Distributional Dynamics of Fama-French Factors in European Markets

WILMER LÖFGREN
Distributional Dynamics of Fama-French Factors in European Markets

WILMER LÖFGREN
Abstract

The three-factor model of Fama and French has proved to be a seminal contribution to asset pricing theory, and was recently extended to include two more factors, yielding the Fama-French five-factor model. Other proposed augmentations of the three-factor model includes the introduction of a momentum factor by Carhart. The extensive use of such factors in asset pricing theory and investing motivates the study of the distributional properties of the returns of these factors. However, previous studies have focused on subsets of these six factors on the U.S. market. In this thesis, the distributional properties of daily log-returns of the five Fama-French factors and the Carhart momentum factor in European data from 2009 to 2019 are examined. The univariate distributional dynamics of the factor log-returns are modelled as ARMA-NGARCH processes with skewed $t$ distributed driving noise sequences. The Gaussian and $t$ copula are then used to model the joint distributions of these factor log-returns. The models developed are applied to estimate the one-day ahead Value-at-Risk (VaR) in testing data. The estimations of the VaR are backtested to check for correct unconditional coverage and exponentially distributed durations between exceedances. The results suggest that the ARMA-NGARCH processes are a valid approximation of the factor log-returns, and lead to good estimations of the VaR. The results of the multivariate analysis suggest that constant Gaussian and $t$ copulas might be insufficient to model the dependence structure of the factors, and that there might be a need for more flexible copula models with dynamic correlations between factor log-returns.
Sammanfattning

Acknowledgements

I am very grateful for the support I have received from my thesis supervisor, associate professor Anja Janßen. In our meetings she has provided me with helpful guidance and interesting discussion. Her comments on preliminary versions of this report have been particularly beneficial.

This project was made possible by Jonas Sandefeldt at QQM Fund Management. I am very thankful that he took the time to introduce me to previous research on the Fama-French factors, and that I was given the opportunity to design a thesis project were I could learn more about subjects that I find interesting.

Lastly, I want to express my gratitude to my family and friends. The encouragement they have provided me during not only this thesis project, but the entirety of my studies at KTH, has been immense.
# Contents

1 Introduction 1

2 Background 3
   2.1 Factor Models and Asset Pricing 3
      2.1.1 The Empirical Motivation of Factor Models 3
      2.1.2 The Fama-French Factors 5
      2.1.3 The Momentum Factor 8
      2.1.4 Previous Studies of Factor Dynamics and Dependence 9
   2.2 Mathematical Background 11
      2.2.1 Skewed Student’s $t$ distribution 11
      2.2.2 ARMA-GARCH models 13
      2.2.3 Nonlinear-asymmetric GARCH 18
      2.2.4 Modelling Dependence with Copulas 19
      2.2.5 Risk Measures 23
      2.2.6 Backtesting Predicted Value-at-Risk 26

3 Methods 29
   3.1 Data 29
   3.2 Univariate Modelling 30
   3.3 Dependence Modelling 32

4 Results 34
   4.1 Factor Dynamics 34
      4.1.1 Factor Properties 34
      4.1.2 Modelling the Factors 39
      4.1.3 Dynamic Estimation of VaR 43
   4.2 Factor Dependence 47
      4.2.1 Factor Correlations and Dependence Modelling 47
      4.2.2 Dynamic Estimation of VaR 52
## CONTENTS

5 Discussion  
5.1 Factor Dynamics ............................................. 58  
5.2 Factor Dependence .......................................... 60  

6 Conclusions ................................................. 63

Bibliography .................................................. 65
Chapter 1

Introduction

The three-factor model of Fama and French [1] is a linear model where three market factors are used as explanatory variables for the returns of stock portfolios. The introduction of this model has proved to be a seminal contribution to empirical asset pricing theory, and the model is prevalent in literature. The three factors of this model are a general market factor, a firm size related factor, and a factor related to book-to-market equity. In this thesis, these three factors will be referred to as MKT, SMB, and HML.

In later papers, extensions to this three-factor model have been proposed. The addition of a momentum factor, which aims to capture effects related to a momentum in stock performance, in Carhart [2] is one such extension. This factor will be referred to as MOM in this thesis. Fama and French [3] have also proposed a five-factor model formed by augmenting their original three-factor model with a factor related to the profitability of firms and a factor related to the investment behavior of firms. These two factors will be referred to as RMW and CMA in this thesis.

With the exception of the general market factor, all these factors are captured by taking the average return of stocks with certain characteristics minus the average return of stocks with opposite characteristics. Besides their use in research within asset pricing theory, these factors are also widely used as a basis for investment strategies [4] [5]. Therefore, it is of both academic and industry relevance to study the distributional dynamics of the returns of these factors.

In two recent papers [6] [7], the distributional properties of the log-returns of the five Fama-French factors and the momentum factor are studied, and the resulting models are applied to estimate risk measures and perform portfolio optimizations. However, in these two papers only subsets of the six factors
are considered. Furthermore, the studies are only based on data from the U.S. market.

The aim of the present thesis is therefore to study the distributional properties of the daily log-returns of all six factors in the European market. Both the univariate distributional dynamics and the dependence structure of the factors are studied, and resulting models are applied to estimate the Value-at-Risk (VaR) of the factors in testing data. In particular, the below two research questions are examined.

1. What are the univariate and joint distributional properties of the daily log-returns of the five Fama-French factors and the momentum factor in the European market?

2. Does models developed to answer 1. give reliable estimates of the one-day ahead Value-at-Risk for the six factors?

The disposition of the thesis is as follows: In Chapter 2 a more detailed presentation of the theoretical background of the factors is given, as well as a summary of the two previous studies of the distributional properties of the factors in the U.S. market. Chapter 2 also includes a comprehensive review of the mathematical background of the modelling in this thesis. Chapter 3 presents the data that is used and outlines the methodology of the thesis. In Chapter 4 the results of the analysis is given. Chapter 5 is a discussion of the results, and the conclusions of the thesis are presented in Chapter 6.
Chapter 2

Background

2.1 Factor Models and Asset Pricing

The focus of the present thesis is the modelling of the univariate distributional dynamics and dependence structure of six risk factors on the European equity market. The development of such models is motivated by the widespread industry and academic use of these factors. The factors are used in research related to asset pricing theory, and in investment strategies [4][5]. In the next four subsections, a brief overview of the empirical results that lead to the development of these factor models is first presented. This is followed by the definitions of the factors. Finally, two previous studies of factor dynamics and dependence in the U.S. equity market are summarized.

2.1.1 The Empirical Motivation of Factor Models

The Fama-French three and five-factor models described in section 2.1.2, and factor models including a momentum factor as described in section 2.1.3, are all extensions of the capital asset pricing model (CAPM) of Sharpe [8] and Lintner [9].

The reasoning behind CAPM is based on the mean-variance portfolio allocation model of Markowitz [10]. If investors are believed to follow this model, they will select portfolios that minimize the variance of the portfolio return while maximizing its expected return. Such portfolios are said to be mean-variance-efficient.

To the assumptions implied by the Markowitz mean-variance model [8] and [9] add the assumption that all investors can borrow and lend at a risk-free rate, and that this rate is equal for all investors. Furthermore, they also assume
that at time $t$ all investors agree on the distribution of asset returns in $t + 1$, and that this distribution is the true distribution of returns.

From these assumptions one can show that all efficient portfolios are comprised of positions in the risk-free asset and in a single portfolio of risky assets known as the tangency portfolio. This means that all investors share one risky portfolio, and this tangency portfolio must therefore be the value-weight market portfolio of risky assets. Furthermore, this means that the weight of any risky asset in the tangency portfolio is equal to the total market value of the risky asset divided by the total market value of all risky assets.

Since the market portfolio has to be efficient, the following equation that holds for any minimum variance portfolio must hold for the market portfolio as well

$$E[R_i] = R_f + \beta_i (E[R_m] - R_f).$$

In this equation $R_i$ is the return of risky asset $i$, with $i = 1, 2, ..., N$ if there are $N$ risky assets in total. $R_f$ is the risk-free rate and $\beta_i$ is the market beta of risky asset $i$

$$\beta_i = \frac{\text{Cov}(R_i, R_m)}{\sigma^2_{R_m}}.$$

This market beta is also the ordinary least squares estimate of the slope in a linear regression model of the return of asset $i$ with the market return being the explanatory variable. Thus, the market beta is often interpreted as a measure of how sensitive asset $i$ is to the outcomes of the market return.

Empirical work to test implications of the CAPM have rejected the model. The model can be tested by cross-sectional regression, where one fits a linear regression model to the cross-section of average returns of risky assets against estimates of the market betas of the assets. The CAPM implies that such a regression will have the risk-free rate, $R_f$, as its intercept, and that the slope will be the market return in excess of the risk-free rate, $E[R_m] - R_f$. Often, the slope is found to be flatter and the intercept higher, than what is implied by the CAPM [11].

A time-series regression test based on the following model introduced in Jensen [12] fits the CAPM to time series data of asset returns

$$R_{it} - R_{ft} = \alpha_i + \beta_i (R_{Mt} - R_{ft}) + \epsilon_{it}.$$

The CAPM states that the expected value of an asset’s excess return, $R_{it} - R_{ft}$, is explained by its market beta multiplied by the expected value of the excess
market return, $R_{Mt} - R_{ft}$. The testable implication of CAPM is thus that $\alpha_i$, often called *Jensen’s alpha*, should equal zero. In Jensen’s original paper, $\alpha$ was introduced as a risk-adjusted estimate of how much the active management of a fund contributes to its returns.

An early rejection of CAPM by both cross-sectional regression and time series regression is found in [13]. In Black [14] a version of CAPM that does not assume risk-free borrowing or lending is introduced. Black’s version of CAPM instead allows unrestricted short sales of risky assets, and in early empirical works like the cross-sectional regression tests in [15], and the time-series regression tests in [16], it fairs better than the Lintner-Sharpe CAPM.

However, later empirical work show that a fundamental flaw of CAPM is that the market beta fails to explain a lot of the variation in expected return, and that there seems to exist other risk factors besides the market that have explanatory power in expected returns of stocks. In [17] stocks with high earnings-price ratios (E/P) are shown to have higher returns than what is implied by the CAPM. A similar effect due to company size is shown in [18]. In [19] stocks with high book-to-market equity ratios (B/M), the ratio of the book value of a stock to its market value, have average returns that are higher than what their market betas imply.

Empirical results such as these, suggesting a need to add additional risk factors beside the market factor, was the motivation behind the development of the factor models that are presented in section 2.1.2 and 2.1.3. Starting in section 2.1.2, the excess market return factor, $R_{Mt} - R_{ft}$, will be denoted MKT to conform to the naming procedure of the other factors. A more comprehensive overview of the theory behind CAPM and some of the empirical results leading to the development of asset pricing models with several risk factors can be found in [11].

### 2.1.2 The Fama-French Factors

The previous section is a review of the history of theoretical and empirical work concerning the CAPM. The empirical rejection of this asset pricing model is the background to the introduction of the factor models of Fama-French. Based on empirical works suggesting that average stock returns vary with size and book-to-market (B/M) equity ratio, Fama and French introduced a three-factor model in 1993 [1].

$$E[R_{it}] - R_{ft} = \beta_{iM} E[MKT_t] + \beta_{is} E[SMB_t] + \beta_{ih} E[HML_t] + \epsilon_{it}$$
In this extension of the CAPM, the SMB (Small minus Big) and HML (High minus Low) factors are added. The SMB factor is the difference between the returns on portfolios of small stocks and portfolios of big stocks (small and big by market capitalization). HML is the difference between the returns on portfolios of high and low B/M stocks. In practice, these factors are captured by taking the average returns of stock with certain characteristics minus the average return of stocks with opposite characteristics. The three betas of this model are the slopes in the multiple regression model of $R_{it} - R_{ft}$ with explanatory variables $MKT_t$, $SMB_t$, and $HML_t$.

To capture the SMB and HML factors, [1] uses a 2x3 sorting (two size groups and three B/M groups) scheme of the stocks in a market, leading to the creation of six portfolios. Combinations of these six portfolios are then used to proxy SMB and HML.

To form the two size groups, stocks are ranked on market size, the median of which is then used to create a group of small (S) stocks and a group of big (B) stocks. The three B/M groups are formed on the 30th and 70th quantiles of B/M. The lower 30% are called low (L), the middle 40% are called medium (M), and the top 30% are called high (H). The reason for sorting size into two groups and B/M into three groups comes from results of an earlier paper by Fama and French [20], where large differences in B/M are found to have a stronger role in average stock returns.

The six portfolios made from the intersection of these groups are called SL, SM, SH, BL, BM, and BH. For example, stocks with a market capitalization belonging to the larger half of the market and a B/M belonging to top 30% will form the BH portfolio. In [1] the two size groups and three B/M groups are constructed once every year in June.

Thus, in this three-factor model the SMB portfolio is the difference in the average return of the three portfolios of small stocks and the average return of the three portfolio of big stocks. The HML portfolio is similarly defined as the difference in the average return of the two portfolios of high B/M stock and the average return of the two portfolios of low B/M stock.

By time series regression (as introduced in the section 2.1.1) the Fama-French three-factor is shown in [1], [21] and [22] to outperform CAPM in describing average stock returns in several large markets.

In 2015 Fama and French [3] introduced a five-factor asset pricing model, in which their earlier three-factor model is expanded by adding a CMA (Conservative minus Aggressive) factor, and a RMW (Robust minus Weak) factor. The time-series regression model of the Fama-French five-factor model is thus
formulated as
\[
R_{it} - R_{ft} = \alpha_i + \beta_{iM} \text{MKT}_t + \beta_{iS} \text{SMB}_t + \beta_{iH} \text{HML}_t \\
+ \beta_{ir} \text{RMW}_t + \beta_{ic} \text{CMA}_t + \epsilon_{it}.
\]

The CMA factor is the difference in returns on portfolios of stocks of firms with low and high levels of investment. The RMW factor is the difference between returns on portfolios of stocks with robust and weak profitability.

The introduction of these two new factors was motivated by evidence that these factors relate to variations in average returns that the three-factor model does not capture. In [23] a proxy for expected profitability is shown to explain cross-sectional average returns with explanatory power similar to that of B/M. In [24] a negative relation is found between expected investment and returns.

The 2x3 sorting scheme employed for the three-factor model is again used in the practical application of the five-factor model. The stocks are resorted at the end of June each year. Stocks are sorted into two size groups, small (S) and big (B), based on the median market capitalization. The 30\text{th} and 70\text{th} quantiles of B/M, investment (Inv), and operating profitability (OP) are used to create three groups each for these three properties. For the sort in year \( t \), the Inv is measured by the change in total assets from the fiscal year ending in year \( t - 2 \) to the fiscal year ending in \( t - 1 \), divided by the total assets in \( t - 2 \). The OP is formed using accounting data for the fiscal year ending in year \( t - 1 \) and is measured by revenues minus costs divided by book equity.

The B/M groups are high (H), neutral (N), and low (L). The OP groups are robust (R), neutral (N), and weak (W). The Inv groups are conservative (C), neutral (N), and aggressive (A). Using the intersection of the size groups and the other groups, three sets of 2x3 sorts are created. Thus 18 portfolios are formed, and combinations of these are used to proxy the factors. Since the size grouping is used in the sorting of the three other properties, three types of size factors can be formed, and the SMB factor is the average of these

\[
\text{SMB}_{B/M} = \frac{SH + SN + SL}{3} - \frac{BH + BN + BL}{3},
\]

\[
\text{SMB}_{OP} = \frac{SR + SN + SW}{3} - \frac{BR + BN + BW}{3},
\]

\[
\text{SMB}_{Inv} = \frac{SC + SN + SA}{3} - \frac{BC + BN + BA}{3},
\]

\[
\text{SMB} = \frac{\text{SMB}_{B/M} + \text{SMB}_{OP} + \text{SMB}_{Inv}}{3}.
\]
The other three factors are formed by taking the average return of the two portfolios belonging to the 70\textsuperscript{th} percentile (measured by the factor property) minus the average return of the two bottom 30\% portfolios

\[
\text{HML} = \frac{SH + BH}{2} - \frac{SL + BL}{2},
\]
\[
\text{RMW} = \frac{SR + BR}{2} - \frac{SW + BW}{2},
\]
\[
\text{CMA} = \frac{SC + BC}{2} - \frac{SA + BA}{2}.
\]

In [3] a data set of the U.S. market during the period July 1963-December 2013 is used, and it is found that the adding the RMW and CMA factor makes the HML factor redundant. A high positive correlation (0.7) between CMA and HML suggests that these two factors are closely related. It is also found that the performance of the five-factor model is not sensitive to the way the factors are defined, i.e. the sorting procedure used to find the portfolios that are used to proxy the factors does not impact the model’s performance.

2.1.3 The Momentum Factor

In the Carhart [2] four-factor model a momentum factor is added to the Fama-French three-factor model. The momentum factor in [2] is defined as the average return of stocks belonging to the 70\textsuperscript{th} percentile of eleven-month returns lagged one month (i.e. at month \( t \) the momentum of a stock is its return over the period \( t - 12 \) to \( t - 1 \)) minus the average return of the stocks with the lowest 30\% momentum.

The motivation behind adding this factor is the evidence in Jegadeesh and Titman [25], where significant positive returns over periods up to one year are found when one buys stocks with a good past performance and sell stocks that have performed poorly in the past. Such a momentum effect suggests that there are variations in the average returns of stocks that are explained by a momentum factor. The data set used in [25] is from the U.S. market and covers the period July 1965-December 1989. A more recent study showing the existence of momentum effects are found in [26] where momentum effects due to both returns momentum and post-earnings announcement momentum are found to exist in the Swedish equity market, and the effects are significant when the risk factors in the Fama-French three-factor model are taken into account. The data set used in [26] covers the period January 1990-June 2005.
In the paper introducing their five-factor model [3], Fama and French note that adding a momentum factor in their five-factor model does not help model performance in their data set, with the regression slope for the momentum factor being close to zero.

2.1.4 Previous Studies of Factor Dynamics and Dependence

Two previous studies that are close to the present one are Christoffersen and Langlois [6] and Zhao et al. [7]. However, these papers use data from the U.S. market, whereas in this thesis the data comes from European markets. Furthermore, these two studies examine the factor dynamics and dependence of subsets of the six factors that are analyzed in this thesis.

In Christoffersen and Langlois [6] the dynamics and dependence of weekly log-returns of the four factors in Carhart [2], i.e. the three factors of Fama and French [1] with an added momentum factor, are investigated. The data set spans July 1963-December 2010 and comes from the U.S. market.

When examining the entire data set, excess kurtosis is reported for all four factors log-returns. Skewness is found to be most profound in the MKT and the MOM factor, where the skewness is negative for both factors. The weekly log-returns of the SMB, HML, and MOM factors are found to have significant autocorrelations.

The univariate dynamics of the factors are modelled using an AR(3) -NGARCH(1, 1) model. Two sets of models are used. One with innovations that have a normal distribution (the innovations are also commonly called driving noise and are the stochastic elements of the AR-NGARCH model), and one where the innovations have the skewed $t$ distribution of [27]. The use of the skewed $t$ distribution is found to give a better fit to the data, and the models of MKT and MOM are found to have innovation distributions with negative skewness.

For the MKT factor and the MOM factor, the leverage parameter of the NGARCH(1, 1) process is found to be significantly different from zero. For MKT it is positive and for MOM it is negative. This implies that a negative return for the MKT factor increases the variance in MKT returns more than a positive return of equal absolute value, while the the opposite is true for the MOM factor.

The largest negative correlation between the weekly factor log-returns is reported to be -0.31 for MKT and HML. The largest positive correlation is 0.05 for MKT and SMB. The correlations in daily log-returns are also re-
ported, with the largest negative correlation being -0.31 for MKT and HML. The largest positive correlation in daily log-returns is 0.07 for MOM and SMB.

The dependence structure of the four factors is further investigated using threshold correlations following the definition found in [28]. The threshold correlations of the factors are found to be far from what is implied by a fitted bivariate Gaussian copula, and indicate the need for the dependence structure of the factor log-returns to be modelled by some copula with tail dependence. The threshold correlations also indicate that there are asymmetric tail dependence in some of the factors.

To model the joint distribution of factor log-returns, the univariate innovations are connected with copulas. Three types of copulas are considered: the standard Gaussian and $t$ copulas, and the skewed $t$ copula of [29] which allows for asymmetric tail dependence. The skewed $t$ copula if found to give the best fit to data.

A dynamic approach to dependence is also implemented by the use of the dynamic conditional correlation (DCC) model of [30]. This model allows the conditional correlation matrix of a copula to change over time. The results from the use of the DCC model indicate that the factor correlations are not constant. For example, over the period 2006-2010 the conditional copula correlation matrix of the skewed $t$ copula has a correlation between the MKT and MOM factors that ranges from -0.5 to 0.5.

In Zhao et al. [7] the five factors in in the Fama-French five-factor model [3], i.e MKT, SMB, HML, CMA, and RMW, are modelled. The data set used comes from the U.S. market and covers the period January 1965-August 2017, with January 1965-December 1983 used as a training set and January 1984-December 1999 used as a testing set. The period January 2000-August 2017 is used as out-of-sample data.

Firstly, a set of 328 predictors, including ARMA-models and recurrent neural networks, are fitted to the training set for each factor, and their performance in predicting returns on the testing set is then evaluated. The best performing ones are then reduced further using principal component analysis to get a set of uncorrelated well-performing predictors for each factor. The dimension of these sets are between six and nine. These uncorrelated well-performing predictors are then combined using three different techniques, with the best performing combination method being the dynamic model averaging (DMA) introduced in [31].

These combined predictors are then used to get predictions of the monthly expected return for each factor. The dependencies of the factors are, as in Christoffersen and Langlois [6], modelled using the skewed $t$ copula of [29].
The use of an asymmetric copula is motivated by asymmetric tail dependence found by rejection of the null hypothesis that the upper and lower tail dependence coefficients are equal for all the bivariate factor log-returns. The test of this null hypothesis follows methods explained in [32].

The largest negative correlation between the factors is reported to be -0.53 for SMB and RMW. The largest positive correlation is 0.62 for CMA and HML.

To apply the copula, the univariate dynamic distributions of the daily log-returns of the factors is modelled using an AR-GJR-GARCH model with innovations following the skewed \( t \) distribution of Hansen [27]. The GJR-GARCH is similar to the NGARCH used in Christoffersen and Langlois [6] in that it allows for leverage type effects. The order of the AR-GJR-GARCH models is selected based on BIC. A five-year rolling window approach is used to refit the copula once a month.

As in Christoffersen and Langlois [6], the conditional correlation matrix of the skewed \( t \) copula is allowed to change over time to model dynamic dependencies. This is achieved using the generalized autoregressive score (GAS) introduced in [33]. The DCC model used in Christoffersen and Langlois [6] is also considered, as well as a generalization of the DCC model called the asymmetric generalized dynamic conditional correlation (AG-DCC) model [34].

### 2.2 Mathematical Background

#### 2.2.1 Skewed Student’s \( t \) distribution

The Student’s \( t \) distribution, commonly just called the \( t \) distribution, is often implemented in the modelling of financial and insurance data. The distribution has in its standard form one parameter, the degrees of freedom \( \nu \), which controls the shape of the distribution. If a random variable \( T \) follows a \( t \) distribution with \( \nu \) degrees of freedom, one can write \( T \sim t_\nu \).

This distribution can be generalized to a location-scale family by introducing parameters \( \mu \) and \( \sigma \), \( X = \mu + \sigma T \). For the random variable \( X \) one can write \( X \sim t_\nu(\mu, \sigma^2) \). The expected value of \( X \) is then \( \mu \) given that \( \nu > 1 \), otherwise the expected value is undefined. The variance of \( X \) is \( \sigma \nu / (\nu - 2) \) given that \( \nu > 2 \), otherwise the variance is undefined.

The main reason that the Student’s \( t \) distribution is common in the modelling of financial and insurance data is that this distribution has heavy tails [35], a property not formally defined but often a distribution is said to have a
heavy left tail if
\[
\lim_{x \to -\infty} \frac{F(x)}{e^{-\lambda(-x)}} = \infty \quad \forall \lambda > 0.
\]

This slow decay of tail probabilities is often empirically found in financial data. The rate of the decay of tail probability of the Student’s t distribution is controlled by \( \nu \). This is perhaps most easily illustrated by the fact that the Student’s t distribution has regularly varying tails with index \(-\nu\) [35]. For the left tail this means that
\[
\lim_{t \to -\infty} \frac{F(tx)}{F(t)} = x^{-\nu}.
\]

Since the Student’s t distribution is symmetric, a corresponding expression for the right tail has the same limit. Thus, a small \( \nu \) gives a slow decay of tail probability while larger values of \( \nu \) reduces the kurtosis of the distribution. In fact, the limit distribution of the Student’s t as \( \nu \to \infty \) is the normal distribution.

A generalization of symmetric distributions such as the Student’s t distribution was introduced in [36]. In this generalization, skewness is introduced in symmetric distributions by adding a skewness factor \( \gamma \in (0, \infty) \). Consider a unimodal probability density function which is symmetric around 0, \( f(x) = f(|x|) \). The transformation into a probability density function allowing for skewness \( f_\gamma(x) \) has the following form
\[
f_\gamma(x) = \frac{2}{\gamma + \gamma^{-1}} \left( f(\gamma^{-1}x)I_{(0,\infty)}(x) + f(\gamma x)I_{(-\infty,0)}(x) \right),
\]

where \( I_{(a,b)} \) is the indicator function for the interval \((a,b)\). Note that setting \( \gamma = 1 \) gives the symmetric original distribution \( f_\gamma(x) = f(x) \). Another thing to note is that inverting \( \gamma \) mirrors the distribution around 0
\[
f_\gamma(x) = f_{1/\gamma}(-x).
\]

That \( \gamma \) controls the skewness is best illustrated by the expression
\[
\frac{P(x \geq 0 \mid \gamma)}{P(x < 0 \mid \gamma)} = \gamma^2
\]

where one can see that if \( \gamma > 1 \) probability mass is moved to the right of 0, while \( \gamma \in (0, 1) \) moves probability mass to left.

The moments of the transformed distribution has below form and exists if and only if the corresponding moment exists for the original distribution (i.e.
when \( \gamma = 1 \)

\[
E[x^r \mid \gamma] = M_r \frac{\gamma^{r+1} + (-1)^r \gamma^{-(r+1)}}{\gamma + \gamma^{-1}},
\]

\[
M_r = 2 \int_0^\infty x^r f(x) dx.
\]

In particular, the above expressions can be used to give \( f_\gamma(x) \) unit variance by dividing with the standard deviation. In this thesis independent and identically distributed standardized skewed Student’s \( t \) distributed random variables as defined here are used in the modelling of univariate factor dynamics.

It is important to note that there are other ways to introduce skewness into the Student’s \( t \) distribution, and the method introduced in [36] is by no means the definite one. In [27] another skewed Student’s \( t \) distribution is introduced. The reason the skewed distribution of [36] was chosen in this thesis is the simple expression for \( f_\gamma(x) \) that does not change form for different values of \( \gamma \), and the simple interpretation of \( \gamma \) as a skewness parameter with the absolute value and sign of \( \gamma - 1 \) giving the amount and side of the skewness introduced.

### 2.2.2 ARMA-GARCH models

Time series of financial returns are prone to exhibit periods of different variability, a phenomenon often referred to as volatility clustering. A well-known early mentioning of how this phenomenon is manifested as clusters of returns of similar absolute value is found in [37].

A seminal contribution to the modelling of such heteroscedasticity in time series was the introduction of the Autoregressive Conditional Heteroscedasticity, ARCH(\( p \)), model by Engle [38] in 1982. This model was further developed into the Generalized Autoregressive Conditional Heteroskedastic model, GARCH(\( p, q \)), that was introduced in 1986 by Bollerslev [39]. The GARCH(\( p, q \)) model is an extension of the ARCH(\( p \)) model, where a GARCH(\( p, 0 \)) reduced to an ARCH(\( p \)).

What makes both these models powerful tools in the modelling of financial time series is that they have a constant unconditional variance, while the variance conditioned on the past is a time series model of previous events. This allows for a good model fit to empirical data exhibiting a dynamic variability.

In this thesis, a version of the GARCH model called Nonlinear-asymmetric GARCH, NGARCH, is used and will be introduced in section 2.2.3. Here the definition of the GARCH(\( p, q \)) process is first given. Using the notation of [40], a GARCH(\( p, q \)) process is defined as a solution \( \{X_t\}_{t \in \mathbb{Z}} \) to the below equations
\[ X_t = \sigma_t \epsilon_t, \quad t \in \mathbb{Z}, \]
\[ \sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i X_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2, \quad t \in \mathbb{Z}, \]

where \( \{\epsilon_t\}_{t \in \mathbb{Z}} \) is a sequence of independent and identically distributed (iid) random variables, \( \omega > 0, \quad p > 0, \quad q \geq 0, \quad \alpha_1, \ldots, \alpha_{p-1} \geq 0, \quad \alpha_p > 0, \quad \beta_1, \ldots, \beta_{q-1} \geq 0, \) and \( \beta_q > 0. \) The sequence \( \{\epsilon_t\}_{t \in \mathbb{Z}} \) is generally referred to as the driving noise sequence or the innovations of the process. Furthermore, the standard approach is to model \( \{\epsilon_t\}_{t \in \mathbb{Z}} \) as coming from a distribution with zero mean and unit variance. It is worth noting that some papers interchange \( p \) and \( q \) in the definition of a GARCH(\( p,q \)), this includes the original paper [39] introducing the model. It should also be noted that a GARCH(1,1) process is usually sufficient to model the conditional variance in financial time series data, and in practice it is therefore common to only consider the case when \( (p,q) = (1,1) \) [41].

For a GARCH(\( p,q \)) process to be a reasonable model it is natural to require that it is causal, meaning that it does not depend on future outcomes. More formally, one requires that \( \sigma_t \) should be measurable by the \( \sigma \)-algebra generated by \( \{\epsilon_{t-h}\}_{h>0} \). Luckily, for GARCH(\( p,q \)) processes causality is implied by another property called strict stationarity, meaning that if one focuses on finding strictly stationary solutions one does not need to worry about causality.

The GARCH(\( p,q \)) process is strictly stationary if

\[ ((X_{t_1}, \sigma_{t_1}), \ldots, (X_{t_N}, \sigma_{t_N})) \overset{d}{=} ((X_{t_1+h}, \sigma_{t_1+h}), \ldots, (X_{t_N+h}, \sigma_{t_N+h})) \]

for any times \( t_1, \ldots, t_N \in \mathbb{Z} \), and every integer \( h \) and positive integer \( N \). The existence of strictly stationary and unique solutions of GARCH(\( p,q \)) processes is studied in [42]. A simple and sufficient way of ensuring the existence of a unique strictly stationary solution is to require that \( \mathbb{E}[\epsilon_t^2] < \infty \) and \( \mathbb{E}[\epsilon_t^2] \sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j < 1 \), and this gives the following unconditional expected values [42] [40]

\[ \mathbb{E}[\sigma_t^2] = \frac{\omega}{1 - \mathbb{E}[\epsilon_t^2] \sum_{i=1}^{p} \alpha_i - \sum_{j=1}^{q} \beta_j}, \]
\[ \mathbb{E}[X_t^2] = \mathbb{E}[\sigma_t^2] \mathbb{E}[\epsilon_t^2]. \]
Furthermore, given these requirements it is shown in [39] and [43] that
the squares of the GARCH\((p, q)\) process, \(\{X_t^2\}_{t \in \mathbb{Z}}\) follow an ARMA equation,
and that the causality of this ARMA equation allows the GARCH\((p, q)\) process
to be expressed as an ARCH\((\infty)\) process. From this ARCH\((\infty)\) expression
one can easily see that \(\sigma_t\) is measurable with respect to the past and that the
expectation and variance of \(X_t\) conditioned on \(\{X_s \mid s < t\}\) is

\[
E[X_t \mid X_s : s < t] = E[\epsilon_t] \sigma_t, \\
\text{Var} (X_t \mid X_s : s < t) = \sigma_t^2 \text{Var} (\epsilon_t).
\]

Since \(\{X_t^2\}_{t \in \mathbb{Z}}\) follows an ARMA equation, its autocorrelation function
(ACF) will be that of an ARMA process. The ACF of ARMA processes is
derived in detail in section 3.3 in [44]. Given \(E[\epsilon_t] = 0\), which is the standard
in practical applications, it is easy to show that the ACF of \(\{X_t\}_{t \in \mathbb{Z}}\) is zero
for any lag. This leads to the very important conclusion that GARCH\((p, q)\)
processes allow one to model time series where there is autocorrelation in the
squares and absolute value of the series, but not in the raw series.

Another important property of GARCH\((p, q)\) processes from a modellers
perspective is that the moments of a strictly stationary solution can differ from
the moments of its driving noise sequence. It might seem natural to assume
that the kurtosis of \(\{X_t\}_{t \in \mathbb{Z}}\) is inherited from \(\{\epsilon_t\}_{t \in \mathbb{Z}}\). However, this is not
exactly the case, and using Jensen’s inequality one can show that the kurtosis
of a stationary solution (provided that it has a finite fourth moment) is in fact
always greater or equal to the kurtosis of the driving noise

\[
E[X_t^4] = E[\epsilon_t^4]E[\sigma_t^4] \geq E[\epsilon_t^4] (E[\sigma_t^2])^2 = \frac{E[\epsilon_t^4]}{(E[\epsilon_t^2])^2} (E[X_t^2])^2.
\]

Thus, if an empirical data set is found to have excess kurtosis it is not
necessary true that \(\{\epsilon_t\}_{t \in \mathbb{Z}}\) should be modelled as having a more heavy tailed
distribution than the normal.

Given an empirical data set of \(n\) observations that is believed to have been
generated by a GARCH\((p, q)\) process, the question becomes how one can estimate
the parameters of the model. The perhaps most common method for fitting a GARCH model to an empirical data set is (Gaussian) quasi-maximum
likelihood estimation (QMLE). In QMLE, one maximizes a likelihood function
that would be the correct if the driving noise sequence had a standard normal distribution, regardless of what distribution one has specified for the
noise. Let \( \varphi = (\omega, \alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q) \) denote the vector of parameters for a GARCH\((p, q)\) process, and let \( \Phi \) be the parameter space of parameter values that give a unique strictly stationary solution. Let \( \varphi_0 \) denote the true parameter vector of the GARCH\((p, q)\) process that has generated the data.

To apply QMLE, one first has to choose a set of starting values \( X_0, \ldots, X_{1-p}, \tilde{\sigma}_0, \ldots, \tilde{\sigma}_{1-q} \), and for \( t = 1, \ldots, n \) define the recursive expression for the volatility process

\[
\tilde{\sigma}_t^2 = \tilde{\sigma}_t^2(\varphi) = \omega + \sum_{i=1}^{p} \alpha_i X_{t-i}^2 + \sum_{j=1}^{q} \beta_j \tilde{\sigma}_{t-j}^2 .
\]

The observant reader might notice that since a set of starting values, which are guesses and thus most likely not part of the strictly stationary solution, this sequence is not strictly stationary. However, a result that is to be stated soon will show that asymptotically this is a valid approximation. The QMLE of \( \varphi_0 \) is any \( \hat{\varphi}_n \) solving

\[
\hat{\varphi}_n = \arg \max_{\varphi \in \Phi} \hat{L}_n(\varphi) = \arg \min_{\varphi \in \Phi} \frac{1}{n} \sum_{t=1}^{n} \hat{l}_t(\varphi),
\]

where

\[
\hat{l}_t(\varphi) = \frac{\epsilon_t^2}{\tilde{\sigma}_t^2} + \log \tilde{\sigma}_t^2 .
\]

Under rather weak assumptions, were the only assumptions on the distribution of the noise terms is that \( \epsilon_t^2 \) has a non-degenerate distribution and \( E[\epsilon_t^2] = 1 \) (in reality this is only for ease of interpretation and \( E[\epsilon_t^2] < \infty \) is sufficient), it is shown in Francq and Zakoïan [45] that the QMLE is strongly consistent

\[
\hat{\varphi}_n \rightarrow \varphi_0 \text{ a.s. when } n \rightarrow \infty,
\]

and that the starting values that raised alarm for giving a non-stationary sequence are of no importance asymptotically.

After adding the additional assumptions that \( \varphi_0 \) is in the interior of \( \Phi \), and that the noise sequence has a fourth moment, it can be shown that the QMLE is asymptotically normal [45]

\[
\sqrt{n} (\hat{\varphi}_n - \varphi_0) \overset{d}{\rightarrow} N \left( 0, \left( E[\epsilon_t^4] - 1 \right) J^{-1} \right),
\]
where

\[ J = E_{\varphi_0} \left[ \frac{\partial^2 l_t(\varphi_0)}{\partial \varphi \partial \varphi^T} \right] . \]

The correlation matrix, \((E[\epsilon_t^2] - 1) J^{-1}\), is calculated at \(\varphi_0\) and its exact value is thus not attainable. However, a consistent estimator can be achieved using

\[ \frac{1}{n} \sum_{i=1}^{n} \frac{\partial^2}{\partial \varphi_i \partial \varphi_j} \tilde{l}_t(\varphi_{ij}^*) \to J(i, j) \] in probability,

where \(\varphi^*\) is between \(\hat{\varphi}_n\) and \(\varphi_0\), and the sample estimate of \(E[\epsilon_t^2]\) [45][46].

These theoretical results give support for the use of QMLE when fitting GARCH\((p, q)\) processes. However, in practical applications it is often found that finding the maximum of the (quasi-)likelihood is a challenging nonlinear optimization problem, where there can be several local maxima [41].

As previously noted, a GARCH\((p, q)\) process allows one to model time series with conditional heteroscedasticity and autocorrelation in the squares. However, for many empirical data set ordinary autocorrelation is present as well. For such data a pure GARCH model might be too restrictive and a better model can be achieved by modelling the mean of the time series by an ARMA equation. The resulting process is referred to as an ARMA\((P, Q)\)-GARCH\((p, q)\) model and follows the below equations

\[ X_t = \mu + \sum_{i=1}^{P} \phi_i (X_{t-i} - \mu) + \sum_{j=1}^{Q} \theta_j \sigma_{t-j} \epsilon_{t-j} + \sigma_t \epsilon_t, \]

\[ \sigma_t^2 = \omega + \sum_{i=1}^{P} \alpha_i X_{t-i}^2 + \sum_{j=1}^{Q} \beta_j \sigma_{t-j}^2, \]

where \(t \in \mathbb{Z}\), and \(\{\epsilon_t\}_{t \in \mathbb{Z}}\) is iid and generally modelled to have zero mean and unit variance. As mentioned previously, a GARCH\((1, 1)\) process is usually sufficient to model the conditional variance in financial time series data, and in practice it is therefore common to only consider ARMA\((p, q)\)-GARCH\((1, 1)\) processes.

The introduction of a dynamic mean allows for a much wider range of applications, but a more complicated model also introduces greater difficulties in the derivation of mathematical properties such as the asymptotic distribution of estimated parameters. However, by adding additional assumptions to the assumptions required for asymptotic normality of the QMLE parameter vector for a GARCH process, it can be shown that the QMLE parameter vector for an ARMA-GARCH process is also a consistent estimator of the true \(\varphi_0\) and is asymptotically normal [45].
2.2.3 Nonlinear-asymmetric GARCH

Since the introduction of the GARCH($p$, $q$) process by Bollerslev [39] there has been a continued development of different types of GARCH models. For financial data, several successful extensions of the original GARCH process are designed to capture what is known as the leverage effect. The leverage effect describes an empirically found phenomenon where large negative returns increase variability more than positive returns of equal absolute value. Simply put, bad news tends to increase uncertainty more than good news. An early example of empirical evidence for the leverage effect in data from the U.S. market covering the period 1928-1984 is found in [47]. In the standard GARCH($p$, $q$) process, the variance as a function of $\epsilon_t$ is symmetric and thus it has no leverage effect. Finding GARCH type models that allow for a leverage effect (or an opposite effect, where positive values impact the variance more than negative ones) translates to transforming $\sigma_t(\epsilon_t)$ into an asymmetric function, $\sigma_t(\epsilon_t) \neq \sigma_t(-\epsilon_t)$.

One such asymmetric model is the Nonlinear-Asymmetric GARCH, NGARCH, process introduced in 1993 by Engle and Ng [48]. In [48] eight asymmetric GARCH and ARCH type models are compared and applied to a data set of daily Japanese stock returns covering the years 1980-1988, where strong evidence of a leverage effect is found.

The NGARCH is also nested in the parametric family of GARCH models introduced by Hentschel [49]. This family of GARCH type models is derived by first creating a simple asymmetric GARCH model and then applying a Box-Cox transformation to the conditional standard deviation, leading to a large set of GARCH models with different types of asymmetry.

The NGARCH($p$, $q$) process is defined as a solution $\{X_t\}_{t \in \mathbb{Z}}$ to the equations

$$
X_t = \sigma_t \epsilon_t \\
\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i \sigma_{t-i}^2 (\epsilon_{t-i} - \eta)^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2
$$

where $\{\epsilon_t\}_{t \in \mathbb{Z}}$ is a sequence of iid random variables, which commonly are defined to have mean zero and unit variance. In the original paper introducing the NGARCH process [48] the $\eta$ parameter is added to $\epsilon_{t-i}$ rather than subtracted. Here the definition in [49] is followed instead. The appendix of [49] includes a discussion of the stationarity of the nested family of GARCH models. Proof of
asymmetric normality of QMLE parameters for similar asymmetric GARCH models can be found in [46].

The NGARCH process provides a simple yet flexible parameterization of asymmetry completely controlled by the single parameter \( \eta \). When \( \eta = 0 \) the NGARCH process reduces to the standard symmetric GARCH. It is easy to see that a positive value of \( \eta \) creates a leverage effect, where negative outcomes of the driving noise makes the predicted conditional variance larger than what positive outcomes of equal size does. A negative value of \( \eta \) creates the opposite effect, where positive outcomes of the driving noise increase the variability more than negative outcomes.

### 2.2.4 Modelling Dependence with Copulas

The data to be analysed in this thesis consists of log-returns of six risk factors in a set of several European stock markets. Since the data is multivariate, it is not sufficient to only model the univariate properties of these log-returns. It is also important to model their joint distributions in order to assess their dependence. This can be achieved by the use of copulas.

Copulas allow one to specify the marginal distributions separately from the dependence structure in multivariate data. It can be viewed as introducing the dependence structure implicit in a multivariate distribution to a set of univariate distributions. For instance, one can model two variables as having different marginal \( t \) distributions, and have a dependence stemming from a bivariate normal distribution. The use of copulas thus gives great flexibility in the modelling of joint distributions.

The reasoning behind copulas relies on the use of the probability and quantile transforms [35]. The probability transform says that if a random variable \( X \) has cumulative distribution function \( F_X \), and \( F_X \) is continuous, then

\[
F_X(X) \overset{d}{=} U(0,1).
\]

The quantile transform is the reverse statement. If \( Y \) is a random variable with continuous cumulative distribution function \( F_Y \), and \( U \) is uniformly distributed on the interval \([0, 1]\), then

\[
F_Y^{-1}(U) \overset{d}{=} Y.
\]

By applying the probability transform on the components of an \( n \)-dimensional random vector \( \mathbf{X} = (X_1, ..., X_n) \), one obtains a vector of \( n \) uniformly distributed random variables with a dependence inherited from the distribution of \( \mathbf{X} \). The quantile transform can then be used to introduce this
dependence structure into a set of random variables \(Y_1, \ldots, Y_n\) with specified univariate distributions.

A copula is the joint cumulative distribution function \(C : [0, 1]^n \to [0, 1]\) of a random vector \(U\) whose components are uniformly distributed.

\[
C(u_1, \ldots, u_n) = P(U_1 \leq u_1, \ldots, U_n \leq u_n), \quad (u_1, \ldots, u_n) \in [0, 1]^n.
\]

If one for example performs the probability transform on the components of a random vector \(X\) with continuous marginal distributions to create a random vector \(U\) with uniformly distributed components, then the joint cumulative distribution function of \(U\) is called the copula of \(X\). A key result in the theoretical framework for multivariate modelling with copulas is Sklar’s theorem\(^{[50]}\), which states that for any \(n\)-dimensional multivariate cumulative distribution function \(F\) with univariate marginal distributions \(F_1, \ldots, F_n\), there exits a copula \(C\) such that \(F\) can be decomposed into \(F_1, \ldots, F_n\) and \(C\).

In this thesis, two types of copulas are considered. The Gaussian (normal) copula and the \(t\) copula. Both copulas are invariant under standardization of their marginal distributions, and thus one only has to consider the copulas of the standard multivariate Gaussian and \(t\) distributions\(^{[29]}\). The Gaussian copula \(C^G_R\) is the cumulative distribution function of the random vector \((\Phi(X_1), \ldots, \Phi(X_n))\), where \(X \sim N_n(0, R)\) and \(\Phi\) is the cumulative distribution function of the univariate standard normal distribution,

\[
C^G_R(u_1, \ldots, u_n) = P(\Phi(X_1) \leq u_1, \ldots, \Phi(X_n) \leq u_n) = \Phi^n_R(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_n)).
\]

Likewise, the \(t\) copula \(C^t_{\nu,R}\) formed by the \(n\)-dimensional standard Student’s \(t\) distribution with correlation matrix \(R\) and degrees of freedom \(\nu\) is the cumulative distribution function of the random variable \((t_{\nu}(X_1), \ldots, t_{\nu}(X_n))\), where \(X \sim t_{\nu,R}^n\) and \(t_{\nu}\) is the univariate Student’s \(t\) distribution with \(\nu\) degrees of freedom,

\[
C^t_{\nu,R}(u_1, \ldots, u_n) = P(t_{\nu}(X_1) \leq u_1, \ldots, t_{\nu}(X_n) \leq u_n) = t_{\nu,R}^n(t_{\nu}^{-1}(u_1), \ldots, t_{\nu}^{-1}(u_n)).
\]

Analogous to the univariate \(t\) distribution, \(\nu > 2\) is required for the correlation matrix of a multivariate \(t\)-distribution to be defined. However, the distribution can also be defined for \(\nu \leq 2\), and \(R\) is usually then also called the correlation matrix.
Both the multivariate normal distribution and the multivariate Student’s $t$ distribution are elliptical distributions. There are some properties of elliptical distributions that are worth noting. Some can be very beneficial when one uses such distributions to form copulas to model dependence, while some can be problematic. Firstly, all elliptical distributions are symmetric, and thus any tail dependence will be symmetric.

Tail dependence is a measure of how dependent a pair of random variables, $(X_1, X_2)$, are in their extremes. Often one looks at the coefficient of upper tail dependence, $\lambda_u$, and the coefficient of lower tail dependence, $\lambda_l$,

\[
\lambda_u = \lim_{x \to \infty} P(X_2 \geq x \mid X_1 \geq x), \\
\lambda_l = \lim_{x \to -\infty} P(X_2 \leq x \mid X_1 \leq x).
\]

These coefficients thus correspond to the limiting conditional probability of both random variables exceeding a certain extreme given that one does so. If such a limit is strictly positive, one says that $X_1$ and $X_2$ are asymptotically dependent in the corresponding tail. For elliptical distributions symmetry implies that $\lambda_u = \lambda_l$, and the coefficient of upper and lower tail dependence can simply be denoted the coefficient of tail dependence, $\lambda$. This symmetry can be problematic when one models dependence in financial data with copulas formed by elliptical distributions, since there are studies that find asymmetric tail dependence in financial return data [28].

A bivariate Gaussian copula $C^{Ga}_\rho$ with correlation $|\rho| < 1$ has zero tail dependence, while a bivariate $t$ copula gives the following [29] coefficient of tail dependence

\[
\lambda = 2t_{\nu+1} \left( -\sqrt{\nu + 1} \frac{\sqrt{1 - \rho}}{1 + \rho} \right).
\]

This formula shows that the tail dependence decreases as $\nu$ increases. In fact, analogous to the univariate case, the multivariate $t$ distribution reduces to the multivariate normal distribution, and thus loses all tail dependence, when $\nu \to \infty$. A general expression for the coefficient of tail dependence for elliptical distributions can be found in [51].

A beneficial result for elliptical distributions is the simple relation between Kendall’s tau and the (linear) correlation. Kendall’s tau is a measure of rank correlation and for the random vector $(X_1, X_2)$ it is defined [35] as

\[
\tau(X_1, X_2) = P((X_1 - X'_1)(X_2 - X'_2) > 0) - P((X_1 - X'_1)(X_2 - X'_2) < 0)
\]
where \((X'_1, X'_2)\) is an independent copy of \((X_1, X_2)\). A noteworthy result for Kendall’s tau is that it does not depend only on the copula of the joint distribution of \((X_1, X_2)\), and not on the marginal distributions of \(X_1\) and \(X_2\) [29].

For elliptical distributions, Kendall’s tau has the below simple relation to the correlation \(\rho\) between \(X_1\) and \(X_2\) [52]

\[
\tau(X_1, X_2) = \frac{2}{\pi} \arcsin \rho.
\]

From this equation one can transform an estimate of Kendall’s tau to an estimate of the correlation. This is a very favorable results since simulations indicate that for elliptical distributions the sample estimate of Kendall’s tau performs better than the sample estimate of the correlation [52] [35]. Using the Kendall’s tau transform, one can thus get a better estimate of the correlation matrix of a data set from an elliptical distribution.

Generally, maximum likelihood is used to estimate the parameters of a copula. However, for elliptical distributions one can use a method-of-moments approach, where the favorable properties of the Kendall’s tau measure allows one to estimate the correlation matrix of the copula, and any remaining parameters can then be estimated by maximum likelihood [29][52][35]. This approach is shown in [53] to work well.

For all the flexibility that copulas allow in specifying joint distributions, there are two properties often found in the dependence in empirical data that common models for copulas are unable to capture.

Firstly, the dependence structure can differ between different variables in multivariate data. In a three-dimensional data sample from a random vector \(X = (X_1, X_2, X_3)\), one might find that the dependence between \(X_1\) and \(X_2\) is best modelled by a \(t\) copula with degree of freedom \(\nu = 3\), while the corresponding best fit for variables \(X_2\) and \(X_3\) is a \(t\) copula with degree of freedom \(\nu = 10\). If one tries to model the dependence in this data with a three-dimensional \(t\) copula one can only specify one value for \(\nu\), and the resulting fit is therefore likely to be rather poor. This problem of incorrectly assuming one copula to match the dependence for several variables can be avoided by combining several copulas. A comprehensive study of such pair-copula constructions can be found in Aas et al. [54].

The second property commonly found in empirical data that a standard copula cannot model is non-constant correlation. In financial data it is often the case that the correlation of two variables is dynamic, and during different periods of time the correlation can differ greatly. A well-known study of this
phenomenon is found in Engle [30], where the Dynamic Conditional Correlation (DCC) model for time-varying correlations is introduced. Another model for changing correlations in time series data is the Generalized Autoregressive Score (GAS) model proposed in Creal, Koopman, and Lucas [33].

In [55] several statistical tests for the goodness-of-fit of copulas are proposed and tested. The results of these tests are very promising. However, these tests are very computationally heavy, and are therefore not feasible when dealing with the data sets considered in this thesis.

\section*{2.2.5 Risk Measures}

One of the main motivations for developing models of the dynamics of asset prices is the need to estimate the risk associated with holding a certain position. This raises the question of how one in a very general setting can best measure financial risk. This can of course be a very subjective question, and many alternatives exist. The estimated variance, or volatility, has historically often been used to quantify the riskiness of a portfolio. The classic examples are found in the Sharpe ratio and in the Markowitz model with its mean-variance portfolio optimization. However, variance has three main drawbacks as a risk measure. Firstly, it does not differentiate between positive and negative outcomes. Secondly, the variance is not easily interpreted in monetary units. Knowing the variance of a portfolio does not immediately give an estimation of possible capital losses. Finally, looking only at the second order properties of a distribution can also be very misleading if one does not keep in mind that higher order properties give distributions very different properties. For a simple example, let us assume that we have been asked to assess the one-day ahead risk of the (excess) market return, the $\text{MKT}$ factor, using the data set of 10 years of daily returns (not the log-returns) that is used in this thesis and will be introduced in section 3.1.

If today is $t$ and we assume the return for MKT tomorrow $r_{t+1}$ to be normally distributed and use maximum likelihood to fit a normal distribution to the entire 10-year data set we get the parameters $(\hat{\mu}, \hat{\sigma}) = (0.029\%, 1.1\%)$. The probability of a loss greater or equal to 5%, $P(r_{t+1} \leq -0.05)$, is then equal to $2.4 \cdot 10^{-6}$. If we instead assume a location-scale Student’s $t$ distribution and again use maximum likelihood to estimate the parameters we get $(\hat{\mu}, \hat{\sigma}, \hat{\nu}) = (0.051\%, 0.74\%, 3.3)$, and the corresponding probability of a loss greater or equal to 5% is then equal to $2.4 \cdot 10^{-3}$. Note that the standard deviation of the $t$ distribution is not equal to $\hat{\sigma} = 0.74\%$, but $\hat{\sigma} \sqrt{\hat{\nu} / (\hat{\nu} - 2)} = 1.9\%$, i.e. the two fitted distributions have no greater difference in variance, but they
give a factor $10^3$ difference in the probability of one-day losses greater than 5%.

To get some perspective of the different risk levels these distributions imply, one can view all $n$ future daily returns as Bernoulli trials with $p = P(r_{t+k} \leq -0.05)$ where $k = 1, 2, ..., n$. The number of days until a loss greater or equal to 5% will then have a geometric distribution with expected value $1/p$ days. In the data set for the MKT factor, each year has 260-261 business days, and thus the assumption of normality gives us that a one-day loss greater or equal to 5% is expect to happen roughly once every 1,600 years, whereas the Student’s $t$ distribution imply that the frequency is once every 1.7 years. In the 10-year period of data for MKT factor, one-day losses greater or equal to 5% occurs three times.

Two risk measures that have been developed to overcome the shortcomings of using variance as a risk measure are Value-at-Risk, $\text{VaR}_p(X)$ and Expected Shortfall, $\text{ES}_p(X)$. The main benefit of these two risk measures is that they in one number summarize the risk associated with the distribution of the future value of a portfolio. Another benefit is that both risk measures are given in monetary units, and can be interpreted as the minimum amount of capital invested in risk-free assets needed to ensure that a portfolio is acceptable from a risk perspective.

The Value-at-Risk (VaR) at level $p \in (0, 1)$ of a portfolio with value $X$ at a future time 1 is defined as

$$\text{VaR}_p = \min \{m : P(mR_0 + X < 0) \leq p \}$$

where $R_0$ is the risk-free rate of return. This means that the VaR of a portfolio with value $X$ at time 1 is the smallest amount of capital needed to be invested in a risk-free asset to ensure that the probability of a negative position at time 1 is not greater than $p$. The general view taken is to imagine that at time 0 one takes a risk-free loan corresponding the current portfolio value $V_0$, and uses this capital to buy the portfolio. The net value at time 1 is then $X = V_1 - V_0R_0$. By defining the discounted loss to be $L = -X/R_0$, the $\text{VaR}_p(X)$ can be rewritten as the $(1 - p)$-quantile of $L$

$$\text{VaR}_p(X) = F_L^{-1}(1 - p).$$

This formulation of VaR as a quantile function is beneficial both for the intuitive understanding of VaR, and for calculation of VaR. The following two propositions provide helpful methods for the calculation of quantile function in different settings [35].
Proposition 2.1. If $g : \mathbb{R} \to \mathbb{R}$ is nondecreasing and left continuous, then for any random variable $X$ it holds that $F_{g(X)}^{-1}(p) = g \left( F_X^{-1}(p) \right)$ for all $p \in (0, 1)$.

Proposition 2.2. For any random variable $X$ with continuous cumulative distribution function $F_X$ it holds that $F_X^{-1}(p) = -F_X^{-1}(1-p)$ for all $p \in (0, 1)$.

Three properties that have natural interpretations for any risk measure $\rho(X)$ that gives the minimum amount of money needed to be invested in risk-free assets to make the portfolio acceptable, and hold for VaR are listed below.

1. **Translation invariance** $\rho(X + cR_0) = \rho(X) - c, \quad \forall c \in \mathbb{R}$

2. **Monotonicity** $X_1 \leq X_0 \implies \rho(X_1) \leq \rho(X_0)$

3. **Positive homogeneity** $\rho(\lambda X) = \lambda \rho(X), \quad \forall \lambda \geq 0$

For a risk measure with translation invariance it holds that an investment of $c$ in a risk-free asset will reduce the risk by the same amount. If a risk measure is translation invariant and monotone it is called a monetary measure of risk [35]. Positive homogeneity simply states that doubling the position in a risky asset will double the risk, and also implies that the risk measure has the normalization property, $\rho(0) = 0$.

A drawback with the VaR as a measure of the risk associated with a financial asset $X$ is that it ignores most of the left tail of $X$. It does not give any information as to how severe the losses beyond the level $p$ can get. The risk measure Expected shortfall (ES), also known as Conditional VaR (CVaR), is the average VaR below a level $p$, and thus provides a summary of the entire left tail. The ES at level $p$ is defined as

$$\text{ES}_p(X) = \frac{1}{p} \int_0^p \text{VaR}_u(X) \, du.$$  

If $X$ has a continuous distribution function then the ES at level $p$ has the alternative representation

$$\text{ES}_p(X) = \mathbb{E}[L \mid L \geq F_L^{-1}(1-p)]$$

which motivates calling the risk measure the expected shortfall.

Besides the three properties discussed earlier that hold for the VaR, the ES also has a fourth property that can be beneficial for a risk measure.

4. **Subadditivity** $\rho(X_1 + X_0) \leq \rho(X_1) + \rho(X_0)$

This property means that potential benefits of diversification is considered. A risk measure for which all these four properties hold, such as ES, is called a coherent risk measure [35].
2.2.6 Backtesting Predicted Value-at-Risk

As pointed out in section 2.2.5, one of the main motivations for finding models for the dynamics and dependence of the factors considered in this thesis, or any financial asset, is the need for reliable estimations of risk measures. After finding such models it therefore of interest to test how well the models estimate risk measures in new data not used when fitting. This can be done by backtesting the estimated risk measure, where the model is used to predict the risk measure for several periods in new data, and these estimated values of the risk measure are then examined to see how well they align with what was realized in the data.

The Value-at-Risk is a risk measure that is particularly well suited for backtesting since correct predictions of this risk measure has implications that are easy and intuitive to test. If a predicted one-day ahead VaR at level \( p \) is correct, then the probability that the following day results in a loss exceeding this predicted VaR should be equal to \( p \). Such an event is referred to as an exceedance or a violation of the VaR. The first step for testing VaR estimations of a model, is therefore to define a hit sequence of such exceedances

\[
I_t = \begin{cases} 
1, & \text{if } L_t > \text{VaR}_p(X_t) \\
0, & \text{else}
\end{cases}
\]

If the model one uses gives the true VaR at level \( p \), then this hit sequence is an iid Bernoulli\((p)\) sequence. For such a series, both the unconditional expectation of \( I_t \), \( E[I_t] \), and the expectation conditional on the previous outcomes \( E[I_t | \psi_{t-1}] \), with \( \psi_{t-1} = \{I_x\}_{x=1}^{t-1} \), is equal to \( p \). A backtesting procedure for the predicted one-day ahead VaR should therefore check these two properties.

To test the null hypothesis that \( E[I_t] = p \) one can compare the number of realized VaR exceedances during a test period to what would be expected given the null. This is referred to as checking for correct unconditional coverage. The sum of \( n \) Bernoulli\((p)\) distributed random variables has a binomial distribution, \( B(n, p) \). Thus, the null can be tested by comparing the number of realized VaR exceedances during a test period to the corresponding \( B(n, p) \) distribution (with \( n \) being the number of days in the forecasted period) [56].

To test whether the probability of an exceedance at time \( t \) is still \( p \) when conditioned on previous outcomes, \( \psi_{t-1} \), the duration test of Christoffersen and Pelletier [57] can be used. Let \( t_i \) denote the day of exceedance number \( i \), and define \( D_i \) to be the number of days between exceedance \( i \) and \( i - 1 \).
\[ D_i = t_i - t_{i-1} \]

Under the null hypothesis of \( I_t \) being an iid Bernoulli sequence, the durations \( D_i \) are memory-less. The probability of an exceedance tomorrow should be \( p \), and this should not change if the last exceedance was recent or if it occurred a long time ago. This is equivalent to expecting \( D_i \) to follow a geometric distribution since it is the only memory-less discrete distribution.

To investigate the properties of durations of time between events for data from a particular distribution, one can investigate the hazard function \( \lambda(x) \) of said distribution. The hazard function is defined as being the probability of an event happening at a time \( t \), given that there has not been any event leading up to time \( t \). For the geometric distribution, where the time pasted since the last event does not influence the current probability of an event, the hazard function is flat.

\[
\lambda_G(d) = \frac{(1 - p)^{d-1}p}{1 - \sum_{i=0}^{d-2}(1 - p)^i p} = \frac{(1 - p)^{d-1}p}{(1 - p)^{d-1}} = p
\]

The continuous analogue of the geometric distribution is the exponential distribution. Indeed, these two distributions are the only memory-less distributions, and the exponential distribution can be viewed as the limiting distribution of the geometric distribution. The null hypothesis of correct conditional coverage can thus be formalized as testing if the durations follow an exponential distribution.

In order to test whether the durations are exponentially distributed one can make use of the more general Weibull distribution

\[
f_W(d; a, b) = \begin{cases} 
\frac{b}{a} \left( \frac{d}{a} \right)^{b-1} e^{-(d/a)^b}, & d \geq 0, \\
0, & d < 0,
\end{cases}
\]

with \( a > 0, b > 0 \). The Weibull distribution reduces to an exponential distribution when \( b = 1 \), and this is thus a testable null hypothesis. Other values of
$b$ will give different types of duration properties. The hazard function of the Weibull distribution has the following form

$$
\lambda_W(d) = \frac{b}{a} d^{b-1}.
$$

The parameter $b$ controls the shape of the hazard function, and as mentioned $b = 1$ corresponds to the flat hazard function of an exponential distribution, while $b < 1$ corresponds to a decreasing hazard function, indicating an excessive number of very short and very long durations. If $b > 1$, then the hazard function is increasing, indicating that the probability of an exceedance increases as the time since the last exceedance increases.

To test the null hypothesis $H_0 : b = b_0 = 1$ a likelihood ratio test is implemented where the maximum likelihood estimate of $b$, $\hat{b}$, is compared to the null, $b_0 = 1$. To implement the test, the log-likelihood is calculated by the below expression, where $D_1$, the number of days until the first exceedance, is left-censored, and $D_N$, the number of days between the last exceedance and the last point in time in the testing period, is right-censored. For these two durations one can only say that the no-hit duration lasted at least $D_1$ and $D_N$ days, and their contribution to the likelihood is given by the survival function $1 - F(D_i)$.

$$
l(D; b) = \ln (1 - F(D_1; b)) + \sum_{i=2}^{N-1} \ln f(D_i; b) + \ln (1 - F(D_N; b))
$$

Note that the parameter $a$ is not included in this expression. This is because for a given value of $b$, the first-order condition $\frac{d}{d a} l(D; a, b) = 0$ for a maximum likelihood estimated parameter reduces to

$$
\hat{a} = \left( \frac{N - 2}{\sum_{i=1}^{N} D_i^b} \right)^b
$$

and thus the numerical maximum likelihood estimation can be done over only the $b$ parameter. The maximum likelihood estimate $\hat{b}$, is compared to the null, $b_0 = 1$, by the likelihood-ratio test, where $-2[l(b_0) - l(\hat{b})]$ is asymptotically $\chi^2(1)$ distributed.
Chapter 3

Methods

3.1 Data

The data used in this thesis comes from the Data Library of Kenneth R. French and was collected in October 2019. It consists of ten years of daily returns for the six factor portfolios introduced in sections 2.1.2 and 2.1.3 on a market of sixteen European countries\(^1\). The starting date of the daily returns is July 31, 2009 and the end date is July 31, 2019. Each year has between 260 and 261 observations and the total number of observations is \(n = 2,609\).

The returns are in U.S. dollars and the excess market return factor is the return of a region’s value-weight market portfolio minus the U.S. one month T-bill rate. The excess market return factor will be denoted MKT in this thesis to have all six factors uniformly named. The four other Fama-French factors are constructed following the reasoning explained in section 2.1.2. The stocks in a region are first sorted into two market capitalization groups and three groups each for book-to-market equity, operating profitability, and investment. This is the 2x3 approach of [3], and gives three sets of six value-weight portfolios formed on: 1. size and book-to-market. 2. size and operating profitability. 3. size and investment. This sorting is done yearly at the end of June.

The limits for the size sorting in a region are given by the top 90% of the market capitalization, and the bottom 10% (note that in section 2.1.2 the median market capitalization was used instead of the 10th and 90th percentiles). These two groups are called Big (B) and Small (S). The sorting limits for the book-to-market, operating profitability, and investment in a region are the 30th

---

\(^1\)The countries in alphabetical order: Austria, Belgium, Denmark, Finland, France, Germany, Great Britain, Greece, Ireland, Italy, Netherlands, Norway, Portugal, Spain, Sweden, Switzerland.
and 70th percentiles.

When the stocks have been sorted, the resulting portfolios are combined to proxy the four Fama-French factors (all factors besides MKT) according to their definitions in section 2.1.2.

The momentum factor, MOM, is formed daily using size and the lagged momentum. At the end of day \( t - 1 \) the lagged momentum of a stock is its cumulative return between day \( t - 250 \) to day \( t - 20 \). This correspond to the eleven-month returns lagged one month as formulated in [2]. The breakpoints for the lagged momentum in a region are the 30th and 70th percentiles of the lagged momentum returns of the big stocks. The corresponding three groups are called Losers (L), Neutral (N), and Winners (W). A 2x3 sort on size and momentum gives six value-weight portfolios. The MOM factor is the average return of the two winner portfolios minus the average return of the two loser portfolios.

\[
\text{MOM} = \frac{SW + BW}{2} - \frac{SL + BL}{2}
\]

The raw data of returns for all six factors was transformed into log-returns, and thus the returns modelled in this thesis are log-returns. All data analysis was performed in MATLAB and R. In R the packages \texttt{rugarch} [58], \texttt{copula} [59], and \texttt{WeightedPortTest} [60], were used.

### 3.2 Univariate Modelling

In order to assess the properties of the factor log-returns, the first step in the univariate modelling was to test the factor log-returns for autocorrelation, conditional heteroscedasticity, and a non-normal distribution.

The autocorrelations of the factor log-returns were investigated by the use of the Ljung-Box statistic and the autocorrelation function (ACF) plot. Heteroscedasticity was examined by the Weighted McLeod-Li statistic introduced in [61], which tests for nonlinear GARCH-type effects in data, and the ACF plot of the squared log-returns. The distributional properties of the log-returns was visually inspected by normal QQ-plots.

The modelling of the daily log-returns of the factors was conducted by fitting ARMA(\( p, q \))-NGARCH(1, 1) processes to the training data of each factor

\[
\begin{align*}
    r_t &= \mu + \sum_{i=1}^{p} \phi_i (r_{t-i} - \mu) + \sum_{j=1}^{q} \theta_j \sigma_{t-j} \epsilon_{t-j} + \sigma_t \epsilon_t, \\
    \sigma_t^2 &= \omega + \alpha \sigma_{t-1}^2 (\epsilon_{t-1} - \eta)^2 + \beta \sigma_{t-1}^2.
\end{align*}
\]
with the driving noise, $\epsilon_t$, being iid with mean zero and unit variance. The driving noise was modelled as either having a standard normal distribution or a standardized skewed $t$ distribution as described in section 2.2.1. Furthermore, for each of these assumed distributions, a set of 9 different ARMA orders were fitted, $(p, q) \in \{0, 1, 2\}^2$.

A rolling window approach was used where a model was fitted using the past five years of data, and then tested on the data for the following year. This approach simulates a yearly refitting of the models using data from the past five years. Since 10 years of data was used in total, there were five periods of training and testing data. These are summarized in Table 3.1. The one year testing periods were $n = 261$ days for period 1, 2, 4, and 5, and $n = 260$ days for period 3. The five year training periods were $n = 1,305$ days for period 1, 4, and 5, and $n = 1,306$ days for periods 2 and 3.

All models were fitted using QMLE with a set of algorithms for nonlinear optimization with randomized starting values. For each of the two distributions of the driving noise, a best order of the ARMA model was selected based on BIC. The choice of BIC as model selection criteria was influenced by the use of BIC for model selection in [7]. Furthermore, AIC was also considered, but it was found that in most cases the best model by AIC was also the best model by BIC, and to minimize the risk of overfitting it was decided to follow BIC which punishes additional parameters more than AIC.

The best model by BIC with normally distributed driving noise and the best model by BIC with skewed $t$ distributed driving noise were then compared by residual diagnostics, i.e. by testing if the implied driving noise of the estimated models had the specified distribution and was iid. The autocorrelation of the residuals $\{\hat{\epsilon}_t\}$ was investigated by an inspection of the ACF plot. Similarly, the remains of any GARCH-type nonlinear effects was tested using an ACF plot of the squared residuals.

<table>
<thead>
<tr>
<th>Period</th>
<th>Training Period</th>
<th>Testing Period</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>July 31, 2012 - July 31, 2017</td>
<td>Aug. 1, 2017 - July 31, 2018</td>
</tr>
<tr>
<td>5</td>
<td>July 31, 2013 - July 31, 2018</td>
<td>Aug. 1, 2018 - July 31, 2019</td>
</tr>
</tbody>
</table>

Table 3.1: The five periods of training and testing data used.
The distribution assumption on $\hat{\epsilon}_t$ was tested by an inspection of the QQ-plot, where the quantiles of the residuals was plotted against the corresponding quantiles of the assumed normal or skewed $t$ distribution. The Kolmogorov-Smirnov and Anderson-Darling statistical tests were also considered, but the simple QQ-plot was deemed the best method of verifying or rejecting the assumed driving noise distribution.

To test the models of the factor log-returns in new data, the one-day ahead VaR for the factors was estimated for each day of the five one-year testing periods. The predictions of the VaR was backtested using the methods outlined in section 2.2.6. It should be noted that these VaR predictions were made as if one had taken a long position directly in each of these factors. This is a purely theoretical endeavor. Nonetheless, making and backtesting VaR predictions constitutes a good test of how well the ARMA($p$, $q$)-NGARCH($1$, $1$) models can predict the distributions of the one-day ahead log-returns in real data.

### 3.3 Dependence Modelling

In the multivariate modelling, the training and testing periods in Table 3.1 were again used to simulate the same rolling-window approach, were five years of data is used for model estimation, and the fitted models used to make one-day ahead predictions during the next year.

To model the joint distribution of the log-returns of the factors, univariate models for each factors were first specified. Dependence in the driving noise sequence for the factors was then introduced by the use of Gaussian and $t$ copulas. This approach allows one to create a joint distribution for the log-returns of factors without interfering with the models that capture their univariate dynamics. The same copula-based joint distributional modelling is used in [6], and a comprehensive overview of this type of modelling can be found in [32].

To make the modelling more straightforward and for ease of interpretation it was deemed necessary to specify the same model for the univariate dynamics of all six factor for all periods, rather than selecting univariate models by BIC as described in section 3.2.

The results of the univariate modelling suggested that an ARMA($1$, $1$)-NGARCH($1$, $1$) model with skewed $t$-distributed driving noise should be able to capture the univariate dynamics of all six factors in all periods reasonable well.

The first step in the multivariate modelling for each period was thus to fit ARMA($1$, $1$)-NGARCH($1$, $1$) models with skewed $t$ distributed driving noise to the log-returns in the training data of each factor. With these models fitted,
the implied driving noise sequences for each factor $i = 1, 2, ..., 6$ were formed by taking the realized log-returns minus the corresponding ARMA-implied conditional mean and dividing by the NGARCH-implied conditional standard deviation

$$
\hat{\epsilon}_{it} = \frac{r_{it} - \hat{\mu}_{it}}{\hat{\sigma}_{it}}.
$$

Using the skewed $t$ distributions, the probability transform was then used to create $U(0, 1)$-samples from the driving noise sequence for each factor. Gaussian and $t$ copulas were then fitted to these uniformly distributed samples by the Kendall’s tau approach explained in section 2.2.4.

The joint distribution for the log-returns were then approximated by simulation. First, a large number ($n = 10^6$) of random samples from the fitted copula were drawn. These were then transformed to samples of the joint distribution of the driving noise terms by the use of the quantile transform with the skewed $t$ distribution. With a joint distribution for the driving noise terms found, the joint distribution of the log-returns is simply found from the conditional mean and variance implied by the fitted ARMA(1, 1)-NGARCH(1, 1) model for each factor.

To test the modelled joint distribution of log-returns, the one-day ahead VaR for portfolios consisting of several factors was estimated for each day of the five one-year testing periods. The predicted VaR was backtested using the same procedure as in the univariate case.

The portfolios used for testing the predicted VaR were equally weighted $(1/N)$ portfolios. These were refitted once a week (every five days). As mentioned before, simulation size for the estimation of the joint distributions of the driving noise terms was $n = 10^6$, and the VaR was then estimated by the empirical VaR of the simulated joint distribution of log-returns.
Chapter 4

Results

4.1 Factor Dynamics

4.1.1 Factor Properties

When reviewing previous [6] [7] research on the dynamics of daily log-returns of the five Fama-French factors and the momentum factor in U.S. market data, it is clear that the data used in this thesis, presented in section 3.1, should be checked for three main properties. The properties in question are autocorrelation, heteroscedasticity, and a non-normal distribution.

The autocorrelation of the log-returns data was investigated using the Ljung-Box test statistic, $Q_{LB}(h)$ with lags $h = 3$ and $h = 10$ to test the null hypothesis that there is no autocorrelation in the log-returns. Furthermore, sample ACF plots were constructed to graphically examine the autocorrelations in the data.

To check for heteroscedasticity in the log-returns, the Weighted McLeod-Li test of [61] was implemented using lag $h = 20$. This test is conducted by first fitting an ARMA$(p, q)$ with $(p, q) \in \{0, 1, 2, ..., 5\}^2$ using AIC as model selection criteria. The squared residuals of this fit are then tested by a weighted Ljung-Box test statistic, $Q_{WML}(h)$, to test the null hypothesis that there is no correlation in the squared residuals. The sample ACF of the squared log-returns was also investigated and plotted together with the sample ACF of the log-returns. As explained in section 2.2.2, a GARCH process with a zero mean driving noise sequence has zero autocorrelation in its raw form, while the autocorrelations of the squared series is that of an ARMA process. The distributions of the log-returns were studied by their sample variance, excess kurtosis, and skewness. Furthermore, QQ-plots of the quantiles of standard-
ized log-returns against the quantiles of a standard normal distribution were constructed to graphically examine the distributional properties of the log-returns.

In Figure 4.1 the visual analysis of the MKT, SMB, and HML factors for the testing data of period 1 are summarized in six plots. The left plots in the figure are the normal QQ-plots. The middle plots are the time series of log-returns. The right plots are the sample ACF of log-returns and squared log-returns. The corresponding plots for testing periods 2-5 were very similar and are therefore not presented.

Figure 4.1: Analysis of raw data in period 1. The left plots are QQ-plots of standardized log-returns against N(0,1). The middle plots are the time series of log-returns. The right plots are the ACF of the log-returns (solid) and the squared log-returns (dotted) with 95% C.I. for iid (blue dotted horizontal). Top-down: MKT, SMB, HML.

The normal QQ-plots for all three factors indicate that the distributions of the log-returns have heavier tails than the normal distribution. The plotted time series of log-returns indicate a larger variability in the log-returns of the MKT factor. Furthermore, from a visual inspection, all three time series seem
to have periods of different variability. The sample ACF plots indicate that log-returns of the MKT factor lacks autocorrelation, or has a small negative autocorrelation at lag 3. However, the squared log-returns of the MKT factor are highly correlated. For the SMB factor, the ACF plot indicates that there is a negative autocorrelation at lag one, and a high correlation for the squared log-returns. The HML factor seems to have a significant positive autocorrelation at lag 1, and a significant, albeit smaller than for the other two factors, correlation in the the squared log-return.

In Figure 4.2 the corresponding plots to those in Figure 4.1 are presented for the remaining three factors. There are some striking differences between the factors. The normal QQ-plot for the CMA factor indicates that its distribution is close to a normal distribution. The RMW factor seems to have a heavy left tail, and the MOM factor seems to have heavy left and right tails. The time series plots indicate a very small variance in the CMA factor, and clear clustering of variability in the MOM factor.

The ACF plots indicate a high autocorrelation at lag 1 for the CMA and MOM factors, and a smaller but significant autocorrelation at lag 1 for the RMW factor. The autocorrelations for the squared log-returns are highly correlated for both the CMA and the MOM factor. However, there does not seem to be any correlation between the squared log-returns of the RMW factor, a result unique for this factor.

In Tables 4.1-4.3, the $p$-values of the Ljung-Box and Weighted McLeod-Li tests, and sample moments for all six factors for testing periods 1, 3, and 5 are presented. Presenting results for every other testing period rather than every testing period is done to limit the intersection of data. For example, testing period 1 shares 3 out of its 5 years of data with testing period 3, and 1 year with testing period 5.
Figure 4.2: Analysis of raw data in period 1. The left plots are QQ-plots of standardized log-returns against N(0,1). The middle plots are the time series of log-returns. The right plots are the ACF of the log-returns (solid) and the squared log-returns (dotted) with 95% C.I. for iid (blue dotted horizontal). Top-down: CMA, RMW, MOM.

The $p$-values of the Ljung-Box statistic indicates that the MKT factor exhibits autocorrelation in period 1 and 3, but not in period 5. The Weighted McLeod-Li statistic gives strong support for conditional heteroscedasticity in the MKT factor for all periods. From the sample moments of the MKT factor it is noteworthy that in all periods it has the highest variance of all factors. Furthermore, there is excess kurtosis in all periods and it increases from 2.54 in period 1 to 11.01 in period 5. The skewness is negative for all periods and increases in absolute value, from -0.24 in period 1 to -1.10 in period 5. The skewness of the MKT factor in period 5 is the largest negative skewness of any factor in all periods.

For the SMB factor the autocorrelation is significant in all periods, and so is the Weighted McLeod-Li test, indicating conditional heteroscedasticity. The sample moments of the SMB factor shows excess kurtosis and some negative skewness in all periods.
The HML factor has significant autocorrelation in all three periods when using lag $h = 3$ in the Ljung-Box test. However, when using lag $h = 10$ the $p$-values of the Ljung-Box statistic is only under 0.05 in period 1. The Weighted McLeod-Li test indicates significant heteroscedasticity for all three periods. The sample moments of the HML factor shows excess kurtosis and positive skewness in all periods. The skewness in period 1 is 0.212, making it the largest positive skewness of any factor in any period.

For the CMA factor the Ljung-Box test indicates that the autocorrelation is significant when using lags $h = 3$ and $h = 10$ in period 1 and 3. The same tests in period 5 indicates no significant autocorrelation. The CMA factor has some excess kurtosis. However, it is the lowest of all factors in all three periods, with a maximum of 0.846 in period 1 and a minimum of 0.827 in period 5. The skewness is slightly negative in period 1 and 3, and positive in period 5.

The $p$-values of the Ljung-Box statistic using lag $h = 3$ for the RMW factor shows significant autocorrelation for period 1 and 5, but not in period 3. When using lag $h = 10$, the autocorrelation is significant for all three periods.
The Weighted McLeod-Li test indicates significant heteroscedasticity in all three periods. However, it is worth noting that the $p$-value of $Q_{WML}(20)$ for the RMW factor in period 1 is 0.031, which is the largest for any factor in any of the three periods. The sample moments of the RMW factor shows excess kurtosis and negative skewness in all periods.

For the MOM factor, the Ljung-Box and Weighted McLeod-Li tests give strong support of autocorrelation and heteroscedasticity in all three periods. The sample moments of the MOM factor shows excess kurtosis and negative skewness in all three periods.

### 4.1.2 Modelling the Factors

The Tables 4.4-4.9 give the estimated models for each factor that was found using the model selection procedure presented in section 3.2. Each table gives the QMLE parameters, with estimated standard deviations in parenthesis. Hyphenation is used in the Tables to indicate that a certain model does not include the corresponding parameter. If the models of a factor never included a parameter, then the corresponding parameter column is not included in the table for that factor.

Furthermore, significance levels are given for the ARMA-parameters, $\phi_i$, $\theta_i$, the leverage parameter of the NGARCH(1, 1), $\eta$, and the skewness parameter of the skewed $t$ distribution, $\gamma$. The significance levels are indicated in the tables by $\ast \ast \ast$, $\ast \ast$, and $\ast$, representing significance at the 1%, 5%, and 10% level respectively. For $\phi_i$, $\theta_i$, and $\eta$, the null hypothesis is that the parameter in question is equal to zero. For $\gamma$, the null hypothesis is that it is equal to one. The significance levels are calculated using the asymptotic normality of the QMLE parameters discussed in section 2.2.2 and 2.2.3.

The models for the log-returns of the MKT factor have skewed $t$ distributed driving noise in all five periods, with a $\nu$ around 7. Furthermore, it is noteworthy that the parameter $\eta$ is positive and significant at the 1% in all periods. This is a strong indication that the MKT factor has a leverage effect, which is in agreement with the results of [6]. Furthermore, the skewness parameter, $\gamma$, is negative and significant at the 1% level in all periods. This together with the low sample skewness in Tables 4.1-4.3 indicates that the log-returns of the MKT factor has some negative skewness. For the MKT log-returns, the ACF plot in Figure 4.1 and the relatively high $p$-values of the Ljung-Box statistic in Tables 4.1-4.3, suggest that there is no real autocorrelation. The results in Table 4.4 support this, as an ARMA parameter is only included in the model for period 4, and this parameter is far from significant. Finally, it should be
Table 4.4: MKT. Top-down: estimated model parameters for period 1-5.

Table 4.5: SMB. Top-down: estimated model parameters for period 1-5.

noted that the ω parameters for the MKT log-returns are the largest of any of the models, confirming the results in section 4.1.1 suggesting that this factor has the largest unconditional variance.

It is evident in Table 4.5 that the SMB factor has the most stable log-return model, with the same model specification selected for all five periods. Furthermore, the β₁ and the η parameters are negative and significant at the 1% level for all five periods. Additionally, there seems to be no skewness in the driving noise as γ = 1 cannot be rejected in any of the models. A negative η parameter which is significant at the 1% level in all periods suggest that there is an opposite leverage effect for the SMB factor, were positive results increase variability more than negative results.

The ARMA(p, q)-NGARCH(1, 1) models for the log-returns of the HML factor in Table 4.6 are fairly stable. An AR(1) specification with a φ₁ parameter which is positive and significant at the 1% level is found in the first four
Table 4.6: HML. Top-down: estimated model parameters for period 1-5.

<table>
<thead>
<tr>
<th>(\mu)</th>
<th>(\phi_1)</th>
<th>(\omega)</th>
<th>(\alpha)</th>
<th>(\beta)</th>
<th>(\eta)</th>
<th>(\gamma)</th>
<th>(\nu)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1.364e-04)</td>
<td>(2.875e-02)</td>
<td>(5.2593e-07)</td>
<td>(1.725e-02)</td>
<td>(1.601e-02)</td>
<td>(1.210e-01)</td>
<td>(4.313e-02)</td>
<td>(2.249)</td>
</tr>
<tr>
<td>(1.239e-04)</td>
<td>(2.865e-02)</td>
<td>(2.862e-07)</td>
<td>(8.874e-03)</td>
<td>(7.043e-03)</td>
<td>(1.497e-01)</td>
<td>(1.816e-02)</td>
<td>(2.296)</td>
</tr>
<tr>
<td>(1.285e-04)</td>
<td>(2.833e-02)</td>
<td>(3.072e-07)</td>
<td>(9.471e-03)</td>
<td>(7.084e-03)</td>
<td>(1.483e-01)</td>
<td>(4.061e-02)</td>
<td>(1.642)</td>
</tr>
<tr>
<td>2.113e-05</td>
<td>1.026e-01***</td>
<td>7.529e-07</td>
<td>7.471e-02</td>
<td>8.831e-01</td>
<td>1.799e-01</td>
<td>1.074</td>
<td>7.604</td>
</tr>
<tr>
<td>(1.330e-04)</td>
<td>(2.848e-02)</td>
<td>(2.562e-06)</td>
<td>(7.460e-02)</td>
<td>(8.526e-02)</td>
<td>(1.516e-01)</td>
<td>(4.636e-02)</td>
<td>(2.572)</td>
</tr>
<tr>
<td>(1.020e-04)</td>
<td>(1.362e-06)</td>
<td>(4.521e-02)</td>
<td>(4.881e-02)</td>
<td>(1.348e-01)</td>
<td>(4.053e-02)</td>
<td>(1.818)</td>
<td></td>
</tr>
</tbody>
</table>

Table 4.7: CMA. Top down: parameters for period 1-5.

<table>
<thead>
<tr>
<th>(\mu)</th>
<th>(\phi_1)</th>
<th>(\omega)</th>
<th>(\alpha)</th>
<th>(\beta)</th>
<th>(\eta)</th>
<th>(\gamma)</th>
<th>(\nu)</th>
</tr>
</thead>
<tbody>
<tr>
<td>9.316e-05</td>
<td>1.055e-01***</td>
<td>1.137e-07</td>
<td>4.938e-02</td>
<td>9.294e-01</td>
<td>-1.740e-01</td>
<td>&quot;</td>
<td>&quot;</td>
</tr>
<tr>
<td>(7.049e-05)</td>
<td>(2.881e-02)</td>
<td>(8.324e-07)</td>
<td>(3.188e-02)</td>
<td>(3.152e-02)</td>
<td>(1.768e-01)</td>
<td>(1.962e-01)</td>
<td>(1.962e-01)</td>
</tr>
<tr>
<td>4.808e-05</td>
<td>1.068e-01***</td>
<td>1.045e-07</td>
<td>4.615e-02</td>
<td>9.320e-01</td>
<td>-2.963e-01</td>
<td>&quot;</td>
<td>&quot;</td>
</tr>
<tr>
<td>(7.149e-05)</td>
<td>(2.895e-02)</td>
<td>(8.868e-07)</td>
<td>(3.187e-02)</td>
<td>(2.188e-01)</td>
<td>(2.188e-01)</td>
<td>(2.188e-01)</td>
<td>(2.188e-01)</td>
</tr>
<tr>
<td>5.646e-05</td>
<td>8.665e-02</td>
<td>7.068e-08</td>
<td>3.8547e-02</td>
<td>9.452e-01</td>
<td>-3.791e-01**</td>
<td>1.010</td>
<td>1.279e+01</td>
</tr>
<tr>
<td>(7.215e-05)</td>
<td>(2.783e-02)***</td>
<td>(3.866e-07)</td>
<td>(6.445e-03)</td>
<td>(6.717e-03)</td>
<td>(1.797e-01)</td>
<td>(3.774e-02)</td>
<td>(5.004)</td>
</tr>
<tr>
<td>(7.017e-05)</td>
<td>(2.757e-02)</td>
<td>(3.076e-07)</td>
<td>(6.722e-03)</td>
<td>(5.514e-03)</td>
<td>(1.289e-01)</td>
<td>(3.850e-02)</td>
<td>(2.134)</td>
</tr>
<tr>
<td>(6.492e-05)</td>
<td>(4.846e-08)</td>
<td>(4.007e-03)</td>
<td>(3.787e-03)</td>
<td>(1.674e-01)</td>
<td>(3.825e-02)</td>
<td>(2.170)</td>
<td></td>
</tr>
</tbody>
</table>

periods. The results for the \(\eta\) parameter suggests that, at least during the beginning of the 10-year period analyzed, a leverage effect is present in the HML factor.

The results for the CMA factor in Table 4.7 are unique in that they include the only selected models with normally distributed driving noise sequences. However, starting in period 3, the normal distribution is replaced by the skewed \(t\) distribution, and the results suggest that the parameter \(\nu\), controlling how heavy the tails of the noise are, is decreasing from period 3 to period 5.

The results of the modelling RMW factor log-returns in Table 4.8 are quite unstable, with several different ARMA specifications selected. Furthermore, the leverage parameter \(\eta\) seems to vary greatly too.

The results in Table 4.9 suggest that the log-returns of MOM factor has a fairly stable ARMA specification. However, the other parameters of the ARMA\((p, q)\)-NGARCH\((1, 1)\) model for the MOM factor log-returns are quite
### Table 4.8: RMW. Top down: parameters for period 1-5.

<table>
<thead>
<tr>
<th>$\mu$</th>
<th>$\phi_1$</th>
<th>$\phi_2$</th>
<th>$\theta_1$</th>
<th>$\omega$</th>
<th>$\alpha$</th>
<th>$\beta$</th>
<th>$\eta$</th>
<th>$\gamma$</th>
<th>$\nu$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(9.255e-05)</td>
<td>(2.537e-02)</td>
<td>(2.885e-07)</td>
<td>(1.326e-03)</td>
<td>(6.77e-04)</td>
<td>(1.347e-01)</td>
<td>(3.747e-02)</td>
<td>(1.535)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.899e-04</td>
<td>9.537e-02</td>
<td>-</td>
<td>-</td>
<td>2.54e-08</td>
<td>2.29e-02</td>
<td>6.94e-01</td>
<td>4.29e-01</td>
<td>9.79e-01</td>
<td>9.956</td>
</tr>
<tr>
<td>(8.388e-05)</td>
<td>(2.776e-02)</td>
<td>(2.440e-07)</td>
<td>(6.42e-04)</td>
<td>(2.17e-04)</td>
<td>(1.49e-02)</td>
<td>(3.84e-02)</td>
<td>(2.633)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.545e-04</td>
<td>-</td>
<td>-</td>
<td>2.07e-07</td>
<td>6.00e-02</td>
<td>9.16e-01</td>
<td>1.74e-01</td>
<td>1.02e-01</td>
<td>8.807</td>
<td></td>
</tr>
<tr>
<td>(7.823e-05)</td>
<td>(3.30e-07)</td>
<td>(1.39e-02)</td>
<td>(1.32e-02)</td>
<td>(1.26e-01)</td>
<td>(3.97e-02)</td>
<td>(1.741)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.350e-04</td>
<td>8.756e-02</td>
<td>-</td>
<td>-</td>
<td>2.83e-07</td>
<td>6.03e-02</td>
<td>9.06e-01</td>
<td>1.04e-02</td>
<td>9.77e-01</td>
<td>8.201</td>
</tr>
<tr>
<td>(8.201)</td>
<td>(2.815e-02)</td>
<td>(8.40e-07)</td>
<td>(3.33e-02)</td>
<td>(3.54e-02)</td>
<td>(1.65e-01)</td>
<td>(1.725)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.434e-04</td>
<td>7.126e-02</td>
<td>6.496e-02</td>
<td>-</td>
<td>2.28e-08</td>
<td>1.85e-02</td>
<td>9.74e-01</td>
<td>4.33e-01</td>
<td>9.86e-01</td>
<td>8.297</td>
</tr>
<tr>
<td>(8.390e-05)</td>
<td>(2.762e-02)</td>
<td>(2.71e-02)</td>
<td>(2.48e-07)</td>
<td>(5.58e-04)</td>
<td>(1.40e-02)</td>
<td>(1.959)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table 4.9: MOM. Top down: parameters for period 1-5.

<table>
<thead>
<tr>
<th>$\mu$</th>
<th>$\phi_1$</th>
<th>$\phi_2$</th>
<th>$\theta_1$</th>
<th>$\omega$</th>
<th>$\alpha$</th>
<th>$\beta$</th>
<th>$\eta$</th>
<th>$\gamma$</th>
<th>$\nu$</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.201e-04</td>
<td>-</td>
<td>1.41e-01</td>
<td>8.05e-07</td>
<td>1.31e-01</td>
<td>8.44e-01</td>
<td>1.67e-01</td>
<td>9.54e-01</td>
<td>1.26e+01</td>
<td></td>
</tr>
<tr>
<td>(1.302e-04)</td>
<td>(2.747e-02)</td>
<td>(3.71e-06)</td>
<td>(1.16e-01)</td>
<td>(1.148e-01)</td>
<td>(8.375e-02)</td>
<td>(2.995e-02)</td>
<td>(5.114)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.974e-04</td>
<td>1.41e-01</td>
<td>-</td>
<td>4.23e-07</td>
<td>1.00e-01</td>
<td>8.85e-01</td>
<td>1.12e-01</td>
<td>9.35e-01</td>
<td>1.44e+01</td>
<td></td>
</tr>
<tr>
<td>(1.394e-04)</td>
<td>(2.897e-02)</td>
<td>(8.46e-07)</td>
<td>(2.71e-02)</td>
<td>(2.62e-02)</td>
<td>(1.07e-01)</td>
<td>(3.79e-02)</td>
<td>(4.485)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.319e-04</td>
<td>1.19e-01</td>
<td>-</td>
<td>2.54e-07</td>
<td>8.50e-02</td>
<td>9.15e-01</td>
<td>5.60e-02</td>
<td>9.34e-01</td>
<td>1.03e+01</td>
<td></td>
</tr>
<tr>
<td>(1.481e-04)</td>
<td>(2.838e-02)</td>
<td>(5.495e-07)</td>
<td>(1.62e-02)</td>
<td>(1.50e-02)</td>
<td>(1.17e-01)</td>
<td>(3.82e-02)</td>
<td>(2.828)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1.329e-04)</td>
<td>(2.804e-02)</td>
<td>(6.97e-07)</td>
<td>(1.83e-02)</td>
<td>(1.83e-02)</td>
<td>(1.47e-01)</td>
<td>(3.54e-02)</td>
<td>(1.502)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.496e-04</td>
<td>1.00e-01</td>
<td>-</td>
<td>3.42e-07</td>
<td>6.36e-02</td>
<td>9.21e-01</td>
<td>4.97e-02</td>
<td>8.85e-01</td>
<td>7.371</td>
<td></td>
</tr>
<tr>
<td>(1.284e-04)</td>
<td>(2.769e-02)</td>
<td>(4.63e-07)</td>
<td>(1.29e-02)</td>
<td>(1.26e-02)</td>
<td>(1.62e-01)</td>
<td>(3.54e-02)</td>
<td>(1.285)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
unstable.

In Figures 4.3 and 4.4, residual analysis plots for the estimated models in the training data of period 1 are presented. These plots are very encouraging and suggest that the fitted models approximated the data well. The corresponding residual plots for training periods 2-5 were very similar and are therefore not presented.

Figure 4.3: Analysis of residuals in period 1. The upper plots are the QQ-plots of residuals against the fitted driving noise distribution. The lower plots are the ACF for the residuals (solid) and squared residuals (dotted). Left to right: MKT, SMB, HML.

4.1.3 Dynamic Estimation of VaR

As explained in sections 2.2.5 and 2.2.6, the modelling of the factor log-returns considered in this thesis, or the log-returns of any other financial assets, is often motivated by the need to find and test estimations of risk measures.

Thus, the models developed in the previous section were used to estimate the VaR at level $p = 0.05$ in the testing periods of Table 3.1. The decision to only focus on VaR and not ES was motivated by the favorable backtesting procedures of VaR discussed in section 2.2.6.

By using propositions 2.1 and 2.2, one can show that the one-day ahead VaR at level $p$ at time $t - 1$ for an long position in a portfolio that follows the log-returns modelled in sections 4.1.1 and 4.1.2 is governed by the estimated $p-$quantile of the log-return $r_t$. Thus, an exceedance of the VaR at level $p$ occurs when the realization of $r_t$ falls below the estimated $F_{r_t}^{-1}(p)$. 
Figure 4.4: Analysis of residuals in period 1. The upper plots are the QQ-plots of residuals against the fitted driving noise distribution. The lower plots are the ACF for the residuals (solid) and squared residuals (dotted). Left to right: CMA, RMW, MOM.

\[
F_{r_t}^{-1}(p) = \mu + \sum_{i=1}^{p} \phi_i (r_{t-i} - \mu) + \sum_{j=1}^{q} \theta_j \epsilon_{t-j} + \sigma_t F_{\epsilon_t}^{-1}(p), \\
\sigma_t^2 = \omega + \alpha \sigma_{t-1}^2 (\epsilon_{t-1} - \eta)^2 + \beta \sigma_{t-1}^2.
\]

As explained in section 2.2.6, the first step for backtesting the VaR estimations is to define a hit sequence of exceedances

\[
I_t = \begin{cases} 
1, & \text{if } r_t < F_{r_t}^{-1}(p) \\
0, & \text{else}
\end{cases}
\]

To test the null hypothesis that \(E[I_t] = p\) one simply compares the number of realized VaR exceedances during a test period to the corresponding \(B(n, p)\) distribution (with \(n\) being the number of days in the forecasted period, i.e. \(n = 260\) for period 3, \(n = 261\) for the other four periods, and \(n = 1,304\) for the entire five years of testing periods). In Table 4.10 the realized exceedance fraction for all forecasted periods are summarized.

An approximate 95% confidence interval for \(X/n\), where \(X \sim B(n, 0.05)\), constructed by normal approximation with continuity correction is 2.54% – 7.46% for period 3 (\(n = 260\)), 2.55% – 7.45% for the other four periods (\(n =
<table>
<thead>
<tr>
<th></th>
<th>P1</th>
<th>P2</th>
<th>P3</th>
<th>P4</th>
<th>P5</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>MKT</td>
<td>4.21%</td>
<td>5.36%</td>
<td>1.92%</td>
<td>2.68%</td>
<td>5.36%</td>
<td>3.91%</td>
</tr>
<tr>
<td>SMB</td>
<td>7.66%</td>
<td>6.51%</td>
<td>3.46%</td>
<td>4.21%</td>
<td>8.81%</td>
<td>6.13%</td>
</tr>
<tr>
<td>HML</td>
<td>5.75%</td>
<td>6.51%</td>
<td>3.46%</td>
<td>4.60%</td>
<td>2.68%</td>
<td>4.60%</td>
</tr>
<tr>
<td>CMA</td>
<td>7.28%</td>
<td>7.28%</td>
<td>6.15%</td>
<td>4.60%</td>
<td>2.68%</td>
<td>5.60%</td>
</tr>
<tr>
<td>RMW</td>
<td>2.30%</td>
<td>4.21%</td>
<td>9.23%</td>
<td>3.83%</td>
<td>5.36%</td>
<td>4.98%</td>
</tr>
<tr>
<td>MOM</td>
<td>4.21%</td>
<td>6.13%</td>
<td>5.00%</td>
<td>4.21%</td>
<td>4.60%</td>
<td>4.83%</td>
</tr>
</tbody>
</table>

Table 4.10: Percentage of exceedances of the daily forecasted $\text{VaR}_{0.05}$ in test data using the ARMA($p, q$)-NGARCH(1, 1) models.

261), and 3.86%−6.14% for the entire five-year period ($n = 1, 304$). Referring to Table 4.10, one can see that in 25 of 30 forecasted one-year periods one fails to reject the null hypothesis $p = 0.05$. For the total five-year testing period the null hypothesis cannot be rejected for any of the six factors.

The likelihood ratio test of [57] explained in section 2.2.6 was implemented to test the null hypothesis that a Weibull distribution fitted to the durations between exceedances reduces to the memory-less exponential distribution, $H_0 : b = 1$. In [57] a Monte Carlo simulation indicates that a sample size of at least 750 days (corresponding to roughly three years) is required to enable satisfactory inference when applying this test. Thus this test was only applied to the total five-year ($n = 1, 304$) testing period.

At the $\alpha = 0.05$ level, the null hypothesis of exponentially distributed durations between exceedances was only rejected for the MOM factor. The maximum likelihood estimation of $b$ for the MOM factor was $\hat{b} = 1.39$ giving a $p$-value of 0.002. A $b > 1$ indicates that the probability of an exceedance at time $t$ increases when the time since the last exceedance is long.

In order to properly assess the performance of the ARMA($p, q$)-NGARCH(1, 1) models and the estimated VaR provided by them, there is a need to consider a more simplistic model to use as a reference. The natural choice is to turn to the historical simulation (HS) approach where the empirical distribution of returns is used to approximate the distribution of future returns. Given $n$ observations $x_1, \ldots, x_n$ of some iid random variables $X_1, \ldots, X_n$, the empirical distribution is given by

$$
\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^{n} I(x_i < x).
$$
### Table 4.11: Percentage of exceedances of the daily forecasted VaR<sub>0.05</sub> in test data using the HS model.

<table>
<thead>
<tr>
<th></th>
<th>P1</th>
<th>P2</th>
<th>P3</th>
<th>P4</th>
<th>P5</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>MKT</td>
<td>8.43%</td>
<td>9.58%</td>
<td>0.00%</td>
<td>5.36%</td>
<td>9.20%</td>
<td>6.52%</td>
</tr>
<tr>
<td>SMB</td>
<td>9.96%</td>
<td>8.43%</td>
<td>1.15%</td>
<td>1.92%</td>
<td>7.28%</td>
<td>5.75%</td>
</tr>
<tr>
<td>HML</td>
<td>8.05%</td>
<td>9.20%</td>
<td>2.31%</td>
<td>3.45%</td>
<td>4.21%</td>
<td>5.44%</td>
</tr>
<tr>
<td>CMA</td>
<td>8.05%</td>
<td>7.28%</td>
<td>4.23%</td>
<td>3.07%</td>
<td>3.45%</td>
<td>5.21%</td>
</tr>
<tr>
<td>RMW</td>
<td>2.30%</td>
<td>8.43%</td>
<td>8.46%</td>
<td>0.38%</td>
<td>2.68%</td>
<td>4.45%</td>
</tr>
<tr>
<td>MOM</td>
<td>5.75%</td>
<td>14.18%</td>
<td>0.77%</td>
<td>0.77%</td>
<td>9.96%</td>
<td>6.29%</td>
</tr>
</tbody>
</table>

By the Glivenko–Cantelli theorem [62], these empirical distributions converge uniformly to the true distribution with probability one as \( n \to \infty \). However, this will only be true if the assumption of iid random variables is true. Thus, when using the empirical distribution to approximate the true distribution of the log-returns of some financial asset, one does not take into account any autocorrelation or conditional heteroscedasticity. Given the results of sections 4.1.1 and 4.1.2, one can expect the HS modelling approach to yield poor estimations of the VaR.

In the HS modelling approach, the empirical estimate of the VaR at level \( p \) is given by

\[
\text{VaR}_p(X) = L_{[np]+1,n}
\]

where \( L_{1,n} \geq ... \geq L_{n,n} \) is the ordered sample of losses, and \([\ ]\) is the floor function. In Table 4.11 the unconditional coverage results for the HS estimated VaR are presented. Here \( H_0 : p = 0.05 \) is rejected in 20 out of 30 one-year test periods. For the five-year test periods, the null is rejected for two out of six factors series.

In the likelihood ratio test, exponentially distributed durations \( H_0 : b = 1 \) cannot be rejected for the entire five-year period for the VaR predictions for SMB, HML, and CMA. For the three other factors, \( b = 1 \) is rejected and the maximum likelihood estimated values of \( b \) are all less than one, indicating that the exceedances are clustered.
4.2 Factor Dependence

4.2.1 Factor Correlations and Dependence Modelling

Dependence in the daily log-returns of the factors was modelled by using copulas to estimate joint distributions of these returns. The results from the univariate modelling in section 4.1 suggests that \( \text{ARMA}(p, q) - \text{NGARCH}(1, 1) \) processes can capture the distributional dynamics of the log-returns of the factors rather well. It is thus desirable to keep these processes as the univariate models when creating joint distributions. This can be achieved by first estimating such univariate models and then use copulas to introduce dependence structures in their driving noise terms. In this thesis, the Gaussian and the \( t \) copula were considered and compared.

To limit complexity and have all driving noise terms defined in a similar way, the univariate distributional dynamics of all log-returns were modelled as \( \text{ARMA}(1, 1) - \text{NGARCH}(1, 1) \) processes with iid mean zero and unit variance skewed \( t \) distributed driving noise. The results from the earlier univariate modelling suggests that this model specification is sufficient for all factors in all periods. After estimating the univariate models in a training period, six realized noise series, \( \{\hat{\epsilon}_{it}\}, i = 1, 2, \ldots, 6 \), were obtained. These were then be transformed using the probability transform and the skewed \( t \) distribution to get samples of a six-dimensional uniformly distributed random vector. Copulas were then fitted to either the entire vector or selected elements of this vector, to create joint distributions for all six factors or subsets of the factors.

The plots in Figures 4.5 and 4.6 illustrate the copula fitting procedure when creating a bivariate joint distribution for the driving noise of the MKT and MOM factors in period 1. The left plot in Figure 4.5 is the scatter plot of the uniformly distributed random variables formed by applying the probability transform to the sample driving noise of the MKT and MOM factors. The estimated \( t \) copula for this data has parameters \( (\rho, \nu) = (-0.0152, 2.81) \). Drawing a random sample, of equal size as the data \( (n = 1, 305) \) used for estimation, from this fitted \( t \) copula yields the middle plot in Figure 4.5, and to the right is the corresponding plot for a fitted Gaussian copula \( (\rho = -0.0152) \). Focusing on the corners of these three plots one can see that both the data and fitted \( t \) copula have samples were both variables are in their extremes. The sample from the fitted Gaussian copula, which lacks tail dependence, does not have these shared extremes. Furthermore, the low correlation means that the samples from the Gaussian copula are close to independent.

Figure 4.6 presents the scatter plots of the driving noise corresponding to
the copula plots in Figure 4.5, i.e. the left plot shows the empirical driving noise of the ARMA(1, 1)-NGARCH(1, 1) process fitted to the log-returns of MKT and MOM in training period 1, while the middle and right plot were created by the quantile transform using the samples in Figure 4.5 from the fitted $t$ and Gaussian copulas. Here one should not focus on the individual sample points, but instead look at the general properties of the scatter plots. One can here again see how both the data and the samples from the joint distribution created with the $t$ copula have some outcomes that are shared extremes of the MKT and MOM factors. As expected, the joint distribution formed using the Gaussian copula has not produced such samples.

Figure 4.5: Scatter plots of $U_{MKT}$ and $U_{MOM}$ in period 1, $n = 1, 305$. To the left is the data. In the middle is a sample from a fitted bivariate $t$ copula. To the right is a sample from a fitted bivariate Gaussian copula.

Figure 4.6: Scatter plots of $\epsilon_{MKT}$ and $\epsilon_{MOM}$ in period 1, $n = 1, 305$. To the left is the data. In the middle is a sample from a fitted bivariate $t$ copula and marginal skewed $t$ distribution. To the right is a sample from a fitted bivariate Gaussian copula and marginal skewed $t$ distribution.

In Tables 4.12-4.14, the correlations of the factors, both in the log-returns and in the driving noise sequences, for training period 1, 3, and 5 are given. These tables make it evident that there are strong correlation between the daily
log-returns of these six factors. The largest negative correlation in log-returns is between HML and RMW in period 5, were the correlation is -0.83. This can be compared to a correlation between RMW and HML of 0.08 in monthly returns in U.S. market data from the period July 1964-December 2013 reported in [3], and a correlation of 0.42 in daily log-returns in a similar U.S. market data set reported in [7]. This suggests that by comparing data from periods of varying length, different markets, and considering both daily and monthly returns, one can find very different dependence structures for the factors.

The largest positive correlation in the daily log-returns is the correlation between HML and CMA in period 5, were the correlation is 0.65. This can be compared to a correlation of 0.62 between U.S. daily log-returns for HML and CMA reported in [7], and a correlation of 0.70 in monthly returns in the U.S. market reported in [3].

The large negative correlation between MKT and SMB in all three Tables 4.12-4.14 is worth noting. In [7] the correlation between the daily log-returns of MKT and SMB in U.S. data is reported to be 0.25, and in [3] the correlation in monthly returns is reported to be 0.28. However, in [6] the correlation between the daily log-returns of MKT and SMB in a data set from the U.S. market is also found to be negative and is reported to be -0.18.

Besides including several strong correlations, the results in Tables 4.12-4.14 also suggest that the correlations between the daily log-returns of the six factors are not constant. One example is the correlation between HML and CMA. As mentioned previously, the correlation between these two factors in period 5 is 0.65, making it the largest positive correlation found. However, the corresponding correlation in period 1 is 0.23. While this is also a positive correlation, it is much lower.

A similar result for factors with negative correlation is found in the correlation between MKT and RMW which varies between -0.46 and -0.12. Furthermore, there are also correlations that are positive in one period and negative in another. The correlation between MKT and CMA goes from -0.13 in period 3 to 0.09 in period 5.

Modelling all possible joint distributions that can be formed from the six factors is not a feasible endeavour in this thesis. Instead, a total of four different joint distributions were modelled for all five periods. These joint distributions were the full six dimensional distribution for all six factors, and three bivariate distributions: MKT and SMB, MKT and HML, and MKT and MOM.

The results for the six dimensional copulas for training period 1, 3, and 5 are presented in Tables 4.15-4.17. These tables include the estimated parameters and log-likelihoods of the fitted Gaussian and t copulas. A result that
### Table 4.12: Sample correlations in training period 1.

<table>
<thead>
<tr>
<th></th>
<th>SMB</th>
<th>HML</th>
<th>CMA</th>
<th>RMW</th>
<th>MOM</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Log-returns, ( r_t )</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MKT</td>
<td>-0.7212</td>
<td>0.6325</td>
<td>-0.0660</td>
<td>-0.4555</td>
<td>-0.2246</td>
</tr>
<tr>
<td>SMB</td>
<td>-0.4506</td>
<td>0.0438</td>
<td>0.3192</td>
<td>0.2325</td>
<td></td>
</tr>
<tr>
<td>HML</td>
<td>1</td>
<td>0.2286</td>
<td>-0.7463</td>
<td>-0.3112</td>
<td></td>
</tr>
<tr>
<td>CMA</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>-0.3153</td>
<td>0.0267</td>
</tr>
<tr>
<td>RMW</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>0.3814</td>
</tr>
</tbody>
</table>

| **Driving noise, \( \hat{\epsilon}_t \)** |       |       |       |       |       |
| MKT    | -0.6543 | 0.5728 | 0.0410 | -0.4090 | -0.0121 |
| SMB    | -0.4207 | -0.0022 | 0.2975 | 0.0880 |       |
| HML    | 1      | 0.3128 | -0.7523 | -0.1916 |       |
| CMA    | -      | -      | 1      | -0.3736 | 0.0076 |
| RMW    | -      | -      | -      | 1      | 0.2244 |

### Table 4.13: Sample correlations in training period 3.

<table>
<thead>
<tr>
<th></th>
<th>SMB</th>
<th>HML</th>
<th>CMA</th>
<th>RMW</th>
<th>MOM</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Log-returns, ( r_t )</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MKT</td>
<td>-0.7007</td>
<td>0.4670</td>
<td>-0.1281</td>
<td>-0.3953</td>
<td>-0.4519</td>
</tr>
<tr>
<td>SMB</td>
<td>1</td>
<td>-0.3576</td>
<td>0.0272</td>
<td>0.2717</td>
<td>0.3394</td>
</tr>
<tr>
<td>HML</td>
<td>-</td>
<td>1</td>
<td>0.3472</td>
<td>-0.8195</td>
<td>-0.6254</td>
</tr>
<tr>
<td>CMA</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>-0.3184</td>
<td>-0.0359</td>
</tr>
<tr>
<td>RMW</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>0.5264</td>
</tr>
</tbody>
</table>

| **Driving noise, \( \hat{\epsilon}_t \)** |       |       |       |       |       |
| MKT    | -0.6392 | 0.4377 | -0.0188 | -0.3441 | -0.2272 |
| SMB    | 1      | -0.3186 | 0.0005 | 0.2301 | 0.1992 |
| HML    | -      | 1      | 0.4077 | -0.8058 | -0.4581 |
| CMA    | -      | -      | 1      | -0.3889 | -0.0468 |
| RMW    | -      | -      | -      | 1      | 0.3543 |
should be noted in these tables is that the degrees of freedom for the \( t \) copula is quite stable, \( \nu \in [6.19, 7.20] \), while some of the correlations noticeably change.

The corresponding tables for the bivariate copulas used to create joint distributions for factor pairs MKT and SMB, MKT and HML, and MKT and MOM, are presented in Table 4.18, 4.19, and 4.20. However, these tables do not include the estimated correlation of the copulas. Since Kendall’s tau, rather than maximum likelihood, is used to estimate the correlations, the estimated correlations of the bivariate copulas are equal to the corresponding correlations in the full six dimensional copula. Thus, to limit unnecessary repetition, the copula result tables of the estimated bivariate copulas only include the estimated degrees of freedom of the \( t \) copulas and the log-likelihoods.

The estimated bivariate copulas for MKT and SMB has a higher estimated degrees of freedom than the six dimensional copula, \( \nu \in [7.32, 9.75] \). As one might expect given the results in Tables 4.12-4.14, the correlation in the estimated bivariate copula for MKT and SMB is large and negative.

The results for the estimated bivariate copula for the MKT and HML factors has an even higher degree of freedom \( \nu \in [10.56, 18.93] \), and large positive correlations. The estimated copulas of MKT and MOM has a noticeably lower degree of freedom, \( \nu \in [2.81, 5.63] \), than the estimated six dimensional copulas. One should also note that correlation for the estimated copula of
Table 4.15: Results for the six-dimensional copula in training period 1.

<table>
<thead>
<tr>
<th></th>
<th>SMB</th>
<th>HML</th>
<th>CMA</th>
<th>RMW</th>
<th>MOM</th>
</tr>
</thead>
<tbody>
<tr>
<td>MKT</td>
<td>-0.6761</td>
<td>0.5934</td>
<td>0.0723</td>
<td>-0.4191</td>
<td>-0.0152</td>
</tr>
<tr>
<td>SMB</td>
<td>1</td>
<td>-0.4297</td>
<td>-0.0188</td>
<td>0.2848</td>
<td>0.0643</td>
</tr>
<tr>
<td>HML</td>
<td>-</td>
<td>1</td>
<td>0.3378</td>
<td>-0.7519</td>
<td>-0.2233</td>
</tr>
<tr>
<td>CMA</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>-0.4005</td>
<td>0.0090</td>
</tr>
<tr>
<td>RMW</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>0.2484</td>
</tr>
</tbody>
</table>

\[ \nu \]

<table>
<thead>
<tr>
<th></th>
<th>t copula</th>
<th>Gaussian copula</th>
</tr>
</thead>
<tbody>
<tr>
<td>log-likelihood</td>
<td>1,542</td>
<td>1,368</td>
</tr>
</tbody>
</table>

Table 4.16: Results for the six-dimensional copula in training period 3.

<table>
<thead>
<tr>
<th></th>
<th>SMB</th>
<th>HML</th>
<th>CMA</th>
<th>RMW</th>
<th>MOM</th>
</tr>
</thead>
<tbody>
<tr>
<td>MKT</td>
<td>-0.6628</td>
<td>0.4320</td>
<td>0.0267</td>
<td>-0.3336</td>
<td>-0.2070</td>
</tr>
<tr>
<td>SMB</td>
<td>1</td>
<td>-0.3028</td>
<td>-0.0129</td>
<td>0.2213</td>
<td>0.1722</td>
</tr>
<tr>
<td>HML</td>
<td>-</td>
<td>1</td>
<td>0.4393</td>
<td>-0.7945</td>
<td>-0.4706</td>
</tr>
<tr>
<td>CMA</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>-0.4159</td>
<td>-0.0446</td>
</tr>
<tr>
<td>RMW</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>0.3533</td>
</tr>
</tbody>
</table>

\[ \nu \]

<table>
<thead>
<tr>
<th></th>
<th>t copula</th>
<th>Gaussian copula</th>
</tr>
</thead>
<tbody>
<tr>
<td>log-likelihood</td>
<td>1,656</td>
<td>1,484</td>
</tr>
</tbody>
</table>

MKT and MOM is quite unstable, varying between \( \rho \in [-0.2070, 0.0256] \).

### 4.2.2 Dynamic Estimation of VaR

As in the univariate modelling of section 4.1, the joint distributions created by the estimated Gaussian and \( t \) copulas were used to estimate the one-day ahead VaR in the one year long testing periods, and these predictions were again backtested for correct unconditional coverage and exponentially distributed durations between exceedances. The portfolios considered were equally weighted (1/N) portfolios that were refitted once a week (every 5 days). A historical simulation (HS) model based on the empirical distribution of the past two years was again used for comparison.


<table>
<thead>
<tr>
<th></th>
<th>SMB</th>
<th>HML</th>
<th>CMA</th>
<th>RMW</th>
<th>MOM</th>
</tr>
</thead>
<tbody>
<tr>
<td>MKT</td>
<td>-0.5087</td>
<td>0.2514</td>
<td>0.1630</td>
<td>-0.1525</td>
<td>0.0256</td>
</tr>
<tr>
<td>SMB</td>
<td>1</td>
<td>-0.1672</td>
<td>-0.1211</td>
<td>0.1081</td>
<td>0.0320</td>
</tr>
<tr>
<td>HML</td>
<td>-</td>
<td>1</td>
<td>0.6890</td>
<td>-0.8031</td>
<td>-0.1554</td>
</tr>
<tr>
<td>CMA</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>-0.6369</td>
<td>0.0222</td>
</tr>
<tr>
<td>RMW</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>0.1082</td>
</tr>
</tbody>
</table>

Table 4.17: Results for the six-dimensional copula in training period 5.

<table>
<thead>
<tr>
<th></th>
<th>Period 1</th>
<th>Period 3</th>
<th>Period 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Copula</td>
<td>(t) Gaussian</td>
<td>(t) Gaussian</td>
<td>(t) Gaussian</td>
</tr>
<tr>
<td>(\nu)</td>
<td>9.75</td>
<td>8.68</td>
<td>207</td>
</tr>
<tr>
<td>log-likelihood</td>
<td>393</td>
<td>384</td>
<td>372</td>
</tr>
</tbody>
</table>

Table 4.18: Results for the bivariate copulas for MKT and SMB.

The percentage of realized exceedances of the predicted VaR at level \(p = 0.05\) for the full six factor portfolio are presented in Table 4.21. As in the univariate case, a 95% confidence interval for the number of exceedances given a correct coverage (\(p = 0.05\)) is 2.55%-7.45% for periods 1, 2, 4, and 5. The corresponding interval for period 3 is 2.54%-7.46%, and for the entire five years of testing periods the interval is 3.86%-6.14%.

Using these intervals, correct unconditional coverage in Table 4.21 is rejected in periods 2, 3, and 5 for the joint distributions formed using the \(t\) and Gaussian copulas. Perhaps quite surprisingly, correct unconditional coverage

<table>
<thead>
<tr>
<th></th>
<th>Period 1</th>
<th>Period 3</th>
<th>Period 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Copula</td>
<td>(t) Gaussian</td>
<td>(t) Gaussian</td>
<td>(t) Gaussian</td>
</tr>
<tr>
<td>(\nu)</td>
<td>10.59</td>
<td>10.56</td>
<td>18.93</td>
</tr>
<tr>
<td>log-likelihood</td>
<td>273</td>
<td>265</td>
<td>138</td>
</tr>
</tbody>
</table>

Table 4.19: Results for the bivariate copulas for MKT and HML.
cannot be rejected for the HS model. However, all three modelling approaches fail to achieve exponentially distributed durations between exceedances, as $H_0 : b = 1$ of a fitted Weibull distribution is rejected at the $\alpha = 0.05$ level. For both the Gaussian and $t$ copulas, the maximum likelihood estimated $b$ of a Weibull distribution fitted to the durations is equal to $\hat{b} = 0.78$. For the HS model the corresponding value is $\hat{b} = 0.80$. As explained in section 2.2.6, an estimation were $\hat{b} < 1$ indicates that the exceedances are clustered.

Another result in Table 4.21 that might come as a surprise is that in testing period 3 the realized number of exceedances is higher for the VaR estimated with the $t$ copula than the VaR estimated with the Gaussian copula. Considering that the $t$ copula has tail dependence, whereas the Gaussian copula does not, this can seem quite odd. However, to highlight the differences between the Gaussian and $t$ copulas one has to look further into the extremes of the left tail. Thus for the three bivariate joint distributions, the VaR at level $p = 0.01$ was also considered. However, due to the limited number of realized exceedances with such a low value of $p$, no attempt was made to perform any statistical test on these exceedances.

The backtesting results for the VaR predictions for a portfolio of the MKT and SMB factors are presented in Table 4.22. Here the exceedances of the VaR at level $p = 0.05$ implied by the Gaussian and $t$ copulas are equal, and correct coverage cannot be rejected in any period. Furthermore, exponentially distributed durations ($\hat{b} = 1$) cannot be rejected for these exceedances either. The historical simulation approach fared worse, with correct unconditional coverage rejected in period 2 and 3, and exponentially distributed between exceedances rejected with a maximum likelihood estimated $\hat{b} = 0.70$, indicating clustered exceedances.

Similar results were obtained for the equally weighted two-factor portfolio of the MKT and HML factors. Here the realized exceedances of the VaR at level $p = 0.05$ are again equal for the Gaussian and $t$ copulas. However, correct unconditional coverage is rejected in period 3. For the HS approach, correct unconditional coverage is rejected in period 2 and 3. The likelihood ratio test
Table 4.21: Results for the portfolio of all six factors. Percentage of exceedances of the daily forecasted VaR$_{0.05}$ in test data.

<table>
<thead>
<tr>
<th></th>
<th>P1</th>
<th>P2</th>
<th>P3</th>
<th>P4</th>
<th>P5</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t$</td>
<td>3.45%</td>
<td>2.30%</td>
<td>8.85%</td>
<td>7.28%</td>
<td>1.15%</td>
<td>4.60%</td>
</tr>
<tr>
<td>Gaussian</td>
<td>3.45%</td>
<td>2.30%</td>
<td>8.46%</td>
<td>7.28%</td>
<td>1.15%</td>
<td>4.52%</td>
</tr>
<tr>
<td>HS</td>
<td>4.60%</td>
<td>5.75%</td>
<td>3.85%</td>
<td>5.75%</td>
<td>3.45%</td>
<td>4.68%</td>
</tr>
</tbody>
</table>

again fails to reject exponentially distributed durations between exceedances for the VaR estimations from both copulas. For the HS model, exponentially distributed durations is rejected with a maximum likelihood estimated $\hat{b} = 0.75$. This again indicates that the estimations of the VaR using HS leads to clustered exceedances.

For the portfolio of the MKT and MOM factors, the results in Table 4.24 leads to correct unconditional coverage being rejected in period 5 for both copulas. However, the duration test again fails to reject independent exceedances for the results from both copulas. The HS modelling approach again fares worse, with rejection of correct unconditional coverage in period 2 and 3, and rejection of exponentially distributed durations, with a maximum likelihood estimated $\hat{b} = 0.72$.

For the exceedances of the predicted VaR at level $p = 0.01$ for all three two-factor portfolios in Tables 4.22-4.24, one should note that the use of a $t$ copula always yields the same number or fewer realized exceedances as the Gaussian copula. This is an expected results since the $t$ copula models tail dependence in the joint distributions of the factors, whereas the Gaussian copula does not.
Table 4.22: Results for the portfolio of MKT and SMB. Percentage of exceedances of the daily forecasted VaR\(_p\) in test data.

<table>
<thead>
<tr>
<th></th>
<th>P1</th>
<th>P2</th>
<th>P3</th>
<th>P4</th>
<th>P5</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>(p = 0.05)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(t)</td>
<td>3.45%</td>
<td>4.98%</td>
<td>3.46%</td>
<td>5.36%</td>
<td>7.28%</td>
<td>4.91%</td>
</tr>
<tr>
<td>Gaussian</td>
<td>3.45%</td>
<td>4.98%</td>
<td>3.46%</td>
<td>5.36%</td>
<td>7.28%</td>
<td>4.91%</td>
</tr>
<tr>
<td>HS</td>
<td>7.28%</td>
<td>7.66%</td>
<td>0.38%</td>
<td>6.13%</td>
<td>6.90%</td>
<td>5.67%</td>
</tr>
<tr>
<td>(p = 0.01)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(t)</td>
<td>0.00%</td>
<td>1.15%</td>
<td>0.00%</td>
<td>2.30%</td>
<td>1.15%</td>
<td>0.92%</td>
</tr>
<tr>
<td>Gaussian</td>
<td>0.38%</td>
<td>1.15%</td>
<td>0.00%</td>
<td>2.30%</td>
<td>1.15%</td>
<td>1.00%</td>
</tr>
<tr>
<td>HS</td>
<td>0.38%</td>
<td>3.83%</td>
<td>0.00%</td>
<td>0.38%</td>
<td>2.30%</td>
<td>1.38%</td>
</tr>
</tbody>
</table>

Table 4.23: Results for the portfolio of MKT and HML. Percentage of exceedances of the daily forecasted VaR\(_p\) in test data.

<table>
<thead>
<tr>
<th></th>
<th>P1</th>
<th>P2</th>
<th>P3</th>
<th>P4</th>
<th>P5</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>(p = 0.05)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(t)</td>
<td>3.07%</td>
<td>4.98%</td>
<td>1.15%</td>
<td>3.07%</td>
<td>4.98%</td>
<td>3.45%</td>
</tr>
<tr>
<td>Gaussian</td>
<td>3.07%</td>
<td>4.98%</td>
<td>1.15%</td>
<td>3.07%</td>
<td>4.98%</td>
<td>3.45%</td>
</tr>
<tr>
<td>HS</td>
<td>6.51%</td>
<td>9.96%</td>
<td>0.38%</td>
<td>4.60%</td>
<td>6.90%</td>
<td>5.67%</td>
</tr>
<tr>
<td>(p = 0.01)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(t)</td>
<td>0.38%</td>
<td>0.38%</td>
<td>0.00%</td>
<td>0.38%</td>
<td>0.38%</td>
<td>0.31%</td>
</tr>
<tr>
<td>Gaussian</td>
<td>0.38%</td>
<td>1.15%</td>
<td>0.00%</td>
<td>0.38%</td>
<td>0.77%</td>
<td>0.54%</td>
</tr>
<tr>
<td>HS</td>
<td>1.92%</td>
<td>3.07%</td>
<td>0.00%</td>
<td>1.15%</td>
<td>1.53%</td>
<td>1.53%</td>
</tr>
<tr>
<td></td>
<td>P1</td>
<td>P2</td>
<td>P3</td>
<td>P4</td>
<td>P5</td>
<td>Total</td>
</tr>
<tr>
<td>-------</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>-------</td>
</tr>
<tr>
<td><strong>p = 0.05</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$t$</td>
<td>3.45%</td>
<td>4.21%</td>
<td>5.77%</td>
<td>6.13%</td>
<td>2.30%</td>
<td>4.37%</td>
</tr>
<tr>
<td>Gaussian</td>
<td>3.45%</td>
<td>4.21%</td>
<td>5.38%</td>
<td>6.13%</td>
<td>1.53%</td>
<td>4.14%</td>
</tr>
<tr>
<td>HS</td>
<td>6.13%</td>
<td>8.05%</td>
<td>0.77%</td>
<td>6.51%</td>
<td>4.21%</td>
<td>5.14%</td>
</tr>
<tr>
<td><strong>p = 0.01</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$t$</td>
<td>0.00%</td>
<td>0.38%</td>
<td>0.00%</td>
<td>1.92%</td>
<td>0.38%</td>
<td>0.54%</td>
</tr>
<tr>
<td>Gaussian</td>
<td>0.00%</td>
<td>0.77%</td>
<td>0.00%</td>
<td>2.30%</td>
<td>0.38%</td>
<td>0.69%</td>
</tr>
<tr>
<td>HS</td>
<td>0.38%</td>
<td>4.21%</td>
<td>0.00%</td>
<td>1.92%</td>
<td>0.38%</td>
<td>1.38%</td>
</tr>
</tbody>
</table>

Table 4.24: Results for the portfolio of MKT and MOM. Percentage of exceedances of the daily forecasted $VaR_p$ in testing data.
Chapter 5
Discussion

The results of this thesis project presented in Chapter 4 are promising. It is clear from the univariate modelling results that the distributional dynamics of the daily log-returns of the six factors considered are well approximated as ARMA-NGARCH processes with skewed \( t \) distributed driving noise. Moreover, the joint distributions formed by copula-based modelling of vectors of the driving noise for several univariate processes has the benefit of preserving these favorable univariate ARMA-NGARCH processes while introducing dependence in the factor log-returns.

5.1 Factor Dynamics

The results when implementing the univariate ARMA-NGARCH models to estimate the one-day ahead VaR are among the most positive of any results in this thesis. It is clear that the ARMA-NGARCH models give VaR predictions that are vastly more accurate than the predictions of a historical simulation approach where one uses the empirical distribution of log-returns. The VaR predictions of the ARMA-NGARCH models fares better in statistical backtesting of both the unconditional coverage and the memory-less durations between exceedances.

This implies that the parametric models formed from ARMA-NGARCH processes with skewed \( t \) distributed driving noise, which can model autocorrelation and conditional heteroscedasticity, leads to a greater goodness-of-fit to daily factor log-returns than what can be achieved by a more simplistic approach based on the empirical distribution. These findings motivate the use of such complex parametric models.

Furthermore, the daily log-returns of the MKT factor and the SMB factor
have distributional dynamics that are seemingly rather stable. For both these factors, the parameter $\eta$ that controls leverage-type effects is significant in all five periods of training data. The MKT factor has a positive $\eta$ which indicates a leverage effect where negative outcomes increases variability more than positive outcomes of a similar size. This is the classic leverage effect that has been reported for the MKT factor in U.S. market data [6] and has been studied in other markets as well [48]. The leverage effect in the SMB factor is of opposite nature, and since the SMB factor depends on both long and short equity positions, the interpretation of this opposite leverage effect is less clear. It is however interesting that in [6] no such opposite leverage effect for the SMB factor was found in U.S. market data. Moreover, in [6] evidence of such an opposite leverage effect was found for the distributional dynamics of the MOM factor, while in this study the models for the MOM factor only have positive $\eta$, several of which are significant, indicating an ordinary leverage effect.

However, even if the distributional dynamics of the MKT and SMB factors are rather stable, the stability of the univariate models is one of the key issues in this study. For several of the factors, the model selection procedure based on BIC gives different model specifications for different periods. Furthermore, when comparing the estimated values of parameters for the same factor in different training periods, it is clear that they can be very different. This can of course partly be explained by estimation error. However, it would be unreasonable to believe that the exact parameter values should be constant over any longer period.

Thus, an interesting pursuit for future research would be to extend the analysis of this thesis by also considering methods for the detection of parameter instability in the ARMA-NGARCH processes that approximate the log-returns. Such methods could include statistical means of detecting parameter changes, or dynamical modelling approaches that update models in a more dynamic way than a once a year refitting based on the previous five years as was done in this thesis. Considering external variables to model general changes in market conditions is also a potential way of ensuring that the ARMA-NGARCH models used do not become obsolete.

The possible faults in the model specification and model selection procedure used in this thesis should also be noted. Firstly, in the specification of ARMA-NGARCH processes as approximating models of the factor log-returns, the order of the mean-modelling ARMA-model was limited to $(p, q) = \{0, 1, 2\}^2$. This of course introduces a bias which might be unfavorable. The specification of a NGARCH process to model the conditional variance in factor log-returns is also a possible shortcoming in this study. The results for the
MKT and SMB factor indicate that it is important to use a GARCH-type model that allows for leverage effects. However, there are many other such GARCH models besides the NGARCH, and a comprehensive study of their properties and the properties of the factor log-returns is needed in order to determine which model is best suited for each factor in each training period. Furthermore, there is a possibility that some other model selection criteria than BIC, which was used in this thesis, would have been better.

The theoretical result that QMLE parameters are asymptotically normal as presented in section 2.2.2 and used in section 4.1.2 to calculate significance levels and standard deviations for the estimated univariate models, is interesting from a theoretical perspective, but in the applied setting of this thesis there is a risk that these strong theoretical results are misinterpreted. Firstly, this theoretical result is only valid if the data used is generated by a true ARMA-GARCH process. In reality, the data sets used in this thesis are at best only approximately ARMA-GARCH processes. Furthermore, even if one used a data set that was generated by a true ARMA-GARCH process, the asymptotically normality of QMLE parameters is only valid if the model one is estimating is correctly specified. If, for example, the data comes from an AR(2)-NGARCH(1, 1) process and one wrongfully assumes some other model specification, then the asymptotic results presented in section 2.2.2 are no longer valid. It is thus very important to only view the significance levels and standard deviations presented in section 4.1.2 as approximate results. Any further analysis of the theoretical validity of these types of QMLE results are beyond the scope of this thesis, but constitutes an interesting pursuit for future research.

Finally, it should also be noted that the parameter estimation when fitting a GARCH process to data is a difficult nonlinear optimization problem that is solved numerically. In order to mitigate the risk of finding parameter vectors that are only local maxima, several complex solving algorithms with randomized starting values were implemented. However, even if great care was taken when performing the numerical optimizations, the complex nature of these optimization problems cannot not be ignored. There is thus a possibility that the use of other solving algorithms would give some parameter estimations that differ from the ones reported in this thesis.

5.2 Factor Dependence

The multivariate modelling based on the use of copulas to connect the driving noise terms of univariate models proved to be an easily implemented and flex-
ible augmentation of the ARMA-NGARCH modelling, and has yielded some interesting results in this thesis. The most positive results are those of the bivariate joint distributions created for the factor pairs MKT and SMB, MKT and HML, and MKT and MOM. The joint distributions created for these pairs gave VaR predictions that outperformed the corresponding predictions by the historical simulation approach, i.e. just as in the univariate case the parametric models gave better results than the models based on the empirical distributions. Backtesting of the VaR predictions could not reject memory-less durations between exceedances for any of the bivariate copula models. For the corresponding VaR predictions based on historical simulation, memory-less durations between exceedances was rejected and evidence of clustered exceedances found in all three cases. Furthermore, the unconditional exceedance coverage was also better for the copula-based models.

However, the problem of model stability, which was also discussed in the previous section in the context of the univariate models, became even more evident in the multivariate modelling. The correlations between the factors was particularly unstable and it is clear that if one wants to model the joint distributions of these factors well, there is a need to implement modelling methods that allow for a more dynamical dependence than a yearly updated five-year rolling-window approach. One possible such method is the *dynamic conditional correlation* (DCC) model of [30], which was implemented in [6] the modelling of factor log-returns dependence in the U.S. market. The non-constant correlations between the factor log-returns is probably partly explained by changing correlations between the underlying factors, and partly explained by the fact that the portfolios that are used to proxy the factors are constructed yearly and can thus change composition once a year.

The unstable correlations between factor log-returns might explain the poor VaR predictions of the full six-dimensional copulas in this thesis. Furthermore, since the historical simulation approach uses a daily updated two-year rolling window approach, and thus only uses data from the past two years, the good unconditional coverage of the VaR prediction of the historical simulation model in Table 4.21 might be explained by this method being better at capturing dynamic correlation.

Another interesting result in this thesis is that there appears to exist a wide range of different types of dependencies between the factors in the data used. This is made evident by the differences in the estimated degrees of freedom, \(\nu\), for the bivariate copulas. For the copula for the joint distribution of the MKT and HML factors, the estimated \(\nu\) in training periods 1, 3, and 5 lies between 10.56 and 18.93, whereas the corresponding \(\nu\) for the copula for the
Joint distribution of the MKT and MOM factors lies between 2.81 and 5.63. As \( \nu \) controls tail dependence in a bivariate \( t \) copula, where a smaller \( \nu \) gives a larger tail dependence, this suggests that MKT and MOM has a stronger tail dependence than MKT and HML.

Since the only copulas considered in this thesis were the Gaussian and \( t \) copulas, the possibility of asymmetric tail dependence in the factors has not been examined. This is a weakness of this study considering that both [6] and [7] use a skewed \( t \) copula when modelling the joint distribution of factor log-returns in the U.S. market to allow for asymmetric dependencies. Both these studies also conclude that there are asymmetric dependencies in the U.S. factor log-returns. A possible improvement of the analysis in the present thesis would therefore be to investigate the possible asymmetries in the tail dependencies of the factors, and to implement other copulas besides the Gaussian and \( t \) copula. Another interesting subject for future research would also be to consider the dependence between factors as dynamic, not only in correlation as mentioned previously, but also in the strength of tail dependence.

The evidence in this thesis of different levels of tail dependence in different pairs of factors has important implications for the modelling of joint distributions of factor log-returns. If one uses only one \( t \) copula to model the joint distribution of several (> 2) factor log-returns with different pair-wise dependence structure, then a poor model fit will likely be the result. This suggests that in future research it might be beneficial to implement pair-copula constructions when modelling the joint distributions of more than two factor log-returns.
Chapter 6

Conclusions

The results of the univariate analysis in section 4.1 indicate that there is autocorrelation and conditional heteroscedasticity in the daily factor log-returns. The use of ARMA-NGARCH processes proved useful in the modelling of these log-returns, and it was found that the driving noise sequences of these estimated processes were almost always closer to a skewed $t$ distribution than a normal distribution. It was shown that in two of the factors clear leverage-type effects were present. In the MKT factor this leverage effect was of the traditional type, were large negative outcomes increase the variability in future log-returns more than positive outcomes of similar size. For the SMB factor, an opposite leverage effect was present, in which large positive outcomes increase the variability in future log-returns more than negative outcomes of similar size. The dynamics of these two factors were also the most stable of the six factors. The stability of the estimated models of the log-returns was found to be the most problematic element of the modelling procedure. The results suggest that the parameters of the ARMA-NGARCH models that describe the factor log-returns are not stable. The detection of changes in these parameters is an interesting topic for future research.

The application of the fitted univariate ARMA-NGARCH processes to estimate VaR proved to be very rewarding and clearly outperformed the more simplistic modelling approach of using the empirical distributions of the log-returns. This is possibly the most decisive result of this thesis, as it motivates the use of more complex models and indicate that the fitted ARMA-NGARCH models provide a good approximation of the factor log-returns.

The results of the multivariate analysis in section 4.2 are more indecisive than the corresponding results in section 4.1. The copula-based approach were driving noise sequences of ARMA-NGARCH processes are made dependent
is evidently a very flexible and intuitive way of forming joint distributions of factor log-returns. The importance of modelling the dependence between factors is highlighted by the fact that many of the correlations between the factors were found to be very large. Furthermore, the results of the modelling of the bivariate joint distribution indicate that at least the MKT and MOM factors have fairly strong tail dependence.

The results for the three bivariate joint distributions of factor log-returns analysed in this thesis are promising. The models formed for these joint distribution gave VaR predictions that outperformed predictions based on the empirical distributions of log-returns. This again motivates the use of the complex models developed in this thesis.

However, the joint distribution of factor log-returns seems to be quite dynamic. This is evident in the large fluctuation in some of the correlations between factors. The unstable correlations between the factors is not surprising, as dynamics correlations in financial data and models for such varying dependence structures are well-studied [30] [33]. Furthermore, in previous studies of factor dynamics in the U.S. market, models for dynamic correlation are used [6] [7].

An important result of the copula modelling in this thesis is that the level of tail dependence seems to differ greatly between different pairs of factors. The bivariate $t$ copulas estimated for the joint distribution of the MKT and MOM factors has degree of freedom $\nu \in [2.81, 5.63]$. The corresponding parameter for the $t$ copulas for the MKT and HML factors is much higher, $\nu \in [10.56, 18.93]$. As the degrees of freedom $\nu$ controls the level of tail dependence in a $t$ copula, with a lower $\nu$ increasing the tail dependence, this indicates that the MKT factor has different levels of tail dependence with different factors. This implies that in the modelling of the joint distribution of several factors it might be beneficial to use pair-copula constructions as described in [54], rather than a large-dimensional copula.
Bibliography


