(\nu\pi)^n|R|}} \left(1 + \frac{x^T R^{-1} x}{\nu}\right)^{-\frac{\nu+n}{2}} \, dx. \quad (2.82)$$
2.3. Dependence Structures

2.3.5 Archimedean Copulas

The Archimedean copula family consists of copulas that are constructed from a specific class of generating functions, referred to as $\Psi$. The Archimedean generating function is a non-increasing and continuous function $\Psi : [0, \infty) \to [0, 1]$ which satisfies the conditions $\Psi(0) = 1$ and $\lim_{x \to \infty} \Psi(x) = 0$ and is strictly decreasing on $[0, \inf\{x : \Psi(x) = 0\})$. An $n$-dimensional copula $C$ is called Archimedean if it permits the representation

$$C(u_1, \ldots, u_n) = \Psi(\Psi^{-1}(u_1) + \cdots + \Psi^{-1}(u_n)), \quad (2.83)$$

where $(u_1, \ldots, u_n) \in [0, 1]^n$ for some Archimedean generator $\Psi$ with inverse $\Psi^{-1} : (0, 1) \to [0, \infty)$ (McNeil and Neslehova 2009).

Up until now, most of the research on Archimedean copulas, as well as on other kinds of copulas, is conducted for the bivariate case. In fact, the conditions under which a generator function $\Psi$ defines a $n$-dimensional copula were clarified as late as 2009. The previously defined conditions were all sufficient (such as complete monotonicity and derivatives up to order $n$ which alternate in sign), but they were not necessary.

What was shown in 2009 was that the necessary and sufficient condition is that $\Psi$ should have the analytical property known as $n$-monotonicity\(^{12}\) (McNeil and Neslehova 2009). This insight led to the possibility of exploring a rich variety of new Archimedean copulas, but unfortunately, there is still a limited amount of multivariate Archimedean copulas that are well-developed. However, three of

\(^{12}\)Meaning that $\Psi$ is continuous on $[0, \infty)$, has derivatives that satisfy $(-1)^k \frac{d^k}{dt^k} \Psi(t) \geq 0$ for all $k \in \{0, \ldots, d - 2\}$, $t \in (0, \infty)$ and $(-1)^{d-2} \frac{d^{d-2}}{dt^{d-2}} \Psi(t)$ is decreasing and convex on $(0, \infty)$.
the most commonly used Archimedean copulas, namely the Clayton, Gumbel-Hougaard and Frank copulas (see e.g. (Frees and Valdez 1997), (Embrechts, Lindskog, and McNeil 2001), (Nelsen 2003) and (Hofert, Mächler, and McNeil 2012a)), have been extended from the bivariate to the multivariate case.

One major advantage of Archimedean copulas, in contrast to elliptical copulas, is that they can capture different tail dependencies (i.e. the behavior in the left tail can differ from the behavior in the right tail). Furthermore, they are given explicitly in terms of the generating function and its inverse, and are relatively easy to construct. However, a potential disadvantage is that the exact same dependence structure is assumed between all of the marginal distributions (in comparison to for example the Gaussian or Student’s t copula, where the correlation matrix allows for different dependencies between the variables). The functional symmetry in the Archimedean copulas is referred to as exchangeability, and it is seen as a drawback particularly when considering high-dimensional portfolios. Also, most often the Archimedean copulas only have one parameter that can be optimized to make the copula fit the data in the best possible way, and for a large number of variables, that is not necessarily enough to capture the dependencies. To get around this, one may create nested Archimedean copulas, where the copulas can be nested within each other under certain conditions (Hofert and Mächler 2011). However, nested copulas will not be considered in this thesis due to the limited time available.

**Kendall’s Tau for Archimedean Copulas**

In the general case, Kendall’s Tau can be analytically computed as an $n$-integral of the copula function. However, this integral is usually quite complicated to compute. In the special case of Archimedean copulas, Kendall’s Tau can be computed with an expression containing only a one-dimensional integral

$$
\tau_n(\theta) = \frac{1}{2^{n-1} - 1} \left( -1 + 2^n \frac{(-1)^n}{(n-1)!} \int_0^{\Psi^{(1)}} u^{n-1} \Psi(u) \Psi^{(n)}(u) du \right),
$$

where $\Psi^{(n)}$ is the $n^{th}$ derivative of $\Psi$. This makes it possible to analytically derive explicit expressions for the multivariate Kendall’s Tau, $\tau_n(\theta)$, for a few special cases of Archimedean copulas. As soon as the analytical expressions are obtained, we may use the empirical estimations $\hat{\tau}_n$ (see formula (2.77)) and let $\tau_n(\theta) \approx \hat{\tau}_n$. Thereafter, the copula parameter $\theta$ may be calculated either through $\hat{\theta} = \tau_n^{-1}(\hat{\tau}_n)$ or numerical root finding (Genest, Neslehova, and Ghorbal 2011).

**The Clayton Copula**

The Clayton copula is constructed from the Laplace transform of a Gamma$(1/\theta, 1)$ distribution, and has generating function $\Psi_\theta(u) = (1 + u)^{-1/\theta}$. Through applying for example l’Hôpital’s rule it can be shown that the Clayton copula has lower tail dependence, but not upper tail dependence (Hult,
Lindskog, Hammarlid, and Rehn 2012). For $\theta > 0$ the generator is completely monotone, and the $n$-dimensional Clayton copula is then given by

$$C_{\theta}^{Cl}(u) = (u_1^{-\theta} + \cdots + u_n^{-\theta} - n + 1)^{-1/\theta}.$$  \hspace{1cm} (2.85)

Kendall’s Tau, $\tau_n(\theta)$, for the Clayton copula can be shown to be (Genest, Neslehova, and Ghorbal 2011)

$$\tau_n^{Cl}(\theta) = \frac{1}{2n-1} \left( -1 + 2^{n-1} \prod_{p=0}^{n-1} \left( \frac{1 + p\theta}{2 + p\theta} \right) \right).$$  \hspace{1cm} (2.86)

\textbf{The Gumbel-Hougaard Copula}

The Gumbel-Hougaard (also known as the logistic) copula is one of the oldest multivariate copulas. For the Gumbel-Hougaard copula we need that $\theta \geq 1$, and it can be shown that it has upper tail dependence (Hult, Lindskog, Hammarlid, and Rehn 2012). With generating function $\Psi_\theta(u) = \exp(-u^{1/\theta})$, the $n$-dimensional Gumbel-Hougaard copula is given by

$$C_{\theta}^{G}(u) = \exp \left( - \left( (-\log u_1)^\theta + \cdots + (-\log u_n)^\theta \right)^{1/\theta} \right).$$  \hspace{1cm} (2.87)

Kendall’s Tau, $\tau_n(\theta)$, for the Gumbel-Hougaard copula can be shown to be (Genest, Neslehova, and Ghorbal 2011)
\[ \tau_n^{Gu}(\theta) = \frac{1}{2^{n-1} - 1} \left( -1 + 2^{n-1} \sum C_{m_1,\ldots,m_n} \frac{(m-1)!}{(n-1)!} \left( \frac{1}{2\theta} \right)^{m-1} \prod_{i=1}^{n} \left( \frac{\Gamma(i-1/\theta)}{\Gamma(1-1/\theta)} \right)^{m_i} \right), \]  

(2.88)

where \( m \) and all of the possible combinations of \( m_i;\)'s satisfy the two conditions: (1) \( m = m_1 + \cdots + m_n \) and (2) \( m_1 + 2m_2 + \cdots + nm_n = n \). The sum in formula (2.88) is calculated over all possible \( n \)-tuples \((m_1,\ldots,m_n)\), so that the second condition is fulfilled. Note that \( n = 5 \) in our applications, which gives us 7 different combinations of the 5-tuples \((m_1,\ldots,m_5)\). Finally,

\[ C_{m_1,\ldots,m_n} = \frac{n!}{m_1! \cdots m_n!} \frac{1}{(1)!^{m_1} \cdots (n)!^{m_n}}. \]  

(2.89)

\[ \text{Figure 2.5: Level curves and density plot of a Gumbel-Hougaard copula with marginal distributions } \sim N(0,1) \text{ and } \theta = 2 \]

The Frank Copula

Unlike the Clayton and Gumbel-Hougaard copulas, the Frank copula has neither upper nor lower tail dependence. With generating function \( \Psi_\theta(u) = -\frac{1}{\theta} \log(1 + e^{-\theta u}(e^{-\theta} - 1)) \) and \( \theta > 0 \) (necessary for \( n \geq 3 \)), the Frank copula is given by

\[ C_\theta^F(u) = -\frac{1}{\theta} \log \left( 1 + \prod_{i=1}^{n} \frac{(e^{-\theta u_i} - 1)}{(e^{-\theta} - 1)^{n-1}} \right). \]  

(2.90)
2.3. Dependence Structures

In the bivariate case, Kendall’s Tau, $\tau_n(\theta)$, for the Frank copula can be shown to be (Embrechts, Lindskog, and McNeil 2001)

$$\tau_F^2(\theta) = 1 - \frac{4(1 - D_1(\theta))}{\theta}, \quad (2.91)$$

where $D_1(x)$ is the Debye function of order one, given by

$$D_1(x) = \frac{1}{x} \int_0^x \frac{t}{e^t - 1} \, dt. \quad (2.92)$$

However, in the multivariate case, there does not yet seem to be an explicit expression of $\tau_F^n(\theta)$.

![Figure 2.6: Level curves and density plot of a Frank copula with marginal distributions $\sim N(0, 1)$ and $\theta = 2$](image)

2.3.6 Parameter Estimation

The maximum pseudo-likelihood\[14\] estimator is an extension of the standard Inference Functions for Margins (IFM) method, but with optimization of non-parametric pseudo-likelihoods. It is seen as one of the best and most efficient approaches to determining parameter values, even though it is still quite numerically complicated and demanding regarding computational power and run-time. The main advantage with the pseudo-likelihood method is that no marginal distributions are assumed. In the standard IFM method, the choices of marginal distributions are crucial for the accuracy of the method, and a slight

---

\[14\]The pseudo-likelihood is throughout this thesis defined as the likelihood of observations from the empirical counterpart of the distribution function. Pseudo-observations, that are mentioned in the model validation, Section 2.3.7, are thus observations from the empirical distribution as opposed to observations from an assumed true parametric distribution.
misspecification in marginal distributions may have decisive impact on the estimation of the dependence parameter (Genest, Rémillard, and Beaudoin 2009). The log-pseudo-likelihood $L(\theta)$ to be maximized is

$$L(\theta) = \sum_{i=1}^{T} \log \left( c_{\theta} \{ \hat{F}_1(X_{i1}), \ldots, \hat{F}_n(X_{in}) \} \right),$$  \hspace{1cm} (2.93)$$

where $c_{\theta}$ denotes the density of copula $C_{\theta}$ (assuming it exists), and the marginal distribution $F_j$ is replaced by its empirical counterpart for the pseudo-observations

$$\hat{F}_j(t) = \frac{1}{T-1} \sum_{i=1}^{T} 1(X_{ij} \leq t).$$  \hspace{1cm} (2.94)$$

The densities of the three Archimedean copulas used in this thesis are$^{15}$

**Clayton copula density**

$$c_{\theta}^{C_{\theta}}(u) = \prod_{k=0}^{n-1} (\theta k + 1) \left( \prod_{j=1}^{n} u_j \right)^{-(1+\theta)} \left( 1 + \sum_{j=1}^{n} \Psi^{-1}(u_j) \right)^{-(n+1/\theta)},$$  \hspace{1cm} (2.95)$$

where $\Psi^{-1}(u) = u^{-\theta} - 1$.

**Gumbel-Hougaard copula density**

$$c_{\theta}^{G_{\theta}} = \theta^n \exp \left( - \left( \sum_{j=1}^{n} \Psi^{-1}(u_j) \right)^{1/\theta} \right) \frac{\prod_{j=1}^{n} (-\log u_j)^{\theta-1}}{\sum_{j=1}^{n} \Psi^{-1}(u_j)} \prod_{j=1}^{n} u_j \cdot P_{n,1/\theta}^{G} \left( \left( \sum_{j=1}^{n} \Psi^{-1}(u_j) \right)^{1/\theta} \right),$$  \hspace{1cm} (2.96)$$

For the Gumbel-Hougaard density we have $\Psi^{-1}(u) = -(\log u)^{\theta}$,

$$P_{n,1/\theta}^{G}(x) = \sum_{k=1}^{n} a_{n,k}^{G}(1/\theta) x^k$$ \text{and} \quad a_{n,k}^{G}(1/\theta) = (-1)^{n-k} \sum_{j=k}^{n} (1/\theta)^j s(n,j) S(j,k).$$

$s$ and $S$ denote the Stirling numbers of the first and second kind, respectively.

$^{15}$For more details about the density functions, please see article by Hofert, Mächler and McNeil (Hofert, Mächler, and McNeil 2012b).
2.3. Dependence Structures

Frank copula density

\[ c_F^{\theta} = \left( \frac{\theta}{1 - e^{-\theta}} \right)^{n-1} \left( \exp\left( -\theta \sum_{j=1}^{n} u_j \right) \right) \left( \frac{1}{h_F^{\theta}(u)} \right) \right) \text{Li}_{n-1}(h_F^{\theta}(u)) \] (2.97)

where \( h_F^{\theta}(u) = (1 - e^{-\theta})^{1-n} \prod_{j=1}^{n} (1 - \exp(-\theta u_j)) \) and \( \text{Li}_s(z) = \sum_{k=1}^{\infty} \frac{z^k}{k^s} \).

2.3.7 Model Validation

Goodness of fit tests for multivariate copulas is still an area under development. There are many different - more or less complicated - ways of computing the fit of the models. As summarized by Genest et al. in 2009, there are broadly three groups of tests (Genest, Rémillard, and Beaudoin 2009):

1. Procedures that are developed for specific dependence structures, such as the Gaussian copula or the Clayton family.

2. Statistics applicable on any class of copulas, but whose implementation involves either arbitrary parameters, smoothing parameters (e.g. kernels or weight functions) or ad hoc categorization of the data in order to apply an analogue of the standard chi-squared test.

3. Tests that can be used to all copula structures and that require no strategic choice for their use, referred to as "Blanket Tests".

In order to be consistent, we prefer to use the same test for goodness of fit evaluation for all of the copulas. Since the first category is limited to only specific classes of copulas and the second category includes arbitrary parameters or assumptions regarding the data, we here prefer the third category.

The null hypothesis which is to be validated is \( H_0 : C \in C_0 \), for some class \( C_0 \) of copulas and the actual multivariate dependence structure \( C \). Since the actual multivariate dependence structure is not known, we need to estimate it. The best way of doing this is probably through the "empirical copula" \( C_T(u) \), which is an entirely nonparametric and consistent estimator of the underlying copula \( C \) (Genest, Rémillard, and Beaudoin 2009). \( C_T(u) \) is given by

\[ C_T(u) = \frac{1}{T} \sum_{i=1}^{T} 1(\hat{U}_{i1} \leq u_1, \ldots, \hat{U}_{in} \leq u_n), \quad u = (u_1, \ldots, u_n) \in [0, 1]^n, \] (2.98)

where \((\hat{U}_{i1}, \ldots, \hat{U}_{in})\) are referred to as pseudo-observations, defined as

\[ (\hat{U}_{i1}, \ldots, \hat{U}_{in}) = (\hat{F}_1(X_{i1}), \ldots, \hat{F}_n(X_{in})). \] (2.99)

Here, \( \hat{F}_j \) is the empirical distribution function for the random variable \( X_j \), as defined in formula (2.94). Some tests then emerge from the empirical goodness of fit process.
Chapter 2. Theory

\[ C_T = \sqrt{T}(C_T - C_{\theta_n}), \tag{2.100} \]

which uses the difference between the empirical copula \( C_T \) and the parametric estimation \( C_{\theta_n} \) of the true copula \( C \) under the hypothesis \( H_0 \). Based on this goodness of fit process, the test statistic recommended by (Genest, Rémillard, and Beaudoin 2009), after their thorough review of the current literature, is the Cramér–von Mises statistic

\[ S_T = \int_{[0,1]^n} C_T(u)^2 dC_T(u), \tag{2.101} \]

where large values of \( S_T \) lead to rejection of \( H_0 \). The tests based on \( S_T \) are consistent, so if \( C \notin C_0 \), then \( H_0 \) is rejected with probability 1 as \( T \to \infty \). The asymptotic distribution of \( S_T \) depend both on the family of copulas and on the unknown parameter value \( \theta_n \). As a result, in order to compute approximate \( p \)-values for the test, it is necessary to implement a specially adapted Monte Carlo method (Genest, Rémillard, and Beaudoin 2009).
Chapter 3

Data

In this chapter we briefly describe the nature of the available data as well as the modifications conducted in order to create data series appropriate for modelling. For handling and modifying the raw data, the software program Matlab was used.

3.1 Description

The five currency exchange rates that are used for modelling in this thesis are: EUR/SEK, EUR/NOK, EUR/USD, USD/SEK and USD/NOK. The total period considered is one month, and all of the five currency pairs cover the same month so that the dependencies between the currencies can be modelled correctly. Notable about the five currency pairs is that they certainly are dependent, since a combination of for example EUR/SEK and EUR/USD (short position) will lead to USD/SEK. Therefore it is expected to find strong dependencies in the data.

In the initial data set, each observation contained a time stamp and the best bid and offer price for that specific point in time. An example of what the data looked like is given below.

<table>
<thead>
<tr>
<th>Date</th>
<th>Time stamp</th>
<th>Best bid</th>
<th>Best offer</th>
</tr>
</thead>
<tbody>
<tr>
<td>2016.01.01</td>
<td>00:00:01</td>
<td>9.7801</td>
<td>9.8903</td>
</tr>
<tr>
<td>2016.01.01</td>
<td>00:00:02</td>
<td>9.7904</td>
<td>NaN</td>
</tr>
<tr>
<td>2016.01.01</td>
<td>00:00:04</td>
<td>9.7800</td>
<td>9.8901</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3.1: Example of raw data

Note that due to confidentiality reasons, table 3.1 is only an illustration of what the actual data looked like, and not an extraction from the original data set. As can be seen in the table, each observation covers a specific second. However, not all seconds are represented in the data (note that the third second is missing
in table 3.1), and there were some gaps of several seconds without any observations in the original data set. Furthermore, for some of the seconds that were represented in the data, either the best bid or the best offer was missing. Consequently, all of the missing seconds and the missing bids and offers had to be properly handled.

3.2 Data Modifications

The missing values in our data set were simply assumed to be equal to the most recent quoted prices, so that a missing value for one of the exchange rates was set equal to the most recent given value for that exchange rate. The reason why the most recent quoted price was used as an approximation of the current price, instead of using e.g. interpolation between times \( t - 1 \) and \( t + 1 \) to approximate the price at time \( t \), was that an interpolation or average of prices would have included information which would not yet have been known at time \( t \). Also, in real life, the price which would have been valid at time \( t \), given that there is no price update at time \( t \), is the price from time \( t - 1 \). After filling in the missing information by using the most recent values, the data set was complete with data points for every second and corresponding best bid and offer prices. Correcting table 3.1 in line with this approach, we receive the following complete table:

<table>
<thead>
<tr>
<th>Date</th>
<th>Time stamp</th>
<th>Best bid</th>
<th>Best offer</th>
</tr>
</thead>
<tbody>
<tr>
<td>2016.01.01</td>
<td>00:00:01</td>
<td>9.7801</td>
<td>9.8903</td>
</tr>
<tr>
<td>2016.01.01</td>
<td>00:00:02</td>
<td>9.7904</td>
<td>9.8903</td>
</tr>
<tr>
<td>2016.01.01</td>
<td>00:00:03</td>
<td>9.7904</td>
<td>9.8903</td>
</tr>
<tr>
<td>2016.01.01</td>
<td>00:00:04</td>
<td>9.7800</td>
<td>9.8901</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

**Table 3.2: Example of adjusted data**

Since the focus in this thesis is on the currency exchange rate movements in general, and not on the specific movements of bids and offers in particular, a univariate price vector was created as a simple average of the best bid and offer for each point in time

\[
p_t = \frac{p_{\text{bid},t} + p_{\text{offer},t}}{2}.
\]

Thereafter, to better be able to compare the exchange rate fluctuations, the log-returns of the series were computed. By looking at returns rather than prices, the different series can more easily be set in relation to each other. By transforming into log-returns we also eliminate most of the random walk behavior in the
3.3 Final Data

price series (we get a relatively stable mean value about 0 over time, instead of a highly variable mean value, which can be seen when comparing Figure 3.1 and Figure 3.2 (A)), which is necessary in order to create a series appropriate for further modelling. The log-returns were computes as

$$r_t = \log \left( \frac{p_t}{p_{t-1}} \right) = \log(p_t) - \log(p_{t-1})$$ (3.2)

One problem with the data set, after filling in the missing information and calculating the log-returns, was that it contained long periods consisting of only zeroes. When the Nordic stock exchanges are closed, there are several minutes where there is no change in the currency exchange rates for SEK and NOK. This creates problems in the modelling, since the realized volatility becomes 0 for many minutes. Specifically, it creates a non-invertible Hessian matrix, and hence the estimation of parameters cannot continue. In order to avoid this problem, we limited our data set to only contain trading hours for the Swedish stock exchange, which is 09.00-17.30 (Central European Time). After this modification, almost all of the zeroes in the log-returns were eliminated. However, there were approximately 10 places in total for SEK and NOK where the realized volatility was still 0. In order to proceed with the modelling, we assumed that the realized volatility in these cases was equal to the realized volatility from the minute before. Note, however, that this approach would probably not be optimal in the case of a larger number of zeroes in the realized volatility series. Here, we only had approximately 10 zeroes out of in total 51,000 realized volatility observations for all exchange rates.

For model estimation, we only use the three first weeks of the modified data set, consisting of 7650 minute by minute observations. Thereafter, for testing our results, we use the last week of the data set consisting of 2550 minute by minute observations.

3.3 Final Data

![USD/SEK Exchange Rate](image)

**Figure 3.1**: USD/SEK exchange rate, one month, 09.00-17.30
In figure 3.1 we see the plotted USD/SEK exchange rate after data modifications. The plot shows the whole period of four weeks with second by second observations. The dotted vertical lines show the breaks between the days. Note that the log-returns are computed within each trading day, meaning that the breaks between two trading days are not included in the log-return and volatility series.

In figure 3.2 we see the final series of minute by minute log-returns and realized volatility for USD/SEK. It is notable that there are some peaks both in returns and in volatility, and tendencies to clustering. However, we see that the realized volatility is still rather stable and not as fluctuating as it would be in case of a squared returns measure.
Chapter 4

Results

In this chapter we present the main results of the thesis. Note that all exchange rates are not presented in all parts of the results. The corresponding results for the exchange rates that are not presented in this chapter can be found in Appendix. For the modeling and execution we mainly used the software program R. Several packages were used throughout the implementation, however, the most essential packages were ‘rugarch’ and ‘copula’.

4.1 Time Series Models

4.1.1 ARIMA-RealGARCH

In order to model the currency exchange rate log-returns, ARIMA-RealGARCH models were used. The best models were selected with a combination of AIC/BIC (BIC in case of discrepancies between the tests), JB-tests, LB-tests, ACF and qq-plots. ARIMA models from ARIMA(0, 1, 0) up to ARIMA(5, 1, 5) were tested and RealGARCH was set to RealGARCH(1, 1). Three different distribution functions were considered for the standardized residuals; Normal, Student’s t and GED. This gave us in total 108 possible model specifications for each of the exchange rates. A parsimonious approach has been used throughout the entire model selection, meaning that models with fewer parameters were preferred over more complex models, other aspects being approximately equal. The parsimonious approach decreases the risk of over-fitting the model to the in-sample data, which is otherwise a risk that could lead to promising in-sample results but poor out-of-sample results.

An example of the BIC matrix for USD/SEK, used for model selection, is presented below.
Table 4.1: USD/SEK - BIC matrix with RealGARCH(1,1)

<table>
<thead>
<tr>
<th>(p,q)</th>
<th>(p,0)</th>
<th>(p,1)</th>
<th>(p,2)</th>
<th>(p,3)</th>
<th>(p,4)</th>
<th>(p,5)</th>
</tr>
</thead>
</table>

Even though there is not a large difference in the BIC-values, it is still possible to distinguish some models that perform better than others. Here, the smallest values are the best, and the lowest BIC-value in the matrix is -13.3534, given by an ARIMA(0,1,2) model. In combination with other evaluation criteria like the ACF and qq-plots, the model is selected. A comparison of the ACF before applying a time series model and after applying a time series model can be seen below.

On the left hand side in Figure 4.1, we see the ACF for the actual observations, before model estimation. Note that the observation series is not necessarily covariance stationary, which means that the left ACF should not be interpreted as the exact true autocorrelations, since they might change over time. However, we see a tendency of time dependence, especially for the first lags, which suggests that it would be appropriate to use time series modelling. On the right hand side, we see the ACF for the standardized residuals, after applying an ARIMA(0,1,2) model. Indeed, the ACF graph of the standardized residuals shows that the time series model makes a significant difference and accounts for most of the dependence within the series.
### 4.1. Time Series Models

<table>
<thead>
<tr>
<th></th>
<th>$\psi_1$</th>
<th>$\psi_2$</th>
<th>$d$</th>
<th>$\theta_1$</th>
<th>$\theta_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>EUR/SEK</td>
<td>0.4219</td>
<td>-</td>
<td>1</td>
<td>-0.4554</td>
<td>-</td>
</tr>
<tr>
<td>EUR/NOK</td>
<td>0.5659</td>
<td>0.0202</td>
<td>1</td>
<td>-0.6480</td>
<td>-</td>
</tr>
<tr>
<td>EUR/USD</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>USD/SEK</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>-0.0600</td>
<td>-0.0371</td>
</tr>
<tr>
<td>USD/NOK</td>
<td>0.3018</td>
<td>-</td>
<td>1</td>
<td>-0.3682</td>
<td>-</td>
</tr>
</tbody>
</table>

**Table 4.2: Estimated best fit ARIMA-parameters**

In Table 4.2, the order of all ARIMA models as well as the parameter values are presented\(^1\). For example, the differentiated USD/SEK log-returns are modelled with an ARMA(0, 2) model, while the EUR/USD log-returns are modelled without any ARMA part. As can be seen, the parameter values are in general small. However, all of the parameters were highly significant. For example, the moving average parameters in the USD/SEK model mean that only about 6% and 4% of $\epsilon_t$ and $\epsilon_{t-1}$, respectively, is used to create the forecast for time $t+1$. Also, the positive $\psi_1$ parameters are for all exchange rates counterweighted by slightly larger negative $\theta_1$ parameters, which contributes even more to limiting the effect created by the ARIMA-parameters\(^2\).

![Actual Log-returns vs ARIMA-Forecasts](image)

**Figure 4.2: Actual log-returns vs forecasts**

Figure 4.2 shows the forecasted values of EUR/USD and USD/SEK plotted against the actual out-of-sample log-returns. Since the differentiated EUR/USD log-returns are modelled without any ARMA parameters, all series forecasts are 0. Even though USD/SEK has highly significant ARMA parameter values that help to remove the autocorrelation in the series, the forecasted values are still close to 0. Thus, we see that the time series part of the modelling does not contribute much to forecasting the actual log-return outcomes. However, as stated before, the purpose of the ARIMA modelling is not to forecast exact outcomes but rather to eliminate autocorrelations in the data in order to obtain

---

\(^1\)The $p$-values for the parameter estimates can be found in Appendix A.3.

\(^2\)Note that in some cases the opposite signs of the parameters can amplify the effect on the forecasts. However, in most cases the sign difference creates a counterbalance and thus limits the effect on the forecasts.
white-noise residuals for further density modelling and ES calculations. Still, the ARIMA forecasts will be used, but only as mean estimations $\mu_t$ for each point in time.

<table>
<thead>
<tr>
<th></th>
<th>$\omega$</th>
<th>$\alpha_1$</th>
<th>$\beta_1$</th>
<th>$\eta_1$</th>
<th>$\eta_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>EUR/SEK</td>
<td>-4.8525</td>
<td>0.1948</td>
<td>0.6143</td>
<td>0.0064</td>
<td>0.0908</td>
</tr>
<tr>
<td>EUR/NOK</td>
<td>-4.0320</td>
<td>0.2223</td>
<td>0.6379</td>
<td>0.0257</td>
<td>0.0271</td>
</tr>
<tr>
<td>EUR/USD</td>
<td>0.1836</td>
<td>0.6264</td>
<td>0.6863</td>
<td>0.0014</td>
<td>0.0575</td>
</tr>
<tr>
<td>USD/SEK</td>
<td>-1.0530</td>
<td>0.6185</td>
<td>0.6193</td>
<td>-0.0050</td>
<td>0.0729</td>
</tr>
<tr>
<td>USD/NOK</td>
<td>-1.7217</td>
<td>0.4935</td>
<td>0.6396</td>
<td>0.0235</td>
<td>0.0210</td>
</tr>
</tbody>
</table>

Table 4.3: Estimated best fit log RealGARCH-parameters

In table 4.3, some of the parameters from the log RealGARCH models are presented. Here, $\omega$ is the constant part of the conditional variance, $\alpha_1$ is the parameter for the first order autoregressive part of the conditional variance and $\beta_1$ is the parameter for the first lag of realized volatility. $\eta_1$ and $\eta_2$ are part of the leverage function, capturing the dependence between the returns and future volatility.

<table>
<thead>
<tr>
<th></th>
<th>$\xi$</th>
<th>$\delta$</th>
<th>$\lambda$</th>
<th>$\nu/\kappa$</th>
</tr>
</thead>
<tbody>
<tr>
<td>EUR/SEK</td>
<td>10.0000</td>
<td>1.1073</td>
<td>0.5114</td>
<td>0.7537</td>
</tr>
<tr>
<td>EUR/NOK</td>
<td>10.0000</td>
<td>1.1301</td>
<td>0.4551</td>
<td>0.8552</td>
</tr>
<tr>
<td>EUR/USD</td>
<td>-0.8817</td>
<td>0.4662</td>
<td>0.2502</td>
<td>6.0879</td>
</tr>
<tr>
<td>USD/SEK</td>
<td>0.1873</td>
<td>0.5244</td>
<td>0.3110</td>
<td>1.2552</td>
</tr>
<tr>
<td>USD/NOK</td>
<td>1.1740</td>
<td>0.5880</td>
<td>0.3827</td>
<td>1.0561</td>
</tr>
</tbody>
</table>

Table 4.4: Estimated best fit log RealGARCH-parameters

In table 4.4, we see the rest of the parameters for the final models. $\xi$, $\delta$ and $\lambda$ are part of the log RealGARCH models\(^3\), and $\nu/\kappa$ is the parameter value for the standardized residuals $z_t$ (the selected residual distributions are presented in table 4.7). By calculating the persistence ($P = \beta_1 + \delta \alpha_1$), we see that each model has a persistence between 0 and 1. The exact numbers of the persistences are \{EUR/SEK, EUR/NOK, EUR/USD, USD/SEK, USD/NOK\} = \{0.8300, 0.8891, 0.9783, 0.9436, 0.9298\}.

\(^3\)\(\xi\) is approximately 9.99998540 for EUR/SEK and 9.99999192 for EUR/NOK
In table 4.5 we see the \( p \)-values from the Li-Mak-test on the standardized residuals for 3 different lags. As can be seen, all of the ARCH-effects are sufficiently being accounted for by the models for all of the exchange rates except EUR/SEK. The \( p \)-values for EUR/SEK indicate that the ARCH-effects have not been accounted for sufficiently by the chosen GARCH model. To correct for this we tried to fit RealGARCH(\( \alpha, \beta \)) models up to \( \alpha = 5 \) and \( \beta = 5 \) (even though this was not in the scope of the thesis), which resulted in small improvements in the Li-Mak \( p \)-values. However, regarding the parameter \( p \)-values, all \( \alpha_i \) and \( \beta_j \) for \( i, j > 1 \) showed no significance. Hence, we decided to model EUR/SEK with RealGARCH(1, 1) since it is in line with our parsimonious approach, and over-fitting could lead to worse results on out-of-sample data.

Two of the highly significant RealGARCH-models are illustrated in Figure 4.3. It is notable that the forecasted realized volatility follows the actual realized volatility remarkably well for both EUR/USD and USD/SEK. However, for EUR/SEK and EUR/NOK the patterns in the volatility are not as clear, and thus the movements in the realized volatility are not as well captured by the RealGARCH-models.
In Figure 4.4 we see the less significant RealGARCH-models plotted against the actual realized volatility. Even though the RealGARCH model performs somewhat better than assuming constant volatility, it is not a major difference.

4.1.2 Residual Diagnostics

As distributions for the standardized residuals, we evaluated the performance of the Normal, Standardized Student’s t and Standardized General Error Distribution. Firstly, the performance of the Normal distribution was tested.

<table>
<thead>
<tr>
<th>Exchange Rate</th>
<th>Q-stat</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>EUR/SEK</td>
<td>6407.3</td>
<td>0.00000</td>
</tr>
<tr>
<td>EUR/NOK</td>
<td>122200.7</td>
<td>0.00000</td>
</tr>
<tr>
<td>EUR/USD</td>
<td>1666.1</td>
<td>0.00000</td>
</tr>
<tr>
<td>USD/SEK</td>
<td>951.0</td>
<td>0.00000</td>
</tr>
<tr>
<td>USD/NOK</td>
<td>158483.6</td>
<td>0.00000</td>
</tr>
</tbody>
</table>

Table 4.6: Jarque-Berra test for Normal Distribution

As can be seen in Table 4.6, the Normal distribution could easily be rejected for all of the exchange rates due to heavy tails and excess kurtosis in the standardized residuals\(^4\). Thereafter Student’s t and GED were tested. In order to select between Student’s t and GED, we primarily used the BIC statistics and QQ-plots. Thanks to its flexibility, GED gave the best results for all of the exchange rates except EUR/USD, where Student’s t performed better.

\(^4\)The presented values in the table are from the best models with Normally distributed standardized residuals.
In Table 4.7 we see the selected residual distributions and their parameter values. Here, $\kappa$ is the shape parameter for the GED and $\nu$ is the fitted degrees of freedom for the Student’s t distribution. All distributions have mean 0 and standard deviation 1.

![Empirical Density of Standardized Residuals](image1.png)  
![GED - QQ Plot](image2.png)

**Figure 4.5: USD/SEK diagnostics of standardized residuals**

In Figure 4.5 we see the histogram of the standardized residuals on the left hand side, plotted against a fitted Normal and Generalized Error Distribution. As can be seen, the GED provides a better fit than the Normal distribution, as it better captures the tails and the peak around 0. On the right hand side, we see a QQ-plot between the residuals and the selected GED. It seems to provide a good fit overall, even though the left tails might be somewhat heavier in the actual residuals than in the theoretical distribution, and the right tails might be somewhat lighter in the actual residuals than in the theoretical distribution.
Table 4.8 shows the $Q_{LB}$ statistic and $p$-values for the joint hypothesis that there is no autocorrelation for the $h$ number of lags tested together for USD/SEK (e.g. if $h=3$, then the hypothesis is that the autocorrelation is zero for lag 1, 2 and 3). As can be seen, the $p$-values are fairly high, indicating that we cannot reject $H_0$ at the significance level of 5%.

### 4.2 Dependence Structures

After having modelled the log-returns for the individual currency exchange rates with individually fitted time series models, the next step was to add dependence structures. The standardized residuals are now the time series of interest, since these are the only stochastic parts left of the time series models. The rest of the fluctuations in the exchange rates are at this point sufficiently captured by the individual ARIMA-RealGARCH models. Even though the standardized residuals are assumed to be uncorrelated (no correlation within the series), we now seek a dependence between the standardized residuals of the different exchange rate series.

#### 4.2.1 Correlation Measures

We first calculated the pairwise linear correlations (Pearson correlation matrix) and the pairwise rank correlations (Kendall’s Tau correlation matrix). The linear correlation matrix was used as input in the Gaussian and Student’s $t$ copula, and the average empirical Kendall’s Tau was used to obtain starting values for the maximum likelihood parameter estimation for the Clayton, Gumbel-Hougaard and Frank copulas.
Looking at the linear correlations between the standardized residuals, we see that there are indeed some significant dependencies, even though time series models have been applied and we only look at the residual series. We see that overall, there are positive correlations between the series. However, as expected, the exceptions are the strong negative correlations between EUR/USD and USD/SEK and USD/NOK, which emerge mainly due to that USD is the denominator currency in EUR/USD and the numerator currency in USD/SEK and USD/NOK.

In Kendall’s Tau, Table 4.10, we see the same pattern as in the linear correlation matrix, Table 4.9. There are positive correlations between most of the exchange rates, but negative between EUR/USD and USD/SEK and USD/NOK.

The multivariate empirical Kendall’s Tau was calculated according to the “Coefficient of Agreement” method, and is the average of all possible combinations of the pairwise Kendall’s Tau. Since there are both positive and negative correlations, they somewhat cancel each other out, and the obtained coefficient of agreement was calculated to be 0.0918.

4.2.2 Graphical Analysis

Even though it is hard to illustrate and interpret the dependence in the data graphically with five variables, we still choose to plot some of the variables against each other to get a grasp of the data characteristics.
In Figure 4.6, we see EUR/SEK plotted against USD/SEK to the left and EUR/USD plotted against USD/SEK to the right. The first observation is that the left graph clearly illustrates a positive dependence while the graph to the right shows a negative dependence. Furthermore, it seems like there could be tail dependencies present, even though it is not obvious.

In Figure 4.7, we see EUR/NOK plotted against USD/NOK to the left and EUR/USD plotted against USD/NOK to the right. One outlier observation has been removed from these graphs in order to create approximately equal x- and y-axis scales, so that the graphs can be more easily interpreted. The removed observation had observation number 3571 and value -19.9016 for USD/NOK. Note that this observation is only excluded from these two graphs and is included in all other parts and calculations. The conclusions that can be drawn from the figure are similar to the conclusions from Figure 4.6. On the left hand side we see a strong positive correlation and on the right hand side we see a negative correlation. The extreme values in the graphs seem to be closer together than the observations in the middle, suggesting that there might be tail dependencies present.

---

5If the variables are tail dependent, the dependence in the tails is stronger than the dependence in the center. Thus, the observations in the corners of the graph would be closer together than the observations in the middle, if the variables were tail dependent.
4.2.3 Copulas

The copula parameters were estimated in two ways. First, parameter estimates were computed based on the analytic formulas for Kendall’s Tau. The analytic expressions were set to equal the empirical Kendall’s Tau, 0.0918, and the parameter \( \theta \) was then obtained with numerical root-finding. Thereafter, new parameter estimates were calculated by maximum likelihood estimation of the pseudo-observations, where \( \theta \) was used as a starting value for the optimization. In the case of the Student’s t copula, the parameter \( \nu_{ML} \) was calculated directly with maximum pseudo-likelihood. The pseudo-observations were in all cases calculated by a pseudo-observation transformation, similar to the empirical distribution function. Thus, the parameters were estimated based on the actual observations and not on parametric approximations of the marginal distribution functions.

<table>
<thead>
<tr>
<th>Copula</th>
<th>( \theta )</th>
<th>( \theta_{ML} )</th>
<th>( \nu_{ML} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gaussian</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Student’s t</td>
<td>-</td>
<td>-</td>
<td>2.7944</td>
</tr>
<tr>
<td>Clayton</td>
<td>0.2733</td>
<td>0.2053</td>
<td>-</td>
</tr>
<tr>
<td>Gumbel</td>
<td>1.1011</td>
<td>1.0922</td>
<td>-</td>
</tr>
<tr>
<td>Frank</td>
<td>0.8318</td>
<td>0.7805</td>
<td>-</td>
</tr>
</tbody>
</table>

**Table 4.11: Copula parameter estimations**

In Table 4.11 we see the Kendall’s Tau estimations of the copula parameters, \( \theta \), the maximum pseudo-likelihood estimations of the copula parameters, \( \theta_{ML} \), and the maximum pseudo-likelihood estimation of the degrees of freedom for the Student’s t copula, \( \nu_{ML} \). It is notable that \( \theta \) and \( \theta_{ML} \) are fairly similar for all Archimedean copulas, but in line with the conclusions from the theory chapter, we use the \( \theta_{ML} \) estimations as our final parameter values.

The underlying marginal distributions that were used in the copula modelling are the selected best-fit distributions from Table 4.7. The best fits among the evaluated copulas were given by the Clayton and Student’s t copulas. The Clayton copula performed best when all five currency exchange rates were included, and gave a Cramér-von-Mises (\( S_T \)) statistic value of 2.04. However, when EUR/USD was excluded\(^6\), the Student’s t copula performed best with a \( S_T \) statistic value of 1.37. It is also notable that the two elliptical copulas, Gaussian and Student’s t, gave the worst fit for the original portfolio but the best fit for the portfolio without EUR/USD.

---

\(^6\)Since EUR/USD is the only exchange rate in the original portfolio which can create triangulation, and also the only exchange rate with a significant negative correlation to some of the other exchange rates (USD/SEK and USD/NOK), it was excluded from the original portfolio as an experiment in order to investigate how that would affect the dependence in the new portfolio.
Chapter 4. Results

<table>
<thead>
<tr>
<th>Copula</th>
<th>$S_T$</th>
<th>$S_{T, w/o EUR/USD}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gaussian</td>
<td>2.4123</td>
<td>1.7409</td>
</tr>
<tr>
<td>Student’s t</td>
<td>2.4029</td>
<td>1.3747</td>
</tr>
<tr>
<td>Clayton</td>
<td>2.0366</td>
<td>6.0711</td>
</tr>
<tr>
<td>Gumbel</td>
<td>2.2799</td>
<td>4.5825</td>
</tr>
<tr>
<td>Frank</td>
<td>2.1964</td>
<td>3.2241</td>
</tr>
</tbody>
</table>

Table 4.12: Cramér-von-Mises statistics for copulas

In Table 4.12 we see the $S_T$ statistics for all copulas. All of the exchange rates are included in the middle column, and in the right column EUR/USD is excluded. The parameter estimations used for the copulas where EUR/USD was excluded can be found in appendix A.5.

In the following sections the results based on all five currency exchange rates will be presented, with copula parameters according to table 4.11. For additional graphs without EUR/USD and with the copula parameters based on only the four other exchange rates, see appendix.

Elliptical Copulas

The plots displayed in Figure 4.8 are the simulated EUR/SEK vs USD/SEK and EUR/USD vs USD/SEK. The simulations are conducted with the Gaussian copula, the marginal distributions of the standardized residuals and the calculated linear correlation matrix. Note the significant similarities with the plots of the actual data, Figure 4.6. However, the trends in the tails seem to differ.
4.2. Dependence Structures

The plots displayed in Figure 4.9 are the simulated EUR/SEK vs USD/SEK and EUR/USD vs USD/SEK. The simulations are conducted with the Student’s t copula, the marginal distributions of the standardized residuals, the calculated linear correlation matrix and the calculated degrees of freedom. Note the even stronger similarities with the plots of the actual data, Figure 4.6, since the Student’s t copula also takes the tail dependencies into account.

**Archimedean Copulas**

The two main differences between the Archimedean copulas and the Elliptical copulas are that Archimedean copulas can take asymmetric tail dependencies into account (that is, the dependence can differ between the left and right tail) and that the Archimedean copulas most often only have one parameter that can be adjusted to fit the actual data. There is no correlation matrix used as input to compute the copula, and hence the only differences between the copula simulations for different exchange rates with an Archimedean copula are due to different marginal distributions.

The plots displayed in Figure 4.10 are the simulated EUR/SEK vs USD/SEK and EUR/USD vs USD/SEK. The simulations are conducted with the Clayton copula, the marginal distributions of the standardized residuals and the computed parameter value $\theta$. Note that there are a few more outcomes in the lower left corner of the graph, due to the left tail dependence. However, due to the low
\( \theta \)-value, the tail dependence is not as notable in Figure 4.10 as in the illustration in the theory chapter, Figure 2.4.

(A) EUR/SEK vs USD/SEK  
(B) EUR/USD vs USD/SEK

**Figure 4.11: Gumbel-Hougaard copula with marginal distributions**

The plots displayed in Figure 4.11 are the simulated EUR/SEK vs USD/SEK and EUR/USD vs USD/SEK. The simulations are conducted with the Gumbel-Hougaard copula, the marginal distributions of the standardized residuals and the computed parameter value \( \theta \). Note that there are a few more outcomes in the upper right corner of the graph, due to the right tail dependence. As with the Clayton copula, the \( \theta \)-value is also lower for the Gumbel copula in the fitted copula case (Figure 4.11) compared to the theoretical illustration (Figure 2.5), which makes the tail dependence weaker and less visible than in the theoretical example.

(A) EUR/SEK vs USD/SEK  
(B) EUR/USD vs USD/SEK

**Figure 4.12: Frank copula with marginal distributions**

The plots displayed in Figure 4.12 are the simulated EUR/SEK vs USD/SEK and EUR/USD vs USD/SEK. The simulations are conducted with the Frank copula, the marginal distributions of the standardized residuals and the computed parameter value \( \theta \). Since the Frank copula has no tail dependence, the outcomes are centered.
4.3 Expected Shortfall

In order to calculate ES we first have to define portfolio weights. Since we are also interested in the underlying risk contributions from the different exchange rates, all weights are chosen to be equal. The risk contributions from the different exchange rates can then easily be assessed and compared. Thus, we use a weight of 20% on each of the exchange rates throughout the implementation.

Before calculating $\text{ES}_{0.05}$, we start by examining the number of necessary simulations for convergence. The final $\text{ES}_{0.05}$ calculations are presented in section 4.3.2.

4.3.1 Convergence of Simulations

For simplicity, we only look at the standardized residuals in this section, since these are the only parts of the models that need to be simulated (in the final ES calculations we of course need to integrate the predicted means as well as the predicted standard deviations, but they are calculated from historical data and not simulated). The portfolio outcome $X^z$ for the standardized residuals with equal weights is then described by $X^z = (Z^1 + \cdots + Z^5) / 5$, where $Z^i$ is the simulated outcome from the standardized residual of exchange rate $i$. After simulating and calculating $X^z$, the observations are sorted so that the worst observations can be extracted to calculate desired VaR and ES levels.

In Figure 4.13 (A) we see an example of the simulated $X^z$ outcomes from a Gaussian copula and in (B) we see the sorted outcomes, from where the worst outcomes are extracted. The following figures illustrate convergence rates of $\text{VaR}_{0.05}$ and $\text{ES}_{0.05}$ for the portfolio standardized residuals and all of the five copulas considered in the thesis.
As can be seen in Figure 4.14, both the VaR\(_{0.05}\) and the ES\(_{0.05}\) for \(X^2\) have converged after about 100,000 simulations for all copulas.

Note that the order of the copulas changes between (A) and (B). In (A), for VaR\(_{0.05}\), we have in increasing order: Gumbel-Hougaard, Frank, Student’s t, Gaussian and Clayton. The VaR\(_{0.05}\) level is thus higher with a Gaussian copula than with a Student’s t copula, which would suggest that models with a Gaussian copula are more conservative. This is a well-known result when comparing VaR\(_{0.05}\) levels for Gaussian and Student’s t distributions, but it is indeed not consistent since Student’s t distributions have heavy tails and the Gaussian distribution does not have heavy tails. However, in (B), for the ES\(_{0.05}\), we have in increasing order: Gumbel-Hougaard and Frank in a shared first place, thereafter Gaussian, Student’s t and Clayton. Here, we see that the Student’s t copula have a higher ES\(_{0.05}\) level than the Gaussian copula, which emphasizes the consistency in the ES measure as well as the fact that ES better captures the risk hidden in the tails.

### 4.3.2 Portfolio Expected Shortfall

All of the ES predictions are calculated for the out-of-sample set, which is one week of trading days, minute by minute with 2550 observations in total.
4.3. Expected Shortfall

In Figure 4.15 we see the in-sample portfolio log-returns to the left of the dotted line, and the out-of-sample portfolio log-returns to the right of the dotted line. In the ES estimations, we predict ES for the out-of-sample data with the estimated models based on the in-sample data.

In order to forecast ES, the predicted means $\mu_i^t$ and volatilities $\sigma_i^t$ need to be taken into account for all times $t$ and exchange rates $i$. The predicted means are created from the ARIMA-estimates, and the predicted volatilities are created from the RealGARCH-estimates. The outcome for time $t$ and exchange rate $i$ is then described by $r_i^t = \mu_i^t + \varepsilon_i^t$. Here, the error term $\varepsilon_i^t$ is decomposed into the volatility and standardized residual so that $\varepsilon_i^t = \sigma_i^t z^i$ (note that the standardized residuals are not dependent on time). The $z^i$'s are, as explained in Section 4.2, simulated using copulas and marginal distributions.

The simulated portfolio outcomes for time $t$ are therefore given by $(\mu_1^t + \sigma_1^t Z^1 + \cdots + \mu_5^t + \sigma_5^t Z^5)/5$, where $Z^i$ is a vector consisting of the $10^5$ simulated outcomes for exchange rate $i$. The $\text{ES}_{0.05}$ for each point in time is then calculated as the average of the worst 5% of the portfolio outcomes for that specific point in time. For a more thorough description of the calculation procedure, see Appendix A.6.1.

![Value at Risk and Expected Shortfalls - all copulas](image)

**Figure 4.16:** Value at Risk and Expected Shortfalls - all copulas

In Figure 4.16 we see the final calculated VaR$_{0.05}$ and ES$_{0.05}$ for all copulas with underlying time series models. As can be seen, the colour scale is more prominent in graph (B), implying that the difference between the copulas is larger in the ES graph while the VaR is almost the same for all copulas. Clayton and Student’s t are the two copulas that provide the highest ES$_{0.05}$ and VaR$_{0.05}$ values.
In Figure 4.17 we see a comparison of $\text{ES}_{0.05}$ vs $\text{VaR}_{0.05}$ for Clayton and Student’s t. Note that the $\text{ES}_{0.05}$ series have higher variances over time, as they better capture the tail-risks.

In Figure 4.18 we compare $\text{ES}_{0.05}$ for the Clayton and Student’s t copula. As can be seen in (A), it is hard to distinguish between the calculated $\text{ES}_{0.05}$ values. In (B), we take a closer look at the difference between the calculated Clayton $\text{ES}_{0.05}$ and Student’s t $\text{ES}_{0.05}$. Here, it is possible to detect some differences. The black horizontal line is the mean of the difference. Thus, the mean lies just below 0, which means that in general the Student’s t copula suggests higher $\text{ES}_{0.05}$ values than the Clayton copula. However, the differentiated series has a heavy right tail, which can be seen from the positive peaks. This implies that when the Clayton copula occasionally suggests a higher $\text{ES}_{0.05}$ value, it is significantly higher than what Student’s t suggests.
4.3. Expected Shortfall

In order to graphically evaluate the performance of the predicted $ES_{0.05}$ for each point in time we plot it against the actual realized portfolio loss series values in Figure 4.19. The grey line in (A) is the actual losses ($L = -X$ where $X$ is the log-returns) and the black line is the calculated $ES_{0.05}$, predicted for one step ahead with rolling predictions so that all information up until time $t$ is known to predict $t + 1$. Even though the predicted $ES_{0.05}$ does not capture all of the peaks in the losses, it still follows the underlying increases and decreases quite well. In (B) we see a comparison of the calculated Clayton $ES_{0.05}$ values against a calculated Historical and Normal $ES_{0.05}$. The Historical $ES_{0.05}$ was calculated from the 7550 historical observations in-sample, and the Normal $ES_{0.05}$ was calculated using a Normal distribution with mean and standard deviation equal to the in-sample mean and standard deviation. It is notable that the Historical $ES_{0.05}$ coincides almost exactly with the mean of the calculated Clayton $ES_{0.05}$ (the Historical $ES_{0.05}$ was approximately 0.000321 while the mean of the simulated Clayton $ES_{0.05}$ was 0.000324). However, the Historical $ES_{0.05}$ does not capture the fluctuations of the underlying series. Using only a Normal distribution to model the portfolio outcomes seems to underestimate $ES_{0.05}$ significantly, as the simulated Normal $ES_{0.05}$ was about 20% lower than the Clayton and Historic $ES_{0.05}$, approximately equal to 0.000256.

4.3.3 Component Expected Shortfall

Figure 4.20: Component Expected Shortfall
In Figure 4.20 we see the calculated risk contributions from the component $ES_{0.05}$ values. (A) shows the component $ES_{0.05}$ for the exchange rates with a Clayton copula, (B) shows the component $ES_{0.05}$ with a Student’s t copula. In (A), it is hard to distinguish between the risk contributions from EUR/SEK, USD/SEK and EUR/USD. However, it is clear that EUR/NOK and USD/NOK contribute the most to the portfolio $ES_{0.05}$. In (B), we see an even clearer difference between the exchange rate contributions to the portfolio risk, where EUR/USD contributes the least, followed by USD/SEK and EUR/SEK, and USD/NOK and EUR/NOK clearly contribute the most.

<table>
<thead>
<tr>
<th></th>
<th>Clayton</th>
<th>Student’s t</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Min</td>
</tr>
<tr>
<td>EUR/NOK</td>
<td>0.090</td>
<td>0.056</td>
</tr>
<tr>
<td>USD/NOK</td>
<td>0.082</td>
<td>0.015</td>
</tr>
<tr>
<td>EUR/SEK</td>
<td>0.050</td>
<td>0.022</td>
</tr>
<tr>
<td>USD/SEK</td>
<td>0.056</td>
<td>0.016</td>
</tr>
<tr>
<td>EUR/USD</td>
<td>0.043</td>
<td>0.016</td>
</tr>
</tbody>
</table>

**Table 4.13: Component $ES_{0.05}$, expressed as $10^{-3}$**

In Table 4.13 we see the mean, minimum and maximum of the component $ES_{0.05}$ for the Clayton (left hand side) and Student’s t (right hand side) copula, computed over time. EUR/NOK and USD/NOK are verified as contributing the most to the portfolio risk, whereas EUR/USD contributes the least. For the Student’s t copula, there are even some negative risk contributions for EUR/USD (minimum value $-0.005 \cdot 10^{-3}$), meaning that EUR/USD in some cases decreases the portfolio risk according to the Student’s t model.

### 4.3.4 Backtesting

<table>
<thead>
<tr>
<th></th>
<th>$n_T$</th>
<th>$p_0$</th>
<th>$p_1$</th>
<th>$p_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gaussian</td>
<td>87</td>
<td>1.000</td>
<td>0.016</td>
<td>1.000</td>
</tr>
<tr>
<td>Student’s t</td>
<td>92</td>
<td>1.000</td>
<td>0.692</td>
<td>1.000</td>
</tr>
<tr>
<td>Clayton</td>
<td>86</td>
<td>1.000</td>
<td>0.328</td>
<td>1.000</td>
</tr>
<tr>
<td>Frank</td>
<td>109</td>
<td>0.960</td>
<td>0.000</td>
<td>0.378</td>
</tr>
<tr>
<td>Gumbel</td>
<td>110</td>
<td>0.952</td>
<td>0.000</td>
<td>0.362</td>
</tr>
<tr>
<td>Normal</td>
<td>111</td>
<td>0.941</td>
<td>0.000</td>
<td>0.084</td>
</tr>
</tbody>
</table>

**Table 4.14: Number of VaR exceedances $n_T$ and $p$-values from the backtests**

In Table 4.14 we see the number of VaR$_{0.05}$ exceedances $n_T$, the $p$-value for the VaR-test $p_0$ as well as the $p$-values for the ES Test 1 and 2, $p_1$ and $p_2$, respectively.
4.3. Expected Shortfall

In the table we compare the results of all copula models and one simplistic model where we only used a Normal distribution to model the portfolio outcomes (no copula, no underlying time series models and no RealGARCH). Low \( p \)-values lead to rejection of the models, and as can be seen, none of the models can be rejected looking only at the VaR-test. In fact, all of the models perform really well considering the number of exceedances above VaR\(_{0.05}\). Given a sample size of 2550 and a confidence level of 95\%, we would expect 127.5 exceedances, which is more than what all of the models give. A reason for this can be that the volatility was higher in-sample than out-of-sample, which will be discussed further in section 5.2. Furthermore, note that the Gaussian VaR gets fewer exceedances than Student’s t on the 95\% level, which again shows the inconsistency of VaR and confirms the fact that ES is a more consistent measure.

Assuming that the calculated VaR-levels are correct, ES Test 1 tests how well the estimated models perform in predicting ES. Here, we see that the Student’s t and Clayton copula models get relatively high \( p \)-values, while the rest of the models can easily be rejected on the 5\% significance level. Thus, we cannot reject the hypothesis that the true distribution tail for the portfolio is described by the Student’s t or the Clayton model. Even though the \( p \)-value for the Student’s t model is notably higher, we cannot tell which of the Student’s t and Clayton models that perform the best. Especially not since Test 1 is insensitive to the number of VaR exceedances. Therefore, Student’s t may get a better \( p_1 \)-value than Clayton just because Student’s t has more included exceedances than Clayton. Less exceedances means that the observations that in fact are included are likely to be more extreme, and hence it might be beneficial with a higher number of exceedances in order to get a good value of the Test 1 statistic.

Looking at the \( p \)-values for the second test, \( p_2 \), we see that none of the models can be rejected (the Normal distribution could be rejected if a higher significance level would have been chosen, for example 10\%). However, keep in mind that the second test is known for accepting too many models as it tests for VaR and ES simultaneously, which can lead to acceptance of ES just because the modelled VaR performs well.
Chapter 5

Concluding Remarks

5.1 Conclusion

In this thesis we have investigated the performance of ES models based on ARIMA-RealGARCH, where the dependence between the standardized residuals was modelled by copulas. The data used is high frequency trading data, originally second by second for the realized volatility and later on minute by minute for the returns of five currency exchange rates. After having estimated the models, Value at Risk (VaR$_{0.05}$), Expected Shortfall (ES$_{0.05}$) and Component Expected Shortfall (cES$_{0.05}$) were calculated and backtests were performed. The main findings from the thesis can be summarized as follows:

- Significant AR and/or MA for all exchange rates. However, EUR/USD was chosen as ARIMA$(0, 1, 0)$ due to the lower BIC value and a preferred parsimonious approach. Even though AR and/or MA were relevant, the impacts from the ARIMA parts were small for all exchange rates. When comparing the ARIMA forecasts to actual outcomes it was notable that the forecasts were close to 0 while the actual outcomes were quite volatile.

- Significant RealGARCH$(1, 1)$ parameters. It is clear that the second by second movements in volatility affect the minute by minute volatilities and thus the magnitude of the movements in the returns. This is especially important since the purpose in this thesis has been to capture the risks arising from changes in volatility and not to forecast the exact log-return outcomes.

- For the standardized residuals, heavy tail distributions performed best for all of the exchange rates. Student’s t had the best fit for EUR/USD and the Generalized Error Distribution performed the best for the rest. Thus, even though some of the peaks in the error terms were captured by the $\sigma_t$ from the RealGARCH models, the models were still improved by letting the standardized residuals be modelled with heavy tail distributions.

- After having applied the ARIMA and RealGARCH time series models, there was still significant dependence left between the standardized residuals for the different exchange rates. This was first seen in the Pearson and Kendall’s Tau correlation matrices, and later on in the copula modelling.

- The copulas that gave the best fit were Clayton and Student’s t. The Clayton copula performed the best when all of the exchange rates were included and the Student’s t copula performed the best when EUR/USD was excluded. It is worth noting that among the evaluated copulas, these
were the only copulas with heavy left tails. It is also clear that changing the portfolio components may change the preferred copula, which could also be seen by the fact that the order of the best fitted copulas changed drastically when one of the exchange rates was excluded.

- The final Clayton and Student’s t $ES_{0.05}$ estimations were very similar to each other and it was barely possible to distinguish between them when compared to the actual portfolio losses out-of-sample. A closer look at the difference between the $ES_{0.05}$ calculations showed that the Student’s t copula model in general suggested higher $ES_{0.05}$ values. However, when the Clayton $ES_{0.05}$ occasionally was higher, it was significantly higher.

- Calculating the component $ES_{0.05}$ showed that the risk contributions in the models were the highest from EUR/NOK and USD/NOK. EUR/USD had the lowest risk contributions, and in the case with the Student’s t copula, it even decreased the risk in some of the cases.

- The backtesting showed that none of the models could be rejected when measuring the risk using $VaR_{0.05}$. However, when the risk was measured by $ES_{0.05}$, three of the five copula models were rejected. This result illustrates the known drawback of VaR as a risk measure; namely that it makes it possible to hide risk in the tail, which is not possible when measuring the risk using ES. The only two models that could not be rejected using $ES_{0.05}$ were the ones based on the Student’s t and Clayton copula. This implies that, among the models evaluated in this thesis and with $ES_{0.05}$ as the preferred risk measure, the only copulas that satisfactorily captured the risk were the ones with heavy left tails.

### 5.2 Discussion

The main topics elaborated on in this discussion are the ones we believe affect the results the most, the ones we found created the greatest obstacles and the ones we believe are important for replication and further modelling. We start by looking closer at the data modifications, after which we discuss the possible difficulties of implementing the models in a live setting. Thereafter, we discuss selected aspects of the modelling, the choices made and their implications. Finally, we discuss to which extent the ES results can be generalized and what modifications that could be made to enable greater generalization possibilities.

One of the major and most influential building blocks of this thesis, however not always emphasised, is the modifications of the raw data. Decisions in this part of the thesis will, far down the line, affect the final results and conclusions. In this thesis we chose to fill in the gaps where price quotes where missing with the most recently quoted price, in line with the assumption that the last given price is the real price until new information is obtained. As of today, there does not seem to be any generally accepted way of handling missing information, it mainly seems to depend on the author’s preferences. Other ways of approaching the problem, however, include linear interpolation between the two closest price quotes and using scaling factors to compensate for missing numbers.
Implementing an alternative approach could have significant effects on the following steps of the implementation and different problems and results could possibly arise.

Another important decision made in this thesis was to include only the trading-hours of the Swedish stock exchange, which means that many data points were excluded and that there are in fact time gaps in the time series. However, since there are only 18 time gaps in a series of in total 10200 observations, we argue that it does not affect the quality of our time series modelling significantly. On the other hand, if it would be desirable to include the nights and the weekends, the models would need to be re-specified in order to accurately capture the shifts in the characteristics of the data. Therefore, it is important to keep in mind that the models suggested in this thesis are only valid during trading hours of the Swedish stock exchange.

An important aspect of handling high frequency trading data is the time consumption, computational power and storage space needed. The modifications needed a lot of space and computational power, and the run-times were sometimes exceeding several hours. But even further down in the modelling, some of the optimization procedures and tests were computationally demanding. As a result, it would be hard to implement the multivariate ES models in a live setting with the software programs and computational power available for this thesis. Another aspect that should be considered before implementing the models live is that the prices used in this thesis are, as previously mentioned, quoted prices and not transaction prices. Furthermore, potential liquidity complications were not taken into consideration. For example, one arising complication could be that not all of the assets can be netted out at the quoted best price, or that the actual amount offered is not large enough to carry out the transaction. This would result in other prices and therefore also other log-returns and risks than expected from our models. More on this topic can be found in the future research section 5.3.

An important insight from the volatility modelling was the major difference in the goodness of fit of EUR/USD and USD/SEK volatilities compared to EUR/SEK and EUR/NOK volatilities. As earlier stated, a GARCH model will never perform better than the actual accuracy of the underlying measurement of the volatility. Since realized volatility is a more accurate measure than for example noisy squared returns, we selected realized volatility as our volatility measure. However, realized volatility only becomes an accurate measure if the underlying asset is liquid enough. As can be seen for EUR/USD, the exchange rate is highly liquid and has a high price frequency, even when going as far down as to the millisecond level. Comparing this to EUR/NOK or EUR/SEK we see a significant difference, where there even for minute long sequences were periods with no changes in the price quotations. In the cases of no or few price movements during longer periods, the realized volatility either remained zero or converged to an approximation of the noisy measure of squared returns. In fact, when comparing the EUR/SEK and EUR/NOK volatility graphs (Figure 4.4) with the example graph of squared returns from the theory chapter (Figure 2.1), there were clear similarities. Even though the end results when calculating ES were satisfactory in this case, the liquidity of the underlying assets affects the results. Therefore, the appropriate sampling frequency should always be examined before initiating RealGARCH modelling.
Regarding the dependence modelling, it should be emphasized that many of the explicit expressions for the copula functions only seem to be available for the bivariate case. When considering the multivariate case (>2 assets), it is both complicated and time demanding to find definitions and explicit expressions for the copula functions. Furthermore, density functions, Kendall’s Tau and other calculations tend to become increasingly complex as the number of variables increases. It should also be mentioned that the final choice of copula is extremely dependent on the features of the underlying variables. For example, we saw that when excluding one of the exchange rates from the portfolio, there was a complete new order of the best fitted copulas. Furthermore, the test statistic implemented in this thesis tests the overall fit of the copula and not only the fit in the tails. It could be argued that the actual left tail fit is a more important issue than the overall fit, since the main goal is to accurately fit a risk measure to the tails. Due to this, it can be misleading to choose copula with the Cramér-von Mises test statistic, and a test which instead evaluates the performance in the left tail could be beneficial. For example, the Clayton (heavy left tail) copula performed worse than the Frank (no heavy tails) copula for the portfolio without EUR/USD according to the Cramér-von Mises test statistic. It could have been expected that the Clayton copula would still perform better, since the underlying data actually has heavy tails, but this was likely cancelled out because of the non-symmetry and overall inferior fit of the Clayton copula. However, when backtesting ES for the full portfolio, it turned out that the only two accepted models were the ones based on the heavy tail Student’s t and Clayton copulas. It is also interesting to note that even though the data in this thesis was not extracted from a period of financial distress, the Gaussian copula could easily be rejected when backtesting ES, which implies that it is not an appropriate copula for financial risk modelling.

The fact that the underlying data was extracted from a period of calm markets leads us to the discussion about whether the results are applicable and extendable in a setting of financial distress. In Figure 4.3 we see that some of the volatility forecasts seem to follow the realized volatility series quite accurately, which could indicate that the forecasts would be reliable also in a distressed setting. However, when examining the ES results, we noted that the ES models seemingly overestimated the risk in the portfolio. When comparing the in-sample data to the out-of-sample data it seems as if the there might have been a shift in the overall volatility level between the two periods, where the in-sample period was a bit more volatile. This indicates that a larger shift in the volatility might be difficult to capture and that it for that reason could be beneficial to re-estimate the coefficients of the models on a rolling basis.

Other important topics regarding the generalization of the results, are for example the forecast horizon and the nature of the assets in the portfolio. In this thesis we focused on estimating ES on a one-step-ahead horizon. This turned out to be accurately achievable, however, the question is whether this could be extended to a multi-step-ahead forecast horizon. Due to the model being built on a differentiated series, the spread of the density forecasts may increase unlimitedly as we look further ahead into the future. This indicates that generalizations of the results beyond the one-step-ahead horizon are unlikely to be valid. If it would be desirable to estimate ES for a 1-hour-step-ahead horizon, or even further into the future, the underlying frequency of the data would probably have to change to match the new forecast horizon. Regarding the nature of
the assets in the portfolio, we argue that as soon as there is a change in which assets that constitute the portfolio, the model needs to be re-estimated. If a new asset is added to the portfolio, its marginal distribution needs to be estimated and the dependence structure for the portfolio needs to be re-estimated.

5.3 Future Research

As identified by The Basel Committee on Banking Supervision (Aldridge 2010), the following types of risk affect financial securities:

1. Market risk - induced by price movements of market securities
2. Credit and counterparty risk - addresses the ability and intent of trading counterparties to uphold their obligations
3. Liquidity risk - the ability of the trading operation to quickly unwind positions
4. Operational risk - the risk of financial losses embedded in daily trading operations
5. Legal risk - the risk of litigation expenses

In this thesis we have focused on the market risk, and we therefore first and foremost recommend further research on including the liquidity aspect in the models. For high frequency trading, the liquidity risk is closely related to the market risk. The risk is especially high for thin currencies like NOK, with relatively low supply. Not only is there a risk that it will be problematic to quickly unwind positions, it is also highly likely that the trading volume will affect the prices.

As suggested by Aldridge (Aldridge 2010), VaR can be adjusted to take the liquidity risk into account. Aldridge presents three different liquidity adjustment methods, looking at the spread, trade volume and price movements. Furthermore, there are a number of studies on the dependence structure between trading volume and price volatility, most of them coming to the conclusion that there clearly is a significant dependence (to mention a few: (Sandström and Bosaeus 2015), (Zheng, Moulines, and Abergel 2012) and (Duong and Kalev 2008)).

Moreover, it would be interesting to further develop the current models by using other methods of measuring volatility (for example kernel-based estimators), other time series models (for example multiple RealGARCH models and not just RealGARCH(1, 1)) and evaluating the performance of more flexible and asymmetric dependence structures (for example nested Archimedean copulas). Regarding residual distributions, a future area of interest could be to test whether skewed distributions could have a better fit than the symmetric distributions used in this thesis. Some of the standardized residual distribution plots indicate a slight skewness being present, and taking the skewness into account could therefore improve the fit of the models even further.

Another recommended topic for future research is extreme value theory and for example the peak over threshold method. In this thesis we have focused on modelling the whole distributions of the underlying assets, but since ES only
depends on the tail distributions, an alternative approach would be to create models that are only adapted for measuring downside risks. However, we may argue that in our case it is useful to model the whole distribution, since it is common to have both long and short positions in currencies.
Appendix A

A.1 Descriptive Statistics

A.1.1 EUR/SEK

![EUR/SEK exchange rate](image1)

**Figure A.1**: EUR/SEK exchange rate, one month, 09.00-17.30

![EUR/SEK log-returns and realized volatility](image2)

(A) Log-returns per minute  
(B) Realized volatility per minute

**Figure A.2**: EUR/SEK, one month, 09.00-17.30
A.1.2 EUR/NOK

Figure A.3: EUR/NOK exchange rate, one month, 09.00-17.30

(A) Log-returns per minute  
(b) Realized volatility per minute

Figure A.4: EUR/NOK, one month, 09.00-17.30

A.1.3 EUR/USD

Figure A.5: EUR/USD exchange rate, one month, 09.00-17.30
A.1. Descriptive Statistics

\[ \text{(A) Log-returns per minute} \quad \text{(B) Realized volatility per minute} \]

Figure A.6: EUR/USD, one month, 09.00-17.30

### A.1.4 USD/NOK

\[ \text{(A) Log-returns per minute} \quad \text{(B) Realized volatility per minute} \]

Figure A.7: USD/NOK exchange rate, one month, 09.00-17.30

Figure A.8: USD/NOK, one month, 09.00-17.30
A.2 Time Series Model Selection

A.2.1 EUR/SEK

<table>
<thead>
<tr>
<th>(p,q)</th>
<th>(p,0)</th>
<th>(p,1)</th>
<th>(p,2)</th>
<th>(p,3)</th>
<th>(p,4)</th>
<th>(p,5)</th>
</tr>
</thead>
</table>

**Table A.1: EUR/SEK - BIC matrix with RealGARCH(1,1)**

A.2.2 EUR/NOK

<table>
<thead>
<tr>
<th>(p,q)</th>
<th>(p,0)</th>
<th>(p,1)</th>
<th>(p,2)</th>
<th>(p,3)</th>
<th>(p,4)</th>
<th>(p,5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0,q)</td>
<td>-12.4401</td>
<td>-12.4456</td>
<td>-12.4383</td>
<td>-12.4457</td>
<td>-12.4403</td>
<td>-12.4398</td>
</tr>
<tr>
<td>(1,q)</td>
<td>-12.4452</td>
<td>-12.4406</td>
<td>-12.4389</td>
<td>-12.4448</td>
<td>-12.4393</td>
<td>-12.4391</td>
</tr>
<tr>
<td>(2,q)</td>
<td>-12.4378</td>
<td>-12.4389</td>
<td>-12.5186</td>
<td>-12.4410</td>
<td>-12.4438</td>
<td>-12.4347</td>
</tr>
<tr>
<td>(3,q)</td>
<td>-12.4454</td>
<td>-12.4382</td>
<td>-12.4410</td>
<td>-12.4427</td>
<td>-12.4348</td>
<td>-12.4334</td>
</tr>
<tr>
<td>(4,q)</td>
<td>-12.4401</td>
<td>-12.4393</td>
<td>-12.4432</td>
<td>-12.4348</td>
<td>-12.4348</td>
<td>-12.4412</td>
</tr>
<tr>
<td>(5,q)</td>
<td>-12.4371</td>
<td>-12.4393</td>
<td>-12.4349</td>
<td>-12.4335</td>
<td>-12.4410</td>
<td>-12.4399</td>
</tr>
</tbody>
</table>

**Table A.2: EUR/NOK - BIC matrix with RealGARCH(1,1)**
### A.2.3 EUR/USD

<table>
<thead>
<tr>
<th>(p,q)</th>
<th>(p,0)</th>
<th>(p,1)</th>
<th>(p,2)</th>
<th>(p,3)</th>
<th>(p,4)</th>
<th>(p,5)</th>
</tr>
</thead>
</table>

**Table A.3: EUR/USD - BIC matrix with RealGARCH(1,1)**

### A.2.4 USD/NOK

<table>
<thead>
<tr>
<th>(p,q)</th>
<th>(p,0)</th>
<th>(p,1)</th>
<th>(p,2)</th>
<th>(p,3)</th>
<th>(p,4)</th>
<th>(p,5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0,q)</td>
<td>-12.7023</td>
<td>-12.6094</td>
<td>-12.6085</td>
<td>-12.6075</td>
<td>-12.6064</td>
<td>-12.6053</td>
</tr>
<tr>
<td>(1,q)</td>
<td>-12.6092</td>
<td>-12.6087</td>
<td>-12.6076</td>
<td>-12.6064</td>
<td>-12.6058</td>
<td>-12.7042</td>
</tr>
<tr>
<td>(2,q)</td>
<td>-12.7081</td>
<td>-12.6076</td>
<td>-12.7064</td>
<td>-12.6062</td>
<td>-12.6057</td>
<td>-12.6054</td>
</tr>
<tr>
<td>(3,q)</td>
<td>-12.6075</td>
<td>-12.6064</td>
<td>-12.6057</td>
<td>-12.7056</td>
<td>-12.6043</td>
<td>-12.6029</td>
</tr>
<tr>
<td>(4,q)</td>
<td>-12.6064</td>
<td>-12.6053</td>
<td>-12.6044</td>
<td>-12.6030</td>
<td>-12.7037</td>
<td>-12.6032</td>
</tr>
<tr>
<td>(5,q)</td>
<td>-12.6053</td>
<td>-12.7041</td>
<td>-12.6040</td>
<td>-12.6044</td>
<td>-12.6021</td>
<td>-12.6037</td>
</tr>
</tbody>
</table>

**Table A.4: USD/NOK - BIC matrix with RealGARCH(1,1)**
### A.3 Time Series Model Diagnostics

<table>
<thead>
<tr>
<th></th>
<th>$\psi_1$</th>
<th>$\psi_2$</th>
<th>$\theta_1$</th>
<th>$\theta_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>EUR/SEK</td>
<td>0.0000</td>
<td>-</td>
<td>0.0000</td>
<td>-</td>
</tr>
<tr>
<td>EUR/NOK</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>-</td>
</tr>
<tr>
<td>EUR/USD</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>USD/SEK</td>
<td>-</td>
<td>-</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>USD/NOK</td>
<td>0.0193</td>
<td>-</td>
<td>0.0033</td>
<td>-</td>
</tr>
</tbody>
</table>

Table A.5: $p$-values for the best fit ARIMA-parameters

<table>
<thead>
<tr>
<th></th>
<th>$\omega$</th>
<th>$\alpha_1$</th>
<th>$\beta_1$</th>
<th>$\eta_1$</th>
<th>$\eta_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>EUR/SEK</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.2870</td>
<td>0.0000</td>
</tr>
<tr>
<td>EUR/NOK</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>EUR/USD</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.6216</td>
<td>0.0000</td>
</tr>
<tr>
<td>USD/SEK</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.1600</td>
<td>0.0000</td>
</tr>
<tr>
<td>USD/NOK</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

Table A.6: $p$-values for the best fit RealGARCH-parameters

<table>
<thead>
<tr>
<th></th>
<th>$\xi$</th>
<th>$\delta$</th>
<th>$\lambda$</th>
<th>$\nu$</th>
</tr>
</thead>
<tbody>
<tr>
<td>EUR/SEK</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>EUR/NOK</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>EUR/USD</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>USD/SEK</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>USD/NOK</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

Table A.7: $p$-values for the best fit RealGARCH-parameters
A.3.1 EUR/SEK

![ACF of observations](image1.png) ![ACF of standardized residuals](image2.png)

**Figure A.9:** EUR/SEK autocorrelation diagnostics

![Empirical Density of Standardized Residuals](image3.png) ![GED - QQ Plot](image4.png)

**Figure A.10:** EUR/SEK diagnostics of standardized residuals

<table>
<thead>
<tr>
<th>$h$</th>
<th>$\chi^2$</th>
<th>$df$</th>
<th>$p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>6.7308</td>
<td>1</td>
<td>0.0095</td>
</tr>
<tr>
<td>4</td>
<td>16.0150</td>
<td>2</td>
<td>0.0003</td>
</tr>
<tr>
<td>5</td>
<td>16.0420</td>
<td>3</td>
<td>0.0011</td>
</tr>
<tr>
<td>10</td>
<td>20.3020</td>
<td>8</td>
<td>0.0093</td>
</tr>
</tbody>
</table>

**Table A.8:** Ljung-Box test for EUR/SEK, joint $H_0$ no autocorrelation
A.3.2 EUR/NOK

(A) ACF of observations  (b) ACF of standardized residuals

FIGURE A.11: EUR/NOK autocorrelation diagnostics

(A) Histogram with fitted GED vs Normal  (b) QQ plot with fitted GED

FIGURE A.12: EUR/NOK diagnostics of standardized residuals

<table>
<thead>
<tr>
<th>$h$</th>
<th>$\chi^2$</th>
<th>$df$</th>
<th>$p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>2.1713</td>
<td>1</td>
<td>0.1406</td>
</tr>
<tr>
<td>5</td>
<td>3.3931</td>
<td>2</td>
<td>0.1833</td>
</tr>
<tr>
<td>10</td>
<td>14.6890</td>
<td>7</td>
<td>0.0402</td>
</tr>
</tbody>
</table>

TABLE A.9: Ljung-Box test for EUR/NOK, joint $H_0$ no autocorrelation
A.3.3 EUR/USD

(A) ACF of observations

(B) ACF of standardized residuals

Figure A.13: EUR/USD autocorrelation diagnostics

(A) Histogram with fitted GED vs Normal

(B) QQ plot with fitted GED

Figure A.14: EUR/USD diagnostics of standardized residuals

<table>
<thead>
<tr>
<th>$h$</th>
<th>$\chi^2$</th>
<th>$df$</th>
<th>$p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>1.8380</td>
<td>3</td>
<td>0.6067</td>
</tr>
<tr>
<td>4</td>
<td>2.2865</td>
<td>4</td>
<td>0.6832</td>
</tr>
<tr>
<td>5</td>
<td>2.3335</td>
<td>5</td>
<td>0.8013</td>
</tr>
<tr>
<td>10</td>
<td>5.7858</td>
<td>10</td>
<td>0.8329</td>
</tr>
</tbody>
</table>

Table A.10: Ljung-Box test for EUR/USD, joint $H_0$ no autocorrelation
Appendix A. Appendix

A.3.4 USD/NOK

![ACF of Observations](image1.png)  ![ACF of Standardized Residuals](image2.png)

(A) ACF of observations  (B) ACF of standardized residuals

**Figure A.15**: USD/NOK autocorrelation diagnostics

![Empirical Density of Standardized Residuals](image3.png)  ![GED - QQ Plot](image4.png)

(A) Histogram with fitted GED vs Normal  (B) QQ plot with fitted GED

**Figure A.16**: USD/NOK diagnostics of standardized residuals

<table>
<thead>
<tr>
<th>$h$</th>
<th>$\chi^2$</th>
<th>$df$</th>
<th>$p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>2.8545</td>
<td>1</td>
<td>0.0911</td>
</tr>
<tr>
<td>4</td>
<td>2.8815</td>
<td>2</td>
<td>0.2368</td>
</tr>
<tr>
<td>5</td>
<td>5.7884</td>
<td>3</td>
<td>0.1224</td>
</tr>
<tr>
<td>10</td>
<td>8.3893</td>
<td>8</td>
<td>0.3964</td>
</tr>
</tbody>
</table>

**Table A.11**: Ljung-Box test for USD/NOK, joint $H_0$ no autocorrelation

A.4 Time Series Forecasts

The forecast graphs that were not included in the report.
A.5. Copulas

Copula       $\theta_{ML}$       $\nu_{ML}$
-------------  -------------  ------------
Gaussian      -             -             
Student’s t   -             3.23571       
Clayton       0.46996       -             
Gumbel        1.25318       -             
Frank         2.02450       -             

Table A.12: Parameter estimations without EUR/USD
Appendix A. Appendix

Figure A.19: Elliptical copulas without EUR/USD

Figure A.20: Archimedean copulas without EUR/USD

A.6 Expected Shortfall

A.6.1 Description of Implementation

In order to calculate the portfolio ES_{0.05} and the component ES_{0.05}, cES_{0.05}, the following procedure was used:

1) Import the predicted mean and standard deviation vectors, \( \mu^i \) and \( \sigma^i \), for all exchange rates \( i \) (obtained from the ARIMA and RealGARCH models).

2) Compute the sum of the mean vectors, \( \mu = \sum_{i=1}^{5} \mu^i \) (gives a combined mean vector with 2550 rows).

3) For all selected dependence structures (five copulas in this thesis):
   a) Simulate \( 10^5 \) outcomes of \( (z^1, \ldots, z^5) \) from the copula with underlying marginal distributions.
   b) For each exchange rate:
      i) Multiply the predicted \( \sigma^i \) for each point in time with the \( 10^5 \) simulated outcomes of \( z^i \) (creates five matrices with \( 10^5 \) rows and 2550 columns).
   c) Add the five matrices together so that \( \sigma z = \sum_{i=1}^{5} \sigma^i z^i \) (creates a matrix with \( 10^5 \) rows and 2550 columns).
d) For each point $t$ in time (2550 time points in total):
   
   i) Pick out column $t$ from the combined matrix as well as from the five separate matrices.
   
   ii) Create a new matrix with all extracted columns from time $t$, so that the new matrix consists of $10^5$ rows and 6 columns in total (one column for the combined $\sigma_t z_t$, five for the separate exchange rates $\sigma^i_t z^i_t$).
   
   iii) Sort the matrix with respect to the combined $\sigma_t z_t$ column and extract the 5% worst outcomes based on the combined $\sigma_t z_t$ (not necessarily the worst outcomes for each of the exchange rates separately, only for the combined portfolio).
   
   iv) Compute the average of the 5% extracted outcomes within each column. Save the average values for the combined $\sigma_t z_t$ and all separate exchange rates $\sigma^i_t z^i_t$.
   
   e) Compute the ES$_{0.05}$ through adding the mean value vector $\mu$ to the vector with the 5% averages of of $\sigma_t z_t$. Divide by 5 to obtain ES$_{0.05}$ (equal portfolio weights).
   
   f) Compute the component ES$_{0.05}$, cES$_{0.05}$, through adding the mean value vector $\mu^i$ for exchange rate $i$ to the vector with the 5% average outcomes of $\sigma^i_t z^i_t$. Divide by 5 to obtain cES$_{0.05}$.

(4) Repeat procedure until all copula structures have been simulated.

### A.6.2 Graphs

**Value at Risk vs Expected Shortfall**

**Component Expected Shortfall**

(A) Expected Shortfall vs Value at Risk

(b) Risk contributions

**Figure A.21:** Gaussian copula
Appendix A. Appendix

(A) Expected Shortfall vs Value at Risk  
(B) Risk contributions

FIGURE A.22: Frank copula

(A) Expected Shortfall vs Value at Risk  
(B) Risk contributions

FIGURE A.23: Gumbel copula
References


