ROYAL INSTITUTE
OF TECHNOLOGY

Master Thesis

Imputation of Missing Data with Application to Commodity Futures

Department of Mathematics,
Royal Institute of Technology (KTH)

Author:
Simon Östlund

Supervisors:
Camilla Johansson Landén, KTH
Mikhail Nechaev, NASDAQ

17th May 2016

Imputation of Missing Data with Application to Commodity Futures

Abstract

In recent years additional requirements have been imposed on financial institutions, including Central Counterparty clearing houses (CCPs), as an attempt to assess quantitative measures of their exposure to different types of risk. One of these requirements results in a need to perform stress tests to check the resilience in case of a stressed market/crisis. However, financial markets develop over time and this leads to a situation where some instruments traded today are not present at the chosen date because they were introduced after the considered historical event. Based on current routines, the main goal of this thesis is to provide a more sophisticated method to impute (fill in) historical missing data as a preparatory work in the context of stress testing. The models considered in this paper include two methods currently regarded as state-of-the-art techniques, based on maximum likelihood estimation (MLE) and multiple imputation (MI), together with a third alternative approach involving copulas. The different methods are applied on historical return data of commodity futures contracts from the Nordic energy market. By using conventional error metrics, and out-of-sample log-likelihood, the conclusion is that it is very hard (in general) to distinguish the performance of each method, or draw any conclusion about how good the models are in comparison to each other. Even if the Student's t-distribution seems (in general) to be a more adequate assumption regarding the data compared to the normal distribution, all the models are showing quite poor performance. However, by analysing the conditional distributions more thoroughly, and evaluating how well each model performs by extracting certain quantile values, the performance of each method is increased significantly. By comparing the different models (when imputing more extreme quantile values) it can be concluded that all methods produce satisfying results, even if the g-copula and t-copula models seems to be more robust than the respective linear models.

Keywords: Missing Data, Bayesian Statistics, Expectation Conditional Maximization (ECM), Conditional Distribution, Robust Regression, MCMC, Copulas.

Imputation av Saknad Data med Tillämpning på Råvaruterminer

Sammanfattning

På senare år har ytterliggare krav införts för finansiella institut (t.ex. Clearinghus) i ett försök att fastställa kvantitativa mått på deras exponering mot olika typer av risker. Ett av dessa krav innebär att utföra stresstester för att uppskatta motståndskraften under stressade marknader/kriser. Dock förändras finansiella marknader över tiden vilket leder till att vissa instrument som handlas idag inte fanns under den dåvarande perioden, eftersom de introducerades vid ett senare tillfälle. Baserat på nuvarande rutiner så är målet med detta arbete att tillhandahålla en mer sofistikerad metod för imputation (ifyllnad) av historisk data som ett förberedande arbete i utförandet av stresstester. I denna rapport implementeras två modeller som betraktas som de bäst presterande metoderna idag, baserade på maximum likelihood estimering (MLE) och multiple imputation (MI), samt en tredje alternativ metod som involverar copulas. Modellerna tillämpas på historisk data för terminskontrakt från den nordiska energimarkanden. Genom att använda väl etablerade mätmetoder för att skatta noggrannheten för respektive modell, är det väldigt svårt (generellt) att särskilja prestandan för varje metod, eller att dra några slutsatser om hur bra vaje modell är i jämförelse med varandra. Även om Students t-fördelningen verkar (generellt) vara ett mer adekvat antagande rörande datan i jämförelse med normalfördelningen, så visar alla modeller ganska svag prestanda vid en första anblick. Däremot, genom att undersöka de betingade fördelningarna mer noggrant, för att se hur väl varje modell presterar genom att extrahera specifika kvantilvärden, kan varje metod förbättras markant. Genom att jämföra de olika modellerna (vid imputering av mer extrema kvantilvärden) kan slutsatsen dras att alla metoder producerar tillfredställande resultat, även om g-copula och t-copula modellerna verkar vara mer robusta än de motsvarande linjära modellerna.

Nyckelord: Saknad Data, Bayesiansk Statistik, Expectation Conditional Maximization (ECM), Betingad Sannolikhet, Robust Regression, MCMC, Copulas.

Acknowledgements

I would like to thank NASDAQ, and especially my supervisor at the Risk Management European Commodities department, Mikhail Nechaev, who introduced me to this subject and supported me throughout the project. Finally I would like to thank my supervisor at the Royal Institute of Technology, Camilla Johansson Landén, for feedback and guidance.

Stockholm, Mars 2016

Simon Östlund

Contents

Chapter 1

Introduction

Missing data is a widespread concern throughout all varieties of science. Researches have for a long time tried different ad hoc techniques to manage incomplete data sets, including omitting the incomplete cases and filling in the missing data. The downside of most of these methods is the requirement of a relatively strict assumption regarding the cause of the missing data, and they are hence likely to introduce bias [10].

During the 1970s some major progress was achieved within this area with the introduction of maximum likelihood estimation (MLE) methods and multiple imputation (MI). In 1976 Rubin [34] introduced an approach for managing missing data in a way that remains in extensive use today. Both the MLE and MI approach have received a lot of attention the past 40 years and are still regarded as the current state-of-the-art techniques. In comparison with more traditional methods MLE and MI are theoretically desirable since they require weaker assumptions, and will hence produce more robust estimates with less bias and greater power.

To be able to reduce bias and achieve more power when dealing with missing data it is important to know the characteristics of the missing data and its relation to the observed values. Missing data patterns and missing data mechanisms are two useful terms regarding this matter, and the latter is a cornerstone in Rubin's missing data theory. A missing data pattern refers to the way in which the data is missing and the missing data mechanism describes the relation between the distribution of the missing data and the observed variables. In other words, the missing data pattern describes where the gaps are located and the missing data mechanism puts more emphasis on why or how the data is missing, even if it does not give a causal elucidation[1]. A more thorough introduction to these concepts together with various examples of missing data patterns can be found in Enders [10]. The different types of missing data mechanisms are also explained in more detail in [34], but since they are of great importance, and form the foundation of the assumptions made in this thesis, they are stated below as they are defined in [10].

- MCAR - The formal definition of missing completely at random (MCAR) requires that the distribution of the missing data on a variable Y is unrelated to other measured variables and is unrelated to the values of Y itself.
- MAR - Data are missing at random (MAR) when the distribution of the missing data on a variable Y is related to some other measured variable (or variables) in the analysis model but not to the values of Y itself.
- MNAR - Data are missing not at random (MNAR) when the distribution of the missing data on a variable Y is related to the values of Y itself.

The aim of this thesis will be to analyse three different methods, and discern the approach best suited for imputation of missing values for a specific missing data pattern. The models considered in this paper include two techniques currently regarded as state-of-the-art methods based on maximum likelihood estimation (MLE) and multiple imputation (MI), together with a third alternative approach involving copulas. To asses the validity of the assumption regarding the distribution of the missing data both a *normal* and a *Student's t-distribution* are implemented.

[1] In the sense that it does not explicitly give a more detailed explanation to the reason why the data is missing.

All methods are applied on a real world data set and generate unbiased estimates under the MAR assumption. The performance of the different models are assessed using conventional error metrics: the mean square error (MSE), the mean absolute error (MAE) and out-of-sample likelihood.

The background of this thesis, and a presentation of the problem in its real world context, is presented in Chapter 2, followed by the most vital mathematical theory in Chapter 3. The actual implementation of each method is described in more detail in Chapter 4. In Chapter 5 the performance of each method is presented after being applied to a real data set. Finally, the conclusions, together with some interesting topics for future reference regarding imputation of missing data are found in Chapter 6.

Chapter 2

Background

2.1 The Role of a Central Counterparty Clearing House

The main responsibility of a Central Counterparty clearing house (CCP) is to ensure that all traded contracts are being honored to maintain efficiency and stability on the financial markets. A clearing house acts as a middle man and is the counterparty in all transactions. By taking both long and short positions its portfolio is always perfectly balanced [26]. The essential function of a risk management department within a CCP is to assess the risk associated with this role. Counterparty risk is the risk when one of the participants in an agreement can not fulfill its obligations in time of settlement, also known as default risk or credit risk. When a participant defaults on its obligation the clearing house needs to offset its position as soon as possible to limit its exposure to the market risk incurred.

Market risk is defined in McNeil et al. [24] as

> *"The risk of a change in the value of a financial position due to changes in the value of the underlying components on which that position depends".*

The components in this definition refer to stock and bond prices, exchange rates, commodity prices, and different types of derivative products. A technique used to analyze the exposure to market risk during various extreme scenarios is stress testing.

2.2 Stress Testing under EMIR/ESMA

In recent years additional requirements has been imposed on financial institutions, including CCPs, as an attempt to assess quantitative measures of their exposure to different types of risk. The *European Securities and Markets Authority* (ESMA) is an independent EU authority operating to maintain and improve the stability of the European Union's financial system by ensuring the consistent and correct application of the *EU's Markets Infrastructure Regulation* (EMIR) [11]. *Nasdaq Clearing AB (Nasdaq Clearing)* was approved by the *Swedish Financial Supervisory Authority* (SFSA) as a CCP under EMIR the 18th of March 2014 [17]. An important role of ESMA is to review and validate the risk models applied by the CCPs, including an annual EU-wide stress test performed in collaboration with the the European Systemic Risk Board (ESRB) [12].

Historical stress testing supposes that historically observed price changes are used to simulate a possible change of market value in the specified portfolio. The idea is to take extreme price changes observed at a date of a crisis (like the default of Lehman Brothers in 2008) and apply these changes to the current portfolio in an attempt to analyze what would be the corresponding impact on the profit/loss distribution if it should happen again.

The simplicity of the idea has a downside which makes it difficult to implement. Financial markets develop over time and this leads to a situation when some instruments traded today are not present at the chosen date because they were introduced after the considered historical event.

This thesis pays attention to the important preparatory work and analysis of the historical data used in stress testing of a financial portfolio, and more specific, the common problem regarding missing data.

2.3 Missing Data within Stress Testing

The lack of historical information may have an adverse effect on the parameter estimates and power of the considered model. Hence, interpretations and conclusions based on obtained results from the model are more likely to be erroneous. Even if stress testing is a tool used when historical data is missing, there is a chance the magnitude and the frequency of the potential losses are heavily underestimated when using incomplete data.

2.4 Current Routines

At the moment there are two simple methods used to manage the problem of missing data. The missing data are either set to zero or equal to the maximum of all observed price changes for similar instruments at the considered date.

At a first glance it may seem meaningless to impute the value zero, and at an intuitive level it would mean the same thing as not regarding the missing cases at all. However, when considering the distribution of relative returns of an instrument with an expected value close to zero it yields less variability of the data. Imputation of values near the center of the distribution could therefore have a significant effect on the variance and the standard deviation. Hence, the covariance and the correlations are also attenuated when the variability is reduced. Even if the data are MCAR this approach will result in biased parameter estimates. According to Enders [10] different studies conclude that expected value imputation is probably the least appealing approach available.

On the contrary to neglecting the missing data, by putting the missing cases equal to the maximum of all observed price changes could lead to an overestimation of the exposure to severe losses. The distribution becomes more skewed and parameter estimates are likely to be biased. Current routines hence need to be replaced by a more sophisticated method to acquire more reliable results.

2.5 Modelling Approach

The aim in this thesis is to impute missing values through the conditional distribution. This approach retains the variability of the data and hence do not affect correlations in such extent as the imputation by the corresponding conditional expected value. Another advantage of this approach is that specific quantile values are tractable and can be extracted from the considered parametric model, which is an beneficial property regarding stress testing.

Chapter 3

Mathematical Background

This chapter will introduce the most vital mathematical theory behind the algorithms used in this thesis. The idea is to present the theory in a framework as general as possible. Therefore, the implementation of each method will be thoroughly described in Chapter 4.

3.1 Maximum Likelihood Estimation (MLE)

The aim of maximum likelihood estimation (MLE) is to specify a distribution that describes the total population in a complete data set. The starting point is the corresponding probability density function $f_Y(y)$, or $f_Y(y;\boldsymbol{\theta})$, where $Y = y$ is a realization of the random variable and $\boldsymbol{\theta}$ is the vector of parameter values to be estimated. Additionally, the goal of MLE is to identify the parameter values most likely to have produced a particular sample of data, i.e the parameter values which maximize the likelihood $L_i = f_Y(y_i;\boldsymbol{\theta})$. However, the algorithm requires the entire sample to be included in the estimation, not just a single data point. The joint probability of independent events is represented by the product of the individual likelihoods

$$L(\boldsymbol{\theta};\mathbf{Y}) = L(\boldsymbol{\theta};y_1...y_N) = f(y_1...y_N;\boldsymbol{\theta}) = \prod_{i=1}^{N} f_Y(y_i;\boldsymbol{\theta}). \tag{3.1}$$

Since the product of many small numbers ($L_i \in [0,1]$) produce an even smaller number, which is prone to rounding errors and is more difficult to work with, a more common metric to use is the natural logarithm of Equation (3.1), i.e the log-likelihood

$$\log\big(L(\boldsymbol{\theta};\mathbf{Y})\big) = \sum_{i=0}^{N} \log\big(f_Y(y_i;\boldsymbol{\theta})\big).$$

The natural logarithm is a strictly increasing function which means

$$\underset{\theta\in\boldsymbol{\theta}}{\operatorname{argmax}}\big(L(\boldsymbol{\theta};\mathbf{Y})\big) = \underset{\theta\in\boldsymbol{\theta}}{\operatorname{argmax}}\Big(\log\big(L(\boldsymbol{\theta};\mathbf{Y})\big)\Big)$$

The MLE can be obtained through an iterative process which repeatedly computes the sample log-likelihood with different parameter values until convergence is reached. The process is usually ended when the difference between the previous log-likelihood and the current estimation is below a certain threshold (tolerance level), or when the process has reached a certain maximum number of iterations. However, to calculate the MLE in a more efficient manner, the first order partial derivative of the log-likelihood function can be used by letting

$$\frac{\partial \log\big(L(\boldsymbol{\theta};\mathbf{Y})\big)}{\partial\theta} = 0,$$

which is then solved for every $\theta \in \boldsymbol{\theta}$ [10].

3.2 Bayesian Inference

In probability theory there are two fundamentally different approaches regarding the interpretation of the set of parameters to be estimated. They consist of the *Frequentist inference* and the *Bayesian inference*. In the Frequentist analysis approach the data is interpreted as a repeatable random sample, and the estimated parameters to be fixed, i.e $f(\mathbf{Y};\boldsymbol{\theta})$. However, in the Bayesian framework the data is assumed to be observed from a realized sample, and the parameters to be estimated are referred to as random variables, i.e $f(\boldsymbol{\theta}|\mathbf{Y}=\mathbf{y})$. More specific, the Bayesian framework consist of three essential steps

1. specify prior distributions for the included parameters
2. assign a likelihood function to include the information held by the data regarding the different parameter values, and
3. combine the prior distributions and the likelihood to form a posterior distribution that describes the relative probability across a range of parameter values [10].

The prior distributions are used to incorporate additional information to the model that is not explained by the data alone. The intuition of assigning a uniform prior is hence that no additional information is available or used. Based on the data, the likelihood function gives information regarding the relative probability for different parameter values. Additionally, the prior information regarding the parameters and the information attained from the data is then used to form the posterior distribution of the parameters given the data via Bayes' theorem, see Equation (3.2).

$$f(\boldsymbol{\theta}|\mathbf{Y}) = \frac{f(\mathbf{Y}|\boldsymbol{\theta})f(\boldsymbol{\theta})}{f(\mathbf{Y})} \tag{3.2}$$

In Equation (3.2) the parameters of interest are given by $\boldsymbol{\theta}$, and $\mathbf{Y}$ is the sample data. The prior distribution assigned to $\boldsymbol{\theta}$ is given by $f(\boldsymbol{\theta})$, the likelihood (i.e the distribution of the data conditional on the parameter) is given by $f(\mathbf{Y}|\boldsymbol{\theta})$, and the marginal distribution of the data is given by $f(\mathbf{Y})$. The posterior distribution (i.e the distribution of the parameter conditional on the data) is then given by $f(\boldsymbol{\theta}|\mathbf{Y})$. The denominator of Equation (3.2) is just a normalizing constant used to ensure that the area under the distribution function integrates to one, in other words

$$\text{Posterior} = \frac{\text{Likelihood} \cdot \text{Prior}}{\text{Normalizing constant}}.$$

The shape of the posterior distribution is not affected by this constant, which often is difficult to calculate. Hence, a more common way to express the posterior distribution is to exclude this constant (see [10]) and write

$$\text{Posterior} \propto \text{Likelihood} \cdot \text{Prior}.$$

3.3 Expectation Conditional Maximization (ECM)

The MLE is an attractive tool when managing data. However, since the MLE is based on the first derivative of the log-likelihood function for every observed event it becomes intractable when working with missing data. Hence, another approach for the maximum likelihood estimation is needed. The expectation conditional maximization (ECM) algorithm is an updated version of the classical expectation maximization (EM) method first introduced by Dempster et al. (1977) [6], and is an appealing solution to the missing data estimation problem. The ECM algorithm functions partially under the Bayesian assumptions and will through an iterative process include the conditional probability distribution of the missing data and provide a corresponding maximum-a-posteriori (MAP) estimate of the parameter vector $\boldsymbol{\theta}$, i.e the mode value from its posterior distribution.

The EMC is a two-step iterative estimation process. The first step is the expectation step (E-step) and the second step is the conditional maximization step (CM-step). The E-step generates samples of missing (or latent) data by taking expectations over complete-data conditional distributions. Thereafter, the CM-step involves individual complete-data maximum likelihood estimation of each parameter, by conditionally keeping the other parameters fixed. The CM-step is what separates the original EM method from the ECM algorithm. In the EM approach the corresponding M-step is maximizing the likelihood function with respect to all parameters at the same time. Hence, even if the ECM method is more computer intensive in the sense of number of iterations, the overall rate of convergence may be improved. The objective of the ECM algorithm is to maximize the observed-data likelihood by iteratively maximizing the complete-data likelihood [25].

Consider the complete-data $\mathbf{Y} = (\mathbf{Y}_{obs}, \mathbf{Y}_{miss})$, where $\mathbf{Y}_{obs}$ refers to the observed-data and $\mathbf{Y}_{miss}$ the missing-data. By using Bayes' theorem, see Equation (3.2), the likelihood of the complete-data can be written as

$$f_Y(\mathbf{Y}|\boldsymbol{\theta}) = f_{Y_{obs},Y_{miss}}(\mathbf{Y}_{obs}, \mathbf{Y}_{miss}|\boldsymbol{\theta}) = f_{Y_{miss}}(\mathbf{Y}_{miss}|\boldsymbol{\theta}, \mathbf{Y}_{obs}) \cdot f_{Y_{obs}}(\mathbf{Y}_{obs}|\boldsymbol{\theta})$$

$$\Rightarrow f_{Y_{obs}}(\mathbf{Y}_{obs}|\boldsymbol{\theta}) = \frac{f_{Y_{obs},Y_{miss}}(\mathbf{Y}_{obs}, \mathbf{Y}_{miss}|\boldsymbol{\theta})}{f_{Y_{miss}}(\mathbf{Y}_{miss}|\boldsymbol{\theta}, \mathbf{Y}_{obs})},$$

where $f_{Y_{miss}}(\mathbf{Y}_{miss}|\boldsymbol{\theta}, \mathbf{Y}_{obs})$ is the conditional distribution of the missing data given the observed data. The relation between the complete-data likelihood $L(\boldsymbol{\theta}; \mathbf{Y})$ and the observed-data likelihood $L(\boldsymbol{\theta}; \mathbf{Y}_{obs})$ is then given by

$$\log\big(L(\boldsymbol{\theta}; \mathbf{Y}_{obs})\big) = \log\big(L(\boldsymbol{\theta}; \mathbf{Y})\big) - \log\big(f_{Y_{miss}}(\mathbf{Y}_{miss}|\boldsymbol{\theta}, \mathbf{Y}_{obs})\big). \tag{3.3}$$

The E-step now yields taking the expectation of Equation (3.3) with respect to $f_{Y_{miss}}(\mathbf{Y}_{miss}|\boldsymbol{\theta}, \mathbf{Y}_{obs})$, i.e

$$\log\big(L(\boldsymbol{\theta}; \mathbf{Y}_{obs})\big) = \mathbb{E}_{\boldsymbol{\theta}_0}\Big[\log\big(L(\boldsymbol{\theta}; \mathbf{Y})\big)\Big] - \mathbb{E}_{\boldsymbol{\theta}_0}\Big[\log\big(f_{Y_{miss}}(\mathbf{Y}_{miss}|\boldsymbol{\theta}, \mathbf{Y}_{obs})\big)\Big], \tag{3.4}$$

for any initial value $\boldsymbol{\theta}_0$. Since the expectation is taken with respect to the missing data $\mathbf{Y}_{miss}$, the left hand side is by definition deterministic and the expectation is omitted. A common ECM notation for the expected log-likelihood (see [31]) is

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}_0, \mathbf{Y}_{obs}) = \mathbb{E}_{\boldsymbol{\theta}_0}\Big[\log\big(L(\boldsymbol{\theta}; \mathbf{Y})\big)\Big]. \tag{3.5}$$

The consecutive CM-step then maximizes Equation (3.5) by

$$Q\Big(\hat{\boldsymbol{\theta}}_{(m+1)}\Big|\hat{\boldsymbol{\theta}}_{(m)}, \mathbf{Y}_{obs}\Big) = \max_{\boldsymbol{\theta}}\Big[Q\Big(\boldsymbol{\theta}\Big|\hat{\boldsymbol{\theta}}_{(m)}, \mathbf{Y}_{obs}\Big)\Big]. \tag{3.6}$$

If $\boldsymbol{\theta}_m = \Big(\theta^{(1)}_{(m)}, ..., \theta^{(N)}_{(m)}\Big)$ is the parameter estimate at iteration m, $m = 0, ..., M$, the ECM sequentially maximize Equation (3.5) through the following scheme

$$Q\Big(\hat{\theta}^{(1)}_{(m+1)}\Big|\hat{\theta}^{(1)}_{(m)},...,\hat{\theta}^{(k)}_{(m)},...,\hat{\theta}^{(N)}_{(m)},\mathbf{Y}_{obs}\Big) = \max_{\theta^{(1)}}\Big[Q\Big(\theta^{(1)}\Big|\hat{\theta}^{(1)}_{(m)},...,\hat{\theta}^{(k)}_{(m)},...,\hat{\theta}^{(N)}_{(m)},\mathbf{Y}_{obs}\Big)\Big]$$
$$\vdots$$
$$Q\Big(\hat{\theta}^{(2)}_{(m+1)}\Big|\hat{\theta}^{(1)}_{(m+1)},\hat{\theta}^{(2)}_{(m)},...,\hat{\theta}^{(k)}_{(m)},...,\hat{\theta}^{(N)}_{(m)},\mathbf{Y}_{obs}\Big) = \max_{\theta^{(2)}}\Big[Q\Big(\theta^{(2)}\Big|\hat{\theta}^{(1)}_{(m+1)},\hat{\theta}^{(2)}_{(m)}...,\hat{\theta}^{(k)}_{(m)},...,\hat{\theta}^{(N)}_{(m)},\mathbf{Y}_{obs}\Big)\Big]$$
$$\vdots$$
$$Q\Big(\hat{\theta}^{(N)}_{(m+1)}\Big|\hat{\theta}^{(1)}_{(m+1)},...,\hat{\theta}^{(k)}_{(m+1)},...,\hat{\theta}^{(N-1)}_{(m+1)},\hat{\theta}^{(N)}_{(m)},\mathbf{Y}_{obs}\Big) = \max_{\theta^{(N)}}\Big[Q\Big(\theta^{(N)}\Big|\hat{\theta}^{(1)}_{(m+1)},...,\hat{\theta}^{(k)}_{(m+1)},...,\hat{\theta}^{(N-1)}_{(m+1)},\hat{\theta}^{(N)}_{(m)},\mathbf{Y}_{obs}\Big)\Big].$$

In other words, the ECM algorithm maximize Equation (3.5) by first maximizing with respect to $\theta^{(1)}_{(m)}$ keeping all other parameters constant. The fundamental in the theory behind the ECM algorithm is that the observed likelihood in Equation (3.4) is increased by maximizing $Q(\boldsymbol{\theta}|\boldsymbol{\theta}_0,\mathbf{Y}_{obs})$ in each iteration. The following theorem and proof is referred to by Robert et al. in [31], but were first stated by Dempster et al. (1977) [6].

Theorem 3.3.1. *The sequence* $\Big(\hat{\theta}_{(m)}\Big)$ *defined by Equation* (3.6) *satisfies*

$$L\Big(\hat{\theta}_{(m+1)}\Big|\mathbf{Y}_{obs}\Big) \geq L\Big(\hat{\theta}_{(m)}\Big|\mathbf{Y}_{obs}\Big)$$

with equality holding if and only if $Q\Big(\hat{\theta}_{(m+1)}\Big|\hat{\theta}_{(m)},\mathbf{Y}_{obs}\Big) = Q\Big(\hat{\theta}_{(m)}\Big|\hat{\theta}_{(m)},\mathbf{Y}_{obs}\Big)$.

Proof. On successive iterations, it follows from the definition of $\hat{\theta}_{m+1}$ that

$$Q\Big(\hat{\theta}_{(m+1)}\Big|\hat{\theta}_{(m)},\mathbf{Y}_{obs}\Big) \geq Q\Big(\hat{\theta}_{(m)}\Big|\hat{\theta}_{(m)},\mathbf{Y}_{obs}\Big).$$

Hence, if it can be showed that

$$\mathbb{E}_{\hat{\theta}_{(m)}}\Big[\log\Big(f_{Y_{miss}}\Big(\mathbf{Y}_{miss}\Big|\hat{\theta}_{(m+1)},\mathbf{Y}_{obs}\Big)\Big)\Big] \leq \mathbb{E}_{\hat{\theta}_{(m)}}\Big[\log\Big(f_{Y_{miss}}\Big(\mathbf{Y}_{miss}\Big|\hat{\theta}_{(m)},\mathbf{Y}_{obs}\Big)\Big)\Big], \quad (3.7)$$

it will follow from Equation (3.4) that the value of the likelihood is increased at each iteration. By defining the *Shannon entropy*, first introduced by Shannon [36] in 1948, as

$$\begin{aligned} H\Big(\hat{\theta}_{(m+1)}\Big|\hat{\theta}_{(m)}\Big) &= -\mathbb{E}_{\hat{\theta}_{(m)}}\Big[\log\Big(f_{Y_{miss}}\Big(\mathbf{Y}_{miss}\Big|\hat{\theta}_{(m+1)},\mathbf{Y}_{obs}\Big)\Big)\Big] \\ &= -\int_{\mathbf{Y}_{miss}}\log\Big(f_{Y_{miss}}\Big(\mathbf{Y}_{miss}\Big|\hat{\theta}_{(m+1)},\mathbf{Y}_{obs}\Big)\Big)\cdot f_{Y_{miss}}\Big(\mathbf{Y}_{miss}\Big|\hat{\theta}_{(m)},\mathbf{Y}_{obs}\Big)\mathrm{d}\mathbf{Y}_{miss}, \end{aligned}$$

and by using the *Gibbs inequality*[2] [2]

[2] For a continuous probability distribution $F=\int f_X(x)dx$, and any other continuous probability distribution $Q=\int q_X(x)dx$, the Gibbs inequality states

$$-\int\log(f_X(x))\cdot f_X(x)dx \leq -\int\log(q_X(x))\cdot f_X(x)dx,$$

with equality only if $f=q$.

$$-\int_{\mathbf{Y}_{miss}} \log\Big(f_{Y_{miss}}\Big(\mathbf{Y}_{miss}\Big|\hat{\theta}_{(m)}, \mathbf{Y}_{obs}\Big)\Big) \cdot f_{Y_{miss}}\Big(\mathbf{Y}_{miss}\Big|\hat{\theta}_{(m)}, \mathbf{Y}_{obs}\Big)\mathrm{d}\mathbf{Y}_{miss}$$

$$\leq$$

$$-\int_{\mathbf{Y}_{miss}} \log\Big(f_{Y_{miss}}\Big(\mathbf{Y}_{miss}\Big|\hat{\theta}_{(m+1)}, \mathbf{Y}_{obs}\Big)\Big) \cdot f_{Y_{miss}}\Big(\mathbf{Y}_{miss}\Big|\hat{\theta}_{(m)}, \mathbf{Y}_{obs}\Big)\mathrm{d}\mathbf{Y}_{miss}$$

$$\Rightarrow$$

$$H\Big(\hat{\theta}_{(m)}\Big|\hat{\theta}_{(m)}\Big) \leq H\Big(\hat{\theta}_{(m+1)}\Big|\hat{\theta}_{(m)}\Big),$$

the theorem is established. □

However, Theorem 3.3.1 does not confirm that the sequence $\hat{\theta}_{(m)}$ converges to its corresponding MLE estimate. The following complementing theorem is given by Roberts et al. [31] to ensure convergence to a stationary point, i.e either a local maximum or a saddle point.

Theorem 3.3.2. *If the expected complete-data likelihood $Q\big(\theta|\theta_0, \mathbf{Y}_{obs}\big)$ is continuous in both θ and θ_0, then every limit point of an EM sequence $\hat{\theta}_{(m)}$ is a stationary point of $L(\theta|\mathbf{Y}_{obs})$ and $L\Big(\hat{\theta}_{(m)}\Big|\mathbf{Y}_{obs}\Big)$ converges monotonically to $L\Big(\hat{\theta}\Big|\mathbf{Y}_{obs}\Big)$ for some stationary point $\hat{\theta}$.*

Be aware that Theorem 3.3.2 only guarantees convergence to a stationary point and not a global maximum. This problem can be solved by running the ECM algorithm a number of times using different initial values [31].

In Appendix A a short example [28] of the derivation of the ECM updating formula for the one parameter case can be found.

3.4 Regression Analysis

The basic idea of the linear regression model is to build a set of regression equations to estimate the conditional expectation of a set of random variables $Y = (Y_1, ..., Y_n)$, see Equation (3.8).

$$y_i = \sum_{j=0}^{k} x_{i,j}\beta_j + e_i, i = 1, ..., N. \tag{3.8}$$

The predicted value y_i is a realisation of the dependent random variable Y_i whose value depends on the explanatory variables (covariates) $x_{i,j}$. In contrast to the observed covariates, the residuals, e_i, are random variables and are assumed to be independent between observations, and with conditional expected value and variance

$$\mathbb{E}[e_i|x] = 0 \text{ and } \mathbb{E}\Big[e_i^2\Big|x\Big] = \sigma^2,$$

where σ is typically unknown [22]. However, this representation of the residuals is under the assumption of *homoskedasticity*[3]. This is often an unverified assumption, and in many cases *heteroskedasticity*[4] is a more appropriate premise [16].

To bring more power to the estimated regression model it is strongly recommended to construct a regression equation with high correlation between the explanatory variables and the dependent variable, but with little to none correlation among the covariates themselves. Additionally, the presence of *multicollinearity*[5] can otherwise induce uncertainty regarding the inference to be

[3] Homoskedasticity is when $\mathbb{E}\Big[e_i^2\Big|x\Big] = \sigma^2$, i.e the residual variance does not depend on the observation.

[4] Heteroskedasticity is when $\mathbb{E}\Big[e_i^2\Big|x\Big] = \sigma_i^2$, i.e the residual variance depends on the observation.

[5] Multicollinearity arises when two or more covariates are perfectly correlated.

made about the model, even if the phenomenon in general does not reduce power or influence the reliability of the model [22].

However, the first covariate, $x_{i,0}$, is usually set to 1. This means that the first coefficient β_0, called the intercept, in (3.8) acts as a constant term. Hence, in contrast to the other coefficients the intercept is somewhat hard to interpret since it does not indicate how much a specific covariate contributes to the estimation of the dependent variable. However, the intercept is essential and is included to reduce bias by shifting the regression line, and therefore guaranteeing that the conditional expected value of the residual is $\mathbb{E}[e_i|x] = 0$ [15].

3.5 Copulas

The name *copula* is a Latin noun that means a "link, tie, bond" [27]. In probability theory a copula is a function used to describe and model the relation between one-dimensional distribution functions, when the corresponding joint distribution is intractable. The name and theory behind copulas (in the mathematical context) was first introduced by Sklar in 1959. Still, an informative introduction to copulas (which is some what easy to digest) can be found in Embrechts et al. [8], in Lindskog et al. [7] and in Embrechts [9]. More extensive literature concerning this subject is presented by Nelsen [27] and by Dall'Aglio et al. [4]. However, the theorems and corresponding proofs behind the theory of copulas in this paper is stated as in McNeil et al. [24].

A copula is defined in McNeil et al. [24] as:

Definition 3.5.1. *A d-dimensional copula is a distribution function on* $[0,1]^d$ *with standard uniform marginal distributions. The notation* $C(u) = C(u_1, ..., u_d)$ *is reserved for the multivariate distribution functions that are copulas. Hence C is a mapping of the form* $C : [0,1]^d \rightarrow [0,1]$*, i.e. a mapping of the unit hypercube into the unit interval. The following three properties must hold.*

1. $C(u_1, ..., u_d)$ is increasing in each component u_i.
2. $C(1, ..., 1, u_i, 1, ..., 1) = u_i$ for all $i \in 1, ..., d, u_i \in [0,1]$.
3. For all $(a_1, ..., a_d), (b_1, ..., b_d) \in [0,1]^d$ with $a_i \leq b_i$, the following must apply
$$\sum_{i_1=1}^{2} ... \sum_{i_d=1}^{2} (-1)^{i_1+...+i_d} C(u_{1i_1}, ..., u_{di_d}) \geq 0, \tag{3.9}$$
where $u_{j1} = a_j$ and $u_{j2} = b_j$ for all $j \in 1, ..., d$.

The first property is clearly required of any multivariate distribution function and the second property is the requirement of uniform marginal distributions. The third property is less obvious, but the so-called rectangle inequality in Equation (3.9) *ensures that if the random vector* $(U_1, ..., U_d)'$ *has distribution function C, then* $P(a_1 \leq U_1 \leq b_1, ..., a_d \leq U_d \leq b_d)$ *is non-negative. These three properties characterize a copula; if a function C fulfills them, then it is a copula.*

Before stating the famous *Sklar's Theorem*, which is of great importance in the the study of multivariate distribution functions, the operations of probability and quantile transformation together with the properties of generalized inverses must be introduced, see Proposition 3.5.1. The proof and the underlying theory to Proposition 3.5.1 can be found in [24].

Proposition 3.5.1. *Let G be a distribution function and let* $G^{\leftarrow}$ *denote its generalized inverse, i.e. the function* $G^{\leftarrow}(y) = inf\{x : G(x) \geq y\}$.

1. **Quantile transformation**. If $U \sim U(0,1)$ has a standard uniform distribution, then $P(G^{\leftarrow}(U) \leq x) = G(x)$.
2. **Probability transformation**. If Y has distribution function G, where G is a continuous univariate distribution function, then $G(Y) \sim U(0,1)$.

Finally, the following theorem stated by Sklar in 1959 emphasize that all multivariate distribution functions contain copulas, and how copulas can be used in conjunction with univariate distribution functions to form multivariate distribution functions. The proof of the theorem can be found in Appendix B.

Theorem 3.5.1. ***(Sklar 1959)*** *Let F be a joint distribution function with margins $F_1, ..., F_d$. Then there exists a copula $C : [0,1]^d \to [0,1]$ such that, for all $x_1, ..., x_d$ in $\bar{\mathbb{R}} = [-\infty, \infty]$,*

$$F(x_1, ..., x_d) = C\big(F_1(x_1), ..., F_d(x_d)\big) \tag{3.10}$$

If the margins are continuous, then C is unique; otherwise C is uniquely determined on $\mathrm{Ran}F_1 \times \mathrm{Ran}F_2 \times ... \times \mathrm{Ran}F_d$, where $\mathrm{Ran}F_i = F_i\big(\bar{\mathbb{R}}\big)$ denotes the range of F_i. Conversely, if C is a copula and $F_1, ..., F_d$ are univariate distribution functions, then the function F defined in Equation (3.10) is a joint distribution function with margins $F_1, ..., F_d$.

The copula of a distribution is invariant under strictly increasing transformations of the marginals. This is a very useful property since in conjunction with Sklar's Theorem (Theorem 3.5.1) the copula of a distribution then becomes a very powerful and dynamic tool to represent the dependence structure of that distribution, see Proposition 3.5.2.

Proposition 3.5.2. *Let $(X_1, ..., X_d)$ be a random vector with continuous margins and copula C and let $T_1, ..., T_d$ be strictly increasing functions. Then $\big(T_1(X_1), ..., T_d(X_d)\big)$ also has copula C.*

The proof and the underlying theory to Proposition 3.5.2 can be found in [24].

3.5.1 Elliptical Distributions

The multivariate probability distributions used to sample from the corresponding copula in this thesis belongs to the family of elliptical distributions. The following definition, theorem and example of the elliptical distribution is stated as in [7].

Definition 3.5.2. *If $\mathbf{X}$ is a d-dimensional random vector and, for some $\mu \in \mathbb{R}^d$ and some $d \times d$ non-negative definite, symmetric matrix Σ, the characteristic function $\varphi_{\mathbf{X}-\mu(\mathbf{t})}$ of $\mathbf{X} - \mu$ is a function of the quadratic form $\mathbf{t}^T\Sigma\mathbf{t}$, $\varphi_{\mathbf{X}-\mu(\mathbf{t})} = \phi(\mathbf{t}^T\Sigma\mathbf{t})$, then $\mathbf{X}$ has an elliptical distribution with parameters μ, Σ and ϕ, and is denoted $\mathbf{X} \sim E_d(\mu, \Sigma, \phi)$.*

Remark. *If $d = 1$, the class of elliptical distributions coincides with the class of one-dimensional symmetric distributions. A function ϕ is called a characteristic generator.*

Theorem 3.5.2. $\mathbf{X} \sim E_d(\mu, \Sigma, \phi)$ *with $rank(\Sigma) = k$ if and only if there exist a random variable* R ≥ 0 *independent of $\mathbf{U}$, a k-dimensional random vector uniformly distributed on the unit hypersphere $\mathbf{z} \in \mathbb{R}^k | \mathbf{z}^T\mathbf{z} = 1$, and an $d \times k$ matrix A with $AA^T = \Sigma$, such that*

$$X \overset{d}{=} \mu + RA\mathbf{U}.$$

Example 3.5.1. Let $\mathbf{X} \sim \mathcal{N}_d(0, \mathbf{I}_d)$. Since the components $X_i \sim \mathcal{N}(0,1)$, $i = 1, ..., d$, are independent and the characteristic function of X_i is $\exp\Big(-\frac{t_i^2}{2}\Big)$, the characteristic function of $\mathbf{X}$ is

$$\exp\Big(-\frac{1}{2}\Big(t_1^2 + \cdots + t_2^d\Big)\Big) = \exp\Big(-\frac{1}{2}\mathbf{t}^T\mathbf{t}\Big).$$

From Theorem 3.5.2 it then follows that $\mathbf{X} \sim E_d(0, \mathbf{I}_d, \phi)$, where $\phi(u) = \exp\big(-\frac{u}{2}\big)$.

Chapter 4

Methodology

In this chapter the mathematical theory from Chapter 3 is implemented in three different methods for imputation of missing data. More information regarding the underlying data and the corresponding results are presented in Chapter 5 where they are put in the context of the solution of a real world problem.

4.1 Stochastic Regression Imputation

Stochastic regression is, as the name implies, an evolved version of the ordinary regression analysis. In resemblance with the expected value imputation, described in Chapter 2, ordinary regression imputation also suffers from lost variability and is therefore likely to induce bias. Stochastic regression modelling effectively restores this variability in the data by adding a random variable to the regression equation

$$\hat{y}_i = \sum_{j=0}^{k} x_{i,j}\hat{\beta}_j + z_i,\ i = 1, ..., N, \tag{4.1}$$

where $\hat{y}_i$ is the estimate of the predicted variable, $x_{i,j}$ is the known covariate, $\hat{\beta}_j$ is the corresponding parameter estimate and z_i is the added noise variable. The matrix representation of Equation (4.1) is given by

$$\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{z}, \tag{4.2}$$

where

$$\hat{\mathbf{Y}} = \begin{bmatrix} \hat{y}_1 \\ \vdots \\ \hat{y}_N \end{bmatrix},\ \mathbf{X} = \begin{bmatrix} 1 & x_{1,1} & \dots & x_{1,j} & \dots & x_{1,k} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 1 & x_{N,1} & \dots & x_{N,j} & \dots & x_{N,k} \end{bmatrix}, \hat{\boldsymbol{\beta}} = \begin{bmatrix} \hat{\beta}_0 \\ \vdots \\ \hat{\beta}_k \end{bmatrix},\ \text{and } \mathbf{z} = \begin{bmatrix} z_1 \\ \vdots \\ z_N \end{bmatrix}.$$

The coefficients in Equation (4.2) are estimated using the ECM algorithm, which is introduced in Chapter 3. In contrast to the widespread use of the *Ordinary Least Squares (OLS)* method, the ECM algorithm make use of information attained from the missing data in the estimation process.

Two different assumptions regarding the distribution of the data is used during the estimation and imputation phases. First a normality assumption is applied by regarding the data as realisations from a *normal* distribution as in Enders (2010) [10]. However, since the data is known to be heavy tailed, this procedure is further developed by also implementing the assumption that the data belongs to a *Student's t*-distribution (t-distribution).

The parameters are estimated using a full Bayesian implementation of the ECM algorithm. In other words, the posterior distribution is derived by specifying prior distributions for the model parameters and assigning a likelihood function for the data as in Section 3.2. However, to make the derivation easier to follow the likelihood function is first assigned to the data, and then the prior distributions are specified for each parameter.

4.1.1 The Likelihood Function

Consider the complete d-variate data vector $\mathbf{y}_i$. Next, the assumption that the complete data belongs to a multivariate t-distribution

$$\begin{aligned} \mathbf{y}_i|\nu,\boldsymbol{\beta},\boldsymbol{\Psi} &\sim t(\nu,\mathbf{X}_{i,0:k}\boldsymbol{\beta},\boldsymbol{\Psi}) \\ f_{\mathbf{y}_i}(\mathbf{y}_i|\nu,\boldsymbol{\beta},\boldsymbol{\Psi}) &\propto |\boldsymbol{\Psi}|^{\frac{1}{2}}\left(1+\frac{1}{\nu}(\mathbf{y}_i-\mathbf{X}_{i,0:k}\boldsymbol{\beta})'\boldsymbol{\Psi}(\mathbf{y}_i-\mathbf{X}_{i,0:k}\boldsymbol{\beta})\right)^{-\frac{\nu+d}{2}}, \end{aligned} \tag{4.3}$$

with shape matrix[6] $\boldsymbol{\Psi}$, is applied by using a mixture of normal distributions together with a prior distribution of an auxiliary weight parameter w_i. More specific, the corresponding prior distribution of the weight vector and the normal mixture assigned to the observations are given by

$$\begin{aligned} w_i|\nu &\sim Gamma\left(\frac{\nu}{2},\ \frac{\nu}{2}\right)=\frac{\left(\frac{\nu}{2}\right)^{\left(\frac{\nu}{2}\right)}}{\Gamma\left(\frac{\nu}{2}\right)}(w_i)^{\frac{\nu}{2}-1}\mathrm{e}^{-\frac{\nu}{2}w_i} \\ f_{w_i}(w_i|\nu) &\propto (w_i)^{\frac{\nu}{2}-1}\mathrm{e}^{-\frac{\nu}{2}w_i} \end{aligned} \tag{4.4}$$

and

$$\begin{aligned} \mathbf{y}_i|\nu,\boldsymbol{\beta},\boldsymbol{\Psi},w_i &\sim \mathcal{N}\left(\mathbf{X}_{i,0:k}\boldsymbol{\beta},(w_i\boldsymbol{\Psi})^{-1}\right)=\frac{|w_i\boldsymbol{\Psi}|^{\frac{1}{2}}}{\sqrt{2\pi}}\mathrm{e}^{-\frac{w_i}{2}(\mathbf{y}_i-\mathbf{X}_{i,0:k}\boldsymbol{\beta})'\boldsymbol{\Psi}(\mathbf{y}_i-\mathbf{X}_{i,0:k}\boldsymbol{\beta})} \\ f_{\mathbf{y}_i}(\mathbf{y}_i|\nu,\boldsymbol{\beta},\boldsymbol{\Psi},w_i) &\propto |w_i\boldsymbol{\Psi}|^{\frac{1}{2}}\mathrm{e}^{-\frac{w_i}{2}(\mathbf{y}_i-\mathbf{X}_{i,0:k}\boldsymbol{\beta})'\boldsymbol{\Psi}(\mathbf{y}_i-\mathbf{X}_{i,0:k}\boldsymbol{\beta})}, \end{aligned}$$

where $|\cdot|$ is the determinant and $\Gamma(\cdot)$ is the *Gamma function.* If the $N\times d$ vector $\mathbf{Y}=\left[\mathbf{y}^1,...,\mathbf{y}^i,...,\mathbf{y}^N\right]'$ of observations is assumed to be conditionally independent, the total likelihood is

$$\begin{aligned} f_{\mathbf{Y}}(\mathbf{Y}|\nu,\boldsymbol{\beta},\boldsymbol{\Psi},\boldsymbol{\omega}) &= \prod_{i=1}^{N} f_{\mathbf{y}_i}(\mathbf{y}_i|\nu,\boldsymbol{\beta},\boldsymbol{\Psi},w_i) \\ &\propto |\boldsymbol{\Psi}|^{\frac{N}{2}}\cdot\prod_{i=1}^{N}(w_i)^{\frac{d}{2}}\cdot\mathrm{e}^{-\frac{1}{2}\sum_{i=1}^{N}w_i(\mathbf{y}_i-\mathbf{X}_{i,0:k}\boldsymbol{\beta})'\boldsymbol{\Psi}(\mathbf{y}_i-\mathbf{X}_{i,0:k}\boldsymbol{\beta})}. \end{aligned}$$

In consistency with Section 3.2 the sign $\propto$ means "proportional to" and indicates that some normalizing constants have been omitted [30].

4.1.2 The Prior Distribution

The parameter $\boldsymbol{\beta}$ is assigned a uniform prior distribution (see [23]) given by

$$\boldsymbol{\beta}\propto constant.$$

The specified joint prior distribution for the parameter $\boldsymbol{\Psi}$ is given by an uninformative version of the *inverse Wishart distribution* with density

$$\begin{aligned} \boldsymbol{\Psi} &\sim \frac{|\boldsymbol{\Lambda}|^{\frac{\nu}{2}}}{2^{\frac{\nu\cdot d}{2}}\Gamma_d\left(\frac{\nu}{2}\right)}|\boldsymbol{\Psi}|^{-\frac{\nu+d+1}{2}}\mathrm{e}^{-\frac{1}{2}\mathrm{tr}\left(\boldsymbol{\Lambda}\boldsymbol{\Psi}^{-1}\right)}, \\ \boldsymbol{\Psi} &\propto |\boldsymbol{\Psi}|^{-\frac{\nu+d+1}{2}}\mathrm{e}^{-\frac{1}{2}\mathrm{tr}\left(\boldsymbol{\Lambda}\boldsymbol{\Psi}^{-1}\right)}, \end{aligned} \tag{4.5}$$

[6] The relation between the MAP estimate of the shape matrix and the covariance matrix is for the normal distribution given by $\hat{\boldsymbol{\Sigma}}=\hat{\boldsymbol{\Psi}}^{-1}$, and for the t-distribution $\hat{\boldsymbol{\Sigma}}=\frac{\nu}{\nu-2}\hat{\boldsymbol{\Psi}}^{-1}$, $(\nu>2)$.

where $\Gamma_d\left(\frac{\nu}{2}\right)$ is the *multivariate gamma function* and $\mathrm{tr}(\cdot)$ is the *trace* [7]. In this case the word uninformative means that no initial assumptions regarding the distribution of $\boldsymbol{\Psi}$ are applied. More specific, setting $\nu = 0$ and $\boldsymbol{\Lambda} = 0$ in Equation (4.5) is akin to say that no additional information regarding the distribution is used [10]. This means that the improper[8] prior distribution of $\boldsymbol{\Psi}$ (see [30]) is proportional to

$$f_{\boldsymbol{\Psi}}(\boldsymbol{\Psi}) \propto |\boldsymbol{\Psi}|^{-\frac{d+1}{2}}.$$

4.1.3 The Posterior Distribution

By using Bayes' theorem (see Equation (3.2)) and assuming that $\boldsymbol{\omega}, \boldsymbol{\beta}, \boldsymbol{\Psi}$ are a-priori independent finally gives the posterior distribution

$$\begin{aligned} f_{\nu,\boldsymbol{\beta},\boldsymbol{\Psi},\boldsymbol{\omega}}(\nu,\boldsymbol{\beta},\boldsymbol{\Psi},\boldsymbol{\omega}|\mathbf{Y}) &\propto f_{\mathbf{Y}}(\mathbf{Y}|\nu,\boldsymbol{\beta},\boldsymbol{\Psi},\boldsymbol{\omega}) \cdot f_{\boldsymbol{\beta}}(\boldsymbol{\beta}) \cdot f_{\boldsymbol{\Psi}}(\boldsymbol{\Psi}) \cdot f_{\boldsymbol{\omega}}(\boldsymbol{\omega}|\nu) \\ &\propto |\boldsymbol{\Psi}|^{\frac{N-d-1}{2}} \prod_{i=1}^{N} (w_i)^{\frac{d+\nu}{2}-1} \mathrm{e}^{-\frac{1}{2}\sum_{i=1}^{N} w_i\left((\mathbf{y}_i-\mathbf{X}_{i,0:k}\boldsymbol{\beta})'\boldsymbol{\Psi}(\mathbf{y}_i-\mathbf{X}_{i,0:k}\boldsymbol{\beta})+\nu\right)} \end{aligned} \tag{4.6}$$

Given the full posterior distribution in Equation (4.6), the posterior conditional distribution for each parameter can be derived.

The posterior distribution of the weights, conditional on $(\mathbf{Y}, \nu, \boldsymbol{\beta}, \boldsymbol{\Psi})$, is attained by omitting everything in Equation (4.6) not depending on $\boldsymbol{\omega}$, i.e

$$f_{\boldsymbol{\omega}}(\boldsymbol{\omega}|\mathbf{Y},\nu,\boldsymbol{\beta},\boldsymbol{\Psi}) \propto \prod_{i=1}^{N} (w_i)^{\frac{d+\nu}{2}-1} \cdot \mathrm{e}^{-\frac{1}{2}\sum_{i=1}^{N} w_i\left((\mathbf{y}_i-\mathbf{X}_{i,0:k}\boldsymbol{\beta})'\boldsymbol{\Psi}(\mathbf{y}_i-\mathbf{X}_{i,0:k}\boldsymbol{\beta})+\nu\right)}. \tag{4.7}$$

By inspection, the Equation (4.7) is proportional to the closed form expression of the gamma distribution (see Equation (4.4)), hence

$$\begin{aligned} w_i|\mathbf{y}_i,\nu,\boldsymbol{\beta},\boldsymbol{\Psi} &\sim Gamma\left(\frac{d+\nu}{2}, \frac{(\mathbf{y}_i-\mathbf{X}_{i,0:k}\boldsymbol{\beta})'\boldsymbol{\Psi}(\mathbf{y}_i-\mathbf{X}_{i,0:k}\boldsymbol{\beta})+\nu}{2}\right) \\ f_{\boldsymbol{\omega}}(\boldsymbol{\omega}|\mathbf{Y},\nu,\boldsymbol{\beta},\boldsymbol{\Psi}) &= \prod_{i=1}^{N} f(w_i|\mathbf{y}_i,\nu,\boldsymbol{\beta},\boldsymbol{\Psi}). \end{aligned}$$

The corresponding conditional expectation of the weights is then given by

$$\mathrm{E}\big[w_i\big|\mathbf{y}_i,\nu,\boldsymbol{\beta},\boldsymbol{\Psi}\big] = \frac{d+\nu}{(\mathbf{y}_i-\mathbf{X}_{i,0:k}\boldsymbol{\beta})'\boldsymbol{\Psi}(\mathbf{y}_i-\mathbf{X}_{i,0:k}\boldsymbol{\beta})+\nu}. \tag{4.8}$$

Continuing with the posterior conditional distribution of the shape matrix $\boldsymbol{\Psi}$, and comparing Equation (4.6) with the *Wishart distribution*

$$\frac{|\boldsymbol{\Psi}|^{\frac{\nu-d-1}{2}} \mathrm{e}^{-\frac{1}{2}\mathrm{tr}\left(\boldsymbol{\Lambda}^{-1}\boldsymbol{\Psi}\right)}}{2^{\frac{\nu\cdot d}{2}} |\boldsymbol{\Lambda}|^{\frac{\nu}{2}} \Gamma_d\left(\frac{\nu}{2}\right)},$$

gives

[7] The trace of an $n \times n$ square matrix A is defined to be the sum of the elements on the main diagonal of A, i.e $\mathrm{tr}(A) = a_{11} + a_{22} + ... + a_{nn} = \sum_{i=1}^{n} a_{ii}$

[8] Improper in the sense that it does not integrate to 1.

$$\boldsymbol{\Psi}|\mathbf{Y},\nu,\boldsymbol{\beta},\boldsymbol{\omega} \sim Wishart(\boldsymbol{\Lambda}^{-1},N)$$
$$f_{\boldsymbol{\Psi}}(\boldsymbol{\Psi}|\mathbf{Y},\nu,\boldsymbol{\beta},\boldsymbol{\omega}) \propto |\boldsymbol{\Psi}|^{\frac{N-d-1}{2}} \cdot \mathrm{e}^{-\frac{1}{2}\mathrm{tr}(\boldsymbol{\Psi}\boldsymbol{\Lambda})},$$

where

$$\sum_{i=1}^{N} w_i(\mathbf{y}_i - \mathbf{X}_{i,0:k}\boldsymbol{\beta})\boldsymbol{\Psi}(\mathbf{y}_i - \mathbf{X}_{i,0:k}\boldsymbol{\beta})' = \mathrm{tr}(\boldsymbol{\Psi}\boldsymbol{\Lambda})$$

and

$$\boldsymbol{\Lambda} = \sum_{i=1}^{N} w_i(\mathbf{y}_i - \mathbf{X}_{i:k}\boldsymbol{\beta})(\mathbf{y}_i - \mathbf{X}_{i,0:k}\boldsymbol{\beta})'.$$

The MAP estimate of the posterior conditional distribution of the shape matrix $\boldsymbol{\Psi}$ is then given by

$$\mathsf{mode}(\boldsymbol{\Psi}|\mathbf{Y},\nu,\boldsymbol{\beta},\boldsymbol{\omega}) = (N-d-1)\cdot\boldsymbol{\Lambda}^{-1}, \tag{4.9}$$

for $N > d-1$. Finally, by examining Equation (4.6) it can be seen that the posterior conditional distribution of the coefficient vector $\boldsymbol{\beta}$ is normal (see [30]), and the MAP estimate is given by

$$\mathsf{mode}(\boldsymbol{\beta}|\mathbf{Y},\nu,\boldsymbol{\Psi},\boldsymbol{\omega}) = \left(\sum_{i=1}^{N} w_i(\mathbf{X}_{i,0:k})'\boldsymbol{\Psi}\mathbf{X}_{i,0:k}\right)^{-1}\sum_{i=1}^{N} w_i(\mathbf{X}_{i,0:k})'\boldsymbol{\Psi}\mathbf{y}_i. \tag{4.10}$$

4.1.4 Imputation Procedure

The iterative ECM method is used to attain the MAP estimates of the corresponding posterior conditional distributions. However, until now the estimation of the degrees of freedom parameter, ν, has not been considered. In comparison with the other parameters the estimation of the degrees of freedom parameter, ν, is a bit more tedious.

Estimation of the Degrees of Freedom Parameter, ν

Given the marginal distribution of the **observed** weights (see Equation (4.4)) the corresponding log-likelihood of ν, ignoring constants, is given by

$$L_{Gamma}(\nu|\boldsymbol{\omega}) \propto -N\cdot\ln\left(\Gamma\left(\frac{\nu}{2}\right)\right) + \frac{N\cdot\nu}{2}\cdot\ln\left(\frac{\nu}{2}\right) + \frac{\nu}{2}\cdot\sum_{i=1}^{N}\Big(\ln(w_i) - w_i\Big). \tag{4.11}$$

However, the representation in Equation (4.11) is assuming that the weights, w_i, are known. Hence, when the weights are unknown the last term in Equation (4.11) is replaced by its conditional expectation, i.e by

$$\begin{aligned}\mathrm{E}\left[\sum_{i=1}^{N}\Big(\ln(w_i) - w_i\Big)|\mathbf{Y},\nu,\boldsymbol{\beta},\boldsymbol{\Psi}\right] &= \phi\left(\frac{d+\nu}{2}\right) - \ln\left(\frac{d+\nu}{2}\right)\\ &+ \sum_{i=1}^{N}\Big(\ln\big(\mathrm{E}\big[w_i\big|\mathbf{Y},\nu,\boldsymbol{\beta},\boldsymbol{\Psi}\big]\big) - \mathrm{E}\big[w_i\big|\mathbf{Y},\nu,\boldsymbol{\beta},\boldsymbol{\Psi}\big]\Big),\end{aligned}$$

where $\phi(\cdot)$ is the *digamma function*[9] and $\mathrm{E}\big[w_i\big|\mathbf{Y},\nu,\boldsymbol{\beta},\boldsymbol{\Psi}\big]$ is given by Equation (4.8).

[9] The digamma function is defined as $\phi(x) = \frac{\mathrm{d}\ln\big(\Gamma(x)\big)}{\mathrm{d}x}$.

The degrees of freedom parameter is then attained by maximizing Equation (4.11) with respect to ν. In other words, at iteration $(m+1)$ of the ECM algorithm the value of ν is obtained by finding the solution to the following equation

$$0 = -\phi\left(\frac{\nu}{2}\right) + \ln\left(\frac{\nu}{2}\right) + 1 + \phi\left(\frac{d+\nu^{(m)}}{2}\right) - \ln\left(\frac{d+\nu^{(m)}}{2}\right) + \frac{1}{N}\sum_{i=1}^{N}\left(\ln\left(w_i^{(m)}\right) - w_i^{(m)}\right). \tag{4.12}$$

The solution to Equation (4.12) can be found by a one-dimensional search using a half interval method. In this thesis Equation (4.12) is solved using the MATLAB function *Bisection* [3], see also [1] for more information. Additionally, for a more thorough derivation of Equation (4.12) see [23].

Treating Missing Values

Until now, the missing data $\mathbf{Y}_{miss}$ in $\mathbf{Y} = \left(\mathbf{Y}_{obs}, \mathbf{Y}_{miss}\right)$ has not been covered in the estimation process. In the parameter estimation process the missing data is temporary filled in using the conditional expectation

$$\mathrm{E}[\mathbf{Y}_{miss}|\mathbf{Y}_{obs} = \mathbf{y}_{obs}] = \mathbf{X}\boldsymbol{\beta} - \boldsymbol{\Sigma}_{\mathbf{Y}_{obs},\mathbf{Y}_{miss}}\boldsymbol{\Sigma}^{-1}_{\mathbf{Y}_{obs},\mathbf{Y}_{miss}}(\mathbf{y}_{obs} - \mathbf{X}\boldsymbol{\beta}),$$

and then the conditional covariance of $\mathbf{Y}_{miss}|\mathbf{Y}_{obs} = \mathbf{y}_{obs}$ is updated via

$$\boldsymbol{\Sigma}_{\mathbf{Y}_{miss}|\mathbf{Y}_{obs}=\mathbf{y}_{obs}} = \boldsymbol{\Sigma}_{\mathbf{Y}_{miss}} - \boldsymbol{\Sigma}_{\mathbf{Y}_{obs},\mathbf{Y}_{miss}}\boldsymbol{\Sigma}^{-1}_{\mathbf{Y}_{obs},\mathbf{Y}_{obs}}\boldsymbol{\Sigma}_{\mathbf{Y}_{miss},\mathbf{Y}_{obs}}.$$

Implementing the ECM Algorithm

Consider a $N \times d$ vector $\mathbf{y} = (\mathbf{y}_{obs}, \mathbf{y}_{miss})$, where $N = (N_{obs}, N_{miss})$ is the number of observation and missing values in $\mathbf{y}$. The ECM algorithm is then used to obtain the MAP estimates of each parameter, see Algorithm (1). The E-step corresponds to Equation (4.8) and the CM-steps to Equations (4.9), (4.10) and (4.12).

Algorithm 1: The ECM Algorithm

1: Initialize $\mathbf{y}_{miss} \leftarrow \mathbf{0}$, $\boldsymbol{\Psi}^{(0)}_{obs} \leftarrow \mathbf{I}$, $\boldsymbol{\Psi}^{(0)} = \left(\boldsymbol{\Psi}^{(0)}_{obs}, \boldsymbol{\Psi}^{(0)}_{miss}\right)$, $\boldsymbol{\omega}^{(0)} \leftarrow \mathbf{1}$ and $\nu^{(0)} \leftarrow 4$.

2: **while** $\left|\boldsymbol{\beta}^{(m+1)} - \boldsymbol{\beta}^{(m)}\right| \geq$ tolerance **do**

3: $\quad \boldsymbol{\beta}^{(m+1)} \leftarrow \left(\sum_{i=1}^{N} w_i^{(m)} \mathbf{y}_i (\mathbf{X}_{i,0:k})' \boldsymbol{\Psi}^{(m)} \mathbf{X}_{i,0:k}\right)^{-1} \sum_{i=1}^{N} w_i^{(m)} (\mathbf{X}_{i,0:k}) \boldsymbol{\Psi}^{(m)} \mathbf{y}_i$

4: $\quad \boldsymbol{\Lambda}^{(m+1)} \leftarrow \sum_{i=1}^{N} w_i^{(m)} \left(\mathbf{y}_i - \mathbf{X}_{i,0:k}\boldsymbol{\beta}^{(m+1)}\right)\left(\mathbf{y}_i - \mathbf{X}_{i,0:k}\boldsymbol{\beta}^{(m+1)}\right)'$

5: $\quad \boldsymbol{\Psi}^{(m+1)} \leftarrow (N - d - 1)\left(\boldsymbol{\Lambda}^{(m+1)}\right)^{-1}$

6: $\quad$ Solve for ν:

$$0 = -\,\phi\left(\frac{\nu}{2}\right) + \ln\left(\frac{\nu}{2}\right) + 1 + \phi\left(\frac{d + \nu^{(m)}}{2}\right) - \ln\left(\frac{d + \nu^{(m)}}{2}\right) + \frac{1}{N}\sum_{i=1}^{N}\left[\ln\left(w_i^{(m)}\right) - w_i^{(m)}\right]$$

7: $\quad$ Set $\nu^{(m+1)} = \nu$

8: $\quad$ **for** $i = 1$ to N **do**

9: $\qquad w_i^{(m+1)} \leftarrow \frac{d+\nu^{(m+1)}}{\nu^{(m+1)} + (\mathbf{y}_i - \mathbf{X}_{i,0:k}\boldsymbol{\beta}^{(m+1)})'\boldsymbol{\Psi}^{(m+1)}(\mathbf{y}_i - \mathbf{X}_{i,0:k}\boldsymbol{\beta}^{(m+1)})}$

10: $\quad$ **end for**

11: $\quad$ Treating missing values

12: $\quad$ **for** $i = 1$ to N_{miss} **do**

13: $\qquad \mathbf{y}^{(m+1)}_{miss,i} = \mathbf{X}_{i,0:k}\boldsymbol{\beta}^{(m+1)} - \left(\boldsymbol{\Psi}^{(m+1)}_{obs,miss}\right)^{-1} \boldsymbol{\Psi}^{(m+1)}_{obs,obs}\left(\mathbf{y}_{obs,i} - \mathbf{X}_{i,0:k}\boldsymbol{\beta}^{(m+1)}\right)$

14: $\quad$ **end for**

15: $\quad \left(\boldsymbol{\Psi}^{(m+1)}_{miss|obs}\right)^{-1} \leftarrow \left(\boldsymbol{\Psi}^{(m+1)}_{miss}\right)^{-1} - \left(\boldsymbol{\Psi}^{(m+1)}_{obs,miss}\right)^{-1} \boldsymbol{\Psi}^{(m+1)}_{obs,obs}\left(\boldsymbol{\Psi}^{(m+1)}_{miss,obs}\right)^{-1}$

16: $\quad$ Set $\boldsymbol{\Psi}^{(m+1)}_{miss} \leftarrow \boldsymbol{\Psi}^{(m+1)}_{miss|obs}$

17: $\quad$ Set $\boldsymbol{\Psi}^{(m+1)} \leftarrow \left(\boldsymbol{\Psi}^{(m+1)}_{obs}, \boldsymbol{\Psi}^{(m+1)}_{miss}\right)$

18: **end while**

The MAP estimate for the normal regression, i.e letting $\nu \to \infty$, is obtained by neglecting the weight update in lines 8-10. The ECM routine described above is implemented through an updated version of the MATLAB function, *mvsregress* [30]. In the original version of *mvsregress* the degrees of freedom, ν, is assumed to be known, and the function hence exclude the estimation in line 6 in Algorithm 1.

The estimated coefficients in Equation (4.1) and the degrees of freedom parameter, with different amount of missing data (NaN), can be found in Table 4.1.

Normal distribution					Student's t-distribution				
NaN	β_0	β_1	β_2	ν	NaN	β_0	β_1	β_2	ν
0	0.0205	0.5744	0.5394	∞	0	0.0153	0.5949	0.5058	2.8947
3	0.0217	0.5741	0.5401	∞	3	0.0164	0.5944	0.5068	2.9135
10	0.0197	0.5734	0.5412	∞	10	0.0140	0.5933	0.5088	2.9328
25	0.0213	0.5723	0.5401	∞	25	0.0135	0.5917	0.5074	2.9705
50	0.0337	0.5563	0.5501	∞	50	0.0162	0.5813	0.5173	3.3667
100	0.0088	0.5426	0.5598	∞	100	0.0106	0.5724	0.5253	5.0469
150	0.0171	0.5370	0.5579	∞	150	0.0127	0.5585	0.5394	6.2348
250	0.0193	0.5342	0.5851	∞	250	0.0191	0.5494	0.5636	9.3356

Table 4.1: Estimated parameter values for the normal distribution assumption and the Student's t-distribution with different amount of missing data (NaN).

The predicted values are now computed using Equation (4.1) together with the parameter estimates from Table 4.1. The random noise variable z_i is then added to every output to restore variability to the imputed data. In the first model the noise term is assumed to have a normal distribution with zero mean and a corresponding residual variance. In the model where the data is assumed to be heavy tailed the noise is generated by the t-distribution with zero mean and corresponding residual variance and degree of freedom, i.e

$$
\begin{aligned}
z_i &\sim \mathcal{N}\big(0, \mathbf{\Psi}^{-1}\big) \\
z_i &\sim t\left(\nu, 0, \frac{\nu}{\nu - 2}\mathbf{\Psi}^{-1}\right).
\end{aligned}
$$

4.2 Multiple Imputation (MI)

Multiple imputation (MI) is an alternative to maximum likelihood estimation (MLE) and is the other method currently regarded as a state-of-the-art missing data technique [10]. The MI method tends to produce similar analysis results as the MLE approach if the imputation model contains the same variables as the subsequent analysis model. However, in the case when additional variables are used that are not included in the subsequent analysis the two approaches may provide different estimates, standard errors, or both. However, the main difference between the methods is that the analysis from the MI approach does not depend solely on one imputed set of missing values but of several augmented chains. For more information regarding the advantages and disadvantages of the two methods please see [10]. The imputation approach outlined in this section functions under the same assumption as in Section 4.1, i.e. that the data is missing at random (MAR). In line with Section 4.1 both a normal distribution and a t-distribution is assigned to the data. To attain the corresponding t-distribution, the approach used in Section 4.1, is also applied here.

MI is actually a broad term and include a variety of underlying imputation techniques, each superior depending on the current missing data pattern and mechanism assumption. A multiple imputation analysis consists of three distinct steps common to all multiple imputation procedures:

- The imputation phase
- The analysis phase
- The pooling phase

In the imputation phase the underlying algorithm is used to estimate a multiple set of predicted values for each missing data point. In other words, a multiple set of parallel series are generated for each set of missing data. The underlying MI procedure used in this thesis is an iterative version of the stochastic regression imputation technique from Section 4.1. However, the mathematical principle behind the update of the parameter estimates rely heavily on Bayesian methodology. The goal of the analysis phase is, as its name implies, to individually analyze the filled-in data sets and estimate the corresponding parameters. With exception of the initial step, this amounts to sequentially applying the same statistical procedures as if the data had been complete. Additionally, the analysis phase should hence produce a parameter estimate for each imputed parallel set of values from the imputation phase. The purpose of the final step, the pooling phase, is to estimate a single set of results by combining all the estimated parameters.

4.2.1 The Imputation Phase

The imputation phase is a data augmentation algorithm consisting of a two-step procedure, the actual imputation step (I-step) and the posterior step (P-step). The I-step uses the stochastic regression imputation technique from Section 4.1 to fill-in the set of missing values. More specific, the I-step uses an estimate of the mean vector and covariance matrix to construct a set of regression equations to predict the missing values using the observed covariates.

The objective of the imputation phase is to generate a large number κ of complete data sets of length $N = N_{obs} + N_{miss}$, each of which contains augmented chains of length N_{miss} of unique estimates of the missing values, see Table 4.2. To be able to attain these unique imputations, new estimates of the regression coefficients are required for each I-step. The purpose of the P-step is therefore to draw new candidates of the mean vector and covariance matrix from the respective posterior distribution. The imputation step is hence an iterative process which alternates between the I-step and the P-step.

$Y^*_{\kappa,1}$	$\dots$	$Y^*_{\kappa,N_{obs}}$	$Y^*_{\kappa,N_{obs}+1}$	$\dots$	$Y^*_{\kappa,N_{obs}+N_{miss}}$
$Y^*_{\kappa-1,1}$	$\dots$	$Y^*_{\kappa-1,N_{obs}}$	$Y^*_{\kappa-1,N_{obs}+1}$	$\dots$	$Y^*_{\kappa-1,N_{obs}+N_{miss}}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$Y^*_{2,1}$	$\dots$	$Y^*_{2,N_{obs}}$	$Y^*_{2,N_{obs}+1}$	$\dots$	$Y^*_{2,N_{obs}+N_{miss}}$
$Y^*_{1,1}$	$\dots$	$Y^*_{1,N_{obs}}$	$Y^*_{1,N_{obs}+1}$	$\dots$	$Y^*_{1,N_{obs}+N_{miss}}$

Table 4.2: An overview of the κ number of complete data sets of length $N = N_{obs} + N_{miss}$, each of which contains augmented chains of length N_{miss} of unique estimates of the missing values.

The I-Step

As described above, the purpose of the I-step is to construct a set of regression equations from an estimate of the mean vector and the covariance matrix. Since the procedure is an iterative process alternating between the I-step and the P-step, the first iteration requires an initial estimate of the mean vector and the covariance matrix. Typically, a good starting point is to use the estimated parameters from the ECM algorithm used in Section 4.1. From these initial parameters estimates a first set of missing values are predicted from the observed covariates by using stochastic regression imputation. In other words, the initial phase of the I-step produces a first set of complete data, which the rest of the analysis is based upon.

Since the data is assigned a mixture of normal distributions together with an auxiliary weight variable, w_i, the corresponding unbiased[10] Monte Carlo integration estimate of the expected value and covariance of the complete data is given by

$$\hat{\mu}_{\mathbf{Y}} = \frac{1}{N}\sum_{i=1}^{N} w_i y_i \tag{4.13}$$

and

$$\hat{Cov}(\mathbf{Y}, \mathbf{X}_j) = \frac{1}{N-1}\sum_{i=1}^{N} w_i (y_i - \hat{\mu}_{\mathbf{Y}})(x_{i,j} - \hat{\mu}_{\mathbf{X}_j}). \tag{4.14}$$

The value of the auxiliary weight parameter w_i is attained together with the initial parameter estimates using the ECM algorithm. Note that under the normal assumption the weight $w_i = 1$, which is consistent with Section 4.1.

Given the estimated regression equation

$$\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{z},$$

where

$$\hat{\mathbf{Y}} = \begin{bmatrix} \hat{y}_1 \\ \vdots \\ \hat{y}_N \end{bmatrix}, \; \mathbf{X} = \begin{bmatrix} 1 & x_{1,1} & \dots & x_{1,j} & \dots & x_{1,k} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 1 & x_{N,1} & \dots & x_{N,j} & \dots & x_{N,k} \end{bmatrix} = \begin{bmatrix} \mathbf{1} & \mathbf{X}_1 & \dots & \mathbf{X}_k & \dots & \mathbf{X}_k \end{bmatrix},$$

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} \hat{\beta}_0 \\ \vdots \\ \hat{\beta}_k \end{bmatrix}, \text{ and } \mathbf{z} = \begin{bmatrix} z_1 \\ \vdots \\ z_N \end{bmatrix},$$

[10] Unbiased in the sense that $\hat{\mu}_{(\cdot)} = \frac{1}{N}\sum_{i=1}^{N}(\cdot) \to \mathrm{E}[\cdot]$ as $N \to \infty$.

then the corresponding mean vector and covariance matrix is

$$\hat{\boldsymbol{\mu}} = \begin{bmatrix} \hat{\mu}_{\mathbf{Y}} \\ \hat{\mu}_{\mathbf{X}_1} \\ \vdots \\ \hat{\mu}_{\mathbf{X}_k} \end{bmatrix}, \ \hat{\boldsymbol{\Sigma}} = \begin{bmatrix} \hat{Cov}(\mathbf{Y},\mathbf{Y}) & \hat{Cov}(\mathbf{Y},\mathbf{X}_1) & \dots & \hat{Cov}(\mathbf{Y},\mathbf{X}_k) \\ \hat{Cov}(\mathbf{X}_1,\mathbf{Y}) & \hat{Cov}(\mathbf{X}_1,\mathbf{X}_1) & \dots & \hat{Cov}(\mathbf{X}_1,\mathbf{X}_k) \\ \vdots & \vdots & \ddots & \vdots \\ \hat{Cov}(\mathbf{X}_k,\mathbf{Y}) & \hat{Cov}(\mathbf{X}_k,\mathbf{X}_1) & \dots & \hat{Cov}(\mathbf{X}_k,\mathbf{X}_k) \end{bmatrix}.$$

By using Equation (4.13) and (4.14) the regression coefficients are calculated by solving the following set of equations

$$\hat{Cov}(\mathbf{Y},\mathbf{X}_i) = \sum_{j=1}^{k} \hat{\beta}_j \cdot \hat{Cov}(\mathbf{X}_j,\mathbf{X}_i).$$

and

$$\hat{\beta}_0 = \hat{\mu}_{\mathbf{Y}} - \hat{\boldsymbol{\mu}}_{\mathbf{X}_{1:k}} \hat{\boldsymbol{\beta}}_{1:k},$$

or by using the subvector and submatrix of the mean vector and covariance matrix

$$\hat{\boldsymbol{\beta}}_{1:k} = \begin{bmatrix} \hat{Cov}(\mathbf{X}_1,\mathbf{X}_1) & \dots & \hat{Cov}(\mathbf{X}_1,\mathbf{X}_k) \\ \vdots & \ddots & \vdots \\ \hat{Cov}(\mathbf{X}_k,\mathbf{X}_1) & \dots & \hat{Cov}(\mathbf{X}_k,\mathbf{X}_k) \end{bmatrix}^{-1} \begin{bmatrix} \hat{Cov}(\mathbf{Y},\mathbf{X}_1) \\ \vdots \\ \hat{Cov}(\mathbf{Y},\mathbf{X}_k) \end{bmatrix}$$

and

$$\hat{\beta}_0 = \hat{\mu}_{\mathbf{Y}} - \hat{\boldsymbol{\mu}}_{\mathbf{X}_{1:k}} \hat{\boldsymbol{\beta}}_{1:k}.$$

The P-Step

In this step a new set of mean values and covariances are generated from the respective posterior distribution. From Section 4.1 it is known that the posterior distribution of the shape matrix $\boldsymbol{\Psi}$ is

$$\boldsymbol{\Psi} \sim Wishart(\boldsymbol{\Lambda}^{-1}, N),$$

where $\boldsymbol{\Lambda} = (N-d-1)\boldsymbol{\Sigma}$. If a random matrix has a Wishart distribution with parameters $\boldsymbol{\Lambda}^{-1}$ and N, then the inverse of that random matrix has an inverse Wishart distribution with parameters $\boldsymbol{\Lambda}$ and N. Hence, the posterior distribution of the covariance matrix $\boldsymbol{\Sigma}$ is given by

$$\boldsymbol{\Sigma}^* \sim Wishart^{-1}(\hat{\boldsymbol{\Lambda}}, N),$$

where the * denotes that the parameter is generated through simulation from the posterior distribution. From Equation (4.6) it can be seen that the posterior distribution of the expected value is multivariate normal, i.e

$$\boldsymbol{\mu}^* \sim \mathcal{MN}\big(\hat{\boldsymbol{\mu}}, (N-d-1)^{-1}\boldsymbol{\Sigma}^*\big).$$

The alternating steps of the method is presented in a more concise and perspicuous manner in Algorithm 2 below.

Algorithm 2: The Data Augmentation Algorithm of Multiple Imputation

Initial I-Step:

Initialize $\hat{\boldsymbol{\beta}}^{(0)}$, $\mathbf{w}$ and ν via Algorithm 1.

Estimate $\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}}^{(0)}$

$\boldsymbol{\Psi}^{(t)} \leftarrow (N-d-1)\left(\sum_{i=1}^{N} w_i \left(\mathbf{y}_i - \mathbf{X}_{i,0:k}\boldsymbol{\beta}^{(t)}\right)\left(\mathbf{y}_i - \mathbf{X}_{i,0:k}\boldsymbol{\beta}^{(t)}\right)'\right)^{-1}$

Draw $\mathbf{z}^* \sim \mathcal{N}\left(0, \left(\hat{\boldsymbol{\Psi}}^{(0)}\right)^{-1}\right)$ or $\mathbf{z}^* \sim t\left(\nu, 0, \frac{\nu}{\nu-2}\left(\hat{\boldsymbol{\Psi}}^{(0)}\right)^{-1}\right)$

Set $\hat{\mathbf{Y}}^{(0)} = \mathbf{X}\hat{\boldsymbol{\beta}}^{(0)} + \mathbf{z}^*$

Estimate $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$

for $t = 1$ to T **do**

P-Step:

Set $\hat{\boldsymbol{\Lambda}} = (N-d-1)\hat{\boldsymbol{\Sigma}}$

Draw $\boldsymbol{\Sigma}^{*(t)} \sim Wishart^{-1}(\hat{\boldsymbol{\Lambda}}, N)$

Draw $\boldsymbol{\mu}^{*(t)} \sim \mathcal{MN}\left(\hat{\boldsymbol{\mu}}, (N-d-1)^{-1}\boldsymbol{\Sigma}^{*(t)}\right)$

I-Step:

Compute $\hat{\boldsymbol{\beta}}_{1:k}^{(t)}$ using $\boldsymbol{\Sigma}^{*(t)}$

Compute $\hat{\beta}_0^{(t)}$ using $\boldsymbol{\mu}^{*(t)}$

Estimate $\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}}^{(t)}$

$\boldsymbol{\Psi}^{(t)} \leftarrow (N-d-1)\left(\sum_{i=1}^{N} w_i \left(\mathbf{y}_i - \mathbf{X}_{i,0:k}\boldsymbol{\beta}^{(t)}\right)\left(\mathbf{y}_i - \mathbf{X}_{i,0:k}\boldsymbol{\beta}^{(t)}\right)'\right)^{-1}$

Draw $\mathbf{z}^* \sim \mathcal{N}\left(0, \left(\hat{\boldsymbol{\Psi}}^{(t)}\right)^{-1}\right)$ or $\mathbf{z}^* \sim t\left(\nu, 0, \frac{\nu}{\nu-2}\left(\hat{\boldsymbol{\Psi}}^{(t)}\right)^{-1}\right)$

Set $\hat{\mathbf{Y}}^{(t)} = \mathbf{X}\hat{\boldsymbol{\beta}}^{(t)} + \mathbf{z}^*$

Estimate $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$

end for

Convergence

The data augmentation algorithm used in the imputation phase belongs to a family of Markov Chain Monte Carlo (MCMC) procedures. The objective of a MCMC algorithm is to generate random draws from a distribution. By iteratively alternating between the I-step and the P-step, and sequentially generating new samples, a so-called data augmentation chain is created:

$$\mathbf{Y}_1^*, \boldsymbol{\theta}_1^*, \mathbf{Y}_2^*, \boldsymbol{\theta}_2^*, \mathbf{Y}_3^*, \boldsymbol{\theta}_3^*, \ldots, \mathbf{Y}_t^*, \boldsymbol{\theta}_t^*, \ldots, \mathbf{Y}_T^*, \boldsymbol{\theta}_T^*,$$

where $\mathbf{Y}_t^*$ represents the imputed values at I-step t, and $\boldsymbol{\theta}_t^*$ contains the generated parameter values at P-step t. Note that the length of the total number of generated complete sets, T, is much greater than the final number of unique independent complete sets, κ.

In contrast to maximum likelihood based algorithms which reach convergence when the parameter estimates do not change over consecutive iterations, data augmentation converges when the distributions become stationary[11]. The complexity of this definition is due to the fact that each step in the data augmentation chain is dependent on the former step. Hence, even if the data augmentation chain appears to be sequentially increased by random samples, the dependence between each I-step and P-step induces a correlation between the simulated parameters from consecutive P-steps. Additionally, analyzing data sets from consecutive I-steps inherit the same problem and is therefore also inappropriate. Consequently, to be able to generate a multiple of independent copies of the missing data it is essential to ensure when the data augmentation

[11] Stationary in the sense that the expected value and variance do not change over time, i.e there is no visible trend present.

chain has reached stationarity.

Monitoring the evolution of the chain over a large number of cycles is one way to accomplish this. In Figure 4.1-4.4 the first 200 iterations of the data augmentation chain of the expected values and the elements of the covariance matrix can be seen for the case with the greatest amount of missing values (in this case NaN=250). The horizontal line represents the corresponding mean of the total chain of $T = 10,000$ iterations.

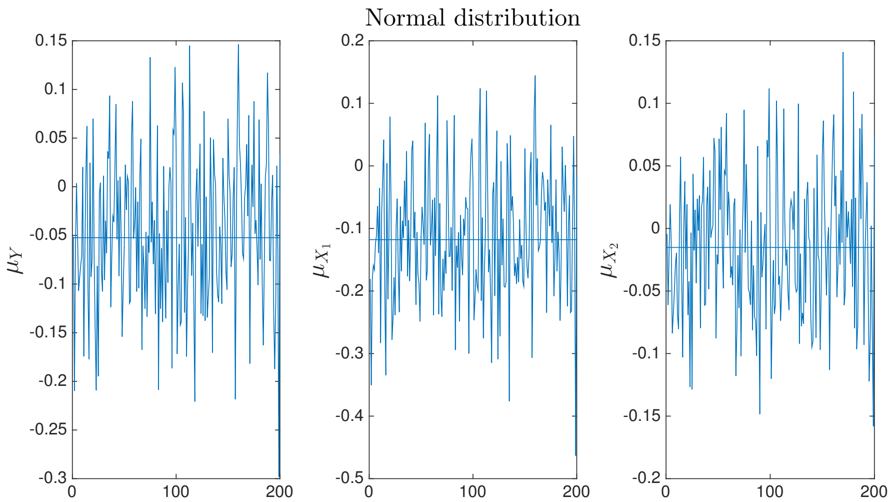

Figure 4.1: The evolution of the data augmentation chain for the first 200 sampled expected values under the normal assumption.

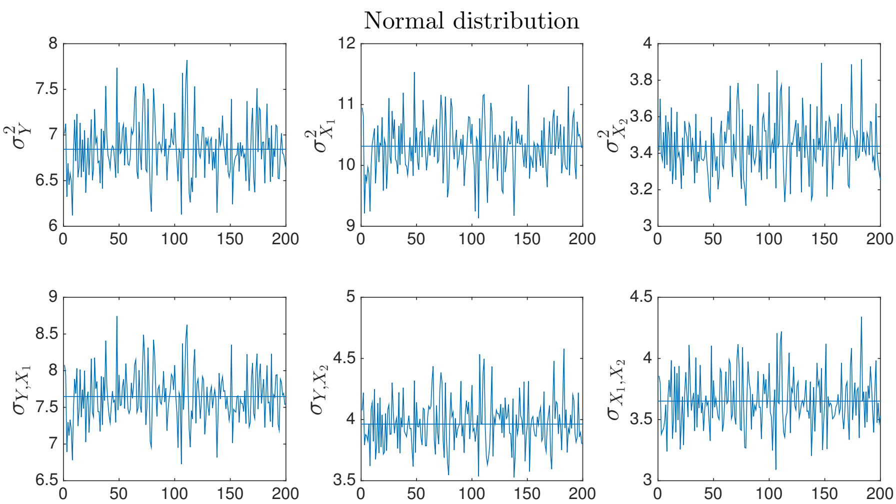

Figure 4.2: The evolution of the data augmentation chain for the first 200 sampled elements of the covariance matrix under the normal assumption.

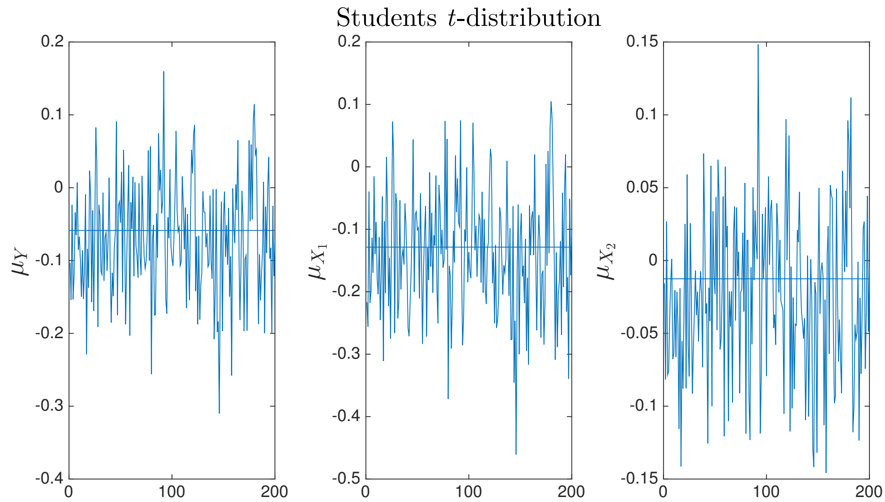

Figure 4.3: The evolution of the data augmentation chain for the first 200 sampled expected values under the Student's t model.

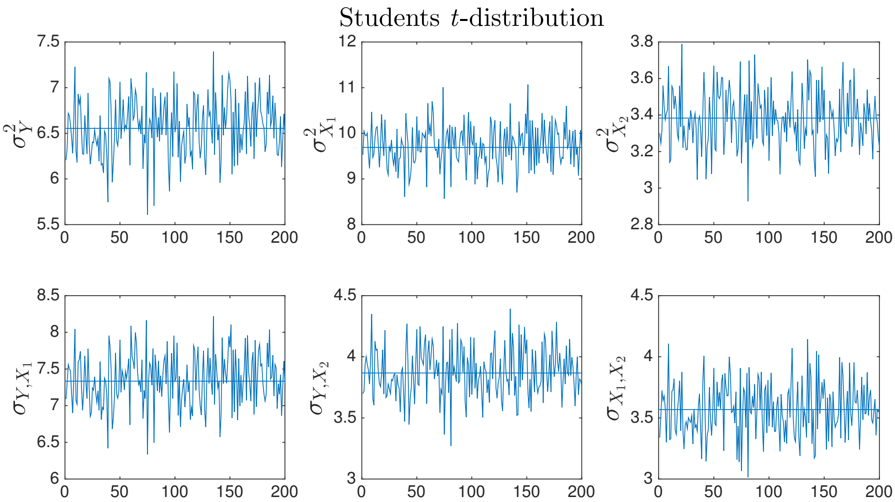

Figure 4.4: The evolution of the data augmentation chain for the first 200 sampled elements of the covariance matrix under the Student's t model.

In Figure 4.1-4.4 there is no visible long-term trends, and the parameters evolve in a seemingly random fashion. The absence of trend is an ideal situation and suggests that the parameters almost instantly converges to a stable distribution.

Another important diagnostic tool for assessing convergence is the auto-correlation function which quantifies the magnitude and duration of the induced dependency. The Pearson correlation is used to measure the k-lag auto-correlation between sets of parameter values separated by k iterations in the data augmentation chain. In Figure 4.5-4.8 the sample auto-correlation of the complete data augmentation chain of the expected values and elements of the covariance matrix can be seen.

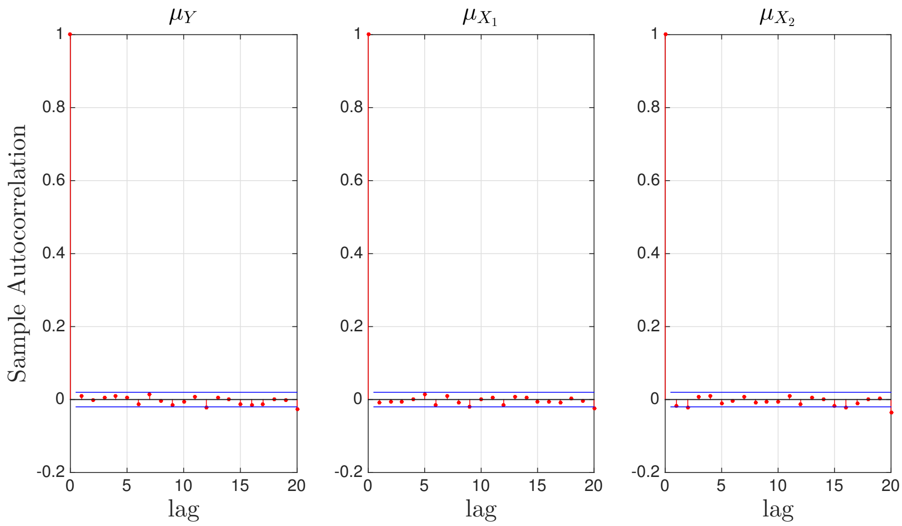

Figure 4.5: The auto-correlation function (acf) of the complete data augmentation chain for the sampled expected values under the normal distribution.

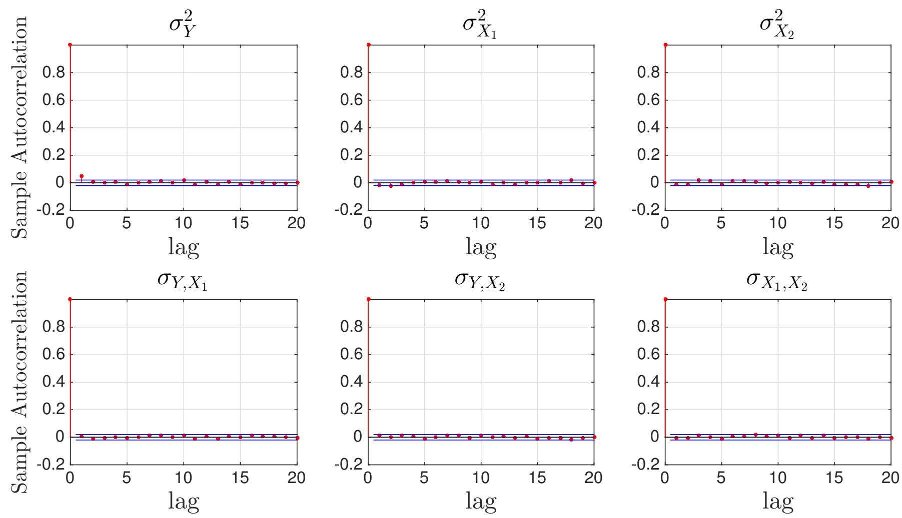

Figure 4.6: The auto-correlation function (acf) of the complete data augmentation chain for the sampled elements of the covariance matrix under the normal assumption.

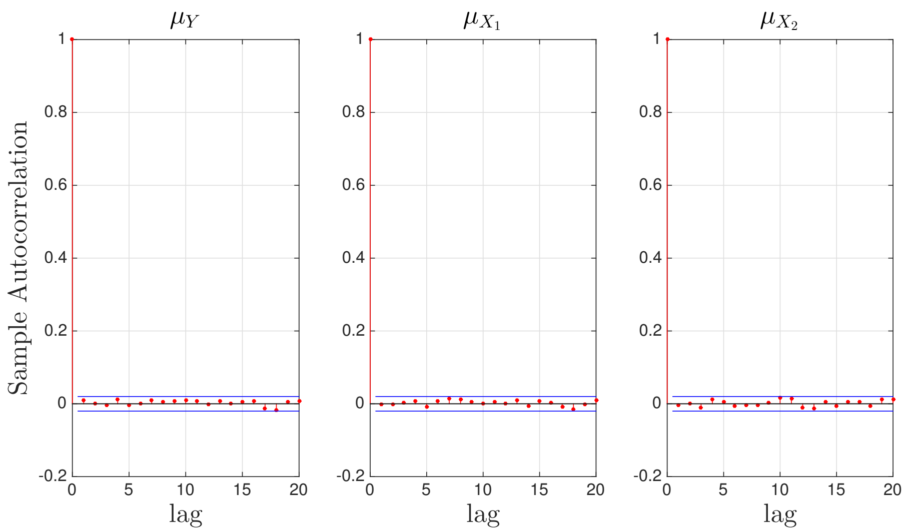

Figure 4.7: The auto-correlation function (acf) of the complete data augmentation chain for the sampled expected values under the Student's t model.

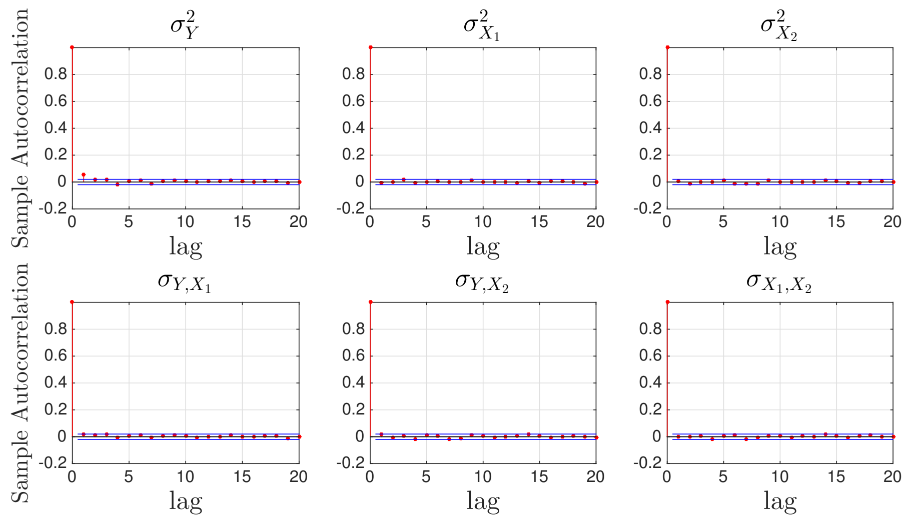

Figure 4.8: The auto-correlation function (acf) of the complete data augmentation chain for the sampled elements of the covariance matrix under the Student's t model.

In Figure 4.5-4.8 the auto-correlation plots of the different parameters confirm the behavior in Figure 4.1-4.4. The auto-correlation drops to within sampling error of zero almost immediately. This suggests that the distribution of the parameters becomes stable after a very small number of iterations.

Final Set of Imputations

As stated earlier, the main goal of the imputation phase is to generate a large number κ of complete data sets, each of which contains unique estimates of the missing values. After analyzing the complete data augmentation chains there are two approaches for attaining the final set of independent imputations, κ, namely by sequential and parallel sampling. Sequential data augmentation chains refers to saving and analysing the data sets at regular intervals, e.g. every 20^{th} I-step. Parallel data augmentation chains are generated by sampling several chains and save and analyse the imputed data at the final I-step. In this thesis the final set of imputations are generated by using sequential sampling. To be certain that the data augmentation chain has reached stationarity the generated samples are saved and analyzed every 100^{th} I-step. Hence, a total of $T = 10,000$ iterations are required to attain $\kappa = 100$ imputed samples for each missing value.

4.2.2 The Analysis Phase

In the analysis phase the κ data sets are analyzed using the same procedure as if the data had been complete. This phase is in a way integrated in Algorithm 2, and the mean vector and covariance matrix of the corresponding κ final complete data sets are saved and analyzed. The regression coefficients are then computed through these κ sets of independently sampled parameters and are combined to a single estimate in the subsequent pooling phase.

4.2.3 The Pooling Phase

In the final step of the MI procedure the κ estimates of each parameter from the analysis phase are pooled into a single point estimate. In contrast to the ECM algorithm in Section 4.1 the MI approach relies on the results from multiple estimates rather than a single data set. In 1987 Rubin [35] defined the MI point estimate as the arithmetic mean value of the κ independent estimates

$$\bar{\theta} = \frac{1}{\kappa}\sum_{i=1}^{\kappa}\hat{\theta}_i,$$

where $\hat{\theta}_i$ is the parameter estimate from data set i and $\bar{\theta}$ is the pooled estimate [10]. In Figure 4.9-4.10, 10,000 samples generated from the corresponding posterior distribution of each regression coefficient are plotted in a histogram with a fitted normal distribution and Student's t-distribution. Additionally, in Table 4.3 the pooled estimates of the regression coefficients are listed for different amount of missing data (NaN).

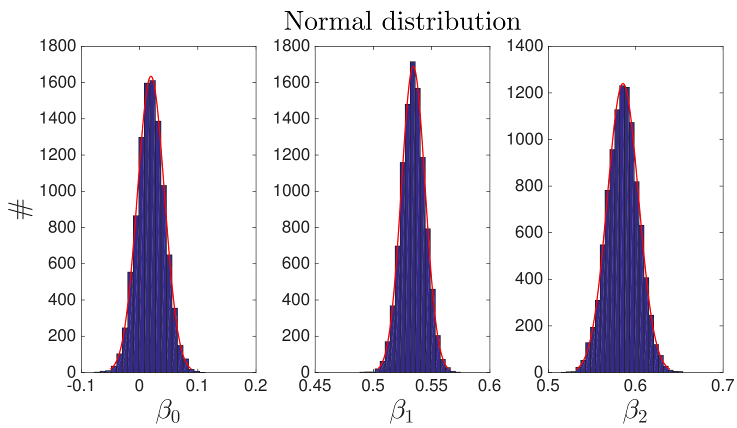

Figure 4.9: A histogram of 10,000 sampled regression coefficients under the normal assumption with a fitted normal distribution.

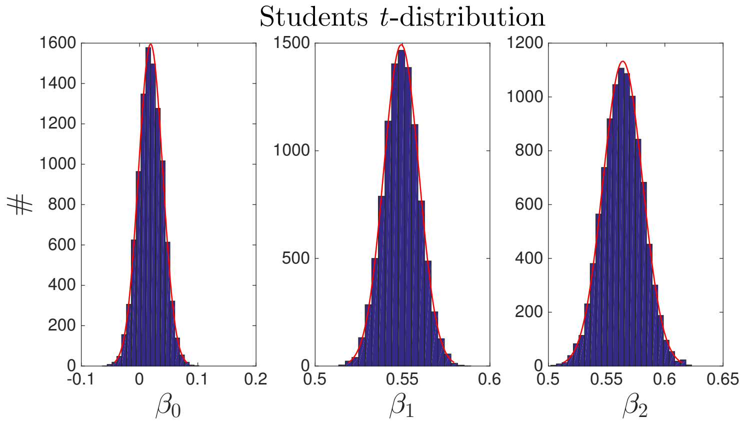

Figure 4.10: A histogram of 10,000 sampled regression coefficients under the Student's t model with a fitted Student's t-distribution.

Normal distribution				
NaN	β_0	β_1	β_2	ν
3	0.0216	0.5719	0.5425	∞
10	0.0205	0.5746	0.5399	∞
25	0.0200	0.5720	0.5402	∞
50	0.0376	0.5552	0.5520	∞
100	0.0077	0.5438	0.5584	∞
150	0.0177	0.5369	0.5580	∞
250	0.0184	0.5349	0.5836	∞

Student's t-distribution				
NaN	β_0	β_1	β_2	ν
3	0.0173	0.5954	0.5055	2.9135
10	0.0130	0.5929	0.5091	2.9328
25	0.0097	0.5929	0.5063	2.9705
50	0.0186	0.5807	0.5169	3.3667
100	0.0097	0.5726	0.5242	5.0469
150	0.0123	0.5579	0.5387	6.2348
250	0.0164	0.5479	0.5661	9.3356

Table 4.3: The pooled parameter values for the normal distribution assumption and the Student's t-distribution with different amount of missing data (NaN).

The predicted values are now computed using Equation (4.1) together with the parameter estimates from Table 4.1. The random noise variable z_i is then added to every output to restore variability to the imputed data. In the first model the noise term is assumed to have a normal distribution with zero mean and a corresponding residual variance. In the model where the data is assumed to be heavy tailed the noise is generated by the t-distribution with zero mean and corresponding residual variance and degree of freedom, i.e

$$\begin{aligned} z_i &\sim \mathcal{N}\big(0, \boldsymbol{\Psi}^{-1}\big) \\ z_i &\sim t\left(\nu, 0, \frac{\nu}{\nu - 2}\boldsymbol{\Psi}^{-1}\right). \end{aligned}$$

4.3 A Copula Approach

The use of copulas has become very popular over the years since they offer a dynamic way to model the dependence structure between random variables. It is an appealing approach since the marginal behavior of each random variable and the joint dependent structure can be modelled separately. This gives the opportunity to a custom made modeling approach for each problem. Although the use of copulas to model the dependence structure between random variables has increased (especially in finance), the literature regarding copulas in the context of missing data is very limited. The earliest articles found concerning this topic is the imputation of missing data by using the Gaussian copula (g-copula) by Käärik (2005) [19] and Käärik (2006) [20]. This work is developed further by Friedman et al. (2009) [13] and Friedman et al. (2012) [14], where the Student's t-copula (t-copula), among other models (including the Gaussian copula), are applied to heavy tailed data.

In consistency with Section 4.1-4.2, both a normal assumption and heavy tailed model is applied to the data in this section. The construction of the dependence structure used in this thesis stems from the definition of a copula as a multivariate distribution function defined on the d-dimensional unit cube $[0,1]^d$, with uniformly distributed marginals [7]. In line with the assumptions regarding the data, a multivariate *normal* distribution and a multivariate t-distribution is used to generate samples from the corresponding copula. Both these distributions belong to the family of elliptical distributions, and the corresponding copulas are hence referred to as elliptical copulas, see Definition 3.5.2. Elliptical distributions are easy to simulate from and this helpful property is inherited by the elliptical copulas through *Sklar's* theorem (see, Theorem 3.5.1) [7], even if elliptical copulas in general do not have simple closed forms [24].

Copulas are usually used to model the dependence structure between marginal distributions when the corresponding joint distribution function is intractable. However, in this thesis copulas are used to model the dependency of conditional distribution functions directly, without the use of Bayes' theorem, see Section 3.3. The following feature of the (t-copula) is stated in Demarta and McNeil (2004) [5]:

> *"If a random vector* $\mathbf{Y}$ *has the t-copula* $C^t_{\nu,P}$ *and univariate t margins with the same degree of freedom parameter* ν*, then it has a multivariate t-distribution with* ν *degrees of freedom. If, however, Equation* (3.10) *is used to combine any other set of univariate distribution functions using the t-copula we obtain multivariate distribution functions* F *which have been termed meta-t-distribution functions".*

The modelling approach is based on known results regarding the *Arellano-Valle and Bolfarine's generalized t-distribution* first given by Kots and Nadarajah (2004) [21], but is here applied as in [13] and [14], see Definition 4.3.1.

Definition 4.3.1. *A d-dimensional random vector* $\mathbf{Y} = (Y_1, ..., Y_d)'$ *is said to have an Arellano-Valle and Bolfarine's generalized t-distribution with degrees of freedom* ν*, mean vector* $\mathbf{m}$*, and symmetric positive definite* $d \times d$ *covariance matrix,* $\mathbf{R}$*, and additional parameter* $\lambda > 0$ *if it can be represented via*

$$\mathbf{Y} = \mathbf{m} + V^{\frac{1}{2}}\mathbf{G}, \tag{4.15}$$

where V *has the inverse gamma distribution given by the probability density function*

$$f_V(v) = \frac{\left(\frac{\lambda}{2}\right)^{\frac{\nu}{2}}}{\Gamma\left(\frac{\nu}{2}\right)}\left(\frac{1}{v}\right)^{\frac{\nu}{2}+1}\exp\left(-\frac{\lambda}{2v}\right), \ v > 0$$

and $\mathbf{G} \sim N(\mathbf{0}, \mathbf{R})$ *is distributed independently of* V*. A common notation for the distribution of* $\mathbf{Y}$ *is* $\mathbf{Y} \sim t(\mathbf{m}, \mathbf{R}, \lambda, \nu)$*, and the probability density function of* $\mathbf{Y}$ *is denoted by* $t(\mathbf{y}; \mathbf{m}, \mathbf{R}, \lambda, \nu)$*.*

Note that the matrix, $\mathbf{R}$, is the covariance of the normally distributed variable $\mathbf{G}$, and not the t-distributed random variable $\mathbf{Y}$.

Proposition 4.3.1. *If* $\mathbf{Y} \sim t(\mathbf{y}; \mathbf{m}, \mathbf{R}, \lambda, \nu)$*, then*

$$\begin{aligned} \mu_{\mathbf{Y}} &= \mathbf{m}, \;\; \nu > 1 \\ Cov(\mathbf{Y}) = \mathbf{\Sigma} &= \frac{\lambda}{\nu - 2}\mathbf{R}, \;\; \nu > 2. \end{aligned} \tag{4.16}$$

Definition 4.3.2. *In the special case* $\lambda = \nu$*, the generalized* t*-distribution reduces to the (usual) multivariate* t*-distribution with mean* $\mathbf{m}$*, degrees of freedom* ν*, and dependence structure matrix,* $\mathbf{R}$.

In the parameter estimation of the corresponding copula, the (usual) multivariate t-distribution is used. However, as to be seen, the conditional distributions of (usual) multivariate t-distributions are, in general, not usual but generalized t-distributions [13] [14].

Consider the d-dimensional t-distributed random variable $\mathbf{Y}$ with degrees of freedom ν and let

$$\mathbf{Y} = \begin{bmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{bmatrix}, \text{ with sizes } \begin{bmatrix} d_1 \times 1 \\ (d - d_1) \times 1 \end{bmatrix},$$

$$\mathbf{m} = \begin{bmatrix} \mathbf{m}_1 \\ \mathbf{m}_2 \end{bmatrix}, \text{ with sizes } \begin{bmatrix} d_1 \times 1 \\ (d - d_1) \times 1 \end{bmatrix},$$

and

$$\mathbf{\Sigma} = \begin{bmatrix} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} \\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22} \end{bmatrix}, \text{ with sizes } \begin{bmatrix} d_1 \times d_1 & d_1 \times (d - d_1) \\ (d - d_1) \times d_1 & (d - d_1) \times (d - d_1) \end{bmatrix},$$

then Proposition 4.3.2 holds.

Proposition 4.3.2. *The marginal density function of* $\mathbf{Y}_1$ *is given by*

$$f(\mathbf{y}_1) = \frac{\Gamma\left(\frac{\nu + d_1}{2}\right)}{(\pi\lambda)^{\frac{d_1}{2}}\Gamma\left(\frac{\nu}{2}\right)|\mathbf{R}_{11}|^{\frac{1}{2}}}\left(1 + \frac{1}{\lambda}(\mathbf{y}_1 - \mathbf{m}_1)'\mathbf{R}_{11}^{-1}(\mathbf{y}_1 - \mathbf{m}_1)\right)^{-\frac{\nu + d_1}{2}},$$

i.e. $\mathbf{Y}_1 \sim t(\mathbf{m}_1, \mathbf{R}_{11}, \lambda, \nu)$.

Now, let the corresponding conditional expectation, $\mathbf{m}_{\mathbf{Y}_1|\mathbf{Y}_2=\mathbf{y}_2}$, conditional covariance, $\mathbf{\Sigma}_{\mathbf{Y}_1|\mathbf{Y}_2=\mathbf{y}_2}$, and extra degrees of freedom parameter $q_{\mathbf{Y}_1|\mathbf{Y}_2=\mathbf{y}_2}$ be defined as

$$\mathbf{m}_{\mathbf{Y}_1|\mathbf{Y}_2=\mathbf{y}_2} = \mathbf{m}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{y}_2 - \mathbf{m}_2), \tag{4.17}$$

$$\boldsymbol{\Sigma}_{\mathbf{Y}_1|\mathbf{Y}_2=\mathbf{y}_2} = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}, \text{ and} \tag{4.18}$$

$$q_{\mathbf{Y}_1|\mathbf{Y}_2=\mathbf{y}_2} = (\mathbf{y}_2 - \mathbf{m}_2)'\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{y}_2 - \mathbf{m}_2).$$

The first step in computing $\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}$ and $\boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$ in Equation (4.17) and (4.18) is to invert the matrix $\boldsymbol{\Sigma}$ via

$$\begin{aligned}\boldsymbol{\Sigma}^{-1} = \boldsymbol{\Psi} &= \begin{bmatrix}\boldsymbol{\Psi}_{11} & \boldsymbol{\Psi}_{12}\\ \boldsymbol{\Psi}_{21} & \boldsymbol{\Psi}_{22}\end{bmatrix}\\ &= \begin{bmatrix}(\boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21})^{-1} & -(\boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21})^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\\ -(\boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12})^{-1}\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1} & (\boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12})^{-1}\end{bmatrix},\end{aligned}$$

and then calculate

$$\begin{aligned}-\boldsymbol{\Psi}_{11}^{-1}\boldsymbol{\Psi}_{12} &= \underbrace{(-(\boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}))(-(\boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21})^{-1})}_{\mathbf{I}}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1} = \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\\ \boldsymbol{\Psi}_{11}^{-1} &= \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21},\end{aligned}$$

where $\mathbf{I}$ is the identity matrix [33].

Proposition 4.3.3. *The conditional density function of* $\mathbf{Y}_1$*, given* $\mathbf{Y}_2$*, is*

$$\mathbf{Y}_1|\mathbf{Y}_2 = \mathbf{y}_2 \sim t\big(\mathbf{m}_{\mathbf{Y}_1|\mathbf{Y}_2=\mathbf{y}_2}, \mathbf{R}_{\mathbf{Y}_1|\mathbf{Y}_2=\mathbf{y}_2}, \lambda + q_{\mathbf{Y}_1|\mathbf{Y}_2=\mathbf{y}_2}, \nu + (d - d_1)\big).$$

Based on the Proposition 4.3.2-4.3.3 of a generalized t-distribution, the conditional density is extracted directly, which make these distributions very convenient to work with. However, under the assumption $\lambda = \nu$ the conditional distribution can be expressed as an (usual) multivariate t-distribution by rescaling the conditional covariance as follows

$$\mathbf{Y}_1|\mathbf{Y}_2 = \mathbf{y}_2 \sim t\left(\mathbf{m}_{\mathbf{Y}_1|\mathbf{Y}_2=\mathbf{y}_2}, \frac{\nu + q_{\mathbf{Y}_1|\mathbf{Y}_2=\mathbf{y}_2}}{\nu + (d - d_1)} \cdot \mathbf{R}_{\mathbf{Y}_1|\mathbf{Y}_2=\mathbf{y}_2}, \nu + (d - d_1)\right). \tag{4.19}$$

Additionally, for a thorough derivation and presentation of the multivariate and conditional Student's t-distribution please see Roth (2013) [32].

The copula approach in this thesis uses the same estimation technique as in [13] [14], i.e. by starting with raw data, the approach is to

1. estimate marginal (univariate) probability distribution functions for each variable separately
2. transform the data to a space referred to as the *working space*
3. estimate the conditional distribution of generalized t-distributed random vectors in the working space
4. compute the conditional probability distribution on the raw variable space.

More specific, suppose that there are N observations of d random variables on the raw variable space $\mathcal{Y} = \mathbb{R}^d$. Assume, for the moment, that the degrees of freedom, ν, of a (usual) multivariate t-distribution is known. The first step in the estimation of the conditional distribution of the raw variables is to estimate the univariate cumulative distribution function (cdf), F_j, for each variable in the raw space. Additionally, as in [13] and [14] the family of *Johnson SU distributions* is used to model the marginal characteristics of the raw data. This is a four-parameter family of distributions suited to fit normally distributed data as well as heavy tailed data. In contrast to [13] and [14] the marginal distributions of the data are fitted using the *Johnson Curve Toolbox* [18]. The univariate distributions are then used to map each observation to a point in $\mathcal{U} = [0, 1]^d$, by using the probability transformation $u_{i,j} = F_j(y_{i,j})$, see Proposition 3.5.1, for $i = 1, .., N$ and $j = 1, ..., d$. Let T_ν be the univariate cdf of a t random variable with ν degrees of freedom. By making use of the inverse function, $T_\nu^{-1}(u_{i,j})$, each value $u_{i,j} \in \mathcal{U}$ is transformed to a point in the working space $w_{i,j} \in \mathcal{W}$, i.e

$$\mathcal{Y} \xrightarrow{F_j,\ j=1,\ldots,d} \mathcal{U} \xrightarrow{T_\nu^{-1}} \mathcal{W}. \tag{4.20}$$

The distribution in the working space is then estimated via a (usual) multivariate t-distribution, i.e letting $\lambda = \nu$. Note, that the procedure above is assuming complete data. However, the estimation approach is exactly the same when dealing with incomplete data.

4.3.1 Copula Estimation

This subsection describes in more detail the techniques used to estimate the parameters of the copulas, including the degrees of freedom ν and covariance matrix $\mathbf{\Sigma}$. As described above, the first step is to estimate the marginal distributions of the considered variables. Even if some of the variables have missing values the marginal distributions are estimated using the existing data. The data is then transformed to the working space using the inverse of a univariate Student's t cumulative distribution function (t-cdf) with ν degrees of freedom, see the transformation in Equation (4.20). Until now, the parameter, ν has been assumed to be known. In [13] and [14] the method of *k-fold cross validation* is used to estimate the degrees of freedom parameter, ν, in the set, $\{3, 10, 20, \infty\}$ of parameter values. However, this method is considered to be less suitable when the range of possible parameter values is more narrow. In this thesis, the data is known to be heavy tailed, and the question is rather which value the degrees of freedom should be assigned in the set, $\{3, 4, 5, \infty\}$. Note, for the case when $\nu = \infty$, the normal assumption regarding the data is applied. The problem of choosing the parameter becomes more obvious when it is time to estimate the degrees of freedom of the (usual) multivariate t-distribution in the common working space. The estimated degrees of freedom parameter in the working space should coincide with the one in the transformation when using the inverse of the t-cdf in the former step. The method used to estimate the parameter, ν, in this thesis is to initially use each of the values in the subset $\{3, 4, 5\}$ separately in the transformation to the working space. After all observations have been transformed it is time to estimate the corresponding degrees of freedom of the (usual) multivariate distribution function on the working space.

Estimation of the parameters ν and $\mathbf{\Sigma}$

Before the parameters, ν and $\mathbf{\Sigma}$ can be estimated on the common working space the variables with missing observations must be augmented. In [13] and [14] the *regularized EM* method, based on ridge regression, is used to **temporarily** fill in the missing values of the considered random variable to form complete data sets. However, in this thesis the stochastic regression imputation method, from Section 4.1, is used to temporarily construct complete data sets on the working space, see Algorithm 1. The covariance matrix, $\mathbf{\Sigma}$, is then computed on the working space data by using Equation (4.14), where the corresponding weights are attained from the ECM algorithm.

Finally, armed with complete data it is time to continue the estimation of the degrees of freedom parameter, ν. In this thesis the degree of freedom parameter of the (usual) multivariate t-distribution is estimated using the maximum likelihood estimation algorithm[12]. There is a problem of determining, ν, since the estimate of this parameter should also be used in the transformation to the working space via T_ν^{-1}. The method is hence to use each of the values of ν in the subset $\{3, 4, 5\}$ and see how much the parameter estimate of the degrees of freedom parameter of the (usual) multivariate t-distribution varies. In Table 4.4 it can be seen how the final number of degrees of freedom of the (usual) multivariate t-distribution is affected by the initial assumption regarding the degrees of freedom in the transformation, T_ν^{-1}, when $\nu \in \{3, 4, 5\}$, and with different amount of missing data (NaN). Here ν_{MLE} denotes the final maximum likelihood estimation of the degrees of freedom of the (usual) multivariate t-distribution on the working space.

	T_3^{-1}	T_4^{-1}	T_5^{-1}	NaN
ν_{MLE}	5.4202	5.4202	5.4202	0
ν_{MLE}	5.4148	5.4236	5.4040	3
ν_{MLE}	5.4159	5.4352	5.4361	10
ν_{MLE}	5.4690	5.3632	5.3720	25
ν_{MLE}	5.3241	5.1620	5.0132	50
ν_{MLE}	5.8079	6.4288	6.8772	100
ν_{MLE}	5.8712	6.5297	5.7734	150
ν_{MLE}	4.6699	6.2609	6.4192	250

Table 4.4: Overview of the effect on ν_{MLE} by assigning different values of ν during the transformation via T_ν^{-1} with different amount of missing data (NaN).

Remember, the parameter ν depends on the underlying data which is temporarily filled in using the stochastic regression imputation method in Section 4.1. However, it is considered to work as an indicator of which value of the degrees of freedom is best suited to be assigned to the corresponding copula. In this thesis the degrees of freedom is set to $\nu = 5$ for all cases of missing data. Note, as stated earlier, even if the (usual) multivariate t-distribution is used to estimate the dependence structure in the working space, the final conditional distribution of the (usual) multivariate t-distribution is in fact a generalized t-distribution.

4.3.2 Copula Simulation

In the copula estimation procedure the stochastic regression imputation method is used to temporarily fill in the missing values in the working space. In this section the complete data sets are obtained, containing both the observed values and the augmented set of imputed values sampled from the corresponding conditional distribution.

The goal is first to generate samples from the estimated copula on the working space, i.e Equation (4.19). This is made possible by using the representation in Equation (4.15). Note, in the copula estimation $\mathbf{\Sigma}$ was computed, and not $\mathbf{R}$. However, $\mathbf{R}$ is obtained from the relation in Equation (4.16) as follows

$$\mathbf{\Sigma} = \frac{\nu}{\nu - 2}\mathbf{R} \to \mathbf{R} = \frac{\nu - 2}{\nu}\mathbf{\Sigma}.$$

[12] Here the function, *copulafit*, in MATLAB is used to calculate the maximum likelihood estimation (MLE) of ν.

Under the normal assumption (i.e letting $\nu \to \infty$) the covariance matrix $\mathbf{\Sigma}$ is equal to $\mathbf{R}$. The generated samples on the working space are then transformed back to the raw data space.

Consider a $N \times d$ dimensional vector $\mathbf{Y}$ and a $N_{miss} \times 1$ dimensional vector of missing values $\mathbf{Y}_{miss}$. In Algorithm 3 is the simulation procedure presented when the data is assumed to be heavy tailed. The expectation, $\mathbf{m}$., and covariance, $\mathbf{\Sigma}$, are computed using Equation (4.13) and Equation (4.14) in Section 4.2. Additionally, the normal assumption regarding the data is applied by letting $\nu \to \infty$.

Algorithm 3: Copula Simulation

for $j = 1$ to d **do**
  Set $F_j(\mathbf{Y}_j) = \mathbf{U}_j$
  Set $\mathbf{W}_j = T_\nu^{-1}(\mathbf{U}_j)$
end for
Use Algorithm 1 in Section 4.1 to construct complete data sets on $\mathcal{W}$
Estimate $\mathbf{m}$ and $\mathbf{\Sigma}$ on $\mathcal{W}$
for $i = 1$ to N_{miss} **do**
  Set $\mathbf{m}_1(\mathbf{w}_{i,2}) = \mathbf{m}_1 + \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}(\mathbf{w}_{i,2} - \mathbf{m}_2)$
  Set $q(\mathbf{w}_{i,2}) = (\mathbf{w}_{i,2} - \mathbf{m}_2)\mathbf{\Sigma}_{22}^{-1}(\mathbf{w}_{i,2} - \mathbf{m}_2)$
  Set $\mathbf{R}_{\mathbf{W}_1|\mathbf{W}_2=\mathbf{w}_{i,2}} = \frac{\nu+(d-d_1)-2}{\nu+(d-d_1)}(\mathbf{\Sigma}_{11} - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21})$
  Generate $\mathbf{G}_i \sim \mathcal{N}\left(\left(\frac{\nu+q_{\mathbf{W}_1|\mathbf{W}_2=\mathbf{w}_2}}{\nu+(d-d_1)} \cdot \mathbf{R}_{\mathbf{W}_1|\mathbf{W}_2=\mathbf{w}_{i,2}}\right), 1\right)$
  Generate $\frac{1}{V_i} \sim Gamma\big(\nu + (d - d_1), \nu + (d - d_1)\big)$
  Set $\mathbf{W}_i = \mathbf{m}_1(\mathbf{w}_{i,2}) + V_i^{\frac{1}{2}}\mathbf{G}_i$
  Set $\mathbf{U}_i = T_\nu(\mathbf{W}_i)$
  Set $\mathbf{Y}_{miss,i} = F^{-1}(\mathbf{U}_i)$
end for

Chapter 5

Results

In this chapter the performance of each model is evaluated after being applied to a real data set. The data is provided by NASDAQ and consist of 4 years of daily relative returns[13] of monthly, quarterly and yearly commodity futures contracts from the Nordic Power market, with 2 month to maturity. The actual time period stretches between the 3^{rd} of January 2008 and the 5^{th} of July 2012. The relative returns are turned into percentage returns for numerical stability (i.e the relative returns are multiplied by 100).

The model setup has been to impute the missing return values of the quarterly futures contract conditioned upon the returns of the corresponding monthly and yearly futures contracts. The data was split into two sets, one set used to estimate the models and a second out-of-sample set used to evaluate the performance of each model. The missing data was planned to be missing in a uniform monotone pattern for the cases of $(0, 3, 10, 25, 50, 100, 150, 250)$ removed values. In other words, known data for the quarterly futures contract are removed in a cumulative manner where the data in the case of 3 missing values are also missing in the case of 10 missing values, and so on. Since the different imputation procedures are to be used in the preparatory work of stress testing, the considered data range has intentionally been chosen because of the increased volatility during this time period. Additionally, since the data is assumed to be missing during the crisis, the data is removed from 2008 and forward, and not from 2012 and backward.

Each model has separately been tested by assuming a normal distribution, and a more heavy tailed assumption by also assigning a Student's t-distribution for the data. In this chapter the performance of each model is evaluated by calculating the *Mean Absolute Error* (MAE), the *Root Mean Square Error* (RMSE) and the out-of-sample log-likelihood. The MAE and the RMSE measure are defined as follows

$$\begin{aligned} Mean\ Absolute\ Error &= \frac{1}{N_{miss}} \sum_{i=1}^{N_{miss}} \left| Y^i_{o-o-s} - Y^i_{pre} \right| \\ Root\ Mean\ Square\ Error &= \sqrt{\frac{1}{N_{miss}} \sum_{i=1}^{N_{miss}} \left(Y^i_{o-o-s} - Y^i_{pre} \right)^2}, \end{aligned}$$

where Y^i_{o-o-s} is the observed out-of-sample value and Y^i_{pre} the corresponding predicted value. Both the MAE and the RMSE are two accuracy measures. The MAE, as its name indicates, measures the absolute distance to the true value and weights all errors equally. In contrast, the RMSE is obtained by taking the square root of the *Mean Square Error* (MSE). The MSE is defined as the second moment of the errors and is given by

$$MSE = \frac{1}{N_{miss}} \sum_{i=1}^{N_{miss}} \left(Y^i_{o-o-s} - Y^i_{pre} \right)^2 = Var\left[Y^i_{o-o-s} - Y^i_{pre} \right] + \left(Bias\left(Y^i_{o-o-s} - Y^i_{pre} \right) \right)^2.$$

[13] The relative returns are calculated as $\frac{V_{t+1} - V_t}{V_t}$, where V_t is the value at time t.

Additionally, if the bias is zero the MSE is equal to the variance of the error, and thus the RMSE is equal to the standard deviation. Hence, the RMSE is preferred to the MSE as it is on the same scale as the data. For more information regarding error metrics see [37].

In Figure 5.1-5.2 and in Table 5.1-5.2 the calculated MAE and RMSE for each method can be seen, where "SR" and "MI" stands for "Stochastic Regression" and "Multiple Imputation". Since the MAE and the RMSE is calculated after generating predicted values from the corresponding conditional distribution, and both are very sensitive when used on small samples (typically $N_{miss} < 100$), these error metrics have been averaged over 100 iterations for a more clear interpretation.

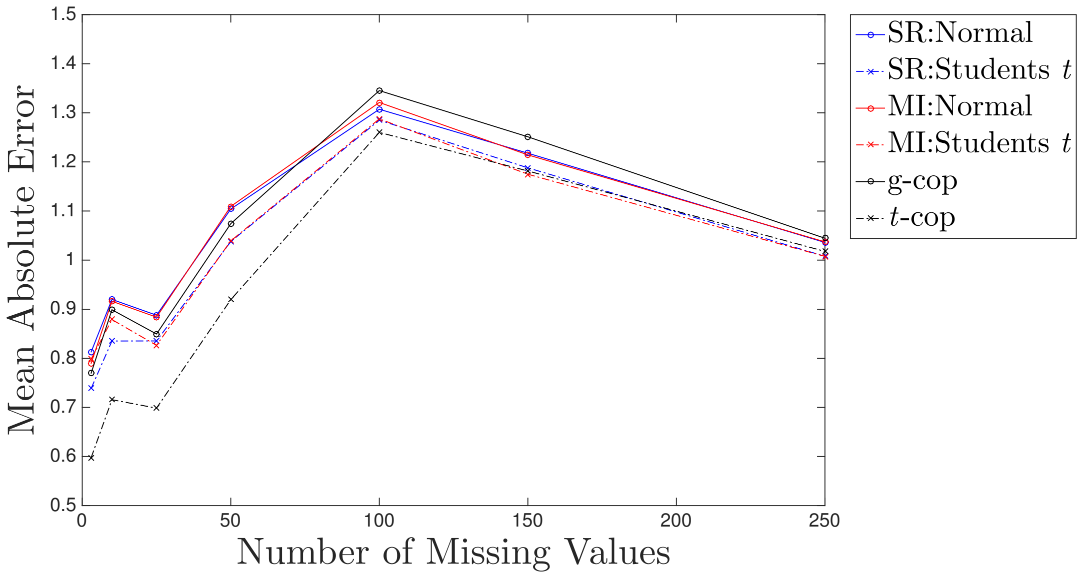

Figure 5.1: The Mean Absolute Error (MAE) for each model with different amount of missing data (NaN).

Mean Absolute Error (MAE)							
NaN	3	10	25	50	100	150	250
SR:normal	0.8130	0.9199	0.8878	1.1041	1.3073	1.2177	1.0360
SR:Student's t	0.7388	0.8352	0.8354	1.0378	1.2851	1.1879	1.0078
MI:normal	0.7900	0.9162	0.8838	1.1094	1.3211	1.2143	1.0371
MI:Student's t	0.7987	0.8789	0.8265	1.0391	1.2876	1.1749	1.0075
g-cop	0.7699	0.8989	0.8495	1.0742	1.3453	1.2512	1.0445
t-cop	0.5977	0.7157	0.6986	0.9195	1.2601	1.1816	1.0178

Table 5.1: The Mean Absolute Error (MAE) for each model with different amount of missing data (NaN).

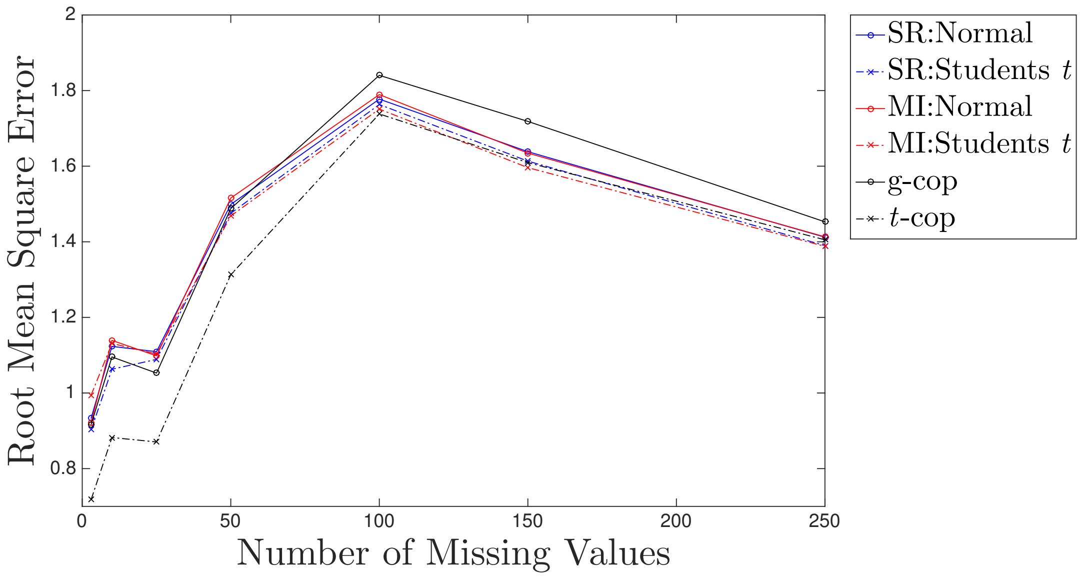

Figure 5.2: The Root Mean Square Error (RMSE) for each model with different amount of missing data (NaN).

Root Mean Square Error (RMSE)							
NaN	3	10	25	50	100	150	250
SR:normal	0.9336	1.1233	1.1093	1.4988	1.7774	1.6385	1.4121
SR:Student's t	0.9025	1.0629	1.0889	1.4754	1.7632	1.6134	1.3900
MI:normal	0.9224	1.1392	1.0984	1.5157	1.7894	1.6346	1.4131
MI:Student's t	0.9948	1.1311	1.1033	1.4685	1.7511	1.5962	1.3879
g-cop	0.9170	1.0954	1.0527	1.4879	1.8404	1.7185	1.4527
t-cop	0.7183	0.8822	0.8705	1.3129	1.7382	1.6095	1.4052

Table 5.2: The Root Mean Square Error (RMSE) for each model with different amount of missing data (NaN).

To test how well the models conform/fit the out-of-sample data, the out-of-sample log-likelihood for each model is calculated. The out-of-sample log-likelihood measures how well each model is able to predict the i^{th} observation. The likelihood functions of the stochastic regression and the multiple imputation model is given by Equation 4.3, where the normal case is implemented by letting $\nu \to \infty$. However, the likelihood for the copulas are not as easy to compute since a transformation from the working space $\mathcal{W}$ back to the raw variable space $\mathcal{Y}$ is required, see Proposition C.1 in Appendix C.

The out-of-sample log-likelihood for each model with different amount of missing data can be seen in Table 5.3.

Out-of-sample log-likelihood							
NaN	3	10	25	50	100	150	250
SR:normal	-2.9991	-11.9770	-28.0976	-97.1829	-336.5288	-448.7840	-571.0222
SR:Student's t	-2.6538	-12.2629	-28.7245	-76.2985	-228.2967	-325.0647	-449.6862
MI:normal	-3.0054	-11.9641	-28.1054	-97.4623	-336.0377	-448.8132	-570.5890
MI:Student's t	-2.6512	-12.2703	-28.7040	-76.4121	-228.3159	-325.2982	-450.3115
g-cop	-3.0067	-11.8347	-27.1434	-74.7699	-281.6717	-392.1448	-488.9255
t-cop	-2.9745	-13.6448	-32.2567	-88.7305	-221.7049	-335.5773	-433.8382

Table 5.3: The out-of-sample log-likelihood for each model with different amount of missing data (NaN).

Based on these metrics it is very hard to distinguish the performance of each method, or draw any conclusion about how good the models are in comparison to each other. Even if the Student's t-distribution seems (in general) to be a more adequate assumption regarding the data compared to the normal distribution (see Figure 5.1-5.2, Table 5.1-5.2 and Table 5.3), all the models are showing quite poor performance. However, this may not come as a surprise since they are implemented on very volatile data. Note that the accuracy of each method based on the MAE, RMSE and the out-of-sample log-likelihood when the models are applied on more stable data can be viewed in Appendix E. The results from using more stable data are included to emphasize that the performance of each model is very data dependent/sensitive. Moreover, even if the conditional distribution is used, and the data is strongly correlated, the conditional data does not have the same impact during a stressed market as it would during a more stable market. Therefore a more thorough analysis of the conditional distributions is required to be able to say more about the accuracy of each method. The idea is to look deeper into the tails and see how well each model performs by extracting certain quantile values from the respective conditional distribution. In other words, by considering the set of quantiles, $\alpha = \{0.001, 0.005, 0.05, 0.5, 0.95, 0.995, 0.999\}$, more extreme values can be imputed. Also, to investigate how sensitive each model is this approach is implemented by using two different scenarios, and is applied on the most volatile consecutive sequence of values in the out-of-sample data. The specific period stretches between the 3^{rd} of March in 2008 back to the 25^{th} of Feburary 2008. In the first scenario all the data back until the 4^{th} of March in 2008 is used to calibrate each model. In the second scenario 250 values after the 3^{rd} of March in 2008 are assumed to be missing (removed). In Table 5.4-5.21 the values from the most volatile period in the out-of-sample set can be seen together with the conditional values[14] and the corresponding estimated quantile values for $\alpha = \{0.05, 0.5, 0.95\}$ for each method. The absolute error[15] (distance) between the true value and the estimated value together with the percentage error[16] is also presented for the two different scenarios described above. The imputed values for the more extreme quantiles in the set $\alpha = \{0.001, 0.005, 0.995, 0.999\}$ can be seen in Table D.1-D.24 in Appendix D.

[14] *Month* denotes the corresponding relative returns for the monthly future contracts, and *Year* the yearly future contracts.

[15] $|\Delta| = |Y_{o-o-s} - Y_{pre}|$.

[16] $\Delta\% = \frac{Y_{o-o-s} - Y_{pre}}{Y_{o-o-s}}$.

SR:normal, $\alpha = 0.95$									
				$N_{miss} = 0$			$N_{miss} = 250$		
Date	Monthly	Yearly	Y_{o-o-s}	Y_{pre}	$\|\Delta\|$	$\Delta\%$	Y_{pre}	$\|\Delta\|$	$\Delta\%$
2008-03-03	-7.9176	-2.4645	-10.3876	-4.4107	5.9769	0.5754	-4.5136	5.8740	0.5655
2008-02-29	-4.8392	0	-5.9767	-1.3406	4.6361	0.7757	-1.4499	4.5268	0.7574
2008-02-28	0.4615	0.3806	1.3294	1.8233	0.4939	-0.3716	1.5231	0.1937	-0.1457
2008-02-27	7.4380	2.5366	7.4603	6.8969	0.5634	0.0755	6.4223	1.0380	0.1391
2008-02-26	-6.3467	-2.6591	-7.6246	-3.6420	3.9826	0.5223	-3.8158	3.8089	0.4995
2008-02-25	-7.7143	-0.5666	-8.2100	-3.2545	4.9554	0.6036	-3.2771	4.9329	0.6008

Table 5.4: An overview of the values from the most volatile period in the out-of-sample set together with the conditional values, and the corresponding estimated quantile values for $\alpha = 0.95$ for the SR:normal method. The absolute error (distance) between the true value and the estimated value together with the percentage error is also presented for two different scenarios.

SR:normal, $\alpha = 0.5$									
				$N_{miss} = 0$			$N_{miss} = 250$		
Date	Monthly	Yearly	Y_{o-o-s}	Y_{pre}	$\|\Delta\|$	$\Delta\%$	Y_{pre}	$\|\Delta\|$	$\Delta\%$
2008-03-03	-7.9176	-2.4645	-10.3876	-5.7323	4.6553	0.4482	-5.5538	4.8338	0.4653
2008-02-29	-4.8392	0	-5.9767	-2.6622	3.3145	0.5546	-2.4902	3.4865	0.5834
2008-02-28	0.4615	0.3806	1.3294	0.5017	0.8277	0.6226	0.4828	0.8466	0.6368
2008-02-27	7.4380	2.5366	7.4603	5.5753	1.8850	0.2527	5.3821	2.0782	0.2786
2008-02-26	-6.3467	-2.6591	-7.6246	-4.9636	2.6610	0.3490	-4.8560	2.7687	0.3631
2008-02-25	-7.7143	-0.5666	-8.2100	-4.5761	3.6338	0.4426	-4.3173	3.8926	0.4741

Table 5.5: An overview of the values from the most volatile period in the out-of-sample set together with the conditional values, and the corresponding estimated quantile values for $\alpha = 0.5$ for the SR:normal method. The absolute error (distance) between the true value and the estimated value together with the percentage error is also presented for two different scenarios.

SR:normal, $\alpha = 0.05$									
				$N_{miss} = 0$			$N_{miss} = 250$		
Date	Monthly	Yearly	Y_{o-o-s}	Y_{pre}	$\|\Delta\|$	$\Delta\%$	Y_{pre}	$\|\Delta\|$	$\Delta\%$
2008-03-03	-7.9176	-2.4645	-10.3876	-7.0539	3.3337	0.3209	-6.5940	3.7936	0.3652
2008-02-29	-4.8392	0	-5.9767	-3.9838	1.9929	0.3334	-3.5304	2.4463	0.4093
2008-02-28	0.4615	0.3806	1.3294	-0.8199	2.1493	1.6167	-0.5574	1.8868	1.4193
2008-02-27	7.4380	2.5366	7.4603	4.2537	3.2066	0.4298	4.3419	3.1184	0.4180
2008-02-26	-6.3467	-2.6591	-7.6246	-6.2852	1.3394	0.1757	-5.8962	1.7284	0.2267
2008-02-25	-7.7143	-0.5666	-8.2100	-5.8977	2.3122	0.2816	-5.3575	2.8524	0.3474

Table 5.6: An overview of the values from the most volatile period in the out-of-sample set together with the conditional values, and the corresponding estimated quantile values for $\alpha = 0.05$ for the SR:normal method. The absolute error (distance) between the true value and the estimated value together with the percentage error is also presented for two different scenarios.

SR:Student's t, $\alpha = 0.95$									
				$N_{miss} = 0$			$N_{miss} = 250$		
Date	Monthly	Yearly	Y_{o-o-s}	Y_{pre}	$\|\Delta\|$	$\Delta\%$	Y_{pre}	$\|\Delta\|$	$\Delta\%$
2008-03-03	-7.9176	-2.4645	-10.3876	-4.6754	5.7122	0.5499	-4.6151	5.7724	0.5557
2008-02-29	-4.8392	0	-5.9767	-1.6094	4.3673	0.7307	-1.5601	4.4165	0.7390
2008-02-28	0.4615	0.3806	1.3294	1.6782	0.3488	-0.2624	1.4767	0.1473	-0.1108
2008-02-27	7.4380	2.5366	7.4603	6.8586	0.6018	0.0807	6.4262	1.0341	0.1386
2008-02-26	-6.3467	-2.6591	-7.6246	-3.8597	3.7650	0.4938	-3.8921	3.7325	0.4895
2008-02-25	-7.7143	-0.5666	-8.2100	-3.5783	4.6317	0.5642	-3.4146	4.7954	0.5841

Table 5.7: An overview of the values from the most volatile period in the out-of-sample set together with the conditional values, and the corresponding estimated quantile values for $\alpha = 0.95$ for the SR:Student's t method. The absolute error (distance) between the true value and the estimated value together with the percentage error is also presented for two different scenarios.

SR:Student's t, $\alpha = 0.5$									
				$N_{miss} = 0$			$N_{miss} = 250$		
Date	Monthly	Yearly	Y_{o-o-s}	Y_{pre}	$\|\Delta\|$	$\Delta\%$	Y_{pre}	$\|\Delta\|$	$\Delta\%$
2008-03-03	-7.9176	-2.4645	-10.3876	-5.8722	4.5154	0.4347	-5.6107	4.7769	0.4599
2008-02-29	-4.8392	0	-5.9767	-2.8062	3.1705	0.5305	-2.5557	3.4210	0.5724
2008-02-28	0.4615	0.3806	1.3294	0.4814	0.8480	0.6379	0.4811	0.8483	0.6381
2008-02-27	7.4380	2.5366	7.4603	5.6618	1.7985	0.2411	5.4307	2.0296	0.2721
2008-02-26	-6.3467	-2.6591	-7.6246	-5.0564	2.5682	0.3368	-4.8877	2.7370	0.3590
2008-02-25	-7.7143	-0.5666	-8.2100	-4.7751	3.4349	0.4184	-4.4101	3.7998	0.4628

Table 5.8: An overview of the values from the most volatile period in the out-of-sample set together with the conditional values, and the corresponding estimated quantile values for $\alpha = 0.5$ for the SR:Student's t method. The absolute error (distance) between the true value and the estimated value together with the percentage error is also presented for two different scenarios.

SR:Student's t, $\alpha = 0.05$									
				$N_{miss} = 0$			$N_{miss} = 250$		
Date	Monthly	Yearly	Y_{o-o-s}	Y_{pre}	$\|\Delta\|$	$\Delta\%$	Y_{pre}	$\|\Delta\|$	$\Delta\%$
2008-03-03	-7.9176	-2.4645	-10.3876	-7.0690	3.3186	0.3195	-6.6062	3.7814	0.3640
2008-02-29	-4.8392	0	-5.9767	-4.0030	1.9737	0.3302	-3.5512	2.4255	0.4058
2008-02-28	0.4615	0.3806	1.3294	-0.7153	2.0447	1.5381	-0.5144	1.8438	1.3869
2008-02-27	7.4380	2.5366	7.4603	4.4650	2.9953	0.4015	4.4352	3.0251	0.4055
2008-02-26	-6.3467	-2.6591	-7.6246	-6.2532	1.3714	0.1799	-5.8832	1.7414	0.2284
2008-02-25	-7.7143	-0.5666	-8.2100	-5.9719	2.2381	0.2726	-5.4057	2.8043	0.3416

Table 5.9: An overview of the values from the most volatile period in the out-of-sample set together with the conditional values, and the corresponding estimated quantile values for $\alpha = 0.05$ for the SR:Student's t method. The absolute error (distance) between the true value and the estimated value together with the percentage error is also presented for two different scenarios.

MI:normal, $\alpha = 0.95$									
				$N_{miss} = 0$			$N_{miss} = 250$		
Date	Monthly	Yearly	Y_{o-o-s}	Y_{pre}	$\|\Delta\|$	$\Delta\%$	Y_{pre}	$\|\Delta\|$	$\Delta\%$
2008-03-03	-7.9176	-2.4645	-10.3876	-4.4163	5.9713	0.5748	-4.4920	5.8956	0.5676
2008-02-29	-4.8392	0	-5.9767	-1.3384	4.6382	0.7761	-1.4267	4.5500	0.7613
2008-02-28	0.4615	0.3806	1.3294	1.8261	0.4967	-0.3736	1.5293	0.1999	-0.1503
2008-02-27	7.4380	2.5366	7.4603	6.9058	0.5545	0.0743	6.4146	1.0457	0.1402
2008-02-26	-6.3467	-2.6591	-7.6246	-3.6485	3.9762	0.5215	-3.8008	3.8238	0.5015
2008-02-25	-7.7143	-0.5666	-8.2100	-3.2539	4.9560	0.6037	-3.2465	4.9635	0.6046

Table 5.10: An overview of the values from the most volatile period in the out-of-sample set together with the conditional values, and the corresponding estimated quantile values for $\alpha = 0.95$ for the MI:normal method. The absolute error (distance) between the true value and the estimated value together with the percentage error is also presented for two different scenarios.

MI:normal, $\alpha = 0.5$									
				$N_{miss} = 0$			$N_{miss} = 250$		
Date	Monthly	Yearly	Y_{o-o-s}	Y_{pre}	$\|\Delta\|$	$\Delta\%$	Y_{pre}	$\|\Delta\|$	$\Delta\%$
2008-03-03	-7.9176	-2.4645	-10.3876	-5.7180	4.6696	0.4495	-5.5226	4.8650	0.4684
2008-02-29	-4.8392	0	-5.9767	-2.6501	3.3266	0.5566	-2.4619	3.5148	0.5881
2008-02-28	0.4615	0.3806	1.3294	0.5071	0.8223	0.6186	0.4927	0.8367	0.6294
2008-02-27	7.4380	2.5366	7.4603	5.5730	1.8873	0.2530	5.3735	2.0868	0.2797
2008-02-26	-6.3467	-2.6591	-7.6246	-4.9515	2.6731	0.3506	-4.8313	2.7934	0.3664
2008-02-25	-7.7143	-0.5666	-8.2100	-4.5607	3.6493	0.4445	-4.2803	3.9297	0.4786

Table 5.11: An overview of the values from the most volatile period in the out-of-sample set together with the conditional values, and the corresponding estimated quantile values for $\alpha = 0.5$ for the MI:normal method. The absolute error (distance) between the true value and the estimated value together with the percentage error is also presented for two different scenarios.

MI:normal, $\alpha = 0.05$									
				$N_{miss} = 0$			$N_{miss} = 250$		
Date	Monthly	Yearly	Y_{o-o-s}	Y_{pre}	$\|\Delta\|$	$\Delta\%$	Y_{pre}	$\|\Delta\|$	$\Delta\%$
2008-03-03	-7.9176	-2.4645	-10.3876	-7.0585	3.3291	0.3205	-6.5621	3.8255	0.3683
2008-02-29	-4.8392	0	-5.9767	-3.9866	1.9901	0.3330	-3.5017	2.4749	0.4141
2008-02-28	0.4615	0.3806	1.3294	-0.8206	2.1500	1.6173	-0.5511	1.8805	1.4146
2008-02-27	7.4380	2.5366	7.4603	4.2561	3.2042	0.4295	4.3258	3.1345	0.4202
2008-02-26	-6.3467	-2.6591	-7.6246	-6.2892	1.3354	0.1751	-5.8722	1.7524	0.2298
2008-02-25	-7.7143	-0.5666	-8.2100	-5.9018	2.3081	0.2811	-5.3183	2.8916	0.3522

Table 5.12: An overview of the values from the most volatile period in the out-of-sample set together with the conditional values, and the corresponding estimated quantile values for $\alpha = 0.05$ for the MI:normal method. The absolute error (distance) between the true value and the estimated value together with the percentage error is also presented for two different scenarios.

MI:Student's t, $\alpha = 0.95$									
				$N_{miss} = 0$			$N_{miss} = 250$		
Date	Monthly	Yearly	Y_{o-o-s}	Y_{pre}	$\|\Delta\|$	$\Delta\%$	Y_{pre}	$\|\Delta\|$	$\Delta\%$
2008-03-03	-7.9176	-2.4645	-10.3876	-2.9315	7.4561	0.7178	-4.5888	5.7988	0.5582
2008-02-29	-4.8392	0	-5.9767	0.1342	6.1108	1.0224	-1.6037	4.3729	0.7317
2008-02-28	0.4615	0.3806	1.3294	3.4165	2.0871	-1.5700	1.4619	0.1325	-0.0997
2008-02-27	7.4380	2.5366	7.4603	8.5919	1.1316	-0.1517	6.3855	1.0748	0.1441
2008-02-26	-6.3467	-2.6591	-7.6246	-2.1177	5.5070	0.7223	-3.8454	3.7792	0.4957
2008-02-25	-7.7143	-0.5666	-8.2100	-1.8323	6.3777	0.7768	-3.4599	4.7500	0.5786

Table 5.13: An overview of the values from the most volatile period in the out-of-sample set together with the conditional values, and the corresponding estimated quantile values for $\alpha = 0.95$ for the MI:Student's t method. The absolute error (distance) between the true value and the estimated value together with the percentage error is also presented for two different scenarios.

MI:Student's t, $\alpha = 0.5$									
				$N_{miss} = 0$			$N_{miss} = 250$		
Date	Monthly	Yearly	Y_{o-o-s}	Y_{pre}	$\|\Delta\|$	$\Delta\%$	Y_{pre}	$\|\Delta\|$	$\Delta\%$
2008-03-03	-7.9176	-2.4645	-10.3876	-5.8754	4.5122	0.4344	-5.5796	4.8080	0.4629
2008-02-29	-4.8392	0	-5.9767	-2.8076	3.1691	0.5302	-2.5964	3.3802	0.5656
2008-02-28	0.4615	0.3806	1.3294	0.4820	0.8473	0.6374	0.4614	0.8680	0.6530
2008-02-27	7.4380	2.5366	7.4603	5.6655	1.7948	0.2406	5.3766	2.0837	0.2793
2008-02-26	-6.3467	-2.6591	-7.6246	-5.0591	2.5655	0.3365	-4.8389	2.7857	0.3654
2008-02-25	-7.7143	-0.5666	-8.2100	-4.7776	3.4323	0.4181	-4.4488	3.7612	0.4581

Table 5.14: An overview of the values from the most volatile period in the out-of-sample set together with the conditional values, and the corresponding estimated quantile values for $\alpha = 0.5$ for the MI:Student's t method. The absolute error (distance) between the true value and the estimated value together with the percentage error is also presented for two different scenarios.

MI:Student's t, $\alpha = 0.05$									
				$N_{miss} = 0$			$N_{miss} = 250$		
Date	Monthly	Yearly	Y_{o-o-s}	Y_{pre}	$\|\Delta\|$	$\Delta\%$	Y_{pre}	$\|\Delta\|$	$\Delta\%$
2008-03-03	-7.9176	-2.4645	-10.3876	-7.0711	3.3165	0.3193	-6.5805	3.8071	0.3665
2008-02-29	-4.8392	0	-5.9767	-4.0036	1.9730	0.3301	-3.5949	2.3818	0.3985
2008-02-28	0.4615	0.3806	1.3294	-0.7149	2.0443	1.5378	-0.5379	1.8673	1.4047
2008-02-27	7.4380	2.5366	7.4603	4.4674	2.9929	0.4012	4.3783	3.0820	0.4131
2008-02-26	-6.3467	-2.6591	-7.6246	-6.2551	1.3695	0.1796	-5.8404	1.7842	0.2340
2008-02-25	-7.7143	-0.5666	-8.2100	-5.9732	2.2367	0.2724	-5.4472	2.7628	0.3365

Table 5.15: An overview of the values from the most volatile period in the out-of-sample set together with the conditional values, and the corresponding estimated quantile values for $\alpha = 0.05$ for the MI:Student's t method. The absolute error (distance) between the true value and the estimated value together with the percentage error is also presented for two different scenarios.

g-cop, $\alpha = 0.95$									
				$N_{miss} = 0$			$N_{miss} = 250$		
Date	Monthly	Yearly	Y_{o-o-s}	Y_{pre}	$\|\Delta\|$	$\Delta\%$	Y_{pre}	$\|\Delta\|$	$\Delta\%$
2008-03-03	-7.9176	-2.4645	-10.3876	-4.2304	6.1572	0.5927	-4.3909	5.9967	0.5773
2008-02-29	-4.8392	0	-5.9767	-1.4468	4.5299	0.7579	-1.4808	4.4959	0.7522
2008-02-28	0.4615	0.3806	1.3294	1.7823	0.4529	0.3407	1.4994	0.1700	0.1279
2008-02-27	7.4380	2.5366	7.4603	7.3363	0.1241	0.0166	6.7530	0.7073	0.0948
2008-02-26	-6.3467	-2.6591	-7.6246	-3.6314	3.9932	0.5237	-3.7924	3.8322	0.5026
2008-02-25	-7.7143	-0.5666	-8.2100	-3.0412	5.1687	0.6296	-3.0559	5.1541	0.6278

Table 5.16: An overview of the values from the most volatile period in the out-of-sample set together with the conditional values, and the corresponding estimated quantile values for $\alpha = 0.95$ for the g-copula method. The absolute error (distance) between the true value and the estimated value together with the percentage error is also presented for two different scenarios.

g-cop, $\alpha = 0.5$									
				$N_{miss} = 0$			$N_{miss} = 250$		
Date	Monthly	Yearly	Y_{o-o-s}	Y_{pre}	$\|\Delta\|$	$\Delta\%$	Y_{pre}	$\|\Delta\|$	$\Delta\%$
2008-03-03	-7.9176	-2.4645	-10.3876	-5.7383	4.6493	0.4476	-5.7453	4.6423	0.4469
2008-02-29	-4.8392	0	-5.9767	-2.6817	3.2950	0.5513	-2.5241	3.4526	0.5777
2008-02-28	0.4615	0.3806	1.3294	0.5522	0.7772	0.5846	0.5399	0.7895	0.5939
2008-02-27	7.4380	2.5366	7.4603	5.5107	1.9497	0.2613	5.2722	2.1881	0.2933
2008-02-26	-6.3467	-2.6591	-7.6246	-5.0631	2.5615	0.3359	-5.0473	2.5774	0.3380
2008-02-25	-7.7143	-0.5666	-8.2100	-4.4069	3.8030	0.4632	-4.2660	3.9440	0.4804

Table 5.17: An overview of the values from the most volatile period in the out-of-sample set together with the conditional values, and the corresponding estimated quantile values for $\alpha = 0.5$ for the g-copula method. The absolute error (distance) between the true value and the estimated value together with the percentage error is also presented for two different scenarios.

g-cop, $\alpha = 0.05$									
				$N_{miss} = 0$			$N_{miss} = 250$		
Date	Monthly	Yearly	Y_{o-o-s}	Y_{pre}	$\|\Delta\|$	$\Delta\%$	Y_{pre}	$\|\Delta\|$	$\Delta\%$
2008-03-03	-7.9176	-2.4645	-10.3876	-7.4611	2.9265	0.2817	-7.3039	3.0837	0.2969
2008-02-29	-4.8392	0	-5.9767	-4.0092	1.9675	0.3292	-3.6225	2.3542	0.3939
2008-02-28	0.4615	0.3806	1.3294	-0.6387	1.9681	1.4805	-0.3821	1.7115	1.2874
2008-02-27	7.4380	2.5366	7.4603	3.9256	3.5348	0.4738	4.0158	3.4445	0.4617
2008-02-26	-6.3467	-2.6591	-7.6246	-6.6826	0.9420	0.1236	-6.4893	1.1353	0.1489
2008-02-25	-7.7143	-0.5666	-8.2100	-5.9350	2.2750	0.2771	-5.5921	2.6179	0.3189

Table 5.18: An overview of the values from the most volatile period in the out-of-sample set together with the conditional values, and the corresponding estimated quantile values for $\alpha = 0.05$ for the g-copula method. The absolute error (distance) between the true value and the estimated value together with the percentage error is also presented for two different scenarios.

t-cop, $\alpha = 0.95$									
				$N_{miss} = 0$			$N_{miss} = 250$		
Date	Monthly	Yearly	Y_{o-o-s}	Y_{pre}	$\|\Delta\|$	$\Delta\%$	Y_{pre}	$\|\Delta\|$	$\Delta\%$
2008-03-03	-7.9176	-2.4645	-10.3876	-4.9212	5.4664	0.5262	-4.3924	5.9952	0.5772
2008-02-29	-4.8392	0	-5.9767	-2.0292	3.9474	0.6605	-1.4568	4.5199	0.7563
2008-02-28	0.4615	0.3806	1.3294	1.2423	0.0871	0.0655	1.5377	0.2083	0.1567
2008-02-27	7.4380	2.5366	7.4603	6.2975	1.1628	0.1559	6.1186	1.3417	0.1798
2008-02-26	-6.3467	-2.6591	-7.6246	-4.2011	3.4236	0.4490	-3.7665	3.8582	0.5060
2008-02-25	-7.7143	-0.5666	-8.2100	-3.9074	4.3025	0.5241	-3.2295	4.9805	0.6066

Table 5.19: An overview of the values from the most volatile period in the out-of-sample set together with the conditional values, and the corresponding estimated quantile values for $\alpha = 0.95$ for the t-copula method. The absolute error (distance) between the true value and the estimated value together with the percentage error is also presented for two different scenarios.

t-cop, $\alpha = 0.5$									
				$N_{miss} = 0$			$N_{miss} = 250$		
Date	Monthly	Yearly	Y_{o-o-s}	Y_{pre}	$\|\Delta\|$	$\Delta\%$	Y_{pre}	$\|\Delta\|$	$\Delta\%$
2008-03-03	-7.9176	-2.4645	-10.3876	-5.8175	4.5701	0.4400	-5.7156	4.6720	0.4498
2008-02-29	-4.8392	0	-5.9767	-2.9775	2.9991	0.5018	-2.8468	3.1299	0.5237
2008-02-28	0.4615	0.3806	1.3294	0.5076	0.8217	0.6181	0.5080	0.8214	0.6179
2008-02-27	7.4380	2.5366	7.4603	5.4621	1.9982	0.2678	5.1876	2.2727	0.3046
2008-02-26	-6.3467	-2.6591	-7.6246	-5.0123	2.6124	0.3426	-4.9805	2.6441	0.3468
2008-02-25	-7.7143	-0.5666	-8.2100	-4.9899	3.2201	0.3922	-4.8050	3.4049	0.4147

Table 5.20: An overview of the values from the most volatile period in the out-of-sample set together with the conditional values, and the corresponding estimated quantile values for $\alpha = 0.5$ for the t-copula method. The absolute error (distance) between the true value and the estimated value together with the percentage error is also presented for two different scenarios.

t-cop, $\alpha = 0.05$									
				$N_{miss} = 0$			$N_{miss} = 250$		
Date	Monthly	Yearly	Y_{o-o-s}	Y_{pre}	$\|\Delta\|$	$\Delta\%$	Y_{pre}	$\|\Delta\|$	$\Delta\%$
2008-03-03	-7.9176	-2.4645	-10.3876	-6.6062	3.7814	0.3640	-6.7073	3.6803	0.3543
2008-02-29	-4.8392	0	-5.9767	-3.8308	2.1459	0.3590	-4.0086	1.9681	0.3293
2008-02-28	0.4615	0.3806	1.3294	-0.2210	1.5504	1.1662	-0.5444	1.8738	1.4095
2008-02-27	7.4380	2.5366	7.4603	4.5426	2.9178	0.3911	3.9672	3.4931	0.4682
2008-02-26	-6.3467	-2.6591	-7.6246	-5.7389	1.8857	0.2473	-5.8987	1.7260	0.2264
2008-02-25	-7.7143	-0.5666	-8.2100	-5.9278	2.2821	0.2780	-6.0259	2.1840	0.2660

Table 5.21: An overview of the values from the most volatile period in the out-of-sample set together with the conditional values, and the corresponding estimated quantile values for $\alpha = 0.05$ for the t-copula method. The absolute error (distance) between the true value and the estimated value together with the percentage error is also presented for two different scenarios.

By investigating Table 5.4-5.21 and Table D.1-D.24, it can be seen that the performance of each method can be significantly increased by imputing more extreme quantile values. By letting the median ($\alpha = 0.5$) of the corresponding conditional distribution be a representative for the standard method (where samples are randomly drawn from the corresponding conditional distribution), the standard method can be compared with the approach of imputing more extreme quantile values. In Figure 5.1-5.2 and Table 5.1-5.2 it can be seen that the standard method is insufficient when applied on more volatile data. However, the standard approach can be used to compute a first estimate of the final imputed value. This value can then be substituted by lower extreme values ($\alpha = 0.05$) or higher extreme values ($\alpha = 0.95$) depending on whether the corresponding conditional values are greater or smaller then this first estimate. In other words, the results in Table 5.4-5.21 and Table D.1-D.24 indicate that it is required to extract lower quantile values if the median value is greater than the corresponding Monthly future value, and vice versa. Even if there is an obvious trend in the results the dependence structure regarding the conditional values, quantile values and median values is very vague and requires to be more thoroughly analyzed before any real conclusions can be made.

Additionally, by calculating the *absolute difference* between the percentage error ($\Delta\%$) for $N_{miss} = 0$ and $N_{miss} = 250$ it can be concluded that the g-copula and t-copula models are more robust than the respective linear models, see Table 5.22.

			$\|\Delta\%_{N_{miss}=0} - \Delta\%_{N_{miss}=250}\|$					
α	0.999	0.995	0.95	0.5	0.05	0.005	0.001	Average
SR:normal	0.1150	0.0944	0.0572	0.0219	0.0744	0.1116	0.1322	0.1011
SR:Student's t	0.6452	0.2646	0.0413	0.0275	0.0655	0.2988	0.6794	0.2889
MI:normal	0.1131	0.0917	0.0554	0.0230	0.0778	0.1167	0.1363	0.0877
MI:Student's t	0.5839	0.2670	0.4402	0.0312	0.0632	0.2927	0.6725	0.3358
g-cop	0.1178	0.1041	0.0558	0.0146	0.0587	0.0912	0.1193	0.0802
t-cop	0.2348	0.1862	0.0669	0.0159	0.0655	0.1371	0.1739	0.1257

Table 5.22: The mean absolute difference (MD) between the percentage error ($\Delta\%$) for $N_{miss} = 0$ and $N_{miss} = 250$.

Although the more heavy tailed assumption in general seems to be more accurate (see Figure 5.1-5.2 and Table 5.1-5.2), using the g-copula and impute values from the extreme quantile $\alpha = 0.0001$ seems to give very satisfying results, see Table 5.23.

				g-cop, $\alpha = 0.0001$					
					$N_{miss} = 0$			$N_{miss} = 250$	
Date	Monthly	Yearly	Y_{o-o-s}	Y_{pre}	$\|\Delta\|$	$\Delta\%$	Y_{pre}	$\|\Delta\|$	$\Delta\%$
2008-03-03	-7.9176	-2.4645	-10.3876	-10.0661	0.3215	0.0309	-9.7969	0.5907	0.0569
2008-02-29	-4.8392	0	-5.9767	-5.9064	0.0702	0.0118	-5.2828	0.6939	0.1161
2008-02-28	0.4615	0.3806	1.3294	-2.1607	3.4901	2.6253	-1.6686	2.9980	2.2552
2008-02-27	7.4380	2.5366	7.4603	2.1809	5.2794	0.7077	2.4865	4.9738	0.6667
2008-02-26	-6.3467	-2.6591	-7.6246	-9.1100	1.4854	0.1948	-8.8006	1.1760	0.1542
2008-02-25	-7.7143	-0.5666	-8.2100	-8.2017	0.0083	0.0010	-7.6464	0.5636	0.0686

Table 5.23: An overview of the values from the most volatile period in the out-of-sample set together with the conditional values, and the corresponding estimated quantile values for $\alpha = 0.0001$ for the g-copula method. The absolute error (distance) between the true value and the estimated value together with the percentage error is also presented for two different scenarios.

Chapter 6

Conclusions

In this thesis three different methods have been investigated for imputation of missing data concerning commodity futures returns in the context of the preparatory work regarding stress testing. To retain the variability of the data the chosen approach has been to impute historical return data by the conditional distribution rather than the conditional expectation. This approach also has the advantage that certain extreme quantile values can be extracted from each parametric model, and used in different stress testing scenarios.

The two first models implemented in this paper are currently regarded as state-of-the-art techniques for managing missing data. The first method, Stochastic Regression Imputation, uses a Bayesian approach to find the corresponding MAP-estimate of the coefficients in the regression equation. The second model, Multiple Imputation (MI), uses a full[17] implementation of the Bayesian theory, and by simulation generates new covariances and mean values used to construct multiple sets of augmented data chains. Although the first and second approach tend to produce similar results when the regression equations includes the same variables, the latter has the advantage of not depending solely on one imputed value for each missing data event. The third alternative method implemented in this thesis is based on elliptical copulas and the beneficial properties of the Arellano-Valle and Bolfarine's generalized t-distribution [21]. In contrast to the first two approaches the marginal distribution of each random variable is modelled separately, and the dependence structure is applied in a non-linear manner in the third model.

The characteristics of the data is known to be non-normal. However, both a normal distribution and a more heavy tailed assumption, in form of a t-distribution, has been assigned to the missing data to study the validity of the model assumptions for each method separately. However, it is well known that the normal assumption is a very naive approach since it does not regard any possible tail dependence. This means that the probability that two random variables simultaneously attain extreme values is zero, even if the variables are correlated. In contrast, the tail dependence under the t-distribution increases with correlation, and decreases when the number of degrees of freedom increases. However, to attain the accuracy of each method more thoroughly some conventional error metrics like the Mean Absolute Error, Root Mean Square Error and out-of-sample log-likelihood are used.

Based on these metrics it is very hard to distinguish the performance of each method, or draw any conclusion about how good the models are in comparison to each other. Even if the Student's t-distribution seems (in general) to be a more adequate assumption regarding the data compared to the normal distribution (see Figure 5.1-5.2, Table 5.1-5.2 and Table 5.3), all the models are showing quite poor performance. However, by analyzing the conditional distributions more thoroughly, i.e. by looking deeper into the tails and extracting more extreme quantile values the performance of each method is increased significantly. Additionally, by calculating the absolute difference between the percentage error ($\Delta\%$) for $N_{miss} = 0$ and $N_{miss} = 250$ it can be concluded that the g-copula and t-copula models are more robust than the respective linear models, see Table 5.22.

Note, even if the results in Table 5.4-5.21, Table 5.23 and Table D.1-D.24 show very good performance the models are calibrated when the true data is known. Hence, these results are

[17] In the sense that the distribution of each parameter is considered and not only a single estimate.

just presented to show that the models can actually perform very well if the distributions are more thoroughly investigated. Therefore, an important and interesting topic for future studies would be to investigate how sensitive the conditional distributions are, concerning the conditional data, to see how far into the tails it is required to go before any reasonable results can be obtained.

Moreover, the aim of this thesis has been to provide a more sophisticated method used to impute missing historical data in the preparatory work of portfolio stress testing. However, the methods used are only implemented on one type of data and missing data pattern. Hence, in the future it would be interesting to apply these methods on a broader set of data/financial instruments, and in a context where the missing data is missing in a non-monotone univariate pattern.

Finally, a natural way of proceeding this work would be to investigate the influence these imputation methods have on the profit and loss distribution compared to current routines. In other words, it would be interesting to see how the expected future losses and the amount of capital needed to be secured is affected when these new tractable methods are implemented in the context of portfolio stress testing.

Bibliography

[1] Burden, R. L. and Faires, J. D. 2011.*Solutions of Equations in One Variable - The Bisection Method*. Beamer Presentation Slides prepared by John Carroll Dublin City University. Available at: https://www.math.ust.hk/~mamu/courses/231/Slides/ch02_1.pdf, (Retrieved 2016-02-26).

[2] Carter, Tom. 2014. *An introduction to information theory and entropy*. California State University. Available at: http://csustan.csustan.edu/~tom/Lecture-Notes/Information-Theory/info-lec.pdf, (Retrieved 2016-01-20).

[3] Chakrabarti, Brato. 2011. *Bisection Method*.
Available at: http://www.mathworks.com/matlabcentral/fileexchange/33748-bisection-method/all_files, (Retrieved 2016-02-26).

[4] Dall'Aglio, G., Kotz, S., Salinetti, G. 1991. *Advances In Probability Distributions with Given Marginals : Beyond the Copulas*. Vol. 67. Dordrecht: Springer Science+Business Media Dordrecht.

[5] Demarta, Stefano and McNeil, Alexander J. 2004. *The t Copula and Related Copulas*. Department of Mathematics Federal Institute of Technology, ETH, Zürich.
Available at: http://clone.risknet.de/fileadmin/eLibrary/t-Copula-Demarta-ETH.pdf, (Retrieved 2016-01-20).

[6] Dempster, Arthur P., Laird, Nan M. and Dempster, A., and Rubin, Donald B. 1977. *Maximum Likelihood from Incomplete Data via the EM Algorithm*. Journal of the Royal Statistical Society. Series B (Methodological), Vol. 39, No. 1, pp. 1-38.

[7] Embrechts, Paul, Lindskog, Filip and McNeil, Alexander. 2001. *Modelling Dependence with Copulas and Applications to Risk Management*. Department of Mathematics, ETHZ, CH-8092 Zürich, Switzerland. Available at: https://people.math.ethz.ch/~embrecht/ftp/copchapter.pdf, (Retrieved 2016-01-17).

[8] Embrechts, P., McNeil A. and Straumann D. 2009. *Correlation and Dependency in Risk Management: Properties and Pitfalls*. The Journal of Risk and Insurance. Series B, Vol. 76, No. 3, pp. 639-650.

[9] Embrechts, Paul. 2009. *Copulas: A Personal View*. The Journal of Risk and Insurance. Series B, Vol. 76, No. 3, pp. 639-650.

[10] Enders, Craig K. 2010. *Applied Missing Data Analysis*. New York: The Guilford Press.

[11] European Securities and Markets Authority. 2015. *WHO WE ARE*.
Available at: https://www.esma.europa.eu/about-esma/who-we-are, (Retrieved 2016-01-20).

[12] European Securities and Markets Authority. 2015. *2016 Work Programme*.
Available at: https://www.esma.europa.eu/sites/default/files/library/2015/10/2015-1475_esma_2016_work_programme.pdf, (Retrieved 2016-01-20).

[13] Friedman, C., Huang, J., Zhang, Y. and Cao, W. 2009. *Joint and conditional transformed t-mixture models with applications to financial and economic data*. The Journal of Risk, Vol. 11, No. 3, pp. 25-54.

[14] Friedman, C., Huang, J., Zhang, Y. and Cao, W. 2012. *Missing value imputation via copula and transformation methods, with applications to financial and economic data*. Int. J. Data Analysis Techniques and Strategies, Vol. 4, No. 4, pp. 315-339.

[15] Frost, Jim. 2013. *Regression Analysis: How to Interpret the Constant (Y Intercept)*. The Minitab Blog. [Blog]. 11 July. Available at: http://blog.minitab.com/blog/adventures-in-statistics/regression-analysis-how-to-interpret-the-constant-y-intercept, (Retrieved 2015-11-11).

[16] Hansen, Bruce E. 2000. *Econometrics*. University of Wisconsin. Available at: http://www.ssc.wisc.edu/~bhansen/econometrics/, (Retrieved 2016-01-20).

[17] Hedensiö, Martin. 2014. *NASDAQ OMX CLEARING RECEIVES EMIR AUTHORISATION*. NASDAQ OMX. Available at: http://www.nasdaqomx.com/newsroom/pressreleases/pressrelease?messageId=1253167&displayLanguage=en, (Retrieved 2016-01-20).

[18] Jones, D. L. 2014. *The Johnson Curve Toolbox for Matlab: analysis of non-normal data using the Johnson system of distributions*. College of Marine Science, University of South Florida, St. Petersburg, Florida, USA. Available at: http://www.mathworks.com/matlabcentral/fileexchange/46123-johnson-curve-toolbox, (Retrieved 2016-01-17).

[19] Käärik, Ene. 2005. *Handling dropouts by copulas*. WSEAS Transactions on Biology and Biomedicine, Vol. 1, No. 2, pp. 93-97.

[20] Käärik, Ene. 2006. *Imputation Algorithm Using Copulas*. Metodoloski zvezki, Vol. 3, No. 1, pp. 109-120.

[21] Kotz, S. and Nadarajah S. 2004. *Multivariate t-Distributions and Their Applications*. Cambridge.

[22] Lang, Harald. 2014. *Elements of Regression Analysis*. Royal Institute of Technology, Stockholm.

[23] Liu, Chuanhai. 1994. *Statistical Analysis Using the Multivariate t Distribution*. Department of Statistics Harvard University. Available at: www.geog.ucsb.edu/~dongmei/chuanhai.ps, (Retrieved 2016-02-15).

[24] McNeil, Alexander J., Rüdiger Frey., Embrechts Paul. 2005. *Quantitative Risk Management: Concepts, techniques and Tools*. Princeton and Oxford: Princeton University Press.

[25] Meng, Xiao L. and Rubin, Donald B. 1993. *Maximum Likelihood via the ECM Algorithm: A General Framework*. Biometrika, Vol. 80, No. 2, pp. 267-278.

[26] NASDAQ OMX Group. April 2015. *NASDAQ Clearing as a Counterparty*. Available at: http://www.nasdaqomx.com/digitalAssets/98/98491_nasdaq-clearing-as-a-counterparty-april-2015.pdf, (Retrieved 2016-01-20).

[27] Nelsen, Roger B. 2006. *An Introduction to Copulas*. 2nd. Ed. New York: Springer.

[28] Olsson, Jimmy. 2014. *SF2955 Computer Intensive Methods in Mathematical Statistics* [Exam and Solutions]. Department of mathematics, Royal Institute of Technology (KTH), Stockholm. Available at: https://www.math.kth.se/matstat/gru/sf2955/Exam/exam_SF2955_20140818.pdf and https://www.math.kth.se/matstat/gru/sf2955/Exam/exam_SF2955_20140818_solutions.pdf (Retrieved 2016-04-20).

[29] Peel, David and Mclachlan, Geoffrey J. 2000. *Robust mixture modelling using the t distributions*. Statistics and Computing, Vol. 10, No. 4, pp. 339-348.

[30] Piché, Robert. 2011. *Robust multivariate linear regression using the Student-t distribution*. Professor of Automation Science and Engineering at Tampere University of Technology. Available at: http://www.mathworks.com/matlabcentral/fileexchange/38212-mvsregress/content/mvsregress_package/mvsregress.m, (Retrieved 2016-01-17).

[31] Robert, Christian P. and Casella, George. 2004. *Monte Carlo Statistical Methods*. 2nd. Ed. New York: Springer.

[32] Roth, Michael. 2013. *On the Multivariate t Distribution*. Technical report from Automatic Control at Linköpings University. Available at: http://users.isy.liu.se/en/rt/roth/student.pdf, (Retrieved 2016-02-15).

[33] Lindgren, Georg, Rootzén, Holger and Sandsten, Maria. 2013. *Appendix A Some Probability and Statistics*. Stationary Stochastic Processes for Scientists and Engineers, Vol. 1, No. 1, pp. 261-270. CRC Press Inc. Available at: http://www.math.chalmers.se/~rootzen/highdimensional/SSP4SE-appA.pdf, (Retrieved 2016-02-03).

[34] Rubin, Donald B. 1976. *Inference and Missing Data*. Biometrika, Vol. 63, No. 3, pp. 581-592.

[35] Rubin, Donald B. 1987. *Multiple imputation for nonresponse in surveys*. Hoboken, NJ: Wiley.

[36] Shannon, Claude E. 1948. *A Mathematical Theory of Communication*. Bell System Technical Journal, Vol. 27, No. 3, pp. 379-423, 623-656.

[37] Walther, Bruno A. and Moore, Joslin L. 2005. *The concepts of bias, precision and accuracy, and their use in testing the performance of species richness estimators, with a literature review of estimator performance*. Ecography, Vol. 28, No. 6, pp. 815-829.

Appendices

Appendix A

The following example is taken from [28] and explicitly show how the posterior conditional distribution, used in the ECM algorithm for the case with one parameter, is derived.

Example A.1. A latent data model comprises N independent observations $(Y_i)_{i=1}^N$ of a random variable

$$Y \sim \text{Exp}(X) = x\mathrm{e}^{-xy},\ y \geq 0,$$

where the intensity X is unobserved and itself exponentially distributed, i.e.,

$$X \sim \text{Exp}(\lambda) = \lambda\mathrm{e}^{-\lambda x},\ x \geq 0.$$

The parameter λ is unknown. For all the observations $(Y_i)_{i=1}^N$ the corresponding latent variables $(X_i)_{i=1}^N$ are mutually independent (otherwise the observations would not be independent).

(a) Compute the conditional distribution of X given Y.

Answer: Write

$$f(x|y) \propto f(y,x) = f(y|x)f(x) = x\mathrm{e}^{-xy}\lambda\mathrm{e}^{-\lambda x} \propto x\mathrm{e}^{-x(y+\lambda)},$$

from which can be concluded that

$$X|Y = y \sim \mathit{Gamma}(2, y+\lambda),$$

where

$$\mathit{Gamma}(k,\mu) = \frac{\mu^k}{\Gamma(k)} x^{k-1}\mathrm{e}^{-\mu x},$$

and Γ is the gamma function.

(b) Derive the ECM-updating formula for the unknown parameter λ based on the N observations.

Answer: For this model, the complete data likelihood w.r.t. λ is, as the pairs $\{(X_i, Y_i)\}_{i=1}^N$ are independent, given by

$$f_\lambda(x_{1:N}, y_{1:N}) = \prod_{i=1}^N f_\lambda(x_i, y_i) = \lambda^N \mathrm{e}^{-\sum_{i=1}^N x_i(y_i+\lambda)} \prod_{i=1}^N x_i,$$

yielding the complete data log-likelihood

$$\log\big(f_\lambda(x_{1:N}, y_{1:N})\big) \overset{\text{c.}}{=} N\log(\lambda) - \lambda \sum_{i=1}^N x_i,$$

where $\overset{\text{c.}}{=}$ means that the equality holds up to a constant that does not depend on λ. This yields the ECM intermediate quantity

$$Q_{\lambda_m}(\lambda) = \mathbb{E}_{\lambda_m}\big[\log\big(f_\lambda(X_{1:N}, Y_{1:N})\big)|Y_{1:N}\big] \overset{\text{c.}}{=} N\log(\lambda) - \lambda \sum_{i=1}^N \mathbb{E}_{\lambda_m}\big[X_i|Y_i\big].$$

For $X \sim Gamma(k, \mu)$ it holds that $\mathbb{E}[X] = \frac{k}{\mu}$, which implies that

$$\mathbb{E}_m\left[X_i|Y_i\right] = \frac{2}{Y_i + \lambda_m},$$

providing

$$Q_{\lambda_m}(\lambda) \stackrel{c.}{=} N\log(\lambda) - 2\lambda \sum_{i=1}^{N} \frac{1}{Y_i + \lambda_m}.$$

Finally, in the M-step, $\mathcal{Q}_{\lambda_m}(\lambda)$ is maximized w.r.t. λ:

$$\frac{\mathrm{d}Q_{\lambda_m}(\lambda)}{\mathrm{d}\lambda} = \frac{N}{\lambda} - 2\sum_{i=1}^{N} \frac{1}{Y_i + \lambda_m} = 0 \Leftrightarrow \lambda = \frac{N}{2\sum_{i=1}^{N} \frac{1}{Y_i+\lambda_m}}.$$

This yields the ECM updating formula

$$\lambda_{m+1} = \frac{N}{2\sum_{i=1}^{N} \frac{1}{Y_i+\lambda_m}}$$

for λ.

Appendix B

The proof to Theorem 3.5.1 (*Sklar 1959*).

Proof. This proof is restricted to prove the existence and uniqueness of a copula in the case when $F_1, ..., F_d$ are continuous and the converse statement in its general form. For a full proof see Nelsen [27]. For any $x_1, ..., x_d$ in $\bar{\mathbb{R}} = [-\infty, \infty]$, if X has distribution function F, then $P(F(X) \leq F(x)) = P(X \leq x)$, hence

$$F(x_1, ..., x_d) = P\big(F_1(X_1) \leq F_1(x_1), ..., F_d(X_d) \leq F_d(x_d)\big).$$

Since $F_1, ..., F_d$ are continuous, Proposition 3.5.1 and Definition 3.5.1 imply that the distribution function of $\big(F_1(X_1), ..., F_d(X_d)\big)$ is a copula, which is denoted by C, and thus the identity in Equation (3.10) obtained. If Equation (3.10) is evaluated at the arguments $x_i = F_i^{\leftarrow}(u_i)$, $0 \leq u_i \leq 1$, $i = 1, ..., d$, and use $F_i\big(F_i^{\leftarrow}(u_i)\big) = u_i$, the following is obtained

$$C(u_1, ..., u_d) = F(F_1^{\leftarrow}(u_1), ..., F_d^{\leftarrow}(u_d)),$$

which gives an explicit representation of C in terms of F and its margins, and thus shows uniqueness. For the converse statement assume that C is a copula and that $F_1, ..., F_d$ are univariate distribution functions. A random vector with distribution function as in Equation (3.10) is then constructed by taking U to be a random vector with distribution function C and setting $X := \big(F_1^{\leftarrow}(U_1), ..., F_d^{\leftarrow}(U_d)\big)$. By using the fact that if F is right-continuous, i.e. $F_i(x_i) \geq y_i \iff F_i^{\leftarrow}(y_i) \leq x_i$, it is finally verified that

$$\begin{aligned} P(X_1 \leq x_1, ..., X_d \leq x_d) &= P(F_1^{\leftarrow}(U_1) \leq x_1, ..., F_d^{\leftarrow}(U_d) \leq x_d) \\ &= P(U_1 \leq F_1(x_1), ..., U_d \leq F_d(x_d)) \\ &= C(F_1(x_1), ..., F_d(x_d)). \end{aligned}$$

□

Appendix C

Proposition C.1. *Let* $\mathbf{W}(\mathbf{Y})$ *denote the vector with* i^{th} *component*

$$\mathbf{W}_i(\mathbf{Y}_i) = T_\nu^{-1}\big(F_i(\mathbf{Y}_i)\big),$$

and let $\left|\frac{\partial \mathbf{W}}{\partial \mathbf{Y}}\right|$ *denote the determinant of the Jacobian matrix associated with the transformation* (4.20)*. Then the conditional distribution on the raw variables is given by*

$$f_{\mathcal{Y}}(\mathbf{Y}_1|\mathbf{Y}_2) = f_{\mathcal{W}}\big(\mathbf{W}_1(\mathbf{Y}_1)|\mathbf{W}_2(\mathbf{Y}_2)\big)\left|\frac{\partial \mathbf{W}_1}{\partial \mathbf{Y}_1}\right|,$$

where

$$\frac{\partial \mathbf{W}_1}{\partial \mathbf{Y}_1}(y_1) = \frac{f_1(y_1)}{t_\nu\big(T_\nu^{-1}\big(F_1(y_1)\big)\big)}.$$

Given $f_{\mathcal{Y}}(\mathbf{Y}_1|\mathbf{Y}_2)$*, it is easy to extract the conditional distribution for any particular component, say the* j^{th} *component* $y_{1,j}$*, of* $\mathbf{Y_1}$*, given* $\mathbf{Y_2}$*. The distribution of* $y_{1,j}$*, given* $\mathbf{Y_2}$*, is then given by*

$$f_{\mathcal{Y}}(y_{1,j}|\mathbf{Y_2}) = f_{\mathcal{W}}\big(w_{1,j}(y_{1,j})|\mathbf{W}_2(\mathbf{Y}_2)\big)\left|\frac{f_1(y_{1,j})}{t_\nu\big(T_\nu^{-1}\big(F_1(y_{1,j})\big)\big)}\right|$$

For a full proof of Proposition C.1 please see [13] and [14].

Appendix D

In Table D.1-D.24 the values from the most volatile period in the out-of-sample set can be seen together with the conditional values and the corresponding estimated quantile values for $\alpha = \{0.001, 0.005, 0.995, 0.999\}$ for each method.

SR:normal, $\alpha = 0.999$									
				$N_{miss} = 0$			$N_{miss} = 250$		
Date	Monthly	Yearly	Y_{o-o-s}	Y_{pre}	$\|\Delta\|$	$\Delta\%$	Y_{pre}	$\|\Delta\|$	$\Delta\%$
2008-03-03	-7.9176	-2.4645	-10.3876	-3.2494	7.1382	0.6872	-3.5995	6.7881	0.6535
2008-02-29	-4.8392	0	-5.9767	-0.1792	5.7974	0.9700	-0.5358	5.4408	0.9103
2008-02-28	0.4615	0.3806	1.3294	2.9847	1.6553	-1.2451	2.4371	1.1077	-0.8333
2008-02-27	7.4380	2.5366	7.4603	8.0582	0.5979	-0.0801	7.3364	0.1239	0.0166
2008-02-26	-6.3467	-2.6591	-7.6246	-2.4807	5.1439	0.6746	-2.9017	4.7230	0.6194
2008-02-25	-7.7143	-0.5666	-8.2100	-2.0932	6.1168	0.7450	-2.3630	5.8469	0.7122

Table D.1: An overview of the values from the most volatile period in the out-of-sample set together with the conditional values, and the corresponding estimated quantile values for $\alpha = 0.999$ for the SR:normal method. The absolute error (distance) between the true value and the estimated value together with the percentage error is also presented for two different scenarios.

SR:normal, $\alpha = 0.995$									
				$N_{miss} = 0$			$N_{miss} = 250$		
Date	Monthly	Yearly	Y_{o-o-s}	Y_{pre}	$\|\Delta\|$	$\Delta\%$	Y_{pre}	$\|\Delta\|$	$\Delta\%$
2008-03-03	-7.9176	-2.4645	-10.3876	-3.6627	6.7249	0.6474	-3.9248	6.4628	0.6222
2008-02-29	-4.8392	0	-5.9767	-0.5925	5.3841	0.9009	-0.8612	5.1155	0.8559
2008-02-28	0.4615	0.3806	1.3294	2.5713	1.2420	-0.9342	2.1118	0.7824	-0.5886
2008-02-27	7.4380	2.5366	7.4603	7.6449	0.1846	-0.0247	7.0111	0.4492	0.0602
2008-02-26	-6.3467	-2.6591	-7.6246	-2.8940	4.7306	0.6204	-3.2270	4.3976	0.5768
2008-02-25	-7.7143	-0.5666	-8.2100	-2.5065	5.7034	0.6947	-2.6883	5.5216	0.6726

Table D.2: An overview of the values from the most volatile period in the out-of-sample set together with the conditional values, and the corresponding estimated quantile values for $\alpha = 0.995$ for the SR:normal method. The absolute error (distance) between the true value and the estimated value together with the percentage error is also presented for two different scenarios.

SR:normal, $\alpha = 0.005$									
				$N_{miss} = 0$			$N_{miss} = 250$		
Date	Monthly	Yearly	Y_{o-o-s}	Y_{pre}	$\|\Delta\|$	$\Delta\%$	Y_{pre}	$\|\Delta\|$	$\Delta\%$
2008-03-03	-7.9176	-2.4645	-10.3876	-7.8019	2.5857	0.2489	-7.1828	3.2048	0.3085
2008-02-29	-4.8392	0	-5.9767	-4.7318	1.2449	0.2083	-4.1191	1.8575	0.3108
2008-02-28	0.4615	0.3806	1.3294	-1.5679	2.8973	2.1794	-1.1462	2.4756	1.8622
2008-02-27	7.4380	2.5366	7.4603	3.5057	3.9546	0.5301	3.7531	3.7072	0.4969
2008-02-26	-6.3467	-2.6591	-7.6246	-7.0332	0.5914	0.0776	-6.4850	1.1397	0.1495
2008-02-25	-7.7143	-0.5666	-8.2100	-6.6457	1.5642	0.1905	-5.9463	2.2636	0.2757

Table D.3: An overview of the values from the most volatile period in the out-of-sample set together with the conditional values, and the corresponding estimated quantile values for $\alpha = 0.005$ for the SR:normal method. The absolute error (distance) between the true value and the estimated value together with the percentage error is also presented for two different scenarios.

SR:normal, $\alpha = 0.001$									
				$N_{miss} = 0$			$N_{miss} = 250$		
Date	Monthly	Yearly	Y_{o-o-s}	Y_{pre}	$\|\Delta\|$	$\Delta\%$	Y_{pre}	$\|\Delta\|$	$\Delta\%$
2008-03-03	-7.9176	-2.4645	-10.3876	-8.2152	2.1724	0.2091	-7.5081	2.8795	0.2772
2008-02-29	-4.8392	0	-5.9767	-5.1451	0.8316	0.1391	-4.4445	1.5322	0.2564
2008-02-28	0.4615	0.3806	1.3294	-1.9812	3.3106	2.4903	-1.4715	2.8009	2.1069
2008-02-27	7.4380	2.5366	7.4603	3.0924	4.3679	0.5855	3.4278	4.0325	0.5405
2008-02-26	-6.3467	-2.6591	-7.6246	-7.4465	0.1781	0.0234	-6.8103	0.8143	0.1068
2008-02-25	-7.7143	-0.5666	-8.2100	-7.0591	1.1509	0.1402	-6.2716	1.9383	0.2361

Table D.4: An overview of the values from the most volatile period in the out-of-sample set together with the conditional values, and the corresponding estimated quantile values for $\alpha = 0.001$ for the SR:normal method. The absolute error (distance) between the true value and the estimated value together with the percentage error is also presented for two different scenarios.

SR:Student's t, $\alpha = 0.999$									
				$N_{miss} = 0$			$N_{miss} = 250$		
Date	Monthly	Yearly	Y_{o-o-s}	Y_{pre}	$\|\Delta\|$	$\Delta\%$	Y_{pre}	$\|\Delta\|$	$\Delta\%$
2008-03-03	-7.9176	-2.4645	-10.3876	-0.7847	9.6029	0.9245	-3.3557	7.0319	0.6769
2008-02-29	-4.8392	0	-5.9767	2.2813	8.2580	1.3817	-0.3007	5.6759	0.9497
2008-02-28	0.4615	0.3806	1.3294	5.5690	4.2396	-3.1891	2.7361	1.4067	-1.0581
2008-02-27	7.4380	2.5366	7.4603	10.7493	3.2890	-0.4409	7.6856	0.2253	-0.0302
2008-02-26	-6.3467	-2.6591	-7.6246	0.0311	7.6557	1.0041	-2.6327	4.9919	0.6547
2008-02-25	-7.7143	-0.5666	-8.2100	0.3125	8.5224	1.0381	-2.1552	6.0548	0.7375

Table D.5: An overview of the values from the most volatile period in the out-of-sample set together with the conditional values, and the corresponding estimated quantile values for $\alpha = 0.999$ for the SR:Student's t method. The absolute error (distance) between the true value and the estimated value together with the percentage error is also presented for two different scenarios.

SR:Student's t, $\alpha = 0.995$									
				$N_{miss} = 0$			$N_{miss} = 250$		
Date	Monthly	Yearly	Y_{o-o-s}	Y_{pre}	$\|\Delta\|$	$\Delta\%$	Y_{pre}	$\|\Delta\|$	$\Delta\%$
2008-03-03	-7.9176	-2.4645	-10.3876	-2.9356	7.4520	0.7174	-3.8788	6.5088	0.6266
2008-02-29	-4.8392	0	-5.9767	0.1304	6.1071	1.0218	-0.8238	5.1529	0.8622
2008-02-28	0.4615	0.3806	1.3294	3.4181	2.0887	-1.5712	2.2130	0.8836	-0.6647
2008-02-27	7.4380	2.5366	7.4603	8.5984	1.1381	-0.1526	7.1626	0.2978	0.0399
2008-02-26	-6.3467	-2.6591	-7.6246	-2.1198	5.5048	0.7220	-3.1558	4.4688	0.5861
2008-02-25	-7.7143	-0.5666	-8.2100	-1.8384	6.3715	0.7761	-2.6783	5.5317	0.6738

Table D.6: An overview of the values from the most volatile period in the out-of-sample set together with the conditional values, and the corresponding estimated quantile values for $\alpha = 0.995$ for the SR:Student's t method. The absolute error (distance) between the true value and the estimated value together with the percentage error is also presented for two different scenarios.

SR:Student's t, $\alpha = 0.005$									
				$N_{miss} = 0$			$N_{miss} = 250$		
Date	Monthly	Yearly	Y_{o-o-s}	Y_{pre}	$\|\Delta\|$	$\Delta\%$	Y_{pre}	$\|\Delta\|$	$\Delta\%$
2008-03-03	-7.9176	-2.4645	-10.3876	-8.8089	1.5787	0.1520	-7.3425	3.0451	0.2931
2008-02-29	-4.8392	0	-5.9767	-5.7428	0.2338	0.0391	-4.2875	1.6891	0.2826
2008-02-28	0.4615	0.3806	1.3294	-2.4552	3.7846	2.8469	-1.2507	2.5801	1.9408
2008-02-27	7.4380	2.5366	7.4603	2.7251	4.7352	0.6347	3.6989	3.7615	0.5042
2008-02-26	-6.3467	-2.6591	-7.6246	-7.9931	0.3684	-0.0483	-6.6195	1.0051	0.1318
2008-02-25	-7.7143	-0.5666	-8.2100	-7.7117	0.4982	0.0607	-6.1420	2.0680	0.2519

Table D.7: An overview of the values from the most volatile period in the out-of-sample set together with the conditional values, and the corresponding estimated quantile values for $\alpha = 0.005$ for the SR:Student's t method. The absolute error (distance) between the true value and the estimated value together with the percentage error is also presented for two different scenarios.

SR:Student's t, $\alpha = 0.001$									
				$N_{miss} = 0$			$N_{miss} = 250$		
Date	Monthly	Yearly	Y_{o-o-s}	Y_{pre}	$\|\Delta\|$	$\Delta\%$	Y_{pre}	$\|\Delta\|$	$\Delta\%$
2008-03-03	-7.9176	-2.4645	-10.3876	-10.9598	0.5722	-0.0551	-7.8656	2.5220	0.2428
2008-02-29	-4.8392	0	-5.9767	-7.8937	1.9171	-0.3208	-4.8106	1.1661	0.1951
2008-02-28	0.4615	0.3806	1.3294	-4.6061	5.9355	4.4648	-1.7738	3.1032	2.3343
2008-02-27	7.4380	2.5366	7.4603	0.5742	6.8861	0.9230	3.1758	4.2845	0.5743
2008-02-26	-6.3467	-2.6591	-7.6246	-10.1440	2.5193	-0.3304	-7.1426	0.4820	0.0632
2008-02-25	-7.7143	-0.5666	-8.2100	-9.8626	1.6527	-0.2013	-6.6651	1.5449	0.1882

Table D.8: An overview of the values from the most volatile period in the out-of-sample set together with the conditional values, and the corresponding estimated quantile values for $\alpha = 0.001$ for the SR:Student's t method. The absolute error (distance) between the true value and the estimated value together with the percentage error is also presented for two different scenarios.

MI:normal, $\alpha = 0.999$									
				$N_{miss} = 0$			$N_{miss} = 250$		
Date	Monthly	Yearly	Y_{o-o-s}	Y_{pre}	$\|\Delta\|$	$\Delta\%$	Y_{pre}	$\|\Delta\|$	$\Delta\%$
2008-03-03	-7.9176	-2.4645	-10.3876	-3.2495	7.1381	0.6872	-3.5800	6.8076	0.6554
2008-02-29	-4.8392	0	-5.9767	-0.1810	5.7957	0.9697	-0.5149	5.4618	0.9139
2008-02-28	0.4615	0.3806	1.3294	2.9865	1.6571	-1.2465	2.4420	1.1126	-0.8369
2008-02-27	7.4380	2.5366	7.4603	8.0620	0.6017	-0.0807	7.3281	0.1322	0.0177
2008-02-26	-6.3467	-2.6591	-7.6246	-2.4792	5.1455	0.6748	-2.8884	4.7362	0.6212
2008-02-25	-7.7143	-0.5666	-8.2100	-2.0963	6.1137	0.7447	-2.3351	5.8749	0.7156

Table D.9: An overview of the values from the most volatile period in the out-of-sample set together with the conditional values, and the corresponding estimated quantile values for $\alpha = 0.999$ for the MI:normal method. The absolute error (distance) between the true value and the estimated value together with the percentage error is also presented for two different scenarios.

MI:normal, $\alpha = 0.995$									
				$N_{miss} = 0$			$N_{miss} = 250$		
Date	Monthly	Yearly	Y_{o-o-s}	Y_{pre}	$\|\Delta\|$	$\Delta\%$	Y_{pre}	$\|\Delta\|$	$\Delta\%$
2008-03-03	-7.9176	-2.4645	-10.3876	-3.6621	6.7255	0.6475	-3.8955	6.4921	0.6250
2008-02-29	-4.8392	0	-5.9767	-0.5880	5.3886	0.9016	-0.8298	5.1469	0.8612
2008-02-28	0.4615	0.3806	1.3294	2.5753	1.2459	-0.9372	2.1213	0.7919	-0.5957
2008-02-27	7.4380	2.5366	7.4603	7.6512	0.1909	-0.0256	7.0026	0.4577	0.0613
2008-02-26	-6.3467	-2.6591	-7.6246	-2.8942	4.7304	0.6204	-3.2063	4.4184	0.5795
2008-02-25	-7.7143	-0.5666	-8.2100	-2.5024	5.7076	0.6952	-2.6474	5.5625	0.6775

Table D.10: An overview of the values from the most volatile period in the out-of-sample set together with the conditional values, and the corresponding estimated quantile values for $\alpha = 0.995$ for the MI:normal method. The absolute error (distance) between the true value and the estimated value together with the percentage error is also presented for two different scenarios.

MI:normal, $\alpha = 0.005$									
				$N_{miss} = 0$			$N_{miss} = 250$		
Date	Monthly	Yearly	Y_{o-o-s}	Y_{pre}	$\|\Delta\|$	$\Delta\%$	Y_{pre}	$\|\Delta\|$	$\Delta\%$
2008-03-03	-7.9176	-2.4645	-10.3876	-7.8047	2.5829	0.2487	-7.1525	3.2351	0.3114
2008-02-29	-4.8392	0	-5.9767	-4.7424	1.2343	0.2065	-4.0895	1.8872	0.3158
2008-02-28	0.4615	0.3806	1.3294	-1.5728	2.9022	2.1831	-1.1354	2.4648	1.8541
2008-02-27	7.4380	2.5366	7.4603	3.5000	3.9603	0.5308	3.7466	3.7137	0.4978
2008-02-26	-6.3467	-2.6591	-7.6246	-7.0328	0.5918	0.0776	-6.4617	1.1629	0.1525
2008-02-25	-7.7143	-0.5666	-8.2100	-6.6576	1.5524	0.1891	-5.9080	2.3020	0.2804

Table D.11: An overview of the values from the most volatile period in the out-of-sample set together with the conditional values, and the corresponding estimated quantile values for $\alpha = 0.005$ for the MI:normal method. The absolute error (distance) between the true value and the estimated value together with the percentage error is also presented for two different scenarios.

MI:normal, $\alpha = 0.001$									
				$N_{miss} = 0$			$N_{miss} = 250$		
Date	Monthly	Yearly	Y_{o-o-s}	Y_{pre}	$\|\Delta\|$	$\Delta\%$	Y_{pre}	$\|\Delta\|$	$\Delta\%$
2008-03-03	-7.9176	-2.4645	-10.3876	-8.2223	2.1652	0.2084	-7.4848	2.9028	0.2794
2008-02-29	-4.8392	0	-5.9767	-5.1545	0.8222	0.1376	-4.4170	1.5597	0.2610
2008-02-28	0.4615	0.3806	1.3294	-1.9843	3.3137	2.4927	-1.4638	2.7932	2.1011
2008-02-27	7.4380	2.5366	7.4603	3.0931	4.3672	0.5854	3.4210	4.0393	0.5414
2008-02-26	-6.3467	-2.6591	-7.6246	-7.4510	0.1737	0.0228	-6.7950	0.8296	0.1088
2008-02-25	-7.7143	-0.5666	-8.2100	-7.0709	1.1390	0.1387	-6.2359	1.9740	0.2404

Table D.12: An overview of the values from the most volatile period in the out-of-sample set together with the conditional values, and the corresponding estimated quantile values for $\alpha = 0.001$ for the MI:normal method. The absolute error (distance) between the true value and the estimated value together with the percentage error is also presented for two different scenarios.

MI:Student's t, $\alpha = 0.999$									
				$N_{miss} = 0$			$N_{miss} = 250$		
Date	Monthly	Yearly	Y_{o-o-s}	Y_{pre}	$\|\Delta\|$	$\Delta\%$	Y_{pre}	$\|\Delta\|$	$\Delta\%$
2008-03-03	-7.9176	-2.4645	-10.3876	-0.7908	9.5968	0.9239	-3.3153	7.0723	0.6808
2008-02-29	-4.8392	0	-5.9767	2.2760	8.2527	1.3808	-0.3314	5.6452	0.9445
2008-02-28	0.4615	0.3806	1.3294	5.5668	4.2374	-3.1875	2.7279	1.3985	-1.0520
2008-02-27	7.4380	2.5366	7.4603	3.0931	4.3672	0.5854	3.4210	4.0393	0.5414
2008-02-26	-6.3467	-2.6591	-7.6246	0.0261	7.6507	1.0034	-2.5741	5.0505	0.6624
2008-02-25	-7.7143	-0.5666	-8.2100	0.3056	8.5155	1.0372	-2.1846	6.0254	0.7339

Table D.13: An overview of the values from the most volatile period in the out-of-sample set together with the conditional values, and the corresponding estimated quantile values for $\alpha = 0.999$ for the MI:Student's t method. The absolute error (distance) between the true value and the estimated value together with the percentage error is also presented for two different scenarios.

MI:Student's t, $\alpha = 0.995$									
				$N_{miss} = 0$			$N_{miss} = 250$		
Date	Monthly	Yearly	Y_{o-o-s}	Y_{pre}	$\|\Delta\|$	$\Delta\%$	Y_{pre}	$\|\Delta\|$	$\Delta\%$
2008-03-03	-7.9176	-2.4645	-10.3876	-2.9315	7.4561	0.7178	-3.8397	6.5479	0.6304
2008-02-29	-4.8392	0	-5.9767	0.1342	6.1108	1.0224	-0.8546	5.1221	0.8570
2008-02-28	0.4615	0.3806	1.3294	3.4165	2.0871	-1.5700	2.1993	0.8699	-0.6543
2008-02-27	7.4380	2.5366	7.4603	8.5919	1.1316	-0.1517	7.1124	0.3479	0.0466
2008-02-26	-6.3467	-2.6591	-7.6246	-2.1177	5.5070	0.7223	-3.1008	4.5239	0.5933
2008-02-25	-7.7143	-0.5666	-8.2100	-1.8323	6.3777	0.7768	-2.7055	5.5044	0.6705

Table D.14: An overview of the values from the most volatile period in the out-of-sample set together with the conditional values, and the corresponding estimated quantile values for $\alpha = 0.995$ for the MI:Student's t method. The absolute error (distance) between the true value and the estimated value together with the percentage error is also presented for two different scenarios.

MI:Student's t, $\alpha = 0.005$									
				$N_{miss} = 0$			$N_{miss} = 250$		
Date	Monthly	Yearly	Y_{o-o-s}	Y_{pre}	$\|\Delta\|$	$\Delta\%$	Y_{pre}	$\|\Delta\|$	$\Delta\%$
2008-03-03	-7.9176	-2.4645	-10.3876	-8.8020	1.5856	0.1526	-7.3144	3.0732	0.2959
2008-02-29	-4.8392	0	-5.9767	-5.7352	0.2415	0.0404	-4.3320	1.6447	0.2752
2008-02-28	0.4615	0.3806	1.3294	-2.4542	3.7836	2.8461	-1.2776	2.6070	1.9611
2008-02-27	7.4380	2.5366	7.4603	2.7207	4.7396	0.6353	3.6339	3.8264	0.5129
2008-02-26	-6.3467	-2.6591	-7.6246	-7.9888	0.3642	-0.0478	-6.5748	1.0498	0.1377
2008-02-25	-7.7143	-0.5666	-8.2100	-7.7012	0.5087	0.0620	-6.1826	2.0273	0.2469

Table D.15: An overview of the values from the most volatile period in the out-of-sample set together with the conditional values, and the corresponding estimated quantile values for $\alpha = 0.005$ for the MI:Student's t method. The absolute error (distance) between the true value and the estimated value together with the percentage error is also presented for two different scenarios.

MI:Student's t, $\alpha = 0.001$									
				$N_{miss} = 0$			$N_{miss} = 250$		
Date	Monthly	Yearly	Y_{o-o-s}	Y_{pre}	$\|\Delta\|$	$\Delta\%$	Y_{pre}	$\|\Delta\|$	$\Delta\%$
2008-03-03	-7.9176	-2.4645	-10.3876	-10.9567	0.5691	-0.0548	-7.8535	2.5341	0.2440
2008-02-29	-4.8392	0	-5.9767	-7.8957	1.9190	-0.3211	-4.8696	1.1070	0.1852
2008-02-28	0.4615	0.3806	1.3294	-4.6070	5.9364	4.4655	-1.8041	3.1335	2.3571
2008-02-27	7.4380	2.5366	7.4603	0.5706	6.8897	0.9235	3.1185	4.3418	0.5820
2008-02-26	-6.3467	-2.6591	-7.6246	-10.1399	2.5152	-0.3299	-7.1100	0.5147	0.0675
2008-02-25	-7.7143	-0.5666	-8.2100	-9.8642	1.6542	-0.2015	-6.7256	1.4844	0.1808

Table D.16: An overview of the values from the most volatile period in the out-of-sample set together with the conditional values, and the corresponding estimated quantile values for $\alpha = 0.001$ for the MI:Student's t method. The absolute error (distance) between the true value and the estimated value together with the percentage error is also presented for two different scenarios.

g-cop, $\alpha = 0.999$									
				$N_{miss} = 0$			$N_{miss} = 250$		
Date	Monthly	Yearly	Y_{o-o-s}	Y_{pre}	$\|\Delta\|$	$\Delta\%$	Y_{pre}	$\|\Delta\|$	$\Delta\%$
2008-03-03	-7.9176	-2.4645	-10.3876	-3.0274	7.3602	0.7086	-3.3603	7.0273	0.6765
2008-02-29	-4.8392	0	-5.9767	-0.3955	5.5811	0.9338	-0.6462	5.3305	0.8919
2008-02-28	0.4615	0.3806	1.3294	2.9383	1.6089	1.2103	2.3652	1.0358	0.7791
2008-02-27	7.4380	2.5366	7.4603	9.1939	1.7336	0.2324	8.2151	0.7548	0.1012
2008-02-26	-6.3467	-2.6591	-7.6246	-2.4748	5.1498	0.6754	-2.8218	4.8028	0.6299
2008-02-25	-7.7143	-0.5666	-8.2100	-1.9276	6.2823	0.7652	-2.1309	6.0791	0.7405

Table D.17: An overview of the values from the most volatile period in the out-of-sample set together with the conditional values, and the corresponding estimated quantile values for $\alpha = 0.999$ for the g-copula method. The absolute error (distance) between the true value and the estimated value together with the percentage error is also presented for two different scenarios.

g-cop, $\alpha = 0.995$									
				$N_{miss} = 0$			$N_{miss} = 250$		
Date	Monthly	Yearly	Y_{o-o-s}	Y_{pre}	$\lvert\Delta\rvert$	$\Delta\%$	Y_{pre}	$\lvert\Delta\rvert$	$\Delta\%$
2008-03-03	-7.9176	-2.4645	-10.3876	-3.4290	6.9586	0.6699	-3.7541	6.6335	0.6386
2008-02-29	-4.8392	0	-5.9767	-0.7563	5.2204	0.8735	-0.9942	4.9825	0.8337
2008-02-28	0.4615	0.3806	1.3294	2.5219	1.1925	0.8970	2.0280	0.6986	0.5255
2008-02-27	7.4380	2.5366	7.4603	8.4985	1.0382	0.1392	7.6473	0.1870	0.0251
2008-02-26	-6.3467	-2.6591	-7.6246	-2.8642	4.7604	0.6243	-3.1820	4.4427	0.5827
2008-02-25	-7.7143	-0.5666	-8.2100	-2.3016	5.9083	0.7197	-2.5178	5.6922	0.6933

Table D.18: An overview of the values from the most volatile period in the out-of-sample set together with the conditional values, and the corresponding estimated quantile values for $\alpha = 0.995$ for the g-copula method. The absolute error (distance) between the true value and the estimated value together with the percentage error is also presented for two different scenarios.

g-cop, $\alpha = 0.005$									
				$N_{miss} = 0$			$N_{miss} = 250$		
Date	Monthly	Yearly	Y_{o-o-s}	Y_{pre}	$\lvert\Delta\rvert$	$\Delta\%$	Y_{pre}	$\lvert\Delta\rvert$	$\Delta\%$
2008-03-03	-7.9176	-2.4645	-10.3876	-8.6685	1.7191	0.1655	-8.2964	2.0912	0.2013
2008-02-29	-4.8392	0	-5.9767	-4.8883	1.0884	0.1821	-4.3007	1.6760	0.2804
2008-02-28	0.4615	0.3806	1.3294	-1.3326	2.6619	2.0024	-0.9202	2.2496	1.6922
2008-02-27	7.4380	2.5366	7.4603	3.1298	4.3305	0.5805	3.3406	4.1197	0.5522
2008-02-26	-6.3467	-2.6591	-7.6246	-7.7988	0.1741	0.0228	-7.4075	0.2171	0.028
2008-02-25	-7.7143	-0.5666	-8.2100	-6.9938	1.2161	0.1481	-6.4244	1.7856	0.2175

Table D.19: An overview of the values from the most volatile period in the out-of-sample set together with the conditional values, and the corresponding estimated quantile values for $\alpha = 0.005$ for the g-copula method. The absolute error (distance) between the true value and the estimated value together with the percentage error is also presented for two different scenarios.

g-cop, $\alpha = 0.001$									
				$N_{miss} = 0$			$N_{miss} = 250$		
Date	Monthly	Yearly	Y_{o-o-s}	Y_{pre}	$\lvert\Delta\rvert$	$\Delta\%$	Y_{pre}	$\lvert\Delta\rvert$	$\Delta\%$
2008-03-03	-7.9176	-2.4645	-10.3876	-9.2252	1.1624	0.1119	-8.7791	1.6085	0.1549
2008-02-29	-4.8392	0	-5.9767	-5.3029	0.6737	0.1127	-4.6234	1.3532	0.2264
2008-02-28	0.4615	0.3806	1.3294	-1.6993	3.0287	2.2783	-1.1991	2.5285	1.9020
2008-02-27	7.4380	2.5366	7.4603	2.6755	4.7849	0.6414	2.9850	4.4753	0.5999
2008-02-26	-6.3467	-2.6591	-7.6246	-8.3305	0.7058	0.0926	-7.8680	0.2433	0.0319
2008-02-25	-7.7143	-0.5666	-8.2100	-7.4705	0.7395	0.0901	-6.8082	1.4017	0.1707

Table D.20: An overview of the values from the most volatile period in the out-of-sample set together with the conditional values, and the corresponding estimated quantile values for $\alpha = 0.001$ for the g-copula method. The absolute error (distance) between the true value and the estimated value together with the percentage error is also presented for two different scenarios.

t-cop, $\alpha = 0.999$									
				$N_{miss} = 0$			$N_{miss} = 250$		
Date	Monthly	Yearly	Y_{o-o-s}	Y_{pre}	$\|\Delta\|$	$\Delta\%$	Y_{pre}	$\|\Delta\|$	$\Delta\%$
2008-03-03	-7.9176	-2.4645	-10.3876	-3.4317	6.9559	0.6696	-2.0144	8.3732	0.8061
2008-02-29	-4.8392	0	-5.9767	-0.5387	5.4379	0.9099	0.8110	6.7877	1.1357
2008-02-28	0.4615	0.3806	1.3294	2.3205	0.9911	0.7455	3.1877	1.8583	1.3978
2008-02-27	7.4380	2.5366	7.4603	7.4587	0.0016	0.0002	7.6590	0.1987	0.0266
2008-02-26	-6.3467	-2.6591	-7.6246	-2.8736	4.7510	0.6231	-1.5943	6.0303	0.7909
2008-02-25	-7.7143	-0.5666	-8.2100	-2.1065	6.1035	0.7434	-0.4631	7.7468	0.9436

Table D.21: An overview of the values from the most volatile period in the out-of-sample set together with the conditional values, and the corresponding estimated quantile values for $\alpha = 0.999$ for the t-copula method. The absolute error (distance) between the true value and the estimated value together with the percentage error is also presented for two different scenarios.

t-cop, $\alpha = 0.995$									
				$N_{miss} = 0$			$N_{miss} = 250$		
Date	Monthly	Yearly	Y_{o-o-s}	Y_{pre}	$\|\Delta\|$	$\Delta\%$	Y_{pre}	$\|\Delta\|$	$\Delta\%$
2008-03-03	-7.9176	-2.4645	-10.3876	-4.0660	6.3216	0.6086	-2.9376	7.4500	0.7172
2008-02-29	-4.8392	0	-5.9767	-1.1605	4.8162	0.8058	-0.0045	5.9722	0.9992
2008-02-28	0.4615	0.3806	1.3294	1.8980	0.5686	0.4277	2.5748	1.2454	0.9369
2008-02-27	7.4380	2.5366	7.4603	7.0302	0.4301	0.0577	7.0842	0.3761	0.0504
2008-02-26	-6.3467	-2.6591	-7.6246	-3.4305	4.1942	0.5501	-2.4498	5.1748	0.6787
2008-02-25	-7.7143	-0.5666	-8.2100	-2.8775	5.3325	0.6495	-1.4824	6.7275	0.8194

Table D.22: An overview of the values from the most volatile period in the out-of-sample set together with the conditional values, and the corresponding estimated quantile values for $\alpha = 0.995$ for the t-copula method. The absolute error (distance) between the true value and the estimated value together with the percentage error is also presented for two different scenarios.

t-cop, $\alpha = 0.005$									
				$N_{miss} = 0$			$N_{miss} = 250$		
Date	Monthly	Yearly	Y_{o-o-s}	Y_{pre}	$\|\Delta\|$	$\Delta\%$	Y_{pre}	$\|\Delta\|$	$\Delta\%$
2008-03-03	-7.9176	-2.4645	-10.3876	-7.2424	3.1452	0.3028	-7.5755	2.8121	0.2707
2008-02-29	-4.8392	0	-5.9767	-4.5235	1.4531	0.2431	-4.9661	1.0106	0.1691
2008-02-28	0.4615	0.3806	1.3294	-0.8312	2.1605	1.6252	-1.4892	2.8186	2.1202
2008-02-27	7.4380	2.5366	7.4603	3.7367	3.7236	0.4991	2.7895	4.6708	0.6261
2008-02-26	-6.3467	-2.6591	-7.6246	-6.3278	1.2969	0.1701	-6.7285	0.8962	0.1175
2008-02-25	-7.7143	-0.5666	-8.2100	-6.6755	1.5345	0.1869	-7.0169	1.1930	0.1453

Table D.23: An overview of the values from the most volatile period in the out-of-sample set together with the conditional values, and the corresponding estimated quantile values for $\alpha = 0.005$ for the t-copula method. The absolute error (distance) between the true value and the estimated value together with the percentage error is also presented for two different scenarios.

t-cop, $\alpha = 0.001$									
				$N_{miss} = 0$			$N_{miss} = 250$		
Date	Monthly	Yearly	Y_{o-o-s}	Y_{pre}	$\|\Delta\|$	$\Delta\%$	Y_{pre}	$\|\Delta\|$	$\Delta\%$
2008-03-03	-7.9176	-2.4645	-10.3876	-7.7172	2.6704	0.2571	-8.1221	2.2655	0.2181
2008-02-29	-4.8392	0	-5.9767	-5.0440	0.9326	0.1560	-5.6073	0.3694	0.0618
2008-02-28	0.4615	0.3806	1.3294	-1.3195	2.6489	1.9926	-2.1577	3.4871	2.6231
2008-02-27	7.4380	2.5366	7.4603	3.0303	4.4300	0.5938	1.8021	5.6582	0.7584
2008-02-26	-6.3467	-2.6591	-7.6246	-6.7714	0.8532	0.1119	-7.2429	0.3817	0.0501
2008-02-25	-7.7143	-0.5666	-8.2100	-7.2278	0.9822	0.1196	-7.6668	0.5432	0.0662

Table D.24: An overview of the values from the most volatile period in the out-of-sample set together with the conditional values, and the corresponding estimated quantile values for $\alpha = 0.001$ for the t-copula method. The absolute error (distance) between the true value and the estimated value together with the percentage error is also presented for two different scenarios.

Appendix E

In Figure E.1-E.2 and in Table E.1-E.2 the calculated MAE and RMSE for each method can be seen when the models are applied on more stable data. Since the MAE and the RMSE is calculated after generating predicted values from the corresponding conditional distribution, and both is very sensitive when used on small samples (typically $N_{miss} < 100$), these error metrics have been averaged over 100 iterations for a more distinguishable interpretation.

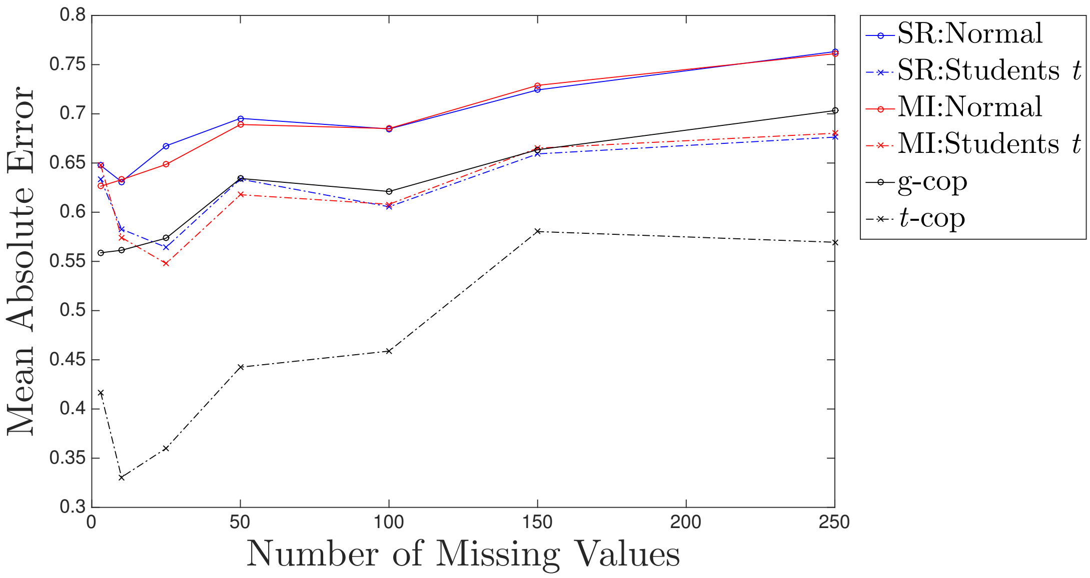

Figure E.1: The Mean Absolute Error (MAE) for each model with different amount of missing data (NaN).

Mean Absolute Error (MAE)							
NaN	3	10	25	50	100	150	250
SR:normal	0.6475	0.6310	0.6675	0.6954	0.6846	0.7243	0.7632
SR:Student's t	0.6337	0.5828	0.5643	0.6336	0.6056	0.6593	0.6763
MI:normal	0.6264	0.6331	0.6486	0.6891	0.6851	0.7289	0.7612
MI:Student's t	0.6479	0.5743	0.5479	0.6178	0.6081	0.6651	0.6802
g-cop	0.5588	0.5613	0.5736	0.6343	0.6211	0.6635	0.7033
t-cop	0.4169	0.3306	0.3600	0.4425	0.4588	0.5803	0.5695

Table E.1: The Mean Absolute Error (MAE) for each model with different amount of missing data (NaN).

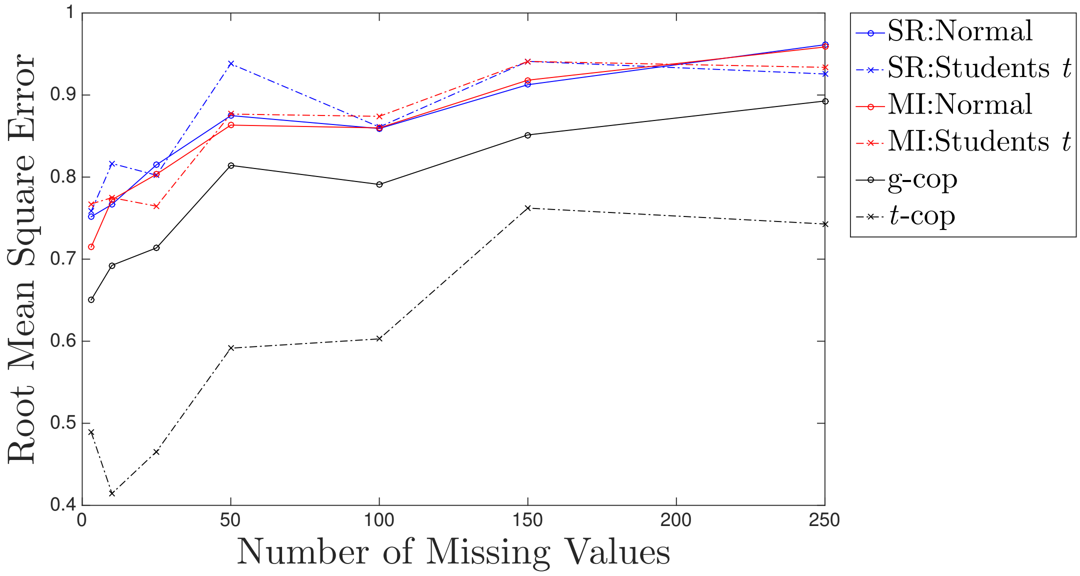

Figure E.2: The Root Mean Square Error (RMSE) for each model with different amount of missing data (NaN).

Root Mean Square Error (RMSE)							
NaN	3	10	25	50	100	150	250
SR:normal	0.7517	0.7667	0.8150	0.8749	0.8592	0.9126	0.9613
SR:Student's t	0.7583	0.8166	0.8023	0.9380	0.8611	0.9410	0.9257
MI:normal	0.7148	0.7724	0.8035	0.8634	0.8600	0.9181	0.9586
MI:Student's t	0.7673	0.7749	0.7646	0.8768	0.8740	0.9410	0.9337
g-cop	0.6504	0.6925	0.7138	0.8142	0.7911	0.8514	0.8927
t-cop	0.4893	0.4144	0.4655	0.5915	0.6028	0.7622	0.7428

Table E.2: The Root Mean Square Error (RMSE) for each model with different amount of missing data (NaN).

The out-of-sample log-likelihood for each model with different amount of missing data can be seen in Table E.3.

Out-of-sample log-likelihood							
NaN	3	10	25	50	100	150	250
SR:normal	-2.2510	-7.0579	-17.8985	-39.8011	-77.6194	-127.7380	-227.6115
SR:Student's t	-1.6950	-3.8313	-9.9953	-29.2104	-53.2038	-101.5000	-192.0945
MI:normal	-2.2520	-7.0597	-17.8979	-39.8081	-77.5979	-127.7068	-227.6826
MI:Student's t	-1.6967	-3.8322	-10.0025	-29.3128	-53.2682	-101.3712	-192.1108
g-cop	-1.9771	-5.5584	-14.1795	-34.2358	-64.3550	-110.2361	-201.2402
t-cop	-1.2434	-0.9233	-4.0002	-20.4623	-30.6634	-80.5289	-177.2783

Table E.3: The out-of-sample log-likelihood for each model with different amount of missing data (NaN).

The performance, based on these metrics, shows that the alternative t-copula method outperforms the more established methods when the models are applied on less volatile data. The distribution of the error is less bias and has lower variance compared to the more established methods (see Figure E.1 and E.2). Additionally, the out-of-sample log-likelihood is greater (less negative), for all cases of missing data, which indicates that the alternative t-copula model conforms much more efficiently to the out of sample data (see Table E.3).