

Copula-GARCH modelling with Factor Structure for stock returns

Author

Hespera Henzell

Supervisor

Dr. Pavel Krupskiy

*A thesis submitted for the degree
of Master of Science*

School of Mathematics and Statistics
The University of Melbourne

May, 2022

Acknowledgements

With many, many thanks to my supervisor, Dr. Pavel Krupskiy. Thank you for your support, your generosity with your time, and for always steering this project in interesting directions.

Thank you also to all the staff and students in the Mathematics Department. Many of you have been kind enough to give me your thoughts over the course of this project, and they have always been fruitful.

Thank you to my family, for all your support.

Contents

1	Introduction	9
1.1	Joint models for stock returns	9
1.2	Sklar's theorem	11
1.2.1	The Copula Invariance Principle	12
1.2.2	The Inversion Method	12
1.2.3	Using uniform and normal scores to check dependency	13
1.3	IFM method	15
1.4	Desirable Features of Copula-GARCH models	15
1.4.1	Overview of Copula-GARCH models	17
1.5	Structure of the thesis	19
2	Modelling the margins	20
2.1	Marginal Models	20
2.1.1	ARMA models	20
2.1.2	ARCH and GARCH models	21
2.1.3	ARMA-GARCH models with external covariates	24
2.1.4	Modelling in R with <code>rugarch</code>	24
2.1.5	Diagnostics for ARMA-GARCH models	24
2.2	Modelling	28
3	Fitting the copulas	33
3.1	Estimating marginal distributions	33
3.2	Checking scores for dependence	34
3.3	Estimating normal and t -copulas	36
3.3.1	Copulas with 1, 2 and 3 factor structure	38
3.3.2	Nested models	40
3.3.3	Bifactor models	42
3.3.4	Diagnostics to select the best copula model	44
3.4	Modelling and results	48
3.5	Radially asymmetric copulas	51

3.5.1	Squared copulas with factor structure	53
3.5.2	Chi-squared copula	53
3.5.3	<i>F</i> -copula	55
3.5.4	Fitting chi-squared copulas with Kendall's Tau	57
3.5.5	Fitting models with Minimum Distance Estimators (MDEs)	58
3.5.6	Diagnostics	59
3.5.7	Results	60
3.5.8	Simulation Study	64
4	Discussion	66
Appendices		69
A	Stock tickers & Marginal Models	70
B	Gradient for Copulas With 1, 2 or 3 Factor Structure	72
B.1	1 Factor normal model	72
B.2	2 Factor normal model	74
B.3	3 Factor normal model	75
B.4	1, 2 & 3 factor <i>t</i> -model	77
B.4.1	Loadings matrix parameters	77
B.4.2	Shape/degrees of freedom parameter	78
C	Results for normal & <i>t</i>-copulas	81
D	Results for Chi-squared & <i>F</i>-copulas	91
E	Constructing the bivariate <i>F</i>-copula	107
F	Pearson correlation in the multivariate Chi-squared & <i>F</i>-distributions	109
F.1	Pearson correlation for the Chi-squared distribution	109
F.2	Pearson correlation for the <i>F</i> -distribution	110
G	Simulation study results	113
H	Normal scores scatterplots	115
Bibliography		115

List of Figures

1.1	Metlife log-returns, 2010-2019	10
1.2	Normal scores for bivariate normal and Student t -copulas	14
2.1	Simulations of ARCH(1)- t and GARCH(1,1)- t models	23
2.2	ACF and PACF plots for AR(2), MA(2) and ARMA(2,2) processes . . .	26
2.3	Example of ACF and PACF plots when choosing models	30
2.4	QQ-plots of residuals from fitted ARMA-GARCH- t models	32
3.1	Normal scores of select stocks	35
3.2	Structure of the nested graph fitted to 40 stocks in 4 sectors	41
3.3	Structure of the bifactor model fitted to 40 stocks in 4 sectors	43
3.4	Chi-squared and reflected chi-squared distributions	52
H.1	Normal scores scatterplots of log stock returns in the technology industry	116
H.2	Normal scores scatterplots of log stock returns in the pharmaceutical industry	117
H.3	Normal scores scatterplots of log stock returns in the financial industry .	118
H.4	Normal scores scatterplots of log stock returns in the consumer staples industry	119
H.5	Normal scores scatterplots of log stock returns between stocks in the technology and pharmaceutical industries	120
H.6	Normal scores scatterplots of log stock returns between stocks in the consumer staples and financial industries	121

List of Tables

3.1	Performance of different t -copula models by AIC and BIC	48
3.2	LRT for the elliptical copulas	49
3.3	Time taken to fit elliptical models	51

3.4	Time to fit chi-squared and F -copulas	60
C.1	Results for normal copula with 1 factor structure	81
C.2	Results for t -copula with 1 factor structure	82
C.3	Results for normal copula with 2 factor structure	83
C.4	Results for t -copula with 2 factor structure	84
C.5	Results for normal copula with 3 factor structure	85
C.6	Results for t -copula with 3 factor structure	86
C.7	Results for nested normal copula	87
C.8	Results for bifactor normal	88
C.9	Results for nested t -copula	89
C.10	Results for bifactor t -copula	90
D.1	Results for 1 factor χ^2 copula, fitted with pairwise likelihood	92
D.2	Results for 2 factor χ^2 copula, fitted with pairwise likelihood	93
D.3	Results for 3 factor χ^2 copula, fitted by pairwise likelihood	94
D.4	Results for 1 factor χ^2 copula, fitted with Kendall's Tau	95
D.5	Results for 2 factor χ^2 copula, fitted with Kendall's Tau	96
D.6	Results for 3 factor χ^2 copula, fitted with Kendall's Tau	97
D.7	Results for 1 factor χ^2 copula, fitted with MDEs	98
D.8	Results for 2 factor χ^2 copula, fitted with MDEs	99
D.9	Results for 3 factor χ^2 copula, fitted with MDEs	100
D.10	Results for 1 factor $F_{1,\nu}$ copula, fitted with pairwise likelihood	101
D.11	Results for 2 factor $F_{1,\nu}$ copula, fitted with pairwise likelihood	102
D.12	Results for 3 factor $F_{1,\nu}$ copula, fitted with pairwise likelihood	103
D.13	Results for 1 factor $F_{1,5}$ copula, fitted with MDEs	104
D.14	Results for 2 factor $F_{1,5}$ copula, fitted with MDEs	105
D.15	Results for 3 factor $F_{1,5}$ copula, fitted with MDEs	106
G.1	Simulation study results for χ_1^2 model fitted with pairwise likelihood	113
G.2	Simulation study results for χ_1^2 model fitted with Kendall's Tau	113
G.3	Simulation study results for χ_1^2 model fitted with MDE method	114
G.4	Simulation study results for $F_{1,5}$ -model fitted with MDE method	114
G.5	Simulation study results for $F_{1,5}$ -model fitted with pairwise likelihood	114

Notation

This list is incomplete. However it contains most of the notation used consistently throughout this thesis. Efforts were made to use standard notation, and not to reuse notation, wherever possible.

d for number of dimensions (number of stocks)

n for sample size (number of days stock prices were taken on)

C and c for the copula distribution function and density

U_i for Uniformly distributed random variable i , typically on $[0, 1]$, not to be confused with:

U_ν for chi-squared random variable, with ν degrees of freedom

χ_ν^2 also for chi-squared random variable, with ν degrees of freedom

S_i for t -distributed random variable with mean 0

Z_i for standard normal random variable

V_i for standard normal random variable

W_i for chi-squared or F random variables

Σ for covariance matrix, or scale matrix where covariance does not exist

\mathbf{L} for loadings matrix

D for diagonal matrix containing values such that $\Sigma = \mathbf{L}^T \mathbf{L} + D$

$\det(\Sigma)$ for determinant of Σ

\mathbf{I}_{k^*} for $k^* \times k^*$ identity matrix

$\phi(\cdot)$ for density function of standard normal distribution

$\phi(\dots; \Sigma)$ for density function of multivariate standard normal distribution with covariance matrix Σ

$\Phi(\cdot)$ for distribution function of univariate standard normal distribution

$\Phi(\dots; \Sigma)$ for distribution function of multivariate standard normal distribution with covariance matrix Σ

$\Phi^{-1}(u)$ for inverse transformation of quantile u to univariate standard normal density

ν is for degrees of freedom

t_ν for univariate density of t -distribution with ν degrees of freedom, also to refer to the

t_ν -distribution

$t_\nu(\dots; \Sigma)$ for density function of multivariate t -distribution with ν degrees of freedom and covariance matrix Σ

$T_\nu^{-1}(u)$ for inverse transformation of quantile u to density of univariate t -distribution with ν degrees of freedom

$\Gamma(\cdot)$ for Gamma function

$F(\cdot)$ for Digamma function

t for time t (not to be confused with t_ν)

y_t for A random variable taken at time t in a time series

ε_t for residual from the ARMA-GARCH models at time t

h_t for conditional variance function of a GARCH model at time t

ω for intercept in ARMA and GARCH models

α, β for coefficients in the ARMA and GARCH models

ℓ for log-likelihood

\mathcal{L} for likelihood

\hat{r}_k for k -th sample autocorrelation

$Q(\hat{r})$ for Ljung-Box Q -statistic

c_ρ^χ for density of the bivariate chi-squared copula with Pearson correlation ρ^2

c_ρ^N for density of the bivariate normal copula with Pearson correlation ρ

c_ρ^t for density of the bivariate t -copula with Pearson correlation ρ

c_ρ^t for density of the bivariate t -copula with Pearson correlation ρ

c_ρ^F for density of the bivariate F -copula based on the above t -copula

τ for Kendall's Tau

ρ for Pearson correlation

ρ_S for Spearman rank correlation

ρ^- for lower semicorrelation

ρ^+ for upper semicorrelation

$\mathbf{1}$ is an indicator variable

q for quantile

q_f for reflected quantile

$\chi^2, (\chi_\nu^2)$ for chi-squared distribution (with degrees of freedom ν)

$F_{1,\nu}$ for F -distribution with 1 and ν degrees of freedom

\mathbb{E} for expectation

VAR for variance, COV for covariance, CORR for correlation

Abbreviations

ACF for Autocorrelation function

AIC for Akaike Information Criterion

ARCH model for Autoregressive conditional heteroskedastic model

ARMA model for Autoregressive moving average model

BIC for Bayesian Information Criterion

EDF for Empirical distribution function

GARCH for Generalised autoregressive conditional heteroskedastic model

IFM for Inference method for margins

LRT for Likelihood Ratio Test

MDE for Minimum Distance Estimator

PACF for Partial autocorrelation function

QQ plot for Quantile-quantile plot

VaR for Value at risk

VIX for Ticker for the CBOE volatility index

1: Introduction

1.1 Joint models for stock returns

Joint models for stocks in a portfolio need to fit the whole shape of the data well. Clearly, investors are interested in metrics like the Expected Return, or the amount of profit they are predicted to make on average. To model these, we need a good fit at the centre of the data. Then, in order not to lose a portfolio, chasing profit needs to be balanced against the risk of catastrophic losses. These are in the bottom tail. One risk metric is the Value at Risk (VaR). This is a quantile of the joint distribution of the stocks in a portfolio; typically the 5% quantile. The idea is to examine the maximum loss expected if we exclude the most extreme losses. A useful model for this metric needs to have the 5% quantile fall fairly close to reality. Once we know the VaR, we can calculate the Expected Shortfall, or how much money we expect a portfolio to lose once we have fallen below the VaR. To estimate this, we need a model that gives a good approximation of the shape at the very tail of the distribution. Modelling for inference in the tail is quite challenging, due to there being little data there.

Unfortunately, stocks behave in ways that make modelling difficult in general. Three phenomena are of particular interest to this project. The first is volatility clustering. This refers to the way stock returns go through periods of higher and lower volatility (Mandelbrot 1963, Bollerslev & Andersen 2018), and can be seen in figure 1.1. Thus the variance of a log stock return* is correlated in time, which needs to be accounted for to accurately measure the VaR, particularly during periods of heightened volatility.

Secondly, stock prices also tend to rise somewhat independently, but fall together. Thirdly, periods of high market volatility are associated with falling prices and negative returns. This is called the leverage effect, and its discovery is attributed to Black (1976, cited by Bollerslev & Andersen 2018). So stock returns will exhibit temporal dependence between

*The reason we model with the logarithm of the stock return, rather than the raw stock return, is explained in Chapter 2

observations, but they will also exhibit this other, asymmetrical dependence between stocks.

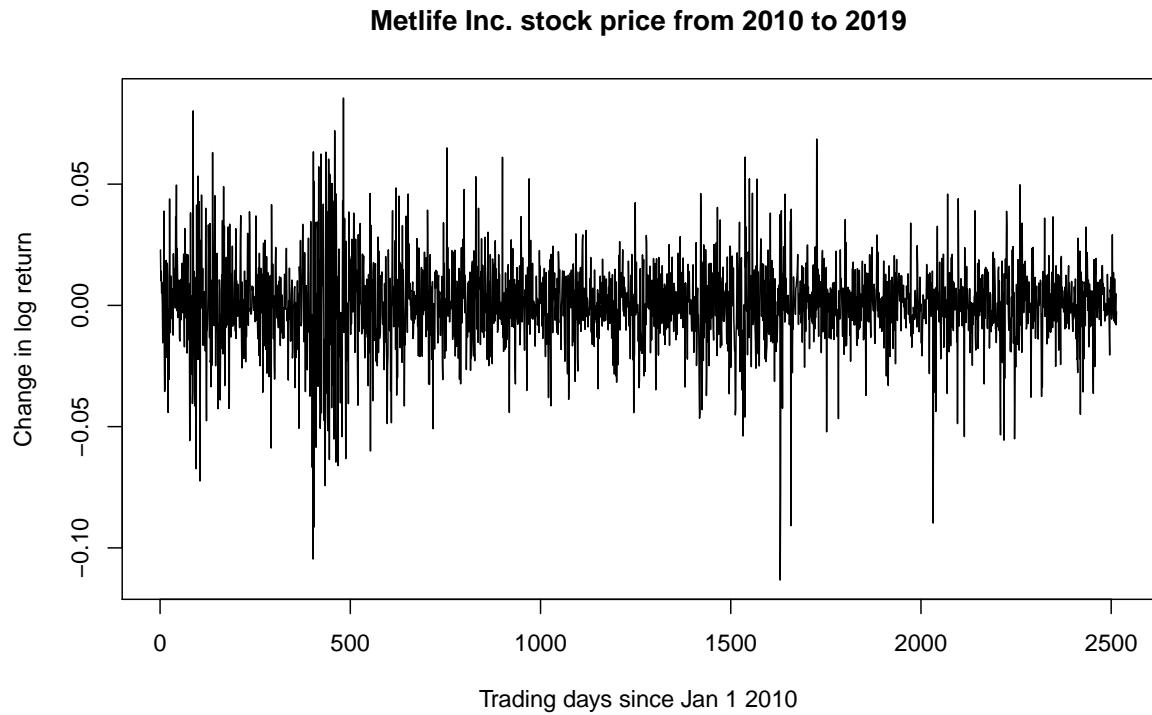


Figure 1.1: The log stock return of Metlife (ticker: MET) between 2010 and 2019. We can see volatility clustering, with two periods of high variance in the first 500 trading days, followed by a smaller returns over the next 100 days

Thus choosing a good joint model for a stock portfolio is not a trivial task. Alongside fitting the centre and the tail of the data well, it needs to be robust to these three phenomena.

A joint distribution function specifies both the marginal distributions, or margins, and the dependence between variables. A problem arises when we want to use univariate margins that do not combine easily into a joint distribution that models the dependence well. In this case we can utilise copulas.

A copula is a function that when applied to the univariate distribution functions give the joint distribution, first defined by Abe Sklar (1959). As such, it separates modelling the dependence and the margins in joint distributions. Provided we can find the right copula, we can give a set of fitted margins a dependence structure from a completely different distribution, allowing us great flexibility in modelling the joint distribution.

The Copula-GARCH model is a popular method for financial instruments. It is a flexible multivariate model, able to model both the complex dependencies between stocks, and the autocorrelation in the log stock returns (Jondeau & Rockinger 2006, Patton 2006). The univariate distributions are residuals from ARMA-GARCH models, fitted to the log stock returns. These models are used to remove temporal dependence from time series. If an ARMA-GARCH model is well fitted, its residuals should be stationary. These marginal models are explained further in chapter 2. The copula is fitted to the quantiles of the residuals from the ARMA-GARCH models.

While copulas can simplify dependence modelling, it is still far from trivial, and poor choices can be costly. A misspecified models may severely underestimate risk. Some have blamed global financial crisis of 2007-2008 on the inappropriate use of normal copulas for mortgage-backed securities, which underestimated the true risk of a large number of mortgage defaults at one time (Donnelly & Embrechts 2010).

This thesis focuses on designing copulas which model the dependence between stocks well, including the covariance in the data, and the strong dependence in the lower tail that occurs when the stock price drops.

1.2 Sklar's theorem

According to Sklar's theorem (Sklar 1959, Nelson 2006), if $G(X_1, \dots, X_n)$ is the joint distribution function for random variables X_1, \dots, X_n with marginal distribution functions F_{X_1}, \dots, F_{X_n} , then there exists some copula C such that for all (x_1, \dots, x_n) in the Cartesian product $\mathbb{R}_1 \times \dots \times \mathbb{R}_n$,

$$G(x_1, \dots, x_n) = C(F_{X_1}(x_1), \dots, F_{X_n}(x_n)) \quad (1.1)$$

Moreover, C is unique if the marginal distributions are continuous.

Since univariate distribution functions are uniformly distributed, the copula itself is joint distribution function of n univariate random variables. As such, it has identical mathematical properties to the multivariate uniform distribution function. Let U_1, \dots, U_n be n random variables on $[0, 1]^n$. An n -dimensional copula, C , is a continuous, non-decreasing function that maps $[0, 1]^n$ to $[0, 1]$, with the properties:

- If any of u_1, \dots, u_n is zero, then $C(u_1, \dots, u_n) = 0$
- $C(1, \dots, 1, u_j, 1, \dots, 1) = u_j$ for any $j = 1, \dots, n$.

- For any n -dimensional hypercube B with all vertices in the unit cube, such that $B = \prod_{i=1}^n [x_i, y_i] \subset [0, 1]^n$, the volume of C in B is non-negative, i.e.,

$$\sum_{\mathbf{b} \in B} \text{sgn}(\mathbf{b}) C(\mathbf{b}) \geq 0$$

where

$$\text{sgn}(\mathbf{b}) = \begin{cases} 1 & \text{an even number of } b_k = x_k \text{ for } k = 1, \dots, n \\ -1 & \text{an odd number of } b_k = x_k \text{ for } k = 1, \dots, n \end{cases}$$

1.2.1 The Copula Invariance Principle

Assume X_1, \dots, X_n is continuous. Schweizer and Wolff (1981) showed that if f is a strictly increasing function then $f(X_1), \dots, f(X_n)$ has the same copula as X_1, \dots, X_n . Transforming a distribution to its quantiles (the probability integral transform) is a strictly increasing function. So is transforming those quantiles to another continuous distribution function. Thus, by the *Copula Invariance Principle*, we can transform the data to a new set of marginal distributions without changing the copula. The dependency the copula contains is really the dependency between the quantiles of the random variables that form it.

1.2.2 The Inversion Method

Again, let us assume X_1, \dots, X_n are continuous. Let $F_{X_1}^{-1}, \dots, F_{X_n}^{-1}$ be the inverse functions of F_{X_1}, \dots, F_{X_n} , respectively. Moreover, let $F_{X_i}(X_i) \sim U_i$ for $i = 1, \dots, n$. Then U_i gives the quantiles of X_i , and is uniformly distributed on $[0, 1]$. Then we can rewrite equation (1.1) as:

$$C(u_1, \dots, u_n) = G(F_{X_1}^{-1}(u_1), \dots, F_{X_n}^{-1}(u_n)) \quad (1.2)$$

The Inversion Method exploits this to make constructing copulas easy. If we obtain the quantiles of all the random variables, we can choose the joint distribution and margins so they will form a copula with the dependence structure we want to use.

For example, say we have fitted marginal distributions and transformed them into the quantiles u_1, \dots, u_n . Then we can transform them into standard normal densities. If we say the joint model is multivariate normal with covariance matrix Σ . Then, using formula (1.2), we will have a normal copula that holds the joint dependency for the model in Σ , like so:

$$C(u_1, \dots, u_n) = \Phi(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_n); \Sigma)$$

1.2.3 Using uniform and normal scores to check dependency

Say we have a set of n data points, either from a known distribution, or an unknown one. If they are from a known distribution, we can transform them to a set of quantiles. Otherwise we can rank the observations, and divide them by $n+1$ to create the *empirical distribution function* with the set of *empirical quantiles*[†]:

$$\left\{ \frac{1}{n+1}, \dots, \frac{n}{n+1} \right\}$$

By either method, since quantiles have a uniform distribution, we have transformed the data to a *uniform score*. A *normal score transformation* is transforming the quantiles to quantiles of the normal distribution. We are left with the *normal score*.

If we have joint dependency of a certain type, then the bivariate density between two of the marginal distributions should be consistent with this. Since we cannot visualise the multivariate dependence between all variables on a flat piece of paper, we instead plot the bivariate dependence between two variables, to check if it matches the multivariate pattern of dependence we hypothesise that it has.

Dependency can be checked by plotting the scores of two variables against each other. It is common to use uniform scores, however in this thesis we follow the approach of Nikoloulopoulos et al. (2012), who used *normal scores scatterplots* to determine bivariate distributions. It can be easier to identify distributions on normal scores scatterplots rather than uniform scores scatterplots.

As the name implies, a normal scores scatterplots is a scatterplot of the normal scores of two variables. It should show an ellipsis if the dependency in the data is bivariate normal. Different distributions show characteristic shapes of dependence on a score plot. Figure 1.2 shows examples of bivariate normal and t -copulas plotted on a normal scores scatterplot. The t -copula forms a diamond shape on the normal score plot, with sharp tails. If the dependence in the data is well modelled by a particular distribution, it follows that the normal scores scatterplot of the data would look similar to the normal scores scatterplot of that distribution.

[†]An alternate definition of the empirical distribution function and quantiles divides the ranks by n instead of $n+1$, giving the empirical quantiles values in the set $\left\{ \frac{1}{n}, \dots, \frac{n-1}{n}, 1 \right\}$

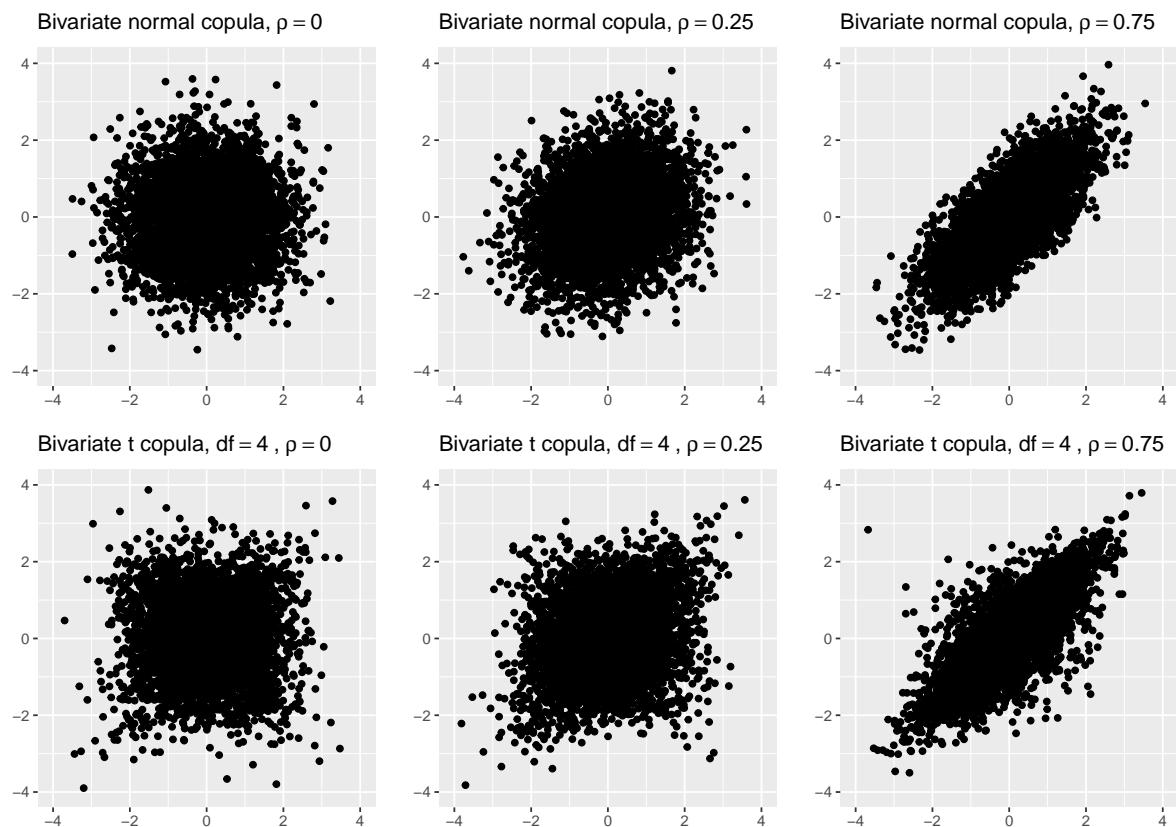


Figure 1.2: Normal scores for bivariate normal and Student t -copulas

1.3 IFM method

The full process of fitting a copula to data involves estimating all dependence parameters as well as the parameters for this univariate margins. In general, copula parameters do not have simple analytic formulas for their estimators, and must be fit through some optimisation procedure.

The *inference function for margins (IFM) method* (Joe & Xu 1996, Xu 1996) is an alternative to fitting multivariate models with full likelihoods. It can be used on models where some parameters are associated with the margins, and as such is perfect for copulas. Rather than fitting the whole likelihood at once, the univariate margins are fit first. This fixes all the parameters in the univariate margins. The joint likelihood is then fitted, which gives us estimators for the joint dependency parameters. Under regularity conditions, that hold for the models in this thesis, the IFM estimators are asymptotically normal and unbiased (see Joe & Xu (1996) for details on regularity conditions). Moreover, Xu's simulation studies have shown the IFM estimators to be almost as efficient as maximum likelihood estimators in many cases (Joe & Xu 1996, Xu 1996).

The trade off for losing some efficiency compared to maximum likelihood is a potentially massive gain in computational tractability, with fitting a reduced number of parameters in stages making solving some problems even feasible. In this thesis we follow the IFM method by fitting the marginal distributions with maximum likelihood, and then fit a variety of copulas with different dependence structures to these margins. Without the IFM method we would have to re-fit the margins for every model.

We follow the IFM method for the majority of the thesis, however we do explore fitting some copulas with Kendall's Tau and minimum distance estimators rather than likelihood.

1.4 Desirable Features of Copula-GARCH models

While there is a plethora of options for dependence structures in Copula-GARCH models for stock returns, it remains an active area of research. Designing multivariate models for real world applications is rarely easy, even using copulas. This thesis explored balancing features we would ideally want in a good model, including tractability, a good fit to the data, and interpretability.

Multivariate models tend to suffer from the curse of dimensionality, where large models are many times more difficult to fit than smaller models. For example, take fitting covariance. Adding 1 dimension to a d -dimensional model creates d more covariances between variables, whereas adding 3 more dimensions creates $3d + 3$ more. A dataset of 50 or more variables can be considered to be high dimensional in copula modelling. This may be sufficient to fit say, a moderately sized index of 100 stocks, but has not kept up with the size of modern “big” datasets, which can contain hundreds of thousands of variables.

Creating a tractable model involves creating both a model and a method to fit it, which together are computationally efficient. The typical way of fitting models is with maximum likelihoods, which is difficult for models whose likelihoods increase polynomially or even exponentially with increasing dimensions. The computational power required to fit these models quickly becomes prohibitive. Even alternative methods, based on fitting the model to match certain statistics can become slow when the statistic is difficult to calculate.

In practice, there is a huge pressure for rapid models. Volatile stock prices can update every second, and practitioners need models which can respond quickly. Hence the tractability of a model may limit the size of a model a researcher fits, but much more seriously limit what models are useful in practice.

We also need the models to fit the data well. This is a challenge as, due to stock prices tending to crash together, we expect the multivariate models to have a heavier lower tail than upper tail. On a normal scores scatterplot, this appears as a teardrop shape, pointing to the bottom left. The common, tractable elliptical distributions, such as the normal and Student’s t , do not show this asymmetry, making them a slight misfit. There are two obvious ways to deal with this fit. We can design a distribution that fits the teardrop shape well. We can also choose a distribution that does not fit all of the data well, but fits adequately enough for us to draw inferences from it.

An interpretable model is one where we have an explanation of how an underlying process changes the model. These models can provide us inference on these underlying processes. Moreover, it can be easier to find a cause when the model stops performing well, or predict when it will do so. In contrast, without interpretability, we can find a great fitting model but not fully understand the conditions we need to use it.

1.4.1 Overview of Copula-GARCH models

In this thesis, we fit non-dynamic Copula-GARCH models with factor structure. This is a strong assumption. In general we expect the dependency between stocks to be dynamic: that is, for it to change over time. Early Copula-GARCH models, such as those by Patton (2006) and Jondeau & Rockinger (2006), designed copulas with changing correlation parameters. These copulas had the flexibility to model changing correlation in financial data over many years. However, allowing correlation to change makes modelling more complex.

Structured factor copulas were first proposed by Klüppelberg & Kuhn (2009), when they proposed creating elliptical copulas with structured correlation matrices. Krupskii & Joe (2013, 2015) worked further on a subset of these models, with structures based on dependence on a small number of latent variables that defined the correlation matrices for elliptical copulas. These are the models used in this thesis, and are explained in detail in section 3. The most classic version is the copula with factor structure, where the dependency in the model is assumed to have the same structure as a latent factor model. As these models are based on elliptical copulas, they have well known likelihoods that can be calculated with a series of matrix operations, making them relatively quick to fit. There is now even been a rapid method based on the one factor model able to fit big data (Krupskii & Joe 2020).

Oh & Patton (2017) used a factor model structure without dynamic dependence to model a dataset of 100 stocks, noting that the relatively small number of constant parameters simplified the model sufficiently to fit high dimensional copulas. This is the approach taken in this thesis, with the simplification being justified by picking a stable period in the market, where correlation between stocks is not expected to vary widely. As an alternative, we could have used newer factor Copula-GARCH models that can model dynamic dependence, such as in Oh & Patton (2018) or Opschoor et al. (2020), but we may have been limited to fitting a smaller dataset.

Another popular choice is *Archimedean copulas*. They are a family that for one parameter set $\boldsymbol{\theta}$, have the structure:

$$C(u_1, \dots, u_n; \xi, \boldsymbol{\theta}) = \xi \left(\sum_{j=1}^d \xi^{-1}(u_j; \boldsymbol{\theta}) \right)$$

for some function ξ with the generalised inverse ξ^{-1} . Two important examples for stocks are the Clayton and Gumbel copulas. These copulas have asymmetrical dependence

that is strong in one tail, and show a teardrop shape on the normal scores scatterplot. Archimedean copulas use a small parameter set; typically just one parameter. This makes them quick to fit, but inflexible, as all pairs of stocks have the same dependence. This can be counteracted by creating structures based on lower dimensional Archimedean copulas. One such model is the nested Archimedean copula. The fully nested copula with $d - 1$ nesting levels is given by:

$$C(u_1, \dots, u_n; \xi_k, \boldsymbol{\theta}_k, 0 \leq k \leq d - 1) = \xi_0 (\xi_0^{-1} (u_j) + C(u_2, \dots, u_n; \xi_k, \boldsymbol{\theta}_k, 1 \leq k \leq d - 1))$$

where the nested copulas defined recursively by the same equation (Hofert 2010). In other words, nested copulas of lower dimensions are recursively used as a margin to create a larger nested copula. If, at some nesting levels we use multiple lower dimensional copulas as margins with a symmetric Archimedean copula, we have a partially nested copula. Since we can fit a different parameter sets and even copulas at different levels, the model becomes much more flexible in modelling dependence, but slower to fit. So there is a trade-off between good fit and tractability with the Archimedean copulas.

Another classic way to fit copulas are with structures called *vines*. Like nested Archimedean copulas, vines are based on a tree structure linking copulas which combine into larger structures, and eventually into a joint distribution. However, where Archimedean copulas use whole lower dimensional copulas, vines use lower dimensional *conditional copulas*. Recall that the copula is a continuous multivariate uniform distribution. Hence when you condition a copula on some of its marginal univariate distributions you get a lower dimensional conditional uniform distribution. It is still on a uniform scale, and so can be used in other copulas, or even conditional copulas. This process can be repeated until the original copula is expressed in terms of bivariate conditional copulas. Similar to fully nested Archimedean copulas, vines use these lower dimensional conditional copulas as margins in larger copulas. Creating a vine involves sequentially linking, conditioning and combining bivariate copulas into larger copulas, until a joint distribution is created (see, for example, Joe (2014) or Aas et al. (2009) for details). Since different parameters can be fit at different levels, vines are incredibly flexible models. But they are computationally expensive to fit, and since a joint distribution can be decomposed into a vine structure in many ways, we have the problem of choosing which vine to use. *Truncated vines* assume that after a certain point, all the lower dimensional conditional copulas are independent, which trades some flexibility for tractability.

These popular models illustrate the major trade off between tractability and flexibil-

ity required to fit copulas with current methods. There is active research into all of these methods. We focus on the structured factor copula models, that are created with specific covariance matrices. Their covariance matrices are simple enough to give them tractability in high dimensions, while retaining the flexibility to provide a reasonable fit to the data.

Another small group of models discussed in this thesis are copulas based on squaring elliptical distributions, which have asymmetrical dependence with a sharp upper tail. The copulas we use are based on the chi-squared and F -distributions. Currently the literature on these copulas focuses on geostatistical applications (Bárdossy 2006, Quessy et al. 2016). Like elliptical copulas, they have a covariance matrix, which allows us to fit them with similar methods to factor copulas. However, unlike elliptical copulas, their likelihoods are challenging to fit, forcing us to come up with other methods.

1.5 Structure of the thesis

The main part of thesis is an application of Copula-GARCH models to a set of 40 stocks from 4 different sectors. Since we are not modelling dynamic dependence, we choose a stable period in the market. In chapter 2, we briefly survey the literature of ARMA-GARCH models, before using them to fit the marginal models.

In chapter 3, we fit the residuals of these models with a range of structured factor copula models for elliptical copulas, and compare the fit of the models. Then we extend the covariance structure of these models to chi-squared and F -copulas, to fit factor models with an asymmetric dependence structure.

Next, we try fitting the chi-squared copulas with Kendall's Tau, and fitting both chi-squared and F -copulas with a new minimum distance method where we match the covariance in the models to the covariance in the data. Finally, we perform a small simulation study, testing different methods of fitting the chi-squared and F -copulas.

2: Modelling the margins

2.1 Marginal Models

Before we can create our copulas, we need to fit univariate marginal models to each stock. ARMA-GARCH models are well established as a standard way of fitting log stock returns (Bollerslev & Andersen 2018). The price of stock returns increases with time, and stocks with higher prices are expected to have a larger change in returns. Since we want to compare relative changes in return (in percentage) across time and different stocks, we fit the models to the logarithm of the stock returns, or the log stock returns. During stable periods of the market, the log stock returns are roughly covariance stationary, but correlated in time in the first two moments, due to factors such as volatility clustering. ARMA-GARCH models are able to explain this temporal correlation. Moreover, if we manage to explain all the correlation with the model, the residuals should be identically distributed and independent, at least in the first two moments. This is ideal data to form our copula with: the temporal correlation has been removed, leaving the other correlation between different stocks which we are more interested in. One of the causes of volatility clustering is the leverage effect: where returns are lower at times of increasing volatility in the market (Black 1976, via Andersen & Bollerslev 2018). This may cause an asymmetric effect in the underlying data where a negative error behaves differently to positive error. A volatility covariate is used to model this asymmetric effect.

2.1.1 ARMA models

In a time series, an *Autoregressive (AR)* model predicts the next value in the series based on past values. A *Moving Average (MA)* model predicts the next value based on previous residuals. The *Autoregressive Moving Average (ARMA)* model combines these two models.

Let y be a time series, and y_t be the random variable representing the value of y at time t . Furthermore, let ψ_t be all the information available at t and ε_t be the residual at

t . The ARMA(p,q) model will predict the value of y_t based on the p previous realisations of y and the q previous residuals in the model. The predicted value of y_t is:

$$\mathbb{E}(y_t|\psi_{t-1}) = \omega + \sum_{i=1}^p \alpha_i y_{t-i} + \sum_{i=1}^q \beta_i \varepsilon_{t-i}$$

In the most basic ARMA models, the variance is kept constant, rather than conditioned on previous values. Therefore:

$$\text{Var}(y_t) = \text{Var}(\omega) = \sigma_\omega^2$$

Thus ARMA model can remove autocorrelation in the residuals . However, it can not remove the autocorrelation in the squared residuals caused by volatility clustering (Box et al. 2015).

2.1.2 ARCH and GARCH models

ARMA modelling alone can not remove correlations in the squared residuals caused by volatility clustering. Engle (1982) introduced the class of *Autoregressive Conditional Heteroskedastic (ARCH)* models to deal with this problem. These models are based on conditional variances which are dependent on the size of previous error values.

Now suppose we have a set of temporal data that is stationary in the mean. Let ε_t be the error and ψ_t be all the information available at time t . The class of ARCH(q) models is defined by having:

$$\mathbb{E}(\varepsilon_t|\psi_{t-1}) = 0,$$

$$\text{Var}(\varepsilon_t|\psi_{t-1}) = h_t, \quad h_t = h(\varepsilon_{t-1}, \dots, \varepsilon_{t-q})$$

where h is some variance function.

While ARCH models can remove correlations in the squared residuals, sometimes they are not good enough. In particular, the lag structure of the ARCH model declines quickly (see figure 2.1). Bollerslev (1986), desiring a model that allowed a longer lag structure more typical of empirical data, created the *Generalised Autoregressive Conditional Heteroskedastic (GARCH)* model, which conditions variance on both the previous errors and previous variances. The class of GARCH(p,q) models is defined by having the following conditional distribution of the errors:

$$\mathbb{E}(\varepsilon_t|\psi_{t-1}) = 0,$$

$$\text{Var}(\varepsilon_t | \psi_{t-1}) = h_t, \quad h_t = h(\varepsilon_{t-1}, \dots, \varepsilon_{t-q}, h_{t-1}, \dots, h_{t-p})$$

where h is some variance function. GARCH(0,q) models are ARCH(q) models.

These are the most general structures for the whole class. The ARCH(p) model generally refers to a model with the variance function at time t , h_t , is:

$$h_t = \omega + \sum_{i=1}^p \alpha_i \varepsilon_{t-i}^2$$

Likewise, the GARCH(p,q) model is generally understood to have the variance at time t :

$$h_t = \omega + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^p \beta_i h_{t-i}$$

This is sometimes referred to Bollerslev's GARCH model, or the Standard GARCH model (Hentschel 1995), and are the models used in this thesis.

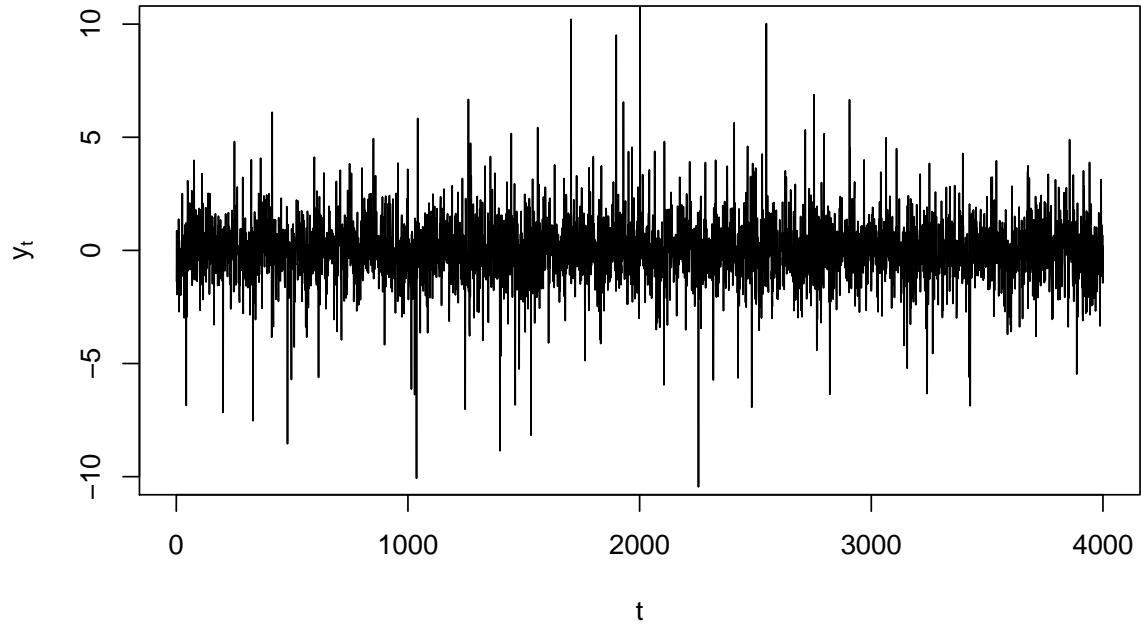
Figure 2.1 shows simulations of a ARCH(1) model and standard GARCH(1,1) model. We can see brief bursts of high variance in the ARCH(1) simulation, which quickly decay. The GARCH(1,1) shows the longer empirical memory Bollerslev desired: since shocks (increases in variance) decay slower than in the ARCH(1) example, we can clearly see periods of high and low variance grouped together. Empirically, GARCH(1,1) have been well established as the model of choice for log stock returns (Bollerslev et al. 1992, Bollerslev & Andersen 2018).

The GARCH (and ARCH) model also requires us to choose a conditional distribution for the errors, with a variance of 1. One of the more simple GARCH(p,q) models is the one Bollerslev originally proposes, where the errors are distributed normally, such that:

$$\begin{aligned} \varepsilon_t | \psi_{t-1} &\sim N(0, h_t), \\ h_t &= \omega + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^p \beta_i h_{t-i}, \\ \omega > 0, \quad \alpha_i, \beta_i &\geq 0, \quad p \geq 0, \quad q > 0 \end{aligned}$$

While the conditional distribution of this model is normal, the unconditional distribution is leptokurtic, but not enough to describe the degree of dependency in tails we see in empirical data (Bollerslev 1986). This leptokurticity has been found to be well modelled

Simulated ARCH(1) model



Simulated GARCH(1,1) model

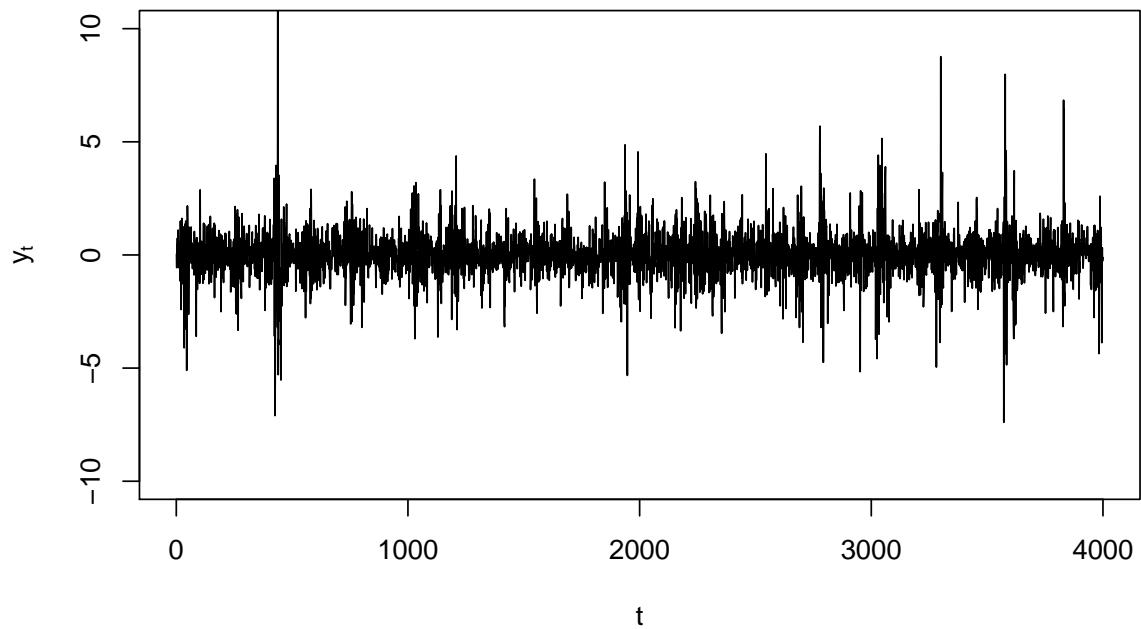


Figure 2.1: Simulations of a standard ARCH(1) model and standard GARCH(1,1) model with conditional Student's t_4 -distributions for the error. The parameters were chosen so the models have the same unconditional variance. However only the GARCH(1,1) model shows clear periods of high and low variance typical of volatility clustering. Parameters of the ARCH(1) model: $\omega = 0.1$, $\alpha_1 = 0.95$. Parameters of the GARCH(1,1) model: $\omega = 0.1$, $\alpha_1 = 0.25$, $\beta_1 = 0.7$.

by the Standardised Student's t -distribution with degrees of freedom estimated to match the thickness in the tails as the conditional distribution of the errors (Bollerslev 1987). The t -distribution is in scale family, and as such any classic Student's t -distribution can be divided by a scale parameter to change its variance. The resulting distribution is called a generalised Student's t -distribution, and has both a shape (or degrees of freedom) and scale parameter. As such, a classic Student's t -distribution with a shape parameter of $\nu > 2$ can be transformed to have variance of 1 by dividing it by its standard deviation. The resulting distribution is the standardised Student's t -distribution.

2.1.3 ARMA-GARCH models with external covariates

The ARMA-GARCH model allows us to fit the conditional expectation of a time series with an ARMA model and the conditional variance with a GARCH model. We can also add extra covariates to either part of the model.

It makes sense to directly add a volatility covariate to the model to explain the leverage effect, allowing use of an otherwise symmetric GARCH model.

The *CBOE Volatility Index* is a measure of volatility in the market based on the stock price of the S&P 500, with the ticker VIX. In our modelling we use a GARCH(1,1) model with VIX as an external covariate. The variance function h_t is:

$$h_t = \omega + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1} + \gamma \text{VIX}_{t-1}$$

2.1.4 Modelling in R with rugarch

The R package `rugarch` provides an environment for simulating, fitting and testing ARMA-GARCH models (Ghalanos 2020). The package allows us to fit models with external covariate, and a variety of conditional distributions for the error, including normal and standardised Student's t -distributions.

2.1.5 Diagnostics for ARMA-GARCH models

ACF and PACF plots

The autocorrelation function (ACF) of a time series finds the size of the correlation between current values and previous lags. Of course, if each point in the time series is

correlated with the previous point, then the previous point will be correlated with the point before it, and so on. Thus autocorrelation does not suddenly cut off at a certain lag, but rather the magnitude of it starts to slowly diminish. In R the command we use to compute the ACF, `acf`, computes the sample autocorrelations of the residuals (Venables & Ripley 2013). The k -th sample autocorrelation, \hat{r}_k , is given by:

$$\hat{r}_k = \frac{\sum_{i=k+1}^n \hat{a}_i \hat{a}_{i-k}}{\sum_{i=1}^n \hat{a}_i^2}, \quad k = 1, 2, \dots \quad (2.1)$$

where \hat{a}_i is the i th residual from an ARMA model.

We can use the partial autocorrelation function (PACF) to measure the autocorrelation at lag k that is not accounted for by autocorrelation between previous lags (Box & Pierce 1970). To find the PACF for lags $1, \dots, k$, we fit the equations:

$$\begin{aligned} y_t &= \phi_{11} y_{t-1} + \varepsilon_t \\ y_t &= \phi_{21} y_{t-1} + \phi_{22} y_{t-2} + \varepsilon_t \\ &\vdots \\ y_t &= \phi_{k1} y_{t-1} + \phi_{k2} y_{t-2} + \cdots + \phi_{kk} y_{t-k} + \varepsilon_t \end{aligned}$$

The PACFs for lag $1, \dots, k$ are $\phi_{11}, \dots, \phi_{kk}$ respectively.

In this project the ACF and PACF is evaluated using ACF and PACF plots, as shown in figure 2.2. These respectively graph the autocorrelation and partial autocorrecation at lag k . ACF and PACF is evaluated using 5 percent hypothesis tests, and the dotted lines in the plots are at the critical values of these tests, which are given by $\frac{2}{\sqrt{n}}$ where n is the size of the sample (Box et al. 2015).

An AR(p) process should show p significant lags on the PACF plot, and tail off on the ACF plot, and a MA(q) process should show q significant lags on an ACF plot and potentially tail off on the PACF plot. An ARMA process should cause the ACF and potentially the PACF plot to tail off. Figure 2.2 shows ACF and PACF plots of these three types of processes. If an ARMA model is correctly fitted, the residuals should not show significant correlation on either plot. In the same manner a correctly fitted GARCH model will remove partial and full autocorrelation in the squared residuals.

ACF and PACF plots do not always behave exactly as expected. Firstly, because they are a visual representation of a series of hypothesis tests, at times a lag may be large enough

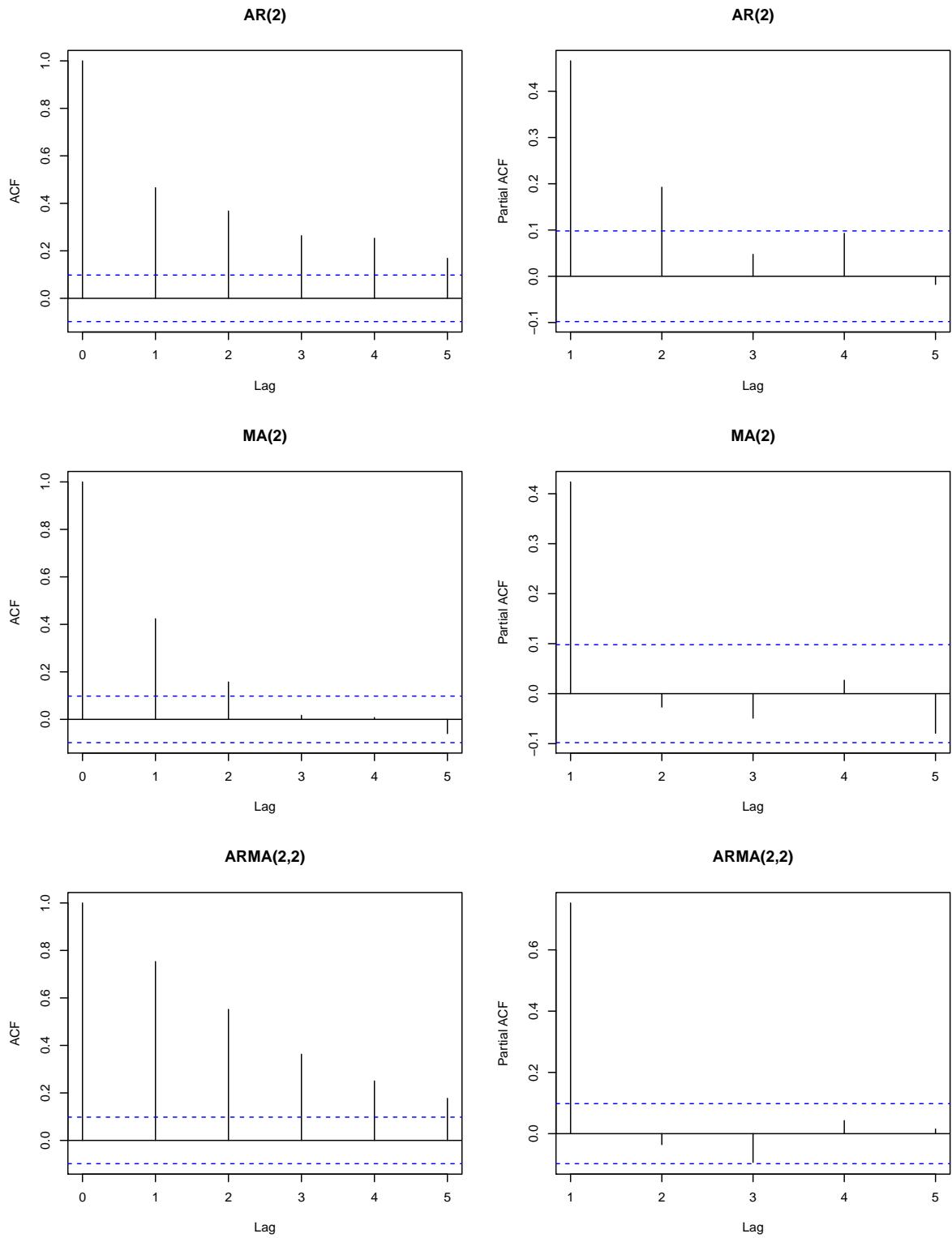


Figure 2.2: ACF and PACF plots of simulated AR(2), MA(2) and ARMA(2,2) processes, sample size = 400. Parameters for the AR(2) simulation: $\alpha_1 = 0.4$, $\alpha_2 = 0.2$, $\omega = 0.5$. Parameters for the MA(2) simulation: $\beta_1 = 0.4$, $\beta_2 = 0.2$, $\omega = 0.5$. Parameters for the ARMA(2,2) simulation: $\alpha_1 = 0.4$, $\alpha_2 = 0.2$, $\beta_1 = 0.4$, $\beta_2 = 0.2$, $\omega = 0.5$

to be significant when there is little correlation in the underlying process. Secondly, due to large variances and dependence between correlations, we may see subtle patterns on the ACF and PACF plots when correlation has actually died out. Strong patterns should be fairly clear, but weaker ones ones may need to be investigated further. In these situations, Box et al. (2015) advocate fitting and comparing multiple models. There is a danger to overfitting ARMA-GARCH models, as it can create serial correlation in the residuals.

Ljung-Box test

If an ARMA(p,q) model can adequately explain the variation in the data, and the parameters are correctly fitted, then the fitted autocorrelations (as defined in equation 2.1) should be some sort of white noise with a variance of 1. Then the statistic:

$$Q(\hat{r}) = n(n+2) \sum_{k=1}^m (n-k)^{-1} \hat{r}_k^2$$

is approximately χ^2_{m-p-q} distributed when n is large (Box & Pierce 1970, Ljung & Box 1978). If, however, there is still correlation present, because the magnitude of autocorrelation in the residuals diminishes rather than disappears, there should be significant correlation in $\hat{r}_1, \dots, \hat{r}_m$, so that the statistic is not approximated by a χ^2_{m-p-q} distribution, even when we have only underfit the model by a small number of lags.

The Q -statistic is slightly biased in variance, but has been found to be sufficient and robust for empirical work (Ljung & Box 1978). The Ljung-Box test is hypothesis test for lack-of-fit test, with the null distribution being that $Q(\hat{r})$ has a χ^2_{m-p-q} distribution (and therefore that the ARMA model is correctly fitted). If the p-value is larger than our desired significance level (typically 0.05), we have no evidence that our model is misspecified.

There have been a number of suggestions for the degrees of freedom parameter $m - p - q$ for the Q -statistic in the literature. The very idea of the test is to detect a pattern over a number of lags, however we should take care not to choose too many, as the degrees of freedom increases, the power of the test decreases. Recent results from simulations also suggest the actual size of the test can become increasing larger than the set α level as the degrees of freedom increase. Interested readers should refer to Hassani & Yeganegi (2020) for an investigation into these issues, and brief survey of suggested degrees of freedom in the literature. We choose 5 lags for a test with a significance level of 0.05. With 5 lags, we should have power over 0.9 and the actual size of the test should not exceed 0.05 (Hassani & Yeganegi 2020).

AIC

The Akaike Information Criterion (AIC) is used in model selection to discourage overfitting. As more parameters are added to a model its likelihood will improve, however if too many parameters are added the model will begin to fit noise. AIC penalises the increase in log-likelihood by the number of parameters added. In `rugarch`, it is calculated with the formula:

$$\text{AIC} = \frac{2k - 2\ell}{n}$$

where k is the number of fitted parameters in the model, n is the length of the time series, and ℓ is the maximum log-likelihood of the fitted model (Ghalanos 2020). When fitting with this criterion, the model with the lowest AIC is chosen.

QQ-plots

QQ-plots can be used on the residuals to ensure the marginal models are correctly specified. Of particular interest are evidence of heavy tails and skew. If for some reason the distribution of the conditional errors does not capture the leptokurtosis in the data we will see heavy tails, and will have to re-select our conditional error distribution. Skewness would indicate that the VIX covariate has failed to fully account for the asymmetry in variance due to the leverage effect, and we need to find a better way of removing asymmetry from the residuals.

The residuals from `rugarch` are generalised Student's t_{ν} -distributions with a scale parameter of $\sigma = (\frac{\nu}{\nu-2})^{1/2}$, but the package we wanted to use to create the QQ-plots, `car`, only creates QQ-plots the classic Student's t -distribution. Therefore residuals were multiplied by $\hat{\sigma}$, where $\hat{\nu}$ is the fitted degrees of freedom for the marginal distribution from `rugarch`, and:

$$\hat{\sigma} = \sqrt{\frac{\hat{\nu}}{\hat{\nu}-2}}$$

The transformed residuals were then compared to theoretical quantiles of the classic Student's $t_{\hat{\nu}}$ -distribution on a QQ-plot.

2.2 Modelling

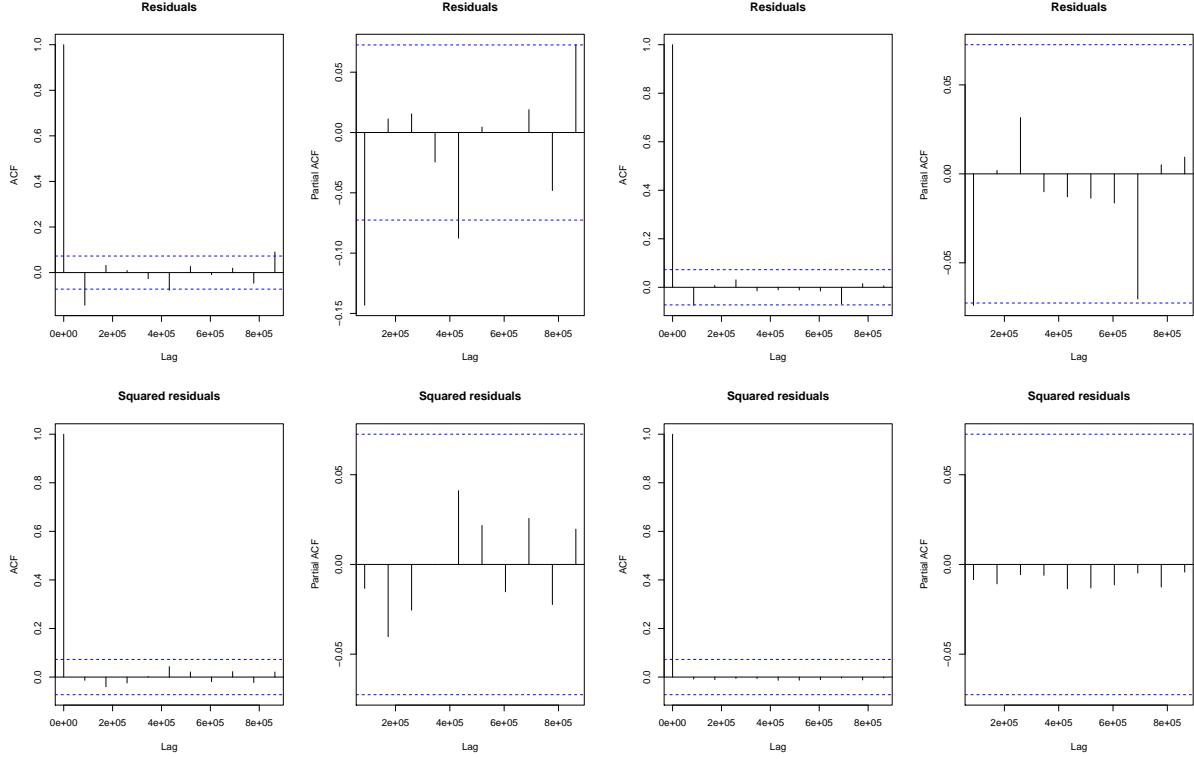
Stock returns for 4 sets of 10 stocks from 4 different sectors of the market for the period from January 1st 2016 to December 31st 2018 were taken from Yahoo Finance. The stock tickers are in Appendix A. The log stock returns were fitted with a Student's t -ARMA-

GARCH model in `rugarch`, with the VIX as an external covariate to model the leverage effect. All 50 stocks were first fit with an ARMA(0,0)-GARCH(1,1) model. The residuals and squared residuals were both checked by calculating the Q -statistic from Ljung-Box tests with 5 degrees of freedom on them, and by viewing the ACF and PACF plots for 10 lags. A dotted line at the critical value for hypothesis tests of no autocorrelation and partial autocorrelation was plot on the ACF and PACF plots, respectively. The AIC and BIC of the models were also calculated. Examples of the ACF and PACF plots are in figure 2.3.

When fitting the models, the Q -statistic and the ACF and PACF plots were analysed together to consider whether to fit with a higher order model. If the residuals or squared-residuals had a p -value below 0.05, or a high early lag, then larger models were fitted. Some stocks had high later lags in the ACF and PACF plots. In the cases where only a higher lag or two was above the dotted line, but the Q -statistic had a normal p -value, and there was not a concerning pattern in the ACF and PACF plots, this was thought to be due to chance.

When models were refitted with higher orders, the PACF and ACF plots were examined to determine what models were most likely. Typically it was not that clear: for most stocks, there were both spikes at lag 1 in the ACF and PACF plots, so ARMA(0,1)-GARCH(1,1), ARMA(0,1)-GARCH(1,1) and ARMA(1,1)-GARCH(1,1) models were all fit and compared. The most typical outcome is that all three models improved the ACF and PACF plots, with the ARMA(0,1)-GARCH(1,1) models having the lowest AIC. One interesting case was the stock ATVI. The ARMA(0,0)-GARCH(1,1) model had a spike at lag 1 in the ACF and PACF plots, but a p -value for the Q -statistic above 0.05, indicating no lack of fit over the first 5 lags. Higher order ARMA models were fit, which removed the spike, but also had higher AICs than the original model. As such, the original ARMA(0,0) model was considered to be the best fit, and the spike at lag 1 likely due to chance.

The only stock that needed higher order GARCH models to be considered was K. It had a spike in the ACF and PACF plots at lag 1 for the squared residuals, and its p -value for the Q -statistic on the squared residuals was 0.03. GARCH models of higher orders were investigated. The best performing was the ARMA(0,0)-GARCH(1,2) model, which lowered the spike in the plots a little to slightly below the dotted line, and improved the p -value of the Q -statistic to 0.16. It had a slightly higher AIC than the GARCH(1,1) model. There was a case for choosing both models, but since GARCH(1,1) models are a



(a) PGR, $Q(\hat{r}) < 0.001$, $Q(\hat{r}^2) = 0.53$

(b) ATVI, $Q(\hat{r}) = 0.42$, $Q(\hat{r}^2) = 0.99$

Figure 2.3: ACF and PACF plots for stocks PGR, first two columns, and ATVI, last two columns, when fitted with an ARMA(0,0)-GARCH(1,1)model. $Q(\hat{r})$ and $Q(\hat{r}^2)$ are the Q -statistics for the Ljung-Box tests on residuals and squared residuals respectively. The Q -statistics, ACF and PACF plot for residuals of PGR indicates they are still correlated, and a higher order ARMA model should be fitted. The plots for the squared residuals show no evidence of correlation. Despite a the first lag in the ACF and PACF plots for the residuals of ATVI touching the blue line, the Q -statistic does not indicate a lack of fit. After comparing with high order models, the ARMA(0,0) model was kept.

standard choice for stocks, we used the ARMA(0,0)-GARCH(1,1) model.

Once the marginal models were fitted, the residuals were transformed from standardised Student's t -distributions to classic Student's t -distributions, and checked against QQ-plots of the t -distribution with the degrees of freedom they had been estimated to be, with the command `qqPlot` in the `car` package. A 95 percent confidence envelope was also plotted on the QQ-plots. Overall, the QQ-plots indicated good fits. The majority of stocks had all observations fall within the confidence lines, including in the tails. The stocks SCHW and KO both had upper tails that fell at the edge of the confidence bands. They could be considered to have some asymmetry left in their fit. All points for the

stock AIG fell within the confidence envelope, but the points appeared to show a slight curve on the QQ-plot, indicating that `rugarch` may have fit the shape parameter a little low. The QQ-plots of these three stocks are shown in figure 2.4. However, none of the QQ-plots indicated a serious misfit, and so the fits were accepted.

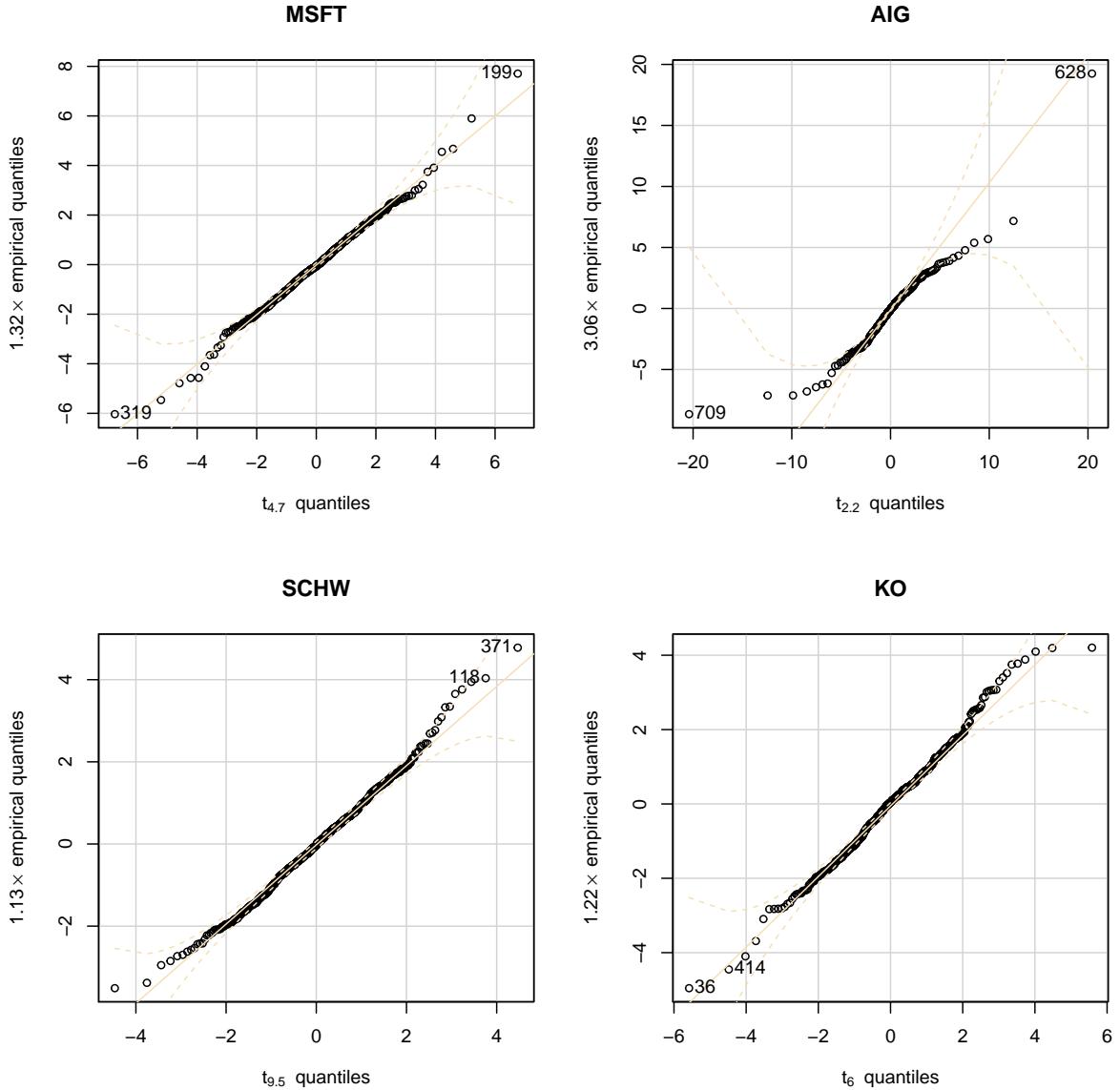


Figure 2.4: QQ-plots of residuals for MSFT, AIG, SCHW and KO with a 95 percent confidence envelope. The fit of MSFT appears to be fine. AIG shows some excess curvature, as if the degrees of freedom may have been slightly overfit, but falls within the confidence envelope. The upper tails of SCHW and KO fall at the edge of the confidence envelope, indicating an possibly heavy upper tail.

3: Fitting the copulas

3.1 Estimating marginal distributions

In the first stage of the IFM method, the marginal distributions of all the stocks were estimated. This involved taking the residuals of the fitted ARMA-GARCH models of the log stock returns. If the ARMA-GARCH models are fitted correctly, the residuals should at least be covariance stationary. From this we assume temporal stationary, and treat each day as a data point where the correlation in time has been removed, but the correlation between the performance of stocks on that day remains.

Since we are interested in the correlations between stocks, we fit our copula to the residuals from the ARMA-GARCH modelling, rather than the log stock returns. Hence, the residuals for each stock are taken to be our marginal distributions. All that remains to be done is convert them to quantiles. The quantiles of the empirical distribution were used. The residuals for each stock were given a rank, and then divided by the number of days n plus 1, to give a set of quantiles with the values:

$$\frac{1}{n+1}, \frac{2}{n+1}, \dots, \frac{n}{n+1}$$

The empirical quantiles are divided by $n + 1$ rather than n to avoid the value of the final quantile being 1. Likewise the first empirical quantile is chosen so it does not have a value of 0. This avoids the situation where transforming these quantiles to those of the normal or t -distribution will result in an infinite number.

The great advantage of using the empirical distribution function is that it speeds up fitting the copulas when performing a quantile transformation takes a long time. This is not important for the normal model, where the quantile transform can be done before fitting the copula, as none of the parameters being optimised effect the normal quantiles. By contrast, it speeds up fitting the t -copula, which involves optimising the degrees of freedom parameter. Since changing the degrees of freedom changes the values of the

t -quantiles, the quantile transform must be performed in every iteration of the fitting procedure. Using the empirical distribution means each stock has the same quantiles, allowing us to perform the quantile transform for one set of quantiles and then sorting it for each stock — a much faster procedure.

3.2 Checking scores for dependence

Once the marginal distributions had been fitted, they were transformed into quantiles, and then to normal distributions and used to create normal scores scatterplots. We could not examine all 780 bivariate scores for the data, but we looked at a selection of scores in and between the four sectors. Examples of scores within and between sectors is in appendix H, with the examples discussed below in figure 3.1. Many scores showed a sharper lower tail than upper tail, with an extreme example being the score of APPL and MFST, or JNJ and LLY. The scores between stocks both in the consumer staples or financial industry tend towards slightly more radial symmetry than see in the rest of the scores scatterplots. For example the score of CL and PEP, in the consumer staples sector, and BAC and JPM, in the financial sector, show equally sharp upper and lower tails. One stock in the financial sector, ANZ.AX, had very little dependence with other stocks, showing up as an ellipsis on many plots (the normal scores scatterplot of ANZ.AX and MET, both stocks from the financial sector, is shown in figure 3.1). ANZ.AX is on a different stock exchange to most of the other stocks, and including it in the dataset was probably a mistake. Even so, a model of this size should be robust enough that one stock behaving differently to the others should not destroy the fit of the overall model. The other stock that is not on the same stock exchange is NESN.SW, however this is for a large, multinational company, and displayed fairly typical dependence with other stocks (the score of MDLZ and NESN.SW two large food companies in the consumer staples sector is show in figure 3.1). Finally, the last line of figure 3.1 shows stocks with fairly typical dependence: AMD and EA, both from the technology sector, show the sharper lower dependence we expect, compared to the upper tail. MET and YUM show reasonably strong lower tail dependence, despite being from different sectors. EA and PGR, also stocks from different sectors, show less dependence, forming a more circular shape on the normal scores scatterplot.

Overall, most of the scores show a sharp, heavy lower tail, which is consistent with stocks tending to crash together. A sharp tail is also indicates we will need a fat tailed joint distribution to better fit the stocks, as we suspected.

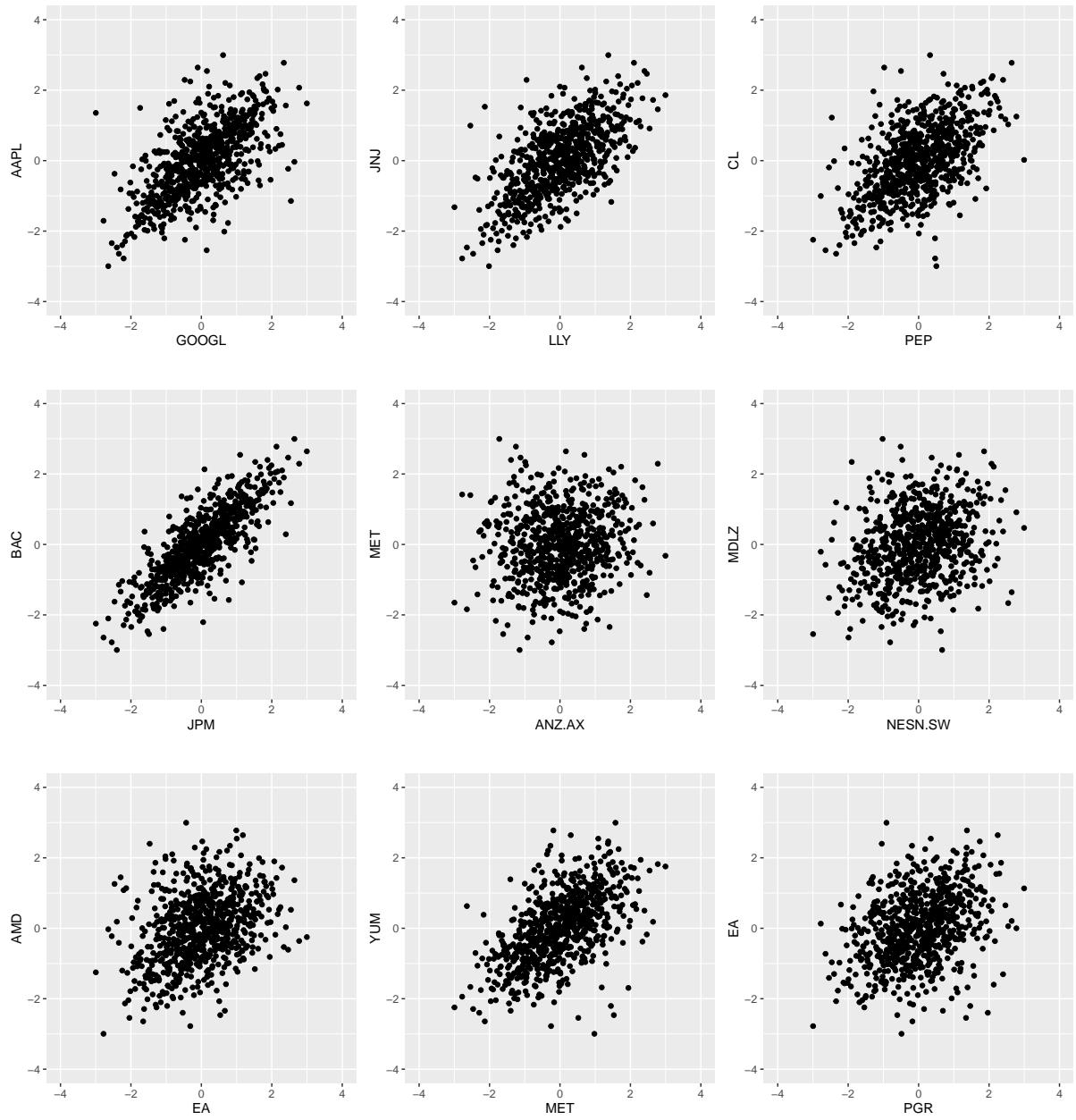


Figure 3.1: Selected normal scores scatterplots for stocks. Refer to section 3.2 for discussion of these plots.

3.3 Estimating normal and t -copulas

In this section, we estimate normal or t -copulas. Our goal is to create a copula that models the dependence in our data well. Normal and t -copulas are symmetric in the upper and lower tail, and therefore will not be able to model the pattern seen in the data of a heavier and sharper lower tail and a lighter and blunter upper tail shown in many of the score plots. Thus the hope with these copulas is that they will model both tails sufficiently for inference, rather than exactly.

We create our copulas using the inversion method. Let \mathbf{F} be the distribution function of the multivariate normal or t -distribution, and F be the distribution function of its univariate counterpart. Then we can construct a copula with the following distribution function:

$$C(\mathbf{u}_1, \dots, \mathbf{u}_d) = \prod_{i=1}^n \mathbf{F}(F^{-1}(u_{i,1}), \dots, F^{-1}(u_{i,d}))$$

where the vector $\mathbf{u}_i = [u_{i,1} \ \dots \ u_{i,d}]^T$ contains the empirical quantiles of the residuals of all d stocks for day i . Through an application of the chain rule, we see the density of the copula is:

$$c(\mathbf{u}_1, \dots, \mathbf{u}_d) = \frac{\partial^d C(\mathbf{u}_1, \dots, \mathbf{u}_d)}{\partial \mathbf{u}_1 \dots \partial \mathbf{u}_d} = \frac{\prod_{i=1}^n \mathbf{f}(F^{-1}(u_{i,1}), \dots, F^{-1}(u_{i,d}))}{\prod_{i=1}^n \prod_{j=1}^d f(F^{-1}(u_{i,j}))} \quad (3.1)$$

Now in the numerator of this equation we have the product of n densities of multivariate distributions, and in the denominator we have a product of nd densities of univariate distributions.

The density of the copula is easily converted to negative log-likelihood to be minimised, like so:

$$\begin{aligned} -\ell(\boldsymbol{\theta}) &= -\log c(\mathbf{u}_1, \dots, \mathbf{u}_d) \\ &= -\sum_{i=1}^n \log (\mathbf{f}(F^{-1}(u_{i,1}), \dots, F^{-1}(u_{i,d}))) + \sum_{i=1}^n \sum_{j=1}^d \log (f(F^{-1}(u_{i,j}))) \end{aligned} \quad (3.2)$$

where $\boldsymbol{\theta}$ is a vector of the dependency parameters we wish to fit.

The first step in fitting a normal copula is transforming our empirical quantiles to normal

scores. Let us denote the transformed quantiles for stock j on day i as $Z_{i,j}$, such that:

$$Z_{i,j} = \Phi^{-1}(u_{i,j})$$

We can assume $(Z_{i,1}, \dots, Z_{i,d})$ are realisations of the random vector (Z_1, \dots, Z_d) on day i . The negative log-likelihood for the normal copulas is:

$$-\ell(\Sigma) = -\sum_{i=1}^n \log(\phi(Z_{i,1}, \dots, Z_{i,d}; \Sigma)) + \sum_{i=1}^n \sum_{j=1}^d \log(\phi(Z_{i,j})) \quad (3.3)$$

So we see that once the quantiles are transformed to the $Z_{i,j}$ -random variables, all that remains is to fit the covariance matrix Σ . We construct Σ by assuming the $Z_{i,j}$ random variables are dependent on latent normal random variables. The relationships we design can be called the *dependency structure* or the *stochastic representation*. Once we fit these parameters, we will have fit a Copula-GARCH model which has t -distributed margins, and a normally distributed dependence structure.

Constructing the dependency structure of the t -copulas involves an extra step. Like before, first we need to transform the empirical quantiles to t -scores. Let us denote the transformed quantiles for stock j on day i as $S_{i,j}$, such that:

$$S_{i,j} = T_\nu^{-1}(u_{i,j}; \nu)$$

The negative log-likelihood of the t -copulas is:

$$-\ell(\Sigma, \nu) = -\sum_{i=1}^n \log(t_\nu(S_{i,1}, \dots, S_{i,d}; \Sigma)) + \sum_{i=1}^n \sum_{j=1}^d \log(t_\nu(S_{i,j})) \quad (3.4)$$

We want to design the covariance or scale matrix Σ , so that it has the same structures as the normal copulas. *Normal variance mixtures* are random vectors with zero means, which are typically non-normally distributed, but share the same Σ matrix as a multivariate normal model (Barndorff-Nielsen et al. 1982). Their dependency structure is created by dividing a normal random vector with zero mean by an independent factor. To create a t_ν -distributed normal variance mixture, we divide by the factor $\sqrt{U_\nu/\nu}$, where U_ν is an independent chi-squared random variable with ν degrees of freedom. Since we are assuming $(Z_{i,1}, \dots, Z_{i,d})$ are realisations of the random vector (Z_1, \dots, Z_d) on day i , then likewise let us assume $(S_{i,1}, \dots, S_{i,d})$ are realisations from (S_1, \dots, S_d) on day i . The stochastic

representation of (S_1, \dots, S_d) is:

$$(S_1, \dots, S_d) = \frac{(Z_1, \dots, Z_d)}{\sqrt{U_\nu/\nu}}$$

where the Z_j random variables have the same stochastic representation as in the normal copulas. Therefore, in the next section, we just give the stochastic representation of the Z_j random variables.

3.3.1 Copulas with 1, 2 and 3 factor structure

In these copulas, we assume the dependency of Z_1, \dots, Z_d are affected by 1, 2 or 3 latent independent standard normal factors, as well as a standard normal error term. For example, in the three factor model, we assume Z_j , is dependent on 3 independent standard normal global factors, as well as a normal error term, like so:

$$Z_j = a_j V_1 + b_j V_2 + c_j V_3 + \psi_j V_{0j}, \quad V_1, V_2, V_3, V_{0j} \stackrel{i.i.d.}{\sim} N(0, 1)$$

Furthermore, we assume $a_j^2 + b_j^2 + c_j^2 + \psi_j^2 = 1$. Then we know that for $j = 1, \dots, d$ and $j \neq k$:

$$\text{VAR}(Z_j) = 1, \quad \text{COV}(Z_j, Z_k) = a_j a_k + b_j b_k + c_j c_k$$

This makes it easy to construct the covariance matrix of the normal model. In the three factor model, we construct a diagonal matrix D and a loadings matrix \mathbf{L} , like so:

$$\mathbf{L} = \begin{bmatrix} 0 & 0 & a_3 & \cdots & a_d \\ 0 & b_2 & b_3 & \cdots & b_d \\ c_1 & c_2 & c_3 & \cdots & c_d \end{bmatrix}, \quad D = \begin{bmatrix} \psi_1^2 & 0 & \cdots & 0 \\ 0 & \psi_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \psi_d^2 \end{bmatrix}$$

We can see the covariance matrix has the structure:

$$\Sigma = \mathbf{L}^T \mathbf{L} + D$$

The values in the upper triangle of \mathbf{L} are set to zero in \mathbf{L} to make the resulting model identifiable. Consider the case of a loadings matrix \mathbf{L}' where no parameters are fixed to zero (and the corresponding diagonal matrix D'). Let \mathbf{Q} be a 3×3 rotation matrices such that $\mathbf{Q}^T \mathbf{Q} = \mathbf{I}_3$. Then:

$$\Sigma = \mathbf{L}'^T \mathbf{L} + D' = \mathbf{L}'^T \mathbf{Q}^T \mathbf{Q} \mathbf{L}' + D'$$

So either \mathbf{L}' or $\mathbf{Q}\mathbf{L}'$ can be used to create Σ . And since there are an infinite number of matrices that satisfy the conditions for \mathbf{Q} , there is an infinite number of alternative loadings matrices, and we will not be able to converge upon one of them.

\mathbf{L} has been created so its columns cannot be rotated into a different loadings matrix without placing non-zero parameters in its upper left triangle. By forcing these parameters to be zero, we make \mathbf{L} unique, which will allow us to converge on a single set of parameters.

The covariance matrix has the same structure in the 1, 2 and 3 factor models. The key difference between these types of models is the number of rows of variables in the loadings matrix: the 1 factor model contains one row, and the 2 factor model contains two.

Once the covariance matrix is calculated, we can substitute it into the equations of the negative log-likelihood for the normal and t -copulas that were given in equations 3.3 and 3.4. These likelihoods involve calculating the log-density of multivariate normal and t -distributions with the covariance or scale matrix Σ . This can easily be done in R using the `mvtnorm` package. Computing this involves finding the determinant and inverse of the $d \times d$ covariance matrix Σ . Luckily Σ is in a form that allows us to decompose its inverse and determinant into forms where we only have to perform these operations on simpler matrices. Denote the number of factors used in the model as k^* . Then, the inverse of determinant of the covariance matrix are (see Krupskii (2014), section 4.4.2 for proofs):

$$\Sigma^{-1} = D^{-1} - D^{-1}\mathbf{L}^T(\mathbf{L}D^{-1}\mathbf{L}^T + \mathbf{I}_{k^*})^{-1}\mathbf{L}D^{-1} \quad (3.5)$$

$$\det(\Sigma) = \det(D) \det(\mathbf{L}D^{-1}\mathbf{L}^T + \mathbf{I}_{k^*}) \quad (3.6)$$

Now we only need to find the inverse and determinant of a diagonal matrix and a $k^* \times k^*$ matrix.

Once we had a formula for the negative log likelihood, the negative log likelihood of the copula could be minimised to estimate the values of a_j , b_j and c_j for $j = 1, \dots, d$ with the Newton-Raphson method in R with the function `nlm`.

Constraints were applied to ensure

$$a_j^2 + b_j^2 + c_j^2 \leq 1, \quad j = 1, \dots, d \quad (3.7)$$

This was done in two stages. The first constraint applied kept the value of variables of interest between -1 and 1 . This was applied to a_1, \dots, a_d , but also the new variables b_1^*, \dots, b_d^* and c_1^*, \dots, c_d^* . The best constraint found was supplying a vector containing real numbers, denoted here as $x_{a_j}, x_{b_j^*}, x_{c_j^*}$, and then transforming them to a number between -1 and 1 with a sigmoidal constraint, like so:

$$a_j = \frac{2}{1 - e^{-x_{a_j}}} - 1, \quad x_{a_j} \in \mathbb{R}$$

Box constraints and constraint involving a non-monotonic transformation were also attempted, but both exhibited problems with convergence.

The second constraint applied was:

$$\begin{aligned} b_j &= b_j^* \times \sqrt{1 - a_j^2}, & a_j^2, b_j^{*2} \leq 1, \quad j = 1, \dots, d \\ c_j &= c_j^* \times \sqrt{1 - a_j^2 - b_j^2}, & a_j^2, b_j^2, c_j^{*2} \leq 1, \quad j = 1, \dots, d \end{aligned}$$

Which is equivalent to constraint 3.7.

To perform Newton-Raphson optimisation `nlm` estimates the gradient of parameters at each iteration with numerical differentiation. This can increase run-time of the algorithm in at least two ways. Firstly, small errors in the estimation can increase the number of iterations, particularly in higher dimensions. Secondly, numerical differentiation requires calculating the value of the function at slightly different points, meaning the function has to be run a number of times to estimate the gradient. To combat this we used analytic gradient for both the normal and t -copulas, which is derived in Appendix B. As expected, supplying the analytic gradient made optimisation much faster.

In the t -copula, the gradient of ν also has to be calculated. This is somewhat difficult, as it involves calculating the gradient of ν when the empirical quantiles were transformed to the t -distribution. R uses Hill's Algorithm 396 (1970) to calculate the quantiles of the t -distribution. An approximation to the gradient of this algorithm was calculated, however finite differencing exhibited slightly better convergence, and hence the latter method was used.

3.3.2 Nested models

In a nested model, the variables of interest are dependent on latent variables, which themselves can be dependent on other latent variables, and so on. The structure of the nested

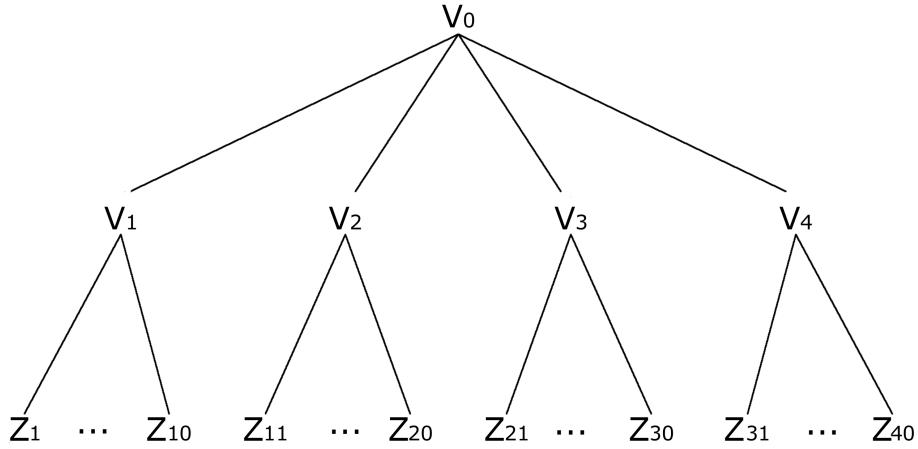


Figure 3.2: The dependence structure of the nested model used to fit 40 stocks from 4 sectors. Each stock Z_1, \dots, Z_{40} is dependent on the appropriate sector specific random variable V_1, \dots, V_4 , which in turn are all dependent on a global random variable V_0

model in the normal case used for this project is shown in figure 3.2. In this case, our four groups of ten stocks from different sectors are dependent on a latent sector specific variable. The four sector specific random variables are then dependent on a global random variable. When structured this way, the nested model is highly interpretable. The global variable V_0 can be seen as a shock to the whole market, with the sector specific variables V_1, \dots, V_4 controlling how vulnerable each sector is to the global shock.

More mathematically, we assume that the global variable V_0 has a standard normal distribution. We assume that the sector specific variables V_1, \dots, V_4 are dependent on both V_0 and independent standard normal random variables V_{01}, \dots, V_{04} that control error, like so:

$$V_k = a_{V_k} V_0 + \psi_{V_k} V_{0k}, \quad k = 1, 2, 3, 4, \quad V_0, V_{01}, \dots, V_{04} \stackrel{iid}{\sim} N(0, 1)$$

Furthermore, we assume that:

$$a_{V_k}^2 + \psi_{V_k}^2 = 1, \quad k = 1, 2, 3, 4$$

Which means that:

$$V_k \sim N(0, 1), \quad \text{VAR}(V_k) = 1, \quad \text{COV}(V_k, V_j) = a_{V_k} a_{V_j}, \quad j, k = 1, 2, 3, 4, \quad j \neq k$$

Then we need to look at what happens within a sector. Let us take the technology sector controlled by the variable V_1 as an example. The variables Z_1 to Z_{10} are transformed residuals from stocks in this sector, and are dependent on V_1 and independent standard normal random variables $V_{11}, \dots, V_{1,10}$ that control the error, like so:

$$Z_k = a_{Z_k} V_1 + \psi_{Z_k} V_{1k}, \quad k = 1, \dots, 10, \quad V_1, V_{11}, \dots, V_{1,10} \stackrel{iid}{\sim} N(0, 1)$$

Again, we assume the the error term V_{1k} is normally distributed with a mean of 0 and independent of all other variables in the model, and that

$$a_{Z_k}^2 + \psi_{Z_k}^2 = 1, \quad k = 1, \dots, 10$$

Hence we have that Z_1 to Z_{10} are normally distributed, where:

$$\text{VAR}(Z_k) = 1, \quad \text{COV}(Z_k, Z_j) = a_{Z_k} a_{Z_j}, \quad j, k = 1, \dots, 10, \quad j \neq k$$

All four sectors have the same structure as this. We have a different pattern of covariance. For example, between stock Z_1 and stock Z_{11} , which are from two different sectors, we have that:

$$\text{COV}(Z_1, Z_{11}) = a_{Z_1} a_{Z_{11}} a_{V_1} a_{V_2}$$

Unlike the factor models, this covariance matrix cannot be decomposed into the form $\Sigma = \mathbf{L}^T \mathbf{L} + D$. Hence the only method used to find this copula utilised `mvtnorm`, and involved inverting the full 40×40 covariance matrix. The gradient for this model is available in Krupskii (2014), but was not implemented in this project.

Like in the previous models, a sigmoidal constraint was used to constrain $a_{Z_1}, \dots, a_{Z_{40}}$ and a_{V_1}, \dots, a_{V_4} to between -1 and 1 .

3.3.3 Bifactor models

In a bifactor model, variables are placed into groups. The variables then depend on a global variable and a group variable. The structure of the bifactor model used here is shown in figure 3.3. The global variable is V_0 . The transformed log stock returns Z_1 to Z_{40} are grouped by sector, with the first ten variables in the first sector (with group variable V_1), the next ten variables in the second sector (with group variable V_2), and so on.

For this model we assume V_0, V_1, V_2, V_3 and V_4 are independent standard normal variables.

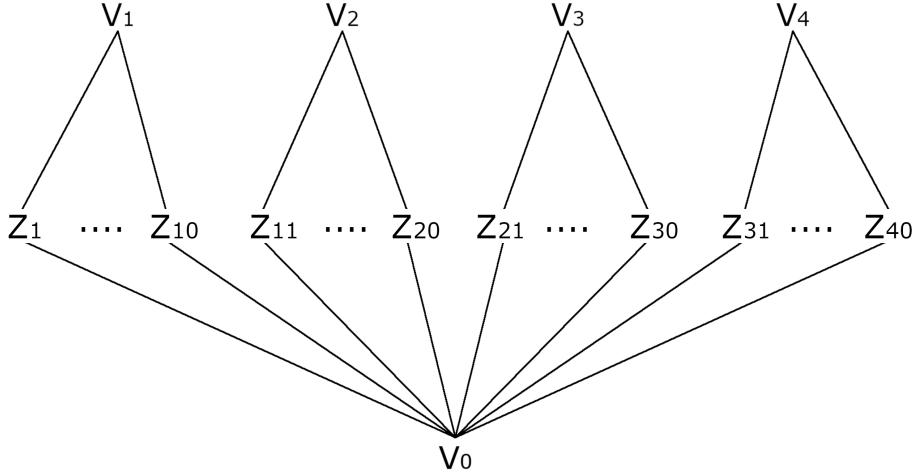


Figure 3.3: Dependency structure of the bifactor model used to fit 40 stocks from 4 sectors. Stocks Z_1, \dots, Z_{40} depend on a global random variable V_0 as well as the appropriate sector specific random variable V_1, \dots, V_4 .

Additionally, for $k = 1, \dots, 40$, each variable Z_k is dependent on another independent standard normal variable V_{0k} , which controls the error in the model. The stochastic representation of Z_k is given by:

$$Z_k = a_k V_0 + b_k V_G + \psi_k V_{0k}, \quad k = 1, \dots, 40$$

where $G = 1, 2, 3$ or 4 depending on which sector stock Z_k belongs to. Again, we assume that

$$a_k^2 + b_k^2 + \psi_k^2 = 1, \quad k = 1, \dots, 40$$

Which means the variables for the stocks are normally distributed, and we have that for $j, k = 1, \dots, 40$ and $j \neq k$:

$$\text{VAR}(Z_k) = 1, \quad \text{COV}(Z_k, Z_j) = \begin{cases} b_k b_j + a_k a_j, & Z_j \text{ and } Z_k \text{ are in the same sector} \\ a_k a_j, & Z_j \text{ and } Z_k \text{ are in different sectors} \end{cases}$$

The covariance matrix for this model can be put into the form $\mathbf{L}^T \mathbf{L} + D$, where:

$$\mathbf{L}^T = \begin{bmatrix} b_1 & 0 & 0 & 0 & a_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ b_{10} & 0 & 0 & 0 & a_{10} \\ 0 & b_{11} & 0 & 0 & a_{11} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & b_{20} & 0 & 0 & a_{20} \\ 0 & 0 & b_{21} & 0 & a_{21} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & b_{30} & 0 & a_{30} \\ 0 & 0 & 0 & b_{31} & a_{31} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & b_{40} & a_{40} \end{bmatrix}, \quad D = \begin{bmatrix} \psi_1^2 & 0 & \dots & 0 \\ 0 & \psi_2^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \psi_d^2 \end{bmatrix}$$

Hence, like in the factor models, we can speed up the fitting procedure by supplying the analytic formula for Σ^{-1} . For this model, we need the constraint:

$$a_i^2 + b_i^2 \leq 1$$

Which we can achieve by applying a sigmoidal constraint to restrict all the parameters to between -1 and 1 , and then setting:

$$b_i = b_i^* \sqrt{1 - a_i^2}, \quad i = 1, \dots, 40$$

Like the nested model, the gradient for this model is available in Krupskii (2014), but was not implemented in this project.

3.3.4 Diagnostics to select the best copula model

Likelihood Ratio Tests

As we used likelihood methods to compare the fit of the models to the data, we could do likelihood ratio tests (LRTs) for models that could be considered to be nested in each other. A smaller model is considered nested in a larger model when the smaller model can be created from the larger model by fixing parameters. The LRT is a hypothesis test, where the null hypothesis is that the parameters are truly fixed as in the smaller model.

Its test statistic is:

$$\lambda(\mathbf{x}) = \frac{\mathcal{L}(\hat{\boldsymbol{\theta}}_0 | \mathbf{x})}{\mathcal{L}(\hat{\boldsymbol{\theta}} | \mathbf{x})}$$

where \mathbf{x} is the data, $\mathcal{L}(\boldsymbol{\theta})$ is the likelihood for parameters $\boldsymbol{\theta}$, and $\hat{\boldsymbol{\theta}}_0$ and $\hat{\boldsymbol{\theta}}$ are the maximum likelihood estimates of the smaller and larger models, respectively. Under the null hypothesis, as the sample size becomes infinitely large the LRT statistic has the limiting distribution:

$$-2 \log \lambda(\mathbf{x}) \sim \chi_r^2$$

where r is the difference in number of parameters between the larger and smaller models. Therefore for a test at significance level α , the larger model is chosen when $-2 \log \lambda(\mathbf{x})$ is greater than the $1 - \alpha$ critical value for the χ_r^2 distribution.

The limiting distribution of the t -distribution as the degrees of freedom ν approach infinity is the normal distribution. As such the normal distribution can be considered a special case of the t -distribution with fixed degrees of freedom, we can compare the models with LRTs. The other models we can compare with LRTs were the 1, 2 and 3 factor models. When the first row of the loadings matrix in the 3 factor model is fixed to zero, it becomes the 2 factor model. This then can be converted into the 1 factor model in the same way. Therefore the 1 factor model is nested in the 2 factor model, which is nested in the 3 factor model. Lastly we can say the 1 factor model is nested in the bifactor model, as setting all the sector specific parameters in the bifactor model to zero will give us the 1 factor model.

AIC and BIC

The Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) are criteria used in model selection to discourage overfitting. As more parameters are added to a model its likelihood will improve, however if too many parameters are added the model will begin to fit noise. Both criteria penalise the increase in log likelihood by the number of parameters added, with BIC penalising parameters more than AIC. Their formulas are:

$$\text{AIC} = 2k - 2\ell, \quad \text{BIC} = k \log n - 2\ell$$

where k is the number of fitted parameters in the model, n is the sample size and ℓ is the maximum log-likelihood of the fitted model. When fitting using one of these criteria, the model with the lowest AIC or BIC is chosen. Note this is different to the formula used by `rugarch` to calculate the AIC of the ARMA-GARCH models in the

margins in section 2.1.5. That formula was re-scaled by a constant, however re-scaling does not effect which model will have the higher or lower AIC, which is what is important.

Unlike likelihood ratio tests, AIC and BIC can be used to compare the likelihoods of models with different structures fit to the same set of data. The BIC penalises the number of parameters more heavily than the AIC.

Differences in within and between sector statistics

We used four statistics to evaluate the fit of the theoretical models to the data: the Pearson correlation, the Spearman rank correlation, and the upper and lower semicorrelation. For each statistics, we calculated the difference and absolute difference between the statistic for the theoretical model, and the sample statistic. The former was used as a measure of bias and the latter as a measure of variability of the fitted model. The difference was calculated by subtracting the theoretical statistic from the sample statistic. Hence if it is positive, the theoretical statistic is smaller than the sample statistic, and vice versa.

Pearson correlation

As a measure of fit in the centre of the distributions, we compared the fitted Sigma matrices estimated with both normal and t -copulas to the Pearson correlation in the transformed data. The Sigma matrices of the copulas contain the Pearson correlation in the model (assuming $\nu > 2$ so that the covariance exists), which can be quickly verified for the normal model like so:

$$\Sigma_{ij} = \frac{\text{COV}(Z_i, Z_j)}{1} = \frac{\text{COV}(Z_i, Z_j)}{\sqrt{\text{VAR}(Z_i)}\sqrt{\text{VAR}(Z_j)}} = \text{CORR}(Z_i, Z_j)$$

And for the t -distributed model:

$$\Sigma_{ij} = \frac{\text{COV}(S_i, S_j)}{\nu/(\nu - 2)} = \frac{\text{COV}(S_i, S_j)}{\sqrt{\text{VAR}(S_i)}\sqrt{\text{VAR}(S_j)}} = \text{CORR}(S_i, S_j)$$

If the models fit the centre of the data well, the correlations of the fitted model should match the correlations in the transformed data. Both the difference and absolute difference between the fitted Sigma matrix and the correlation of the transformed data were calculated, with the former being a measure of accuracy of the fitted models, and the latter a measure of variance in the fitted models. The distance was measured by subtracting the fitted Sigma matrix from the correlation in the data.

In the case of the t -copulas, this is only a good way to evaluate the fit of the model when the theoretical covariance exists. If the fitted value of ν is smaller than 2, we should use other methods.

Spearman rank correlation

Spearman rank correlation, or Spearman's Rho, denoted here as ρ_S , is a measure of correlation between the ranks of two random variables. It is an alternative way to measure correlation in the dataset, that is distribution free, and more robust to outliers than Pearson correlation. The population version of the statistic for a dataset (x_i, \tilde{x}_i) for $i = 1, \dots, n$ is given by

$$\hat{\rho}_S = \frac{\sum_{i=1}^n r_i \tilde{r}_i - n[(n+1)/2]^2}{n(n^2-1)/12}$$

where r_i and \tilde{r}_i are the ranks of (x_i, \tilde{x}_i) , or $r_i = k$ if x_i is the k -th smallest observation among x_1, \dots, x_n (and likewise for \tilde{r}_i) (Joe 2014). The population version of Spearman rank correlation for random variables X and \tilde{X} can be calculated a few ways. Perhaps the easiest is as the Pearson correlation between their distribution functions, like so:

$$\rho_S = \text{CORR}\left(F_X(X), F_{\tilde{X}}(\tilde{X})\right)$$

The exact Spearman rank correlation for both the multivariate normal and t -distribution is known, however calculating the latter statistic is not particularly simple (Heinen & Valdesogo 2020). Therefore the empirical Spearman rank correlation for the model was found instead. This was done by simulating a new dataset from the fitted model, and calculating the population version of the statistic for the dataset.

Semicorrelation

Semicorrelation was used as a diagnostic to compare fit in the tails. Semicorrelations are statistic designed for data where observations in the lower tail have a different correlation than observations in the upper tail (Erb et al. 1994). Recall the normal bivariate score plots indicated heavier lower tail dependence is not unusual in our data. The upper and lower semicorrelations ρ^+ and ρ^- for bivariate uniform random variables are:

$$\begin{aligned}\rho^+ &= \text{CORR}(U_1, U_2 \mid U_1 > 0.5, U_2 > 0.5) \\ \rho^- &= \text{CORR}(U_1, U_2 \mid U_1 < 0.5, U_2 < 0.5)\end{aligned}$$

The data was transformed to a uniform score, and the semicorrelations between stocks were calculated using the formulas above. As it would be difficult to derive an analytical

Model	Number of parameters	Maximum log-likelihood	AIC	BIC
1 factor structure normal	40	6638	-13197	-13013
1 factor structure t	41	7049	-14016	-13828
2 factor structure normal	79	7340	-14521	-14159
Nested normal	44	7518	-14948	-14746
3 factor structure normal	117	7799	-15365	-14828
2 factor structure t	80	7700	-15240	-14873
Bifactor normal	80	7818	-15476	-15108
Nested t	45	7862	-15634	-15427
3 factor structure t	118	8135	-16034	-15492
Bifactor t	81	8147	-16131	-15759

Table 3.1: Comparison of the performance of the different models on the 40 stocks. Models are ordered from highest to lowest BIC. The bifactor t -copula is the best fitting model by both AIC and BIC.

formula for the semicorrelation in the fitted model, it was found empirically instead. First data was simulated from the fitted model, before being transformed to a uniform score, and used to calculate the empirical semicorrelation.

3.4 Modelling and results

All models were fit to the empirical quantiles of the residuals from the ARMA-GARCH models, and compared. Table 3.1 shows the AIC and BIC of the different models. From top to bottom, the table is ordered by decreasing BIC. When we just compare the copulas with the same distribution, but different Sigma matrices, we see the three models with the lowest AIC and BIC are the nested, 3 factor and bifactor models, in descending order. When we compare the copulas with the same Sigma matrices but different distributions, we see the t -copulas always have a lower AIC and BIC than their normal counterparts. Despite the bifactor normal copula having 35 more parameters than the nested t -copula, without the former model having a parameter allowing its tails to become leptokurtic, it still has a slightly smaller maximum log-likelihood than the nested copula.

The only two models which were not ordered the same way by both descending AIC and BIC (their ranks were swapped), were the normal copula with 3 factor structure and the t -copula with 2 factor structure. We ranked the table by descending BIC as it encourages smaller models than AIC. This is one way to encourage tractability, but it is a crude one, as the speed of fitting these models does not just depend on the number of parameters. In this case, both models took a similar amount of time to fit (see table 3.3).

Smaller model	Larger model	$-2 \log \lambda(\mathbf{x})$	p-value
1 factor structure normal	2 factor structure normal	1404	< 0.001
2 factor structure normal	3 factor structure normal	918	< 0.001
3 factor structure normal	3 factor structure t	672	< 0.001
1 factor structure normal	Bifactor normal	2360	< 0.001
Bifactor normal	Bifactor t	658	< 0.001
Nested normal	Nested t	688	< 0.001

Table 3.2: Likelihood ratio tests for the elliptical copulas, shown the likelihood ratio statistic, and p -values. All the LRTs favoured the larger models.

All the likelihood ratio tests favoured the larger model. A selection of the LRTs are shown in table 3.2, but can be inferred from the table showing log-likelihoods. Recall the LRT is twice the difference between the log-likelihood of the larger and smaller model. The differences in log-likelihoods between these models are so large that the resulting LRT statistic for all models was well above the critical value for the test at even a 0.999 level of confidence. For example, the 0.999 quantile of a χ^2_1 distribution is only about 10.8. The difference between the likelihoods of the normal copulas and their t -counterparts were all in the hundreds, meaning the t -models were easily superior to the their smaller counterparts. The other models compared usually had a difference close to 40 parameters. The 0.999 quantile of the χ^2_{40} distribution, for example, is about 73.4, and 0.999 quantile of chi-squared distributions with similar degrees of freedom is close to that. By contrast, the difference between likelihoods of models we were comparing was typically in the hundreds, if not thousands, so the larger models were always accepted.

The rest of the results, for the difference in statistics, are shown in tables in appendix C. These are shown for averaged over all the stocks, as well as between and within sectors. All of the t -copulas had fitted degrees of freedom well above 2, so we could compare theoretical and sample Pearson correlation. When we examine the overall differences, all models tend to have fairly close differences in both Spearman rank correlation and Pearson correlation, with differences for both statistics being around ± 0.02 , and absolute differences being in the range of $0.02 - 0.05$. Thus, all models had fairly good fits at the centre of the data.

When averaged over all stocks, all normal copulas showed a difference and absolute difference between lower semicorrelation of about 0.11 and 0.11 respectively, regardless of covariance matrix structure. All the t -copula showed a difference and absolute difference of 0.08 and 0.09. Since this number is positive, both models are underestimating the lower

semicorrelation, with the t -copula underestimating it less. This indicates both models will underestimate the risk in the crucial, heavy lower tail of the joint log stock returns. Normal models and t -models had fairly similar fits for difference in upper semicorrelation (0.01 and -0.02) respectively.

When we look at the pattern of results sector by sector, we can see the 3 factor, nested and bifactor models have the best fit within and between sectors, for all the four statistics calculated. There is a clear pattern of these models producing smaller differences across the between and within sector statistics, with less variability, as measured by the absolute difference. For example, the worst fit for lower semicorrelation is for the normal model with 1 factor structure, where both the difference and absolute difference is 0.20 for sector 1. Both the difference and absolute difference are 0.15 in the normal model with 3 factor structure, the nested normal model and the bifactor normal model. The rest of the differences in lower semicorrelation are comparatively smaller in the latter models, compared to the 1 and 2 factor models. This also occurs in the difference in upper correlation, Spearman rank correlation, and covariances. The extra flexibility in these models allow them to fit between and in sector statistics better.

Table 3.3 shows the time it took to fit various models. Due to variability, we would need to fit many datasets to properly evaluate the relative speed of each model, but we can still use our results to infer which models should be faster. Factors we expect to make fitting take longer is an increase in parameters, and increase in iterations (which is associated with the number of parameters), whether the gradient is supplied, and if it is a t -copula. The results are consistent with this. The t -copulas take much longer to fit than their normal counterparts, with the perfect example of this being the nested t -copula taking approximately 5 times as long to fit as the nested normal, despite running for fewer iterations. Since we need to transform the t -copula to t -scores every iteration, this will always be the case, unless we fix the degrees of freedom parameter. However, it may be possible to slightly improve on this time by finding a quicker way to calculate the gradient of the transformation of quantiles to t -scores than finite differencing.

The bifactor t -copula took the longest time to fit, at 24.1 seconds. However, the gradient of this model was not supplied. The bifactor model has a similar number of parameters to the 2 factor model, and assuming its gradient is not much harder to calculate, it seems likely it would take a similar amount of time to fit with a supplied gradient.

If we want the best fit, the bifactor t -copula would be the natural choice, as it has the

Model	Time (s)	Iterations	Number of parameters	Gradient supplied ?
1 factor normal	1.5	8	40	yes
1 factor t	2.9	17	41	yes
2 factor normal	3.1	28	79	yes
2 factor t	6.1	32	80	yes
3 factor normal	5.4	117	47	yes
3 factor t	9.2	51	118	yes
Nested normal	1.8	23	44	no
Nested t	9.9	19	45	no
Bifactor normal	19.8	23	80	no
Bifactor t	24.1	28	81	no

Table 3.3: Time take to fit elliptical models, in seconds

lowest AIC and BIC, provides a better fit to the lower semicorrelation than the normal models, and is one of the better models for fitting both Pearson correlation and Spearman rank correlation. As the models are currently implemented, the 3 factor t -copula would be a good choice if we want a rapid model that still offers a reasonable fit to the data, not just overall, but also between and within sectors. Moreover, if we decide we just want a model with a good central fit, and are not that worried about the fit in the tails, we could further speed up fitting by choosing a normal model.

3.5 Radially asymmetric copulas

All the copula modelling so far was done using copulas with radial symmetry: that is symmetry between the upper and lower tails of the copula. However, since we expect stocks to be more dependent when their price is falling, we expect the lower tail to be heavier than the upper tail. Indeed, we see heavier lower tails in many of the normal scores scatterplots, such as between AAPL and GOOGL in figure H.1, which has a particularly sharp lower tail. In this section we explore two copulas that arise from squaring elliptical copulas, which fit stronger dependence in one tail. We shall say these two copulas are both *squared copulas*.

These squared copulas arise from non-monotonic transformation of elliptical distributions. Any non-monotonic transformation will create a new copula. Since these copulas are based on elliptical copulas, we can easily simulate from them, by simulating from the elliptical distribution and applying the non-monotonic transform to the results. We can infer their properties and calculate statistics based on their relationship with the underlying elliptical distribution. If this is too hard, since they can be simulated from,

we can just make these inferences from their empirical distributions.

We know the chi-squared distribution as a sum of squared normal random variables. The chi-squared copula has the dependency structure of the multivariate chi-squared distribution, and has been used in geostatistical applications (Bárdossy 2006, Quessy et al. 2016). In this thesis, we also use another copula which has the dependency structure of a multivariate t_ν^2 -distribution, or $F_{1,\nu}$ -distribution. We call it the F -copula. The distributions in this section are not to be confused with the Wishart distribution of Hotelling's T^2 distribution, which are constructed through different methods to any distributions in this section. The way the copulas in this section are constructed are discussed shortly.

In figure 3.4 we have three normal scores scatterplots: from left to right, they are a

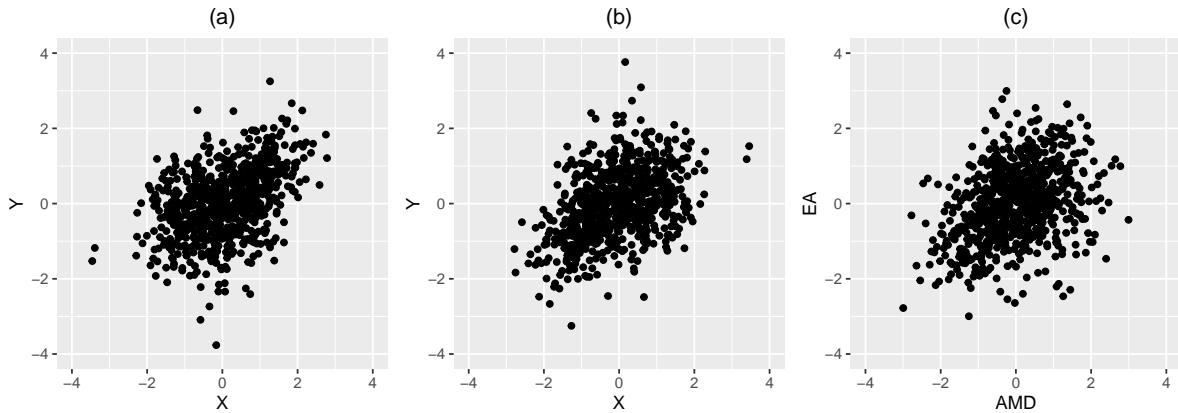


Figure 3.4: Normal scores of simulated bivariate chi-squared distributions, compared to normal scores scatter plots of log stock returns. (a) is a bivariate chi-squared distribution with margins X and Y , where $(X, Y) \sim \chi_1^2$ with correlation parameter $\rho = 0.75$. (b) is a reflected bivariate chi-squared distribution with margins X and Y , where $(-X, -Y) \sim \chi_1^2$ with correlation parameter $\rho = 0.75$. (c) is a normal score scatter plot of the residuals from stocks EA and AMD. The shape of the real data on the normal score scatter plot in (c) quite clearly matches the normal score scatter plot of the reflected bivariate χ^2 distribution in (b)

bivariate chi-squared distribution, a reflected bivariate chi-squared distribution, and the residuals from two stocks. We can see that unlike a multivariate normal or t -distribution, the reflected bivariate chi-squared copula models the heavy lower tail and light upper tail seen in the actual data. To account for this, we fit the copulas in this section to the *reflected* residuals from the marginal ARMA-GARCH models.

3.5.1 Squared copulas with factor structure

The stochastic representation of both the normal and t -copulas was based off a normal random vector (Z_1, \dots, Z_d) , where variable Z_j was created from the residuals of the ARMA-GARCH models for the j -th log stock return. In this section we take the stochastic representations of the 1, 2 and 3 factor models, and simply square them. We are left with dependency structures of chi-squared and F -copulas, which still have covariance based off a factor structure.

Since the both the chi-squared and F -copulas have a heavy upper tail, and the residuals from the ARMA-GARCH models have a heavy lower tail, we fit the copula to the flipped quantiles of the residuals of the stocks. We flip the quantiles by subtracting them from one: if q is the quantile of the residuals then the corresponding flipped quantile, q_f , will be:

$$q_f = 1 - q$$

We then transform flipped quantiles to quantiles of the normal distribution, which gives us the random variables Z_1, \dots, Z_d . We assume Z_1, \dots, Z_d have the same dependence on latent normal random variables as in the 1, 2 and 3 factor structure we saw in the normal factor model. Therefore they will have the same structured covariance matrix.

3.5.2 Chi-squared copula

If we let (Z_1, \dots, Z_d) be a d -dimensional multivariate normal distribution, where Z_1, \dots, Z_d are standard normal, with correlation matrix Σ , then the central chi-squared copula is the dependence structure of $(W_1, \dots, W_d) = (Z_1^2, \dots, Z_d^2)$. When applied to the marginal distribution functions, the central chi-squared copula will give a multivariate chi-squared distribution. The central chi-squared copulas are the only copulas considered in this thesis, and are from here on referred to just as chi-squared copulas.

The density of the chi-squared copula has previously been derived by Quessy et al. (2016). As constructing the copulas is rather involved, we leave interested readers to refer to the original paper for details. We follow the same method to derive the density of the bivariate F -copula in Appendix E. In brief — the authors find a distribution function for the copula using the normal distribution. This function is then found in terms of the normal copula. Importantly, the copula is constructed using the inversion method. While constructing a chi-squared random variable involves a non-monotonic transform, this is only performed when defining the dependency structure of the copula. The only

transformation applied to the actual quantiles is transforming them into the quantiles of a chi-squared distribution. This is a strictly increasing transformation, and hence by the copula invariance principle, does not transform the copula.

If we let $\Xi = \{-1, 1\}^d$, then the full chi-squared copula density, c_{Σ}^{χ} is given by:

$$c_{\Sigma}^{\chi}(u_1, \dots, u_d) = \frac{1}{2^d} \sum_{\epsilon \in \Xi} \left(\prod_{j=1}^d \epsilon_j \right) c_{\Sigma}^N \left(\frac{1 + \epsilon_1 u_1}{2}, \dots, \frac{1 + \epsilon_d u_d}{2} \right)$$

where c_{Σ}^N is the normal copula density.

The size of Ξ grows exponentially as the numbers of variables increases, making the complexity of the likelihood $O(2^d)$. As this is not computationally tractable, we use a pairwise likelihood approach to fit the copulas.

The density of the bivariate central chi-squared copula is also derived in Quessy et al. (2016). Using the notation from that paper, it is:

$$c_{\rho}^{\chi}(u_j, u_k) = \frac{1}{2} \left\{ c_{\rho}^N \left(\frac{1+u_j}{2}, \frac{1+u_k}{2} \right) + c_{\rho}^N \left(\frac{1+u_j}{2}, \frac{1-u_k}{2} \right) \right\}$$

where c_{ρ}^{χ} is the bivariate chi-squared copula with Pearson correlation ρ^2 and c_{ρ}^N is the bivariate normal copula with covariance ρ . Recall that we showed the density of copulas constructed by the inversion method in equation 3.1 in section 3.3. We can apply this to obtain the density of the bivariate normal copula:

$$c_{\rho}^N(a, b) = \frac{\phi_{\rho}(\Phi^{-1}(a), \Phi^{-1}(b))}{\phi(\Phi^{-1}(a)) \phi(\Phi^{-1}(b))}$$

Note that due to symmetry, $\phi(\Phi^{-1}(\frac{1+u_k}{2})) = \phi(\Phi^{-1}(\frac{1-u_k}{2}))$. Hence the log-likelihood of the bivariate chi-squared copula is:

$$\begin{aligned} & \log c_{\rho}^{\chi}(u_j, u_k) \\ &= \log \left\{ \phi_{\rho} \left(\Phi^{-1} \left(\frac{1+u_j}{2} \right), \Phi^{-1} \left(\frac{1+u_k}{2} \right) \right) + \phi_{\rho} \left(\Phi^{-1} \left(\frac{1+u_j}{2} \right), \Phi^{-1} \left(\frac{1-u_k}{2} \right) \right) \right\} \\ & \quad - \log \phi \left(\Phi^{-1} \left(\frac{1+u_j}{2} \right) \right) - \log \phi \left(\Phi^{-1} \left(\frac{1+u_k}{2} \right) \right) - \log 2 \end{aligned}$$

When taking pairwise composite likelihoods, it is traditional to take the sum of the log-likelihoods. Furthermore, we want to take this sum over each of the n days. Thus, the

log-likelihood we end up maximising is:

$$\sum_{i=1}^n \sum_{j=2}^d \sum_{k=1}^{j-1} \log c_\rho^\chi(u_{i,j}, u_{i,k})$$

where $u_{i,j}$ is the empirical quantile of the flipped, fitted residual of stock j on day i . Hence, for one day, the computational complexity is $O(d^2)$, which is tractable in higher dimensions.

Composite likelihoods yield asymptotically normal and consistent estimators (see Varin (2008) for a survey of the literature). In the case of the chi-squared copula Quessy et al. (2016) found the pairwise likelihood to be slightly inferior to, but competitive with using the full likelihood in a simulation.

We fit chi-squared copulas with 1, 2 and 3 factors. We let the dependency structure of the copulas be:

$$(W_1, \dots, W_d) = (Z_1^2, \dots, Z_d^2)$$

Where the Z_j random variables had the same stochastic representation as for the 1, 2, and 3 factor models for the elliptical copulas, for $j = 1, \dots, d$. Hence,

$$(Z_1, \dots, Z_d) \sim N(0, \Sigma)$$

where Σ is defined as the for 1, 2 and 3 factor elliptical models. Hence the random vector (W_1, \dots, W_d) has a multivariate chi-squared distribution with one degree of freedom and the correlation structure:

$$\text{CORR}(W_j, W_k) = \text{CORR}(Z_j, Z_k)^2$$

This is shown in appendix F.

3.5.3 F -copula

If we let (Z_1, \dots, Z_d) be a multivariate standard normal distribution with non-negative correlations, and U_ν be an independent chi-squared distribution with ν degrees of freedom, then:

$$\frac{(Z_1, \dots, Z_d)}{\sqrt{U_\nu/\nu}}$$

has a multivariate t -distribution with ν degrees of freedom. Moreover,

$$(W_1, \dots, W_d) = \frac{(Z_1^2, \dots, Z_d^2)}{U_\nu / \nu}$$

has a multivariate $F_{1,\nu}$ distribution. The F -copula is then the dependency structure of (W_1, \dots, W_d) .

In this project we have frequently used the t -distribution in order to model sharper tails (as seen on the normal scores scatterplots) than the normal distribution. This property extends to the multivariate $F_{1,\nu}$ -distribution — it has a sharper upper tail than the chi-squared distribution with one degree of freedom. The density in the lower tail of the multivariate $F_{1,\nu}$ distribution comes from the centre of the t_ν -distribution is created from, and the density in the heavy upper tail comes from the tails of the t_ν -distribution. Since t -distributions have less mass in the centre and more in the tails the lower their degrees of freedom, it follows that the lower ν is, the heavier and sharper the upper tail of $F_{1,\nu}$ is on the normal scores scatterplot. As the normal distribution is a special case of the t -distribution with infinite degrees of freedom, the χ_1^2 distribution is a special case of the $F_{1,\nu}$ distribution where $\nu = \infty$. Hence the χ_1^2 distribution has a blunter upper tail than any other $F_{1,\nu}$ -distribution. The density of the bivariate F -copula, c_ρ^F , in terms of the bivariate t -copula with correlation or scale parameter ρ is, c_ρ^t , is:

$$c_\rho^F(u_1, u_2) = \frac{1}{2} c_\rho^t\left(\frac{1+u_1}{2}, \frac{1+u_2}{2}\right) + \frac{1}{2} c_\rho^t\left(\frac{1+u_1}{2}, \frac{1-u_2}{2}\right) \quad (3.8)$$

This is derived in Appendix E.

Using equation 3.1 in section 3.3 for normal variance mixture t -copulas, the bivariate t -copula density is:

$$c_\rho^t(a, b) = \frac{\mathbf{t}_{\nu, \rho}(T_\nu^{-1}(a), T_\nu^{-1}(b))}{t_\nu(T_\nu^{-1}(a)) t_\nu(T_\nu^{-1}(b))}$$

And of course, due to the same computational issues as for the normal model, we use the pairwise likelihood summed over all n days of observations:

$$\sum_{i=1}^n \sum_{j=2}^d \sum_{k=1}^{j-1} \log c_\rho^t(u_{i,j}, u_{i,k})$$

The $F_{1,\nu}$ -distribution has a more complicated correlation structure than the chi-squared distribution. Correlation is not actually defined in a multivariate $F_{1,\nu}$ distribution for

$\nu < 4$. Where it is defined, correlation in the model depends on both Σ and the degrees of freedom ν . The correlation between W_i and W_j is given by:

$$\text{CORR}(W_i, W_j) = \frac{(\nu - 2) \text{CORR}(Z_i, Z_j)^2 + 1}{\nu - 1} \quad (3.9)$$

This is shown in appendix F.

3.5.4 Fitting chi-squared copulas with Kendall's Tau

Kendall's Tau is a type of correlation based on the difference between concordance and discordance in the distribution (Joe 2014). If we let (X_1, X_2) and $(\tilde{X}_1, \tilde{X}_2)$ be two independent, identically distributed random vectors, then they are considered concordant if $(X_1 - \tilde{X}_1)(X_2 - \tilde{X}_2) > 0$, and discordant if $(X_1 - \tilde{X}_1)(X_2 - \tilde{X}_2) < 0$. The population version of Kendall's Tau is:

$$\tau = \Pr \left[(X_1 - \tilde{X}_1)(X_2 - \tilde{X}_2) > 0 \right] - \Pr \left[(X_1 - \tilde{X}_1)(X_2 - \tilde{X}_2) < 0 \right]$$

And the sample version of Kendall's Tau, for n observations of paired data $(x_{11}, x_{12}), \dots, (x_{n1}, x_{n2})$ is:

$$\hat{\tau} = \frac{2}{n(n-1)} \sum_{i=2}^n \sum_{j=1}^{i-1} [\mathbf{1} \{(x_{i1} - x_{j1})(x_{i2} - x_{j2}) > 0\} - \mathbf{1} \{(x_{i1} - x_{j1})(x_{i2} - x_{j2}) < 0\}]$$

Where $\mathbf{1}$ is an indicator variable. Note that like Spearman rank correlation, the sample version of Kendall's Tau depends on the ranks of different observations rather than their absolute values, and as such is a distribution free method that is invariant to a probability integral transformation.

Quessy et al. (2016) have found an analytic form for Kendall's Tau for the central chi squared copula. Let us assume we have a standard normal bivariate distribution (Z_1, Z_2) such that:

$$(Z_1, Z_2) \sim N \left(0, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right)$$

Then the population version of Kendall's Tau for chi-squared copula with the dependence structure (Z_1^2, Z_2^2) is given by:

$$\tau(C_\rho^\chi) = \left(\frac{2}{\pi} \sin^{-1}(\rho) \right)^2$$

The covariance matrix of the chi-square copula can then be fit to minimise the squared distance between the theoretical value of Kendall's Tau and the sample value for Kendall's Tau, like so:

$$\sum_{i=1}^n \sum_{j=1}^{i-1} \left(\hat{\tau}_{ij} - \tau(C_{\rho_{ij}}^\chi)_{ij} \right)^2$$

Unfortunately we could not extend fitting with Kendall's Tau to the $F_{1,\nu}$ copula. It does not have the same value of Kendall's Tau as the Chi-squared copula, and to the best of the author's knowledge, there is no quick formula to calculate it in the literature.

3.5.5 Fitting models with Minimum Distance Estimators (MDEs)

We also tried a new method of fitting the chi-squared and F -copulas, where the Pearson covariance in the data was matched with the theoretical Pearson covariance in the model. This is an example of minimum distance estimation.

Chi-squared copulas

Since the correlation between (Z_1, \dots, Z_d) in the elliptical copulas was denoted as Σ , we can say the correlation in the chi-squared copulas is held in the matrix Σ^2 , such that:

$$\Sigma_{ij}^2 = \text{CORR}(W_i, W_j) = \text{CORR}(Z_i, Z_j)^2$$

The reflected quantiles of the residuals were transformed to χ_1^2 margins, and then used to calculate the sample Pearson correlation (denoted here as $\hat{\Sigma}^2$). Then the parameters of the 1,2 and 3 factor models were optimised to minimise the sum of squared errors between the empirical and theoretical correlation, or:

$$\sum_{i=1}^n \sum_{j=1}^{i-1} \left(\hat{\Sigma}_{ij}^2 - \Sigma_{ij}^2 \right)^2$$

F -copula

The F -copula was more challenging to fit by a minimum distance method, as we need to fit not only a scale matrix, but the degrees of freedom.

While ν does effect the correlation, it is really a tail measurement. Thus estimating ν from a central measure, like correlation, is not a sound approach. It must be fit by some other method.

We tried to fit this model by estimating a loadings matrix \mathbf{L} that matched the Pearson correlation in the data, such that:

$$(\mathbf{L}^T \mathbf{L})_{ij} = \text{CORR}(W_i, W_j) = \frac{\Sigma_{ij}^2(\nu - 2) + 1}{\nu - 1}$$

Then the degrees of freedom, ν , was estimated by pairwise likelihood. The correlation parameter for the bivariate t -copulas used to calculate the pairwise likelihood (as in equation 3.8) for stocks i and j was calculated as:

$$\hat{\Sigma}_{ij} = \max \left\{ 0, \sqrt{\frac{(\nu - 1)(\mathbf{L}^T \mathbf{L})_{ij} - 1}{\nu - 2}} \right\}$$

As we can see in the formula above, we need a constraint to prevent the value of $\hat{\Sigma}_{ij}$ becoming negative. This copula provided a very poor fit, which likelihoods up to 20 percent worse than the full pairwise copula. A modified two stage approach was tried, which provided an even worse fit.

After estimating the degrees of freedom parameter failed, we choose to fix it, and instead estimate the loadings matrix such that $(\mathbf{L}^T \mathbf{L})_{ij} = \Sigma_{ij}$. This method rapidly estimated Σ_{ij} . This allowed us to quickly fit several models, with different, fixed degrees of freedom, which we could then compare with diagnostic measures, including tail measure and likelihood.

Once the degrees of freedom is fixed, it becomes a shape parameter controlling the heaviness of the bottom tail that the modeller can control. Since fitting with likelihood is so computationally expensive, this provides a method of fitting where we can take advantage of the heavier tail in the F -distribution, while still fitting the model rapidly.

The great limitation of fitting by correlation is we need to assume $\nu > 4$, as correlation does not exist below these values.

3.5.6 Diagnostics

We did not calculate the AIC and BIC for these models.

We compared models by upper and lower tail semicorrelation, empirical Spearman rank correlation and covariance, which are explained in section 3.3.4. The only difference is the covariance in the F -copulas is not given by the scale matrix Σ . The theoretical Pearson

correlation of the model, denoted here as \mathbf{C} , can be calculated from the scale matrix and degrees of freedom with the formula:

$$\mathbf{C} = \frac{(\nu - 1)\Sigma^2 + 1}{\nu - 2}$$

The differences between theoretical and sample statistics were calculated by subtracting the theoretical statistic from the sample statistic. When the difference is negative, the theoretical statistic is larger than the sample one.

3.5.7 Results

Model	Fitting Method	Time (s)	Number of parameters
1 Factor χ^2	Pairwise likelihood	49	40
2 Factor χ^2	Pairwise likelihood	974	79
3 Factor χ^2	Pairwise likelihood	3120	117
1 Factor $F_{1,\nu}$	Pairwise likelihood	514	41
2 Factor $F_{1,\nu}$	Pairwise likelihood	2133	80
3 Factor $F_{1,\nu}$	Pairwise likelihood	21353	118

Table 3.4: Time taken to fit chi-squared and F -copulas, in seconds

We fit chi-squared copulas to the residuals from the ARMA-GARCH models with three methods: pairwise likelihood, with Kendall's Tau and with MDEs. We also fit F -copulas to the residuals with pairwise likelihood and MDEs. Since we fit the copulas to the flipped residuals, the heavy tail that reflects the tendency of stock prices crashing together is now the upper tail. Hence, it becomes important to match the theoretical upper semicorrelation to the sample upper semicorrelation.

There is a pattern in these results that indicate the chi-squared and F -copulas are not good models to approximate the whole shape of the dependency. All models provided poor fits of lower semicorrelation (or the light tail in the dependency of the data), often underestimating it by about 0.18. Moreover, there was a pattern where models that fit the correlation well fit the upper tail semicorrelation poorly, and vice versa.

The best model and method for fitting the upper tail was the chi-squared copula with 3 factor structure, fit by pairwise likelihood. The chi-squared copulas with 1 and 2 factor structure fit by pairwise likelihood fitted larger upper semicorrelation than in the sample,

with all differences in semicorrelation being negative. The difference was quite large for the 1 factor model, which had a difference and absolute difference in upper semicorrelation of -0.10 and 0.11 respectively, and very reasonable for the 2 factor model, which had a difference and absolute difference of -0.03 and 0.06 respectively. The 3 factor model fit the tail very well, with a small amount of variability, with a difference and absolute difference of -0.01 and 0.05 respectively. The 3 factor model also had good results sector-by-sector, with the worst difference in upper semicorrelation being only -0.03, between sectors 3 and 4. The 3 factor model, fitted with pairwise likelihood, appears to be the only unbiased method for fitting the heavy tail of the residuals among the asymmetric and elliptical copulas. Furthermore, we have good between and within sector results and a relatively low absolute difference overall in upper semicorrelation, compared to the 1 and 2 factor models. The model is not just fitting good estimates overall — it is fitting results between and within sectors with less bias and variability.

The chi-squared copulas fit by pairwise likelihood overfit the covariances in the model. However, the 2 and 3 factor models had the lowest difference in covariances of all the chi-squared copulas fit with pairwise likelihood and Kendall's Tau, with both having mean differences between theoretical and sample pairwise covariances of -0.09, and mean absolute differences of 0.10 and 0.09 respectively. The models also had a decent fit for Spearman rank correlation, with both having a mean difference of 0.07.

All the models fit by pairwise likelihood were slow and computationally expensive. We have a simulation study that show this in the next section. Table 3.4 shows the time the copulas fit with pairwise likelihood took to fit. The results are in line with what we would expect for a dataset of this size. The 2 factor model took 16 minutes to fit, and the 3 factor model took 52 minutes. The gradient was not calculated or implemented, which would likely have sped up fitting considerably. However, we know any pairwise likelihood method will become increasingly hard to fit as the number of dimensions increase, and this method will likely never be able to fit truly large data sets. To fit this 40 dimensional model, we needed to calculate 780 pairwise log-likelihoods. To fit a high-dimensional model of 80 variables, we would need to calculate 3160. Since this model cannot be scaled up to fit large data, we probably cannot justify the relatively poor fit to the overall shape of the dependence. It would be better to choose a more flexible, slow method to fit medium dimensional stock data.

The chi-squared copulas with 1, 2 and 3 factor structure fit with Kendall's Tau all provided similar differences in the four statistics measured, when taken across all 40

variables. It still had a decent fit for upper semicorrelation, with a difference of about -0.07 for all models. Unsurprisingly, the fit for difference in Spearman rank correlation was the best out of all models. Since both Kendall's Tau and Spearman rank correlation are distribution free, rank based methods of calculating correlation, it makes sense that a model based on minimising the differences in Kendall's Tau would also minimise the differences in Spearman rank correlation. The 3 factor model fit with Kendall's Tau had a mean difference in Spearman rank correlation of 0.00, and mean absolute difference of 0.04. It is clear when we look at the sector by sector results, that the model becomes better at fitting Spearman rank correlation within and between sectors as more factors are added.

The MDE method for the chi-squared copula yielded the lowest differences in Pearson correlation, but worse fits for Spearman rank correlation, and upper semicorrelation, compared to the models fit with Kendall's Tau and pairwise likelihood. It tended to underfit upper semicorrelation, with the 1 factor model having a difference of 0.05, and the 2 and 3 factor models having a difference of 0.04. This is a better result than the elliptical t -copulas, which underfit semicorrelation in the heavy tail* by about 0.08 in all models. However, the t -copulas had much better fits for the other tail, as well as the Spearman rank correlation.

While pairwise likelihood is the most efficient method of fitting the chi-squared copulas, fitting with Kendall's Tau and MDE has the great advantage of speed, with rapid fitting in the simulation study explained in the next section. They could be scaled up to fit higher dimensional data sets.

The F -copulas fit by pairwise likelihood had similar results to the Kendall's Tau method for upper semicorrelation, as all models had a difference of around -0.07 and an absolute difference of a similar magnitude. The F -copulas had slightly worse results for fitting the correlation in the model, with differences for all models in pairwise correlation of about -0.22[†]. The semicorrelation was fine, but Kendall's Tau was better. The F -copulas fit by pairwise likelihood were extremely slow to fit, as seen in table 3.4, even on our moderate sized dataset. The 3 factor model took almost 6 hours to fit! For this dataset, fitting a chi-squared copula with Kendall's Tau very clearly gives a faster and better result than

*Recall the t -copulas were fit with the quantiles from the margins, and the chi-squared copulas to the reflected quantiles of the margins, so we need to compare lower semicorrelation in the t -copulas with upper semicorrelation in the chi-squared copulas and vice versa

[†]Since this difference was so high, we double checked our calculations for the theoretical correlations by simulating new data from the fitted model. The simulations showed empirical correlation consistent with the theoretical correlation, indicating our calculations are correct.

fitting a F -copula with pairwise likelihood, despite the F -copula having greater flexibility to fit the tails.

The F -copulas fit with MDE methods yielded the worst estimates of upper semicorrelation, with all three models having a difference and absolute difference of 0.09 and 0.11. Notably, the estimates of Spearman rank correlation were very poor, with differences of about 0.24 for all models. The elliptical t -copulas provided a slightly better fit to the heavy tail than the F -copulas fit with MDEs, while having superior results fitting the light tail and Spearman rank correlation.

We needed to specify the degrees of freedom, ν , when fitting the F -copulas with MDEs. We set $\nu = 5$. This is about as heavy a tail we can fit with this method, as we require that $\nu > 4$ so that Pearson correlation exists in the theoretical model. We wanted to see if the model could both fit the Pearson correlation well, and fit an upper tail as heavy, or heavier than in the dataset. If the tail was heavier, we could then try higher values of ν to blunten it. Recall that the chi-squared copula fit by MDEs fit the Pearson correlation well but underfit the upper semicorrelation, and the chi-squared distribution is the limiting distribution of the $F_{1,\nu}$ distribution as ν approaches infinity. If the $F_{1,5}$ -copula fit with MDEs both fit the Pearson correlation well and overfit the upper semicorrelation, it would indicate there may be an F -copula with ν between these two extremes that provided good fits for both statistics. Unfortunately the $F_{1,5}$ -copula also underfit the upper semicorrelation, worse than the chi-squared copula fit with MDEs. Moreover, even though it was fitted to match the Pearson correlation in the data, it was not that close, with a difference of -0.05 for Pearson correlation for 1, 2 and 3 factor models. This points to a fundamental flaw in the model. If we look at the Pearson correlation in the F -copula (see equation 3.9), we can see the minimum Pearson correlation in the theoretical model is $\frac{1}{\nu-1}$ [‡]. However, some pairs of variables in the sample may have low Pearson correlation once they are transformed to F -scores. The model does not have the flexibility to fit this.

One important distinction to make is about whether the models overfit or underfit the semicorrelation in the heavy tail. When the models overfit, we would expect them to tend towards overestimating the risk of all stocks crashing together, whereas when they underestimate, we would expect them to tend towards the opposite. If we cannot choose models that perfectly fit the tail due to tractability issues, we may want the flexibility to choose whether we tend towards overestimating, or underestimating the risk. The best, tractable models that overestimated the risk are the factor chi-squared models fit with

[‡]Where it exists

Kendall's Tau. However, they should only be used for tail inference, as they have fairly poor fits to the centre of the data.

In this whole thesis, the best models that under the heavy tail are the chi-squared models fit with Minimum Distance Estimators, and the t -copulas. In our data, chi-squared copulas have closer semicorrelation to the heavy tail than the t -copulas, which is particularly apparent in the sector-by-sector results. Both copulas also estimate the Pearson correlation well, which is totally unsurprising in the case of the MDE estimator, as it was calculated entirely from the Pearson correlation. However the chi-squared copula has very poor results estimating semicorrelation in the light tail, and Spearman rank correlation, indicating it really does not fit the overall shape of the dataset. In our normal scores scatterplots, we saw some bivariate scatter plots have blunt light tails, whereas some have very sharp ones. Designing a tractable model, able to be scaled up into high dimensions, that has the flexibility to fit both of these tails would be an interesting challenge. From the models in this thesis: the t -copula has more flexibility in the light tail, but cannot fit a blunt light tail and a heavy lower one. The chi-squared copulas can only fit a blunt light tail.

3.5.8 Simulation Study

Finally, we compared all the fitting methods for chi-squared copulas tried in thesis in a simulation study, to test the speed of the models.

For the chi-squared copulas, 100 datasets for each of a 1, 2 and 3 factor model were simulated from a 10 dimensional χ^2 -distribution with correlation matrix Σ , which was simulated to have close values to the correlation in the stocks. The datasets were then fit with pairwise likelihood, Kendall's Tau and the MDE estimator. The same differences in statistics we used to evaluate all other models were calculated. However, since we know the scale matrix Σ of the simulated F -distributions, we could compare how well the different models fit them, rather than finding the difference between theoretical and fitted correlation. Results are in appendix G.

For the chi-squared copulas, the pairwise likelihood method was the best at estimating the correlation matrix, however both the method using Kendall's Tau and the Minimum Distance Estimator were better at recovering Spearman rank correlation and the difference in upper and lower semicorrelation. That said, all results were fairly close. We see the elapsed time for fitting the pairwise likelihood increases greatly as the number of parameters in the model increases, with the 1 factor model taking an average of 0.75

seconds to fit, and the 3 factor model taking an average of 34.41 seconds to fit. The Kendall's Tau and MDE estimators followed this pattern, but were much faster, with the 3 factor model estimated by both Kendall's Tau and the MDE being fit in less than a second. When the underlying dataset is truly a chi-squared copula, fitting with Kendall's Tau and MDEs give very similar results. We saw this was not the case when fitting the real world data, where the two different methods fit different parts of the joint model well.

100 datasets were also simulated from a 10-dimensional $F_{1,5}$ distribution with the same 1, 2 and 3 factor correlation matrices as for the chi-squared datasets. Then the distributions were fit by both pairwise likelihood and the MDE. In the case of the MDE, ν was fixed at 5. By comparison, the pairwise likelihood method needed to estimate the degrees of freedom.

For the F -models, the pairwise likelihood method was superior compared to the MDE method on all diagnostics. However, it took many times as long to fit. All three MDE models could be fit in less than a second. The 1 factor pairwise likelihood model needed 19.94 seconds to fit, whereas the 3 factor one needed 194.33 seconds to fit. The pairwise likelihood method had a large disadvantage computationally, but greater flexibility, as it also estimated the degrees of freedom parameter. Even if we fixed the degrees of freedom parameter, there is no reason to suspect it would take less time to fit than the χ_1^2 copula. As both these copulas suffer the curse of dimensionality, we know they will not be able to be scaled up to fit higher dimensional datasets. We saw this with our dataset of 40 variables.

4: Discussion

Out of the elliptical copulas, the nested, bifactor and 3 factor t -copulas were all good models for estimating both heavy tail in the data, and the centre of the data. These models provided the best overall fit to all statistics tested. The bifactor model was slower to fit, but this is probably largely because the gradient was not implemented. There is already a model based on elliptical copulas with 1 factor structure in high dimensions (Krupskii & Joe 2020). It stands to reason that while likely difficult, creating models for elliptical copulas with these more complicated covariance structures is possible.

Overall, we did not find any models that fit the shape of the whole dependency well. The elliptical copulas are symmetric, which does not match the asymmetric shape of the data. The chi-squared copula will always lack the flexibility to fit anything but a blunt light tail. Additionally, without a parameter controlling tail sharpness, the density in the tails has to be fit from the correlation parameters. We saw in our data that there was no chi-squared copula that fit both the heavy tail and the correlation in the data extremely well.

On the other hand, both the t -copulas and chi-squared copulas fit parts of the data well. If the methods used in this thesis can be refined and used to fit higher dimensional datasets quickly, these models could be extremely useful in calculating certain metrics, such as the Value at Risk, on larger datasets.

We only fit a medium dimensional data set of 40 stocks. We can infer from this what methods are suitable to be scaled up to fit higher dimensional data sets. The pairwise likelihood methods are clearly unsuitable for this. While the 3 factor chi-squared copula fitted with pairwise likelihood had good fits to the centre and fat tail in the data, its computational burden means it will only ever be able to fit small and medium datasets. In that case, we may as well choose a slow fitting, but highly flexible model that is well established as a standard way to fit Copula-GARCH models, like a vine.

Another possible area of research is into approximations of the multivariate chi-squared likelihood. If we can find an approximation that can be calculated quickly, we may be able to design a method that competes with fitting chi-squared copulas by Kendall's Tau.

The F -copulas were investigated, but did not show promising results. Partially this is due to the fact that the chi-squared copula turned out to have the flexibility to fit the heavy tail in the data well, meaning we did not need a sharper model. It would be interesting to test chi-squared copulas on other financial datasets to see if this remains the case. The chi-squared copula has previously been used for geostatistical data, and a literature review did not yield any applications to financial data sets. If this pattern holds true, the chi-squared copula is potentially a rich field of research for financial applications. Fitting with Kendall's Tau and MDE estimators was fairly rapid, and provided decent fits of the heavy tail. It would also be interesting to see if these copulas remain good estimators of the tail in other applications.

The chi-squared and F -copulas also lack the interpretability of the elliptical copulas, due to their correlation structures. For the chi-squared copula, in the 1 factor model we can still interpret the correlation as being based on one global, latent factor. In the 2 and 3 factor models, we can possibly say we have dependence on latent factors, but they interact. This is a stretch. We really chose factor structure because it makes creating the correlation matrix simple. For the F -copulas, the structure of the correlation matrix destroys any interpretability.

The chi-squared copula relies on the covariance matrix to fit both correlation, and the heavy tail in the data. The hope of the F -copulas was that with the extra parameter that controls heaviness in the upper tail, it would have more flexibility to fit both the tail and the centre of the data. That turned out not to be the case, probably largely due to its inflexible correlation structure. However, we can investigate other non-monotonic transformations of elliptical copulas, to see if they have the flexibility to fit both the heavy tail and correlation in the data. Two other attractive copulas from the geostatistical literature are non-central chi-squared copulas (Quessy et al. 2016), and copulas constructed by taking the absolute value of underlying normal random variables (Bárdossy & Li 2008). The former copula may be particularly promising. Quessy et al. (2016) showed the normal copula is the limiting distribution of the non-central chi-squared copula when its the non-centrality parameters approach infinity. The (central) chi-squared copula used in our analysis provided better fits in the heavy tails than the normal copula, whereas the normal copula provided good fits to the centre, and light tail of the data. The non-central

copula may be able to combine these properties in a way that provides a better overall fit to the data. Quessy et al. (2016) also provide the formula for Kendall's Tau for the non-central chi-squared copula in their paper, although not in such an easy form as used in this thesis, paving the way for a potential rapid method of fitting.

Appendices

A: Stock tickers & Marginal Models

Ticker	Sector	ARMA order	GARCH order
GOOGL	Technology	(0,0)	(1,1)
AAPL	Technology	(0,0)	(1,1)
MSFT	Technology	(0,1)	(1,1)
IBM	Technology	(0,0)	(1,1)
T	Technology	(1,1)	(1,1)
INTC	Technology	(0,0)	(1,1)
EA	Technology	(0,0)	(1,1)
AMD	Technology	(0,0)	(1,1)
NVDA	Technology	(0,0)	(1,1)
ATVI	Technology	(0,0)	(1,1)
JPM	Financial	(0,1)	(1,1)
BAC	Financial	(0,0)	(1,1)
C	Financial	(0,0)	(1,1)
AIG	Financial	(0,0)	(1,1)
MET	Financial	(0,1)	(1,1)
ANZ.AX	Financial	(0,0)	(1,1)
CS	Financial	(0,0)	(1,1)
PGR	Financial	(1,0)	(1,1)
MCO	Financial	(0,0)	(1,1)
SCHW	Financial	(1,1)	(1,1)
GSK	Pharmaceutical	(0,0)	(1,1)
DD	Pharmaceutical	(0,0)	(1,1)
NVS	Pharmaceutical	(0,0)	(1,1)
MRK	Pharmaceutical	(0,0)	(1,1)
LLY	Pharmaceutical	(0,0)	(1,1)
JNJ	Pharmaceutical	(0,0)	(1,1)
PFE	Pharmaceutical	(0,0)	(1,1)
BMY	Pharmaceutical	(0,0)	(1,1)
AMGN	Pharmaceutical	(0,0)	(1,1)
ABT	Pharmaceutical	(0,0)	(1,1)
KO	Consumer staples	(0,0)	(1,1)
PEP	Consumer staples	(0,0)	(1,1)
NESN.SW	Consumer staples	(0,0)	(1,1)
MDLZ	Consumer staples	(0,0)	(1,1)
YUM	Consumer staples	(0,0)	(1,1)
CPB	Consumer staples	(0,0)	(1,1)
TSN	Consumer staples	(0,0)	(1,1)
K	Consumer staples	(0,0)	(1,1)
CL	Consumer staples	(0,1)	(1,1)
HSY	Consumer staples	(0,1)	(1,1)

B: Gradient for Copulas With 1, 2 or 3 Factor Structure

B.1 1 Factor normal model

Due to the constraints used, in the 1 factor normal model, the values that are optimised to get the maximum likelihood of the copula are a vector of real numbers x_{a_1} to x_{a_d} . These are then transformed from \mathbb{R}^d to $[0, 1]^d$ using the following sigmoidal transformation:

$$a_j = \frac{2}{1 + e^{x_{a_j}}} - 1, \quad j = 1, \dots, d$$

a_1 to a_d are then used to create the covariance matrix Σ . So we need to find the gradient with respect to x_{a_1} to x_{a_d} . Since in the likelihood only Σ depends on these terms, the gradient of x_{a_j} is:

$$\begin{aligned} & \frac{\partial}{\partial x_{a_j}} \{-\log c(\mathbf{u}_1, \dots, \mathbf{u}_d; \Sigma)\} \\ &= \frac{\partial}{\partial x_{a_j}} \left\{ -\sum_{i=1}^T \log (\phi(\Phi^{-1}(u_{i,1}), \dots, \Phi^{-1}(u_{i,d}); \Sigma)) + \sum_{i=1}^T \sum_{j=1}^d \log (\phi(\Phi^{-1}(u_{i,j}))) \right\} \\ &= \frac{\partial}{\partial x_{a_j}} \left\{ -\sum_{i=1}^T \log (\phi(\Phi^{-1}(u_{i,1}), \dots, \Phi^{-1}(u_{i,d}); \Sigma)) \right\} \end{aligned}$$

We know the log-density of a d -dimensional multivariate normal random vector \mathbf{x} with zero mean and covariance matrix Σ is:

$$\log \phi_{\{1:d\}}(\mathbf{x}) = -\frac{1}{2} \log \det(\Sigma) - \frac{d}{2} \log 2\pi - \frac{1}{2} \mathbf{x}^T \Sigma^{-1} \mathbf{x} \quad (\text{B.1})$$

We also need the following equations results from matrix algebra:

$$\frac{\partial \Sigma^{-1}}{\partial a_j} = -\Sigma^{-1} \frac{\partial \Sigma}{\partial a_j} \Sigma^{-1} \quad (\text{B.2})$$

$$\frac{\partial \det(\Sigma)}{\partial a_j} = \det(\Sigma) \operatorname{tr} \left(\Sigma^{-1} \frac{\partial \Sigma}{\partial a_j} \right) \quad (\text{B.3})$$

So the gradient becomes:

$$\begin{aligned} & \frac{\partial}{\partial x_{a_j}} \{-\log c(\mathbf{u}_1, \dots, \mathbf{u}_d; \Sigma)\} \\ &= \frac{\partial}{\partial x_{a_j}} \frac{1}{2} \left\{ T \log \det(\Sigma) + \sum_{i=1}^T \Phi^{-1}(\mathbf{u}_i^T) \Sigma^{-1} \Phi^{-1}(\mathbf{u}_i) \right\} \\ &= \left\{ \frac{T}{2 \det(\Sigma)} \frac{\partial}{\partial a_j} \{\det(\Sigma)\} + \frac{1}{2} \sum_{i=1}^T \Phi^{-1}(\mathbf{u}_i^T) \frac{\partial \Sigma^{-1}}{\partial a_j} \Phi^{-1}(\mathbf{u}_i) \right\} \frac{\partial a_j}{\partial x_{a_j}} \quad (\text{Chain rule}) \\ &= \left\{ \frac{T}{2} \operatorname{tr} \left(\Sigma^{-1} \frac{\partial \Sigma}{\partial a_j} \right) - \frac{1}{2} \sum_{i=1}^T \Phi^{-1}(\mathbf{u}_i^T) \Sigma^{-1} \frac{\partial \Sigma}{\partial a_j} \Sigma^{-1} \Phi^{-1}(\mathbf{u}_i) \right\} \frac{\partial a_j}{\partial x_{a_j}} \quad (\text{from B.2,B.3}) \end{aligned}$$

Where:

$$\frac{\partial a_j}{\partial x_{a_j}} = \frac{2e^{-a_j}}{(1 + e^{-a_j})^2}$$

In the 1 factor model, the covariance matrix is:

$$\Sigma = \begin{bmatrix} 1 & a_1 a_2 & a_1 a_3 & \dots & a_1 a_d \\ a_1 a_2 & 1 & a_2 a_3 & \dots & a_2 a_d \\ a_1 a_3 & a_2 a_3 & 1 & \dots & a_3 a_d \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_1 a_d & a_2 a_d & a_3 a_d & \dots & 1 \end{bmatrix}$$

So it is quite easy to see that:

$$\frac{\partial \Sigma}{\partial a_1} = \begin{bmatrix} 0 & a_2 & a_3 & \dots & a_d \\ a_2 & 0 & 0 & \dots & 0 \\ a_3 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_d & 0 & 0 & \dots & 0 \end{bmatrix}, \quad \frac{\partial \Sigma}{\partial a_2} = \begin{bmatrix} 0 & a_1 & 0 & \dots & 0 \\ a_1 & 0 & a_3 & \dots & a_d \\ 0 & a_3 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & a_d & 0 & \dots & 0 \end{bmatrix}, \quad \dots, \quad \frac{\partial \Sigma}{\partial a_d} = \begin{bmatrix} 0 & 0 & 0 & \dots & a_1 \\ 0 & 0 & 0 & \dots & a_2 \\ 0 & 0 & 0 & \dots & a_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_1 & a_2 & a_3 & \dots & 0 \end{bmatrix}$$

B.2 2 Factor normal model

In the 2 factor normal model we need to find the gradient for $x_{a_2}, \dots, x_{a_d}, x_{b_1^*}, \dots, x_{b_d^*}$. There is no x_{a_1} since a_1 is set to zero in this model. Like before we have that:

$$b_j^* = \frac{2}{1 + e^{x_{b_j^*}}} - 1, \quad a_j = \frac{2}{1 + e^{x_{a_j}}} - 1, \quad j = 1, \dots, d$$

However now we have the complication that b_j depends on a_j , like so:

$$b_j = b_j^* \times \sqrt{1 - a_j^2}, \quad j = 1, \dots, d$$

For x_{a_1} to x_{a_d} it is still the case that:

$$\begin{aligned} & \frac{\partial}{\partial x_{a_j}} \{-\log c(\mathbf{u}_1, \dots, \mathbf{u}_d; \Sigma)\} \\ &= \left\{ \frac{T}{2} \text{tr} \left(\Sigma^{-1} \frac{\partial \Sigma}{\partial a_j} \right) - \frac{1}{2} \sum_{i=1}^T \Phi^{-1}(\mathbf{u}_i^T) \Sigma^{-1} \frac{\partial \Sigma}{\partial a_j} \Sigma^{-1} \Phi^{-1}(\mathbf{u}_i) \right\} \frac{\partial a_j}{\partial x_{a_j}} \end{aligned} \quad (\text{B.4})$$

However, $\frac{\partial \Sigma}{\partial a_j}$ needs to be changed to take into account that b_j now depends on a_j . For example, we now have that $\frac{\partial \Sigma}{\partial a_2}$ is:

$$\frac{\partial \Sigma}{\partial a_2} = \begin{bmatrix} 0 & \frac{\partial b_1 b_2}{\partial a_2} & 0 & \dots & 0 \\ \frac{\partial b_1 b_2}{\partial a_2} & 0 & a_3 + \frac{\partial b_2 b_3}{\partial a_2} & \dots & a_d + \frac{\partial b_2 b_d}{\partial a_2} \\ 0 & a_3 + \frac{\partial b_2 b_3}{\partial a_2} & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & a_d + \frac{\partial b_2 b_d}{\partial a_2} & 0 & \dots & 0 \end{bmatrix}$$

Where, remembering that $b_2 = b_2^* \sqrt{1 - a_2^2}$,

$$\begin{aligned} \frac{\partial b_1 b_2}{\partial a_2} &= b_1 \frac{\partial}{\partial a_2} b_2^* \sqrt{1 - a_2^2} \\ &= -\frac{a_2 b_1 b_2^*}{\sqrt{1 - a_2^2}} \\ &= \frac{a_2 b_1 b_2}{a_2^2 - 1} \end{aligned}$$

The gradient for $x_{b_1^*}$ to $x_{b_d^*}$ is quite similar to the gradient for x_{a_1} to x_{a_d} in the 1 factor normal model, except we have the extra term $\frac{\partial b_j}{\partial b_j^*}$ to account for the transformation from

b_j^* to b_j . Hence the gradient is:

$$\begin{aligned} & \frac{\partial}{\partial x_{b_j^*}} \{-\log c(\mathbf{u}_1, \dots, \mathbf{u}_d; \Sigma)\} \\ &= \left\{ \frac{T}{2} \text{tr} \left(\Sigma^{-1} \frac{\partial \Sigma}{\partial b_j} \right) - \frac{1}{2} \sum_{i=1}^T \Phi^{-1} (\mathbf{u}_i^T) \Sigma^{-1} \frac{\partial \Sigma}{\partial b_j} \Sigma^{-1} \Phi^{-1} (\mathbf{u}_i) \right\} \frac{\partial b_j}{\partial b_j^*} \frac{\partial b_j^*}{\partial x_{b_j^*}} \end{aligned} \quad (\text{B.5})$$

Where:

$$\frac{\partial b_j}{\partial b_j^*} = \sqrt{1 - a_j^2}, \quad j = 1, \dots, d, \quad a_1 = 0$$

And:

$$\frac{\partial \Sigma}{\partial b_1} = \begin{bmatrix} 0 & b_2 & b_3 & \dots & b_d \\ b_2 & 0 & 0 & \dots & 0 \\ b_3 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ b_d & 0 & 0 & \dots & 0 \end{bmatrix}, \quad \frac{\partial \Sigma}{\partial b_2} = \begin{bmatrix} 0 & b_1 & 0 & \dots & 0 \\ b_1 & 0 & b_3 & \dots & b_d \\ 0 & b_3 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & b_d & 0 & \dots & 0 \end{bmatrix}, \dots, \quad \frac{\partial \Sigma}{\partial b_d} = \begin{bmatrix} 0 & 0 & 0 & \dots & b_1 \\ 0 & 0 & 0 & \dots & b_2 \\ 0 & 0 & 0 & \dots & b_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ b_1 & b_2 & b_3 & \dots & 0 \end{bmatrix}$$

B.3 3 Factor normal model

In the 3 factor normal model we need to find the gradient for

$x_{a_3}, \dots, x_{a_d}, x_{b_2^*}, \dots, x_{b_d^*}, x_{c_1^*}, \dots, x_{c_d^*}$. There is no x_{a_1} , x_{a_2} or x_{b_1} as a_1 , a_2 and b_1 are set to zero in the 3 factor model. We now need to take into account the following constraints:

$$\begin{aligned} b_j &= b_j^* \times \sqrt{1 - a_j^2}, \quad j = 1, \dots, d \\ c_j &= c_j^* \times \sqrt{1 - a_j^2 - b_j^2}, \quad j = 1, \dots, d \end{aligned}$$

The gradient of x_{a_3}, \dots, x_{a_d} is given by the same formula as the 1 and 2 factor normal model given in B.4. However since both b_j and c_j are dependent on a_j , $\frac{\partial \Sigma}{\partial a_j}$ is changed.

For example, the value of $\frac{\partial \Sigma}{\partial a_3}$ is now:

$$\frac{\partial \Sigma}{\partial a_3} = \begin{bmatrix} 0 & 0 & \frac{\partial c_1 c_3}{\partial a_3} & 0 & \dots & 0 \\ 0 & 0 & \frac{\partial b_2 b_3}{\partial a_3} + \frac{\partial c_2 c_2}{\partial a_3} & 0 & \dots & 0 \\ \frac{\partial c_1 c_3}{\partial a_3} & \frac{\partial b_2 b_3}{\partial a_3} + \frac{\partial c_2 c_2}{\partial a_3} & 0 & a_4 + \frac{\partial b_3 b_4}{\partial a_3} + \frac{\partial c_3 c_4}{\partial a_3} & \dots & a_d + \frac{\partial b_3 b_d}{\partial a_3} + \frac{\partial c_3 c_d}{\partial a_3} \\ 0 & 0 & a_4 + \frac{\partial b_3 b_4}{\partial a_3} + \frac{\partial c_3 c_4}{\partial a_3} & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & a_d + \frac{\partial b_3 b_d}{\partial a_3} + \frac{\partial c_3 c_d}{\partial a_3} & 0 & \dots & 0 \end{bmatrix}$$

The value of $\frac{\partial b_j b_k}{\partial a_j}$ is the same as in the 2 factor model. The value of $\frac{\partial c_j c_k}{\partial a_j}$, for $j = 3, \dots, d$, $k = 1, \dots, d$ and $k \neq j$, is:

$$\begin{aligned}\frac{\partial(c_j c_k)}{\partial a_j} &= c_k \frac{\partial}{\partial a_j} c_j^* \sqrt{1 - a_j^2 - b_j^2} \\ &= -\frac{a_j c_j^* c_k}{\sqrt{1 - a_j^2 - b_j^2}} \\ &= \frac{a_j c_j c_k}{a_j^2 + b_j^2 - 1}\end{aligned}$$

The gradient of x_{b_2}, \dots, x_{b_d} is the same as in the 2 factor normal model, which is given in B.5. However, $\frac{\partial \Sigma}{\partial b_j}$ must be modified to contain term that accounts for the fact that c_j depends on b_j . For example, $\frac{\partial \Sigma}{\partial b_2}$ is now:

$$\frac{\partial \Sigma}{\partial b_2} = \begin{bmatrix} 0 & \frac{\partial c_1 c_2}{\partial b_2} & 0 & \dots & 0 \\ \frac{\partial c_1 c_2}{\partial b_2} & 0 & b_3 + \frac{\partial c_2 c_3}{\partial b_2} & \dots & b_d + \frac{\partial c_2 c_d}{\partial b_2} \\ 0 & b_3 + \frac{\partial c_2 c_3}{\partial b_2} & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & b_d + \frac{\partial c_2 c_d}{\partial b_2} & 0 & \dots & 0 \end{bmatrix}$$

Where, for $k = 1, \dots, d$, $j = 2, \dots, d$, $k \neq j$:

$$\begin{aligned}\frac{\partial c_j c_k}{\partial b_j} &= c_k \frac{\partial}{\partial b_j} c_j^* \sqrt{1 - a_j^2 - b_j^2} \\ &= -\frac{b_j c_k c_j^*}{\sqrt{1 - a_j^2 - b_j^2}} \\ &= \frac{b_j c_j c_k}{a_j^2 + b_j^2 - 1}\end{aligned}$$

The gradient for $x_{c_1^*}$ to $x_{c_d^*}$ is very similar to the gradient for x_{b_1} to x_{b_d} in the 2 factor normal model. The formula is:

$$\begin{aligned}&\frac{\partial}{\partial x_{c_j^*}} \{-\log c(\mathbf{u}_1, \dots, \mathbf{u}_d; \Sigma)\} \\ &= \left\{ \frac{T}{2} \text{tr} \left(\Sigma^{-1} \frac{\partial \Sigma}{\partial c_j} \right) - \frac{1}{2} \sum_{i=1}^T \Phi^{-1}(\mathbf{u}_i^T) \Sigma^{-1} \frac{\partial \Sigma}{\partial c_j} \Sigma^{-1} \Phi^{-1}(\mathbf{u}_i) \right\} \frac{\partial c_j}{\partial c_j^*} \frac{\partial c_j^*}{\partial x_{c_j^*}}\end{aligned}\tag{B.6}$$

Where:

$$\frac{\partial c_j}{\partial c_j^*} = \sqrt{1 - a_j^2 - b_j^2}, \quad j = 1, \dots, d, \quad a_1, a_2, b_1 = 0$$

And:

$$\frac{\partial \Sigma}{\partial c_1} = \begin{bmatrix} 0 & c_2 & c_3 & \dots & c_d \\ c_2 & 0 & 0 & \dots & 0 \\ c_3 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c_d & 0 & 0 & \dots & 0 \end{bmatrix}, \quad \frac{\partial \Sigma}{\partial c_2} = \begin{bmatrix} 0 & c_1 & 0 & \dots & 0 \\ c_1 & 0 & c_3 & \dots & c_d \\ 0 & c_3 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & c_d & 0 & \dots & 0 \end{bmatrix}, \dots, \quad \frac{\partial \Sigma}{\partial c_d} = \begin{bmatrix} 0 & 0 & 0 & \dots & c_1 \\ 0 & 0 & 0 & \dots & c_2 \\ 0 & 0 & 0 & \dots & c_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c_1 & c_2 & c_3 & \dots & 0 \end{bmatrix}$$

B.4 1, 2 & 3 factor t -model

B.4.1 Loadings matrix parameters

Since the t models are constructed and optimised in a very similar way to the normal models, the same methods are used to find the gradients of parameters in the loadings matrix. Firstly, since we use the same constraints on the parameters in the loadings matrix as in the normal models, we still need $\frac{\partial x_{aj}}{\partial a_j}$, $\frac{\partial x_{bj}^*}{\partial b_j^*}$, $\frac{\partial x_{cj}^*}{\partial c_j^*}$, $\frac{\partial b_j^*}{\partial b_j}$ and $\frac{\partial c_j^*}{\partial c_j}$, and these are the same as for the normal models.

Secondly, the 1, 2 and 3 factor t -models have the same structure for their Σ matrices as their normal counterparts, the values of $\frac{\partial \Sigma}{\partial a_j}$, $\frac{\partial \Sigma}{\partial b_j}$ and $\frac{\partial \Sigma}{\partial c_j}$ are the same as the normal models, for $j = 1, \dots, d$.

The gradient of the negative log likelihood of the t -copula with respect to the parameters supplied to make the loadings model is, for x_{a_1}, \dots, x_{a_d} :

$$\begin{aligned} & \frac{\partial}{\partial x_{a_j}} \{-\log c(\mathbf{u}_1, \dots, \mathbf{u}_d; \Sigma)\} \\ &= \frac{\partial}{\partial x_{a_j}} \left\{ -\sum_{i=1}^n \log (\mathbf{t}_\nu(T_\nu^{-1}(u_{i,1}), \dots, T_\nu^{-1}(u_{i,d}); \Sigma, \nu)) \right. \\ & \quad \left. + \sum_{i=1}^n \sum_{j=1}^d \log (t_\nu(T_\nu^{-1}(u_{i,j}); \nu)) \right\} \\ &= \frac{\partial}{\partial x_{a_j}} \left\{ -\sum_{i=1}^n \log (\mathbf{t}_{\{1:d\}}(T_\nu^{-1}(u_{i,1}), \dots, T_\nu^{-1}(u_{i,d}); \Sigma, \nu)) \right\} \end{aligned}$$

The log-density of a d -dimensional multivariate t -distributed random vector \mathbf{x} with zero mean, ν degrees of freedom and covariance matrix Σ is:

$$\begin{aligned}\log \mathbf{t}_{\{1:d\}}(\mathbf{x}) &= \log \Gamma\left(\frac{\nu+d}{2}\right) - \log \Gamma\left(\frac{\nu}{2}\right) - \frac{1}{2} \log \det(\Sigma) - \frac{d}{2} \log \nu - \frac{d}{2} \log \pi \\ &\quad - \frac{\nu+d}{2} \log \left\{1 + \frac{1}{\nu} \mathbf{x}^T \Sigma^{-1} \mathbf{x}\right\}\end{aligned}\tag{B.7}$$

Firstly, let us denote \mathbf{t}_i as:

$$\mathbf{t}_i = T_\nu^{-1}(\mathbf{u}_i)\tag{B.8}$$

For $i = 1, \dots, T$, let us define c_i as:

$$c_i = 1 + \frac{1}{\nu} \mathbf{t}_i^T \Sigma^{-1} \mathbf{t}_i$$

The gradient for the negative log likelihood with respect to x_{a_j} is:

$$\begin{aligned}& \frac{\partial}{\partial x_{a_j}} \{-\log c(\mathbf{u}_1, \dots, \mathbf{u}_d; \Sigma)\} \\&= \frac{\partial}{\partial x_{a_j}} \frac{1}{2} \left\{ T \log \det(\Sigma) + \frac{\nu+d}{2} \sum_{i=1}^T \log c_i \right\} \\&= \left\{ \frac{T}{2 \det(\Sigma)} \frac{\partial}{\partial a_j} \{\det(\Sigma)\} + \frac{\nu+d}{2} \sum_{i=1}^T \frac{1}{c_i} \mathbf{t}_i^T \frac{\partial \Sigma^{-1}}{\partial a_j} \mathbf{t}_i \right\} \frac{\partial a_j}{\partial x_{a_j}} \quad (\text{Chain rule}) \\&= \left\{ \frac{T}{2} \text{tr} \left(\Sigma^{-1} \frac{\partial \Sigma}{\partial a_j} \right) - \frac{\nu+d}{2} \sum_{i=1}^T \frac{1}{c_i} \mathbf{t}_i^T \Sigma^{-1} \frac{\partial \Sigma}{\partial a_j} \Sigma^{-1} \mathbf{t}_i \right\} \frac{\partial a_j}{\partial x_{a_j}} \quad (\text{from B.2,B.3})\end{aligned}$$

In the 2 and 3 factor models the gradient of log likelihood with respect to $x_{b_j^*}$ and $x_{c_j^*}$ is the same as above, just with the term $\frac{\partial a_j}{\partial x_{a_j}}$ replaced with $\frac{\partial b_j}{\partial b_j^*} \frac{\partial b_j^*}{\partial x_{b_j^*}}$ and $\frac{\partial c_j}{\partial c_j^*} \frac{\partial c_j^*}{\partial x_{c_j^*}}$ respectively.

B.4.2 Shape/degrees of freedom parameter

We also need a gradient for the degrees of freedom parameter ν in the t -model. In the t -model, every iteration involves transforming univariate quantiles to t_ν -scores, using the degrees of freedom parameter. Rather than find an exact solution, the gradient of this transformation was approximated using finite differencing.

To take the finite difference, we approximate the gradient of ν by finding a sufficiently

small δ such that:

$$\frac{\partial t_{ij}}{\partial \nu} \approx \frac{T_{\nu+\delta}^{-1}(u_{ij}) - T_{\nu-\delta}^{-1}(u_{ij})}{2\delta}$$

where $T_\nu^{-1}(\cdot)$ is the inverse distribution function of the t_ν -distribution. We know δ is sufficiently small when making it smaller does not significantly change the result. However, if it is too small the computer will not be able to distinguish between $T_{\nu+\delta}^{-1}(x)$ and $T_{\nu-\delta}^{-1}(x)$. We tested various values of ν with successively smaller values of δ to find one where our approximation of $\frac{\partial t_{ij}}{\partial \nu}$ stabilised (i.e. stopped changing value as δ decreased in size). The value of $\delta = 10^{-7}$ was chosen.

Now we have a method approximate $\frac{\partial t_{ij}}{\partial \nu}$, let us find the gradient for ν in the copula. Let the d dimensional random vector \mathbf{t} have a multivariate t -distribution with ν degrees of freedom and zero means and covariance matrix Σ . Then the likelihood of ν is:

$$L(\nu) = \frac{\Gamma\left(\frac{\nu+d}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\nu^{d/2}\pi^{d/2}\det(\Sigma)^{1/2}} \left[1 + \frac{1}{\nu}\mathbf{t}^T\Sigma^{-1}\mathbf{t}\right]^{-\frac{\nu+d}{2}}$$

Its log likelihood is:

$$\begin{aligned} l(\nu) &= \log \Gamma\left(\frac{\nu+d}{2}\right) - \Gamma\left(\frac{\nu}{2}\right) - \frac{d}{2}\log\nu - \frac{d}{2}\log\pi - \frac{1}{2}\log\det(\Sigma) \\ &\quad - \frac{\nu+d}{2}\log\left(1 + \frac{1}{\nu}\mathbf{t}^T\Sigma^{-1}\mathbf{t}\right) \end{aligned}$$

Let $F(\cdot)$ be the digamma function, which is the derivative of the logarithm of the gamma function. The first derivative of the log likelihood with respect to ν is:

$$\begin{aligned} \frac{\partial l(\nu)}{\partial \nu} &= \frac{1}{2}F\left(\frac{\nu+d}{2}\right) - \frac{1}{2}F\left(\frac{\nu}{2}\right) - \frac{d}{2\nu} - \frac{1}{2}\log\left(1 + \frac{1}{\nu}\mathbf{t}^T\Sigma^{-1}\mathbf{t}\right) + \frac{(\nu+d)\mathbf{t}^T\Sigma^{-1}\mathbf{t}}{2\nu(\nu + \mathbf{t}^T\Sigma^{-1}\mathbf{t})} \\ &\quad - \frac{(\nu+d)\mathbf{t}^T\Sigma^{-1}\frac{\partial \mathbf{t}}{\partial \nu}}{\nu + \mathbf{t}^T\Sigma^{-1}\mathbf{t}} \end{aligned}$$

Now let the random variable t have a univariate t -distribution with a mean of zero and ν degrees of freedom. Its likelihood with respect to ν is:

$$L(\nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\nu\pi}} \left(1 + \frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

Its log likelihood is

$$l(\nu) = \log\Gamma\left(\frac{\nu+1}{2}\right) - \log\Gamma\left(\frac{\nu}{2}\right) - \frac{1}{2}\log\nu - \frac{1}{2}\log\pi - \frac{\nu+1}{2}\log\left(1 + \frac{t^2}{\nu}\right)$$

Again, using F to denote the digamma function, the first derivative of its log likelihood with respect to ν is:

$$\begin{aligned}\frac{\partial l(\nu)}{\partial \nu} = & \frac{1}{2} F\left(\frac{\nu+1}{2}\right) - \frac{1}{2} F\left(\frac{\nu}{2}\right) - \frac{1}{2\nu} - \frac{1}{2} \log\left(1 + \frac{t^2}{\nu}\right) + \frac{(\nu+1)t^2}{2\nu(\nu+t^2)} \\ & - \frac{(\nu+1)t \frac{\partial t}{\partial \nu}}{\nu+t^2}\end{aligned}$$

Using the notation \mathbf{t}_i as defined in equation B.8, the copula density is given by:

$$c(t_1, \dots, t_d) = \left\{ \prod_{i=1}^T \frac{\Gamma\left(\frac{\nu+d}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right) \nu^{d/2} \pi^{d/2} \det(\boldsymbol{\Sigma})^{1/2}} \left[1 + \frac{1}{\nu} \mathbf{t}_i^T \boldsymbol{\Sigma}^{-1} \mathbf{t}_i\right]^{-\frac{\nu+d}{2}} \right\} \times \\ \left\{ \prod_{i=1}^T \prod_{j=1}^d \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right) \sqrt{\nu\pi}} \left(1 + \frac{t_{ij}^2}{\nu}\right)^{-\frac{\nu+1}{2}} \right\}^{-1}$$

This means the negative log-likelihood of the copula with respect to ν is:

$$\begin{aligned}-l(\nu) = & T d \log \Gamma\left(\frac{\nu+1}{2}\right) - T(d-1) \log \Gamma\left(\frac{\nu}{2}\right) - \frac{\nu+1}{2} \sum_{i=1}^T \sum_{j=1}^d \log \left(1 + \frac{t_{ij}^2}{\nu}\right) \\ & - T \log \Gamma\left(\frac{\nu+d}{2}\right) + \frac{T}{2} \log \det(\boldsymbol{\Sigma}) + \frac{\nu+d}{2} \sum_{i=1}^T \log \left(1 + \frac{1}{\nu} \mathbf{t}_i^T \boldsymbol{\Sigma}^{-1} \mathbf{t}_i\right)\end{aligned}$$

Taking the partial derivative of this gives us the gradient for ν :

$$\begin{aligned}-\frac{\partial l(\nu)}{\partial \nu} = & \frac{T d}{2} F\left(\frac{\nu+1}{2}\right) - \frac{T(d-1)}{2} F\left(\frac{\nu}{2}\right) - \frac{1}{2} \sum_{i=1}^T \sum_{j=1}^d \log \left(1 + \frac{t_{ij}^2}{\nu}\right) + \frac{\nu+1}{2\nu} \sum_{i=1}^T \sum_{j=1}^d \frac{t_{ij}^2}{\nu + t_{ij}^2} \\ & - \frac{T}{2} F\left(\frac{\nu+d}{2}\right) + \frac{1}{2} \sum_{i=1}^T \log \left(1 + \frac{1}{\nu} \mathbf{t}_i^T \boldsymbol{\Sigma}^{-1} \mathbf{t}_i\right) - \frac{\nu+d}{2\nu} \sum_{i=1}^T \frac{\mathbf{t}_i^T \boldsymbol{\Sigma}^{-1} \mathbf{t}_i}{\nu + \mathbf{t}_i^T \boldsymbol{\Sigma}^{-1} \mathbf{t}_i} \\ & - (\nu+1) \sum_{i=1}^T \sum_{j=1}^d \frac{t_{ij} \frac{\partial t_{ij}}{\partial \nu}}{\nu + t_{ij}^2} + (\nu+d) \sum_{i=1}^T \frac{\mathbf{t}_i^T \boldsymbol{\Sigma}^{-1} \frac{\partial \mathbf{t}_i}{\partial \nu}}{\nu + \mathbf{t}_i^T \boldsymbol{\Sigma}^{-1} \mathbf{t}_i}\end{aligned}$$

C: Results for normal & t -copulas

Table C.1: Mean difference/mean absolute difference for different statistics for the normal copula with 1 factor structure

ρ^-				
Sector 1	0.20 / 0.20	0.13 / 0.13	0.12 / 0.12	0.12 / 0.12
Sector 2	0.13 / 0.13	0.13 / 0.14	0.09 / 0.10	0.07 / 0.08
Sector 3	0.12 / 0.12	0.09 / 0.10	0.15 / 0.15	0.13 / 0.13
Sector 4	0.12 / 0.12	0.07 / 0.08	0.13 / 0.13	0.13 / 0.13
All sectors: 0.11 / 0.12				
ρ^+				
	Sector 1	Sector 2	Sector 3	Sector 4
Sector 1	0.06 / 0.08	0.02 / 0.06	0.00 / 0.05	0.00 / 0.05
Sector 2	0.02 / 0.06	0.10 / 0.10	-0.05 / 0.07	-0.02 / 0.06
Sector 3	0.00 / 0.05	-0.05 / 0.07	0.04 / 0.08	0.01 / 0.06
Sector 4	0.00 / 0.05	-0.02 / 0.06	0.01 / 0.06	0.05 / 0.08
All sectors: 0.01 / 0.06				
ρ_S				
	Sector 1	Sector 2	Sector 3	Sector 4
Sector 1	0.09 / 0.10	0.00 / 0.03	-0.02 / 0.05	-0.01 / 0.03
Sector 2	0.00 / 0.03	0.07 / 0.07	-0.03 / 0.04	-0.03 / 0.04
Sector 3	-0.02 / 0.05	-0.03 / 0.04	0.08 / 0.08	0.03 / 0.04
Sector 4	-0.01 / 0.03	-0.03 / 0.04	0.03 / 0.04	0.08 / 0.08
All sectors: 0.01 / 0.05				
$\Sigma - \hat{\Sigma}$				
	Sector 1	Sector 2	Sector 3	Sector 4
Sector 1	0.07 / 0.08	-0.02 / 0.03	-0.04 / 0.05	-0.03 / 0.04
Sector 2	-0.02 / 0.03	0.06 / 0.06	-0.05 / 0.06	-0.05 / 0.06
Sector 3	-0.04 / 0.05	-0.05 / 0.06	0.06 / 0.08	0.01 / 0.04
Sector 4	-0.03 / 0.04	-0.05 / 0.06	0.01 / 0.04	0.06 / 0.06
All sectors: -0.01 / 0.05				

Table C.2: Mean difference/mean absolute difference for different statistics for the t -copula with 1 factor structure, $\hat{\nu} = 20.8$

ρ^-				
	Sector 1	Sector 2	Sector 3	Sector 4
Sector 1	0.17 / 0.17	0.09 / 0.10	0.09 / 0.01	0.08 / 0.09
Sector 2	0.09 / 0.10	0.10 / 0.12	0.05 / 0.08	0.04 / 0.06
Sector 3	0.09 / 0.1	0.05 / 0.08	0.12 / 0.12	0.10 / 0.10
Sector 4	0.08 / 0.09	0.04 / 0.06	0.10 / 0.10	0.10 / 0.10
All sectors: 0.08 / 0.09				
ρ^+				
	Sector 1	Sector 2	Sector 3	Sector 4
Sector 1	0.02 / 0.06	-0.02 / 0.06	-0.04 / 0.06	-0.03 / 0.06
Sector 2	-0.02 / 0.06	0.06 / 0.07	-0.08 / 0.09	-0.06 / 0.08
Sector 3	-0.04 / 0.06	-0.08 / 0.09	0.01 / 0.07	-0.02 / 0.06
Sector 4	-0.03 / 0.06	-0.06 / 0.08	-0.02 / 0.06	0.02 / 0.06
All sectors: -0.02 / 0.06				
ρ_S				
	Sector 1	Sector 2	Sector 3	Sector 4
Sector 1	0.08 / 0.09	-0.01 / 0.03	-0.04 / 0.05	-0.02 / 0.04
Sector 2	-0.01 / 0.03	0.06 / 0.06	-0.05 / 0.06	-0.04 / 0.05
Sector 3	-0.04 / 0.05	-0.05 / 0.06	0.06 / 0.07	0.02 / 0.04
Sector 4	-0.02 / 0.04	-0.04 / 0.05	0.02 / 0.04	0.07 / 0.07
All sectors: 0.00 / 0.05				
$\Sigma - \hat{\Sigma}$				
	Sector 1	Sector 2	Sector 3	Sector 4
Sector 1	0.06 / 0.07	-0.04 / 0.05	-0.06 / 0.06	-0.04 / 0.05
Sector 2	-0.04 / 0.05	0.04 / 0.04	-0.07 / 0.07	-0.06 / 0.07
Sector 3	-0.06 / 0.06	-0.07 / 0.07	0.05 / 0.07	0.00 / 0.04
Sector 4	-0.04 / 0.05	-0.06 / 0.07	0.00 / 0.04	0.05 / 0.06
All sectors: -0.02 / 0.05				

Table C.3: Mean difference/mean absolute difference for different statistics for the normal copula with 2 factor structure

ρ^-				
	Sector 1	Sector 2	Sector 3	Sector 4
Sector 1	0.20 / 0.20	0.12 / 0.12	0.13 / 0.13	0.13 / 0.13
Sector 2	0.12 / 0.12	0.09 / 0.10	0.12 / 0.12	0.10 / 0.10
Sector 3	0.13 / 0.13	0.12 / 0.12	0.13 / 0.13	0.10 / 0.11
Sector 4	0.13 / 0.13	0.10 / 0.10	0.10 / 0.11	0.11 / 0.11
All sectors: 0.11 / 0.11				
ρ^+				
	Sector 1	Sector 2	Sector 3	Sector 4
Sector 1	0.06 / 0.08	0.01 / 0.06	0.01 / 0.05	0.01 / 0.05
Sector 2	0.01 / 0.06	0.05 / 0.06	-0.02 / 0.06	0.00 / 0.06
Sector 3	0.01 / 0.05	-0.02 / 0.06	0.02 / 0.06	-0.01 / 0.05
Sector 4	0.01 / 0.05	0.00 / 0.06	-0.01 / 0.05	0.03 / 0.06
All sectors: 0.01 / 0.05				
ρ_S				
	Sector 1	Sector 2	Sector 3	Sector 4
Sector 1	0.10 / 0.10	-0.01 / 0.03	-0.01 / 0.03	0.01 / 0.03
Sector 2	-0.01 / 0.03	0.01 / 0.03	0.01 / 0.03	0.01 / 0.03
Sector 3	-0.01 / 0.03	0.01 / 0.03	0.04 / 0.05	-0.02 / 0.03
Sector 4	0.01 / 0.03	0.01 / 0.03	-0.02 / 0.03	0.03 / 0.04
All sectors: 0.01 / 0.03				
$\Sigma - \hat{\Sigma}$				
	Sector 1	Sector 2	Sector 3	Sector 4
Sector 1	0.08 / 0.09	-0.03 / 0.04	-0.03 / 0.04	-0.01 / 0.03
Sector 2	-0.03 / 0.04	0.00 / 0.02	0.00 / 0.02	-0.01 / 0.02
Sector 3	-0.03 / 0.04	0.00 / 0.02	0.02 / 0.04	-0.03 / 0.04
Sector 4	-0.01 / 0.03	-0.01 / 0.02	-0.03 / 0.04	0.02 / 0.03
All sectors: -0.01 / 0.03				

Table C.4: Mean difference/mean absolute difference for different statistics for the t -copula with 2 factor structure, $\hat{\nu} = 22.5$

ρ^-				
	Sector 1	Sector 2	Sector 3	Sector 4
Sector 1	0.17 / 0.17	0.08 / 0.10	0.09 / 0.10	0.09 / 0.10
Sector 2	0.08 / 0.10	0.06 / 0.09	0.08 / 0.10	0.06 / 0.08
Sector 3	0.09 / 0.10	0.08 / 0.10	0.09 / 0.10	0.07 / 0.08
Sector 4	0.09 / 0.10	0.06 / 0.08	0.07 / 0.08	0.08 / 0.08
All sectors: 0.08 / 0.09				
ρ^+				
	Sector 1	Sector 2	Sector 3	Sector 4
Sector 1	0.03 / 0.06	-0.02 / 0.06	-0.03 / 0.06	-0.02 / 0.05
Sector 2	-0.02 / 0.06	0.03 / 0.05	-0.06 / 0.07	-0.03 / 0.06
Sector 3	-0.03 / 0.06	-0.06 / 0.07	-0.02 / 0.06	-0.05 / 0.07
Sector 4	-0.02 / 0.05	-0.03 / 0.06	-0.05 / 0.07	0.00 / 0.06
All sectors: -0.02 / 0.06				
ρ_S				
	Sector 1	Sector 2	Sector 3	Sector 4
Sector 1	0.09 / 0.09	-0.02 / 0.04	-0.02 / 0.04	0.00 / 0.03
Sector 2	-0.02 / 0.04	0.01 / 0.03	0.00 / 0.03	0.00 / 0.02
Sector 3	-0.02 / 0.04	0.00 / 0.03	0.03 / 0.04	-0.02 / 0.03
Sector 4	0.00 / 0.03	0.00 / 0.02	-0.02 / 0.03	0.03 / 0.04
All sectors: 0.00 / 0.03				
$\Sigma - \hat{\Sigma}$				
	Sector 1	Sector 2	Sector 3	Sector 4
Sector 1	0.07 / 0.07	-0.04 / 0.05	-0.04 / 0.05	-0.02 / 0.03
Sector 2	-0.04 / 0.05	-0.01 / 0.02	-0.02 / 0.03	-0.02 / 0.03
Sector 3	-0.04 / 0.05	-0.02 / 0.03	0.00 / 0.03	-0.05 / 0.05
Sector 4	-0.02 / 0.03	-0.02 / 0.03	-0.05 / 0.05	0.02 / 0.03
All sectors: -0.02 / 0.04				

Table C.5: Mean difference/mean absolute difference for different statistics for the normal copula with 3 factor structure

ρ^-				
	Sector 1	Sector 2	Sector 3	Sector 4
Sector 1	0.15 / 0.15	0.13 / 0.14	0.14 / 0.14	0.13 / 0.13
Sector 2	0.13 / 0.14	0.08 / 0.10	0.11 / 0.12	0.10 / 0.10
Sector 3	0.14 / 0.14	0.11 / 0.12	0.12 / 0.12	0.10 / 0.11
Sector 4	0.13 / 0.13	0.10 / 0.10	0.10 / 0.11	0.11 / 0.11
All sectors: 0.11 / 0.11				
ρ^+				
	Sector 1	Sector 2	Sector 3	Sector 4
Sector 1	0.01 / 0.06	0.03 / 0.06	0.02 / 0.06	0.01 / 0.05
Sector 2	0.03 / 0.06	0.05 / 0.06	-0.02 / 0.06	0.00 / 0.05
Sector 3	0.02 / 0.06	-0.02 / 0.06	0.01 / 0.06	-0.01 / 0.05
Sector 4	0.01 / 0.05	0.00 / 0.05	-0.01 / 0.05	0.03 / 0.06
All sectors: 0.01 / 0.05				
ρ_S				
	Sector 1	Sector 2	Sector 3	Sector 4
Sector 1	0.01 / 0.03	0.01 / 0.02	0.01 / 0.02	0.00 / 0.02
Sector 2	0.01 / 0.02	0.01 / 0.02	0.01 / 0.02	0.01 / 0.02
Sector 3	0.01 / 0.02	0.01 / 0.02	0.03 / 0.04	-0.01 / 0.03
Sector 4	0.00 / 0.02	0.01 / 0.02	-0.01 / 0.03	0.04 / 0.05
All sectors: 0.01 / 0.02				
$\Sigma - \hat{\Sigma}$				
	Sector 1	Sector 2	Sector 3	Sector 4
Sector 1	-0.01 / 0.03	-0.01 / 0.02	0.00 / 0.02	-0.02 / 0.02
Sector 2	-0.01 / 0.02	0.00 / 0.02	-0.01 / 0.02	-0.01 / 0.02
Sector 3	0.00 / 0.02	-0.01 / 0.02	0.01 / 0.03	-0.03 / 0.04
Sector 4	-0.02 / 0.02	-0.01 / 0.02	-0.03 / 0.04	0.03 / 0.03
All sectors: -0.01 / 0.02				

Table C.6: Mean difference/mean absolute difference for different statistics for the t -copula with 3 factor structure, $\hat{\nu} = 23.6$

ρ^-				
	Sector 1	Sector 2	Sector 3	Sector 4
Sector 1	0.12 / 0.12	0.10 / 0.11	0.11 / 0.11	0.09 / 0.10
Sector 2	0.10 / 0.11	0.06 / 0.08	0.08 / 0.10	0.07 / 0.08
Sector 3	0.11 / 0.11	0.08 / 0.10	0.09 / 0.09	0.07 / 0.08
Sector 4	0.09 / 0.10	0.07 / 0.08	0.07 / 0.08	0.08 / 0.09
All sectors: 0.08 / 0.09				
ρ^+				
	Sector 1	Sector 2	Sector 3	Sector 4
Sector 1	-0.02 / 0.07	0.00 / 0.05	-0.01 / 0.05	-0.02 / 0.05
Sector 2	0.00 / 0.05	0.03 / 0.05	-0.06 / 0.07	-0.03 / 0.06
Sector 3	-0.01 / 0.05	-0.06 / 0.07	-0.02 / 0.05	-0.05 / 0.07
Sector 4	-0.02 / 0.05	-0.03 / 0.06	-0.05 / 0.07	0.00 / 0.06
All sectors: -0.02 / 0.05				
ρ_S				
	Sector 1	Sector 2	Sector 3	Sector 4
Sector 1	0.00 / 0.03	0.00 / 0.02	0 / 0.02	-0.01 / 0.02
Sector 2	0.00 / 0.02	0.00 / 0.02	0.00 / 0.02	0.00 / 0.02
Sector 3	0.00 / 0.02	0.00 / 0.02	0.02 / 0.03	-0.02 / 0.03
Sector 4	-0.01 / 0.02	0.00 / 0.02	-0.02 / 0.03	0.04 / 0.04
All sectors: 0.00 / 0.02				
$\Sigma - \hat{\Sigma}$				
	Sector 1	Sector 2	Sector 3	Sector 4
Sector 1	-0.01 / 0.04	-0.02 / 0.03	-0.02 / 0.03	-0.02 / 0.03
Sector 2	-0.02 / 0.03	-0.01 / 0.02	-0.02 / 0.03	-0.02 / 0.03
Sector 3	-0.02 / 0.03	-0.02 / 0.03	0.00 / 0.03	-0.05 / 0.05
Sector 4	-0.02 / 0.03	-0.02 / 0.03	-0.05 / 0.05	0.02 / 0.03
All sectors: -0.02 / 0.03				

Table C.7: Mean difference/mean absolute difference for different statistics for the nested normal copula

ρ^-				
	Sector 1	Sector 2	Sector 3	Sector 4
Sector 1	0.15 / 0.15	0.16 / 0.16	0.13 / 0.14	0.13 / 0.13
Sector 2	0.16 / 0.16	0.08 / 0.10	0.12 / 0.12	0.09 / 0.10
Sector 3	0.13 / 0.14	0.12 / 0.12	0.10 / 0.11	0.13 / 0.13
Sector 4	0.13 / 0.13	0.09 / 0.10	0.13 / 0.13	0.09 / 0.09
All sectors: 0.11 / 0.12				
ρ^+				
	Sector 1	Sector 2	Sector 3	Sector 4
Sector 1	0.01 / 0.06	0.05 / 0.08	0.01 / 0.06	0.01 / 0.06
Sector 2	0.05 / 0.08	0.05 / 0.06	-0.02 / 0.07	0.00 / 0.06
Sector 3	0.01 / 0.06	-0.02 / 0.07	0.00 / 0.05	0.01 / 0.05
Sector 4	0.01 / 0.06	0.00 / 0.06	0.01 / 0.05	0.01 / 0.05
All sectors: 0.01 / 0.06				
ρ_S				
	Sector 1	Sector 2	Sector 3	Sector 4
Sector 1	0.01 / 0.04	0.05 / 0.06	-0.01 / 0.06	0.01 / 0.05
Sector 2	0.05 / 0.06	0.00 / 0.03	0.01 / 0.05	0.00 / 0.05
Sector 3	-0.01 / 0.06	0.01 / 0.05	0.00 / 0.03	0.03 / 0.04
Sector 4	0.01 / 0.05	0.00 / 0.05	0.03 / 0.04	0.01 / 0.03
All sectors: 0.01 / 0.04				
$\Sigma - \hat{\Sigma}$				
	Sector 1	Sector 2	Sector 3	Sector 4
Sector 1	-0.01 / 0.04	0.03 / 0.05	-0.02 / 0.07	-0.01 / 0.06
Sector 2	0.03 / 0.05	-0.01 / 0.03	-0.01 / 0.05	-0.01 / 0.05
Sector 3	-0.02 / 0.07	-0.01 / 0.05	-0.01 / 0.03	0.01 / 0.03
Sector 4	-0.01 / 0.06	-0.01 / 0.05	0.01 / 0.03	-0.01 / 0.03
All sectors: 0.00 / 0.04				

Table C.8: Mean difference/mean absolute difference for different statistics for the bifactor normal copula

ρ^-				
	Sector 1	Sector 2	Sector 3	Sector 4
Sector 1	0.15 / 0.15	0.14 / 0.15	0.13 / 0.13	0.13 / 0.13
Sector 2	0.14 / 0.15	0.08 / 0.09	0.11 / 0.11	0.08 / 0.09
Sector 3	0.13 / 0.13	0.11 / 0.11	0.10 / 0.11	0.14 / 0.14
Sector 4	0.13 / 0.13	0.08 / 0.09	0.14 / 0.14	0.09 / 0.09
All sectors: 0.11 / 0.11				
ρ^+				
	Sector 1	Sector 2	Sector 3	Sector 4
Sector 1	0.01 / 0.06	0.04 / 0.07	0.01 / 0.05	0.01 / 0.05
Sector 2	0.04 / 0.07	0.05 / 0.06	-0.02 / 0.06	-0.01 / 0.06
Sector 3	0.01 / 0.05	-0.02 / 0.06	0.00 / 0.04	0.02 / 0.06
Sector 4	0.01 / 0.05	-0.01 / 0.06	0.02 / 0.06	0.01 / 0.05
All sectors: 0.01 / 0.05				
ρ_S				
	Sector 1	Sector 2	Sector 3	Sector 4
Sector 1	0.01 / 0.03	0.02 / 0.03	-0.01 / 0.03	0.00 / 0.03
Sector 2	0.02 / 0.03	0.01 / 0.02	0.00 / 0.03	-0.01 / 0.03
Sector 3	-0.01 / 0.03	0.00 / 0.03	0.01 / 0.03	0.05 / 0.06
Sector 4	0.00 / 0.03	-0.01 / 0.03	0.05 / 0.06	0.01 / 0.03
All sectors: 0.01 / 0.03				
$\Sigma - \hat{\Sigma}$				
	Sector 1	Sector 2	Sector 3	Sector 4
Sector 1	-0.01 / 0.03	0.00 / 0.02	-0.03 / 0.04	-0.02 / 0.03
Sector 2	0.00 / 0.02	0.00 / 0.02	-0.02 / 0.03	-0.03 / 0.04
Sector 3	-0.03 / 0.04	-0.02 / 0.03	-0.01 / 0.02	0.03 / 0.05
Sector 4	-0.02 / 0.03	-0.03 / 0.04	0.03 / 0.05	-0.01 / 0.03
All sectors: -0.01 / 0.03				

Table C.9: Mean difference/mean absolute difference for different statistics for the nested t -copula, $\hat{\nu} = 23.0$

ρ^-				
	Sector 1	Sector 2	Sector 3	Sector 4
Sector 1	0.12 / 0.12	0.13 / 0.14	0.10 / 0.11	0.09 / 0.10
Sector 2	0.13 / 0.14	0.05 / 0.08	0.08 / 0.10	0.06 / 0.07
Sector 3	0.10 / 0.11	0.08 / 0.10	0.08 / 0.09	0.10 / 0.10
Sector 4	0.09 / 0.10	0.06 / 0.07	0.10 / 0.10	0.06 / 0.07
All sectors: 0.08 / 0.09				
ρ^+				
	Sector 1	Sector 2	Sector 3	Sector 4
Sector 1	-0.02 / 0.06	0.02 / 0.06	-0.02 / 0.06	-0.02 / 0.06
Sector 2	0.02 / 0.06	0.03 / 0.05	-0.05 / 0.08	-0.04 / 0.07
Sector 3	-0.02 / 0.06	-0.05 / 0.08	-0.03 / 0.06	-0.02 / 0.05
Sector 4	-0.02 / 0.06	-0.04 / 0.07	-0.02 / 0.05	-0.02 / 0.06
All sectors: -0.02 / 0.06				
ρ_S				
	Sector 1	Sector 2	Sector 3	Sector 4
Sector 1	0.01 / 0.04	0.04 / 0.05	-0.02 / 0.06	0.00 / 0.05
Sector 2	0.04 / 0.05	0.00 / 0.03	0.00 / 0.05	-0.01 / 0.05
Sector 3	-0.02 / 0.06	0.00 / 0.05	-0.01 / 0.03	0.02 / 0.04
Sector 4	0.00 / 0.05	-0.01 / 0.05	0.02 / 0.04	0.00 / 0.03
All sectors: 0.00 / 0.04				
$\Sigma - \hat{\Sigma}$				
	Sector 1	Sector 2	Sector 3	Sector 4
Sector 1	-0.01 / 0.04	0.02 / 0.05	-0.04 / 0.07	-0.03 / 0.06
Sector 2	0.02 / 0.05	-0.01 / 0.03	-0.02 / 0.06	-0.03 / 0.05
Sector 3	-0.04 / 0.07	-0.02 / 0.06	-0.02 / 0.04	0.00 / 0.03
Sector 4	-0.03 / 0.06	-0.03 / 0.05	0.00 / 0.03	-0.01 / 0.03
All sectors: -0.01 / 0.05				

Table C.10: Mean difference/mean absolute difference for different statistics for the bifactor t -copula, $\hat{\nu} = 23.4$

ρ^-				
	Sector 1	Sector 2	Sector 3	Sector 4
Sector 1	0.12 / 0.12	0.11 / 0.12	0.10 / 0.11	0.09 / 0.10
Sector 2	0.11 / 0.12	0.05 / 0.08	0.07 / 0.09	0.05 / 0.07
Sector 3	0.10 / 0.11	0.07 / 0.09	0.07 / 0.08	0.11 / 0.11
Sector 4	0.09 / 0.10	0.05 / 0.07	0.11 / 0.11	0.06 / 0.07
All sectors: 0.08 / 0.09				
ρ^+				
	Sector 1	Sector 2	Sector 3	Sector 4
Sector 1	-0.02 / 0.07	0.01 / 0.06	-0.02 / 0.05	-0.02 / 0.06
Sector 2	0.01 / 0.06	0.03 / 0.04	-0.06 / 0.08	-0.04 / 0.07
Sector 3	-0.02 / 0.05	-0.06 / 0.08	-0.03 / 0.05	-0.01 / 0.06
Sector 4	-0.02 / 0.06	-0.04 / 0.07	-0.01 / 0.06	-0.02 / 0.05
All sectors: -0.02 / 0.05				
ρ_S				
	Sector 1	Sector 2	Sector 3	Sector 4
Sector 1	0.00 / 0.03	0.01 / 0.03	-0.02 / 0.04	-0.01 / 0.03
Sector 2	0.01 / 0.03	0.00 / 0.02	-0.01 / 0.03	-0.02 / 0.03
Sector 3	-0.02 / 0.04	-0.01 / 0.03	0.00 / 0.03	0.03 / 0.05
Sector 4	-0.01 / 0.03	-0.02 / 0.03	0.03 / 0.05	0.00 / 0.03
All sectors: 0.00 / 0.03				
$\Sigma - \hat{\Sigma}$				
	Sector 1	Sector 2	Sector 3	Sector 4
Sector 1	-0.02 / 0.04	-0.01 / 0.03	-0.04 / 0.05	-0.03 / 0.04
Sector 2	-0.01 / 0.03	-0.01 / 0.02	-0.03 / 0.04	-0.04 / 0.04
Sector 3	-0.04 / 0.05	-0.03 / 0.04	-0.02 / 0.03	0.01 / 0.04
Sector 4	-0.03 / 0.04	-0.04 / 0.04	0.01 / 0.04	-0.01 / 0.03
All sectors: -0.02 / 0.03				

D: Results for Chi-squared & F -copulas

Note: for the F -copulas, we compare the differences in theoretical Pearson correlation \mathbf{C} to sample Pearson correlation, $\hat{\mathbf{C}}$. The formula to calculate theoretical Pearson correlation from the scale matrix Σ , for $\nu > 4$, is:

$$\mathbf{C} = \frac{(\nu - 1)\Sigma^2 + 1}{\nu - 2}$$

Table D.1: Mean difference/mean absolute difference for different statistics for the 1 factor chi-squared copula, fitted with pairwise likelihood

ρ^-				
Sector 1	0.20 / 0.20	0.20 / 0.21	0.17 / 0.17	0.16 / 0.16
Sector 2	0.20 / 0.21	0.27 / 0.27	0.14 / 0.16	0.15 / 0.15
Sector 3	0.17 / 0.17	0.14 / 0.16	0.20 / 0.20	0.17 / 0.18
Sector 4	0.16 / 0.16	0.15 / 0.15	0.17 / 0.18	0.18 / 0.18
All sectors: 0.16 / 0.17				
ρ^+				
	Sector 1	Sector 2	Sector 3	Sector 4
Sector 1	0.01 / 0.05	-0.09 / 0.10	-0.10 / 0.11	-0.12 / 0.12
Sector 2	-0.09 / 0.10	-0.09 / 0.10	-0.14 / 0.14	-0.17 / 0.18
Sector 3	-0.10 / 0.11	-0.14 / 0.14	-0.05 / 0.06	-0.11 / 0.11
Sector 4	-0.12 / 0.12	-0.17 / 0.18	-0.11 / 0.11	-0.09 / 0.10
All sectors: -0.10 / 0.11				
ρ_S				
	Sector 1	Sector 2	Sector 3	Sector 4
Sector 1	0.07 / 0.08	-0.04 / 0.06	-0.05 / 0.06	-0.06 / 0.07
Sector 2	-0.04 / 0.06	0.01 / 0.04	-0.08 / 0.08	-0.10 / 0.10
Sector 3	-0.05 / 0.06	-0.08 / 0.08	0.05 / 0.07	-0.02 / 0.04
Sector 4	-0.06 / 0.07	-0.10 / 0.10	-0.02 / 0.04	0.01 / 0.05
All sectors: -0.03 / 0.06				
$\Sigma^2 - \hat{\Sigma}^2$				
	Sector 1	Sector 2	Sector 3	Sector 4
Sector 1	-0.10 / 0.11	-0.23 / 0.23	-0.25 / 0.25	-0.26 / 0.26
Sector 2	-0.23 / 0.23	-0.15 / 0.15	-0.25 / 0.25	-0.28 / 0.28
Sector 3	-0.25 / 0.25	-0.25 / 0.25	-0.13 / 0.13	-0.22 / 0.22
Sector 4	-0.26 / 0.26	-0.28 / 0.28	-0.22 / 0.22	-0.17 / 0.17
All sectors: -0.20 / 0.20				

Table D.2: Mean difference/mean absolute difference for different statistics for the 2 factor chi-squared copula, fitted with pairwise likelihood

ρ^-				
Sector 1	0.21 / 0.21	0.21 / 0.22	0.18 / 0.18	0.17 / 0.17
Sector 2	0.21 / 0.22	0.28 / 0.28	0.17 / 0.18	0.17 / 0.17
Sector 3	0.18 / 0.18	0.17 / 0.18	0.20 / 0.20	0.18 / 0.18
Sector 4	0.17 / 0.17	0.17 / 0.17	0.18 / 0.18	0.19 / 0.19
All sectors: 0.18 / 0.18				
ρ^+				
	Sector 1	Sector 2	Sector 3	Sector 4
Sector 1	0.06 / 0.07	-0.02 / 0.06	-0.04 / 0.07	-0.04 / 0.06
Sector 2	-0.02 / 0.06	-0.02 / 0.05	-0.02 / 0.06	-0.04 / 0.07
Sector 3	-0.04 / 0.07	-0.02 / 0.06	-0.03 / 0.05	-0.06 / 0.07
Sector 4	-0.04 / 0.06	-0.04 / 0.07	-0.06 / 0.07	-0.03 / 0.05
All sectors: -0.03 / 0.06				
ρ_S				
	Sector 1	Sector 2	Sector 3	Sector 4
Sector 1	0.14 / 0.14	0.05 / 0.06	0.03 / 0.05	0.05 / 0.06
Sector 2	0.05 / 0.06	0.07 / 0.07	0.06 / 0.07	0.07 / 0.07
Sector 3	0.03 / 0.05	0.06 / 0.07	0.09 / 0.09	0.05 / 0.05
Sector 4	0.05 / 0.06	0.07 / 0.07	0.05 / 0.05	0.10 / 0.10
All sectors: 0.06 / 0.07				
$\Sigma^2 - \hat{\Sigma}^2$				
	Sector 1	Sector 2	Sector 3	Sector 4
Sector 1	0.00 / 0.06	-0.11 / 0.12	-0.14 / 0.14	-0.13 / 0.13
Sector 2	-0.11 / 0.12	-0.08 / 0.08	-0.09 / 0.10	-0.09 / 0.10
Sector 3	-0.14 / 0.14	-0.09 / 0.10	-0.07 / 0.08	-0.12 / 0.12
Sector 4	-0.13 / 0.13	-0.09 / 0.10	-0.12 / 0.12	-0.06 / 0.07
All sectors: -0.09 / 0.10				

Table D.3: Mean difference/mean absolute difference for different statistics for the 3 factor chi-squared copula, fitted with pairwise likelihood

ρ^-				
Sector 1	0.21 / 0.21	0.21 / 0.22	0.18 / 0.18	0.17 / 0.17
Sector 2	0.21 / 0.22	0.28 / 0.28	0.16 / 0.17	0.17 / 0.17
Sector 3	0.18 / 0.18	0.16 / 0.17	0.2 / 0.2	0.18 / 0.18
Sector 4	0.17 / 0.17	0.17 / 0.17	0.18 / 0.18	0.19 / 0.19
All sectors: 0.17 / 0.18				
ρ^+				
	Sector 1	Sector 2	Sector 3	Sector 4
Sector 1	0.01 / 0.05	0.00 / 0.06	0.01 / 0.05	-0.01 / 0.06
Sector 2	0.00 / 0.06	-0.02 / 0.04	-0.02 / 0.06	-0.02 / 0.05
Sector 3	0.01 / 0.05	-0.02 / 0.06	-0.01 / 0.04	-0.03 / 0.06
Sector 4	-0.01 / 0.06	-0.02 / 0.05	-0.03 / 0.06	-0.01 / 0.04
All sectors: -0.01 / 0.05				
ρ_S				
	Sector 1	Sector 2	Sector 3	Sector 4
Sector 1	0.08 / 0.08	0.07 / 0.07	0.08 / 0.08	0.07 / 0.07
Sector 2	0.07 / 0.07	0.07 / 0.07	0.06 / 0.06	0.08 / 0.08
Sector 3	0.08 / 0.08	0.06 / 0.06	0.09 / 0.09	0.05 / 0.06
Sector 4	0.07 / 0.07	0.08 / 0.08	0.05 / 0.06	0.09 / 0.09
All sectors: 0.07 / 0.07				
$\Sigma^2 - \hat{\Sigma}^2$				
	Sector 1	Sector 2	Sector 3	Sector 4
Sector 1	-0.09 / 0.10	-0.10 / 0.10	-0.09 / 0.09	-0.12 / 0.12
Sector 2	-0.10 / 0.10	-0.08 / 0.08	-0.11 / 0.11	-0.08 / 0.09
Sector 3	-0.09 / 0.09	-0.11 / 0.11	-0.09 / 0.09	-0.13 / 0.13
Sector 4	-0.12 / 0.12	-0.08 / 0.09	-0.13 / 0.13	-0.08 / 0.08
All sectors: -0.09 / 0.09				

Table D.4: Mean difference/mean absolute difference for different statistics for the 1 factor chi-squared copula, fitted with Kendall's Tau

ρ^-				
Sector 1	0.20 / 0.20	0.20 / 0.21	0.17 / 0.17	0.16 / 0.16
Sector 2	0.20 / 0.21	0.29 / 0.29	0.15 / 0.16	0.16 / 0.16
Sector 3	0.17 / 0.17	0.15 / 0.16	0.20 / 0.20	0.18 / 0.18
Sector 4	0.16 / 0.16	0.16 / 0.16	0.18 / 0.18	0.18 / 0.18
All sectors: 0.17 / 0.17				
ρ^+				
	Sector 1	Sector 2	Sector 3	Sector 4
Sector 1	0.01 / 0.05	-0.07 / 0.08	-0.09 / 0.10	-0.10 / 0.11
Sector 2	-0.07 / 0.08	-0.03 / 0.08	-0.10 / 0.11	-0.13 / 0.13
Sector 3	-0.09 / 0.10	-0.10 / 0.11	-0.04 / 0.06	-0.08 / 0.09
Sector 4	-0.10 / 0.11	-0.13 / 0.13	-0.08 / 0.09	-0.06 / 0.07
All sectors: -0.07 / 0.09				
ρ_S				
	Sector 1	Sector 2	Sector 3	Sector 4
Sector 1	0.07 / 0.08	-0.02 / 0.04	-0.05 / 0.05	-0.03 / 0.04
Sector 2	-0.02 / 0.04	0.05 / 0.06	-0.05 / 0.06	-0.05 / 0.06
Sector 3	-0.05 / 0.05	-0.05 / 0.06	0.06 / 0.07	0.00 / 0.03
Sector 4	-0.03 / 0.04	-0.05 / 0.06	0.00 / 0.03	0.05 / 0.06
All sectors: -0.01 / 0.05				
$\Sigma^2 - \hat{\Sigma}^2$				
	Sector 1	Sector 2	Sector 3	Sector 4
Sector 1	-0.09 / 0.10	-0.20 / 0.20	-0.23 / 0.23	-0.23 / 0.23
Sector 2	-0.20 / 0.20	-0.11 / 0.11	-0.22 / 0.22	-0.23 / 0.23
Sector 3	-0.23 / 0.23	-0.22 / 0.22	-0.11 / 0.12	-0.18 / 0.18
Sector 4	-0.23 / 0.23	-0.23 / 0.23	-0.18 / 0.18	-0.12 / 0.12
All sectors: -0.17 / 0.18				

Table D.5: Mean difference/mean absolute difference for different statistics for the 2 factor chi-squared copula, fitted with Kendall's Tau

ρ^-				
Sector 1	0.20 / 0.20	0.20 / 0.20	0.17 / 0.17	0.17 / 0.17
Sector 2	0.20 / 0.20	0.28 / 0.28	0.15 / 0.16	0.16 / 0.16
Sector 3	0.17 / 0.17	0.15 / 0.16	0.19 / 0.19	0.17 / 0.17
Sector 4	0.17 / 0.17	0.16 / 0.16	0.17 / 0.17	0.18 / 0.19
All sectors: 0.17 / 0.17				
ρ^+				
	Sector 1	Sector 2	Sector 3	Sector 4
Sector 1	-0.03 / 0.06	-0.10 / 0.11	-0.07 / 0.08	-0.08 / 0.09
Sector 2	-0.10 / 0.11	-0.06 / 0.09	-0.09 / 0.10	-0.12 / 0.12
Sector 3	-0.07 / 0.08	-0.09 / 0.10	-0.08 / 0.08	-0.11 / 0.12
Sector 4	-0.08 / 0.09	-0.12 / 0.12	-0.11 / 0.12	-0.08 / 0.08
All sectors: -0.08 / 0.09				
ρ_S				
	Sector 1	Sector 2	Sector 3	Sector 4
Sector 1	0.03 / 0.04	-0.06 / 0.06	-0.01 / 0.03	0.00 / 0.02
Sector 2	-0.06 / 0.06	0.03 / 0.04	-0.03 / 0.03	-0.03 / 0.04
Sector 3	-0.01 / 0.03	-0.03 / 0.03	0.01 / 0.03	-0.04 / 0.05
Sector 4	0.00 / 0.02	-0.03 / 0.04	-0.04 / 0.05	0.02 / 0.03
All sectors: -0.01 / 0.03				
$\Sigma^2 - \hat{\Sigma}^2$				
	Sector 1	Sector 2	Sector 3	Sector 4
Sector 1	-0.14 / 0.14	-0.24 / 0.24	-0.18 / 0.18	-0.19 / 0.19
Sector 2	-0.24 / 0.24	-0.13 / 0.13	-0.19 / 0.19	-0.20 / 0.21
Sector 3	-0.18 / 0.18	-0.19 / 0.19	-0.16 / 0.17	-0.22 / 0.22
Sector 4	-0.19 / 0.19	-0.20 / 0.21	-0.22 / 0.22	-0.14 / 0.14
All sectors: -0.17 / 0.18				

Table D.6: Mean difference/mean absolute difference for different statistics for the 3 factor chi-squared copula, fitted with Kendall's Tau

ρ^-				
Sector 1	0.19 / 0.19	0.21 / 0.21	0.18 / 0.18	0.16 / 0.16
Sector 2	0.21 / 0.21	0.26 / 0.26	0.16 / 0.16	0.16 / 0.16
Sector 3	0.18 / 0.18	0.16 / 0.16	0.19 / 0.19	0.17 / 0.17
Sector 4	0.16 / 0.16	0.16 / 0.16	0.17 / 0.17	0.18 / 0.18
All sectors: 0.17 / 0.17				
ρ^+				
Sector 1	-0.05 / 0.07	-0.06 / 0.08	-0.05 / 0.07	-0.08 / 0.10
Sector 2	-0.06 / 0.08	-0.08 / 0.09	-0.06 / 0.08	-0.09 / 0.09
Sector 3	-0.05 / 0.07	-0.06 / 0.08	-0.07 / 0.08	-0.11 / 0.11
Sector 4	-0.08 / 0.10	-0.09 / 0.09	-0.11 / 0.11	-0.08 / 0.08
All sectors: -0.07 / 0.08				
ρ_S				
	Sector 1	Sector 2	Sector 3	Sector 4
Sector 1	0.00 / 0.04	-0.01 / 0.02	0 .00/ 0.02	-0.01 / 0.02
Sector 2	-0.01 / 0.02	0.00 / 0.02	-0.01 / 0.02	-0.01 / 0.02
Sector 3	0.00 / 0.02	-0.01 / 0.02	0.01 / 0.03	-0.03 / 0.04
Sector 4	-0.01 / 0.02	-0.01 / 0.02	-0.03 / 0.04	0.03 / 0.03
All sectors: 0.00 / 0.02				
$\Sigma^2 - \hat{\Sigma}^2$				
	Sector 1	Sector 2	Sector 3	Sector 4
Sector 1	-0.17 / 0.17	-0.19 / 0.19	-0.18 / 0.18	-0.21 / 0.21
Sector 2	-0.19 / 0.19	-0.16 / 0.16	-0.18 / 0.18	-0.18 / 0.19
Sector 3	-0.18 / 0.18	-0.18 / 0.18	-0.17 / 0.17	-0.22 / 0.22
Sector 4	-0.21 / 0.21	-0.18 / 0.19	-0.22 / 0.22	-0.15 / 0.15
All sectors: -0.17 / 0.17				

Table D.7: Mean difference/mean absolute difference for different statistics for the 1 factor chi-squared copula, fitted with Minimum Distance Estimators

ρ^-				
Sector 1	0.21 / 0.21	0.22 / 0.22	0.19 / 0.19	0.17 / 0.17
Sector 2	0.22 / 0.22	0.31 / 0.31	0.17 / 0.18	0.17 / 0.18
Sector 3	0.19 / 0.19	0.17 / 0.18	0.21 / 0.21	0.19 / 0.20
Sector 4	0.17 / 0.17	0.17 / 0.18	0.19 / 0.20	0.20 / 0.20
All sectors: 0.18 / 0.18				
ρ^+				
	Sector 1	Sector 2	Sector 3	Sector 4
Sector 1	0.13 / 0.13	0.08 / 0.09	0.05 / 0.08	0.04 / 0.06
Sector 2	0.08 / 0.09	0.10 / 0.11	0.04 / 0.07	0.02 / 0.06
Sector 3	0.05 / 0.08	0.04 / 0.07	0.09 / 0.09	0.05 / 0.07
Sector 4	0.04 / 0.06	0.02 / 0.06	0.05 / 0.07	0.06 / 0.07
All sectors: 0.05 / 0.07				
ρ_S				
	Sector 1	Sector 2	Sector 3	Sector 4
Sector 1	0.23 / 0.23	0.17 / 0.17	0.13 / 0.13	0.15 / 0.15
Sector 2	0.17 / 0.17	0.23 / 0.23	0.13 / 0.14	0.13 / 0.14
Sector 3	0.13 / 0.13	0.13 / 0.14	0.22 / 0.22	0.19 / 0.19
Sector 4	0.15 / 0.15	0.13 / 0.14	0.19 / 0.19	0.21 / 0.20
All sectors: 0.16 / 0.16				
$\Sigma^2 - \hat{\Sigma}^2$				
	Sector 1	Sector 2	Sector 3	Sector 4
Sector 1	0.09 / 0.09	-0.01 / 0.03	-0.04 / 0.05	-0.03 / 0.04
Sector 2	-0.01 / 0.03	0.06 / 0.07	-0.03 / 0.05	-0.04 / 0.05
Sector 3	-0.04 / 0.05	-0.03 / 0.05	0.06 / 0.07	0.02 / 0.04
Sector 4	-0.03 / 0.04	-0.04 / 0.05	0.02 / 0.04	0.06 / 0.06
All sectors: 0.00 / 0.05				

Table D.8: Mean difference/mean absolute difference for different statistics for the 2 factor chi-squared copula, fitted with Minimum Distance Estimators

ρ^-				
Sector 1	0.21 / 0.21	0.22 / 0.22	0.18 / 0.19	0.17 / 0.17
Sector 2	0.22 / 0.22	0.30 / 0.30	0.17 / 0.17	0.17 / 0.17
Sector 3	0.18 / 0.19	0.17 / 0.17	0.22 / 0.22	0.19 / 0.19
Sector 4	0.17 / 0.17	0.17 / 0.17	0.19 / 0.19	0.20 / 0.20
All sectors: 0.18 / 0.18				
ρ^+				
	Sector 1	Sector 2	Sector 3	Sector 4
Sector 1	0.08 / 0.09	0.03 / 0.06	0.06 / 0.08	0.05 / 0.06
Sector 2	0.03 / 0.06	0.07 / 0.09	0.04 / 0.07	0.01 / 0.06
Sector 3	0.06 / 0.08	0.04 / 0.07	0.05 / 0.06	0.01 / 0.06
Sector 4	0.05 / 0.06	0.01 / 0.06	0.01 / 0.06	0.03 / 0.05
All sectors: 0.04 / 0.06				
ρ_S				
	Sector 1	Sector 2	Sector 3	Sector 4
Sector 1	0.18 / 0.18	0.12 / 0.12	0.15 / 0.15	0.17 / 0.17
Sector 2	0.12 / 0.12	0.2 / 0.2	0.14 / 0.14	0.14 / 0.15
Sector 3	0.15 / 0.15	0.14 / 0.14	0.18 / 0.18	0.14 / 0.14
Sector 4	0.17 / 0.17	0.14 / 0.15	0.14 / 0.14	0.17 / 0.17
All sectors: 0.14 / 0.14				
$\Sigma^2 - \hat{\Sigma}^2$				
	Sector 1	Sector 2	Sector 3	Sector 4
Sector 1	0.03 / 0.04	-0.04 / 0.05	0.00 / 0.02	0.01 / 0.02
Sector 2	-0.04 / 0.05	0.05 / 0.05	-0.01 / 0.03	-0.01 / 0.03
Sector 3	0.00 / 0.02	-0.01 / 0.03	0.02 / 0.03	-0.02 / 0.03
Sector 4	0.01 / 0.02	-0.01 / 0.03	-0.02 / 0.03	0.03 / 0.03
All sectors: 0.03 / 0.03				

Table D.9: Mean difference/mean absolute difference for different statistics for the 3 factor chi-squared copula, fitted with Minimum Distance Estimators

ρ^-				
Sector 1	0.21 / 0.21	0.22 / 0.22	0.19 / 0.19	0.17 / 0.18
Sector 2	0.22 / 0.22	0.30 / 0.30	0.17 / 0.18	0.17 / 0.17
Sector 3	0.19 / 0.19	0.17 / 0.18	0.21 / 0.21	0.19 / 0.19
Sector 4	0.17 / 0.18	0.17 / 0.17	0.19 / 0.19	0.19 / 0.20
All sectors: 0.18 / 0.18				
ρ^+				
	Sector 1	Sector 2	Sector 3	Sector 4
Sector 1	0.07 / 0.09	0.06 / 0.09	0.05 / 0.07	0.04 / 0.06
Sector 2	0.06 / 0.09	0.04 / 0.06	0.04 / 0.07	0.02 / 0.06
Sector 3	0.05 / 0.07	0.04 / 0.07	0.04 / 0.06	0.01 / 0.06
Sector 4	0.04 / 0.06	0.02 / 0.06	0.01 / 0.06	0.02 / 0.05
All sectors: 0.04 / 0.06				
ρ_S				
	Sector 1	Sector 2	Sector 3	Sector 4
Sector 1	0.15 / 0.15	0.16 / 0.16	0.16 / 0.16	0.16 / 0.16
Sector 2	0.16 / 0.16	0.15 / 0.15	0.15 / 0.15	0.15 / 0.15
Sector 3	0.16 / 0.16	0.15 / 0.15	0.18 / 0.18	0.14 / 0.14
Sector 4	0.16 / 0.16	0.15 / 0.15	0.14 / 0.14	0.17 / 0.17
All sectors: 0.14 / 0.14				
$\Sigma^2 - \hat{\Sigma}^2$				
	Sector 1	Sector 2	Sector 3	Sector 4
Sector 1	0.00 / 0.05	0.00 / 0.02	0.00 / 0.02	0.00 / 0.02
Sector 2	0.00 / 0.02	0.00 / 0.02	0.00 / 0.02	0.00 / 0.02
Sector 3	0.00 / 0.02	0.00 / 0.02	0.02 / 0.03	-0.02 / 0.03
Sector 4	0.00 / 0.02	0.00 / 0.02	-0.02 / 0.03	0.02 / 0.03
All sectors: 0.00 / 0.02				

Table D.10: Mean difference/mean absolute difference for different statistics for the 1 factor $F_{1,\nu}$ - copula, $\hat{\nu} = 7.0$, fitted with pairwise likelihood

ρ^-				
Sector 1	0.19 / 0.19	0.20 / 0.21	0.17 / 0.17	0.16 / 0.16
Sector 2	0.20 / 0.21	0.28 / 0.28	0.15 / 0.15	0.15 / 0.16
Sector 3	0.17 / 0.17	0.15 / 0.15	0.19 / 0.19	0.17 / 0.18
Sector 4	0.16 / 0.16	0.15 / 0.16	0.17 / 0.18	0.19 / 0.19
All sectors: 0.17 / 0.17				
ρ^+				
	Sector 1	Sector 2	Sector 3	Sector 4
Sector 1	0.02 / 0.06	-0.07 / 0.08	-0.08 / 0.09	-0.09 / 0.10
Sector 2	-0.07 / 0.08	-0.04 / 0.08	-0.10 / 0.10	-0.13 / 0.13
Sector 3	-0.08 / 0.09	-0.10 / 0.10	-0.03 / 0.06	-0.07 / 0.08
Sector 4	-0.09 / 0.10	-0.13 / 0.13	-0.07 / 0.08	-0.05 / 0.07
All sectors: -0.07 / 0.08				
ρ_S				
	Sector 1	Sector 2	Sector 3	Sector 4
Sector 1	0.13 / 0.13	0.04 / 0.05	0.02 / 0.04	0.03 / 0.05
Sector 2	0.04 / 0.05	0.10 / 0.11	0.01 / 0.04	0.01 / 0.04
Sector 3	0.02 / 0.04	0.01 / 0.04	0.12 / 0.12	0.08 / 0.08
Sector 4	0.03 / 0.05	0.01 / 0.04	0.08 / 0.08	0.11 / 0.11
All sectors: 0.05 / 0.06				
$C - \hat{C}$				
	Sector 1	Sector 2	Sector 3	Sector 4
Sector 1	-0.13 / 0.15	-0.25 / 0.25	-0.29 / 0.29	-0.27 / 0.27
Sector 2	-0.25 / 0.25	-0.15 / 0.16	-0.29 / 0.29	-0.26 / 0.26
Sector 3	-0.29 / 0.29	-0.29 / 0.29	-0.19 / 0.19	-0.23 / 0.23
Sector 4	-0.27 / 0.27	-0.26 / 0.26	-0.23 / 0.23	-0.15 / 0.15
All sectors: -0.22 / 0.22				

Table D.11: Mean difference/mean absolute difference for different statistics for the 2 factor $F_{1,\nu}$ - copula, $\hat{\nu} = 7.0$, fitted with pairwise likelihood

ρ^-				
Sector 1	0.20 / 0.20	0.21 / 0.21	0.17 / 0.18	0.16 / 0.16
Sector 2	0.21 / 0.21	0.27 / 0.27	0.15 / 0.16	0.16 / 0.16
Sector 3	0.17 / 0.18	0.15 / 0.16	0.2 / 0.2	0.18 / 0.18
Sector 4	0.16 / 0.16	0.16 / 0.16	0.18 / 0.18	0.19 / 0.19
All sectors: 0.17 / 0.17				
ρ^+				
	Sector 1	Sector 2	Sector 3	Sector 4
Sector 1	-0.03 / 0.06	-0.04 / 0.07	-0.09 / 0.10	-0.10 / 0.10
Sector 2	-0.04 / 0.07	-0.07 / 0.08	-0.10 / 0.11	-0.12 / 0.12
Sector 3	-0.09 / 0.10	-0.10 / 0.11	-0.03 / 0.06	-0.07 / 0.08
Sector 4	-0.10 / 0.10	-0.12 / 0.12	-0.07 / 0.08	-0.05 / 0.06
All sectors: -0.07 / 0.08				
ρ_S				
	Sector 1	Sector 2	Sector 3	Sector 4
Sector 1	0.08 / 0.08	0.09 / 0.10	0.01 / 0.03	0.03 / 0.05
Sector 2	0.09 / 0.10	0.07 / 0.07	0.02 / 0.04	0.02 / 0.04
Sector 3	0.01 / 0.03	0.02 / 0.04	0.11 / 0.11	0.07 / 0.07
Sector 4	0.03 / 0.05	0.02 / 0.04	0.07 / 0.07	0.11 / 0.11
All sectors: 0.05 / 0.06				
$C - \hat{C}$				
	Sector 1	Sector 2	Sector 3	Sector 4
Sector 1	-0.19 / 0.19	-0.19 / 0.19	-0.30 / 0.30	-0.27 / 0.27
Sector 2	-0.19 / 0.19	-0.18 / 0.18	-0.28 / 0.28	-0.25 / 0.25
Sector 3	-0.30 / 0.30	-0.28 / 0.28	-0.19 / 0.20	-0.23 / 0.23
Sector 4	-0.27 / 0.27	-0.25 / 0.25	-0.23 / 0.23	-0.15 / 0.16
All sectors: -0.22 / 0.22				

Table D.12: Mean difference/mean absolute difference for different statistics for the 3 factor $F_{1,\nu}$ - copula, $\hat{\nu} = 7.1$, fitted with pairwise likelihood

ρ^-				
Sector 1	0.19 / 0.19	0.21 / 0.21	0.18 / 0.18	0.16 / 0.16
Sector 2	0.21 / 0.21	0.26 / 0.26	0.15 / 0.15	0.16 / 0.17
Sector 3	0.18 / 0.18	0.15 / 0.15	0.19 / 0.19	0.17 / 0.17
Sector 4	0.16 / 0.16	0.16 / 0.17	0.17 / 0.17	0.18 / 0.18
All sectors: 0.17 / 0.17				
ρ^+				
	Sector 1	Sector 2	Sector 3	Sector 4
Sector 1	-0.04 / 0.06	-0.06 / 0.07	-0.04 / 0.06	-0.06 / 0.08
Sector 2	-0.06 / 0.07	-0.08 / 0.08	-0.07 / 0.08	-0.07 / 0.08
Sector 3	-0.04 / 0.06	-0.07 / 0.08	-0.07 / 0.08	-0.09 / 0.09
Sector 4	-0.06 / 0.08	-0.07 / 0.08	-0.09 / 0.09	-0.07 / 0.07
All sectors: -0.06 / 0.07				
ρ_S				
	Sector 1	Sector 2	Sector 3	Sector 4
Sector 1	0.05 / 0.06	0.05 / 0.06	0.07 / 0.07	0.06 / 0.06
Sector 2	0.05 / 0.06	0.05 / 0.06	0.05 / 0.06	0.08 / 0.09
Sector 3	0.07 / 0.07	0.05 / 0.06	0.08 / 0.08	0.05 / 0.05
Sector 4	0.06 / 0.06	0.08 / 0.09	0.05 / 0.05	0.08 / 0.08
All sectors: 0.06 / 0.06				
$C - \hat{C}$				
	Sector 1	Sector 2	Sector 3	Sector 4
Sector 1	-0.21 / 0.21	-0.24 / 0.24	-0.24 / 0.24	-0.25 / 0.25
Sector 2	-0.24 / 0.24	-0.20 / 0.20	-0.25 / 0.25	-0.20 / 0.20
Sector 3	-0.24 / 0.24	-0.25 / 0.25	-0.24 / 0.24	-0.27 / 0.27
Sector 4	-0.25 / 0.25	-0.20 / 0.20	-0.27 / 0.27	-0.19 / 0.19
All sectors: -0.21 / 0.21				

Table D.13: Mean difference/mean absolute difference for different statistics for the 1 factor $F_{1,\nu}$ - copula, fitted with MDE methods, with $\nu = 5$

ρ^-				
Sector 1	0.22 / 0.22	0.22 / 0.22	0.19 / 0.19	0.18 / 0.18
Sector 2	0.22 / 0.22	0.30 / 0.30	0.17 / 0.18	0.18 / 0.18
Sector 3	0.19 / 0.19	0.17 / 0.18	0.22 / 0.22	0.19 / 0.20
Sector 4	0.18 / 0.18	0.18 / 0.18	0.19 / 0.20	0.20 / 0.20
All sectors: 0.18 / 0.19				
ρ^+				
	Sector 1	Sector 2	Sector 3	Sector 4
Sector 1	0.18 / 0.18	0.10 / 0.13	0.11 / 0.12	0.08 / 0.09
Sector 2	0.10 / 0.13	0.08 / 0.14	0.09 / 0.13	0.04 / 0.09
Sector 3	0.11 / 0.12	0.09 / 0.13	0.15 / 0.15	0.10 / 0.11
Sector 4	0.08 / 0.09	0.04 / 0.09	0.10 / 0.11	0.09 / 0.10
All sectors: 0.09 / 0.11				
ρ_S				
	Sector 1	Sector 2	Sector 3	Sector 4
Sector 1	0.31 / 0.31	0.25 / 0.26	0.25 / 0.25	0.23 / 0.23
Sector 2	0.25 / 0.26	0.27 / 0.29	0.25 / 0.27	0.22 / 0.24
Sector 3	0.25 / 0.25	0.25 / 0.27	0.34 / 0.34	0.29 / 0.29
Sector 4	0.23 / 0.23	0.22 / 0.24	0.29 / 0.29	0.28 / 0.28
All sectors: 0.24 / 0.25				
$C - \hat{C}$				
	Sector 1	Sector 2	Sector 3	Sector 4
Sector 1	0.04 / 0.1	-0.06 / 0.09	-0.08 / 0.09	-0.08 / 0.09
Sector 2	-0.06 / 0.09	-0.02 / 0.08	-0.08 / 0.09	-0.07 / 0.08
Sector 3	-0.08 / 0.09	-0.08 / 0.09	0.00 / 0.09	-0.03 / 0.09
Sector 4	-0.08 / 0.09	-0.07 / 0.08	-0.03 / 0.09	0.01 / 0.08
All sectors: -0.05 / 0.08				

Table D.14: Mean difference/mean absolute difference for different statistics for the 2 factor $F_{1,\nu}$ - copula, fitted with MDE methods, with $\nu = 5$

ρ^-				
Sector 1	0.21 / 0.21	0.22 / 0.23	0.19 / 0.19	0.18 / 0.18
Sector 2	0.22 / 0.23	0.31 / 0.31	0.17 / 0.18	0.17 / 0.18
Sector 3	0.19 / 0.19	0.17 / 0.18	0.21 / 0.21	0.19 / 0.19
Sector 4	0.18 / 0.18	0.17 / 0.18	0.19 / 0.19	0.20 / 0.20
All sectors: 0.18 / 0.19				
ρ^+				
	Sector 1	Sector 2	Sector 3	Sector 4
Sector 1	0.17 / 0.17	0.10 / 0.12	0.11 / 0.12	0.09 / 0.10
Sector 2	0.10 / 0.12	0.08 / 0.13	0.11 / 0.14	0.06 / 0.10
Sector 3	0.11 / 0.12	0.11 / 0.14	0.14 / 0.14	0.09 / 0.10
Sector 4	0.09 / 0.10	0.06 / 0.10	0.09 / 0.10	0.08 / 0.09
All sectors: 0.09 / 0.11				
ρ_S				
	Sector 1	Sector 2	Sector 3	Sector 4
Sector 1	0.31 / 0.31	0.24 / 0.26	0.25 / 0.25	0.24 / 0.24
Sector 2	0.24 / 0.26	0.26 / 0.27	0.26 / 0.28	0.23 / 0.25
Sector 3	0.25 / 0.25	0.26 / 0.28	0.31 / 0.31	0.27 / 0.27
Sector 4	0.24 / 0.24	0.23 / 0.25	0.27 / 0.27	0.27 / 0.27
All sectors: 0.24 / 0.24				
$C - \hat{C}$				
	Sector 1	Sector 2	Sector 3	Sector 4
Sector 1	0.04 / 0.10	-0.07 / 0.09	-0.08 / 0.09	-0.08 / 0.08
Sector 2	-0.07 / 0.09	-0.04 / 0.08	-0.06 / 0.08	-0.06 / 0.06
Sector 3	-0.08 / 0.09	-0.06 / 0.08	-0.03 / 0.06	-0.07 / 0.08
Sector 4	-0.08 / 0.08	-0.06 / 0.06	-0.07 / 0.08	-0.02 / 0.06
All sectors: -0.05 / 0.07				

Table D.15: Mean difference/mean absolute difference for different statistics for the 3 factor $F_{1,\nu}$ - copula, fitted with MDE methods, with $\nu = 5$

ρ^-				
Sector 1	0.22 / 0.22	0.22 / 0.22	0.19 / 0.19	0.18 / 0.18
Sector 2	0.22 / 0.22	0.31 / 0.31	0.17 / 0.18	0.18 / 0.18
Sector 3	0.19 / 0.19	0.17 / 0.18	0.22 / 0.22	0.19 / 0.20
Sector 4	0.18 / 0.18	0.18 / 0.18	0.19 / 0.20	0.20 / 0.20
All sectors: 0.18 / 0.19				
ρ^+				
	Sector 1	Sector 2	Sector 3	Sector 4
Sector 1	0.12 / 0.13	0.10 / 0.13	0.11 / 0.12	0.09 / 0.10
Sector 2	0.10 / 0.13	0.06 / 0.12	0.10 / 0.13	0.06 / 0.09
Sector 3	0.11 / 0.12	0.10 / 0.13	0.13 / 0.14	0.09 / 0.10
Sector 4	0.09 / 0.01	0.06 / 0.09	0.09 / 0.10	0.08 / 0.09
All sectors: 0.09 / 0.11				
ρ_S				
	Sector 1	Sector 2	Sector 3	Sector 4
Sector 1	0.26 / 0.26	0.25 / 0.26	0.24 / 0.24	0.24 / 0.24
Sector 2	0.25 / 0.26	0.25 / 0.26	0.26 / 0.28	0.23 / 0.25
Sector 3	0.24 / 0.24	0.26 / 0.28	0.31 / 0.31	0.26 / 0.26
Sector 4	0.24 / 0.24	0.23 / 0.25	0.26 / 0.26	0.27 / 0.27
All sectors: 0.23 / 0.24				
$C - \hat{C}$				
	Sector 1	Sector 2	Sector 3	Sector 4
Sector 1	-0.03 / 0.06	-0.06 / 0.07	-0.08 / 0.09	-0.07 / 0.08
Sector 2	-0.06 / 0.07	-0.05 / 0.07	-0.06 / 0.08	-0.06 / 0.07
Sector 3	-0.08 / 0.09	-0.06 / 0.08	-0.03 / 0.06	-0.07 / 0.08
Sector 4	-0.07 / 0.08	-0.06 / 0.07	-0.07 / 0.08	-0.02 / 0.06
All sectors: -0.05 / 0.07				

E: Constructing the bivariate F -copula

We follow the same process that (Quessy et al. 2016) used for the bivariate normal copula to construct a central F -copula. This copula is built using the inversion method. If we let the variables W_1 and W_2 have a bivariate $F_{1,\nu}$ distribution, such that

$$(W_1, W_2) = \frac{(Z_1^2, Z_2^2)}{U_\nu/\nu}$$

where (Z_1, Z_2) has a standard bivariate normal distribution with positive correlation ρ , and U_ν is an independent chi-squared distribution with ν degrees of freedom. Then the distribution function of the bivariate F -copula, $C_\rho^F(W_1, W_2)$, with covariance ρ can be expressed as:

$$C_\rho^F(u_1, u_2) = \mathbf{F} \{ F_1^{-1}(u_1), F_2^{-1}(u_2) \}$$

where \mathbf{F} is the distribution function of the bivariate $F_{1,\nu}$ -distribution and $F_{W_1}^{-1}$ and $F_{W_2}^{-1}$ are quantile transformation to univariate $F_{1,\nu}$ distributions. Then we can rewrite the copula distribution function as:

$$C_\rho^F(u_1, u_2) = \Pr \{ F_1(W_1) \leq u_1, F_2(W_2) \leq u_2 \}$$

Now, for $i = 1, 2$, F_i can be decomposed into

$$\begin{aligned} F_i(w_i) &= \Pr(W_1 \leq w_1) = \Pr \left(-\sqrt{w_i} \leq \frac{Z_i}{\sqrt{U_\nu/\nu}} \leq \sqrt{w_i} \right) \\ &= T_\nu(\sqrt{w_i}; \nu) - T_\nu(-\sqrt{w_i}; \nu) = 2 T_\nu(\sqrt{w_i}; \nu) - 1 \end{aligned}$$

If we solve this for w_i we find that:

$$w_i = \left\{ T_\nu^{-1} \left(\frac{1 + F_i(w_i)}{2}; \nu \right) \right\}^2$$

Hence $F^{-1}(u) = \{T_\nu^{-1}(\frac{1+u}{2}; \nu)\}^2$ and $F(u) = 1 - 2 T_\nu(-\sqrt{u}; \nu)$.

Using this identity, we can write the copula distribution function as:

$$\begin{aligned} C_\rho^F(u_1, u_2) &= \Pr\{F_1(W_1) \leq u_1, F_2(W_2) \leq u_2\} \\ &= \Pr\left\{\left|\frac{Z_1}{\sqrt{U_\nu/\nu}}\right| \leq \sqrt{F_1^{-1}(u_1)}, \left|\frac{Z_2}{\sqrt{U_\nu/\nu}}\right| \leq \sqrt{F_2^{-1}(u_2)}\right\} \\ &= \sum_{(\epsilon_1, \epsilon_2) \in \{-1, 1\}^2} \epsilon_1 \epsilon_2 \mathbf{T}_\nu \left\{ T_\nu^{-1}\left(\frac{1 + \epsilon_1 u_1}{2}\right), T_\nu^{-1}\left(\frac{1 + \epsilon_2 u_2}{2}\right) \right\} \end{aligned} \quad (\text{E.1})$$

With the last line being true by the symmetry of the multivariate t -distribution.

Remember that if we use the inversion method, the density of the bivariate t -copula C_ρ^t is given by

$$C_\rho^t(v_1, v_2) = \mathbf{T}_\nu \{T_\nu^{-1}(v_1), T_\nu^{-1}(v_2)\}$$

We can substitute this into line E.1 to write the F -copula distribution function in terms of t -copula densities, like so:

$$C_\rho^F(u_1, u_2) = \sum_{(\epsilon_1, \epsilon_2) \in \{-1, 1\}^2} \epsilon_1 \epsilon_2 C_\rho^t\left(\frac{1 + \epsilon_1 u_1}{2}, \frac{1 + \epsilon_2 u_2}{2}\right)$$

The t -copula is radially symmetric, which means that for $(v_1, v_2) \in [0, 1]^2$,

$$C_\rho^t(v_1, v_2) = C_\rho^t(1 - v_1, 1 - v_2) + v_1 + v_2 - 1$$

which of course, implies that:

$$C_\rho^t(v_1, 1 - v_2) = C_\rho^t(1 - v_1, v_2) + v_1 - v_2$$

Using this relation, the distribution function of the F -copula simplifies to

$$C_\rho^F(u_1, u_2) = 2C_\rho^t\left(\frac{1 + u_1}{2}, \frac{1 + u_2}{2}\right) - 2C_\rho^t\left(\frac{1 + u_1}{2}, \frac{1 - u_2}{2}\right) + u_2$$

Then, by differentiation, the density of the F -copula is given by

$$c_\rho^F(u_1, u_2) = \frac{1}{2} c_\rho^t\left(\frac{1 + u_1}{2}, \frac{1 + u_2}{2}\right) + \frac{1}{2} c_\rho^t\left(\frac{1 + u_1}{2}, \frac{1 - u_2}{2}\right)$$

F: Pearson correlation in the multi-variate Chi-squared & F -distributions

F.1 Pearson correlation for the Chi-squared distribution

Let Z_i and Z_j have a joint bivariate standard normal distribution with correlation ρ , or

$$(Z_i, Z_j) \sim N \left(0, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right)$$

Let us show that the bivariate chi-squared distribution (Z_i^2, Z_j^2) has correlation ρ^2 . First let us find $\mathbb{E}(Z_i^2 Z_j^2)$:

$$\begin{aligned} \mathbb{E}(Z_i^2 Z_j^2) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{z_i^2 z_j^2}{2\pi\sqrt{1-\rho^2}} e^{-\frac{z_i^2 - 2\rho z_i z_j + z_j^2}{2(1-\rho^2)}} dz_j dz_i \\ &= \int_{-\infty}^{\infty} \frac{z_i^2}{\sqrt{2\pi}} e^{-\frac{z_i^2 - \rho^2 z_i^2}{2(1-\rho^2)}} \int_{-\infty}^{\infty} \frac{z_j^2}{\sqrt{2\pi(1-\rho^2)}} e^{-\frac{(z_j - \rho z_i)^2}{2(1-\rho^2)}} dz_j dz_i \\ &= (1 - \rho^2) \int_{-\infty}^{\infty} \frac{z_i^2}{\sqrt{2\pi}} e^{-\frac{z_i^2}{2}} dz_i + \rho^2 \int_{-\infty}^{\infty} \frac{z_i^4}{\sqrt{2\pi}} e^{-\frac{z_i^2}{2}} dz_i \\ &= 1 - \rho^2 + 3\rho^2 \\ &= 1 + 2\rho^2 \end{aligned}$$

Note that since Z_i and Z_j are both standard normal models, for $k = i, j$:

$$\mathbb{E}(Z_k^2) = \text{VAR}(Z_k) = 1$$

Now we can find the covariance between Z_i^2 and Z_j^2 :

$$\text{COV}(Z_i^2, Z_j^2) = \mathbb{E}(Z_i^2 Z_j^2) - \mathbb{E}(Z_i^2)\mathbb{E}(Z_j^2) = 1 + 2\rho^2 - 1 = 2\rho^2$$

Since Z_i^2 and Z_j^2 are both chi-squared random variables with one degree of freedom, both have a variance of 2. Using this, we can find the Pearson correlation between Z_i^2 and Z_j^2 :

$$\text{CORR}(Z_i^2, Z_j^2) = \frac{\text{COV}(Z_i^2, Z_j^2)}{\sqrt{\text{VAR}(Z_i)\text{VAR}(Z_j)}} = \frac{2\rho^2}{2} = \rho^2$$

Hence

$$\text{CORR}(Z_i^2, Z_j^2) = \text{CORR}(Z_i, Z_j)^2$$

F.2 Pearson correlation for the *F*-distribution

Now let us construct a bivariate *t*-distribution with ν degrees of freedom and correlation ρ . Let us re-use the bivariate standard normal (Z_i, Z_j) from the previous section. Let U_ν be an independent chi-squared distribution with ν degrees of freedom. Then let:

$$(S_i, S_j) = \frac{(Z_i^2, Z_j^2)}{\sqrt{U_\nu/\nu}}$$

Then (S_i, S_j) has a bivariate *t*-distribution with ν degrees of freedom. Moreover, if $\nu > 4$:

$$\text{CORR}(S_i, S_j) = \text{CORR}(Z_i, Z_j) = \rho$$

If $\nu < 4$ correlation does not exist. Let us find the covariance between S_i^2 and S_j^2 :

$$\begin{aligned} \text{COV}(S_i^2, S_j^2) &= \mathbb{E}(S_i^2 S_j^2) - \mathbb{E}(S_i^2) \mathbb{E}(S_j^2) \\ &= \mathbb{E}\{\mathbb{E}(S_i^2 S_j^2 | \chi_\nu^2)\} - \mathbb{E}(S_i^2) \mathbb{E}(S_j^2) \\ &= \mathbb{E}\left\{\frac{\mathbb{E}(Z_i^2 Z_j^2)}{(\chi_\nu^2)^2/\nu^2}\right\} - \mathbb{E}\left\{\frac{Z_i^2}{\chi_\nu^2/\nu}\right\} \mathbb{E}\left\{\frac{Z_j^2}{\chi_\nu^2/\nu}\right\} \\ &= \mathbb{E}(Z_i^2 Z_j^2) \mathbb{E}\left\{\frac{\nu^2}{(\chi_\nu^2)^2}\right\} - \mathbb{E}(Z_i^2) \mathbb{E}(Z_j^2) \mathbb{E}\left\{\frac{\nu}{\chi_\nu^2}\right\}^2 \quad (\text{Independence}) \end{aligned}$$

We already found that:

$$\mathbb{E}(Z_i^2 Z_j^2) = 1 + 2\rho^2, \quad \mathbb{E}(Z_i^2) = \mathbb{E}(Z_j^2)$$

To solve this we just need to find the second and forth moment of the inverse chi-squared distribution. The density of this distribution is:

$$f_{1/\chi_\nu^2}(x) = \frac{2^{-\nu/2}}{\Gamma(\nu/2)} x^{-\nu/2-1} e^{-1/2x}, \quad x > 0$$

Using the density, we can calculate the second moment of the inverse chi-squared distribution like so:

$$\begin{aligned}
\mathbb{E}\left(\frac{1}{(\chi_\nu^2)^2}\right) &= \int_0^\infty x^2 \frac{2^{-\nu/2}}{\Gamma(\frac{\nu}{2})} x^{-\nu/2-1} e^{-1/2x} dx \\
&= \frac{2^{(\nu-4)/2} \Gamma(\frac{\nu-4}{2})}{2^{\nu/2} \Gamma(\frac{\nu}{2})} \int_0^\infty \frac{2^{-(\nu-4)/2}}{\Gamma(\frac{\nu-4}{2})} x^{-(\nu-4)/2-1} e^{-1/2x} dx \\
&= \frac{2^{(\nu-4)/2} \Gamma(\frac{\nu-4}{2})}{2^{\nu/2} \Gamma(\frac{\nu}{2})} \\
&= \frac{1}{2^2} \frac{\Gamma(\frac{\nu-4}{2})}{\frac{\nu-2}{2} \frac{\nu-4}{2} \Gamma(\frac{\nu-4}{2})} \\
&= \frac{1}{(\nu-2)(\nu-4)}
\end{aligned} \tag{*}$$

With the integral in (*) equalling one since it is the density of a $\frac{1}{\chi_{\nu-2}^2}$ distribution, integrated over its support. By the same process, the first inverse chi-squared distribution is given by:

$$\begin{aligned}
\mathbb{E}\left(\frac{1}{\chi_\nu^2}\right) &= \int_0^\infty x \frac{2^{-\nu/2}}{\Gamma(\frac{\nu}{2})} x^{-\nu/2-1} e^{-1/2x} dx \\
&= \frac{2^{(\nu-2)/2} \Gamma(\frac{\nu-2}{2})}{2^{\nu/2} \Gamma(\frac{\nu}{2})} \int_0^\infty \frac{2^{-(\nu-2)/2}}{\Gamma(\frac{\nu-2}{2})} x^{-(\nu-2)/2-1} e^{-1/2x} dx \\
&= \frac{2^{(\nu-2)/2} \Gamma(\frac{\nu-2}{2})}{2^{\nu/2} \Gamma(\frac{\nu}{2})} \\
&= \frac{1}{2} \frac{\Gamma(\frac{\nu-2}{2})}{\frac{\nu-2}{2} \Gamma(\frac{\nu-8}{2})} \\
&= \frac{1}{(\nu-2)}
\end{aligned}$$

From these results, we can see that we require $\nu > 4$ for both the second and forth moments to exist. Substituting all our results back into our equation for covariance, we get:

$$\begin{aligned}
\text{COV}(S_i^2 S_j^2) &= \frac{\nu^2(2\rho^2 + 1)}{(\nu-2)(\nu-4)} - \frac{\nu^2}{(\nu-2)^2} \\
&= \frac{2\rho^2\nu^2(\nu-2) + \nu^2(\nu-2) - \nu^2(\nu-4)}{(\nu-2)^2(\nu-4)} \\
&= \frac{2\rho^2\nu^2(\nu-2) + 2\nu^2}{(\nu-2)^2(\nu-4)}
\end{aligned}$$

We still need to find the variance of S_i^2 and S_j^2 to calculate the correlation. For $k = i, j$:

$$\begin{aligned}
\text{VAR}(S_k^2) &= \mathbb{E}(S_k^4) - \mathbb{E}(S_k^2)^2 \\
&= \mathbb{E}(S_k^4|\chi_\nu^2) - \mathbb{E}(S_k^2|\chi_\nu^2)^2 \\
&= \mathbb{E}\left(\frac{\mathbb{E}(Z_k^4)}{(\chi_\nu^2)^2/\nu^2}\right) - \mathbb{E}\left(\frac{\mathbb{E}(Z_k^2)}{\chi_\nu^2/\nu}\right)^2 \\
&= \mathbb{E}(Z_k^4)\mathbb{E}\left(\frac{\nu^2}{(\chi_\nu^2)^2}\right) - \mathbb{E}(Z_k^2)\mathbb{E}\left(\frac{\nu}{\chi_\nu^2}\right)^2 \quad (\text{Independence})
\end{aligned}$$

Using the fact Z_i is a standard normal random variable (meaning it has a forth moment of 3 and second moment of 1) and using the first and second moments of the inverse chi-squared distribution that we have already calculated, we can see the variance is:

$$\text{VAR}(S_k^2) = \frac{3\nu^2}{(\nu-2)(\nu-4)} - \frac{\nu^2}{(\nu-2)^2} = \frac{2\nu^2(\nu-1)}{(\nu-2)^2(\nu-4)}, \quad \nu > 4$$

So our Pearson correlation is given by:

$$\begin{aligned}
\text{CORR}(S_i^2, S_j^2) &= \frac{\text{COV}(S_i^2, S_j^2)}{\sqrt{\text{VAR}(S_i^2)}\sqrt{\text{VAR}(S_j^2)}} \\
&= \frac{2\rho^2\nu^2(\nu-2) + 2\nu^2}{2\nu^2(\nu-1)} \\
&= \frac{\rho^2(\nu-2) + 1}{\nu-1}
\end{aligned}$$

Hence the relationship between correlation in the bivariate $F_{1,\nu}$ distribution (S_i^2, S_j^2) , and the bivariate t_ν -distribution it is constructed from is:

$$(\nu-1) \text{CORR}(S_i^2, S_j^2) = (\nu-2) \text{CORR}(S_i, S_j)^2 + 1$$

G: Simulation study results

Table G.1: Mean difference / absolute mean difference of diagnostic metrics for the χ_1^2 model fitted with pairwise likelihood

	χ_1^2 pairwise likelihood		
	1 factor	2 factor	3 factor
Semicorrelation difference (upper)	0.00 / 0.04	0.00 / 0.04	0.00 / 0.04
Semicorrelation difference (lower)	0.00 / 0.05	0.00 / 0.05	0.00 / 0.05
Difference ρ_S	0.00 / 0.02	0.00 / 0.02	0.00 / 0.02
Difference between Σ^2 and $\hat{\Sigma}^2$	0.00 / 0.02	0.00 / 0.02	0.00 / 0.02
Mean elapsed time (s)	0.75	6.77	34.41

Table G.2: Mean difference / absolute mean difference of diagnostic metrics for the χ_1^2 model fitted with Kendall's Tau

	χ_1^2 Kendall's τ		
	1 factor	2 factor	3 factor
Semicorrelation difference (upper)	0.00 / 0.04	0.00 / 0.04	0.00 / 0.04
Semicorrelation difference (lower)	0.00 / 0.05	0.00 / 0.05	0.00 / 0.05
Difference ρ_S	0.00 / 0.02	0.00 / 0.01	0.00 / 0.01
Difference between Σ^2 and $\hat{\Sigma}^2$	0.00 / 0.02	0.00 / 0.03	0.00 / 0.03
Mean elapsed time (s)	0.01	0.08	0.49

Table G.3: Mean difference / absolute mean difference of diagnostic metrics for the χ^2_1 model fitted with the minimum distance estimator method

	χ^2 MDE		
	1 factor	2 factor	3 factor
Semicorrelation difference (upper)	0.00 / 0.04	0.00 / 0.04	0.00 / 0.04
Semicorrelation difference (lower)	0.00 / 0.05	0.00 / 0.05	0.00 / 0.05
Difference ρ_S	0.00 / 0.03	0.00 / 0.03	0.00 / 0.03
Difference between Σ^2 and $\hat{\Sigma}^2$	0.00 / 0.02	0.00 / 0.03	0.00 / 0.03
Mean elapsed time (s)	0.01	0.10	0.32

Table G.4: Mean difference / absolute mean difference of diagnostic metrics for the $F_{1,5}$ -model fitted with a minimum distance estimator method

	$F_{1,5}$ MDE, ν known		
	1 factor	2 factor	3 factor
Semicorrelation difference (upper)	0.01 / 0.07	0.01 / 0.07	0.02 / 0.07
Semicorrelation difference (lower)	-0.01 / 0.06	0.00 / 0.05	0.00 / 0.05
Difference ρ_S	0.02 / 0.08	0.02 / 0.08	0.02 / 0.08
Difference between Σ^2 and $\hat{\Sigma}^2$	0.04 / 0.11	0.05 / 0.12	0.04 / 0.12
Mean elapsed time (s)	0.01	0.07	0.23

Table G.5: Mean difference / absolute mean difference of diagnostic metrics for the $F_{1,5}$ -model fitted with pairwise likelihood

	$F_{1,5}$ pairwise likelihood		
	1 factor	2 factor	3 factor
Semicorrelation difference (upper)	-0.01 / 0.04	-0.01 / 0.04	-0.01 / 0.04
Semicorrelation difference (lower)	0.00 / 0.05	0.00 / 0.05	0.00 / 0.05
Difference ρ_S	0.00 / 0.02	0.00 / 0.01	0.00 / 0.01
Difference between Σ^2 and $\hat{\Sigma}^2$	0.03 / 0.05	0.03 / 0.05	0.04 / 0.06
Mean elapsed time (s)	19.94	114.21	194.33

H: Normal scores scatterplots

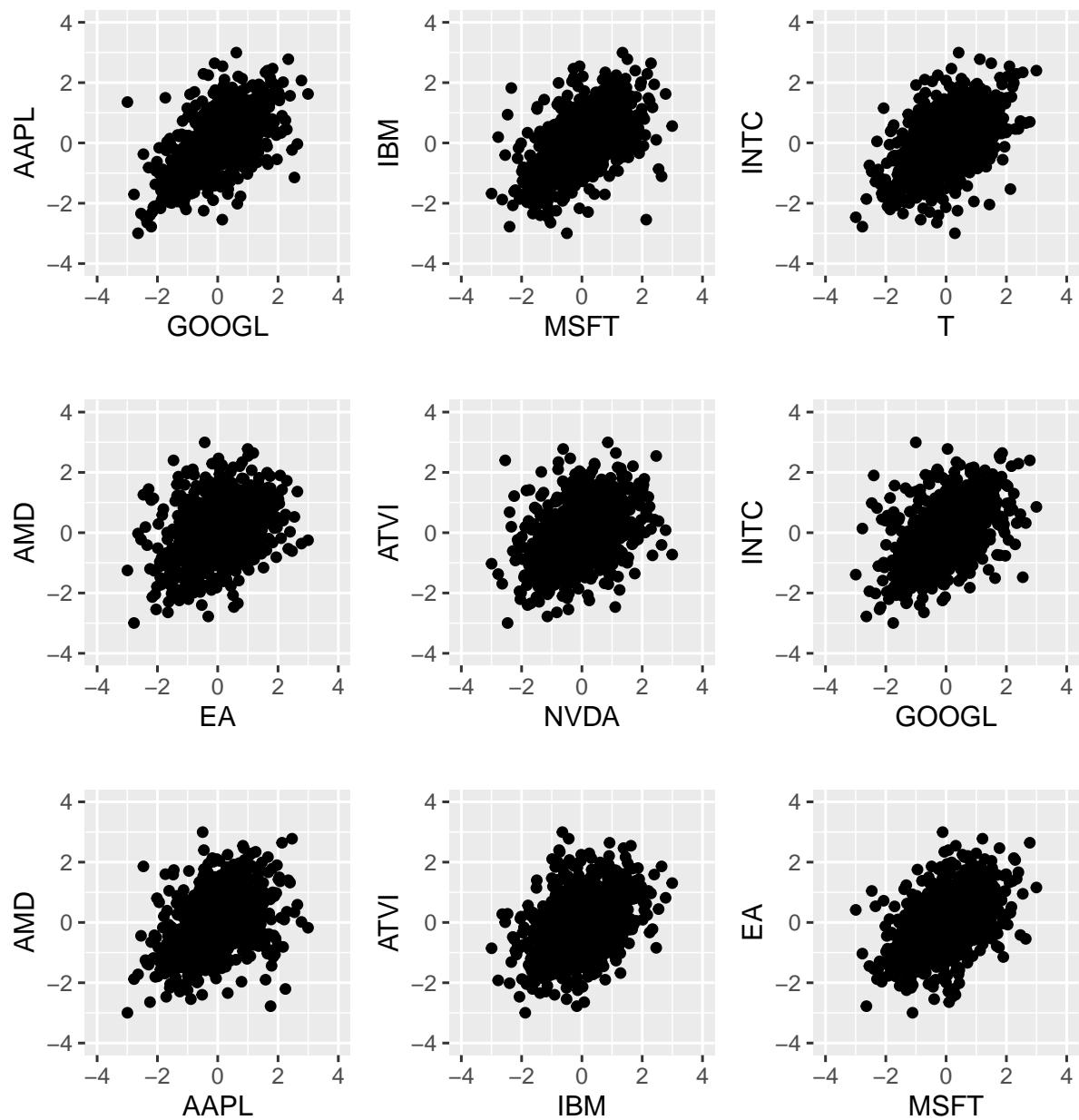


Figure H.1: Normal scores scatterplots of log stock returns in the technology industry

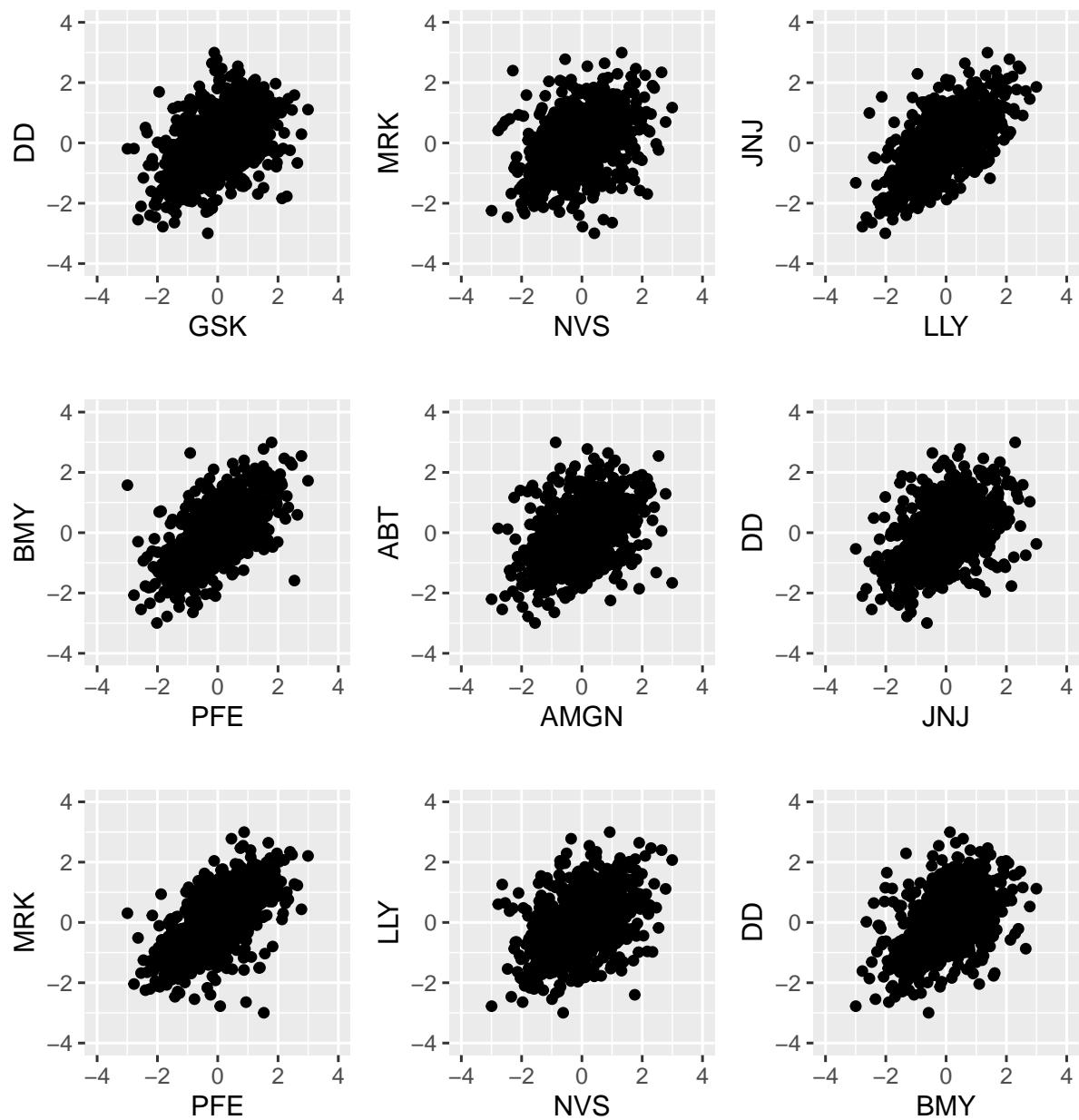


Figure H.2: Normal scores scatterplots of log stock returns in the pharmaceutical industry

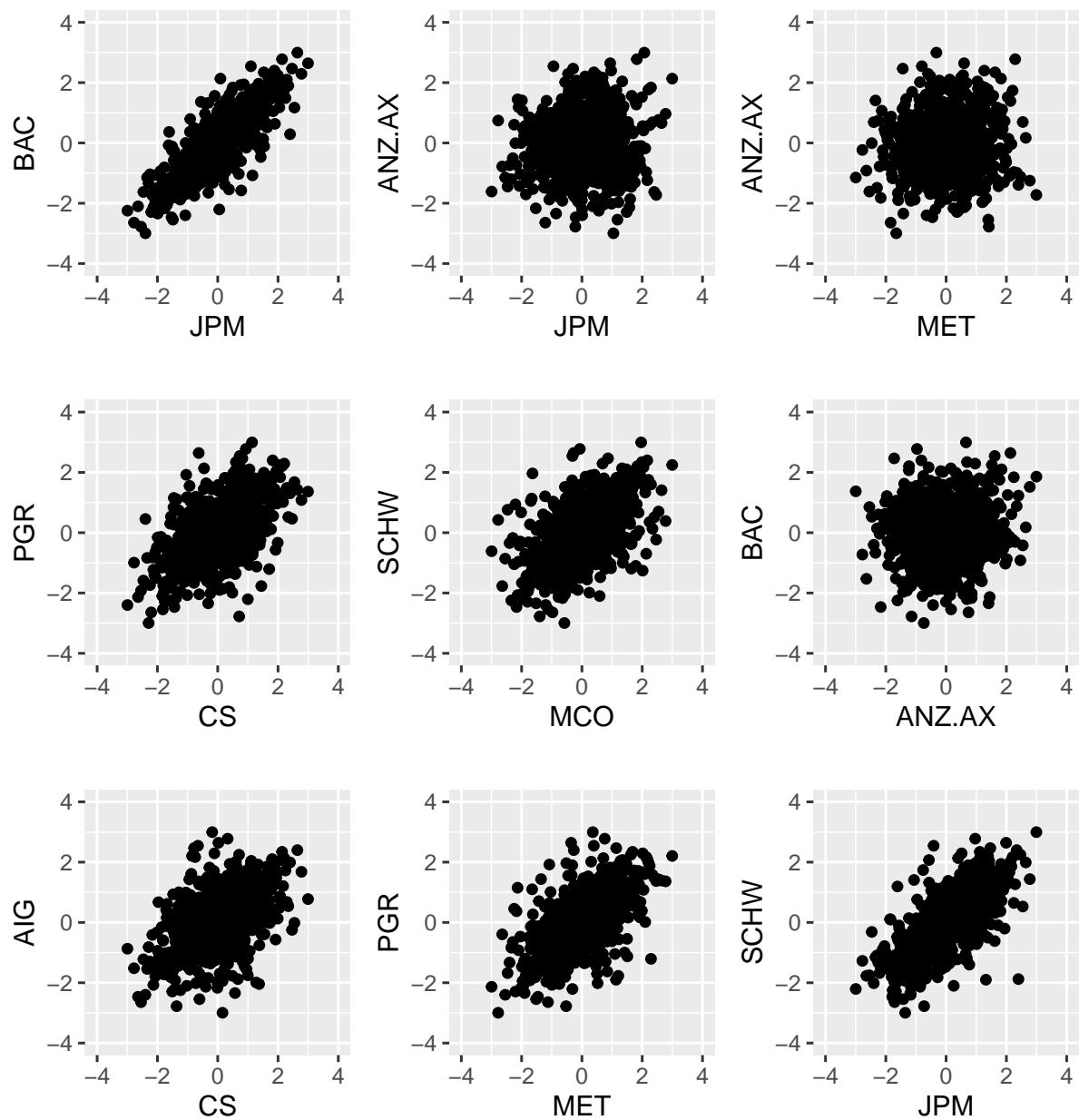


Figure H.3: Normal scores scatterplots of log stock returns in the financial industry

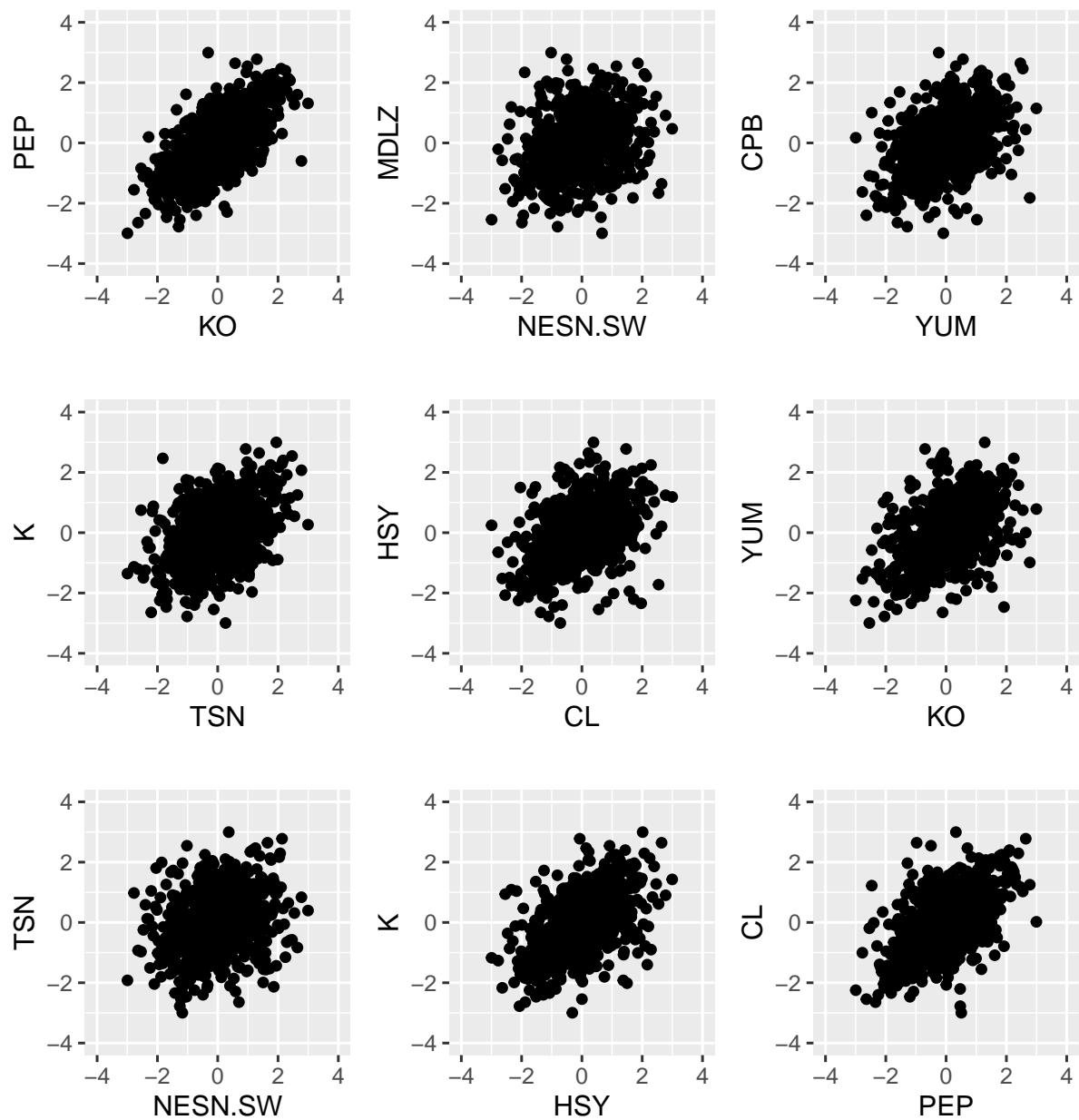


Figure H.4: Normal scores scatterplots of log stock returns in the consumer staples industry

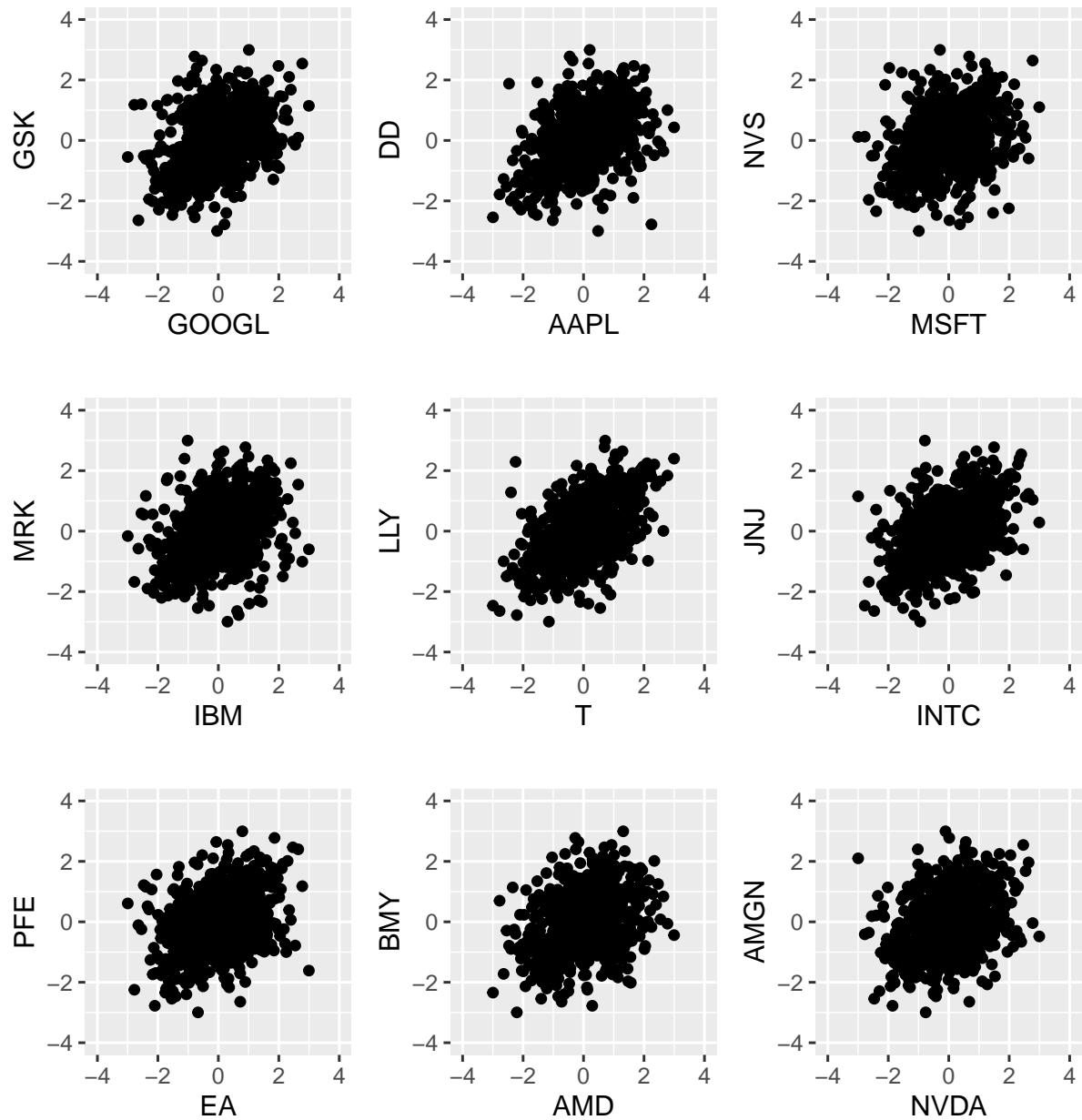


Figure H.5: Normal scores scatterplots of log stock returns between stocks in the technology and pharmaceutical industries. Technology stocks are on the x-axis, pharmaceutical stocks are on the y-axis

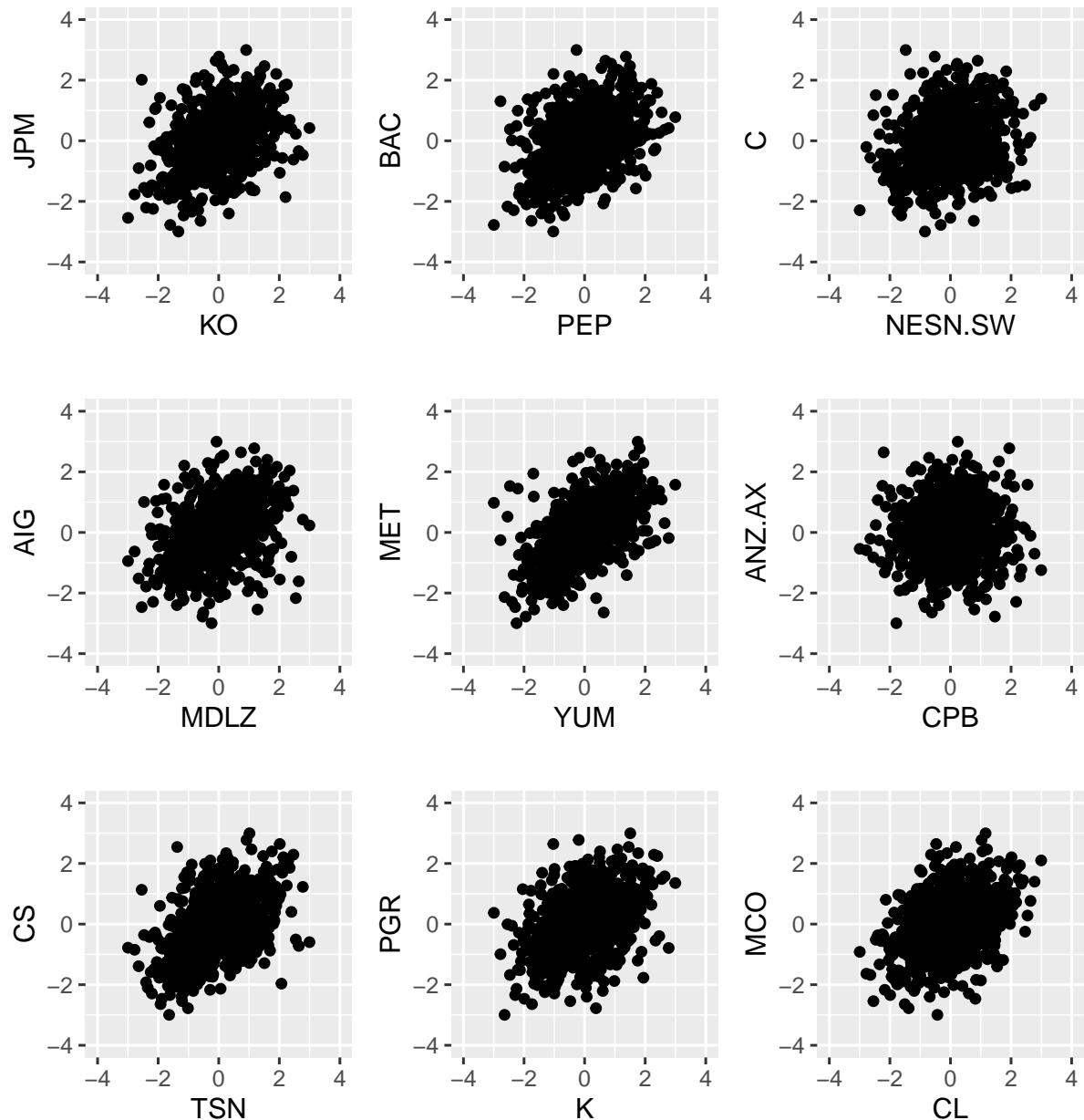


Figure H.6: Normal scores scatterplots of log stock returns between stocks in the technology and pharmaceutical industries. Consumer staples stocks are on the x-axis, financial stocks are on the y-axis

Bibliography

- Aas, K., Czado, C., Frigessi, A. & Bakken, H. (2009), ‘Pair-copula constructions of multiple dependence’, *Insurance: Mathematics and economics* **44**(2), 182–198.
- Bárdossy, A. (2006), ‘Copula-based geostatistical models for groundwater quality parameters’, *Water Resources Research* **42**(11).
- Bárdossy, A. & Li, J. (2008), ‘Geostatistical interpolation using copulas’, *Water Resources Research* **44**(7).
- Barndorff-Nielsen, O., Kent, J. & Sørensen, M. (1982), ‘Normal variance-mean mixtures and z distributions’, *International Statistical Review/Revue Internationale de Statistique* pp. 145–159.
- Bollerslev, T. (1986), ‘Generalized autoregressive conditional heteroskedasticity’, *Journal of econometrics* **31**(3), 307–327.
- Bollerslev, T. (1987), ‘A conditional heteroskedastic time series model for speculative prices and rates of return’, *The Review of Economics and Statistics* **69**(3), 542–547.
- Bollerslev, T. & Andersen, T. G. (2018), *Volatility*, Edward Elgar Publishing.
URL: <https://doi.org/10.4337%2F9781788110624>
- Bollerslev, T., Chou, R. Y. & Kroner, K. F. (1992), ‘Arch modeling in finance: A review of the theory and empirical evidence’, *Journal of econometrics* **52**(1-2), 5–59.
- Box, G. E. P., Jenkins, G. M., Reinsel, G. C., Ljung, G. M. & Ljung, G. M. (2015), *Time Series Analysis : Forecasting and Control*, John Wiley & Sons, Incorporated, New York, UNITED STATES.
- Box, G. E. & Pierce, D. A. (1970), ‘Distribution of residual autocorrelations in autoregressive-integrated moving average time series models’, *Journal of the American statistical Association* **65**(332), 1509–1526.

- Donnelly, C. & Embrechts, P. (2010), ‘The devil is in the tails: actuarial mathematics and the subprime mortgage crisis’, *Astin Bulletin* **40**(1), 1–33.
- Engle, R. F. (1982), ‘Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation’, *Econometrica: Journal of the Econometric Society* pp. 987–1007.
- Erb, C. B., Harvey, C. R. & Viskanta, T. E. (1994), ‘Forecasting international equity correlations’, *Financial Analysts Journal* **50**(6), 32–45.
URL: <http://www.jstor.org/stable/4479787>
- Ghalanos, A. (2020), *rugarch: Univariate GARCH models*. R package version 1.4-4.
- Hassani, H. & Yeganegi, M. R. (2020), ‘Selecting optimal lag order in ljung–box test’, *Physica A: Statistical Mechanics and its Applications* **541**, 123700.
URL: <https://www.sciencedirect.com/science/article/pii/S0378437119320618>
- Heinen, A. & Valdesogo, A. (2020), ‘Spearman rank correlation of the bivariate student t and scale mixtures of normal distributions’, *Journal of Multivariate Analysis* **179**, 104650.
- Hentschel, L. (1995), ‘All in the family nesting symmetric and asymmetric garch models’, *Journal of financial economics* **39**(1), 71–104.
- Hill, G. W. (1970), ‘Algorithm 396: Student’s t-quantiles’, *Communications of the ACM* **13**(10), 619–620.
- Hofert, M. (2010), Construction and sampling of nested archimedean copulas, in ‘Copula theory and its applications’, Springer, pp. 147–160.
- Joe, H. (2014), *Dependence modeling with copulas*, CRC press.
- Joe, H. & Xu, J. J. (1996), The estimation method of inference functions for margins for multivariate models, Technical Report 166, Department of Statistics, The University of British Columbia, Vancouver, Canada.
- Jondeau, E. & Rockinger, M. (2006), ‘The copula-garch model of conditional dependences: An international stock market application’, *Journal of international money and finance* **25**(5), 827–853.
- Klüppelberg, C. & Kuhn, G. (2009), ‘Copula structure analysis’, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **71**(3), 737–753.

Krupskii, P. (2014), Structured factor copulas and tail inference, PhD thesis, The Faculty of Graduate and Postdoctoral Studies (Statistics), Vancouver, Canada.
URL: <https://doi.library.ubc.ca/10.14288/1.0167545>

Krupskii, P. & Joe, H. (2013), ‘Factor copula models for multivariate data’, *Journal of Multivariate Analysis* **120**, 85–101.

Krupskii, P. & Joe, H. (2015), ‘Structured factor copula models: Theory, inference and computation’, *Journal of Multivariate Analysis* **138**, 53–73.

Krupskii, P. & Joe, H. (2020), ‘Flexible copula models with dynamic dependence and application to financial data’, *Econometrics and Statistics* **16**, 148–167.

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

Mandelbrot, B. (1963), ‘The variation of certain speculative prices’, *Journal of Business* **36**(4), 394–419.

Nelson, R. B. (2006), *An Introduction to Copulas*, Springer Science+Business Media, Inc., New York, UNITED STATES.

Nikoloulopoulos, A. K., Joe, H. & Li, H. (2012), ‘Vine copulas with asymmetric tail dependence and applications to financial return data’, *Computational Statistics & Data Analysis* **56**(11), 3659–3673.

Oh, D. H. & Patton, A. J. (2017), ‘Modeling dependence in high dimensions with factor copulas’, *Journal of Business & Economic Statistics* **35**(1), 139–154.

Oh, D. H. & Patton, A. J. (2018), ‘Time-varying systemic risk: Evidence from a dynamic copula model of cds spreads’, *Journal of Business & Economic Statistics* **36**(2), 181–195.

Opschoor, A., Lucas, A., Barra, I. & Van Dijk, D. (2020), ‘Closed-form multi-factor copula models with observation-driven dynamic factor loadings’, *Journal of Business & Economic Statistics* pp. 1–14.

Patton, A. J. (2006), ‘Modelling asymmetric exchange rate dependence’, *International economic review* **47**(2), 527–556.

Quessy, J.-F., Rivest, L.-P. & Toupin, M.-H. (2016), ‘On the family of multivariate chi-square copulas’, *Journal of Multivariate Analysis* **152**, 40–60.

- Schweizer, B., Wolff, E. F. et al. (1981), ‘On nonparametric measures of dependence for random variables’, *The annals of statistics* **9**(4), 879–885.
- Sklar, A. (1959), ‘Distribution functions of n dimensions and margins’, *Publications of the Institute of Statistics of the University of Paris* **8**, 229–231.
- Varin, C. (2008), ‘On composite marginal likelihoods’, *Asta advances in statistical analysis* **92**(1), 1–28.
- Venables, W. N. & Ripley, B. D. (2013), *Modern applied statistics with S-PLUS*, Springer Science & Business Media.
- Xu, J. J. (1996), Statistical modelling and inference for multivariate and longitudinal discrete response data, PhD thesis, University of British Columbia.